# ds4_fusarium_oxysporum_limpieza_de_datos

February 1, 2021

Limpieza de datos

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

## 1 Declaración de variables

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

         X0     X1     X2     X3     X4     X5      X6      X7     X8     X9  \
0     8.889  5.051  4.040  7.677  1.414  8.283   4.646   6.465  2.424  5.455
1     8.945  3.399  3.220  2.683  1.073  4.293   2.504   9.481  2.147  7.335
2     6.949  7.553  3.323  7.855  0.000  7.855   3.625   7.553  2.115  4.834
3     8.067  2.086  3.755  6.676  1.530  5.007   3.477   7.510  0.695  3.616
4     5.898  4.960  3.887  6.166  2.547  4.692   4.155   7.507  2.011  4.960
..      …      …      …      …      …      …       …       …      …      …
995   8.754  8.754  4.377  4.040  1.347  5.724   4.377   4.377  1.852  5.556
996   9.524  9.524  2.381  7.143  4.762  2.381  11.905  11.905  2.381  4.762
997   5.634  4.225  2.113  6.338  1.408  7.746   2.113  10.563  2.113  8.451
998   7.991  4.566  5.708  5.479  2.283  5.708   3.196   7.078  4.795  4.566
999  15.556  7.407  5.185  5.185  0.000  8.148   2.222   8.148  5.926  4.444
```

```
     …    X11    X12    X13     X14     X15     X16    X17    X18    X19  \
0    …  5.051  2.222  3.434   4.040   5.455   3.030  3.232  3.636  4.040
1    …  3.041  2.504  5.009   5.546   6.977   5.903  2.862  4.293  7.513
2    …  3.323  1.511  3.323   6.949   9.668   3.625  1.208  2.115  8.459
3    …  1.391  0.834  4.033   6.259  13.908  14.186  0.974  3.894  6.815
4    …  3.351  2.011  4.692   6.032   8.847   6.702  1.877  4.826  5.496
..   …    …    …    …     …       …       …    …    …    …
995  …  2.694  1.515  5.556   4.714   6.734   5.724  1.684  2.694  7.407
996  …  2.381  2.381  2.381  11.905   7.143   2.381  0.000  0.000  2.381
997  …  6.338  2.113  4.225   4.930   7.042   7.746  0.704  2.817  4.930
998  …  6.393  3.196  6.393   4.795   5.936   4.566  2.055  4.338  4.566
999  …  0.741  3.704  1.481   2.963   5.926   2.222  2.222  0.000  6.667

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

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.268801     5.900585     3.790985     5.649229     1.426726
std       2.338740     2.249480     1.501018     1.912915     1.172710
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.872750     4.576000     2.948500     4.516000     0.741750
50%       8.060000     5.733000     3.666000     5.726500     1.222000
75%       9.386500     7.131500     4.484250     6.697250     1.861250
max      25.641000    17.689000    25.574000    20.690000    12.500000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.128030     4.000106     6.821265     2.508368     5.252795
std       2.318312     1.840948     2.331634     1.272048     1.962849
```

| | | | | | |
|---|---|---|---|---|---|
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 4.622750 | 2.915250 | 5.234500 | 1.691000 | 4.027500 |
| 50% | 5.948500 | 3.721500 | 6.738000 | 2.415000 | 5.112000 |
| 75% | 7.325500 | 4.751250 | 8.173000 | 3.226000 | 6.223250 |
| max | 17.241000 | 23.967000 | 23.077000 | 14.897000 | 21.053000 |

| | X10 | X11 | X12 | X13 | X14 \ |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 8.860051 | 4.905241 | 2.348315 | 3.818471 | 5.705522 |
| std | 2.582738 | 2.166728 | 1.056134 | 1.455084 | 2.295659 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 7.281500 | 3.484000 | 1.661000 | 2.822750 | 4.321750 |
| 50% | 8.900500 | 4.670000 | 2.207000 | 3.849000 | 5.405000 |
| 75% | 10.526000 | 5.882000 | 2.875000 | 4.728500 | 6.805500 |
| max | 18.103000 | 18.803000 | 8.929000 | 9.804000 | 25.439000 |

| | X15 | X16 | X17 | X18 | X19 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 7.894848 | 6.037411 | 1.653739 | 2.925043 | 6.104484 |
| std | 2.557917 | 2.154330 | 1.021725 | 1.308150 | 1.834678 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 6.274750 | 4.865500 | 0.932000 | 2.054500 | 4.887000 |
| 50% | 7.629500 | 5.822500 | 1.537000 | 2.791000 | 6.049000 |
| 75% | 9.265000 | 6.858000 | 2.277250 | 3.681250 | 7.207500 |
| max | 21.260000 | 32.742000 | 6.000000 | 9.483000 | 13.653000 |

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

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 \ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 5.946 | 10.270 | 2.162 | 2.162 | 5.405 | 4.324 | 2.162 | 4.865 | 3.243 | 5.405 |
| 1 | 7.636 | 6.909 | 3.818 | 6.727 | 2.909 | 4.364 | 3.818 | 4.364 | 2.182 | 6.364 |
| 2 | 5.797 | 10.145 | 2.899 | 1.449 | 1.449 | 2.899 | 4.348 | 5.797 | 1.449 | 11.594 |
| 3 | 9.428 | 4.209 | 5.387 | 5.051 | 2.020 | 5.051 | 3.030 | 8.586 | 2.694 | 5.387 |
| 4 | 9.500 | 4.750 | 3.250 | 5.250 | 1.500 | 3.000 | 3.250 | 7.250 | 2.000 | 7.750 |
| .. | … | … | … | … | … | … | … | … | … | |
| 995 | 10.779 | 6.799 | 3.648 | 4.478 | 1.658 | 5.638 | 3.814 | 6.302 | 1.327 | 4.478 |
| 996 | 9.412 | 5.359 | 3.791 | 4.052 | 0.784 | 4.575 | 3.529 | 8.758 | 1.699 | 6.667 |
| 997 | 6.993 | 5.455 | 5.315 | 6.154 | 0.280 | 4.755 | 2.517 | 7.413 | 3.217 | 4.615 |
| 998 | 7.923 | 4.069 | 6.852 | 6.424 | 0.428 | 1.713 | 5.353 | 10.278 | 2.355 | 3.426 |
| 999 | 12.454 | 5.128 | 2.930 | 6.777 | 1.099 | 5.861 | 4.396 | 6.410 | 2.381 | 4.762 |

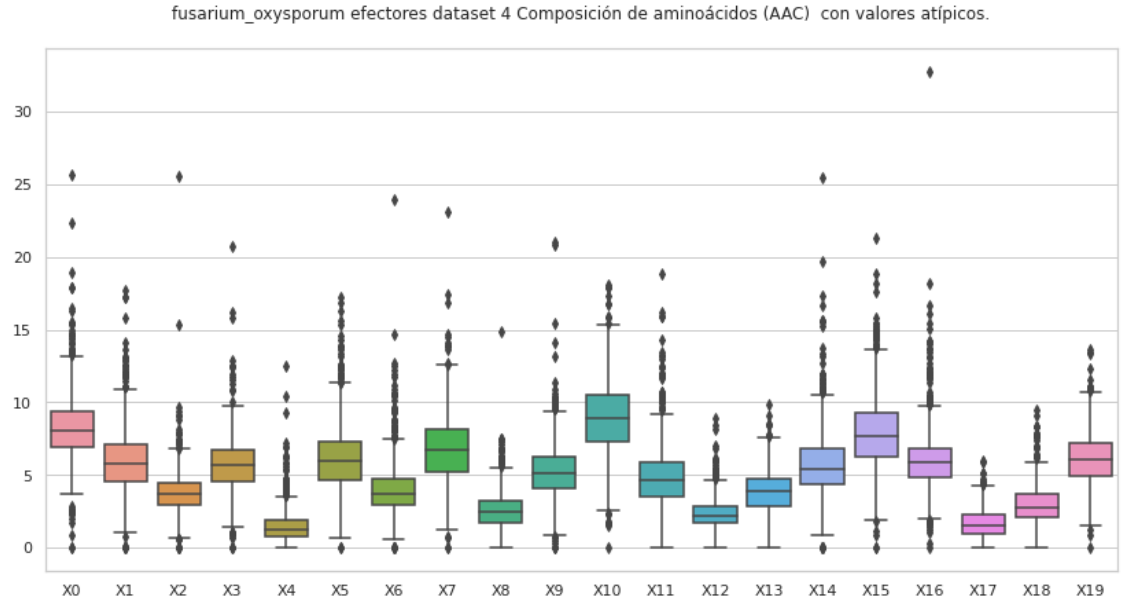
| | … | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | X19 \ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | … | 2.703 | 2.162 | 4.324 | 10.811 | 6.486 | 6.486 | 1.081 | 0.541 | 7.568 |

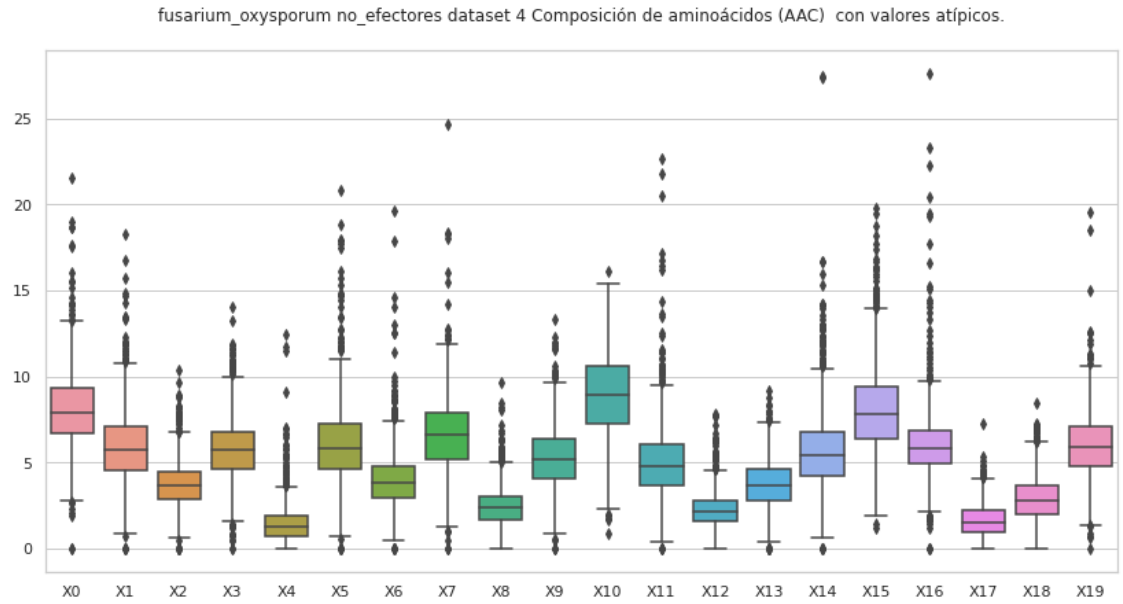
```
1    …  4.000  1.636  4.182   3.091   8.727  5.636  2.727  4.364  5.636
2    …  4.348  2.899  7.246   2.899  15.942  5.797  0.000  0.000  4.348
3    …  4.545  1.178  2.694   5.219   6.566  8.081  2.862  3.367  7.071
4    …  3.750  1.750  5.250   3.250   9.750  7.000  2.750  3.750  7.000
..   …    …    …    …     …     …    …    …    …    …
995  …  6.302  2.322  2.819   6.136   8.458  7.794  0.166  2.156  4.146
996  …  3.660  2.092  5.490   7.059   8.366  4.967  1.046  3.007  5.098
997  …  5.455  2.238  3.357   6.853   5.455  7.692  3.916  3.916  5.734
998  …  2.998  1.713  3.640   3.640   8.565  7.923  3.640  5.567  5.996
999  …  6.960  2.747  2.930   4.029   5.678  5.861  2.198  3.663  5.495

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

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.076510     5.890531     3.757454     5.788568     1.488891
std       2.274743     2.175951     1.372791     1.893998     1.256598
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.679000     4.533000     2.881000     4.662500     0.748250
50%       7.905500     5.746000     3.648000     5.726000     1.282000
75%       9.294500     7.059250     4.446500     6.803750     1.884750
max      21.552000    18.261000    10.345000    14.010000    12.414000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.090524     4.012818     6.692446     2.473274     5.261190
std       2.446491     1.797379     2.282578     1.185002     1.836348
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.586000     2.923000     5.211750     1.695000     4.057500
50%       5.859000     3.829500     6.586500     2.371000     5.186000
```

```
75%       7.256500     4.753000     7.910250     3.032750     6.355000
max      20.779000    19.626000    24.642000     9.677000    13.333000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.846665     5.031270     2.302363     3.751611     5.754727
std        2.503962     2.300357     1.073834     1.461807     2.519570
min        0.909000     0.000000     0.000000     0.000000     0.000000
25%        7.240250     3.640500     1.590750     2.782000     4.224500
50%        8.915500     4.811500     2.162000     3.706500     5.430000
75%       10.602250     6.024000     2.797500     4.627000     6.744000
max       16.092000    22.642000     7.843000     9.174000    27.402000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.142464     6.030595     1.635623     2.915258     6.057195
std        2.600667     2.236678     1.001579     1.339043     1.897945
min        1.227000     0.000000     0.000000     0.000000     0.000000
25%        6.403250     4.911500     0.951250     2.030750     4.813750
50%        7.826000     5.789000     1.534500     2.776000     5.919000
75%        9.416500     6.851000     2.212250     3.704000     7.120750
max       19.777000    27.586000     7.273000     8.491000    19.540000
```






## 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 fusarium_oxysporum dataset 4, sin
valores atípicos.
Valores del documento csv.

        X0     X1     X2     X3     X4     X5     X6      X7     X8     X9  \
0    8.889  5.051  4.040  7.677  1.414  8.283  4.646   6.465  2.424  5.455
1    8.945  3.399  3.220  2.683  1.073  4.293  2.504   9.481  2.147  7.335
2    6.949  7.553  3.323  7.855  0.000  7.855  3.625   7.553  2.115  4.834
4    5.898  4.960  3.887  6.166  2.547  4.692  4.155   7.507  2.011  4.960
5    6.868  4.945  4.121  4.945  1.923  9.615  1.648   7.418  3.297  5.220
..     …      …      …      …      …      …      …       …      …      …
993  7.524  8.252  5.340  6.311  0.971  6.068  2.913   7.767  2.427  7.282
994  8.537  5.285  4.878  4.878  2.033  5.691  3.659   7.317  4.065  3.659
995  8.754  8.754  4.377  4.040  1.347  5.724  4.377   4.377  1.852  5.556
997  5.634  4.225  2.113  6.338  1.408  7.746  2.113  10.563  2.113  8.451
998  7.991  4.566  5.708  5.479  2.283  5.708  3.196   7.078  4.795  4.566

     …    X11    X12    X13    X14    X15    X16    X17    X18    X19  \
0    …  5.051  2.222  3.434  4.040  5.455  3.030  3.232  3.636  4.040
1    …  3.041  2.504  5.009  5.546  6.977  5.903  2.862  4.293  7.513
2    …  3.323  1.511  3.323  6.949  9.668  3.625  1.208  2.115  8.459
4    …  3.351  2.011  4.692  6.032  8.847  6.702  1.877  4.826  5.496
5    …  4.396  1.923  4.121  6.868  4.670  6.319  2.747  4.670  6.319
..   …    …      …      …      …      …      …      …      …      …
993  …  5.340  1.942  3.155  3.641  6.311  6.311  1.214  0.971  8.981
```

```
994  …  2.846  1.626  4.065  7.724  6.911  8.130  1.220  2.033  5.691
995  …  2.694  1.515  5.556  4.714  6.734  5.724  1.684  2.694  7.407
997  …  6.338  2.113  4.225  4.930  7.042  7.746  0.704  2.817  4.930
998  …  6.393  3.196  6.393  4.795  5.936  4.566  2.055  4.338  4.566

          X20
0    efectores
1    efectores
2    efectores
4    efectores
5    efectores
..         …
993  efectores
994  efectores
995  efectores
997  efectores
998  efectores

[864 rows x 21 columns]
```

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

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 864.000000 | 864.000000 | 864.000000 | 864.000000 | 864.000000 | 864.000000 | |
| mean | 8.221344 | 5.855598 | 3.767699 | 5.695245 | 1.382638 | 6.043670 | |
| std | 1.963446 | 1.906061 | 1.155894 | 1.588183 | 0.875733 | 1.953825 | |
| min | 2.273000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 6.957000 | 4.670750 | 2.994500 | 4.643750 | 0.801000 | 4.728500 | |
| 50% | 8.055000 | 5.742500 | 3.673500 | 5.777000 | 1.241500 | 5.971500 | |
| 75% | 9.316000 | 7.044250 | 4.453500 | 6.685250 | 1.826750 | 7.237250 | |
| max | 14.915000 | 12.587000 | 8.185000 | 11.236000 | 4.724000 | 12.568000 | |

| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 864.000000 | 864.000000 | 864.000000 | 864.000000 | 864.000000 | 864.000000 | |
| mean | 3.849593 | 6.856610 | 2.506541 | 5.354251 | 9.093407 | 4.808590 | |
| std | 1.368563 | 2.002142 | 1.049760 | 1.627383 | 2.244119 | 1.728023 | |
| min | 0.000000 | 1.250000 | 0.000000 | 0.000000 | 1.786000 | 0.000000 | |
| 25% | 2.920750 | 5.383500 | 1.782000 | 4.300250 | 7.529000 | 3.595250 | |
| 50% | 3.675000 | 6.785500 | 2.447000 | 5.245500 | 9.050000 | 4.679000 | |
| 75% | 4.604250 | 8.194750 | 3.203000 | 6.271000 | 10.575250 | 5.797000 | |
| max | 9.434000 | 12.632000 | 6.306000 | 10.902000 | 15.909000 | 10.714000 | |

| | X12 | X13 | X14 | X15 | X16 | X17 | \ |
|---|---|---|---|---|---|---|---|
| count | 864.000000 | 864.000000 | 864.000000 | 864.000000 | 864.000000 | 864.000000 | |
| mean | 2.290106 | 3.942192 | 5.624181 | 7.898201 | 5.966410 | 1.712209 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| std | 0.886909 | 1.286826 | 1.849289 | 2.262696 | 1.577833 | 0.931121 |
| min | 0.000000 | 0.000000 | 0.000000 | 1.923000 | 1.000000 | 0.000000 |
| 25% | 1.674000 | 3.009500 | 4.397750 | 6.388000 | 4.978250 | 1.027750 |
| 50% | 2.194000 | 3.938000 | 5.401000 | 7.641500 | 5.831000 | 1.594000 |
| 75% | 2.798750 | 4.780500 | 6.695000 | 9.147750 | 6.789250 | 2.280500 |
| max | 5.357000 | 7.738000 | 11.971000 | 15.385000 | 12.375000 | 4.580000 |

| | X18 | X19 |
|---|---|---|
| count | 864.000000 | 864.000000 |
| mean | 2.901692 | 6.229844 |
| std | 1.078809 | 1.649483 |
| min | 0.000000 | 1.789000 |
| 25% | 2.161500 | 5.079750 |
| 50% | 2.812500 | 6.135000 |
| 75% | 3.650000 | 7.268500 |
| max | 6.769000 | 11.111000 |

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

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | \ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7.636 | 6.909 | 3.818 | 6.727 | 2.909 | 4.364 | 3.818 | 4.364 | 2.182 | 6.364 | |
| 3 | 9.428 | 4.209 | 5.387 | 5.051 | 2.020 | 5.051 | 3.030 | 8.586 | 2.694 | 5.387 | |
| 4 | 9.500 | 4.750 | 3.250 | 5.250 | 1.500 | 3.000 | 3.250 | 7.250 | 2.000 | 7.750 | |
| 5 | 6.371 | 8.315 | 3.240 | 5.940 | 0.324 | 8.099 | 4.752 | 5.940 | 1.728 | 4.752 | |
| 6 | 6.747 | 5.301 | 4.096 | 5.060 | 2.651 | 7.711 | 5.301 | 8.434 | 2.651 | 7.229 | |
| .. | … | … | … | … | … | … | … | … | … | … | |
| 995 | 10.779 | 6.799 | 3.648 | 4.478 | 1.658 | 5.638 | 3.814 | 6.302 | 1.327 | 4.478 | |
| 996 | 9.412 | 5.359 | 3.791 | 4.052 | 0.784 | 4.575 | 3.529 | 8.758 | 1.699 | 6.667 | |
| 997 | 6.993 | 5.455 | 5.315 | 6.154 | 0.280 | 4.755 | 2.517 | 7.413 | 3.217 | 4.615 | |
| 998 | 7.923 | 4.069 | 6.852 | 6.424 | 0.428 | 1.713 | 5.353 | 10.278 | 2.355 | 3.426 | |
| 999 | 12.454 | 5.128 | 2.930 | 6.777 | 1.099 | 5.861 | 4.396 | 6.410 | 2.381 | 4.762 | |

| | … | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | X19 | \ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | … | 4.000 | 1.636 | 4.182 | 3.091 | 8.727 | 5.636 | 2.727 | 4.364 | 5.636 | |
| 3 | … | 4.545 | 1.178 | 2.694 | 5.219 | 6.566 | 8.081 | 2.862 | 3.367 | 7.071 | |
| 4 | … | 3.750 | 1.750 | 5.250 | 3.250 | 9.750 | 7.000 | 2.750 | 3.750 | 7.000 | |
| 5 | … | 5.076 | 1.512 | 2.808 | 7.991 | 10.259 | 5.940 | 0.972 | 3.024 | 5.400 | |
| 6 | … | 4.819 | 4.337 | 2.651 | 3.373 | 7.470 | 3.133 | 1.205 | 4.578 | 6.024 | |
| .. | … | … | … | … | … | … | … | … | … | | |
| 995 | … | 6.302 | 2.322 | 2.819 | 6.136 | 8.458 | 7.794 | 0.166 | 2.156 | 4.146 | |
| 996 | … | 3.660 | 2.092 | 5.490 | 7.059 | 8.366 | 4.967 | 1.046 | 3.007 | 5.098 | |
| 997 | … | 5.455 | 2.238 | 3.357 | 6.853 | 5.455 | 7.692 | 3.916 | 3.916 | 5.734 | |
| 998 | … | 2.998 | 1.713 | 3.640 | 3.640 | 8.565 | 7.923 | 3.640 | 5.567 | 5.996 | |

```
999  …  6.960  2.747  2.930  4.029   5.678  5.861  2.198  3.663  5.495

              X20
1     no_efectores
3     no_efectores
4     no_efectores
5     no_efectores
6     no_efectores
..             …
995   no_efectores
996   no_efectores
997   no_efectores
998   no_efectores
999   no_efectores

[856 rows x 21 columns]
```

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

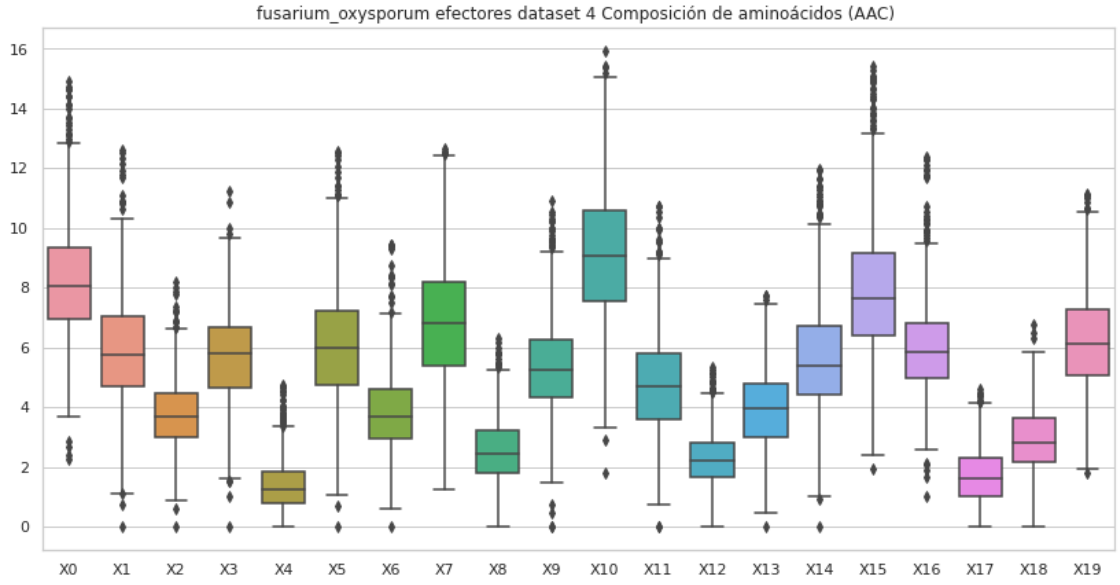
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 856.000000 | 856.000000 | 856.000000 | 856.000000 | 856.000000 | 856.000000 | |
| mean | 8.054053 | 5.929349 | 3.735971 | 5.843374 | 1.363916 | 6.035657 | |
| std | 1.944903 | 1.818389 | 1.177545 | 1.695790 | 0.866682 | 1.986692 | |
| min | 1.923000 | 0.000000 | 0.000000 | 0.800000 | 0.000000 | 0.000000 | |
| 25% | 6.766000 | 4.677500 | 2.918750 | 4.843500 | 0.771750 | 4.787750 | |
| 50% | 7.901000 | 5.796500 | 3.654500 | 5.789000 | 1.265000 | 5.944000 | |
| 75% | 9.203000 | 7.023500 | 4.403000 | 6.781750 | 1.786000 | 7.195500 | |
| max | 14.557000 | 12.033000 | 7.692000 | 11.397000 | 5.042000 | 13.380000 | |

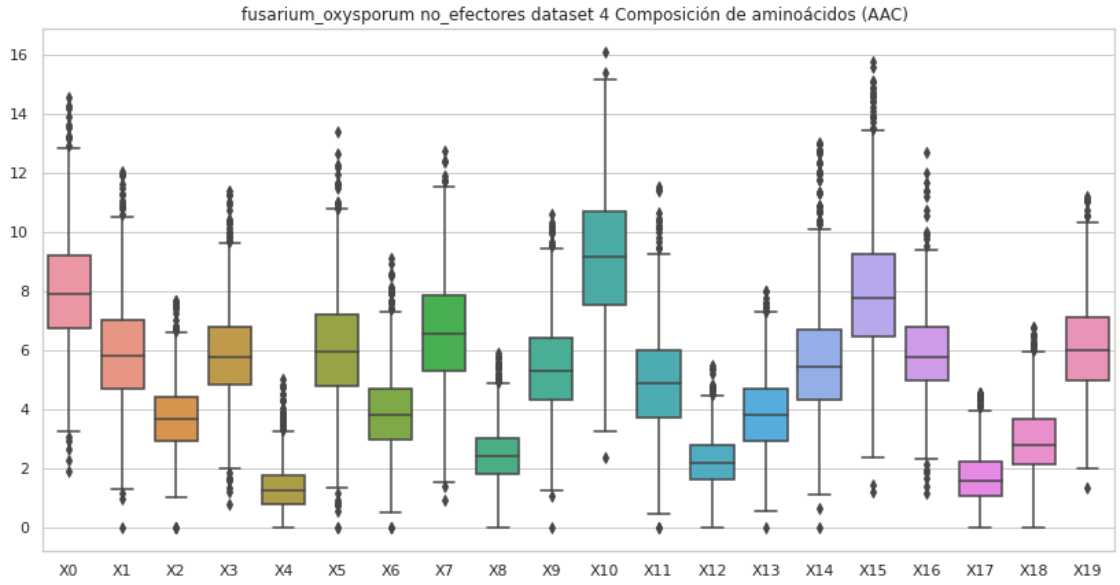
| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 856.000000 | 856.000000 | 856.000000 | 856.000000 | 856.000000 | 856.000000 | |
| mean | 3.923657 | 6.641397 | 2.500959 | 5.388418 | 9.137835 | 4.982606 | |
| std | 1.401390 | 1.926127 | 1.004285 | 1.622322 | 2.248429 | 1.738399 | |
| min | 0.000000 | 0.926000 | 0.000000 | 0.000000 | 2.381000 | 0.000000 | |
| 25% | 2.954250 | 5.292500 | 1.791750 | 4.326750 | 7.559000 | 3.741500 | |
| 50% | 3.829500 | 6.553000 | 2.413000 | 5.306500 | 9.146500 | 4.877000 | |
| 75% | 4.692000 | 7.844250 | 3.043750 | 6.413000 | 10.682000 | 6.001750 | |
| max | 9.140000 | 12.727000 | 5.888000 | 10.628000 | 16.092000 | 11.538000 | |

| | X12 | X13 | X14 | X15 | X16 | X17 | \ |
|---|---|---|---|---|---|---|---|
| count | 856.000000 | 856.000000 | 856.000000 | 856.000000 | 856.000000 | 856.000000 | |
| mean | 2.272143 | 3.884775 | 5.654891 | 8.004923 | 5.885801 | 1.702728 | |
| std | 0.911418 | 1.339633 | 2.014832 | 2.277998 | 1.500809 | 0.898762 | |
| min | 0.000000 | 0.000000 | 0.000000 | 1.227000 | 1.181000 | 0.000000 | |
| 25% | 1.636000 | 2.938250 | 4.340250 | 6.452000 | 4.960750 | 1.073750 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 50% | 2.179500 | 3.815500 | 5.428000 | 7.782500 | 5.774500 | 1.594000 |
| 75% | 2.778000 | 4.680500 | 6.685250 | 9.251750 | 6.780750 | 2.239500 |
| max | 5.495000 | 8.014000 | 13.031000 | 15.761000 | 12.690000 | 4.560000 |

| | X18 | X19 |
|---|---|---|
| count | 856.000000 | 856.000000 |
| mean | 2.931957 | 6.125525 |
| std | 1.199309 | 1.630810 |
| min | 0.000000 | 1.343000 |
| 25% | 2.147250 | 4.975000 |
| 50% | 2.795500 | 6.006000 |
| 75% | 3.678250 | 7.137000 |
| max | 6.790000 | 11.218000 |






## 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
fusarium_oxysporum dataset 4, con valores atípicos.
Valores del documento csv.
```

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.069541  0.011063  0.060058  0.064800  0.026868  0.050576  0.018966
1    0.024059  0.002887  0.007218  0.011548  0.013473  0.025503  0.005774
2    0.028687  0.000000  0.032429  0.032429  0.013720  0.031182  0.008731
3    0.021591  0.004095  0.017868  0.013401  0.010795  0.020102  0.001861
4    0.035692  0.015413  0.037315  0.028392  0.028392  0.045427  0.012168
..        …         …         …         …         …         …         …
995  0.046380  0.007135  0.021406  0.030325  0.029434  0.023190  0.009811
996  0.027310  0.013655  0.020482  0.006827  0.006827  0.034137  0.006827
997  0.026014  0.006504  0.029266  0.035770  0.019511  0.048777  0.009755
998  0.067261  0.019217  0.046122  0.048044  0.053809  0.059574  0.040357
999  0.066257  0.000000  0.022086  0.034706  0.006310  0.034706  0.025241

           X7        X8        X9  …       X74       X75       X76  \
0    0.042673  0.039512  0.090088  … -0.003558 -0.018026  0.011991
1    0.019729  0.008180  0.030315  …  0.020920  0.009122  0.002025
2    0.019956  0.013720  0.033677  … -0.005468  0.022056  0.012153
3    0.009679  0.003723  0.014146  … -0.001891 -0.002650  0.033826
4    0.030014  0.020280  0.056783  …  0.000924 -0.008873  0.002788
..        …         …         …  …         …         …         …
995  0.029434  0.014271  0.064219  … -0.011668  0.005695  0.003179
996  0.013655  0.006827  0.006827  …  0.032470  0.054849  0.032835
997  0.039022  0.029266  0.039022  … -0.058681 -0.053460  0.011576
998  0.038435  0.053809  0.053809  … -0.016974 -0.011343  0.025588
999  0.018931  0.003155  0.050482  … -0.022579 -0.007619  0.007502

           X77       X78       X79       X80       X81       X82         X83
0    0.042574  0.029982 -0.004157  0.028618  0.032674  0.018411  efectores
1    0.013826  0.002177  0.022486  0.011114  0.007156  0.006626  efectores
2    0.003697  0.017453  0.026867 -0.018722 -0.000759  0.017817  efectores
3    0.003463  0.007118  0.032254  0.000164  0.000240  0.033149  efectores
4    0.011796  0.001237  0.003610  0.002608 -0.000800  0.035840  efectores
..        …         …         …         …         …         …         …
995 -0.000221 -0.000090  0.012979 -0.010885  0.005413  0.011613  efectores
996 -0.003045  0.035020  0.012133  0.006029  0.022228  0.017007  efectores
```

```
997 -0.004414  0.028750  0.040168  0.012097  0.031806 -0.010277  efectores
998 -0.022836  0.014347  0.032542 -0.001913 -0.004947 -0.006230  efectores
999  0.005126  0.005534  0.040621 -0.017329 -0.015832  0.014838  efectores

[1000 rows x 84 columns]
```

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

| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.035700 | 0.007329 | 0.027461 | 0.030240 | 0.017867 | |
| std | 0.103356 | 0.009148 | 0.041800 | 0.057291 | 0.070822 | |
| min | -3.123084 | 0.000000 | -1.041028 | -1.041028 | -2.082056 | |
| 25% | 0.027008 | 0.002875 | 0.016412 | 0.017633 | 0.010677 | |
| 50% | 0.035690 | 0.005509 | 0.024962 | 0.026619 | 0.016822 | |
| 75% | 0.045978 | 0.009163 | 0.034753 | 0.037390 | 0.024442 | |
| max | 0.673533 | 0.167601 | 0.502802 | 1.347066 | 0.673533 | |

| | X5 | X6 | X7 | X8 | X9 | … | \ |
|---|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | … | |
| mean | 0.032033 | 0.011866 | 0.026489 | 0.021358 | 0.039640 | … | |
| std | 0.063930 | 0.034724 | 0.054933 | 0.102371 | 0.168898 | … | |
| min | -1.041028 | -1.041028 | -1.041028 | -3.123084 | -5.205140 | … | |
| 25% | 0.021972 | 0.006314 | 0.016007 | 0.013613 | 0.026674 | … | |
| 50% | 0.029576 | 0.010936 | 0.023662 | 0.021111 | 0.039448 | … | |
| 75% | 0.039012 | 0.017207 | 0.032084 | 0.029995 | 0.055803 | … | |
| max | 1.683832 | 0.070401 | 1.010299 | 0.502802 | 0.502802 | … | |

| | X73 | X74 | X75 | X76 | X77 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.017583 | 0.000204 | 0.005349 | 0.010513 | 0.010635 | |
| std | 0.110078 | 0.063703 | 0.111278 | 0.117686 | 0.272558 | |
| min | -0.600806 | -1.867238 | -1.768576 | -3.603945 | -0.268670 | |
| 25% | 0.005840 | -0.006504 | -0.002020 | 0.006266 | -0.006276 | |
| 50% | 0.014540 | 0.002636 | 0.005832 | 0.014752 | 0.002708 | |
| 75% | 0.024137 | 0.011623 | 0.015602 | 0.024042 | 0.010606 | |
| max | 3.359012 | 0.215723 | 2.407965 | 0.365075 | 8.447101 | |

| | X78 | X79 | X80 | X81 | X82 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | -0.002741 | 0.025587 | 0.001035 | 0.005194 | 0.018757 |
| std | 0.180915 | 0.285669 | 0.070035 | 0.125769 | 0.104011 |
| min | -5.378503 | -0.151824 | -1.292359 | -3.375830 | -0.350286 |
| 25% | -0.002870 | 0.006215 | -0.005776 | -0.002686 | 0.006021 |
| 50% | 0.004969 | 0.015142 | 0.003550 | 0.005238 | 0.015359 |

```
75%       0.014499     0.024107     0.011301     0.015474     0.024591
max       0.115660     8.582402     1.421523     1.723205     2.666939

[8 rows x 83 columns]
```

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.034970  0.031791  0.012716  0.025433  0.025433  0.028612  0.019074
1    0.050970  0.019417  0.044902  0.029126  0.027912  0.029126  0.014563
2    0.041181  0.010295  0.010295  0.020590  0.051476  0.041181  0.010295
3    0.040250  0.008625  0.021563  0.021563  0.011500  0.036656  0.011500
4    0.031006  0.004896  0.017135  0.009791  0.017135  0.023662  0.006528
..        …         …         …         …         …         …         …
995  0.037968  0.005841  0.015771  0.019860  0.009930  0.022196  0.004673
996  0.024268  0.002022  0.010449  0.011797  0.014157  0.022583  0.004382
997  0.061989  0.002480  0.054550  0.042152  0.029755  0.065708  0.028515
998  0.043777  0.002366  0.035495  0.009465  0.020114  0.056792  0.013015
999  0.069122  0.006099  0.037610  0.032528  0.016264  0.035577  0.013214

           X7        X8        X9  …       X74       X75       X76  \
0    0.031791  0.015895  0.069940  …  0.010882  0.010036  0.020767
1    0.042475  0.026699  0.072814  … -0.001552  0.007237  0.012916
2    0.082362  0.030886  0.061771  …  0.125212  0.085472  0.002905
3    0.023000  0.019406  0.032344  … -0.005714 -0.001115  0.024434
4    0.025294  0.012239  0.026926  …  0.002938  0.008355 -0.001990
..        …         …         …  …         …         …         …
995  0.015771  0.022196  0.037968  …  0.000248  0.002332  0.021386
996  0.017190  0.009438  0.027302  …  0.002951  0.004277  0.012162
997  0.040913  0.048351  0.076866  …  0.035561  0.014309  0.017358
998  0.018931  0.016564  0.041411  …  0.008021  0.011114  0.016218
999  0.026429  0.038627  0.045742  …  0.004818  0.010733  0.023923

           X77       X78       X79       X80       X81       X82           X83
0   -0.010306  0.007210  0.013361  0.012215  0.020861 -0.014852  no_efectores
1    0.017435  0.012026  0.000785 -0.006062 -0.005635  0.003494  no_efectores
2   -0.033815 -0.020524  0.067947 -0.234934 -0.152283  0.022651  no_efectores
3   -0.001729 -0.007003  0.024135  0.015695  0.011016  0.022309  no_efectores
4    0.004758  0.001362  0.023965  0.012873  0.004145  0.016600  no_efectores
..        …         …         …         …         …         …           …
995  0.008012  0.010168  0.026065  0.013221  0.010532  0.010755  no_efectores
996  0.007334  0.003937  0.015611  0.008735  0.001527  0.017832  no_efectores
997 -0.001999 -0.015216  0.002771  0.034850  0.027022  0.034281  no_efectores
```

```
998 -0.000766 -0.004756  0.034835  0.001008 -0.001523  0.024707  no_efectores
999 -0.004605 -0.008336  0.010589 -0.001986  0.016110  0.025610  no_efectores

[1000 rows x 84 columns]
```

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

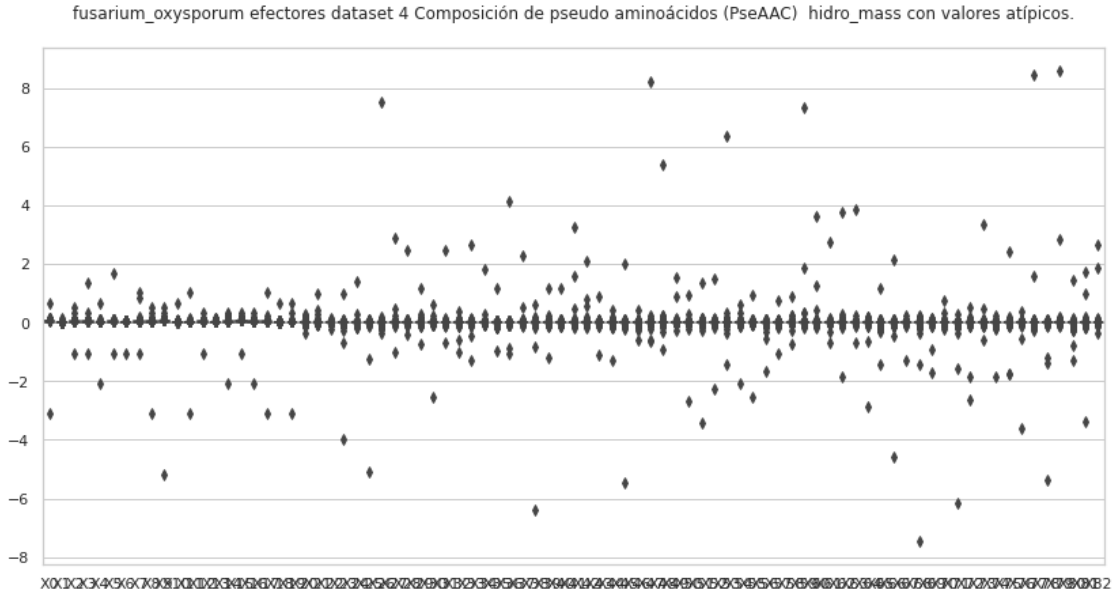
| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.037710 | 0.007607 | 0.028979 | 0.030213 | 0.019013 | |
| std | 0.020829 | 0.013513 | 0.023438 | 0.031026 | 0.018773 | |
| min | -0.235782 | -0.184599 | -0.184599 | -0.314376 | -0.369198 | |
| 25% | 0.027346 | 0.002891 | 0.017312 | 0.017524 | 0.010845 | |
| 50% | 0.036256 | 0.005839 | 0.026435 | 0.026803 | 0.016846 | |
| 75% | 0.045752 | 0.009945 | 0.036553 | 0.037922 | 0.024633 | |
| max | 0.232740 | 0.232740 | 0.465479 | 0.698219 | 0.232740 | |

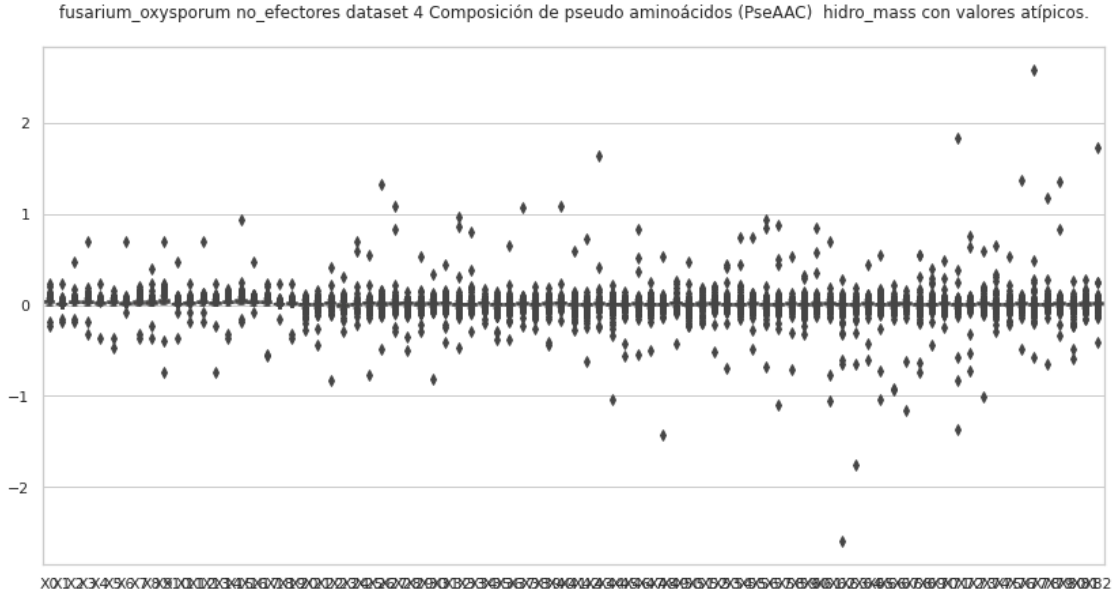
| | X5 | X6 | X7 | X8 | X9 | … | \ |
|---|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | … | |
| mean | 0.030905 | 0.013522 | 0.026152 | 0.024678 | 0.044632 | … | |
| std | 0.025580 | 0.024074 | 0.024748 | 0.025403 | 0.044470 | … | |
| min | -0.471564 | -0.078594 | -0.369198 | -0.369198 | -0.738397 | … | |
| 25% | 0.021283 | 0.006229 | 0.015786 | 0.014218 | 0.027595 | … | |
| 50% | 0.029874 | 0.010930 | 0.024070 | 0.021957 | 0.040883 | … | |
| 75% | 0.038751 | 0.016701 | 0.033596 | 0.031484 | 0.056828 | … | |
| max | 0.157249 | 0.698219 | 0.205585 | 0.393122 | 0.698219 | … | |

| | X73 | X74 | X75 | X76 | X77 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.014027 | 0.002179 | 0.006396 | 0.014605 | 0.003990 | |
| std | 0.042716 | 0.035755 | 0.028076 | 0.049165 | 0.089461 | |
| min | -1.014586 | -0.295211 | -0.209092 | -0.487244 | -0.582671 | |
| 25% | 0.004725 | -0.007814 | -0.002063 | 0.005212 | -0.006412 | |
| 50% | 0.013559 | 0.002227 | 0.005392 | 0.014946 | 0.003829 | |
| 75% | 0.023390 | 0.010371 | 0.015054 | 0.023535 | 0.011594 | |
| max | 0.587532 | 0.655807 | 0.535348 | 1.365365 | 2.583651 | |

| | X78 | X79 | X80 | X81 | X82 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.007660 | 0.016246 | 0.000478 | 0.007516 | 0.016307 |
| std | 0.047809 | 0.054496 | 0.035791 | 0.024289 | 0.060197 |
| min | -0.657149 | -0.330478 | -0.591872 | -0.155220 | -0.414571 |
| 25% | -0.001477 | 0.005812 | -0.005705 | -0.002185 | 0.006127 |
| 50% | 0.005920 | 0.015251 | 0.003439 | 0.006640 | 0.015123 |
| 75% | 0.015094 | 0.025022 | 0.012672 | 0.016418 | 0.024579 |

```
max        1.172648      1.350504      0.260404      0.278126      1.724404

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

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.069541 | 0.011063 | 0.060058 | 0.064800 | 0.026868 | 0.050576 | 0.018966 | |
| 1 | 0.024059 | 0.002887 | 0.007218 | 0.011548 | 0.013473 | 0.025503 | 0.005774 | |
| 2 | 0.028687 | 0.000000 | 0.032429 | 0.032429 | 0.013720 | 0.031182 | 0.008731 | |
| 3 | 0.021591 | 0.004095 | 0.017868 | 0.013401 | 0.010795 | 0.020102 | 0.001861 | |
| 4 | 0.035692 | 0.015413 | 0.037315 | 0.028392 | 0.028392 | 0.045427 | 0.012168 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.046380 | 0.007135 | 0.021406 | 0.030325 | 0.029434 | 0.023190 | 0.009811 | |
| 996 | 0.027310 | 0.013655 | 0.020482 | 0.006827 | 0.006827 | 0.034137 | 0.006827 | |
| 997 | 0.026014 | 0.006504 | 0.029266 | 0.035770 | 0.019511 | 0.048777 | 0.009755 | |
| 998 | 0.067261 | 0.019217 | 0.046122 | 0.048044 | 0.053809 | 0.059574 | 0.040357 | |
| 999 | 0.066257 | 0.000000 | 0.022086 | 0.034706 | 0.006310 | 0.034706 | 0.025241 | |

| | X7 | X8 | X9 | … | X74 | X75 | X76 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.042673 | 0.039512 | 0.090088 | … | -0.003558 | -0.018026 | 0.011991 | |
| 1 | 0.019729 | 0.008180 | 0.030315 | … | 0.020920 | 0.009122 | 0.002025 | |
| 2 | 0.019956 | 0.013720 | 0.033677 | … | -0.005468 | 0.022056 | 0.012153 | |
| 3 | 0.009679 | 0.003723 | 0.014146 | … | -0.001891 | -0.002650 | 0.033826 | |
| 4 | 0.030014 | 0.020280 | 0.056783 | … | 0.000924 | -0.008873 | 0.002788 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.029434 | 0.014271 | 0.064219 | … | -0.011668 | 0.005695 | 0.003179 | |
| 996 | 0.013655 | 0.006827 | 0.006827 | … | 0.032470 | 0.054849 | 0.032835 | |
| 997 | 0.039022 | 0.029266 | 0.039022 | … | -0.058681 | -0.053460 | 0.011576 | |
| 998 | 0.038435 | 0.053809 | 0.053809 | … | -0.016974 | -0.011343 | 0.025588 | |
| 999 | 0.018931 | 0.003155 | 0.050482 | … | -0.022579 | -0.007619 | 0.007502 | |

| | X77 | X78 | X79 | X80 | X81 | X82 | X83 |
|---|---|---|---|---|---|---|---|
| 0 | 0.042574 | 0.029982 | -0.004157 | 0.028618 | 0.032674 | 0.018411 | efectores |
| 1 | 0.013826 | 0.002177 | 0.022486 | 0.011114 | 0.007156 | 0.006626 | efectores |
| 2 | 0.003697 | 0.017453 | 0.026867 | -0.018722 | -0.000759 | 0.017817 | efectores |
| 3 | 0.003463 | 0.007118 | 0.032254 | 0.000164 | 0.000240 | 0.033149 | efectores |
| 4 | 0.011796 | 0.001237 | 0.003610 | 0.002608 | -0.000800 | 0.035840 | efectores |
| .. | … | … | … | … | … | … | |
| 995 | -0.000221 | -0.000090 | 0.012979 | -0.010885 | 0.005413 | 0.011613 | efectores |
| 996 | -0.003045 | 0.035020 | 0.012133 | 0.006029 | 0.022228 | 0.017007 | efectores |
| 997 | -0.004414 | 0.028750 | 0.040168 | 0.012097 | 0.031806 | -0.010277 | efectores |
| 998 | -0.022836 | 0.014347 | 0.032542 | -0.001913 | -0.004947 | -0.006230 | efectores |
| 999 | 0.005126 | 0.005534 | 0.040621 | -0.017329 | -0.015832 | 0.014838 | efectores |

[961 rows x 84 columns]

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

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 961.000000 | 961.000000 | 961.000000 | 961.000000 | 961.000000 | 961.000000 | |
| mean | 0.036945 | 0.006646 | 0.026428 | 0.028136 | 0.018232 | 0.030667 | |
| std | 0.013998 | 0.005550 | 0.014175 | 0.014861 | 0.011016 | 0.012855 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.002301 | |
| 25% | 0.026882 | 0.002870 | 0.016177 | 0.017275 | 0.010413 | 0.021822 | |
| 50% | 0.035238 | 0.005420 | 0.024494 | 0.026002 | 0.016508 | 0.029335 | |
| 75% | 0.045052 | 0.008847 | 0.033790 | 0.036066 | 0.023688 | 0.038190 | |
| max | 0.095195 | 0.031553 | 0.090958 | 0.104260 | 0.089930 | 0.079995 | |

| | X6 | X7 | X8 | X9 | … | X73 | \ |
|---|---|---|---|---|---|---|---|
| count | 961.000000 | 961.000000 | 961.000000 | 961.000000 | … | 961.000000 | |
| mean | 0.012245 | 0.024449 | 0.022429 | 0.042196 | … | 0.014864 | |
| std | 0.008649 | 0.012950 | 0.012512 | 0.022695 | … | 0.016484 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.001871 | … | -0.110985 | |
| 25% | 0.006224 | 0.015700 | 0.013335 | 0.026152 | … | 0.006720 | |
| 50% | 0.010672 | 0.023064 | 0.020686 | 0.038780 | … | 0.014692 | |
| 75% | 0.016344 | 0.030982 | 0.029133 | 0.053934 | … | 0.024020 | |
| max | 0.056890 | 0.089930 | 0.099249 | 0.155441 | … | 0.109104 | |

| | X74 | X75 | X76 | X77 | X78 | X79 | \ |
|---|---|---|---|---|---|---|---|
| count | 961.000000 | 961.000000 | 961.000000 | 961.000000 | 961.000000 | 961.000000 | |
| mean | 0.003063 | 0.007075 | 0.014802 | 0.001935 | 0.005826 | 0.014672 | |
| std | 0.017939 | 0.016884 | 0.015776 | 0.015973 | 0.015310 | 0.015848 | |
| min | -0.076671 | -0.063677 | -0.067994 | -0.069869 | -0.067740 | -0.081660 | |
| 25% | -0.005657 | -0.001387 | 0.006680 | -0.005811 | -0.002319 | 0.006525 | |
| 50% | 0.003005 | 0.005999 | 0.014792 | 0.002857 | 0.005158 | 0.015225 | |
| 75% | 0.011624 | 0.015559 | 0.023926 | 0.010548 | 0.014475 | 0.024085 | |
| max | 0.103273 | 0.121150 | 0.087871 | 0.097941 | 0.095807 | 0.089487 | |

| | X80 | X81 | X82 |
|---|---|---|---|
| count | 961.000000 | 961.000000 | 961.000000 |
| mean | 0.002381 | 0.006097 | 0.015211 |
| std | 0.016771 | 0.015812 | 0.014983 |
| min | -0.113097 | -0.085441 | -0.054309 |
| 25% | -0.005255 | -0.002417 | 0.006497 |
| 50% | 0.003572 | 0.005260 | 0.015639 |
| 75% | 0.010975 | 0.015228 | 0.024564 |
| max | 0.120478 | 0.087101 | 0.090075 |

[8 rows x 83 columns]

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.034970  0.031791  0.012716  0.025433  0.025433  0.028612  0.019074
1    0.050970  0.019417  0.044902  0.029126  0.027912  0.029126  0.014563
3    0.040250  0.008625  0.021563  0.021563  0.011500  0.036656  0.011500
4    0.031006  0.004896  0.017135  0.009791  0.017135  0.023662  0.006528
5    0.022251  0.001131  0.020743  0.028285  0.009806  0.020743  0.006034
..        …         …         …         …         …         …         …
995  0.037968  0.005841  0.015771  0.019860  0.009930  0.022196  0.004673
996  0.024268  0.002022  0.010449  0.011797  0.014157  0.022583  0.004382
997  0.061989  0.002480  0.054550  0.042152  0.029755  0.065708  0.028515
998  0.043777  0.002366  0.035495  0.009465  0.020114  0.056792  0.013015
999  0.069122  0.006099  0.037610  0.032528  0.016264  0.035577  0.013214

           X7        X8        X9  …       X74       X75       X76  \
0    0.031791  0.015895  0.069940  …  0.010882  0.010036  0.020767
1    0.042475  0.026699  0.072814  … -0.001552  0.007237  0.012916
3    0.023000  0.019406  0.032344  … -0.005714 -0.001115  0.024434
4    0.025294  0.012239  0.026926  …  0.002938  0.008355 -0.001990
5    0.016594  0.017726  0.026400  …  0.011565  0.023580  0.003618
..        …         …         …  …         …         …         …
995  0.015771  0.022196  0.037968  …  0.000248  0.002332  0.021386
996  0.017190  0.009438  0.027302  …  0.002951  0.004277  0.012162
997  0.040913  0.048351  0.076866  …  0.035561  0.014309  0.017358
998  0.018931  0.016564  0.041411  …  0.008021  0.011114  0.016218
999  0.026429  0.038627  0.045742  …  0.004818  0.010733  0.023923

           X77       X78       X79       X80       X81       X82           X83
0   -0.010306  0.007210  0.013361  0.012215  0.020861 -0.014852  no_efectores
1    0.017435  0.012026  0.000785 -0.006062 -0.005635  0.003494  no_efectores
3   -0.001729 -0.007003  0.024135  0.015695  0.011016  0.022309  no_efectores
4    0.004758  0.001362  0.023965  0.012873  0.004145  0.016600  no_efectores
5    0.000140  0.023003  0.013136  0.005287  0.018214  0.007305  no_efectores
..        …         …         …         …         …         …           …
995  0.008012  0.010168  0.026065  0.013221  0.010532  0.010755  no_efectores
996  0.007334  0.003937  0.015611  0.008735  0.001527  0.017832  no_efectores
997 -0.001999 -0.015216  0.002771  0.034850  0.027022  0.034281  no_efectores
998 -0.000766 -0.004756  0.034835  0.001008 -0.001523  0.024707  no_efectores
999 -0.004605 -0.008336  0.010589 -0.001986  0.016110  0.025610  no_efectores

[932 rows x 84 columns]
```

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

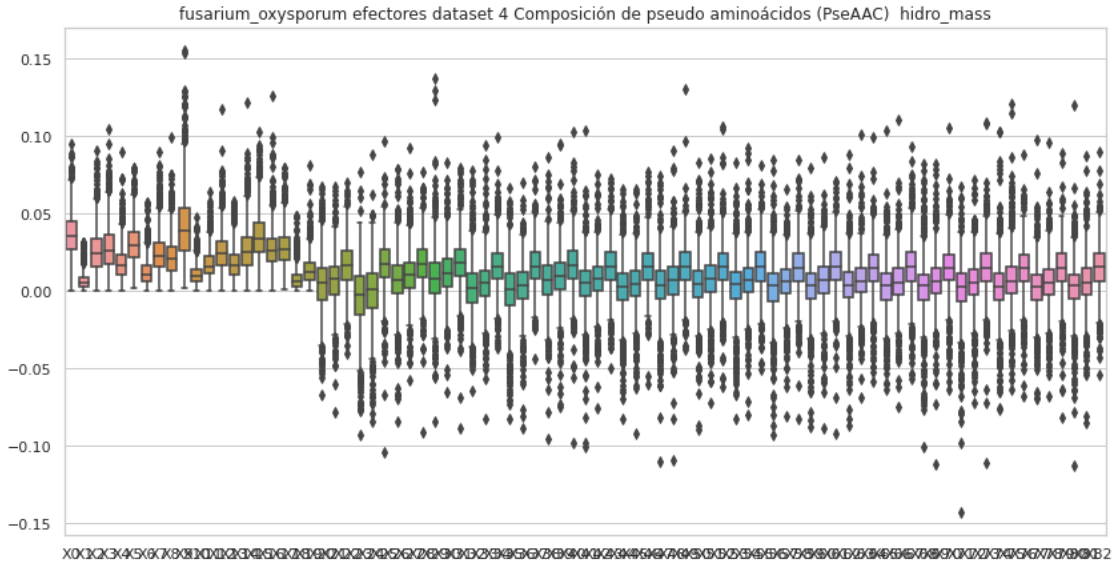
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 932.000000 | 932.000000 | 932.000000 | 932.000000 | 932.000000 | 932.000000 | |
| mean | 0.036401 | 0.006819 | 0.027251 | 0.028147 | 0.017773 | 0.030175 | |
| std | 0.013071 | 0.005799 | 0.014065 | 0.014513 | 0.010134 | 0.012691 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.027108 | 0.002803 | 0.017161 | 0.017253 | 0.010599 | 0.020979 | |
| 50% | 0.035789 | 0.005677 | 0.025670 | 0.026002 | 0.016306 | 0.029393 | |
| 75% | 0.044610 | 0.009315 | 0.034929 | 0.036899 | 0.023410 | 0.036885 | |
| max | 0.096327 | 0.040572 | 0.088996 | 0.089708 | 0.058313 | 0.086894 | |

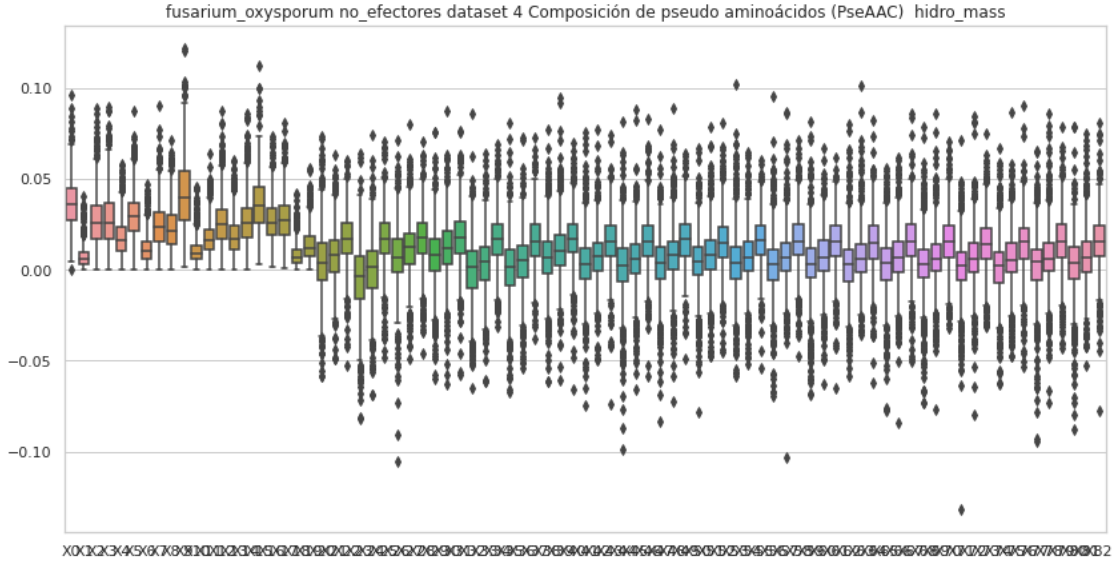
| | X6 | X7 | X8 | X9 | … | X73 | \ |
|---|---|---|---|---|---|---|---|
| count | 932.000000 | 932.000000 | 932.000000 | 932.000000 | … | 932.000000 | |
| mean | 0.011783 | 0.024503 | 0.022931 | 0.041402 | … | 0.014338 | |
| std | 0.007540 | 0.012414 | 0.012074 | 0.020094 | … | 0.014323 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.001576 | … | -0.031042 | |
| 25% | 0.006155 | 0.015595 | 0.013993 | 0.027099 | … | 0.005846 | |
| 50% | 0.010506 | 0.023213 | 0.021361 | 0.039498 | … | 0.013665 | |
| 75% | 0.015811 | 0.031769 | 0.030029 | 0.053948 | … | 0.022818 | |
| max | 0.046632 | 0.089852 | 0.071130 | 0.121623 | … | 0.074691 | |

| | X74 | X75 | X76 | X77 | X78 | X79 | \ |
|---|---|---|---|---|---|---|---|
| count | 932.000000 | 932.000000 | 932.000000 | 932.000000 | 932.000000 | 932.000000 | |
| mean | 0.001338 | 0.006127 | 0.014482 | 0.002355 | 0.007006 | 0.015610 | |
| std | 0.015608 | 0.014690 | 0.014974 | 0.016968 | 0.015876 | 0.014992 | |
| min | -0.062786 | -0.047703 | -0.053447 | -0.094192 | -0.073162 | -0.043193 | |
| 25% | -0.006927 | -0.001593 | 0.005907 | -0.005488 | -0.000622 | 0.006758 | |
| 50% | 0.002256 | 0.005392 | 0.015143 | 0.004192 | 0.006114 | 0.015553 | |
| 75% | 0.009689 | 0.014421 | 0.022968 | 0.011369 | 0.014826 | 0.024682 | |
| max | 0.066972 | 0.086767 | 0.090288 | 0.065431 | 0.085784 | 0.078247 | |

| | X80 | X81 | X82 |
|---|---|---|---|
| count | 932.000000 | 932.000000 | 932.000000 |
| mean | 0.002628 | 0.007266 | 0.015512 |
| std | 0.016387 | 0.015501 | 0.015055 |
| min | -0.087950 | -0.052941 | -0.077525 |
| 25% | -0.004967 | -0.001500 | 0.007525 |
| 50% | 0.003517 | 0.006754 | 0.015266 |
| 75% | 0.012428 | 0.015712 | 0.024340 |
| max | 0.078490 | 0.074923 | 0.080416 |

[8 rows x 83 columns]




# 4 Composición de pseudo aminoácidos (PseAAC) mass

[7]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
```

```
estado = "con valores atípicos.\n"
comp = "mass"
df=""

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores fusarium_oxysporum
dataset 4, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.074760  0.011894  0.064566  0.069663  0.028885  0.054371  0.020389
1    0.049418  0.005930  0.014825  0.023721  0.027674  0.052383  0.011860
2    0.036385  0.000000  0.041131  0.041131  0.017402  0.039549  0.011074
3    0.022423  0.004253  0.018557  0.013918  0.011211  0.020876  0.001933
4    0.038617  0.016676  0.040372  0.030718  0.030718  0.049149  0.013165
..        …         …         …         …         …         …         …
995  0.059664  0.009179  0.027537  0.039011  0.037864  0.029832  0.012621
996  0.040135  0.020067  0.030101  0.010034  0.010034  0.050168  0.010034
997  0.029339  0.007335  0.033006  0.040341  0.022004  0.055010  0.011002
998  0.061640  0.017612  0.042268  0.044029  0.049312  0.054596  0.036984
999  0.061601  0.000000  0.020534  0.032267  0.005867  0.032267  0.023467

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.045876  0.042478  0.096849  …  0.026933  0.009744 -0.000394
1    0.040523  0.016802  0.062267  …  0.013443  0.010725  0.000617
2    0.025311  0.017402  0.042713  …  0.023906  0.030348  0.016944
3    0.010052  0.003866  0.014691  …  0.036935  0.037563  0.033212
4    0.032474  0.021942  0.061436  …  0.021013  0.014736  0.011092
..        …         …         …  …         …         …
995  0.037864  0.018358  0.082612  …  0.012112  0.014876  0.013706
996  0.020067  0.010034  0.010034  … -0.005259  0.022951 -0.005144
997  0.044008  0.033006  0.044008  … -0.002251  0.040098  0.049212
998  0.035223  0.049312  0.049312  …  0.011510  0.010871  0.046327
999  0.017600  0.002933  0.046934  …  0.041857  0.056623  0.028818

          X35       X36       X37       X38       X39       X40        X41
0    0.036181  0.015350 -0.018199  0.012891 -0.004468  0.019793  efectores
1   -0.014518  0.026841 -0.002464  0.004159  0.046187  0.013610  efectores
2    0.007372  0.045146  0.016212  0.015414  0.034076  0.022598  efectores
3    0.038401  0.037697  0.038591  0.035129  0.033497  0.034427  efectores
4    0.017118  0.024395  0.018396  0.003016  0.003906  0.038777  efectores
..        …         …         …         …         …         …
995 -0.014021  0.015294  0.005110  0.004090  0.016696  0.014940  efectores
996  0.045772  0.035587 -0.001299  0.048254  0.017831  0.024993  efectores
997  0.001662  0.004704  0.028038  0.013055  0.045301 -0.011591  efectores
998  0.014140  0.028868  0.010364  0.023450  0.029823 -0.005709  efectores
999 -0.002773  0.031561  0.031352  0.006975  0.037766  0.013795  efectores

[1000 rows x 42 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047559     0.008733     0.034401     0.038004     0.023507
std       0.015217     0.008639     0.016976     0.021353     0.013045
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.038126     0.003797     0.022337     0.023412     0.014388
50%       0.046005     0.007148     0.031455     0.034181     0.021985
75%       0.054815     0.011139     0.043712     0.047893     0.029938
max       0.155723     0.091415     0.138521     0.144115     0.103815

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038539     0.015731     0.031726     0.030225     0.054451  …
std       0.012515     0.010621     0.016419     0.019081     0.026357  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.030276     0.008612     0.021081     0.018011     0.036859  …
```

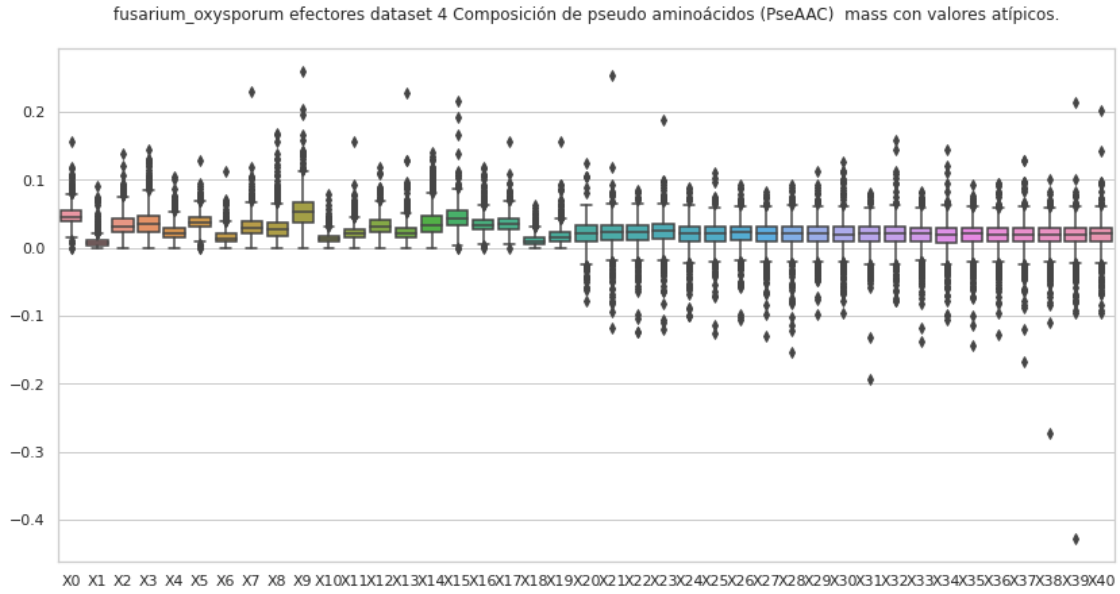
```
50%       0.037493     0.013875     0.030095     0.026776     0.052249  …
75%       0.045686     0.021078     0.039526     0.037204     0.067838  …
max       0.127823     0.112389     0.230868     0.168364     0.259538  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018773     0.018809     0.018513     0.017486     0.017700
std       0.020570     0.020727     0.020401     0.021167     0.021403
min      -0.192445    -0.078943    -0.137401    -0.106823    -0.143037
25%       0.008755     0.008708     0.010020     0.007978     0.009384
50%       0.020727     0.020265     0.021049     0.019293     0.020683
75%       0.030964     0.030747     0.029809     0.029979     0.029963
max       0.081032     0.159311     0.082980     0.143683     0.093448

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.017742     0.017779     0.017321     0.017060     0.018198
std       0.020401     0.021905     0.021953     0.025613     0.021665
min      -0.127455    -0.167487    -0.273582    -0.427935    -0.096503
25%       0.009233     0.008592     0.008919     0.008731     0.008494
50%       0.020196     0.019925     0.019512     0.019778     0.021024
75%       0.029320     0.030106     0.029269     0.029757     0.029725
max       0.095486     0.128152     0.100577     0.214182     0.202452

[8 rows x 41 columns]



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.036638  0.033307  0.013323  0.026646  0.026646  0.029976  0.019984
1    0.054507  0.020764  0.048018  0.031147  0.029849  0.031147  0.015573
2    0.037757  0.009439  0.009439  0.018879  0.047197  0.037757  0.009439
3    0.039854  0.008540  0.021351  0.021351  0.011387  0.036296  0.011387
4    0.052006  0.008212  0.028740  0.016423  0.028740  0.039689  0.010949
..        …         …         …         …         …         …         …
995  0.048068  0.007395  0.019967  0.025143  0.012572  0.028101  0.005916
996  0.040528  0.003377  0.017450  0.019701  0.023641  0.037714  0.007318
997  0.056078  0.002243  0.049349  0.038133  0.026918  0.059443  0.025796
998  0.042139  0.002278  0.034167  0.009111  0.019361  0.054667  0.012528
999  0.069870  0.006165  0.038017  0.032880  0.016440  0.035962  0.013357

           X7        X8        X9  …       X32       X33       X34  \
0    0.033307  0.016653  0.073275  … -0.012745  0.005509  0.016519
```

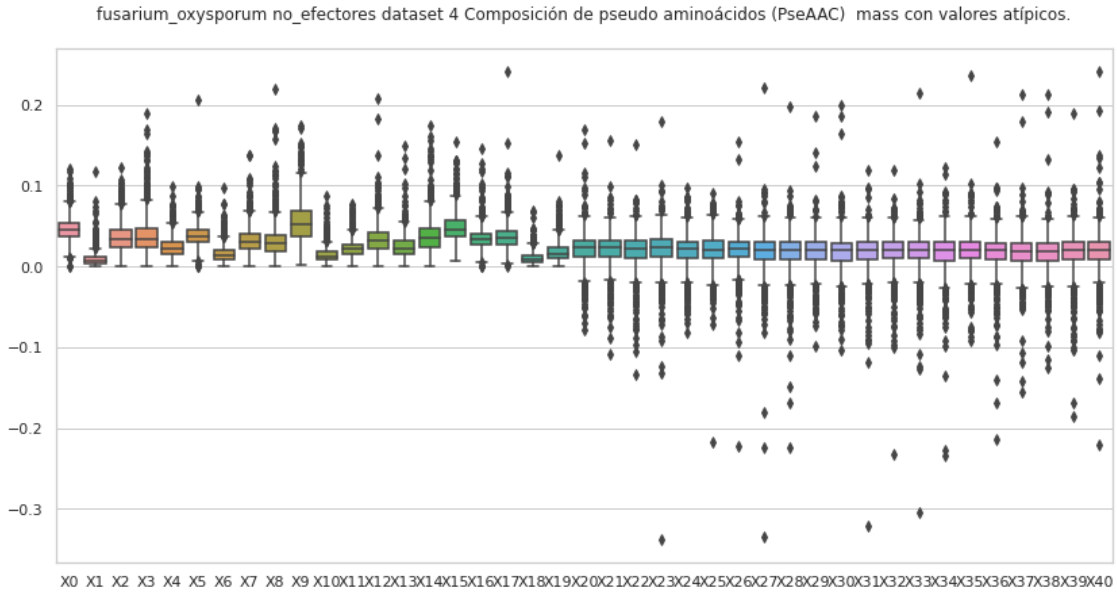
```
1    0.045422  0.028551  0.077867  …  0.022578  0.018990 -0.015400
2    0.075515  0.028318  0.056636  …  0.058068 -0.036471  0.052112
3    0.022774  0.019216  0.032026  …  0.024232  0.027109  0.032319
4    0.042426  0.020529  0.045163  …  0.001129  0.056603  0.016418
..        …         …         …  …         …         …
995  0.019967  0.028101  0.048068  …  0.031523  0.025171  0.028083
996  0.028707  0.015761  0.045594  …  0.029894  0.025938  0.028025
997  0.037012  0.043741  0.069537  … -0.012641 -0.007132 -0.017369
998  0.018222  0.015945  0.039862  …  0.018277  0.005675  0.020871
999  0.026715  0.039045  0.046237  …  0.018918  0.004392  0.021123

          X35       X36       X37       X38       X39       X40           X41
0    0.022948 -0.019101  0.013254  0.021758  0.013998 -0.015561  no_efectores
1    0.051450  0.016189  0.022366  0.013812  0.000839  0.003736  no_efectores
2    0.028010  0.025262 -0.006175  0.002664  0.062298  0.020768  no_efectores
3    0.020723  0.031307  0.033579  0.024194  0.023897  0.022090  no_efectores
4    0.011224  0.029979  0.021269 -0.003338  0.040196  0.027843  no_efectores
..        …         …         …         …         …         …           …
995  0.021606  0.016794  0.026514  0.027076  0.032999  0.013616  no_efectores
996  0.030432  0.032787  0.021884  0.020310  0.026070  0.029779  no_efectores
997  0.009255 -0.006179  0.005517  0.015703  0.002507  0.031013  no_efectores
998 -0.005786  0.052823  0.014101  0.015611  0.033531  0.023782  no_efectores
999  0.021767  0.005750  0.013607  0.024181  0.010703  0.025887  no_efectores

[1000 rows x 42 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.046932     0.009311     0.035766     0.038343     0.023272
std       0.014123     0.009345     0.017473     0.022885     0.012706
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.037209     0.004210     0.023327     0.023548     0.014680
50%       0.046193     0.007520     0.033827     0.034279     0.021384
75%       0.054370     0.011762     0.044934     0.047131     0.030202
max       0.120382     0.117292     0.122433     0.190068     0.098321

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038770     0.015626     0.032374     0.031525     0.054825  …
std       0.014070     0.010200     0.016137     0.020247     0.024604  …
min       0.000000     0.000000     0.000000     0.000000     0.001822  …
25%       0.030026     0.008893     0.021907     0.019364     0.037918  …
50%       0.037403     0.013699     0.030600     0.028085     0.052609  …
```

```
75%       0.045305     0.020095     0.040454     0.039041     0.069548  …
max       0.206220     0.097695     0.137649     0.219844     0.174669  …

                X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018310     0.018470     0.017744     0.017019     0.018868
std       0.022943     0.021886     0.024373     0.023851     0.021431
min      -0.322003    -0.232591    -0.304683    -0.234927    -0.092536
25%       0.009369     0.010286     0.009477     0.007507     0.009563
50%       0.020654     0.020801     0.020228     0.019543     0.020767
75%       0.030356     0.030365     0.029689     0.029828     0.030768
max       0.119806     0.119827     0.215342     0.123269     0.236828

                X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.017698     0.016731     0.017130     0.017586     0.018198
std       0.023354     0.022822     0.022087     0.023944     0.025211
min      -0.214688    -0.154745    -0.125578    -0.184812    -0.221468
25%       0.009110     0.006442     0.007252     0.008235     0.008978
50%       0.019890     0.018635     0.019307     0.020656     0.020548
75%       0.029454     0.028919     0.028566     0.030410     0.030036
max       0.154988     0.212917     0.213077     0.188963     0.241279

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.074760  0.011894  0.064566  0.069663  0.028885  0.054371  0.020389
1    0.049418  0.005930  0.014825  0.023721  0.027674  0.052383  0.011860
2    0.036385  0.000000  0.041131  0.041131  0.017402  0.039549  0.011074
3    0.022423  0.004253  0.018557  0.013918  0.011211  0.020876  0.001933
4    0.038617  0.016676  0.040372  0.030718  0.030718  0.049149  0.013165
..        …         …         …         …         …         …         …
994  0.044597  0.010618  0.025484  0.029731  0.021237  0.038226  0.021237
995  0.059664  0.009179  0.027537  0.039011  0.037864  0.029832  0.012621
997  0.029339  0.007335  0.033006  0.040341  0.022004  0.055010  0.011002
998  0.061640  0.017612  0.042268  0.044029  0.049312  0.054596  0.036984
999  0.061601  0.000000  0.020534  0.032267  0.005867  0.032267  0.023467

           X7        X8        X9  …       X32       X33       X34  \
0    0.045876  0.042478  0.096849  …  0.026933  0.009744 -0.000394
1    0.040523  0.016802  0.062267  …  0.013443  0.010725  0.000617
2    0.025311  0.017402  0.042713  …  0.023906  0.030348  0.016944
3    0.010052  0.003866  0.014691  …  0.036935  0.037563  0.033212
4    0.032474  0.021942  0.061436  …  0.021013  0.014736  0.011092
..        …         …         …  …         …         …         …
994  0.019113  0.014866  0.050968  …  0.010155  0.019554  0.018137
995  0.037864  0.018358  0.082612  …  0.012112  0.014876  0.013706
997  0.044008  0.033006  0.044008  … -0.002251  0.040098  0.049212
998  0.035223  0.049312  0.049312  …  0.011510  0.010871  0.046327
999  0.017600  0.002933  0.046934  …  0.041857  0.056623  0.028818
```

```
          X35       X36       X37       X38       X39       X40        X41
0    0.036181  0.015350 -0.018199  0.012891 -0.004468  0.019793  efectores
1   -0.014518  0.026841 -0.002464  0.004159  0.046187  0.013610  efectores
2    0.007372  0.045146  0.016212  0.015414  0.034076  0.022598  efectores
3    0.038401  0.037697  0.038591  0.035129  0.033497  0.034427  efectores
4    0.017118  0.024395  0.018396  0.003016  0.003906  0.038777  efectores
..        …         …         …         …         …         …
994  0.020122  0.039935  0.037972  0.026363  0.019226  0.010629  efectores
995 -0.014021  0.015294  0.005110  0.004090  0.016696  0.014940  efectores
997  0.001662  0.004704  0.028038  0.013055  0.045301 -0.011591  efectores
998  0.014140  0.028868  0.010364  0.023450  0.029823 -0.005709  efectores
999 -0.002773  0.031561  0.031352  0.006975  0.037766  0.013795  efectores

[858 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  858.000000  858.000000  858.000000  858.000000  858.000000  858.000000
mean     0.045536    0.007704    0.032475    0.034398    0.021947    0.037560
std      0.011721    0.005491    0.014166    0.016343    0.010360    0.010354
min      0.008043    0.000000    0.003172    0.000000    0.000000    0.002812
25%      0.037831    0.003800    0.022003    0.022539    0.014147    0.030239
50%      0.044923    0.006849    0.030671    0.032810    0.021102    0.037090
75%      0.052773    0.010441    0.041076    0.043948    0.028597    0.044408
max      0.089885    0.032993    0.082921    0.099457    0.057895    0.072374

                X6          X7          X8          X9  …         X31  \
count  858.000000  858.000000  858.000000  858.000000  …  858.000000
mean     0.014197    0.029699    0.027178    0.050977  …    0.020801
std      0.007962    0.012478    0.013271    0.020897  …    0.015434
min      0.000000    0.000000    0.000000    0.002596  …   -0.038602
25%      0.008243    0.020417    0.017510    0.035992  …    0.011886
50%      0.013060    0.028714    0.025782    0.050966  …    0.021323
75%      0.018979    0.037214    0.034902    0.064226  …    0.031064
max      0.045034    0.078175    0.085022    0.117369  …    0.078027

               X32         X33         X34         X35         X36         X37  \
count  858.000000  858.000000  858.000000  858.000000  858.000000  858.000000
mean     0.019908    0.020677    0.019315    0.020300    0.019632    0.019675
std      0.014885    0.014957    0.015260    0.015835    0.015471    0.015219
min     -0.032396   -0.041059   -0.034880   -0.041633   -0.043206   -0.041782
25%      0.010930    0.011777    0.010034    0.011816    0.011431    0.011074
50%      0.020787    0.021928    0.020028    0.021368    0.021025    0.020748
75%      0.030376    0.029955    0.029640    0.030344    0.029507    0.030226
```

```
max      0.075738    0.070110    0.067278    0.073614    0.075273    0.067510

              X38         X39         X40
count  858.000000  858.000000  858.000000
mean     0.019749    0.019880    0.019760
std      0.015537    0.014927    0.015401
min     -0.048021   -0.030520   -0.033479
25%      0.011524    0.011223    0.011028
50%      0.020728    0.020850    0.021637
75%      0.029561    0.030026    0.029907
max      0.082913    0.078183    0.070787

[8 rows x 41 columns]
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.036638  0.033307  0.013323  0.026646  0.026646  0.029976  0.019984
1    0.054507  0.020764  0.048018  0.031147  0.029849  0.031147  0.015573
3    0.039854  0.008540  0.021351  0.021351  0.011387  0.036296  0.011387
4    0.052006  0.008212  0.028740  0.016423  0.028740  0.039689  0.010949
5    0.040791  0.002074  0.038025  0.051853  0.017976  0.038025  0.011062
..        …         …         …         …         …         …         …
995  0.048068  0.007395  0.019967  0.025143  0.012572  0.028101  0.005916
996  0.040528  0.003377  0.017450  0.019701  0.023641  0.037714  0.007318
997  0.056078  0.002243  0.049349  0.038133  0.026918  0.059443  0.025796
998  0.042139  0.002278  0.034167  0.009111  0.019361  0.054667  0.012528
999  0.069870  0.006165  0.038017  0.032880  0.016440  0.035962  0.013357

           X7        X8        X9  …       X32       X33       X34  \
0    0.033307  0.016653  0.073275  … -0.012745  0.005509  0.016519
1    0.045422  0.028551  0.077867  …  0.022578  0.018990 -0.015400
3    0.022774  0.019216  0.032026  …  0.024232  0.027109  0.032319
4    0.042426  0.020529  0.045163  …  0.001129  0.056603  0.016418
5    0.030420  0.032494  0.048396  …  0.025951  0.019480  0.019534
..        …         …         …  …         …         …         …
995  0.019967  0.028101  0.048068  …  0.031523  0.025171  0.028083
996  0.028707  0.015761  0.045594  …  0.029894  0.025938  0.028025
997  0.037012  0.043741  0.069537  … -0.012641 -0.007132 -0.017369
998  0.018222  0.015945  0.039862  …  0.018277  0.005675  0.020871
999  0.026715  0.039045  0.046237  …  0.018918  0.004392  0.021123

          X35       X36       X37       X38       X39       X40           X41
0    0.022948 -0.019101  0.013254  0.021758  0.013998 -0.015561  no_efectores
```

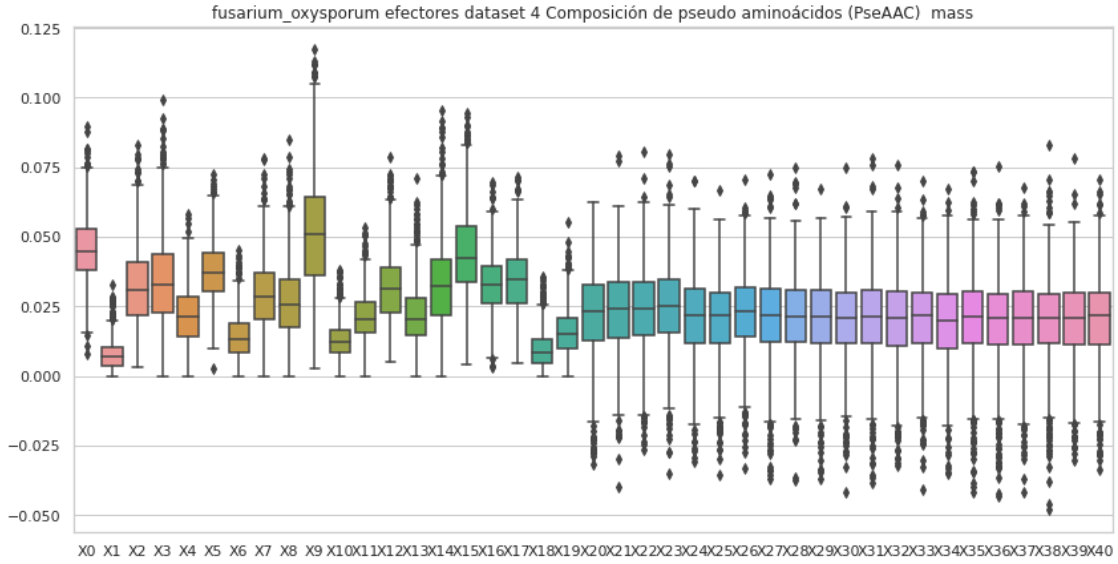
```
1    0.051450  0.016189  0.022366  0.013812  0.000839  0.003736  no_efectores
3    0.020723  0.031307  0.033579  0.024194  0.023897  0.022090  no_efectores
4    0.011224  0.029979  0.021269 -0.003338  0.040196  0.027843  no_efectores
5    0.019965  0.006516  0.020187  0.006633  0.024081  0.013391  no_efectores
..        …         …         …         …         …         …
995  0.021606  0.016794  0.026514  0.027076  0.032999  0.013616  no_efectores
996  0.030432  0.032787  0.021884  0.020310  0.026070  0.029779  no_efectores
997  0.009255 -0.006179  0.005517  0.015703  0.002507  0.031013  no_efectores
998 -0.005786  0.052823  0.014101  0.015611  0.033531  0.023782  no_efectores
999  0.021767  0.005750  0.013607  0.024181  0.010703  0.025887  no_efectores

[862 rows x 42 columns]
```

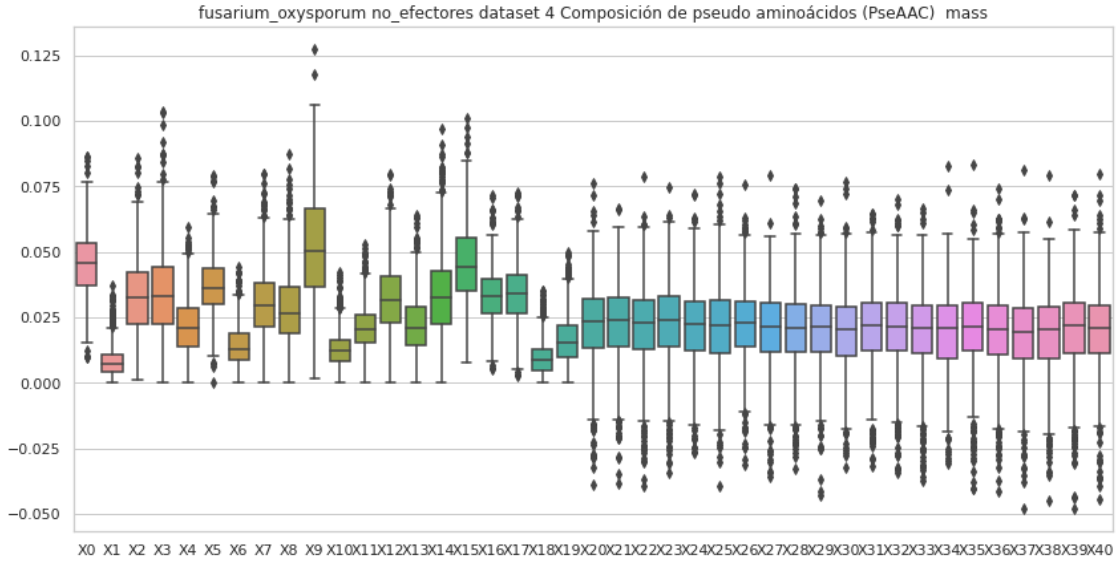
Composición de pseudo aminoácidos (PseAAC)  mass no_efectores fusarium_oxysporum
dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  862.000000  862.000000  862.000000  862.000000  862.000000  862.000000
mean     0.045677    0.008031    0.033158    0.034567    0.021975    0.037203
std      0.011859    0.005902    0.014158    0.016372    0.010546    0.010878
min      0.009820    0.000000    0.001233    0.000000    0.000000    0.000000
25%      0.037093    0.004042    0.022578    0.022562    0.014080    0.029841
50%      0.045589    0.007130    0.032431    0.032933    0.020909    0.036344
75%      0.053187    0.011027    0.042061    0.044382    0.028319    0.043644
max      0.086478    0.037266    0.085853    0.103618    0.059567    0.079306

                X6          X7          X8          X9  …         X31  \
count  862.000000  862.000000  862.000000  862.000000  …  862.000000
mean     0.014162    0.030326    0.028314    0.051637  …    0.020788
std      0.007606    0.012861    0.013775    0.020911  …    0.014695
min      0.000000    0.000000    0.000000    0.001822  …   -0.031506
25%      0.008572    0.021477    0.018894    0.036793  …    0.012210
50%      0.013107    0.029563    0.026472    0.050609  …    0.021836
75%      0.018712    0.038229    0.036921    0.066358  …    0.030631
max      0.044177    0.079663    0.087248    0.127345  …    0.064475

               X32         X33         X34         X35         X36         X37  \
count  862.000000  862.000000  862.000000  862.000000  862.000000  862.000000
mean     0.020806    0.019720    0.019401    0.020304    0.019887    0.018425
std      0.014805    0.015557    0.015244    0.015161    0.015595    0.015569
min     -0.034447   -0.037479   -0.030477   -0.040255   -0.041364   -0.047749
25%      0.012149    0.011338    0.009530    0.012353    0.010682    0.009392
50%      0.021544    0.020722    0.020796    0.021410    0.020617    0.019494
75%      0.030533    0.029782    0.029824    0.030712    0.029429    0.028772
max      0.069908    0.066601    0.082664    0.083081    0.074188    0.081149
```

| | X38 | X39 | X40 |
|---|---|---|---|
| count | 862.000000 | 862.000000 | 862.000000 |
| mean | 0.018767 | 0.020259 | 0.019739 |
| std | 0.015567 | 0.015865 | 0.015530 |
| min | -0.044922 | -0.047948 | -0.044381 |
| 25% | 0.009315 | 0.011323 | 0.011165 |
| 50% | 0.020254 | 0.021955 | 0.021079 |
| 75% | 0.029244 | 0.030810 | 0.029802 |
| max | 0.079107 | 0.071874 | 0.079534 |

[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 fusarium_oxysporum
dataset 4, con valores atípicos.
Valores del documento csv.
```

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.081607 | 0.012983 | 0.070479 | 0.076043 | 0.031530 | 0.059350 | 0.022256 | |
| 1 | 0.030761 | 0.003691 | 0.009228 | 0.014765 | 0.017226 | 0.032607 | 0.007383 | |
| 2 | 0.045943 | 0.000000 | 0.051936 | 0.051936 | 0.021973 | 0.049938 | 0.013983 | |
| 3 | 0.070845 | 0.013436 | 0.058631 | 0.043973 | 0.035423 | 0.065959 | 0.006107 | |
| 4 | 0.052421 | 0.022636 | 0.054803 | 0.041698 | 0.041698 | 0.066717 | 0.017871 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.061639 | 0.009483 | 0.028449 | 0.040302 | 0.039117 | 0.030819 | 0.013039 | |
| 996 | 0.045044 | 0.022522 | 0.033783 | 0.011261 | 0.011261 | 0.056304 | 0.011261 | |
| 997 | 0.045237 | 0.011309 | 0.050892 | 0.062201 | 0.033928 | 0.084820 | 0.016964 | |
| 998 | 0.089617 | 0.025605 | 0.061452 | 0.064012 | 0.071694 | 0.079375 | 0.053770 | |
| 999 | 0.189121 | 0.000000 | 0.063040 | 0.099063 | 0.018011 | 0.099063 | 0.072046 | |

| | X7 | X8 | X9 | … | X53 | X54 | X55 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.050077 | 0.046367 | 0.105718 | … | -0.018691 | -0.014300 | -0.005107 | |
| 1 | 0.025224 | 0.010459 | 0.038759 | … | 0.012604 | 0.007689 | 0.011222 | |
| 2 | 0.031960 | 0.021973 | 0.053933 | … | 0.046001 | 0.019587 | 0.046912 | |
| 3 | 0.031758 | 0.012215 | 0.046416 | … | 0.013064 | -0.012255 | 0.020676 | |
| 4 | 0.044081 | 0.029784 | 0.083396 | … | -0.003926 | -0.009265 | 0.010467 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.039117 | 0.018966 | 0.085346 | … | -0.013369 | 0.019280 | -0.003068 | |
| 996 | 0.022522 | 0.011261 | 0.011261 | … | 0.008285 | -0.018503 | 0.014438 | |
| 997 | 0.067856 | 0.050892 | 0.067856 | … | 0.007412 | 0.048272 | 0.026706 | |
| 998 | 0.051210 | 0.071694 | 0.071694 | … | -0.006532 | -0.001009 | -0.017035 | |
| 999 | 0.054034 | 0.009006 | 0.144092 | … | 0.016503 | 0.098336 | 0.032494 | |

| | X56 | X57 | X58 | X59 | X60 | X61 | X62 |
|---|---|---|---|---|---|---|---|
| 0 | -0.004176 | -0.021153 | 0.049961 | 0.035184 | 0.033583 | 0.038343 | efectores |
| 1 | 0.026748 | 0.011662 | 0.017677 | 0.002783 | 0.014210 | 0.009149 | efectores |
| 2 | -0.008757 | 0.035323 | 0.005922 | 0.027951 | -0.029984 | -0.001215 | efectores |
| 3 | -0.006204 | -0.008696 | 0.011364 | 0.023358 | 0.000537 | 0.000788 | efectores |
| 4 | 0.001357 | -0.013032 | 0.017325 | 0.001816 | 0.003830 | -0.001175 | efectores |
| .. | … | … | … | … | … | … | … |
| 995 | -0.015507 | 0.007569 | -0.000294 | -0.000120 | -0.014466 | 0.007194 | efectores |
| 996 | 0.053554 | 0.090465 | -0.005023 | 0.057761 | 0.009943 | 0.036662 | efectores |
| 997 | -0.102042 | -0.092963 | -0.007675 | 0.049994 | 0.021036 | 0.055308 | efectores |

```
998 -0.022616 -0.015113 -0.030426  0.019116 -0.002549 -0.006591  efectores
999 -0.064448 -0.021746  0.014632  0.015797 -0.049462 -0.045191  efectores

[1000 rows x 63 columns]
```

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

| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.060047 | 0.011121 | 0.041733 | 0.042605 | 0.028111 | |
| std | 0.043541 | 0.012086 | 0.028469 | 0.064460 | 0.034769 | |
| min | -0.931401 | 0.000000 | -0.465701 | -1.862803 | -0.931401 | |
| 25% | 0.040181 | 0.004391 | 0.026815 | 0.028910 | 0.018524 | |
| 50% | 0.057336 | 0.008458 | 0.041567 | 0.042494 | 0.026959 | |
| 75% | 0.076705 | 0.014699 | 0.054144 | 0.057440 | 0.036984 | |
| max | 0.234844 | 0.158330 | 0.361046 | 0.240697 | 0.156563 | |

| | X5 | X6 | X7 | X8 | X9 | … | \ |
|---|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | … | |
| mean | 0.048459 | 0.019324 | 0.038485 | 0.035770 | 0.067502 | … | |
| std | 0.079860 | 0.013488 | 0.053230 | 0.028426 | 0.037135 | … | |
| min | -2.328503 | 0.000000 | -1.397102 | -0.465701 | -0.000000 | … | |
| 25% | 0.031996 | 0.009669 | 0.025477 | 0.021200 | 0.042736 | … | |
| 50% | 0.046238 | 0.017446 | 0.036669 | 0.033306 | 0.063131 | … | |
| 75% | 0.067149 | 0.025661 | 0.050386 | 0.046668 | 0.086802 | … | |
| max | 0.253460 | 0.112611 | 0.601743 | 0.361046 | 0.391407 | … | |

| | X52 | X53 | X54 | X55 | X56 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.002635 | 0.008707 | 0.001477 | 0.005570 | 0.004787 | |
| std | 0.052259 | 0.033809 | 0.053869 | 0.055200 | 0.087820 | |
| min | -1.027743 | -0.665171 | -1.116354 | -1.314135 | -0.225330 | |
| 25% | -0.009415 | -0.003111 | -0.011483 | -0.004972 | -0.010338 | |
| 50% | 0.005436 | 0.009903 | 0.004496 | 0.008480 | 0.004217 | |
| 75% | 0.017487 | 0.022124 | 0.017540 | 0.021720 | 0.017075 | |
| max | 0.560985 | 0.137779 | 0.514599 | 0.300765 | 2.582127 | |

| | X57 | X58 | X59 | X60 | X61 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.009300 | 0.001068 | 0.008373 | 0.005188 | 0.013488 |
| std | 0.091162 | 0.051279 | 0.073251 | 0.072919 | 0.154821 |
| min | -1.244742 | -0.635190 | -0.861095 | -0.411157 | -0.172574 |
| 25% | -0.003572 | -0.010291 | -0.004495 | -0.008989 | -0.004668 |
| 50% | 0.009059 | 0.004110 | 0.008329 | 0.005486 | 0.008101 |
| 75% | 0.022685 | 0.015694 | 0.021225 | 0.017685 | 0.022752 |

```
max       2.445692     1.120254     1.943489     1.787151     4.668298

[8 rows x 62 columns]



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.055188  0.050171  0.020068  0.040137  0.040137  0.045154  0.030102
1    0.069598  0.026514  0.061312  0.039770  0.038113  0.039770  0.019885
2    0.066453  0.016613  0.016613  0.033226  0.083066  0.066453  0.016613
3    0.096520  0.020683  0.051707  0.051707  0.027577  0.087902  0.027577
4    0.042463  0.006705  0.023467  0.013409  0.023467  0.032406  0.008940
..        …         …         …         …         …         …         …
995  0.067516  0.010387  0.028045  0.035316  0.017658  0.039471  0.008310
996  0.036824  0.003069  0.015855  0.017900  0.021480  0.034266  0.006649
997  0.079367  0.003175  0.069843  0.053969  0.038096  0.084129  0.036509
998  0.085225  0.004607  0.069102  0.018427  0.039158  0.110563  0.025337
999  0.122186  0.010781  0.066483  0.057499  0.028750  0.062890  0.023359

           X7        X8        X9  …       X53       X54       X55  \
0    0.050171  0.025085  0.110376  … -0.027427 -0.061146 -0.022410
1    0.057998  0.036456  0.099426  … -0.011628 -0.021242  0.001165
2    0.132905  0.049840  0.099679  …  0.078209 -0.062138 -0.014585
3    0.055154  0.046537  0.077561  …  0.004255 -0.001013  0.025207
4    0.034641  0.016762  0.036876  …  0.019914  0.013577  0.013244
..        …         …         …  …         …         …         …
995  0.028045  0.039471  0.067516  …  0.014803  0.016769  0.015764
996  0.026083  0.014320  0.041426  …  0.006998  0.016691  0.012871
997  0.052382  0.061906  0.098415  … -0.028550 -0.030457 -0.052795
998  0.036854  0.032247  0.080619  …  0.030099 -0.013890 -0.014250
999  0.046718  0.068280  0.080858  … -0.030165 -0.004853  0.002878

           X56       X57       X58       X59       X60       X61           X62
0    0.017173  0.015838 -0.016265  0.011379  0.019277  0.032923  no_efectores
1   -0.002119  0.009882  0.023807  0.016421 -0.008278 -0.007694  no_efectores
2    0.202052  0.137924 -0.054567 -0.033119 -0.379108 -0.245737  no_efectores
3   -0.013701 -0.002674 -0.004146 -0.016793  0.037638  0.026415  no_efectores
4    0.004024  0.011442  0.006516  0.001865  0.017630  0.005677  no_efectores
..        …         …         …         …         …         …           …
995  0.000440  0.004146  0.014247  0.018082  0.023511  0.018729  no_efectores
996  0.004478  0.006489  0.011128  0.005973  0.013254  0.002317  no_efectores
997  0.045530  0.018321 -0.002559 -0.019482  0.044619  0.034597  no_efectores
998  0.015616  0.021637 -0.001491 -0.009260  0.001962 -0.002965  no_efectores
```

```
999  0.008517  0.018972 -0.008140 -0.014735 -0.003511  0.028478  no_efectores

[1000 rows x 63 columns]
```

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

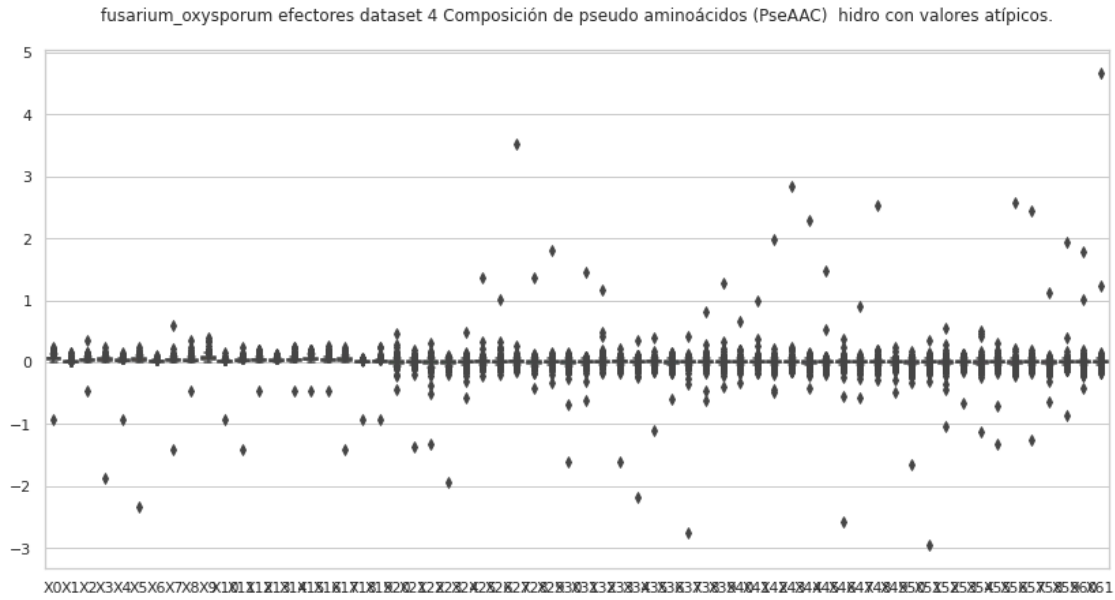
| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.049605 | 0.001351 | 0.033099 | 0.034445 | 0.007579 | |
| std | 0.338407 | 0.335845 | 0.337492 | 0.337977 | 0.671498 | |
| min | -10.599309 | -10.599309 | -10.599309 | -10.599309 | -21.198619 | |
| 25% | 0.040349 | 0.004215 | 0.028448 | 0.029185 | 0.017680 | |
| 50% | 0.056221 | 0.009265 | 0.041956 | 0.042388 | 0.027137 | |
| 75% | 0.075632 | 0.015486 | 0.055417 | 0.057734 | 0.036711 | |
| max | 0.306858 | 0.217923 | 0.435846 | 0.653769 | 0.217923 | |

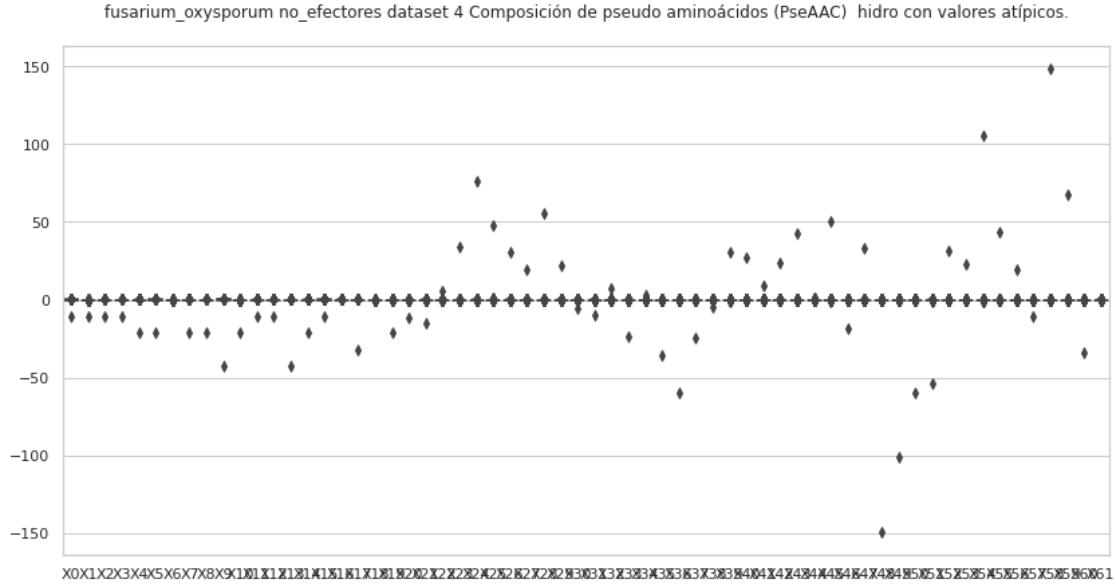
| | X5 | X6 | X7 | X8 | X9 | … | \ |
|---|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | … | |
| mean | 0.029518 | 0.019730 | 0.019524 | 0.016217 | 0.025898 | … | |
| std | 0.672612 | 0.023991 | 0.672162 | 0.671976 | 1.343507 | … | |
| min | -21.198619 | 0.000000 | -21.198619 | -21.198619 | -42.397237 | … | |
| 25% | 0.031790 | 0.010054 | 0.025347 | 0.021713 | 0.043002 | … | |
| 50% | 0.045973 | 0.016843 | 0.037168 | 0.034388 | 0.064595 | … | |
| 75% | 0.065538 | 0.025246 | 0.051657 | 0.048678 | 0.085118 | … | |
| max | 0.423094 | 0.653769 | 0.306858 | 0.385348 | 0.653769 | … | |

| | X52 | X53 | X54 | X55 | X56 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.033050 | 0.031216 | 0.104250 | 0.049292 | 0.021033 | |
| std | 0.985944 | 0.726831 | 3.324688 | 1.373058 | 0.603628 | |
| min | -0.695660 | -0.433002 | -1.281566 | -0.674340 | -0.261991 | |
| 25% | -0.008377 | -0.002375 | -0.009920 | -0.003600 | -0.012585 | |
| 50% | 0.004114 | 0.009657 | 0.003746 | 0.009104 | 0.003507 | |
| 75% | 0.016274 | 0.021525 | 0.015287 | 0.021712 | 0.015494 | |
| max | 31.136531 | 22.967997 | 105.116077 | 43.403869 | 19.047448 | |

| | X57 | X58 | X59 | X60 | X61 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | -0.002310 | 0.150384 | 0.075908 | -0.031944 | 0.009726 |
| std | 0.336698 | 4.691304 | 2.129207 | 1.075535 | 0.032453 |
| min | -10.588873 | -0.545577 | -0.615314 | -33.984084 | -0.245737 |
| 25% | -0.003682 | -0.010560 | -0.002537 | -0.009588 | -0.003498 |
| 50% | 0.008525 | 0.005553 | 0.009597 | 0.005466 | 0.010628 |
| 75% | 0.021871 | 0.017140 | 0.022753 | 0.018199 | 0.024251 |
| max | 0.501267 | 148.347934 | 67.331053 | 0.255254 | 0.349970 |

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

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.081607 | 0.012983 | 0.070479 | 0.076043 | 0.031530 | 0.059350 | 0.022256 | |
| 1 | 0.030761 | 0.003691 | 0.009228 | 0.014765 | 0.017226 | 0.032607 | 0.007383 | |
| 2 | 0.045943 | 0.000000 | 0.051936 | 0.051936 | 0.021973 | 0.049938 | 0.013983 | |
| 3 | 0.070845 | 0.013436 | 0.058631 | 0.043973 | 0.035423 | 0.065959 | 0.006107 | |
| 4 | 0.052421 | 0.022636 | 0.054803 | 0.041698 | 0.041698 | 0.066717 | 0.017871 | |
| .. | … | … | … | … | … | … | … | |
| 994 | 0.087449 | 0.020821 | 0.049971 | 0.058299 | 0.041642 | 0.074956 | 0.041642 | |
| 995 | 0.061639 | 0.009483 | 0.028449 | 0.040302 | 0.039117 | 0.030819 | 0.013039 | |
| 996 | 0.045044 | 0.022522 | 0.033783 | 0.011261 | 0.011261 | 0.056304 | 0.011261 | |
| 997 | 0.045237 | 0.011309 | 0.050892 | 0.062201 | 0.033928 | 0.084820 | 0.016964 | |
| 998 | 0.089617 | 0.025605 | 0.061452 | 0.064012 | 0.071694 | 0.079375 | 0.053770 | |

| | X7 | X8 | X9 | … | X53 | X54 | X55 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.050077 | 0.046367 | 0.105718 | … | -0.018691 | -0.014300 | -0.005107 | |
| 1 | 0.025224 | 0.010459 | 0.038759 | … | 0.012604 | 0.007689 | 0.011222 | |
| 2 | 0.031960 | 0.021973 | 0.053933 | … | 0.046001 | 0.019587 | 0.046912 | |
| 3 | 0.031758 | 0.012215 | 0.046416 | … | 0.013064 | -0.012255 | 0.020676 | |
| 4 | 0.044081 | 0.029784 | 0.083396 | … | -0.003926 | -0.009265 | 0.010467 | |
| .. | … | … | … | … | … | … | … | |
| 994 | 0.037478 | 0.029150 | 0.099942 | … | 0.025472 | 0.040690 | -0.015723 | |
| 995 | 0.039117 | 0.018966 | 0.085346 | … | -0.013369 | 0.019280 | -0.003068 | |
| 996 | 0.022522 | 0.011261 | 0.011261 | … | 0.008285 | -0.018503 | 0.014438 | |
| 997 | 0.067856 | 0.050892 | 0.067856 | … | 0.007412 | 0.048272 | 0.026706 | |
| 998 | 0.051210 | 0.071694 | 0.071694 | … | -0.006532 | -0.001009 | -0.017035 | |

| | X56 | X57 | X58 | X59 | X60 | X61 | X62 |
|---|---|---|---|---|---|---|---|
| 0 | -0.004176 | -0.021153 | 0.049961 | 0.035184 | 0.033583 | 0.038343 | efectores |
| 1 | 0.026748 | 0.011662 | 0.017677 | 0.002783 | 0.014210 | 0.009149 | efectores |
| 2 | -0.008757 | 0.035323 | 0.005922 | 0.027951 | -0.029984 | -0.001215 | efectores |
| 3 | -0.006204 | -0.008696 | 0.011364 | 0.023358 | 0.000537 | 0.000788 | efectores |
| 4 | 0.001357 | -0.013032 | 0.017325 | 0.001816 | 0.003830 | -0.001175 | efectores |
| .. | … | … | … | … | … | … | … |
| 994 | -0.033239 | 0.009305 | -0.000858 | -0.015616 | 0.045670 | 0.038073 | efectores |
| 995 | -0.015507 | 0.007569 | -0.000294 | -0.000120 | -0.014466 | 0.007194 | efectores |
| 996 | 0.053554 | 0.090465 | -0.005023 | 0.057761 | 0.009943 | 0.036662 | efectores |
| 997 | -0.102042 | -0.092963 | -0.007675 | 0.049994 | 0.021036 | 0.055308 | efectores |
| 998 | -0.022616 | -0.015113 | -0.030426 | 0.019116 | -0.002549 | -0.006591 | efectores |

```
[925 rows x 63 columns]
```

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

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 925.000000 | 925.000000 | 925.000000 | 925.000000 | 925.000000 | 925.000000 | |
| mean | 0.058058 | 0.010093 | 0.040457 | 0.042412 | 0.027679 | 0.049146 | |
| std | 0.024923 | 0.007900 | 0.018912 | 0.019214 | 0.014351 | 0.024449 | |
| min | 0.004298 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.002419 | |
| 25% | 0.039690 | 0.004411 | 0.026051 | 0.028057 | 0.018417 | 0.031517 | |
| 50% | 0.055763 | 0.008295 | 0.040383 | 0.041586 | 0.026525 | 0.045034 | |
| 75% | 0.073839 | 0.014060 | 0.053050 | 0.054636 | 0.035756 | 0.065591 | |
| max | 0.166555 | 0.045062 | 0.114750 | 0.113714 | 0.081123 | 0.194610 | |

| | X6 | X7 | X8 | X9 | … | X52 | \ |
|---|---|---|---|---|---|---|---|
| count | 925.000000 | 925.000000 | 925.000000 | 925.000000 | … | 925.000000 | |
| mean | 0.018058 | 0.037513 | 0.034021 | 0.064063 | … | 0.003759 | |
| std | 0.011010 | 0.018435 | 0.017104 | 0.030480 | … | 0.024817 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.002038 | … | -0.119361 | |
| 25% | 0.009523 | 0.024738 | 0.020796 | 0.041800 | … | -0.008801 | |
| 50% | 0.016995 | 0.035885 | 0.032244 | 0.061149 | … | 0.005429 | |
| 75% | 0.024582 | 0.048902 | 0.045301 | 0.084449 | … | 0.016822 | |
| max | 0.058864 | 0.104872 | 0.103738 | 0.167569 | … | 0.124240 | |

| | X53 | X54 | X55 | X56 | X57 | X58 | \ |
|---|---|---|---|---|---|---|---|
| count | 925.000000 | 925.000000 | 925.000000 | 925.000000 | 925.000000 | 925.000000 | |
| mean | 0.009229 | 0.002417 | 0.007766 | 0.004294 | 0.009490 | 0.002489 | |
| std | 0.021536 | 0.023115 | 0.020920 | 0.024203 | 0.022335 | 0.022845 | |
| min | -0.068884 | -0.088615 | -0.071220 | -0.102042 | -0.092963 | -0.102460 | |
| 25% | -0.002625 | -0.010260 | -0.003127 | -0.009060 | -0.001574 | -0.008863 | |
| 50% | 0.009899 | 0.004513 | 0.008551 | 0.005112 | 0.009493 | 0.004167 | |
| 75% | 0.021448 | 0.016909 | 0.021293 | 0.017054 | 0.022598 | 0.015220 | |
| max | 0.101704 | 0.083299 | 0.094712 | 0.095127 | 0.127244 | 0.083837 | |

| | X59 | X60 | X61 |
|---|---|---|---|
| count | 925.000000 | 925.000000 | 925.000000 |
| mean | 0.007694 | 0.003494 | 0.008108 |
| std | 0.021435 | 0.023592 | 0.021493 |
| min | -0.092973 | -0.139426 | -0.100133 |
| 25% | -0.003821 | -0.008246 | -0.003786 |
| 50% | 0.008387 | 0.005489 | 0.008048 |
| 75% | 0.020442 | 0.017015 | 0.021894 |
| max | 0.091191 | 0.091031 | 0.079847 |

[8 rows x 62 columns]

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.055188  0.050171  0.020068  0.040137  0.040137  0.045154  0.030102
1    0.069598  0.026514  0.061312  0.039770  0.038113  0.039770  0.019885
3    0.096520  0.020683  0.051707  0.051707  0.027577  0.087902  0.027577
4    0.042463  0.006705  0.023467  0.013409  0.023467  0.032406  0.008940
5    0.027685  0.001408  0.025808  0.035192  0.012200  0.025808  0.007508
..        …         …         …         …         …         …
995  0.067516  0.010387  0.028045  0.035316  0.017658  0.039471  0.008310
996  0.036824  0.003069  0.015855  0.017900  0.021480  0.034266  0.006649
997  0.079367  0.003175  0.069843  0.053969  0.038096  0.084129  0.036509
998  0.085225  0.004607  0.069102  0.018427  0.039158  0.110563  0.025337
999  0.122186  0.010781  0.066483  0.057499  0.028750  0.062890  0.023359

           X7        X8        X9  …       X53       X54       X55  \
0    0.050171  0.025085  0.110376  … -0.027427 -0.061146 -0.022410
1    0.057998  0.036456  0.099426  … -0.011628 -0.021242  0.001165
3    0.055154  0.046537  0.077561  …  0.004255 -0.001013  0.025207
4    0.034641  0.016762  0.036876  …  0.019914  0.013577  0.013244
5    0.020646  0.022054  0.032846  …  0.018042 -0.008972  0.013788
..        …         …         …  …       …         …         …
995  0.028045  0.039471  0.067516  …  0.014803  0.016769  0.015764
996  0.026083  0.014320  0.041426  …  0.006998  0.016691  0.012871
997  0.052382  0.061906  0.098415  … -0.028550 -0.030457 -0.052795
998  0.036854  0.032247  0.080619  …  0.030099 -0.013890 -0.014250
999  0.046718  0.068280  0.080858  … -0.030165 -0.004853  0.002878

          X56       X57       X58       X59       X60       X61           X62
0    0.017173  0.015838 -0.016265  0.011379  0.019277  0.032923  no_efectores
1   -0.002119  0.009882  0.023807  0.016421 -0.008278 -0.007694  no_efectores
3   -0.013701 -0.002674 -0.004146 -0.016793  0.037638  0.026415  no_efectores
4    0.004024  0.011442  0.006516  0.001865  0.017630  0.005677  no_efectores
5    0.014389  0.029339  0.000174  0.028620  0.006579  0.022661  no_efectores
..        …         …         …         …         …         …
995  0.000440  0.004146  0.014247  0.018082  0.023511  0.018729  no_efectores
996  0.004478  0.006489  0.011128  0.005973  0.013254  0.002317  no_efectores
997  0.045530  0.018321 -0.002559 -0.019482  0.044619  0.034597  no_efectores
998  0.015616  0.021637 -0.001491 -0.009260  0.001962 -0.002965  no_efectores
999  0.008517  0.018972 -0.008140 -0.014735 -0.003511  0.028478  no_efectores

[954 rows x 63 columns]
```

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

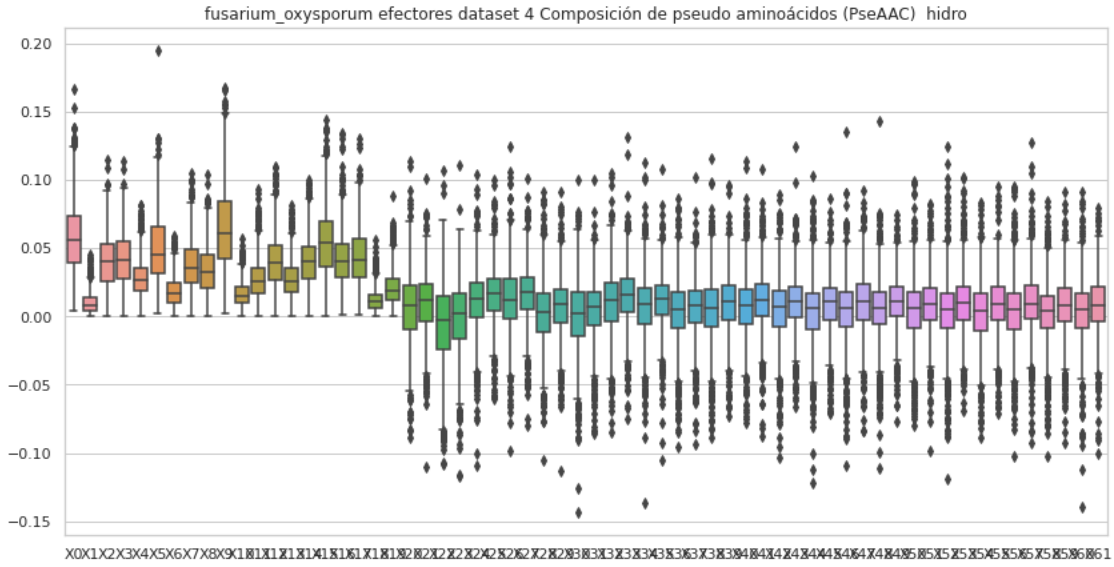
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 954.000000 | 954.000000 | 954.000000 | 954.000000 | 954.000000 | 954.000000 | |
| mean | 0.058279 | 0.011150 | 0.042212 | 0.043020 | 0.027730 | 0.048934 | |
| std | 0.026034 | 0.010996 | 0.019429 | 0.020012 | 0.015019 | 0.025456 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.039947 | 0.004180 | 0.028325 | 0.028692 | 0.017583 | 0.031437 | |
| 50% | 0.055172 | 0.008947 | 0.041338 | 0.042006 | 0.026794 | 0.045111 | |
| 75% | 0.074068 | 0.015073 | 0.054655 | 0.056702 | 0.035786 | 0.064139 | |
| max | 0.223502 | 0.143480 | 0.102525 | 0.133313 | 0.113168 | 0.274504 | |

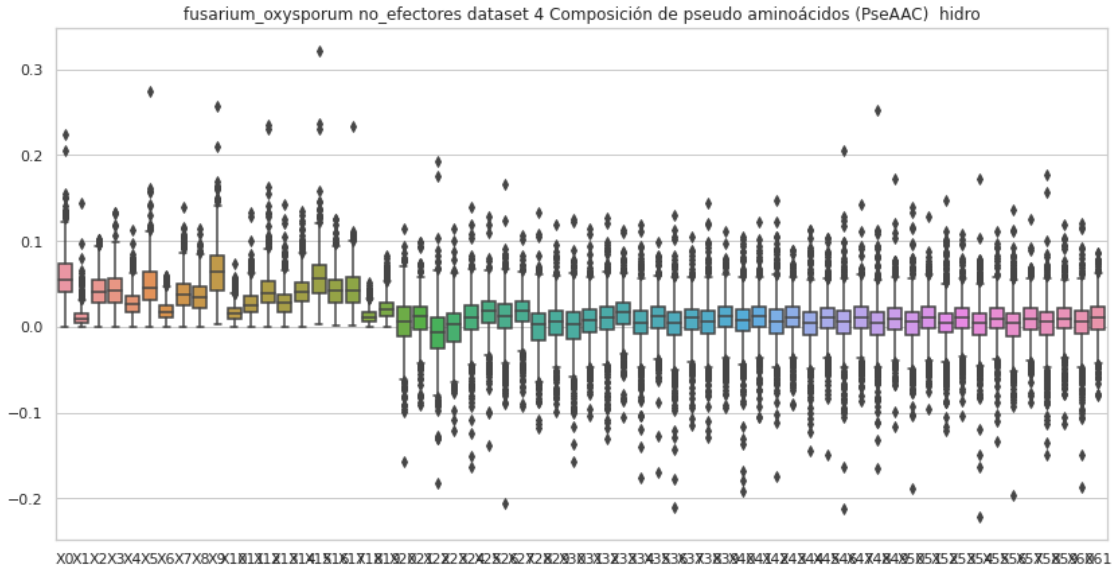
| | X6 | X7 | X8 | X9 | … | X52 | \ |
|---|---|---|---|---|---|---|---|
| count | 954.000000 | 954.000000 | 954.000000 | 954.000000 | … | 954.000000 | |
| mean | 0.018417 | 0.038587 | 0.035841 | 0.065160 | … | 0.003729 | |
| std | 0.011004 | 0.019527 | 0.018122 | 0.030759 | … | 0.026462 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.003083 | … | -0.121255 | |
| 25% | 0.010110 | 0.025245 | 0.021677 | 0.042814 | … | -0.007214 | |
| 50% | 0.016776 | 0.036705 | 0.033979 | 0.063984 | … | 0.004317 | |
| 75% | 0.024856 | 0.050574 | 0.047430 | 0.083698 | … | 0.016137 | |
| max | 0.060163 | 0.140075 | 0.114625 | 0.257205 | … | 0.146565 | |

| | X53 | X54 | X55 | X56 | X57 | X58 | \ |
|---|---|---|---|---|---|---|---|
| count | 954.000000 | 954.000000 | 954.000000 | 954.000000 | 954.000000 | 954.000000 | |
| mean | 0.009017 | 0.002668 | 0.008804 | 0.002210 | 0.008130 | 0.002947 | |
| std | 0.022122 | 0.026070 | 0.022768 | 0.025869 | 0.021967 | 0.027746 | |
| min | -0.103987 | -0.221622 | -0.133998 | -0.196115 | -0.087612 | -0.149492 | |
| 25% | -0.001898 | -0.009037 | -0.002519 | -0.011452 | -0.003036 | -0.009647 | |
| 50% | 0.010052 | 0.004078 | 0.009627 | 0.003697 | 0.008611 | 0.005990 | |
| 75% | 0.021394 | 0.015196 | 0.021799 | 0.015401 | 0.021745 | 0.017073 | |
| max | 0.107642 | 0.173136 | 0.111163 | 0.136809 | 0.125862 | 0.176773 | |

| | X59 | X60 | X61 |
|---|---|---|---|
| count | 954.000000 | 954.000000 | 954.000000 |
| mean | 0.009525 | 0.003618 | 0.009952 |
| std | 0.023660 | 0.027288 | 0.023260 |
| min | -0.113326 | -0.186728 | -0.080876 |
| 25% | -0.001351 | -0.008338 | -0.002841 |
| 50% | 0.009766 | 0.005768 | 0.011007 |
| 75% | 0.022445 | 0.018115 | 0.023906 |
| max | 0.119862 | 0.120251 | 0.086736 |

[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"
```

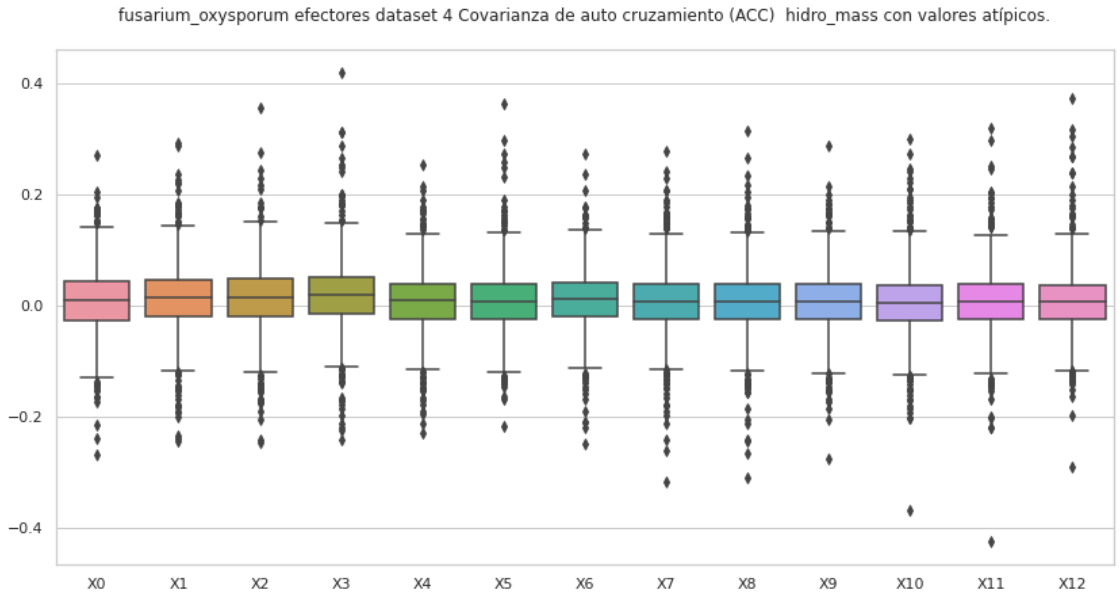
```
comp = "hidro_mass"
df=""

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))
```

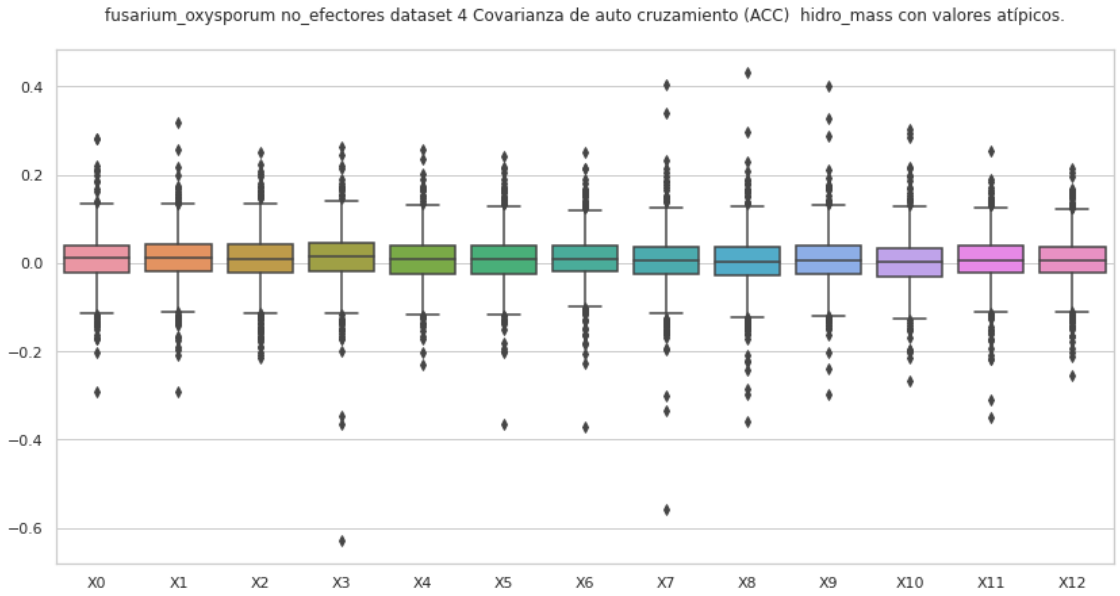
```
efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass efectores fusarium_oxysporum
dataset 4, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.002514 -0.063280  0.045484 -0.017408 -0.088365 -0.049946  0.040684
1    0.025892  0.114121  0.107620  0.178666 -0.060823  0.015862 -0.026677
2   -0.015254 -0.062968  0.064631 -0.002966 -0.016033  0.029092  0.055795
3    0.061754  0.026905  0.005982  0.024770  0.034855  0.073363  0.037252
4   -0.014602 -0.022998 -0.003031 -0.000104 -0.018142  0.018382 -0.012064
..        …         …         …         …         …         …         …
995 -0.008998 -0.021737  0.060831  0.085246 -0.019983 -0.001467 -0.002240
996 -0.269795  0.234772 -0.060960  0.310389  0.061930  0.120341  0.000508
997 -0.016912 -0.103641  0.098139  0.112798  0.005994 -0.105289  0.004904
998 -0.013332  0.074159 -0.041221  0.016725 -0.105068  0.008814 -0.029285
999  0.173667 -0.115145  0.015349 -0.044637  0.044390 -0.013387 -0.147527

           X7        X8        X9       X10       X11       X12        X13
0   -0.104085 -0.017047  0.022324  0.054196  0.024015  0.040474  efectores
```

```
1   -0.028087 -0.035676 -0.045242 -0.064342 -0.083233 -0.047228  efectores
2   -0.017317 -0.068332  0.138964  0.026007 -0.072313  0.016092  efectores
3   -0.008986  0.016100 -0.036955  0.027629  0.041620  0.055806  efectores
4    0.004753 -0.033428  0.011762  0.016421 -0.039024  0.008058  efectores
..        …        …        …        …        …        …        …
995  0.046816  0.025540  0.011101  0.106608  0.054747  0.006770  efectores
996 -0.009896  0.092464  0.213479 -0.369339  0.194116 -0.084821  efectores
997 -0.029013 -0.009154 -0.017590  0.116311 -0.068738 -0.105141  efectores
998  0.004101  0.015825 -0.013806  0.019726 -0.072498  0.007408  efectores
999 -0.066191  0.126133 -0.031623  0.079846  0.249737  0.072723  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008488     0.012432     0.012773     0.018285     0.006988
std       0.058712     0.061115     0.059194     0.062180     0.056393
min      -0.269795    -0.244145    -0.247379    -0.241744    -0.231360
25%      -0.026572    -0.020182    -0.020019    -0.014492    -0.024706
50%       0.008318     0.013110     0.013200     0.019840     0.008713
75%       0.042810     0.045988     0.048318     0.050658     0.038652
max       0.269494     0.292543     0.354185     0.418517     0.252681

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008204     0.010137     0.006518     0.005935     0.006409
std       0.057292     0.054737     0.059158     0.058554     0.056531
min      -0.217379    -0.248852    -0.319175    -0.310764    -0.276546
25%      -0.025244    -0.021590    -0.024183    -0.025424    -0.026020
50%       0.006213     0.012046     0.007537     0.006864     0.006019
75%       0.038562     0.041350     0.037257     0.037247     0.038248
max       0.362091     0.271326     0.276864     0.312919     0.287288

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.004459     0.007373     0.006891
std       0.061024     0.059668     0.059254
min      -0.369339    -0.425312    -0.291072
25%      -0.028105    -0.026330    -0.024872
50%       0.004064     0.005701     0.005384
75%       0.036793     0.038414     0.036706
max       0.299865     0.318647     0.371734
```

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.078359  0.043146  0.079055  0.031518  0.015503  0.061494  0.017258
1   -0.001790  0.068715 -0.037725  0.031892  0.028435  0.002426  0.031217
2   -0.094241 -0.167353 -0.029809  0.050651  0.011501 -0.037355  0.013251
3    0.016754  0.009699  0.034998  0.008028  0.019328 -0.014480  0.055154
4   -0.003720  0.121159  0.044997 -0.059388  0.058817 -0.011158  0.021339
..        …         …         …         …         …         …
995  0.008553  0.016173  0.059000  0.023291 -0.016736 -0.031267 -0.028933
996 -0.008874  0.046922  0.033661 -0.000193 -0.006447  0.002693  0.057325
997 -0.009390 -0.085835  0.012470 -0.023362  0.016244  0.013279 -0.011091
998  0.011906  0.042445  0.021965  0.071344 -0.046389  0.067688  0.079572
999  0.004508  0.078399  0.014382  0.045673  0.086536  0.011496 -0.036729

           X7        X8        X9       X10       X11       X12           X13
0    0.036564  0.092331  0.011791 -0.003005 -0.018494 -0.091950  no_efectores
1    0.022532  0.068664 -0.046674  0.019336 -0.038399  0.040286  no_efectores
2    0.003949 -0.033212  0.003781  0.041692 -0.091715  0.126707  no_efectores
3    0.042914  0.021675  0.057478  0.033847  0.009426  0.005761  no_efectores
4   -0.088845  0.010276 -0.104955 -0.046020  0.016505 -0.073527  no_efectores
..        …         …         …         …         …         …
995  0.021329  0.004369  0.033554  0.031640 -0.017618  0.029623  no_efectores
996  0.022952 -0.045551  0.006032  0.053006 -0.007373  0.027486  no_efectores
997  0.028672  0.006430  0.065755  0.035425 -0.007765 -0.064968  no_efectores
998 -0.004450 -0.010223 -0.035157  0.036381 -0.087081 -0.006716  no_efectores
999  0.059352 -0.008173  0.033531  0.001612 -0.036825  0.005633  no_efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009164     0.010617     0.007901     0.013133     0.006941
std       0.056337     0.054463     0.057374     0.061770     0.054767
min      -0.290225    -0.291437    -0.214284    -0.628848    -0.229668
25%      -0.022044    -0.019032    -0.021266    -0.017761    -0.023927
50%       0.011501     0.011868     0.010244     0.014593     0.008429
75%       0.040371     0.042206     0.041180     0.046790     0.039353
```

```
max       0.282012     0.317877     0.251794     0.262247     0.256692

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008423     0.010483     0.003937     0.003190     0.006631
std       0.056301     0.053399     0.061897     0.060231     0.056975
min      -0.363760    -0.372215    -0.558922    -0.358552    -0.296940
25%      -0.023947    -0.017593    -0.026159    -0.027591    -0.025618
50%       0.007369     0.009473     0.005401     0.004165     0.005931
75%       0.038818     0.038195     0.036283     0.036065     0.038010
max       0.242571     0.252252     0.404301     0.431695     0.402646

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.002885     0.006077     0.006774
std       0.058082     0.055700     0.055194
min      -0.268277    -0.349795    -0.255679
25%      -0.030221    -0.022027    -0.022753
50%       0.003146     0.006522     0.006867
75%       0.033845     0.038290     0.035586
max       0.302242     0.255379     0.214972
```






### 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 fusarium_oxysporum
dataset 4, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.002514 -0.063280  0.045484 -0.017408 -0.088365 -0.049946  0.040684
1    0.025892  0.114121  0.107620  0.178666 -0.060823  0.015862 -0.026677
2   -0.015254 -0.062968  0.064631 -0.002966 -0.016033  0.029092  0.055795
3    0.061754  0.026905  0.005982  0.024770  0.034855  0.073363  0.037252
4   -0.014602 -0.022998 -0.003031 -0.000104 -0.018142  0.018382 -0.012064
..        …         …         …         …         …         …         …
993 -0.127110  0.008322  0.017766 -0.037289 -0.021772  0.000265  0.005239
994 -0.084425 -0.070477  0.012906 -0.066162  0.012945 -0.011572  0.092989
995 -0.008998 -0.021737  0.060831  0.085246 -0.019983 -0.001467 -0.002240
997 -0.016912 -0.103641  0.098139  0.112798  0.005994 -0.105289  0.004904
998 -0.013332  0.074159 -0.041221  0.016725 -0.105068  0.008814 -0.029285

           X7        X8        X9       X10       X11       X12        X13
0   -0.104085 -0.017047  0.022324  0.054196  0.024015  0.040474  efectores
1   -0.028087 -0.035676 -0.045242 -0.064342 -0.083233 -0.047228  efectores
2   -0.017317 -0.068332  0.138964  0.026007 -0.072313  0.016092  efectores
3   -0.008986  0.016100 -0.036955  0.027629  0.041620  0.055806  efectores
4    0.004753 -0.033428  0.011762  0.016421 -0.039024  0.008058  efectores
..        …         …         …         …         …         …         …
993  0.003473 -0.023894 -0.031830  0.002103  0.048642  0.077085  efectores
```

```
994 -0.028423  0.029884  0.002559  0.045542 -0.048364 -0.056534  efectores
995  0.046816  0.025540  0.011101  0.106608  0.054747  0.006770  efectores
997 -0.029013 -0.009154 -0.017590  0.116311 -0.068738 -0.105141  efectores
998  0.004101  0.015825 -0.013806  0.019726 -0.072498  0.007408  efectores

[905 rows x 14 columns]
```

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

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | |
| mean | 0.009330 | 0.012878 | 0.013908 | 0.018661 | 0.007357 | 0.005279 | |
| std | 0.051520 | 0.051036 | 0.049756 | 0.048579 | 0.049345 | 0.048277 | |
| min | -0.164171 | -0.151425 | -0.151210 | -0.126878 | -0.153601 | -0.146151 | |
| 25% | -0.023471 | -0.018033 | -0.017315 | -0.012920 | -0.023490 | -0.023456 | |
| 50% | 0.009164 | 0.013780 | 0.013935 | 0.019525 | 0.008332 | 0.005835 | |
| 75% | 0.042192 | 0.044557 | 0.047293 | 0.048207 | 0.037157 | 0.034534 | |
| max | 0.163835 | 0.184508 | 0.183659 | 0.199434 | 0.171045 | 0.162536 | |

| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | |
| mean | 0.011162 | 0.006277 | 0.003994 | 0.005564 | 0.003667 | 0.006412 | |
| std | 0.046986 | 0.050635 | 0.048707 | 0.048765 | 0.049894 | 0.050166 | |
| min | -0.140390 | -0.159230 | -0.155967 | -0.155608 | -0.171278 | -0.168408 | |
| 25% | -0.019158 | -0.023345 | -0.025164 | -0.024795 | -0.025271 | -0.025271 | |
| 50% | 0.012373 | 0.006627 | 0.004247 | 0.005519 | 0.003602 | 0.005689 | |
| 75% | 0.040565 | 0.035380 | 0.033315 | 0.033670 | 0.033345 | 0.035376 | |
| max | 0.173752 | 0.176126 | 0.173473 | 0.172184 | 0.183741 | 0.184695 | |

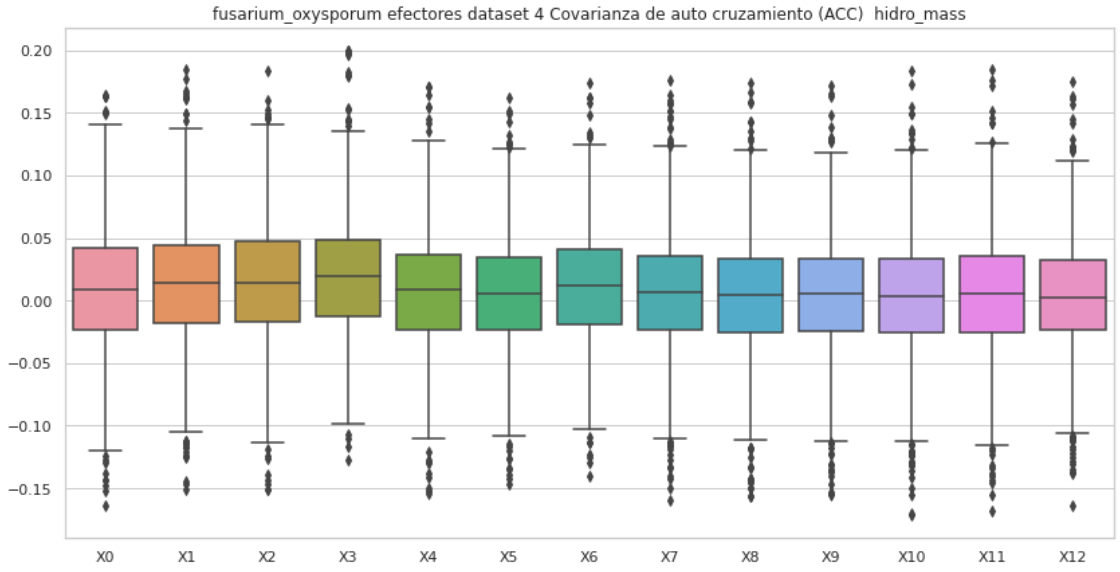
| | X12 |
|---|---|
| count | 905.000000 |
| mean | 0.003269 |
| std | 0.047805 |
| min | -0.163474 |
| 25% | -0.023811 |
| 50% | 0.002835 |
| 75% | 0.031888 |
| max | 0.174366 |

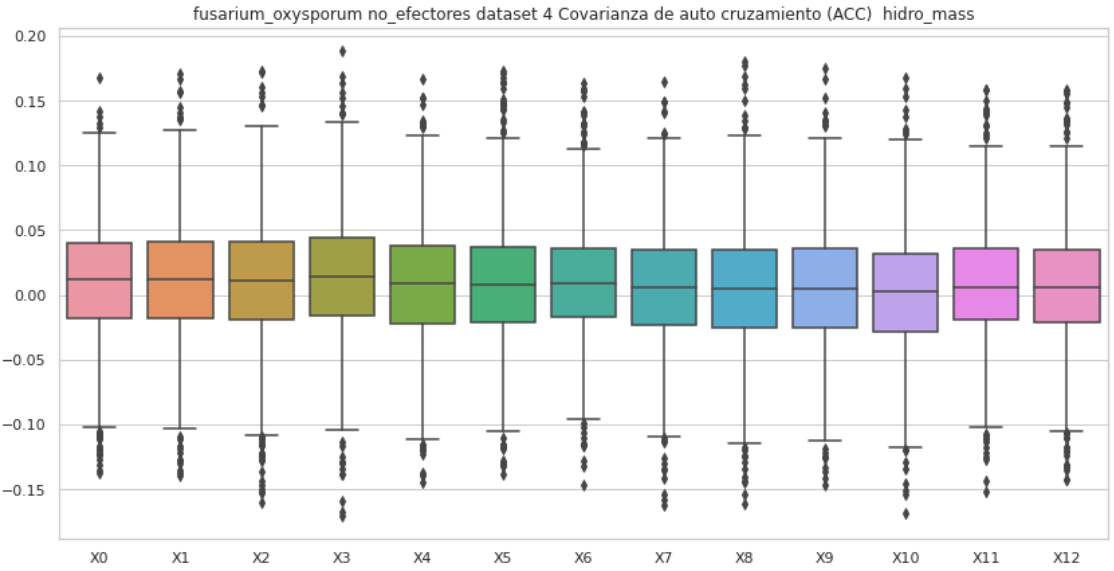
no_efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores fusarium_oxysporum
dataset 4, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.078359  0.043146  0.079055  0.031518  0.015503  0.061494  0.017258
1   -0.001790  0.068715 -0.037725  0.031892  0.028435  0.002426  0.031217
3    0.016754  0.009699  0.034998  0.008028  0.019328 -0.014480  0.055154
4   -0.003720  0.121159  0.044997 -0.059388  0.058817 -0.011158  0.021339
5    0.019457  0.022436  0.009165  0.033283 -0.008717  0.028988  0.005616
..        …         …         …         …         …         …
995  0.008553  0.016173  0.059000  0.023291 -0.016736 -0.031267 -0.028933
996 -0.008874  0.046922  0.033661 -0.000193 -0.006447  0.002693  0.057325
997 -0.009390 -0.085835  0.012470 -0.023362  0.016244  0.013279 -0.011091
998  0.011906  0.042445  0.021965  0.071344 -0.046389  0.067688  0.079572
999  0.004508  0.078399  0.014382  0.045673  0.086536  0.011496 -0.036729

           X7        X8        X9       X10       X11       X12           X13
0    0.036564  0.092331  0.011791 -0.003005 -0.018494 -0.091950  no_efectores
1    0.022532  0.068664 -0.046674  0.019336 -0.038399  0.040286  no_efectores
3    0.042914  0.021675  0.057478  0.033847  0.009426  0.005761  no_efectores
4   -0.088845  0.010276 -0.104955 -0.046020  0.016505 -0.073527  no_efectores
5   -0.010629  0.042915 -0.050987 -0.028890 -0.029756  0.033284  no_efectores
..        …         …         …         …         …         …
995  0.021329  0.004369  0.033554  0.031640 -0.017618  0.029623  no_efectores
996  0.022952 -0.045551  0.006032  0.053006 -0.007373  0.027486  no_efectores
997  0.028672  0.006430  0.065755  0.035425 -0.007765 -0.064968  no_efectores
998 -0.004450 -0.010223 -0.035157  0.036381 -0.087081 -0.006716  no_efectores
999  0.059352 -0.008173  0.033531  0.001612 -0.036825  0.005633  no_efectores

[914 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.009462    0.010698    0.008888    0.014164    0.006960    0.008957
std      0.047941    0.047298    0.049793    0.050565    0.048563    0.049743
min     -0.137389   -0.139878   -0.160116   -0.170587   -0.144798   -0.138308
25%     -0.017987   -0.018175   -0.019109   -0.016030   -0.022095   -0.021590
50%      0.011923    0.011668    0.011079    0.014254    0.008775    0.007692
75%      0.039678    0.040644    0.040683    0.043772    0.037760    0.036857
max      0.167333    0.170396    0.173373    0.188423    0.166366    0.172655

                X6          X7          X8          X9         X10         X11  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.010410    0.005279    0.004987    0.005944    0.002252    0.007167
std      0.044675    0.047904    0.048714    0.048361    0.048930    0.047170
```

| | | | | | | |
|---|---|---|---|---|---|---|
| min | -0.146535 | -0.162299 | -0.161527 | -0.146800 | -0.168810 | -0.151423 |
| 25% | -0.016585 | -0.022787 | -0.025046 | -0.025266 | -0.028633 | -0.019417 |
| 50% | 0.009309 | 0.005825 | 0.004582 | 0.004605 | 0.003080 | 0.006335 |
| 75% | 0.036332 | 0.035188 | 0.035306 | 0.036051 | 0.031791 | 0.036156 |
| max | 0.164105 | 0.164658 | 0.180551 | 0.174836 | 0.167291 | 0.158740 |

| | X12 |
|---|---|
| count | 914.000000 |
| mean | 0.006999 |
| std | 0.048557 |
| min | -0.143016 |
| 25% | -0.021542 |
| 50% | 0.006225 |
| 75% | 0.034905 |
| max | 0.158854 |






# 7 Covarianza de auto cruzamiento (ACC) mass

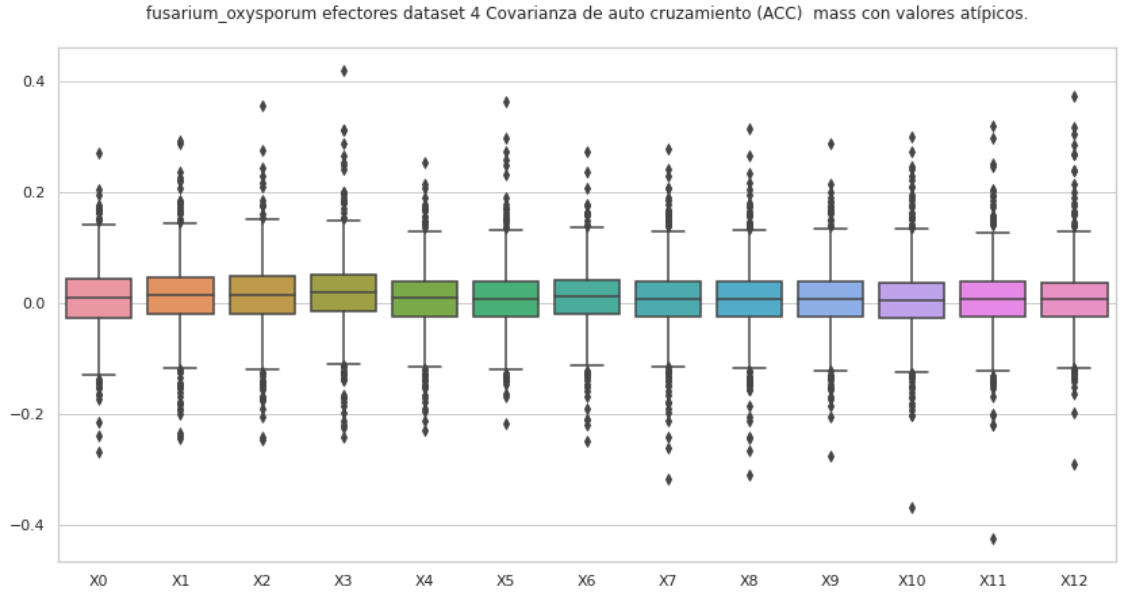
[13]:
```
#mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "mass"
df=""

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")
```

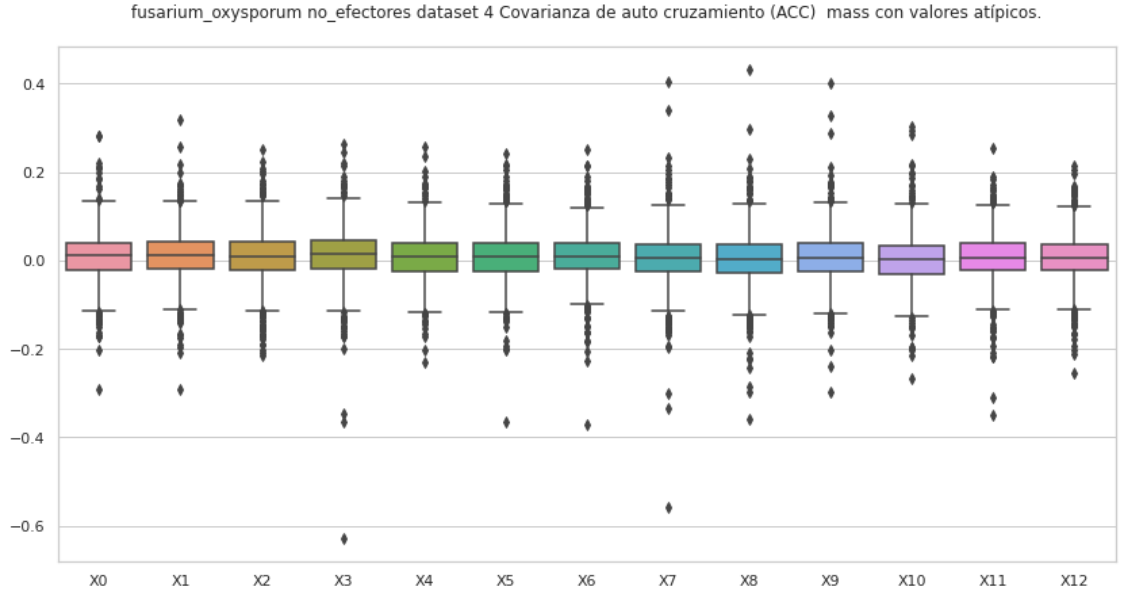
```
    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.002514 -0.063280  0.045484 -0.017408 -0.088365 -0.049946  0.040684
1    0.025892  0.114121  0.107620  0.178666 -0.060823  0.015862 -0.026677
2   -0.015254 -0.062968  0.064631 -0.002966 -0.016033  0.029092  0.055795
3    0.061754  0.026905  0.005982  0.024770  0.034855  0.073363  0.037252
4   -0.014602 -0.022998 -0.003031 -0.000104 -0.018142  0.018382 -0.012064
..        …         …         …         …         …         …         …
995 -0.008998 -0.021737  0.060831  0.085246 -0.019983 -0.001467 -0.002240
996 -0.269795  0.234772 -0.060960  0.310389  0.061930  0.120341  0.000508
997 -0.016912 -0.103641  0.098139  0.112798  0.005994 -0.105289  0.004904
998 -0.013332  0.074159 -0.041221  0.016725 -0.105068  0.008814 -0.029285
999  0.173667 -0.115145  0.015349 -0.044637  0.044390 -0.013387 -0.147527

           X7        X8        X9       X10       X11       X12         X13
0   -0.104085 -0.017047  0.022324  0.054196  0.024015  0.040474  efectores
1   -0.028087 -0.035676 -0.045242 -0.064342 -0.083233 -0.047228  efectores
2   -0.017317 -0.068332  0.138964  0.026007 -0.072313  0.016092  efectores
3   -0.008986  0.016100 -0.036955  0.027629  0.041620  0.055806  efectores
4    0.004753 -0.033428  0.011762  0.016421 -0.039024  0.008058  efectores
..        …         …         …         …         …         …         …
995  0.046816  0.025540  0.011101  0.106608  0.054747  0.006770  efectores
996 -0.009896  0.092464  0.213479 -0.369339  0.194116 -0.084821  efectores
997 -0.029013 -0.009154 -0.017590  0.116311 -0.068738 -0.105141  efectores
998  0.004101  0.015825 -0.013806  0.019726 -0.072498  0.007408  efectores
999 -0.066191  0.126133 -0.031623  0.079846  0.249737  0.072723  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008488     0.012432     0.012773     0.018285     0.006988
```

```
std       0.058712     0.061115     0.059194     0.062180     0.056393
min      -0.269795    -0.244145    -0.247379    -0.241744    -0.231360
25%      -0.026572    -0.020182    -0.020019    -0.014492    -0.024706
50%       0.008318     0.013110     0.013200     0.019840     0.008713
75%       0.042810     0.045988     0.048318     0.050658     0.038652
max       0.269494     0.292543     0.354185     0.418517     0.252681

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008204     0.010137     0.006518     0.005935     0.006409
std       0.057292     0.054737     0.059158     0.058554     0.056531
min      -0.217379    -0.248852    -0.319175    -0.310764    -0.276546
25%      -0.025244    -0.021590    -0.024183    -0.025424    -0.026020
50%       0.006213     0.012046     0.007537     0.006864     0.006019
75%       0.038562     0.041350     0.037257     0.037247     0.038248
max       0.362091     0.271326     0.276864     0.312919     0.287288

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.004459     0.007373     0.006891
std       0.061024     0.059668     0.059254
min      -0.369339    -0.425312    -0.291072
25%      -0.028105    -0.026330    -0.024872
50%       0.004064     0.005701     0.005384
75%       0.036793     0.038414     0.036706
max       0.299865     0.318647     0.371734



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.078359  0.043146  0.079055  0.031518  0.015503  0.061494  0.017258
1   -0.001790  0.068715 -0.037725  0.031892  0.028435  0.002426  0.031217
2   -0.094241 -0.167353 -0.029809  0.050651  0.011501 -0.037355  0.013251
3    0.016754  0.009699  0.034998  0.008028  0.019328 -0.014480  0.055154
4   -0.003720  0.121159  0.044997 -0.059388  0.058817 -0.011158  0.021339
..        …         …         …         …         …         …         …
995  0.008553  0.016173  0.059000  0.023291 -0.016736 -0.031267 -0.028933
996 -0.008874  0.046922  0.033661 -0.000193 -0.006447  0.002693  0.057325
997 -0.009390 -0.085835  0.012470 -0.023362  0.016244  0.013279 -0.011091
998  0.011906  0.042445  0.021965  0.071344 -0.046389  0.067688  0.079572
999  0.004508  0.078399  0.014382  0.045673  0.086536  0.011496 -0.036729

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.036564  0.092331  0.011791 -0.003005 -0.018494 -0.091950  no_efectores
1    0.022532  0.068664 -0.046674  0.019336 -0.038399  0.040286  no_efectores
2    0.003949 -0.033212  0.003781  0.041692 -0.091715  0.126707  no_efectores
3    0.042914  0.021675  0.057478  0.033847  0.009426  0.005761  no_efectores
4   -0.088845  0.010276 -0.104955 -0.046020  0.016505 -0.073527  no_efectores
..        …       …       …       …       …       …           …
995  0.021329  0.004369  0.033554  0.031640 -0.017618  0.029623  no_efectores
996  0.022952 -0.045551  0.006032  0.053006 -0.007373  0.027486  no_efectores
997  0.028672  0.006430  0.065755  0.035425 -0.007765 -0.064968  no_efectores
998 -0.004450 -0.010223 -0.035157  0.036381 -0.087081 -0.006716  no_efectores
999  0.059352 -0.008173  0.033531  0.001612 -0.036825  0.005633  no_efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009164     0.010617     0.007901     0.013133     0.006941
std       0.056337     0.054463     0.057374     0.061770     0.054767
min      -0.290225    -0.291437    -0.214284    -0.628848    -0.229668
25%      -0.022044    -0.019032    -0.021266    -0.017761    -0.023927
50%       0.011501     0.011868     0.010244     0.014593     0.008429
75%       0.040371     0.042206     0.041180     0.046790     0.039353
max       0.282012     0.317877     0.251794     0.262247     0.256692

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008423     0.010483     0.003937     0.003190     0.006631
std       0.056301     0.053399     0.061897     0.060231     0.056975
min      -0.363760    -0.372215    -0.558922    -0.358552    -0.296940
25%      -0.023947    -0.017593    -0.026159    -0.027591    -0.025618
50%       0.007369     0.009473     0.005401     0.004165     0.005931
75%       0.038818     0.038195     0.036283     0.036065     0.038010
max       0.242571     0.252252     0.404301     0.431695     0.402646

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.002885     0.006077     0.006774
std       0.058082     0.055700     0.055194
min      -0.268277    -0.349795    -0.255679
25%      -0.030221    -0.022027    -0.022753
50%       0.003146     0.006522     0.006867
75%       0.033845     0.038290     0.035586
max       0.302242     0.255379     0.214972
```


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


fusarium_oxysporum no_efectores dataset 4 Covarianza de auto cruzamiento (ACC)  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

### 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 fusarium_oxysporum dataset
4, sin valores atípicos.
Valores del documento csv.
```

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.002514 | -0.063280 | 0.045484 | -0.017408 | -0.088365 | -0.049946 | 0.040684 | |
| 1 | 0.025892 | 0.114121 | 0.107620 | 0.178666 | -0.060823 | 0.015862 | -0.026677 | |
| 2 | -0.015254 | -0.062968 | 0.064631 | -0.002966 | -0.016033 | 0.029092 | 0.055795 | |
| 3 | 0.061754 | 0.026905 | 0.005982 | 0.024770 | 0.034855 | 0.073363 | 0.037252 | |
| 4 | -0.014602 | -0.022998 | -0.003031 | -0.000104 | -0.018142 | 0.018382 | -0.012064 | |
| .. | … | … | … | … | … | … | … | |
| 993 | -0.127110 | 0.008322 | 0.017766 | -0.037289 | -0.021772 | 0.000265 | 0.005239 | |
| 994 | -0.084425 | -0.070477 | 0.012906 | -0.066162 | 0.012945 | -0.011572 | 0.092989 | |
| 995 | -0.008998 | -0.021737 | 0.060831 | 0.085246 | -0.019983 | -0.001467 | -0.002240 | |
| 997 | -0.016912 | -0.103641 | 0.098139 | 0.112798 | 0.005994 | -0.105289 | 0.004904 | |
| 998 | -0.013332 | 0.074159 | -0.041221 | 0.016725 | -0.105068 | 0.008814 | -0.029285 | |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|---|---|---|---|---|---|---|---|
| 0 | -0.104085 | -0.017047 | 0.022324 | 0.054196 | 0.024015 | 0.040474 | efectores |
| 1 | -0.028087 | -0.035676 | -0.045242 | -0.064342 | -0.083233 | -0.047228 | efectores |
| 2 | -0.017317 | -0.068332 | 0.138964 | 0.026007 | -0.072313 | 0.016092 | efectores |
| 3 | -0.008986 | 0.016100 | -0.036955 | 0.027629 | 0.041620 | 0.055806 | efectores |
| 4 | 0.004753 | -0.033428 | 0.011762 | 0.016421 | -0.039024 | 0.008058 | efectores |
| .. | … | … | … | … | … | … | … |
| 993 | 0.003473 | -0.023894 | -0.031830 | 0.002103 | 0.048642 | 0.077085 | efectores |
| 994 | -0.028423 | 0.029884 | 0.002559 | 0.045542 | -0.048364 | -0.056534 | efectores |
| 995 | 0.046816 | 0.025540 | 0.011101 | 0.106608 | 0.054747 | 0.006770 | efectores |
| 997 | -0.029013 | -0.009154 | -0.017590 | 0.116311 | -0.068738 | -0.105141 | efectores |
| 998 | 0.004101 | 0.015825 | -0.013806 | 0.019726 | -0.072498 | 0.007408 | efectores |

```
[905 rows x 14 columns]
```

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

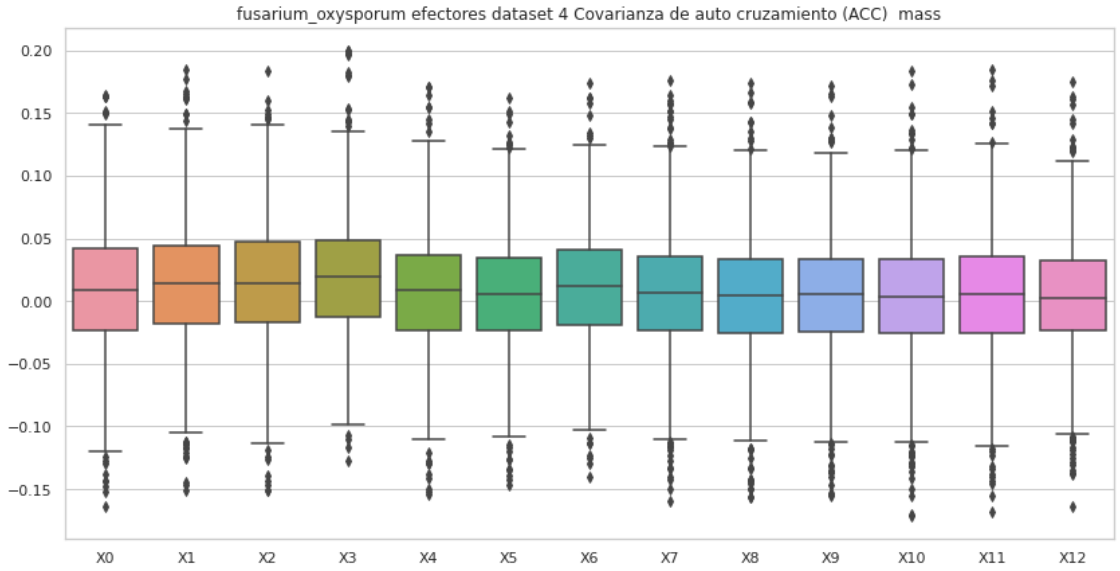
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | |
| mean | 0.009330 | 0.012878 | 0.013908 | 0.018661 | 0.007357 | 0.005279 | |
| std | 0.051520 | 0.051036 | 0.049756 | 0.048579 | 0.049345 | 0.048277 | |
| min | -0.164171 | -0.151425 | -0.151210 | -0.126878 | -0.153601 | -0.146151 | |
| 25% | -0.023471 | -0.018033 | -0.017315 | -0.012920 | -0.023490 | -0.023456 | |
| 50% | 0.009164 | 0.013780 | 0.013935 | 0.019525 | 0.008332 | 0.005835 | |
| 75% | 0.042192 | 0.044557 | 0.047293 | 0.048207 | 0.037157 | 0.034534 | |

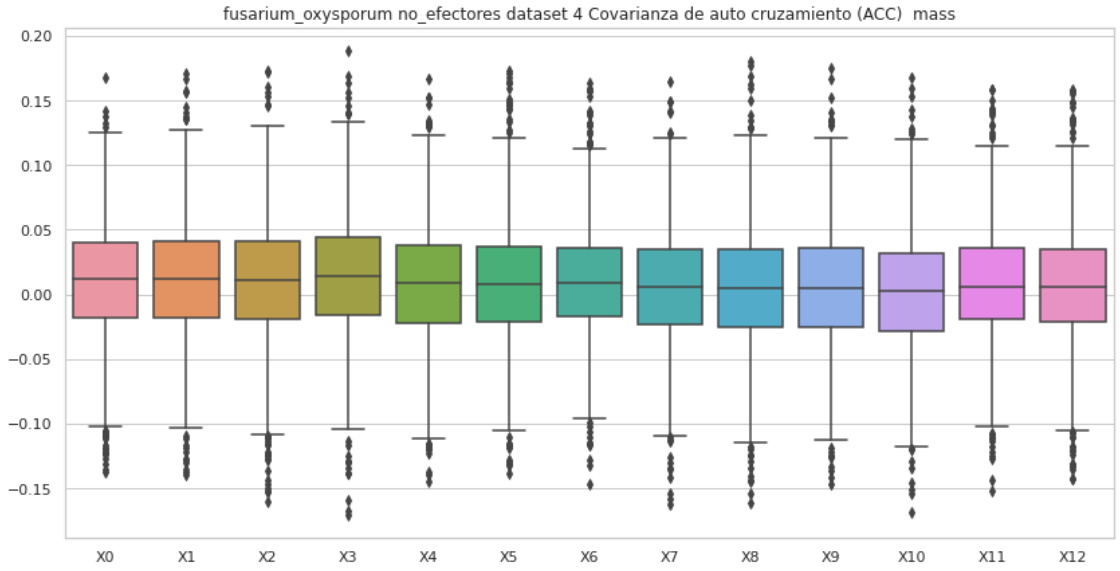
```
max      0.163835    0.184508    0.183659    0.199434    0.171045    0.162536

               X6          X7          X8          X9         X10         X11  \
count  905.000000  905.000000  905.000000  905.000000  905.000000  905.000000
mean     0.011162    0.006277    0.003994    0.005564    0.003667    0.006412
std      0.046986    0.050635    0.048707    0.048765    0.049894    0.050166
min     -0.140390   -0.159230   -0.155967   -0.155608   -0.171278   -0.168408
25%     -0.019158   -0.023345   -0.025164   -0.024795   -0.025271   -0.025271
50%      0.012373    0.006627    0.004247    0.005519    0.003602    0.005689
75%      0.040565    0.035380    0.033315    0.033670    0.033345    0.035376
max      0.173752    0.176126    0.173473    0.172184    0.183741    0.184695

              X12
count  905.000000
mean     0.003269
std      0.047805
min     -0.163474
25%     -0.023811
50%      0.002835
75%      0.031888
max      0.174366
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.078359  0.043146  0.079055  0.031518  0.015503  0.061494  0.017258
1   -0.001790  0.068715 -0.037725  0.031892  0.028435  0.002426  0.031217
3    0.016754  0.009699  0.034998  0.008028  0.019328 -0.014480  0.055154
4   -0.003720  0.121159  0.044997 -0.059388  0.058817 -0.011158  0.021339
5    0.019457  0.022436  0.009165  0.033283 -0.008717  0.028988  0.005616
..        …         …         …         …         …         …         …
995  0.008553  0.016173  0.059000  0.023291 -0.016736 -0.031267 -0.028933
996 -0.008874  0.046922  0.033661 -0.000193 -0.006447  0.002693  0.057325
997 -0.009390 -0.085835  0.012470 -0.023362  0.016244  0.013279 -0.011091
998  0.011906  0.042445  0.021965  0.071344 -0.046389  0.067688  0.079572
999  0.004508  0.078399  0.014382  0.045673  0.086536  0.011496 -0.036729

           X7        X8        X9       X10       X11       X12           X13
0    0.036564  0.092331  0.011791 -0.003005 -0.018494 -0.091950  no_efectores
1    0.022532  0.068664 -0.046674  0.019336 -0.038399  0.040286  no_efectores
3    0.042914  0.021675  0.057478  0.033847  0.009426  0.005761  no_efectores
4   -0.088845  0.010276 -0.104955 -0.046020  0.016505 -0.073527  no_efectores
5   -0.010629  0.042915 -0.050987 -0.028890 -0.029756  0.033284  no_efectores
..        …         …         …         …         …         …           …
```

```
995  0.021329  0.004369  0.033554  0.031640 -0.017618  0.029623  no_efectores
996  0.022952 -0.045551  0.006032  0.053006 -0.007373  0.027486  no_efectores
997  0.028672  0.006430  0.065755  0.035425 -0.007765 -0.064968  no_efectores
998 -0.004450 -0.010223 -0.035157  0.036381 -0.087081 -0.006716  no_efectores
999  0.059352 -0.008173  0.033531  0.001612 -0.036825  0.005633  no_efectores

[914 rows x 14 columns]


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

                X0           X1           X2           X3           X4           X5  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.009462    0.010698    0.008888    0.014164    0.006960    0.008957
std      0.047941    0.047298    0.049793    0.050565    0.048563    0.049743
min     -0.137389   -0.139878   -0.160116   -0.170587   -0.144798   -0.138308
25%     -0.017987   -0.018175   -0.019109   -0.016030   -0.022095   -0.021590
50%      0.011923    0.011668    0.011079    0.014254    0.008775    0.007692
75%      0.039678    0.040644    0.040683    0.043772    0.037760    0.036857
max      0.167333    0.170396    0.173373    0.188423    0.166366    0.172655

                X6           X7           X8           X9          X10          X11  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.010410    0.005279    0.004987    0.005944    0.002252    0.007167
std      0.044675    0.047904    0.048714    0.048361    0.048930    0.047170
min     -0.146535   -0.162299   -0.161527   -0.146800   -0.168810   -0.151423
25%     -0.016585   -0.022787   -0.025046   -0.025266   -0.028633   -0.019417
50%      0.009309    0.005825    0.004582    0.004605    0.003080    0.006335
75%      0.036332    0.035188    0.035306    0.036051    0.031791    0.036156
max      0.164105    0.164658    0.180551    0.174836    0.167291    0.158740

               X12
count  914.000000
mean     0.006999
std      0.048557
min     -0.143016
25%     -0.021542
50%      0.006225
75%      0.034905
max      0.158854
```




# 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 fusarium_oxysporum dataset
4, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.000304 -0.047114  0.051996 -0.056099 -0.104767 -0.025944  0.049423
1    0.106179  0.066390  0.029266  0.047348  0.071189  0.055114  0.059927
2    0.020605  0.072440 -0.036689  0.047272  0.084708 -0.002785  0.013601
3   -0.016620  0.056250 -0.021174 -0.051633  0.011272  0.017370 -0.038016
4   -0.038174  0.016853  0.003386 -0.005425  0.022376  0.040025 -0.037184
..        …         …         …         …         …         …         …
995  0.062777  0.077328  0.013122  0.042313 -0.076870  0.026133  0.005999
996  0.046030 -0.038759 -0.093911 -0.083573  0.060842 -0.089706 -0.114610
997 -0.101995 -0.034409  0.027662 -0.079080 -0.131433  0.017373  0.048871
998 -0.040296  0.021890 -0.027670 -0.070593  0.024978  0.004749 -0.020802
999 -0.111941 -0.015217 -0.048248 -0.055122 -0.052760 -0.014694 -0.015067

           X7        X8        X9       X10       X11       X12        X13
0   -0.062897 -0.041226  0.129442  0.002424 -0.001224  0.115306  efectores
1    0.074907  0.046747 -0.015035  0.069933  0.062963  0.055542  efectores
2   -0.119399  0.027827  0.033533  0.042164  0.031452 -0.014525  efectores
```

```
3   -0.045962  0.000671  0.018462  0.006649  0.006972 -0.005791  efectores
4    0.044542 -0.071611  0.007024  0.018557 -0.033066 -0.036918  efectores
..        …       …         …         …         …         …         …
995  0.009893  0.019548  0.028407 -0.097803  0.031716  0.023584  efectores
996 -0.196504  0.162793  0.239918 -0.024079 -0.101003 -0.102105  efectores
997  0.031698  0.004545  0.049092 -0.038426 -0.153349  0.054965  efectores
998 -0.022580 -0.037919 -0.005087  0.064768 -0.020349  0.052591  efectores
999  0.019630 -0.072857  0.022120 -0.100187  0.046758  0.127547  efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.010475    -0.018154     0.023313     0.027424     0.000954
std       0.067417     0.073563     0.062764     0.065712     0.064827
min      -0.237027    -0.305464    -0.248819    -0.445725    -0.331696
25%      -0.033304    -0.067601    -0.012890    -0.010704    -0.037254
50%       0.006091    -0.020818     0.023305     0.026730    -0.003123
75%       0.047182     0.027828     0.061622     0.064824     0.035136
max       0.369554     0.322270     0.285074     0.315341     0.274947

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.007903     0.022674     0.011715    -0.000876     0.008562
std       0.069924     0.064557     0.059572     0.068383     0.065007
min      -0.522842    -0.255394    -0.363496    -0.411023    -0.314677
25%      -0.044840    -0.014228    -0.020991    -0.035495    -0.029046
50%      -0.008682     0.023269     0.012501    -0.002717     0.007276
75%       0.034597     0.060425     0.044552     0.035303     0.044980
max       0.358439     0.460240     0.267221     0.358184     0.341007

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.012575     0.008674     0.003669
std       0.065444     0.061205     0.061385
min      -0.330059    -0.358681    -0.317678
25%      -0.020551    -0.025116    -0.029462
50%       0.011594     0.008780     0.005476
75%       0.050284     0.039340     0.038327
max       0.637715     0.329628     0.330045
```

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

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.155954 | 0.028858 | -0.028411 | -0.029849 | 0.026986 | 0.044780 | -0.050676 | |
| 1 | 0.093553 | -0.037763 | 0.047974 | -0.005360 | -0.002267 | -0.039666 | -0.055616 | |
| 2 | 0.125491 | -0.119018 | -0.037770 | 0.058711 | -0.138663 | -0.004527 | 0.062589 | |
| 3 | -0.044677 | -0.073973 | -0.009475 | 0.084450 | 0.003392 | -0.038241 | 0.031120 | |
| 4 | 0.030521 | 0.008824 | 0.085461 | 0.057496 | 0.024336 | 0.031340 | 0.030831 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.050524 | -0.028888 | 0.054840 | 0.040303 | -0.109118 | -0.098921 | -0.050692 | |
| 996 | 0.099369 | 0.072313 | 0.152915 | 0.062860 | 0.073408 | 0.104893 | 0.086452 | |
| 997 | -0.033032 | -0.014799 | -0.032832 | 0.043024 | -0.020831 | -0.001155 | -0.030626 | |
| 998 | 0.001939 | -0.039129 | -0.019499 | -0.093605 | 0.010863 | -0.036670 | 0.024312 | |
| 999 | -0.026925 | -0.073587 | -0.081148 | 0.045747 | 0.020914 | 0.005033 | -0.028335 | |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|---|---|---|---|---|---|---|---|
| 0 | 0.033484 | 0.106962 | -0.115925 | -0.110319 | -0.149608 | 0.001096 | no_efectores |
| 1 | -0.053137 | -0.078929 | -0.022931 | 0.051742 | 0.005061 | -0.040319 | no_efectores |
| 2 | -0.087162 | 0.122928 | 0.129421 | -0.047976 | -0.143422 | 0.164991 | no_efectores |
| 3 | -0.075578 | -0.002241 | -0.035431 | 0.005913 | -0.022948 | -0.022862 | no_efectores |
| 4 | 0.084529 | 0.060409 | 0.045012 | -0.007217 | 0.065174 | 0.067240 | no_efectores |
| .. | … | … | … | … | … | … | … |
| 995 | 0.045017 | 0.011191 | 0.024893 | 0.038912 | 0.062463 | 0.065511 | no_efectores |
| 996 | 0.066104 | 0.047073 | 0.075881 | 0.037126 | 0.125561 | 0.011599 | no_efectores |
| 997 | 0.058794 | -0.047358 | -0.024413 | -0.025227 | -0.006200 | 0.011042 | no_efectores |
| 998 | 0.032470 | -0.046232 | -0.028036 | 0.028314 | 0.018750 | 0.028797 | no_efectores |
| 999 | -0.013744 | 0.010687 | 0.009639 | -0.051037 | 0.053140 | -0.044943 | no_efectores |

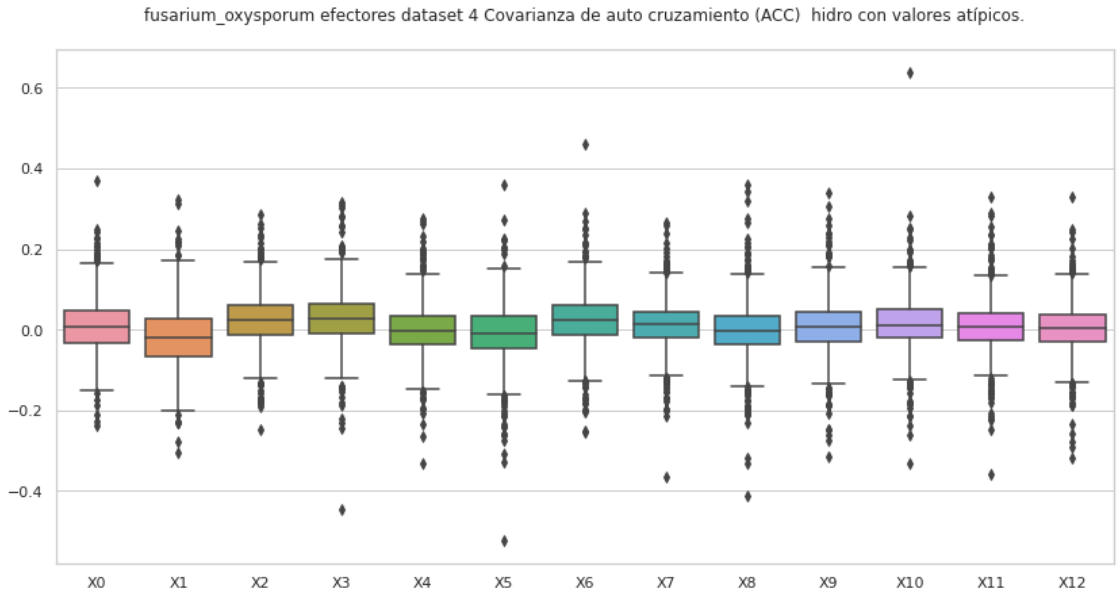
[1000 rows x 14 columns]

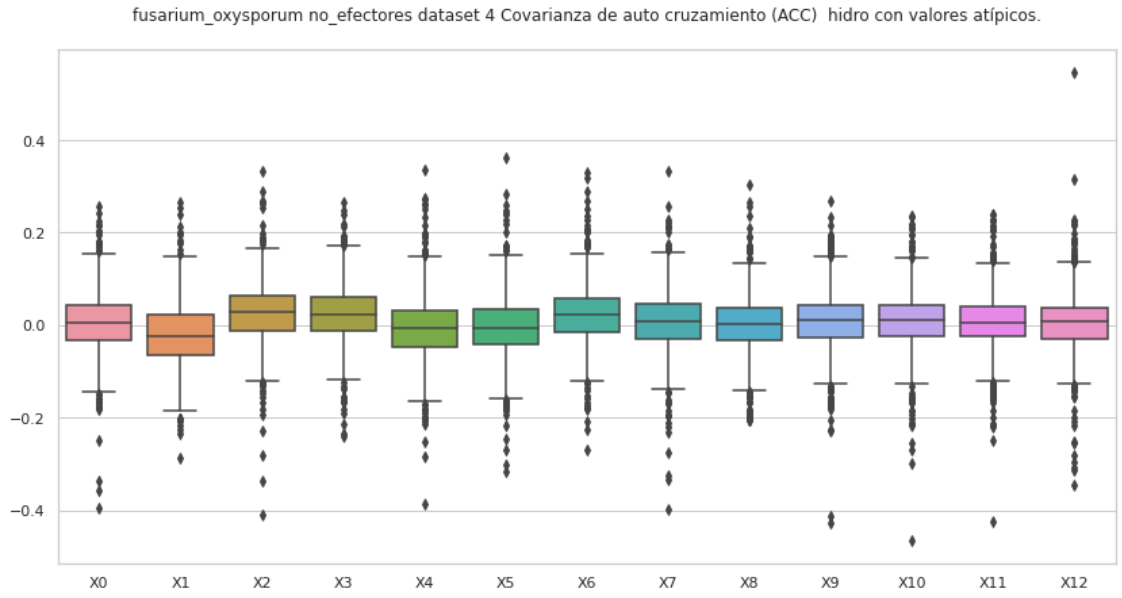
Covarianza de auto cruzamiento (ACC)  hidro no_efectores fusarium_oxysporum
dataset 4, con valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.005885 | -0.019288 | 0.024642 | 0.025270 | -0.006736 | |
| std | 0.065566 | 0.068448 | 0.065043 | 0.062557 | 0.070351 | |
| min | -0.394024 | -0.285733 | -0.410908 | -0.238948 | -0.387048 | |
| 25% | -0.031516 | -0.065169 | -0.011413 | -0.011210 | -0.046618 | |
| 50% | 0.004505 | -0.022631 | 0.027604 | 0.024392 | -0.007416 | |
| 75% | 0.044522 | 0.022953 | 0.062569 | 0.062358 | 0.032486 | |
| max | 0.258113 | 0.265543 | 0.333545 | 0.264360 | 0.335733 | |

| | X5 | X6 | X7 | X8 | X9 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | -0.004221 | 0.023389 | 0.007795 | 0.000711 | 0.008227 | |
| std | 0.068181 | 0.063726 | 0.065523 | 0.062761 | 0.064893 | |
| min | -0.315100 | -0.268307 | -0.398224 | -0.205213 | -0.426027 | |
| 25% | -0.042130 | -0.013833 | -0.029505 | -0.033972 | -0.025830 | |
| 50% | -0.005035 | 0.021832 | 0.008884 | 0.001513 | 0.009922 | |
| 75% | 0.035447 | 0.058001 | 0.046766 | 0.036425 | 0.043828 | |
| max | 0.361143 | 0.331455 | 0.331549 | 0.304676 | 0.267472 | |

| | X10 | X11 | X12 |
|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.008117 | 0.006567 | 0.005358 |
| std | 0.061643 | 0.061537 | 0.067644 |
| min | -0.465384 | -0.423024 | -0.345997 |
| 25% | -0.024843 | -0.024995 | -0.028524 |
| 50% | 0.010377 | 0.005144 | 0.007758 |
| 75% | 0.043154 | 0.039380 | 0.037524 |
| max | 0.235313 | 0.240142 | 0.546230 |






## 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 fusarium_oxysporum dataset 4,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.000304 -0.047114  0.051996 -0.056099 -0.104767 -0.025944  0.049423
1    0.106179  0.066390  0.029266  0.047348  0.071189  0.055114  0.059927
2    0.020605  0.072440 -0.036689  0.047272  0.084708 -0.002785  0.013601
3   -0.016620  0.056250 -0.021174 -0.051633  0.011272  0.017370 -0.038016
4   -0.038174  0.016853  0.003386 -0.005425  0.022376  0.040025 -0.037184
..        …       …       …       …       …       …       …
994 -0.025724 -0.035968 -0.063828 -0.028051 -0.009763  0.031904  0.032880
995  0.062777  0.077328  0.013122  0.042313 -0.076870  0.026133  0.005999
997 -0.101995 -0.034409  0.027662 -0.079080 -0.131433  0.017373  0.048871
998 -0.040296  0.021890 -0.027670 -0.070593  0.024978  0.004749 -0.020802
999 -0.111941 -0.015217 -0.048248 -0.055122 -0.052760 -0.014694 -0.015067

           X7        X8        X9       X10       X11       X12        X13
0   -0.062897 -0.041226  0.129442  0.002424 -0.001224  0.115306  efectores
1    0.074907  0.046747 -0.015035  0.069933  0.062963  0.055542  efectores
2   -0.119399  0.027827  0.033533  0.042164  0.031452 -0.014525  efectores
3   -0.045962  0.000671  0.018462  0.006649  0.006972 -0.005791  efectores
4    0.044542 -0.071611  0.007024  0.018557 -0.033066 -0.036918  efectores
..        …       …       …       …       …       …        …
994  0.119874  0.008302 -0.078148  0.056745 -0.135519 -0.002081  efectores
```

```
995  0.009893  0.019548  0.028407 -0.097803  0.031716  0.023584  efectores
997  0.031698  0.004545  0.049092 -0.038426 -0.153349  0.054965  efectores
998 -0.022580 -0.037919 -0.005087  0.064768 -0.020349  0.052591  efectores
999  0.019630 -0.072857  0.022120 -0.100187  0.046758  0.127547  efectores

[921 rows x 14 columns]
```

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

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 921.000000 | 921.000000 | 921.000000 | 921.000000 | 921.000000 | 921.000000 | |
| mean | 0.009433 | -0.018874 | 0.022932 | 0.026523 | -0.000551 | -0.005443 | |
| std | 0.060839 | 0.066186 | 0.054747 | 0.054817 | 0.054405 | 0.056968 | |
| min | -0.174051 | -0.232516 | -0.156627 | -0.152121 | -0.172916 | -0.214615 | |
| 25% | -0.031265 | -0.066007 | -0.011500 | -0.009214 | -0.035989 | -0.041128 | |
| 50% | 0.006457 | -0.021122 | 0.022988 | 0.026414 | -0.003540 | -0.006905 | |
| 75% | 0.045791 | 0.025901 | 0.057761 | 0.063356 | 0.030732 | 0.033529 | |
| max | 0.199593 | 0.185172 | 0.201408 | 0.207622 | 0.190773 | 0.200041 | |

| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 921.000000 | 921.000000 | 921.000000 | 921.000000 | 921.000000 | 921.000000 | |
| mean | 0.021446 | 0.011837 | -0.001335 | 0.007480 | 0.011432 | 0.008375 | |
| std | 0.053296 | 0.050359 | 0.054894 | 0.052450 | 0.053593 | 0.049654 | |
| min | -0.170317 | -0.165930 | -0.196441 | -0.182948 | -0.182591 | -0.170262 | |
| 25% | -0.013047 | -0.019062 | -0.033177 | -0.025988 | -0.019790 | -0.023673 | |
| 50% | 0.021610 | 0.012021 | -0.002859 | 0.007024 | 0.011547 | 0.008817 | |
| 75% | 0.057377 | 0.042868 | 0.032057 | 0.042500 | 0.045635 | 0.038551 | |
| max | 0.190550 | 0.165553 | 0.187179 | 0.182472 | 0.191610 | 0.184578 | |

| | X12 |
|---|---|
| count | 921.000000 |
| mean | 0.003838 |
| std | 0.051903 |
| min | -0.179193 |
| 25% | -0.027526 |
| 50% | 0.005225 |
| 75% | 0.035688 |
| max | 0.170009 |

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

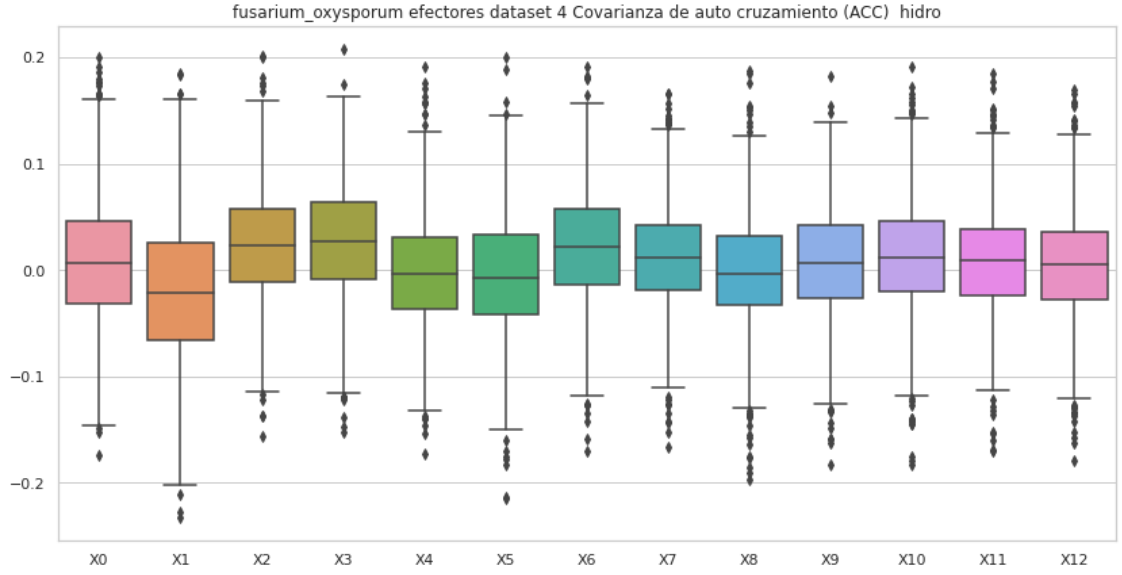
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.155954  0.028858 -0.028411 -0.029849  0.026986  0.044780 -0.050676
1    0.093553 -0.037763  0.047974 -0.005360 -0.002267 -0.039666 -0.055616
2    0.125491 -0.119018 -0.037770  0.058711 -0.138663 -0.004527  0.062589
3   -0.044677 -0.073973 -0.009475  0.084450  0.003392 -0.038241  0.031120
4    0.030521  0.008824  0.085461  0.057496  0.024336  0.031340  0.030831
..        …         …         …         …         …         …         …
995  0.050524 -0.028888  0.054840  0.040303 -0.109118 -0.098921 -0.050692
996  0.099369  0.072313  0.152915  0.062860  0.073408  0.104893  0.086452
997 -0.033032 -0.014799 -0.032832  0.043024 -0.020831 -0.001155 -0.030626
998  0.001939 -0.039129 -0.019499 -0.093605  0.010863 -0.036670  0.024312
999 -0.026925 -0.073587 -0.081148  0.045747  0.020914  0.005033 -0.028335

           X7        X8        X9       X10       X11       X12           X13
0    0.033484  0.106962 -0.115925 -0.110319 -0.149608  0.001096  no_efectores
1   -0.053137 -0.078929 -0.022931  0.051742  0.005061 -0.040319  no_efectores
2   -0.087162  0.122928  0.129421 -0.047976 -0.143422  0.164991  no_efectores
3   -0.075578 -0.002241 -0.035431  0.005913 -0.022948 -0.022862  no_efectores
4    0.084529  0.060409  0.045012 -0.007217  0.065174  0.067240  no_efectores
..        …         …         …         …         …         …           …
995  0.045017  0.011191  0.024893  0.038912  0.062463  0.065511  no_efectores
996  0.066104  0.047073  0.075881  0.037126  0.125561  0.011599  no_efectores
997  0.058794 -0.047358 -0.024413 -0.025227 -0.006200  0.011042  no_efectores
998  0.032470 -0.046232 -0.028036  0.028314  0.018750  0.028797  no_efectores
999 -0.013744  0.010687  0.009639 -0.051037  0.053140 -0.044943  no_efectores

[918 rows x 14 columns]
```

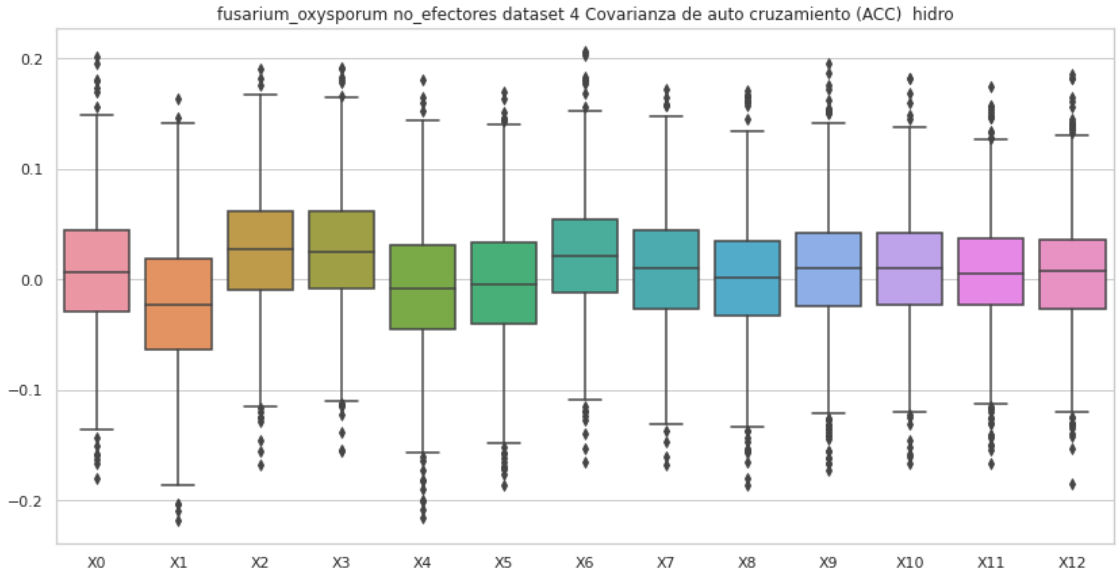
Covarianza de auto cruzamiento (ACC)  no_efectores fusarium_oxysporum dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.007338   -0.020625    0.024737    0.026348   -0.008704   -0.004113
std      0.056542    0.061375    0.054391    0.055224    0.059717    0.056898
min     -0.179728   -0.218123   -0.167969   -0.155047   -0.215316   -0.185831
25%     -0.028717   -0.063946   -0.009479   -0.007932   -0.045361   -0.039621
50%      0.005866   -0.022710    0.027409    0.025076   -0.008305   -0.005172
75%      0.044174    0.019032    0.061466    0.061210    0.030625    0.033249
max      0.201813    0.163769    0.189976    0.191763    0.180623    0.170073

                X6          X7          X8          X9         X10         X11  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.021998    0.008981    0.000765    0.007807    0.008751    0.006402
std      0.054207    0.053115    0.054323    0.053830    0.050942    0.050650
```

| | | | | | | |
|---|---|---|---|---|---|---|
| min | -0.165906 | -0.167786 | -0.185893 | -0.173060 | -0.166111 | -0.166820 |
| 25% | -0.011685 | -0.026930 | -0.032607 | -0.024805 | -0.023262 | -0.022971 |
| 50% | 0.020987 | 0.009593 | 0.001445 | 0.009637 | 0.010571 | 0.005284 |
| 75% | 0.054228 | 0.044124 | 0.034990 | 0.041509 | 0.042177 | 0.037035 |
| max | 0.206278 | 0.172167 | 0.170849 | 0.195867 | 0.182185 | 0.174021 |

| | X12 |
|---|---|
| count | 918.000000 |
| mean | 0.005082 |
| std | 0.053488 |
| min | -0.184588 |
| 25% | -0.026684 |
| 50% | 0.007458 |
| 75% | 0.036089 |
| max | 0.185729 |




fusarium_oxysporum 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