ds5_archaea_limpieza_de_datos

January 19, 2021

Limpieza de datos

```python
[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

```python
[2]: organismo ="archaea"
     dataset = 5
     nombre = ("ds" + str(dataset) + "_" + str(organismo))
     nombre2 = (str(organismo)+ " dataset " + str(dataset))
     r2 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
      ↪transformaciones/sin_filtrar")
     r3 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
      ↪transformaciones/sin_atipicos")

     nom1 = ("/ds" + str(dataset) + "_AAC_efectores_" + str(organismo) + ".txt")
     nom2 = ("/ds" + str(dataset) + "_ACC_hidro_mass_efectores_" + str(organismo) +␣
      ↪".txt")
     nom3 = ("/ds" + str(dataset) + "_ACC_mass_efectores_" + str(organismo) + ".txt")
     nom4 = ("/ds" + str(dataset) + "_ACC_hidro_efectores_" + str(organismo) + ".
      ↪txt")
     nom5 = ("/ds" + str(dataset) + "_PseAAC_hidro_mass_efectores_" + str(organismo)␣
      ↪+ ".txt")
     nom6 = ("/ds" + str(dataset) + "_PseAAC_mass_efectores_" + str(organismo) + ".
      ↪txt")
     nom7 = ("/ds" + str(dataset) + "_PseAAC_hidro_efectores_" + str(organismo) + ".
      ↪txt")
```

```python
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 archaea dataset 5, con valores
atípicos.
Valores del documento csv.

         X0      X1     X2      X3     X4      X5     X6      X7     X8  \
0     6.977  10.078  3.101   8.527  0.775  10.853  2.326   4.651  3.876
1     6.452  11.290  2.419  12.903  0.806   8.065  0.806   5.645  3.226
2    10.938   8.594  2.344   9.375  0.781  10.156  1.562  11.719  1.562
3     4.342   4.342  4.482   5.602  0.000   7.423  2.241   3.922  1.120
4    12.587   9.790  0.000   5.594  0.699  11.888  2.098   6.993  2.797

..      ...     ...    ...     ...    ...     ...    ...     ...    ...
495   5.747  10.345  2.299   5.747  3.448  10.345  1.149   4.598  0.000
496   9.594   8.856  0.369  11.070  0.000   8.487  1.476  10.332  0.738
497  10.897   8.333  0.641   4.487  1.282   2.564  2.564   4.487  3.205
498   7.143   4.911  1.339   4.018  0.000   4.464  1.339   8.929  0.893
499   7.390   4.388  3.464   5.312  0.000   7.621  3.464   7.852  3.695
```

```
            X9   …     X11     X12     X13     X14     X15      X16     X17     X18  \
0      0.775   …    1.550   3.101   1.550   4.651   6.202    8.527   0.775   2.326
1      2.419   …    1.613   0.806   2.419   4.839   8.871    3.226   0.806   3.226
2      3.906   …    0.781   1.562   4.688   6.250   1.562    6.250   0.000   1.562
3     11.485   …    9.244   1.401   2.801   4.762   6.443    5.182   0.420   6.443
4      2.797   …    4.895   1.399   2.797   6.294   6.993    5.594   0.000   2.098
..        …   … …       …       …       …       …       …        …       …
495    5.747   …    8.046   4.598   1.149   1.149   8.046    2.299   0.000   3.448
496    2.583   …    0.738   0.738   1.476   5.166   5.904    7.380   0.369   2.214
497    4.487   …    1.282   1.282   6.410   6.410   6.410    6.410   3.205   5.128
498    8.929   …    0.893   1.786   6.696   3.571   7.589   11.161   0.893   2.232
499    9.007   …    8.776   2.309   2.540   7.159   2.540    4.388   0.231   2.540

          X19         X20
0        8.527    efectores
1        8.871    efectores
2       10.156    efectores
3        4.762    efectores
4        4.895    efectores
..          …           …
495      8.046    efectores
496     15.867    efectores
497      6.410    efectores
498     10.714    efectores
499     10.162    efectores

[500 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   500.00000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      9.15243    6.033006    2.829824    5.953344    0.665144    6.981942
std       4.40091    2.626824    2.164666    2.635075    0.922017    3.667012
min       0.00000    0.000000    0.000000    1.020000    0.000000    0.000000
25%       5.87350    3.985000    1.166000    3.846000    0.000000    4.167000
50%       8.49100    5.784000    2.326000    5.591500    0.307500    6.940000
75%      12.11475    7.692000    4.043500    7.692000    0.990000    9.459000
max      24.02200   14.062000   12.227000   14.286000    4.878000   20.661000

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      2.372540    7.519232    1.768576    5.747138   11.472948    4.505950
std       1.558318    2.951743    1.169474    3.353142    3.236235    4.272688
```

```
min        0.000000     0.667000     0.000000     0.000000     3.529000     0.000000
25%        1.270750     5.269750     0.777250     3.180750     9.158750     1.074000
50%        2.237000     7.043000     1.678500     5.208000    11.392000     2.763000
75%        3.185250     9.503000     2.520250     7.717000    13.725000     7.459000
max        9.474000    16.995000     6.161000    16.320000    20.849000    20.388000

                 X12          X13          X14          X15          X16          X17  \
count     500.000000   500.000000   500.000000   500.000000   500.000000   500.000000
mean        2.104668     3.813400     4.109850     6.052750     5.562200     1.250086
std         1.213013     1.990065     1.853343     2.120261     2.096773     1.046656
min         0.250000     0.000000     0.000000     1.042000     0.588000     0.000000
25%         1.175500     2.419000     2.963000     4.590750     4.074250     0.565750
50%         1.814000     3.654500     3.964000     5.866000     5.466500     1.123500
75%         2.667000     4.930750     5.195000     7.218250     6.880500     1.671500
max         8.187000    14.583000    10.000000    17.347000    14.170000     5.389000

                 X18          X19
count     500.000000   500.00000
mean        3.293224     8.81168
std         1.579787     3.61327
min         0.000000     1.25800
25%         2.213000     5.99575
50%         3.152500     8.31750
75%         4.167000    11.36050
max         9.223000    19.39400
```

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

```
            X0        X1       X2        X3       X4        X5      X6        X7       X8  \
0        9.639     9.639    2.410     5.422    0.000    10.843   3.012     6.024    3.012
1        7.843     3.922    3.922     0.000    0.000     3.922   1.961    15.686    1.961
2        6.952     3.209    4.278     1.604    1.070     8.021   4.813     5.348    3.209
3       14.748    10.432    1.439    11.511    0.000    12.230   1.439     7.554    2.518
4        7.035     1.508    7.035     6.533    0.000     3.015   1.508     8.040    2.513
..         …         …        …         …        …         …       …         …
495      4.124     8.247    4.124     6.186    0.000     9.278   4.124    10.309    3.093
496     20.064     6.688    0.955     2.866    0.000     2.866   0.318    11.146    0.637
497      2.899     4.348   14.493     2.899    2.899     4.348   1.449     5.797    1.449
498     10.490     6.294    1.399     2.098    0.000     1.399   4.196    13.287    1.399
499     11.483     3.110    1.675     3.349    0.000     6.220   1.196    11.005    1.914

            X9    …      X11      X12      X13      X14      X15      X16      X17      X18  \
0        5.422    …    0.602    1.205    3.012    8.434    6.024    4.819    1.205    1.205
```

5

```
1      11.765  …    0.000   3.922   5.882   1.961   7.843    3.922   0.000   1.961
2       7.487  …    8.021   3.209   3.209   6.417   6.417    4.813   0.000   2.139
3       1.799  …    0.719   1.079   1.799   5.036   4.676    4.317   0.719   0.000
4       8.040  …    3.015   2.010   6.030   4.523  12.060   10.050   1.508   2.010
..        …    … …     …       …       …       …       …        …       …       …
495     3.093  …    8.247   1.031   3.093   8.247   5.155    7.216   0.000   1.031
496     4.140  …    0.637   1.592   4.140   3.503   6.051    4.777   0.637   2.548
497     7.246  …   11.594   2.899   0.000   8.696   4.348    2.899   0.000   2.899
498     1.399  …    0.699   0.699   4.895   5.594   3.497    7.692   3.497   2.098
499     6.220  …    0.957   2.153   5.502   4.785   4.067    5.024   1.914   2.153


            X19            X20
0       11.446     no_efectores
1        5.882     no_efectores
2        9.626     no_efectores
3        8.633     no_efectores
4        5.025     no_efectores
..        …                 …
495      7.216     no_efectores
496     12.420     no_efectores
497      5.797     no_efectores
498     13.986     no_efectores
499     13.636     no_efectores

[500 rows x 21 columns]
```

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

```
                 X0            X1            X2            X3            X4            X5  \
count    500.000000    500.000000    500.000000    500.000000    500.000000    500.000000
mean       9.828960      6.124186      2.823640      7.654346      1.070588      8.045028
std        4.242888      2.906082      2.122343      3.724102      1.544732      4.018782
min        0.000000      0.000000      0.000000      0.000000      0.000000      0.000000
25%        6.856500      4.239250      1.493000      5.063000      0.000000      5.459500
50%        9.736000      5.871500      2.405000      7.692000      0.581000      8.023000
75%       12.481250      7.771000      3.640250     10.160750      1.384000     10.217000
max       32.143000     27.273000     16.250000     24.324000     10.145000     25.532000

                 X6            X7            X8            X9           X10           X11  \
count    500.000000    500.000000    500.000000    500.000000    500.000000    500.000000
mean       2.593048      8.188346      1.900252      5.038538      8.889900      3.122132
std        1.837954      3.118614      1.396257      3.272462      3.493552      3.252465
min        0.000000      0.000000      0.000000      0.000000      0.000000      0.000000
25%        1.447750      6.317250      1.024500      2.738000      6.597750      0.868000
50%        2.207000      7.877000      1.751500      4.310000      8.527000      2.064000
```
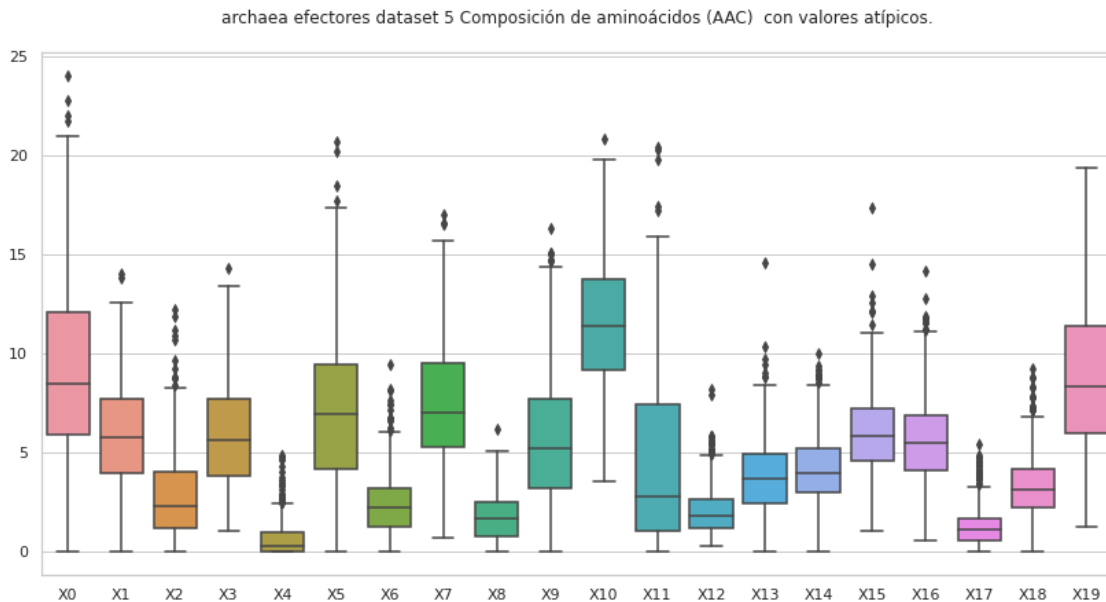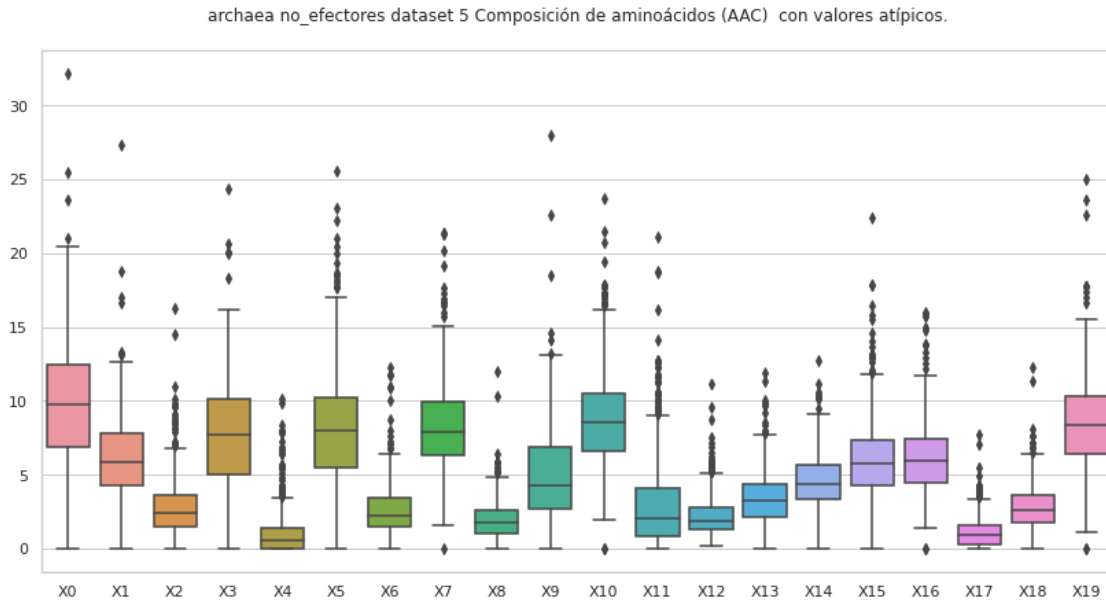
75%	3.458000	9.924250	2.603000	6.916250	10.499250	4.119750
max	12.245000	21.387000	12.000000	28.000000	23.656000	21.127000

	X12	X13	X14	X15	X16	X17 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	2.232516	3.428828	4.480904	6.071160	6.140460	1.083342
std	1.404173	1.844071	1.956244	2.769508	2.427885	1.026028
min	0.178000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.258750	2.125500	3.324750	4.304500	4.478000	0.304500
50%	1.894000	3.268000	4.345000	5.720500	5.932500	0.946000
75%	2.804000	4.379250	5.665500	7.313750	7.393000	1.547000
max	11.111000	11.913000	12.766000	22.424000	16.000000	7.692000

	X18	X19
count	500.000000	500.000000
mean	2.823134	8.460686
std	1.615307	3.329442
min	0.000000	0.000000
25%	1.775000	6.394500
50%	2.624000	8.365000
75%	3.636000	10.288000
max	12.245000	25.000000

archaea efectores dataset 5 Composición de aminoácidos (AAC) con valores atípicos.

archaea no_efectores dataset 5 Composición de aminoácidos (AAC) con valores atípicos.

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)])
```

8

```python
    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 archaea dataset 5, sin valores
atípicos.
Valores del documento csv.

           X0      X1     X2      X3     X4      X5     X6       X7      X8  \
0       6.977  10.078  3.101   8.527  0.775  10.853  2.326    4.651   3.876
1       6.452  11.290  2.419  12.903  0.806   8.065  0.806    5.645   3.226
2      10.938   8.594  2.344   9.375  0.781  10.156  1.562   11.719   1.562
3       4.342   4.342  4.482   5.602  0.000   7.423  2.241    3.922   1.120
4      12.587   9.790  0.000   5.594  0.699  11.888  2.098    6.993   2.797
..        …       …      …       …      …       …      …        …       …
493     8.145   8.597  0.905   9.955  0.000   6.335  5.882    9.050   0.452
496     9.594   8.856  0.369  11.070  0.000   8.487  1.476   10.332   0.738
497    10.897   8.333  0.641   4.487  1.282   2.564  2.564    4.487   3.205
498     7.143   4.911  1.339   4.018  0.000   4.464  1.339    8.929   0.893
499     7.390   4.388  3.464   5.312  0.000   7.621  3.464    7.852   3.695

          X9  …     X11    X12    X13     X14     X15     X16    X17     X18  \
0      0.775  …   1.550  3.101  1.550   4.651   6.202   8.527  0.775   2.326
1      2.419  …   1.613  0.806  2.419   4.839   8.871   3.226  0.806   3.226
2      3.906  …   0.781  1.562  4.688   6.250   1.562   6.250  0.000   1.562
3     11.485  …   9.244  1.401  2.801   4.762   6.443   5.182  0.420   6.443
4      2.797  …   4.895  1.399  2.797   6.294   6.993   5.594  0.000   2.098
..       …    … …   …       …      …       …       …       …      …       …
493    1.357  …   1.357  2.715  1.810   4.977   9.502   5.430  0.452   2.715
```

```
496   2.583   …   0.738   0.738   1.476   5.166   5.904    7.380   0.369   2.214
497   4.487   …   1.282   1.282   6.410   6.410   6.410    6.410   3.205   5.128
498   8.929   …   0.893   1.786   6.696   3.571   7.589   11.161   0.893   2.232
499   9.007   …   8.776   2.309   2.540   7.159   2.540    4.388   0.231   2.540

          X19        X20
0        8.527   efectores
1        8.871   efectores
2       10.156   efectores
3        4.762   efectores
4        4.895   efectores
..         …         …
493     10.407   efectores
496     15.867   efectores
497      6.410   efectores
498     10.714   efectores
499     10.162   efectores

[441 rows x 21 columns]


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

                 X0          X1          X2          X3          X4          X5   \
count    441.000000  441.000000  441.000000  441.000000  441.000000  441.000000
mean       9.363866    6.120961    2.670816    5.954805    0.585624    6.935646
std        4.333685    2.568835    1.909959    2.624439    0.776640    3.529825
min        0.000000    0.621000    0.000000    1.024000    0.000000    0.000000
25%        6.081000    4.101000    1.158000    3.846000    0.000000    4.211000
50%        8.759000    5.941000    2.273000    5.584000    0.289000    7.000000
75%       12.312000    7.692000    3.896000    7.792000    0.889000    9.326000
max       22.006000   13.805000    9.244000   13.400000    3.361000   17.742000

                 X6          X7          X8          X9         X10         X11   \
count    441.000000  441.000000  441.000000  441.000000  441.000000  441.000000
mean       2.320166    7.688070    1.800141    5.717871   11.585805    4.213401
std        1.418692    2.834478    1.152345    3.312948    3.230577    3.946627
min        0.000000    0.667000    0.000000    0.000000    3.614000    0.000000
25%        1.282000    5.572000    0.778000    3.211000    9.302000    1.017000
50%        2.241000    7.389000    1.717000    5.208000   11.502000    2.476000
75%        3.183000    9.722000    2.527000    7.692000   13.839000    7.119000
max        6.736000   15.672000    5.085000   15.038000   20.849000   17.204000

                X12         X13         X14         X15         X16         X17   \
count    441.000000  441.000000  441.000000  441.000000  441.000000  441.000000
mean       2.050841    3.798902    4.145773    5.986317    5.589109    1.262522
```

```
std        1.104153     1.881520     1.849008     1.985780     1.962274     0.977597
min        0.250000     0.000000     0.000000     1.429000     0.990000     0.000000
25%        1.176000     2.471000     2.970000     4.587000     4.242000     0.606000
50%        1.810000     3.676000     3.980000     5.817000     5.537000     1.176000
75%        2.602000     4.923000     5.172000     7.159000     6.897000     1.703000
max        5.628000     9.697000     9.360000    12.183000    11.747000     4.375000


                    X18          X19
count       441.000000   441.000000
mean          3.251293     8.957998
std           1.497945     3.589502
min           0.000000     1.460000
25%           2.232000     6.231000
50%           3.146000     8.564000
75%           4.069000    11.513000
max           7.864000    19.394000
```

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

```
          X0       X1      X2       X3      X4       X5      X6       X7      X8  \
0      9.639    9.639   2.410    5.422   0.000   10.843   3.012    6.024   3.012
1      7.843    3.922   3.922    0.000   0.000    3.922   1.961   15.686   1.961
2      6.952    3.209   4.278    1.604   1.070    8.021   4.813    5.348   3.209
3     14.748   10.432   1.439   11.511   0.000   12.230   1.439    7.554   2.518
4      7.035    1.508   7.035    6.533   0.000    3.015   1.508    8.040   2.513
..       …        …       …        …       …        …       …        …       …
494   10.651    5.917   4.438    9.172   0.592    8.580   1.479   10.947   1.183
495    4.124    8.247   4.124    6.186   0.000    9.278   4.124   10.309   3.093
496   20.064    6.688   0.955    2.866   0.000    2.866   0.318   11.146   0.637
498   10.490    6.294   1.399    2.098   0.000    1.399   4.196   13.287   1.399
499   11.483    3.110   1.675    3.349   0.000    6.220   1.196   11.005   1.914

          X9   …     X11     X12     X13     X14     X15      X16     X17     X18  \
0      5.422   …   0.602   1.205   3.012   8.434   6.024    4.819   1.205   1.205
1     11.765   …   0.000   3.922   5.882   1.961   7.843    3.922   0.000   1.961
2      7.487   …   8.021   3.209   3.209   6.417   6.417    4.813   0.000   2.139
3      1.799   …   0.719   1.079   1.799   5.036   4.676    4.317   0.719   0.000
4      8.040   …   3.015   2.010   6.030   4.523  12.060   10.050   1.508   2.010
..       …   …     …       …       …       …       …        …       …       …
494    3.254   …   2.663   1.479   2.663   5.621   3.550    4.438   2.071   2.959
495    3.093   …   8.247   1.031   3.093   8.247   5.155    7.216   0.000   1.031
496    4.140   …   0.637   1.592   4.140   3.503   6.051    4.777   0.637   2.548
498    1.399   …   0.699   0.699   4.895   5.594   3.497    7.692   3.497   2.098
```

```
499   6.220   …   0.957   2.153   5.502   4.785   4.067   5.024   1.914   2.153

         X19          X20
0     11.446   no_efectores
1      5.882   no_efectores
2      9.626   no_efectores
3      8.633   no_efectores
4      5.025   no_efectores
..        …            …
494   10.651   no_efectores
495    7.216   no_efectores
496   12.420   no_efectores
498   13.986   no_efectores
499   13.636   no_efectores

[415 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   415.000000  415.000000  415.000000  415.000000  415.000000  415.000000
mean     10.070446    6.213275    2.760017    7.961002    0.919643    8.190395
std       3.895501    2.497273    1.778387    3.342429    1.108452    3.546743
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       7.229000    4.554000    1.560000    5.618000    0.000000    5.918500
50%      10.096000    6.211000    2.415000    8.036000    0.588000    8.176000
75%      12.590500    7.819000    3.607500   10.428500    1.262000   10.261000
max      21.012000   13.242000    9.091000   18.301000    5.691000   20.000000

                X6          X7          X8          X9         X10         X11  \
count   415.000000  415.000000  415.000000  415.000000  415.000000  415.000000
mean      2.513501    8.111034    1.976453    4.974781    8.885812    2.920024
std       1.491967    2.717922    1.204204    2.806764    2.980006    2.784070
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       1.477500    6.407500    1.147500    2.854500    7.006500    0.866000
50%       2.273000    7.869000    1.871000    4.310000    8.577000    2.077000
75%       3.374000    9.687000    2.656000    6.822500   10.357500    3.875500
max       7.627000   16.949000    5.882000   13.235000   17.808000   12.727000

               X12         X13         X14         X15         X16         X17  \
count   415.000000  415.000000  415.000000  415.000000  415.000000  415.000000
mean      2.036164    3.383135    4.589494    5.868246    6.068641    1.034299
std       1.111498    1.618050    1.731882    2.333046    2.076796    0.833373
min       0.178000    0.000000    0.000000    0.000000    1.408000    0.000000
25%       1.198000    2.253500    3.502500    4.320500    4.569000    0.386500
```
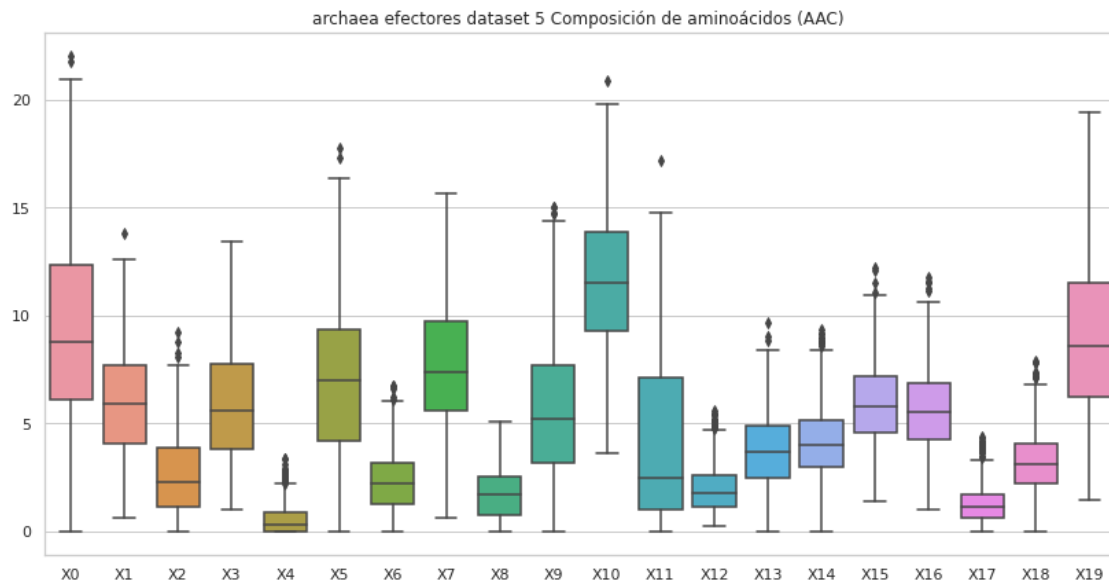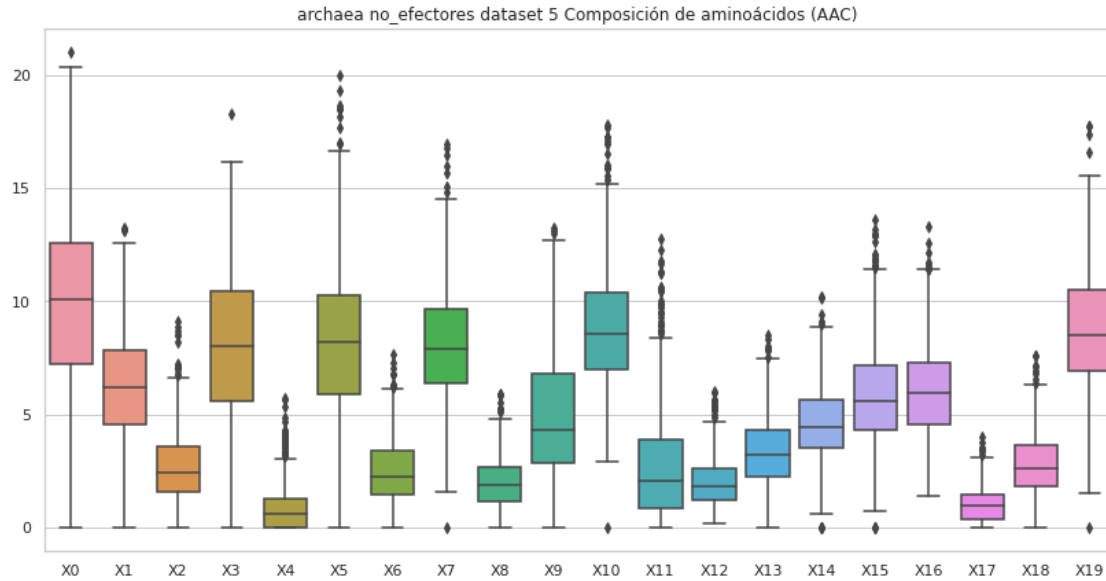
50%	1.818000	3.239000	4.451000	5.612000	5.936000	0.949000
75%	2.605500	4.342000	5.671000	7.171000	7.303500	1.484000
max	6.024000	8.511000	10.227000	13.623000	13.295000	4.000000

	X18	X19
count	415.000000	415.000000
mean	2.845634	8.678024
std	1.444821	2.882019
min	0.000000	0.000000
25%	1.831500	6.897000
50%	2.632000	8.503000
75%	3.636000	10.492000
max	7.619000	17.757000

archaea efectores dataset 5 Composición de aminoácidos (AAC)

archaea no_efectores dataset 5 Composición de aminoácidos (AAC)

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")
```

```python
    #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 archaea dataset
5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6
0	0.032131	0.003570	0.039271	0.049981	0.007140	0.021420	0.017850
1	0.034062	0.004258	0.068125	0.042578	0.012773	0.029804	0.017031
2	0.038258	0.002733	0.032793	0.035525	0.016396	0.040991	0.005465
3	0.038609	0.000000	0.049818	0.066009	0.024909	0.034873	0.009964
4	0.038701	0.002150	0.017201	0.036551	0.008600	0.021501	0.008600
..	…	…	…	…	…	…	
495	0.029837	0.017902	0.029837	0.053707	0.005967	0.023870	0.000000
496	0.020940	0.000000	0.024161	0.018524	0.003222	0.022551	0.001611
497	0.101655	0.011959	0.041858	0.023919	0.059797	0.041858	0.029898
498	0.014235	0.000000	0.008007	0.008897	0.013345	0.017794	0.001779
499	0.031433	0.000000	0.022593	0.032416	0.010805	0.033398	0.015717

	X7	X8	X9	…	X74	X75	X76
0	0.003570	0.007140	0.049981	…	0.030345	0.040621	0.026073
1	0.012773	0.008516	0.059609	…	-0.014710	0.030389	-0.000685
2	0.013664	0.002733	0.021862	…	-0.012054	-0.012787	0.052429
3	0.102127	0.082200	0.120809	…	0.007989	0.006044	0.017487
4	0.008600	0.015050	0.030101	…	0.001252	0.017858	0.020039
..	…	…	…	…	…	…	…
495	0.029837	0.041772	0.071610	…	-0.043543	0.004084	-0.007173
496	0.005638	0.001611	0.014497	…	0.008702	0.003796	0.017636
497	0.041858	0.011959	0.131553	…	-0.040956	-0.016382	-0.062933
498	0.017794	0.001779	0.024911	…	0.026367	0.002834	0.025897
499	0.038309	0.037327	0.030451	…	0.011507	0.015597	0.014704

	X77	X78	X79	X80	X81	X82	X83
0	-0.002740	-0.018590	0.036528	0.068752	0.066852	0.028688	efectores
1	0.031145	0.026239	0.049267	0.041954	0.061963	0.049466	efectores
2	-0.035960	0.015359	0.029611	0.031367	0.006707	0.027408	efectores
3	-0.000425	-0.029927	0.014914	0.023854	-0.008825	0.018088	efectores
4	0.012500	0.012479	0.021411	0.000384	0.024570	-0.006136	efectores
..	…	…	…	…	…	…	
495	-0.014391	-0.062210	0.014221	0.050376	0.080820	-0.001323	efectores
496	0.002872	0.006922	0.034019	0.007539	0.007213	0.021589	efectores

```
497 -0.005947 -0.022547  0.084616  0.053378  0.028270 -0.058229  efectores
498  0.019335  0.000860  0.008928  0.021873  0.004058  0.010641  efectores
499 -0.001066  0.003847  0.010002  0.011830  0.018776  0.024264  efectores

[500 rows x 84 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.033053    0.003820    0.027543    0.034711    0.016669    0.027917
std      0.018955    0.007101    0.021771    0.029914    0.017461    0.014592
min      0.000000    0.000000    0.001285    0.000000    0.000000    0.002551
25%      0.021063    0.000000    0.009670    0.008395    0.007085    0.017712
50%      0.029031    0.000648    0.021705    0.028805    0.012141    0.025005
75%      0.039895    0.004971    0.039885    0.053214    0.020322    0.034553
max      0.229915    0.057479    0.120354    0.172436    0.172436    0.143697

                X6          X7          X8          X9    …         X73  \
count  500.000000  500.000000  500.000000  500.000000    …  500.000000
mean     0.008059    0.028641    0.025973    0.048021    …    0.012986
std      0.008796    0.030256    0.033049    0.032132    …    0.020387
min      0.000000    0.000000    0.000000    0.006060    …   -0.125363
25%      0.002207    0.007834    0.002322    0.025341    …    0.003522
50%      0.006060    0.017498    0.010081    0.040016    …    0.014150
75%      0.010812    0.040621    0.041029    0.060259    …    0.024123
max      0.114958    0.228872    0.247181    0.214525    …    0.089015

                X74         X75         X76         X77         X78         X79  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.005543    0.010247    0.012168    0.006523    0.008807    0.014212
std      0.032151    0.030315    0.021140    0.032705    0.026534    0.026510
min     -0.151316   -0.259160   -0.198191   -0.181949   -0.127898   -0.127461
25%     -0.007308   -0.001131    0.002630   -0.005314   -0.000527    0.005141
50%      0.008518    0.004803    0.014732    0.009541    0.005942    0.014952
75%      0.016304    0.018692    0.023718    0.018857    0.016500    0.025190
max      0.302486    0.208495    0.078401    0.176925    0.159183    0.392967

                X80         X81         X82
count  500.000000  500.000000  500.000000
mean     0.004566    0.009723    0.014218
std      0.035322    0.026849    0.025388
min     -0.248735   -0.166266   -0.096082
25%     -0.003505   -0.001288    0.004064
50%      0.009629    0.005960    0.014744
```

```
75%       0.018706    0.018816    0.024192
max       0.277933    0.177641    0.332001


[8 rows x 83 columns]



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

            X0          X1          X2          X3          X4          X5          X6   \
0      0.041314    0.000000    0.023239    0.046478    0.012910    0.025821    0.012910
1      0.014155    0.000000    0.000000    0.007078    0.010616    0.028310    0.003539
2      0.043060    0.006625    0.009937    0.049685    0.019874    0.033123    0.019874
3      0.038821    0.000000    0.030300    0.032193    0.004734    0.019884    0.006628
4      0.022466    0.000000    0.020861    0.009628    0.019256    0.025675    0.008023
..          …           …           …           …           …           …
495    0.009924    0.000000    0.014886    0.022329    0.007443    0.024810    0.007443
496    0.030739    0.000000    0.004391    0.004391    0.006343    0.017077    0.000976
497    0.021559    0.021559    0.021559    0.032339    0.000000    0.043119    0.010780
498    0.017568    0.000000    0.003514    0.002342    0.008198    0.022253    0.002342
499    0.018769    0.000000    0.005474    0.010166    0.008993    0.017986    0.003128


            X7          X8          X9   …        X74         X75         X76   \
0      0.023239    0.002582    0.028403   …  -0.002224    0.013604    0.024038
1      0.021233    0.000000    0.031849   …   0.028105    0.002216    0.037118
2      0.046373    0.049685    0.062934   …   0.022225    0.023072   -0.014675
3      0.004734    0.001894    0.024618   …  -0.008388    0.011759    0.007470
4      0.025675    0.009628    0.027280   …   0.015130    0.020937    0.022764
..          …           …           …   …        …           …           …
495    0.007443    0.019848    0.014886   …  -0.012378    0.017933    0.019453
496    0.006343    0.000976    0.021468   …   0.012736    0.002698    0.022042
497    0.053898    0.086237    0.097017   …   0.090499    0.014712    0.042823
498    0.002342    0.001171    0.025766   …   0.008487    0.001696    0.015716
499    0.010166    0.001564    0.022288   …   0.015666    0.001234    0.020184


            X77         X78         X79         X80         X81         X82          X83
0     -0.029026   -0.002311    0.001257   -0.019049    0.012621    0.009777    no_efectores
1      0.045409    0.014562    0.059032    0.054930    0.029398    0.046771    no_efectores
2     -0.042902   -0.016630   -0.033172   -0.028778   -0.019921    0.018600    no_efectores
3      0.025046    0.032289    0.025294   -0.003857    0.007218    0.019302    no_efectores
4      0.019514    0.004314    0.024086    0.010692   -0.003632    0.008278    no_efectores
..          …           …           …           …           …           …
495   -0.005855    0.012639    0.034069    0.009992    0.017327    0.034438    no_efectores
496    0.003135   -0.002944    0.022257    0.010424   -0.001139    0.024057    no_efectores
497   -0.080005   -0.012320   -0.016925   -0.057973   -0.020179    0.054680    no_efectores
```

```
498   0.013257   0.002864   0.001786   0.017466   0.006173   0.013407   no_efectores
499   0.012157  -0.003024   0.015766   0.016778   0.004048   0.015117   no_efectores

[500 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	X5
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.033977	0.004879	0.031608	0.034094	0.014822	0.029827
std	0.015580	0.008824	0.025205	0.028258	0.018873	0.017721
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.023634	0.000000	0.015724	0.016119	0.006154	0.019626
50%	0.031479	0.001998	0.029147	0.030418	0.010551	0.026799
75%	0.042460	0.005553	0.041205	0.045656	0.017668	0.036060
max	0.123226	0.082150	0.246451	0.250473	0.250473	0.205376

	X6	X7	X8	X9	…	X73
count	500.000000	500.000000	500.000000	500.000000	…	500.000000
mean	0.008682	0.023134	0.015912	0.034800	…	0.018774
std	0.010840	0.031284	0.026854	0.026294	…	0.022908
min	0.000000	0.000000	0.000000	0.000000	…	-0.163038
25%	0.002418	0.007315	0.002122	0.019843	…	0.008635
50%	0.006216	0.014699	0.007327	0.028661	…	0.019686
75%	0.010718	0.028445	0.017530	0.041461	…	0.028539
max	0.125237	0.424216	0.269956	0.308521	…	0.257635

	X74	X75	X76	X77	X78	X79
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.001864	0.007312	0.015938	0.002683	0.006870	0.017679
std	0.025696	0.026879	0.021141	0.033102	0.037380	0.021574
min	-0.146588	-0.327295	-0.095619	-0.233752	-0.320985	-0.118488
25%	-0.007445	-0.000873	0.005389	-0.006716	-0.002362	0.007795
50%	0.004087	0.005971	0.018535	0.004477	0.005167	0.019389
75%	0.014391	0.015781	0.026368	0.014368	0.016732	0.027424
max	0.192992	0.128021	0.226047	0.452510	0.412160	0.182883

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.001127	0.008049	0.016369
std	0.048667	0.039375	0.027791
min	-0.863266	-0.636708	-0.411351
25%	-0.007076	-0.000847	0.008298
50%	0.004253	0.006872	0.018934
75%	0.014001	0.019213	0.027795

```
max       0.181416      0.139340      0.114648
```

[8 rows x 83 columns]

archaea efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro_mass con valores atípicos.

archaea no_efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro_mass con valores atípicos.
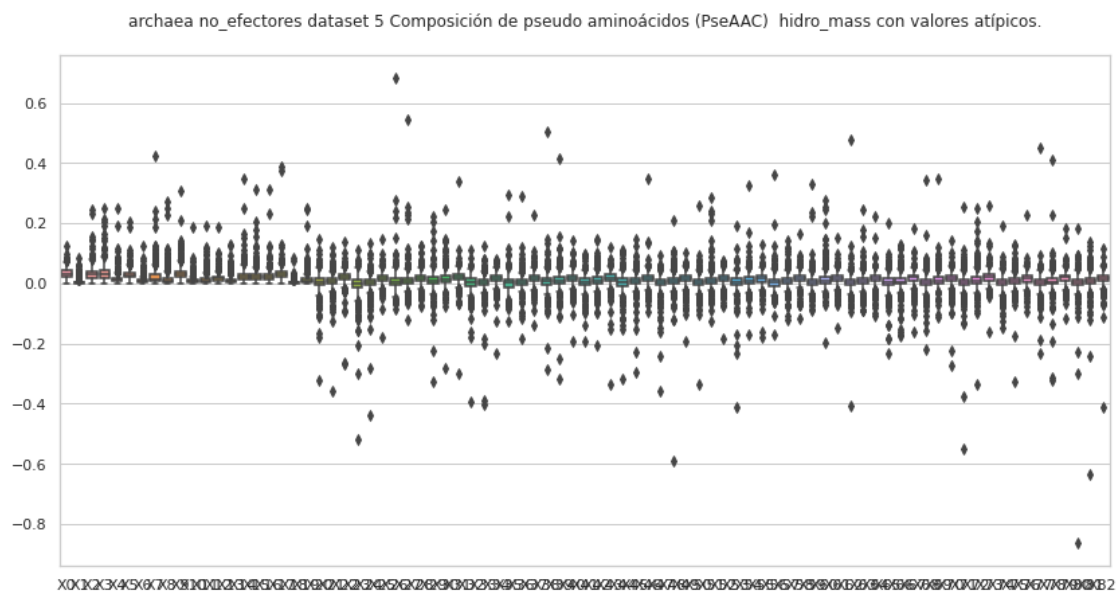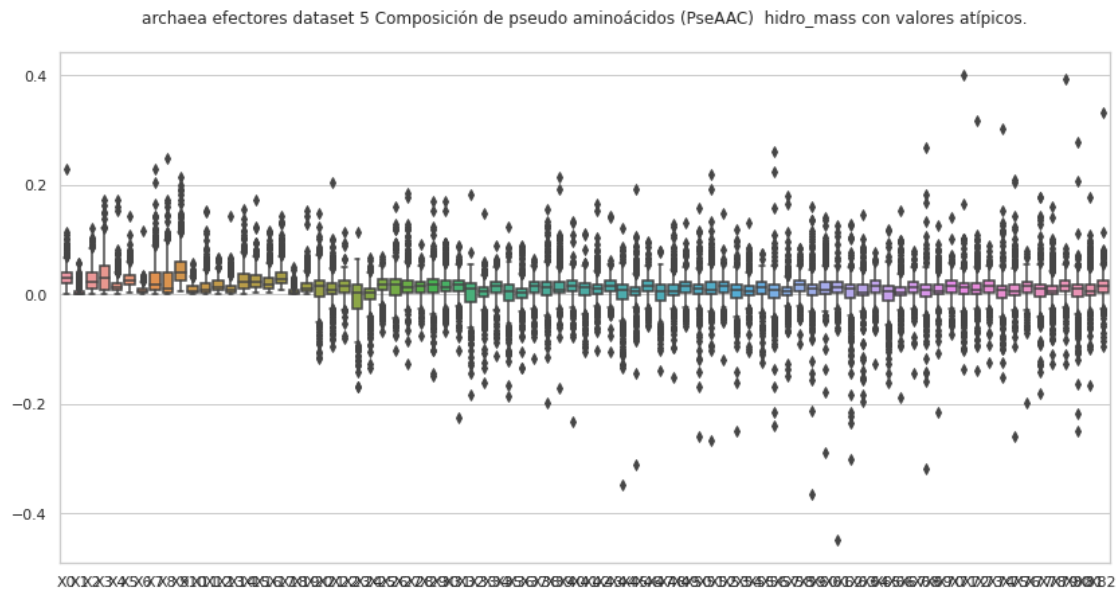
3.1 Composición de pseudo aminoácidos (PseAAC) hidro_mass, sin valores atípicos

```python
#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 archaea dataset
5, sin valores atípicos.
Valores del documento csv.

```
            X0        X1        X2        X3        X4        X5        X6  \
0     0.032131  0.003570  0.039271  0.049981  0.007140  0.021420  0.017850
1     0.034062  0.004258  0.068125  0.042578  0.012773  0.029804  0.017031
2     0.038258  0.002733  0.032793  0.035525  0.016396  0.040991  0.005465
3     0.038609  0.000000  0.049818  0.066009  0.024909  0.034873  0.009964
4     0.038701  0.002150  0.017201  0.036551  0.008600  0.021501  0.008600
..         …         …         …         …         …         …
493   0.020964  0.000000  0.025623  0.016306  0.004659  0.023294  0.001165
494   0.015410  0.000000  0.017464  0.020546  0.023628  0.020546  0.002055
496   0.020940  0.000000  0.024161  0.018524  0.003222  0.022551  0.001611
498   0.014235  0.000000  0.008007  0.008897  0.013345  0.017794  0.001779
499   0.031433  0.000000  0.022593  0.032416  0.010805  0.033398  0.015717

            X7        X8        X9    …       X74       X75       X76  \
0     0.003570  0.007140  0.049981  …  0.030345  0.040621  0.026073
1     0.012773  0.008516  0.059609  … -0.014710  0.030389 -0.000685
2     0.013664  0.002733  0.021862  … -0.012054 -0.012787  0.052429
3     0.102127  0.082200  0.120809  …  0.007989  0.006044  0.017487
4     0.008600  0.015050  0.030101  …  0.001252  0.017858  0.020039
..         …         …         …   …        …         …         …
493   0.003494  0.003494  0.025623  … -0.013797 -0.000714  0.014603
494   0.056502  0.021573  0.040065  …  0.004554 -0.006034  0.000650
496   0.005638  0.001611  0.014497  …  0.008702  0.003796  0.017636
498   0.017794  0.001779  0.024911  …  0.026367  0.002834  0.025897
499   0.038309  0.037327  0.030451  …  0.011507  0.015597  0.014704

            X77       X78       X79       X80       X81       X82       X83
0    -0.002740 -0.018590  0.036528  0.068752  0.066852  0.028688  efectores
1     0.031145  0.026239  0.049267  0.041954  0.061963  0.049466  efectores
2    -0.035960  0.015359  0.029611  0.031367  0.006707  0.027408  efectores
3    -0.000425 -0.029927  0.014914  0.023854 -0.008825  0.018088  efectores
4     0.012500  0.012479  0.021411  0.000384  0.024570 -0.006136  efectores
..         …         …         …         …         …         …
493   0.004957  0.010818  0.014594  0.006657  0.016158  0.009374  efectores
494   0.041919  0.013727  0.003140  0.019904 -0.005409  0.016600  efectores
496   0.002872  0.006922  0.034019  0.007539  0.007213  0.021589  efectores
498   0.019335  0.000860  0.008928  0.021873  0.004058  0.010641  efectores
499  -0.001066  0.003847  0.010002  0.011830  0.018776  0.024264  efectores
```

[398 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5
count	398.000000	398.000000	398.000000	398.000000	398.000000	398.000000
mean	0.029053	0.002216	0.020979	0.024579	0.013023	0.024059
std	0.012374	0.003661	0.015645	0.020268	0.009596	0.009505
min	0.005020	0.000000	0.001285	0.000000	0.000000	0.002551
25%	0.020402	0.000000	0.008210	0.006360	0.006455	0.016893
50%	0.026392	0.000384	0.016802	0.020036	0.010377	0.022356
75%	0.035970	0.003038	0.030828	0.037563	0.016564	0.028500
max	0.082851	0.020081	0.068125	0.087300	0.064602	0.064548

	X6	X7	X8	X9	…	X73
count	398.000000	398.000000	398.000000	398.000000	…	398.000000
mean	0.006210	0.020463	0.016034	0.036829	…	0.015113
std	0.005328	0.019114	0.020233	0.017906	…	0.014542
min	0.000000	0.000000	0.000000	0.006060	…	-0.033699
25%	0.001956	0.006693	0.001784	0.022885	…	0.006695
50%	0.005261	0.012467	0.005845	0.034019	…	0.015014
75%	0.008833	0.030913	0.026622	0.046660	…	0.024094
max	0.030642	0.102127	0.096903	0.120809	…	0.061876

	X74	X75	X76	X77	X78	X79
count	398.000000	398.000000	398.000000	398.000000	398.000000	398.000000
mean	0.006341	0.007136	0.015573	0.008128	0.008088	0.014747
std	0.017036	0.014956	0.013281	0.018208	0.015089	0.014255
min	-0.062745	-0.038717	-0.037411	-0.056055	-0.043962	-0.040495
25%	-0.004299	-0.000833	0.007642	-0.001976	0.000237	0.007091
50%	0.009088	0.004112	0.016840	0.010639	0.005560	0.015215
75%	0.015700	0.012567	0.024237	0.018400	0.013779	0.023472
max	0.062991	0.062326	0.053153	0.091495	0.064331	0.075919

	X80	X81	X82
count	398.000000	398.000000	398.000000
mean	0.008507	0.008583	0.015018
std	0.017299	0.016081	0.014835
min	-0.056739	-0.054481	-0.044129
25%	0.000099	-0.000641	0.007443
50%	0.010983	0.005266	0.015947
75%	0.017836	0.015809	0.023804
max	0.068752	0.080332	0.067801

[8 rows x 83 columns]

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

```
           X0        X1        X2        X3        X4        X5        X6   \
0     0.041314  0.000000  0.023239  0.046478  0.012910  0.025821  0.012910
1     0.014155  0.000000  0.000000  0.007078  0.010616  0.028310  0.003539
3     0.038821  0.000000  0.030300  0.032193  0.004734  0.019884  0.006628
4     0.022466  0.000000  0.020861  0.009628  0.019256  0.025675  0.008023
5     0.056841  0.005545  0.047136  0.047136  0.018023  0.041591  0.012477
..         …         …         …         …         …         …
494   0.031696  0.001761  0.027294  0.025533  0.007924  0.032577  0.003522
495   0.009924  0.000000  0.014886  0.022329  0.007443  0.024810  0.007443
496   0.030739  0.000000  0.004391  0.004391  0.006343  0.017077  0.000976
498   0.017568  0.000000  0.003514  0.002342  0.008198  0.022253  0.002342
499   0.018769  0.000000  0.005474  0.010166  0.008993  0.017986  0.003128

           X7        X8        X9   …       X74       X75       X76   \
0     0.023239  0.002582  0.028403  … -0.002224  0.013604  0.024038
1     0.021233  0.000000  0.031849  …  0.028105  0.002216  0.037118
3     0.004734  0.001894  0.024618  … -0.008388  0.011759  0.007470
4     0.025675  0.009628  0.027280  …  0.015130  0.020937  0.022764
5     0.009704  0.012477  0.033273  …  0.008336  0.012512  0.023703
..         …         …         …  …        …         …         …
494   0.009685  0.007924  0.022892  … -0.008057  0.003412  0.028573
495   0.007443  0.019848  0.014886  … -0.012378  0.017933  0.019453
496   0.006343  0.000976  0.021468  …  0.012736  0.002698  0.022042
498   0.002342  0.001171  0.025766  …  0.008487  0.001696  0.015716
499   0.010166  0.001564  0.022288  …  0.015666  0.001234  0.020184

           X77       X78       X79       X80       X81       X82          X83
0     -0.029026 -0.002311  0.001257 -0.019049  0.012621  0.009777  no_efectores
1      0.045409  0.014562  0.059032  0.054930  0.029398  0.046771  no_efectores
3      0.025046  0.032289  0.025294 -0.003857  0.007218  0.019302  no_efectores
4      0.019514  0.004314  0.024086  0.010692 -0.003632  0.008278  no_efectores
5     -0.004120 -0.000163  0.005811 -0.009555  0.027510  0.015538  no_efectores
..         …         …         …         …         …         …
494    0.002010  0.003760  0.029778  0.003408  0.004952  0.031374  no_efectores
495   -0.005855  0.012639  0.034069  0.009992  0.017327  0.034438  no_efectores
496    0.003135 -0.002944  0.022257  0.010424 -0.001139  0.024057  no_efectores
498    0.013257  0.002864  0.001786  0.017466  0.006173  0.013407  no_efectores
499    0.012157 -0.003024  0.015766  0.016778  0.004048  0.015117  no_efectores

[427 rows x 84 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   427.000000  427.000000  427.000000  427.000000  427.000000  427.000000
mean      0.032531    0.003492    0.027566    0.028469    0.011355    0.026858
std       0.013030    0.004862    0.016576    0.018146    0.007762    0.010429
min       0.002218    0.000000    0.000000    0.000000    0.000000    0.000000
25%       0.022920    0.000000    0.014604    0.013623    0.005844    0.019360
50%       0.031214    0.001892    0.027339    0.027997    0.009737    0.025942
75%       0.040662    0.004573    0.038964    0.039141    0.015395    0.033626
max       0.069020    0.028751    0.086699    0.086010    0.050300    0.063687

                X6          X7          X8          X9    …         X73  \
count   427.000000  427.000000  427.000000  427.000000    …   427.000000
mean      0.006912    0.017238    0.010387    0.029755    …     0.019328
std       0.006247    0.015232    0.012879    0.015196    …     0.013674
min       0.000000    0.000000    0.000000    0.000000    …    -0.018335
25%       0.002327    0.006534    0.001909    0.018958    …     0.010756
50%       0.005513    0.012715    0.006050    0.026748    …     0.020275
75%       0.009368    0.022841    0.013695    0.038749    …     0.027280
max       0.037512    0.097006    0.078980    0.089821    …     0.083165

                X74         X75         X76         X77         X78         X79  \
count   427.000000  427.000000  427.000000  427.000000  427.000000  427.000000
mean      0.003697    0.007757    0.017505    0.003861    0.007544    0.018606
std       0.016041    0.015179    0.014304    0.017041    0.015399    0.014665
min      -0.049829   -0.053372   -0.043153   -0.053280   -0.045132   -0.041112
25%      -0.004795   -0.000507    0.009253   -0.005854   -0.001150    0.010369
50%       0.004655    0.005649    0.019453    0.004490    0.005224    0.019802
75%       0.014124    0.015023    0.026433    0.013223    0.015221    0.027334
max       0.066268    0.083001    0.059395    0.063048    0.072270    0.062844

                X80         X81         X82
count   427.000000  427.000000  427.000000
mean      0.005422    0.009767    0.019334
std       0.017386    0.017166    0.013779
min      -0.065102   -0.060258   -0.032034
25%      -0.003949   -0.000235    0.011569
50%       0.005821    0.007218    0.020068
75%       0.013897    0.018520    0.028115
max       0.093060    0.085430    0.061151
```

[8 rows x 83 columns]

archaea efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro_mass

archaea no_efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro_mass

4 Composición de pseudo aminoácidos (PseAAC) mass

```
[7]:  #mass
      transf = "Composición de pseudo aminoácidos (PseAAC) "
      transf2 = "PseAAC"
      estado = "con valores atípicos.\n"
```

```python
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 archaea dataset 5,
con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6
0	0.058184	0.006465	0.071113	0.090508	0.012930	0.038789	0.032324
1	0.051822	0.006478	0.103644	0.064777	0.019433	0.045344	0.025911
2	0.043686	0.003120	0.037445	0.040565	0.018722	0.046806	0.006241
3	0.035131	0.000000	0.045330	0.060063	0.022665	0.031731	0.009066
4	0.074805	0.004156	0.033247	0.070649	0.016623	0.041559	0.016623
..	…	…	…	…	…	…	…
495	0.038766	0.023260	0.038766	0.069779	0.007753	0.031013	0.000000
496	0.027768	0.000000	0.032040	0.024564	0.004272	0.029904	0.002136
497	0.123088	0.014481	0.050683	0.028962	0.072404	0.050683	0.036202
498	0.025881	0.000000	0.014558	0.016175	0.024263	0.032351	0.003235
499	0.037659	0.000000	0.027068	0.038836	0.012945	0.040013	0.018830

	X7	X8	X9	…	X32	X33	X34
0	0.006465	0.012930	0.090508	…	0.010420	-0.057312	-0.030590

```
1     0.019433   0.012955   0.090688   …   0.053093  -0.026523  -0.023651
2     0.015602   0.003120   0.024963   …   0.034952   0.022102   0.002972
3     0.092927   0.074795   0.109926   …   0.009670   0.004804   0.017354
4     0.016623   0.029091   0.058182   …   0.053737   0.017927   0.049514
..       …          …          …  …        …          …          …
495   0.038766   0.054272   0.093038   …  -0.013678  -0.007883  -0.017838
496   0.007476   0.002136   0.019224   …   0.017191   0.038305   0.044245
497   0.050683   0.014481   0.159290   …   0.042646  -0.028886  -0.078368
498   0.032351   0.003235   0.045291   …   0.044430   0.040855   0.010876
499   0.045898   0.044721   0.036483   …   0.026727   0.030438   0.022961

           X35        X36        X37        X38        X39        X40        X41
0    -0.038107   0.031999  -0.005227   0.047214   0.066146   0.051950   efectores
1    -0.042046  -0.032757   0.015205  -0.001043   0.074953   0.075257   efectores
2     0.021170   0.030655   0.055583   0.059867   0.033812   0.031297   efectores
3     0.007951   0.018663  -0.002570   0.015911   0.013570   0.016459   efectores
4     0.014491  -0.011390   0.014891   0.038733   0.041385  -0.011861   efectores
..       …          …          …          …          …          …
495   0.015323   0.013533  -0.018331  -0.009320   0.018477  -0.001719   efectores
496   0.041614   0.024667   0.031977   0.023387   0.045112   0.028628   efectores
497   0.009145  -0.018017   0.034331  -0.076202   0.102456  -0.070506   efectores
498   0.026886   0.013342   0.043073   0.047084   0.016231   0.019347   efectores
499   0.024011   0.015474   0.034002   0.017616   0.011983   0.029070   efectores

[500 rows x 42 columns]
```

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

```
              X0          X1          X2          X3          X4          X5   \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046100    0.004758    0.036617    0.047109    0.023091    0.038918
std      0.016961    0.007755    0.023483    0.038032    0.018137    0.012257
min      0.000000    0.000000    0.002634    0.000000    0.000000    0.007119
25%      0.034493    0.000000    0.017362    0.015545    0.011102    0.030436
50%      0.044217    0.001079    0.031175    0.038758    0.017936    0.037418
75%      0.054920    0.006571    0.053845    0.069509    0.030032    0.045600
max      0.140613    0.049086    0.115018    0.215069    0.123185    0.098454

              X6          X7          X8          X9   …         X31   \
count  500.000000  500.000000  500.000000  500.000000  …   500.000000
mean     0.010909    0.038714    0.034613    0.067085  …     0.014573
std      0.009247    0.033629    0.040644    0.032872  …     0.029040
min      0.000000    0.000000    0.000000    0.006492  …    -0.125805
25%      0.003602    0.013400    0.003500    0.042470  …     0.000259
50%      0.009255    0.026521    0.014377    0.061597  …     0.019991
```

```
75%        0.015524    0.056990    0.056259    0.087185    …    0.033378
max        0.054522    0.169847    0.215039    0.233837    …    0.134952

                  X32         X33         X34         X35         X36         X37   \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.021487    0.016626    0.017740    0.015499    0.019248    0.019267
std       0.025378    0.027740    0.026810    0.025686    0.025771    0.025908
min      -0.089636   -0.195977   -0.131426   -0.097660   -0.105433   -0.122978
25%       0.008960    0.003930    0.005966    0.002888    0.003843    0.004461
50%       0.025492    0.021883    0.022774    0.019602    0.024638    0.024153
75%       0.036923    0.033884    0.033873    0.033435    0.034810    0.036526
max       0.105707    0.137814    0.088267    0.071287    0.145221    0.085070

                  X38         X39         X40
count   500.000000  500.000000  500.000000
mean      0.018774    0.020485    0.020135
std       0.025367    0.026606    0.028199
min      -0.086607   -0.101826   -0.098057
25%       0.004240    0.008294    0.004864
50%       0.025080    0.024455    0.024174
75%       0.035879    0.034666    0.036857
max       0.077315    0.171722    0.145081

[8 rows x 41 columns]


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

              X0          X1          X2          X3          X4          X5          X6   \
0       0.061371    0.000000    0.034521    0.069042    0.019178    0.038357    0.019178
1       0.023913    0.000000    0.000000    0.011956    0.017935    0.047826    0.005978
2       0.039473    0.006073    0.009109    0.045546    0.018218    0.030364    0.018218
3       0.062484    0.000000    0.048768    0.051816    0.007620    0.032004    0.010668
4       0.028035    0.000000    0.026033    0.012015    0.024030    0.032040    0.010013
..           …           …           …           …           …           …           …
495     0.020998    0.000000    0.031497    0.047245    0.015748    0.052495    0.015748
496     0.043091    0.000000    0.006156    0.006156    0.008892    0.023940    0.001368
497     0.020071    0.020071    0.020071    0.030107    0.000000    0.040143    0.010036
498     0.033076    0.000000    0.006615    0.004410    0.015435    0.041896    0.004410
499     0.038681    0.000000    0.011282    0.020952    0.018535    0.037070    0.006447

              X7          X8          X9   …        X32         X33         X34   \
0       0.034521    0.003836    0.042192    …    0.027822    0.007330   -0.019243
1       0.035869    0.000000    0.053804    …    0.049177    0.048769    0.032229
```

```
2    0.042509   0.045546   0.057691   …   0.015585   0.009060   0.011905
3    0.007620   0.003048   0.039624   …   0.039623   0.027653   0.034498
4    0.032040   0.012015   0.034043   …   0.038795   0.038973   0.035934
..      …          …          …     …  …      …          …          …
495  0.015748   0.041996   0.031497   …   0.019678  -0.010206   0.025994
496  0.008892   0.001368   0.030096   …   0.027777   0.036170   0.033842
497  0.050178   0.080285   0.090321   …   0.009664   0.023268   0.013004
498  0.004410   0.002205   0.048511   …   0.037378   0.037832   0.031182
499  0.020952   0.003223   0.045934   …   0.019198   0.027226   0.020961

            X35        X36        X37        X38        X39        X40           X41
0    -0.020882   0.039161   0.033871   0.035709   0.001867   0.014523   no_efectores
1     0.020852   0.015337   0.026114   0.062706   0.099725   0.079013   no_efectores
2     0.024316   0.032051   0.033608  -0.013453  -0.030408   0.017050   no_efectores
3     0.031823   0.025006   0.029805   0.012024   0.040712   0.031067   no_efectores
4     0.047520   0.014474   0.021851   0.028407   0.030057   0.010331   no_efectores
..       …          …          …          …          …          …
495   0.031999   0.017357   0.033078   0.041158   0.072084   0.072866   no_efectores
496   0.028177   0.042904   0.046757   0.030899   0.031200   0.033724   no_efectores
497  -0.009068   0.047488  -0.019490   0.039868  -0.015757   0.050906   no_efectores
498   0.020477   0.011074   0.040424   0.029589   0.003363   0.025241   no_efectores
499   0.031643   0.018850   0.031422   0.041599   0.032492   0.031157   no_efectores

[500 rows x 42 columns]
```

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

```
              X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046389    0.006270    0.042149    0.046594    0.018936    0.039345
std      0.022898    0.011699    0.029848    0.037797    0.015692    0.017935
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034337    0.000000    0.020695    0.021003    0.008690    0.029325
50%      0.043632    0.002527    0.036941    0.038844    0.014946    0.037162
75%      0.055502    0.007070    0.054733    0.060407    0.024578    0.045682
max      0.336494    0.141365    0.209351    0.336494    0.106024    0.212047

              X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.010997    0.028769    0.020556    0.046169  …    0.019119
std      0.011110    0.027193    0.029611    0.026274  …    0.029856
min      0.000000    0.000000    0.000000    0.000000  …   -0.153184
25%      0.003859    0.010352    0.002920    0.028301  …    0.007104
50%      0.008041    0.020249    0.009735    0.040489  …    0.024610
75%      0.014471    0.039733    0.025453    0.056928  …    0.035070
```

```
max       0.098457      0.238341      0.234209      0.208186  …      0.117321

                X32           X33           X34           X35           X36           X37  \
count    500.000000    500.000000    500.000000    500.000000    500.000000    500.000000
mean       0.023116      0.022363      0.020774      0.022516      0.021885      0.024351
std        0.024652      0.027506      0.031081      0.028220      0.030988      0.025435
min       -0.115029     -0.138290     -0.156137     -0.130233     -0.218371     -0.138220
25%        0.012842      0.012006      0.010465      0.012899      0.011498      0.012884
50%        0.026117      0.026695      0.025801      0.026965      0.024631      0.026931
75%        0.036180      0.037553      0.035083      0.035775      0.035850      0.037994
max        0.163320      0.178022      0.177452      0.211351      0.237866      0.130740

                X38           X39           X40
count    500.000000    500.000000    500.000000
mean       0.020589      0.022978      0.022440
std        0.026732      0.029909      0.031768
min       -0.156903     -0.206207     -0.327694
25%        0.009253      0.010848      0.012515
50%        0.025368      0.025804      0.026462
75%        0.034907      0.036276      0.036720
max        0.194085      0.233876      0.138717

[8 rows x 41 columns]
```
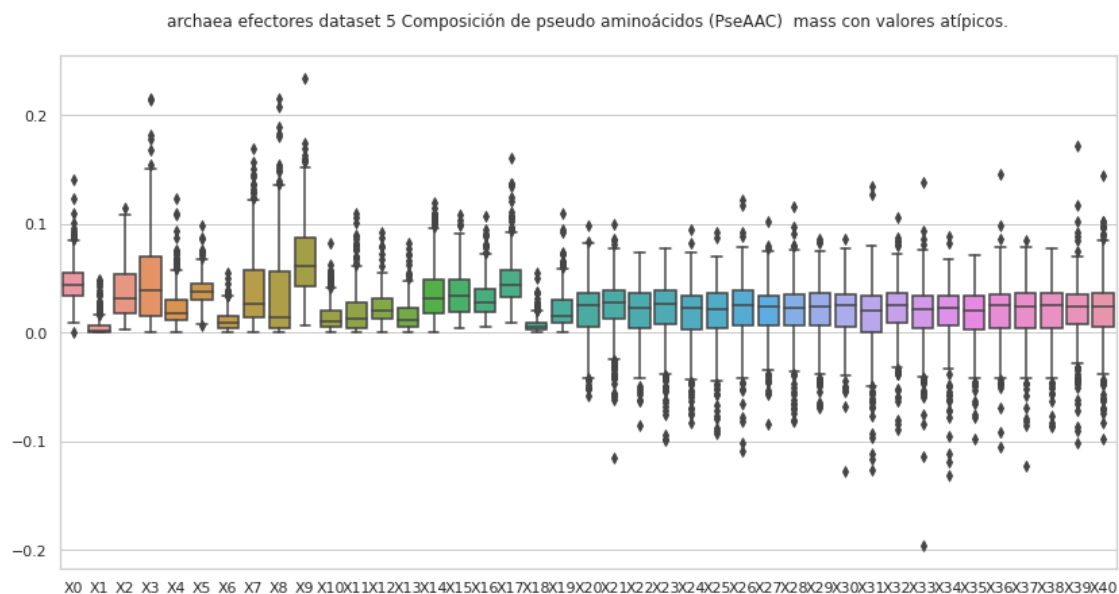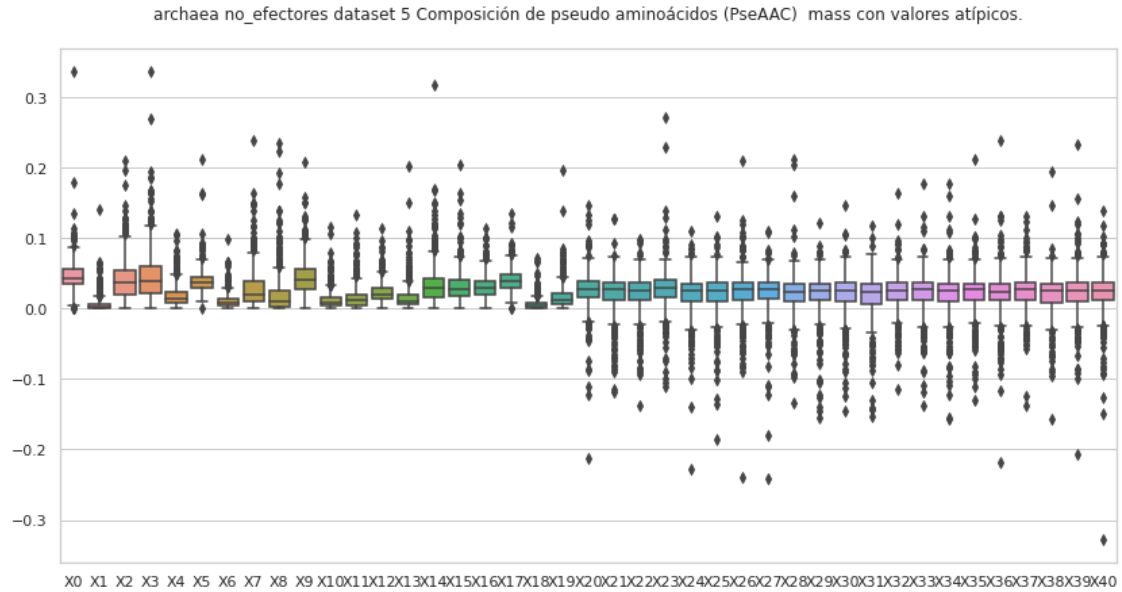
archaea efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) mass con valores atípicos.

archaea no_efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) mass con valores atípicos.

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 archaea dataset 5,
sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
1   0.051822  0.006478  0.103644  0.064777  0.019433  0.045344  0.025911
2   0.043686  0.003120  0.037445  0.040565  0.018722  0.046806  0.006241
3   0.035131  0.000000  0.045330  0.060063  0.022665  0.031731  0.009066
4   0.074805  0.004156  0.033247  0.070649  0.016623  0.041559  0.016623
6   0.047337  0.000000  0.010519  0.026298  0.015779  0.036818  0.000000
..       ...       ...       ...       ...       ...       ...
494 0.030419  0.000000  0.034475  0.040559  0.046643  0.040559  0.004056
495 0.038766  0.023260  0.038766  0.069779  0.007753  0.031013  0.000000
496 0.027768  0.000000  0.032040  0.024564  0.004272  0.029904  0.002136
498 0.025881  0.000000  0.014558  0.016175  0.024263  0.032351  0.003235
499 0.037659  0.000000  0.027068  0.038836  0.012945  0.040013  0.018830

          X7        X8        X9  …       X32       X33       X34  \
1   0.019433  0.012955  0.090688  …  0.053093 -0.026523 -0.023651
2   0.015602  0.003120  0.024963  …  0.034952  0.022102  0.002972
3   0.092927  0.074795  0.109926  …  0.009670  0.004804  0.017354
4   0.016623  0.029091  0.058182  …  0.053737  0.017927  0.049514
6   0.042077  0.036818  0.052596  …  0.001768  0.025714  0.037663
..       ...       ...       ... …       ...       ...       ...
494 0.111538  0.042587  0.079091  …  0.004195  0.018843  0.007177
495 0.038766  0.054272  0.093038  … -0.013678 -0.007883 -0.017838
496 0.007476  0.002136  0.019224  …  0.017191  0.038305  0.044245
498 0.032351  0.003235  0.045291  …  0.044430  0.040855  0.010876
499 0.045898  0.044721  0.036483  …  0.026727  0.030438  0.022961
```

```
           X35        X36        X37        X38        X39        X40        X41
1     -0.042046  -0.032757   0.015205  -0.001043   0.074953   0.075257   efectores
2      0.021170   0.030655   0.055583   0.059867   0.033812   0.031297   efectores
3      0.007951   0.018663  -0.002570   0.015911   0.013570   0.016459   efectores
4      0.014491  -0.011390   0.014891   0.038733   0.041385  -0.011861   efectores
6      0.061706   0.030541   0.001235  -0.021888   0.053823   0.047657   efectores
..          …          …          …          …          …          …
494    0.035723   0.021460   0.024554   0.001283   0.006198   0.032769   efectores
495    0.015323   0.013533  -0.018331  -0.009320   0.018477  -0.001719   efectores
496    0.041614   0.024667   0.031977   0.023387   0.045112   0.028628   efectores
498    0.026886   0.013342   0.043073   0.047084   0.016231   0.019347   efectores
499    0.024011   0.015474   0.034002   0.017616   0.011983   0.029070   efectores

[402 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   402.000000  402.000000  402.000000  402.000000  402.000000  402.000000
mean      0.044782    0.002991    0.031008    0.036868    0.020329    0.037206
std       0.014155    0.004481    0.020316    0.029129    0.014847    0.010164
min       0.011563    0.000000    0.002634    0.000000    0.000000    0.017252
25%       0.034509    0.000000    0.015219    0.013588    0.010526    0.029589
50%       0.043882    0.000716    0.025231    0.030149    0.016535    0.036153
75%       0.053227    0.005089    0.043375    0.054629    0.025196    0.042762
max       0.090856    0.027588    0.103644    0.149800    0.076047    0.066856

                X6          X7          X8          X9   …         X31  \
count   402.000000  402.000000  402.000000  402.000000   …  402.000000
mean      0.009137    0.032392    0.025448    0.058548   …    0.019352
std       0.007259    0.028846    0.031513    0.026423   …    0.020505
min       0.000000    0.000000    0.000000    0.006492   …   -0.064032
25%       0.003240    0.011234    0.002808    0.041013   …    0.008006
50%       0.007560    0.021616    0.009413    0.053244   …    0.023661
75%       0.013304    0.045745    0.039527    0.077291   …    0.034168
max       0.034637    0.127756    0.152118    0.158514   …    0.064497

               X32         X33         X34         X35         X36         X37  \
count   402.000000  402.000000  402.000000  402.000000  402.000000  402.000000
mean      0.024002    0.019881    0.022505    0.019797    0.021983    0.023849
std       0.020344    0.021290    0.019429    0.021628    0.020714    0.020317
min      -0.037849   -0.059348   -0.062229   -0.049607   -0.051006   -0.051326
25%       0.013977    0.009175    0.012848    0.008948    0.010836    0.011096
50%       0.027586    0.023504    0.026180    0.022664    0.026712    0.026432
75%       0.037030    0.033911    0.035870    0.034211    0.035194    0.037703
```

```
max      0.081365     0.093728     0.088267     0.071287     0.065418     0.079151

                X38          X39          X40
count   402.000000   402.000000   402.000000
mean      0.023062     0.023762     0.024090
std       0.019786     0.019201     0.021203
min      -0.033353    -0.044517    -0.046285
25%       0.010756     0.013846     0.011253
50%       0.027030     0.026219     0.026118
75%       0.036652     0.034953     0.037223
max       0.073128     0.092463     0.101631

[8 rows x 41 columns]
```

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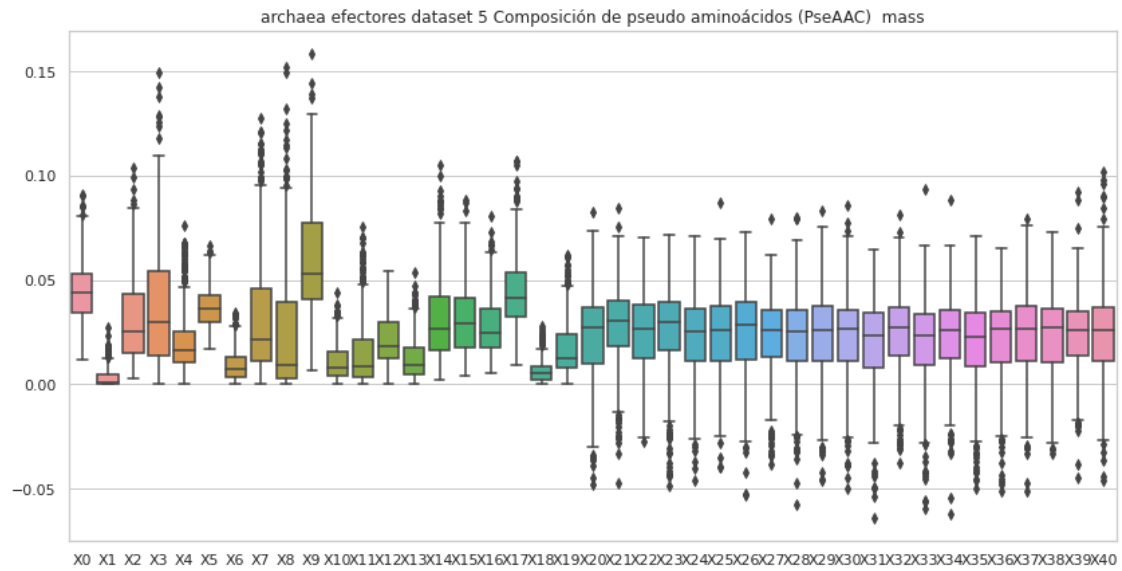sin valores atípicos.
Valores del documento csv.

```
            X0          X1          X2          X3          X4          X5          X6  \
0     0.061371    0.000000    0.034521    0.069042    0.019178    0.038357    0.019178
1     0.023913    0.000000    0.000000    0.011956    0.017935    0.047826    0.005978
2     0.039473    0.006073    0.009109    0.045546    0.018218    0.030364    0.018218
3     0.062484    0.000000    0.048768    0.051816    0.007620    0.032004    0.010668
4     0.028035    0.000000    0.026033    0.012015    0.024030    0.032040    0.010013
..         …           …           …           …           …           …           …
494   0.036015    0.002001    0.031012    0.029012    0.009004    0.037015    0.004002
495   0.020998    0.000000    0.031497    0.047245    0.015748    0.052495    0.015748
496   0.043091    0.000000    0.006156    0.006156    0.008892    0.023940    0.001368
498   0.033076    0.000000    0.006615    0.004410    0.015435    0.041896    0.004410
499   0.038681    0.000000    0.011282    0.020952    0.018535    0.037070    0.006447

            X7          X8          X9    …        X32          X33          X34  \
0     0.034521    0.003836    0.042192    …    0.027822    0.007330   -0.019243
1     0.035869    0.000000    0.053804    …    0.049177    0.048769    0.032229
2     0.042509    0.045546    0.057691    …    0.015585    0.009060    0.011905
3     0.007620    0.003048    0.039624    …    0.039623    0.027653    0.034498
4     0.032040    0.012015    0.034043    …    0.038795    0.038973    0.035934
..         …           …           …    …        …           …           …
494   0.011004    0.009004    0.026010    …    0.041054    0.030902    0.053857
495   0.015748    0.041996    0.031497    …    0.019678   -0.010206    0.025994
496   0.008892    0.001368    0.030096    …    0.027777    0.036170    0.033842
498   0.004410    0.002205    0.048511    …    0.037378    0.037832    0.031182
499   0.020952    0.003223    0.045934    …    0.019198    0.027226    0.020961

            X35         X36         X37         X38         X39         X40          X41
0    -0.020882    0.039161    0.033871    0.035709    0.001867    0.014523   no_efectores
```

```
1    0.020852   0.015337   0.026114   0.062706   0.099725   0.079013   no_efectores
2    0.024316   0.032051   0.033608  -0.013453  -0.030408   0.017050   no_efectores
3    0.031823   0.025006   0.029805   0.012024   0.040712   0.031067   no_efectores
4    0.047520   0.014474   0.021851   0.028407   0.030057   0.010331   no_efectores
..      …          …          …          …          …          …
494  0.029107   0.022253   0.042225   0.032466   0.033835   0.035648   no_efectores
495  0.031999   0.017357   0.033078   0.041158   0.072084   0.072866   no_efectores
496  0.028177   0.042904   0.046757   0.030899   0.031200   0.033724   no_efectores
498  0.020477   0.011074   0.040424   0.029589   0.003363   0.025241   no_efectores
499  0.031643   0.018850   0.031422   0.041599   0.032492   0.031157   no_efectores

[418 rows x 42 columns]
```

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

```
              X0          X1          X2          X3          X4          X5  \
count  418.000000  418.000000  418.000000  418.000000  418.000000  418.000000
mean     0.044480    0.004560    0.037262    0.039208    0.015669    0.036363
std      0.015175    0.006496    0.023377    0.026438    0.010446    0.011046
min      0.004971    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034419    0.000000    0.019848    0.017922    0.008206    0.028919
50%      0.043383    0.002361    0.034996    0.035974    0.013419    0.035414
75%      0.053657    0.005471    0.049663    0.053242    0.022177    0.043191
max      0.104529    0.038901    0.127139    0.129223    0.063139    0.073333

              X6          X7          X8          X9  …         X31  \
count  418.000000  418.000000  418.000000  418.000000  …  418.000000
mean     0.009183    0.023101    0.014176    0.040785  …    0.023572
std      0.007793    0.017878    0.017240    0.019060  …    0.018581
min      0.000000    0.000000    0.000000    0.000000  …   -0.069962
25%      0.003727    0.009292    0.002768    0.026813  …    0.013347
50%      0.007429    0.018438    0.008281    0.038314  …    0.026392
75%      0.012480    0.032044    0.019037    0.052247  …    0.035850
max      0.043326    0.100279    0.102332    0.108329  …    0.076715

             X32         X33         X34         X35         X36         X37  \
count  418.000000  418.000000  418.000000  418.000000  418.000000  418.000000
mean     0.025765    0.025485    0.025890    0.026132    0.026522    0.025754
std      0.017905    0.017531    0.017153    0.018031    0.017638    0.017330
min     -0.047540   -0.052380   -0.038759   -0.058853   -0.024254   -0.037942
25%      0.015846    0.016988    0.016419    0.017110    0.016620    0.017017
50%      0.028362    0.027962    0.027302    0.028131    0.027023    0.027885
75%      0.036448    0.037686    0.035657    0.036173    0.036511    0.036879
max      0.091409    0.071716    0.083702    0.101549    0.098414    0.086132
```
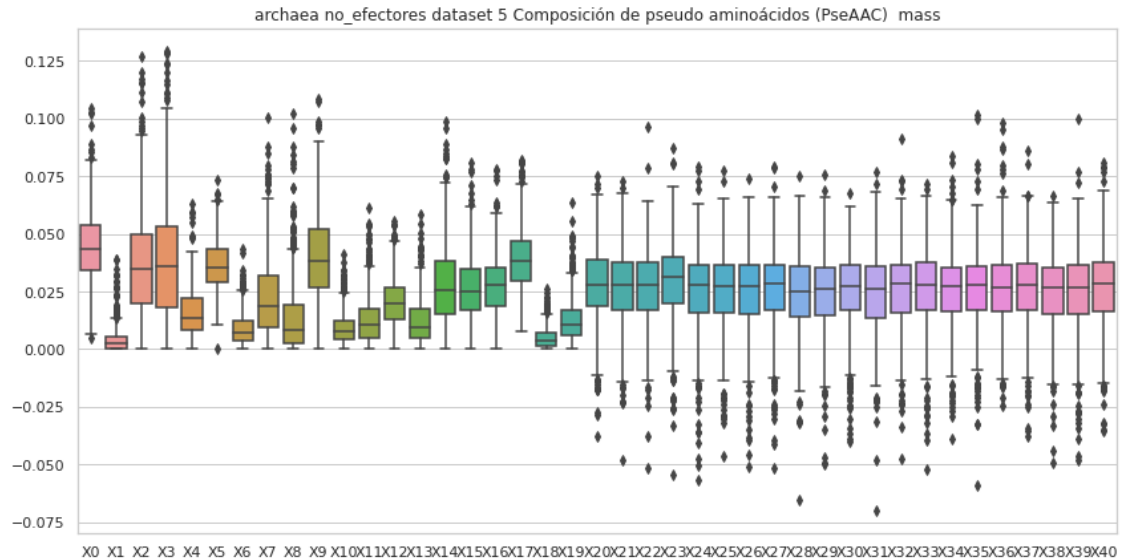
```
             X38         X39         X40
count  418.000000  418.000000  418.000000
mean     0.024181    0.024891    0.026665
std      0.017764    0.019004    0.018043
min     -0.048987   -0.048216   -0.035600
25%      0.015142    0.015426    0.016362
50%      0.026994    0.026557    0.028694
75%      0.035666    0.036287    0.037471
max      0.066569    0.099725    0.081123

[8 rows x 41 columns]
```


archaea efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) mass

archaea no_efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) mass

5 Composición de pseudo aminoácidos (PseAAC) hidro

```python
[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 archaea dataset 5,
con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6   \
0    0.035374  0.003930  0.043235  0.055026  0.007861  0.023583  0.019652
1    0.039122  0.004890  0.078244  0.048902  0.014671  0.034232  0.019561
2    0.080708  0.005765  0.069178  0.074943  0.034589  0.086473  0.011530
3    0.048857  0.000000  0.063041  0.083529  0.031520  0.044129  0.012608
4    0.048983  0.002721  0.021770  0.046261  0.010885  0.027213  0.010885
..        …         …         …         …         …         …
495  0.039810  0.023886  0.039810  0.071658  0.007962  0.031848  0.000000
496  0.045115  0.000000  0.052055  0.039909  0.006941  0.048585  0.003470
497  0.091832  0.010804  0.037813  0.021608  0.054019  0.037813  0.027010
498  0.021924  0.000000  0.012332  0.013703  0.020554  0.027405  0.002741
499  0.053217  0.000000  0.038250  0.054880  0.018293  0.056544  0.026609

           X7        X8        X9   …       X53       X54       X55   \
0    0.003930  0.007861  0.055026  … -0.008047  0.028157  0.053292
1    0.014671  0.009780  0.068463  … -0.005664 -0.016669  0.027893
2    0.028824  0.005765  0.046119  …  0.020137  0.024379  0.035198
3    0.129234  0.104018  0.152874  … -0.001516 -0.032236 -0.011817
4    0.010885  0.019049  0.038098  …  0.058136  0.022805  0.022768
..        …         …         …   … …       …         …         …
495  0.039810  0.055734  0.095545  …  0.027838  0.046491  0.071762
496  0.012146  0.003470  0.031233  … -0.016630  0.010131  0.026529
497  0.037813  0.010804  0.118842  … -0.018268 -0.038597 -0.061328
498  0.027405  0.002741  0.038367  …  0.004165  0.008756 -0.015071
499  0.064859  0.063196  0.051554  …  0.025399  0.002002  0.006436

          X56       X57       X58       X59       X60       X61        X62
0    0.033408  0.044722 -0.003016 -0.020466  0.075692  0.073601   efectores
1   -0.016895  0.034903  0.035771  0.030136  0.048186  0.071167   efectores
2   -0.025428 -0.026975 -0.075861  0.032401  0.066171  0.014149   efectores
3    0.010110  0.007648 -0.000538 -0.037871  0.030185 -0.011167   efectores
4    0.001584  0.022602  0.015820  0.015794  0.000487  0.031097   efectores
..        …         …         …         …         …         …
495 -0.058097  0.005449 -0.019201 -0.083002  0.067213  0.107834   efectores
496  0.018749  0.008178  0.006188  0.014914  0.016242  0.015540   efectores
497 -0.036999 -0.014800 -0.005373 -0.020369  0.048220  0.025538   efectores
```

```
498   0.040610   0.004366   0.029780   0.001325   0.033689   0.006250   efectores
499   0.019481   0.026406  -0.001804   0.006512   0.020029   0.031788   efectores


[500 rows x 63 columns]


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

	X0	X1	X2	X3	X4	X5
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.048284	0.004775	0.037303	0.044986	0.021576	0.040522
std	0.027504	0.007786	0.026310	0.032779	0.017344	0.021764
min	0.000000	0.000000	0.002058	0.000000	0.000000	0.002523
25%	0.028786	0.000000	0.013972	0.012524	0.010726	0.025770
50%	0.043456	0.001154	0.030290	0.045313	0.017640	0.035091
75%	0.059902	0.006455	0.056136	0.066073	0.027286	0.049512
max	0.209588	0.053143	0.138305	0.157191	0.157191	0.176305

	X6	X7	X8	X9	…	X52
count	500.000000	500.000000	500.000000	500.000000	…	500.000000
mean	0.010685	0.036175	0.031386	0.063934	…	0.007563
std	0.009873	0.032486	0.035385	0.033513	…	0.038814
min	0.000000	0.000000	0.000000	0.007326	…	-0.291232
25%	0.003620	0.012547	0.003981	0.040204	…	-0.006980
50%	0.008800	0.024641	0.014702	0.055057	…	0.013819
75%	0.014622	0.052420	0.051964	0.077579	…	0.026199
max	0.104794	0.189336	0.204483	0.224752	…	0.200114

	X53	X54	X55	X56	X57	X58
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.011386	0.012757	0.015459	0.007570	0.012326	0.008571
std	0.027675	0.038686	0.034051	0.039158	0.033361	0.040201
min	-0.196859	-0.172529	-0.129306	-0.204040	-0.193335	-0.266443
25%	-0.001384	-0.000762	0.000188	-0.010428	-0.002073	-0.007547
50%	0.008878	0.017307	0.010248	0.012513	0.007690	0.014268
75%	0.023986	0.028731	0.028146	0.026029	0.024549	0.027492
max	0.122647	0.364938	0.289456	0.302645	0.208604	0.163220

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.011622	0.007873	0.012925
std	0.031687	0.040376	0.032106
min	-0.131282	-0.260593	-0.149296
25%	-0.000886	-0.004607	-0.001711
50%	0.009599	0.015495	0.008981
75%	0.023602	0.028306	0.026827

```
max      0.189142    0.283155    0.180979

[8 rows x 62 columns]



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

            X0          X1          X2          X3          X4          X5          X6    \
0      0.054688    0.000000    0.030762    0.061523    0.017090    0.034180    0.017090
1      0.024051    0.000000    0.000000    0.012026    0.018039    0.048103    0.006013
2      0.081473    0.012534    0.018802    0.094008    0.037603    0.062672    0.037603
3      0.060476    0.000000    0.047201    0.050151    0.007375    0.030976    0.010325
4      0.043371    0.000000    0.040273    0.018587    0.037175    0.049567    0.015490
..          …           …           …           …           …           …
495    0.012922    0.000000    0.019382    0.029074    0.009691    0.032304    0.009691
496    0.069883    0.000000    0.009983    0.009983    0.014420    0.038824    0.002218
497    0.032194    0.032194    0.032194    0.048292    0.000000    0.064389    0.016097
498    0.027608    0.000000    0.005522    0.003681    0.012884    0.034970    0.003681
499    0.027673    0.000000    0.008071    0.014989    0.013260    0.026520    0.004612

            X7          X8          X9    …       X53         X54         X55    \
0      0.030762    0.003418    0.037598    … -0.018116    0.036310    0.017863
1      0.036077    0.000000    0.054116    …  0.010452    0.012281    0.008471
2      0.087741    0.094008    0.119076    …  0.084055    0.122730    0.040880
3      0.007375    0.002950    0.038351    …  0.050795    0.020742    0.036178
4      0.049567    0.018587    0.052665    …  0.009193    0.000004 -0.005006
..          …           …           …   …       …           …           …
495    0.009691    0.025843    0.019382    …  0.038910 -0.020723    0.013214
496    0.014420    0.002218    0.048807    … -0.017378    0.021845    0.001034
497    0.080486    0.128778    0.144875    …  0.007845 -0.164257 -0.145700
498    0.003681    0.001841    0.040492    …  0.002275    0.018484    0.017380
499    0.014989    0.002306    0.032861    …  0.006385    0.024840 -0.000655

            X56         X57         X58         X59         X60         X61              X62
0     -0.002944    0.018008 -0.038422 -0.003059 -0.025216    0.016707    no_efectores
1      0.047753    0.003766    0.077156    0.024743    0.093333    0.049951    no_efectores
2      0.042052    0.043655 -0.081174 -0.031466 -0.054450 -0.037691    no_efectores
3     -0.013067    0.018318    0.039017    0.050300 -0.006009    0.011244    no_efectores
4      0.029209    0.040418    0.037672    0.008329    0.020642 -0.007011    no_efectores
..          …           …           …           …           …           …
495   -0.016116    0.023350 -0.007623    0.016456    0.013010    0.022560    no_efectores
496    0.028954    0.006134    0.007126 -0.006694    0.023699 -0.002590    no_efectores
497    0.135142    0.021970 -0.119471 -0.018397 -0.086570 -0.030133    no_efectores
498    0.013337    0.002665    0.020833    0.004500    0.027447    0.009701    no_efectores
```

40

```
499  0.023098  0.001819  0.017925 -0.004459  0.024738  0.005969  no_efectores
```

[500 rows x 63 columns]

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

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.058422	0.007587	0.048869	0.051131	0.021907	0.050381
std	0.032009	0.012681	0.028236	0.030149	0.018071	0.029077
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.036514	0.000000	0.027532	0.029420	0.011291	0.030960
50%	0.053904	0.003530	0.050026	0.051554	0.018035	0.045939
75%	0.076063	0.008816	0.067575	0.068090	0.027573	0.063287
max	0.302012	0.128611	0.141722	0.178052	0.178052	0.245170

	X6	X7	X8	X9	…	X52 \
count	500.000000	500.000000	500.000000	500.000000	…	500.000000
mean	0.012866	0.033974	0.022073	0.053955	…	0.004340
std	0.012370	0.031957	0.028621	0.028807	…	0.041334
min	0.000000	0.000000	0.000000	0.000000	…	-0.420712
25%	0.004328	0.012726	0.004175	0.034933	…	-0.011003
50%	0.010367	0.024564	0.012219	0.049432	…	0.006917
75%	0.017259	0.046229	0.026566	0.066185	…	0.022772
max	0.117821	0.242232	0.214298	0.204859	…	0.245445

	X53	X54	X55	X56	X57	X58 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.014966	0.004880	0.014083	0.003648	0.011411	0.004450
std	0.032264	0.041697	0.037374	0.037673	0.031900	0.043175
min	-0.197574	-0.391288	-0.237573	-0.371713	-0.186889	-0.458020
25%	0.000599	-0.009065	-0.002249	-0.012041	-0.001592	-0.010843
50%	0.013026	0.007975	0.012670	0.008325	0.010856	0.007549
75%	0.030544	0.024319	0.030326	0.023842	0.027703	0.023957
max	0.247457	0.238814	0.193312	0.147244	0.187113	0.321672

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.010703	0.004653	0.013000
std	0.038651	0.046169	0.039029
min	-0.198282	-0.613662	-0.452611
25%	-0.004879	-0.010822	-0.001859
50%	0.009199	0.007212	0.012318
75%	0.028047	0.023358	0.031054
max	0.292988	0.171388	0.155284

```
[8 rows x 62 columns]
```

archaea efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro con valores atípicos.

archaea no_efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro con valores atípicos.
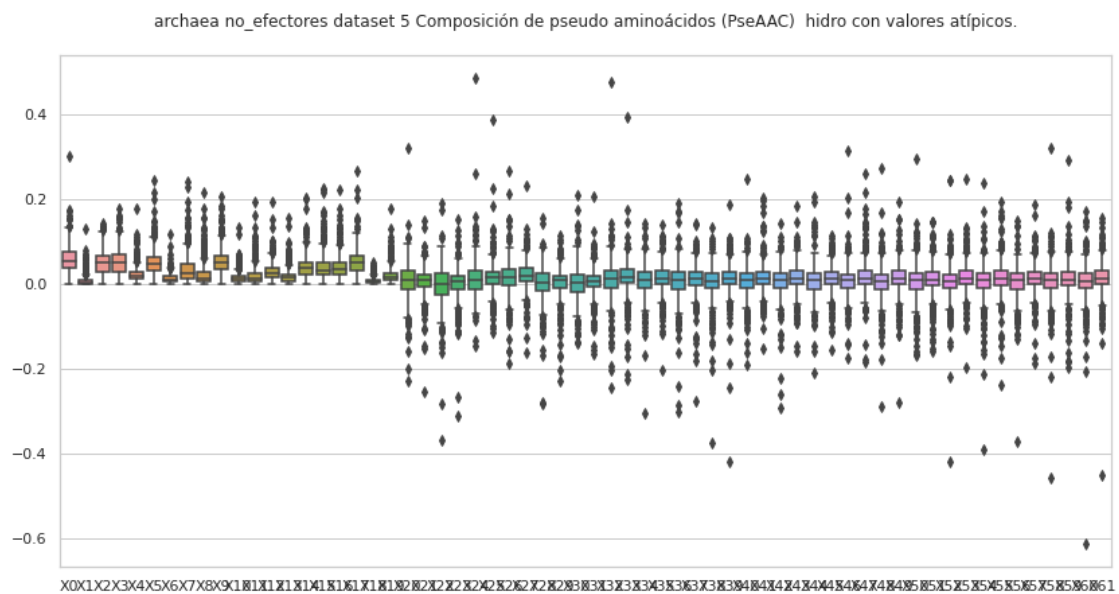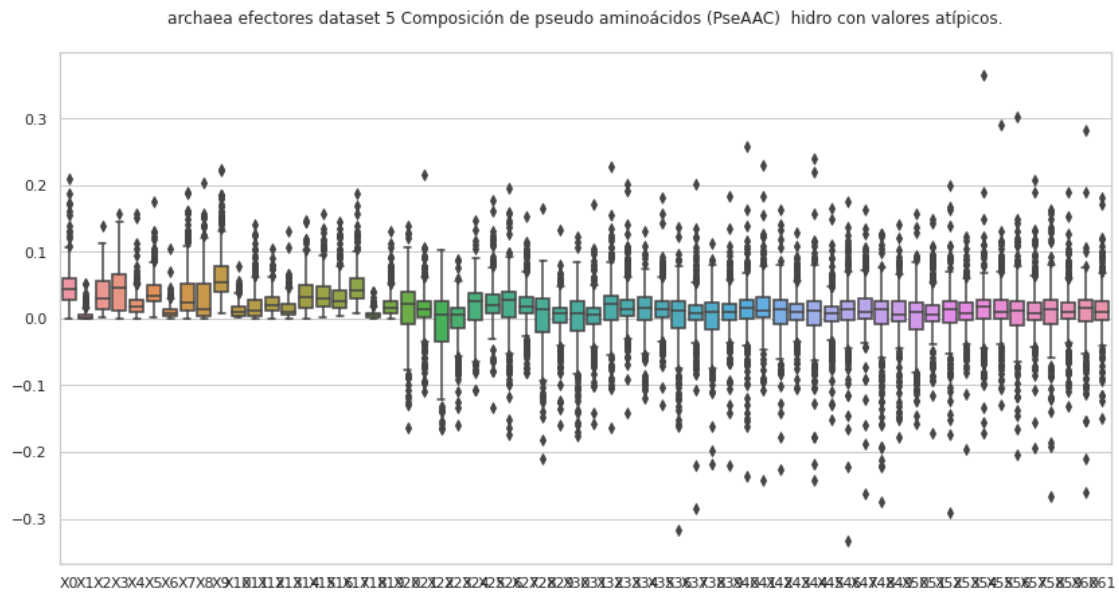
5.1 Composición de pseudo aminoácidos (PseAAC) hidro, sin valores atípicos

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.035374  0.003930  0.043235  0.055026  0.007861  0.023583  0.019652
1    0.039122  0.004890  0.078244  0.048902  0.014671  0.034232  0.019561
2    0.080708  0.005765  0.069178  0.074943  0.034589  0.086473  0.011530
3    0.048857  0.000000  0.063041  0.083529  0.031520  0.044129  0.012608
4    0.048983  0.002721  0.021770  0.046261  0.010885  0.027213  0.010885
..        …         …         …         …         …         …         …
493  0.035410  0.000000  0.043279  0.027541  0.007869  0.039345  0.001967
494  0.018353  0.000000  0.020800  0.024470  0.028141  0.024470  0.002447
496  0.045115  0.000000  0.052055  0.039909  0.006941  0.048585  0.003470
498  0.021924  0.000000  0.012332  0.013703  0.020554  0.027405  0.002741
499  0.053217  0.000000  0.038250  0.054880  0.018293  0.056544  0.026609

           X7        X8        X9  …       X53       X54       X55  \
0    0.003930  0.007861  0.055026  … -0.008047  0.028157  0.053292
1    0.014671  0.009780  0.068463  … -0.005664 -0.016669  0.027893
2    0.028824  0.005765  0.046119  …  0.020137  0.024379  0.035198
3    0.129234  0.104018  0.152874  … -0.001516 -0.032236 -0.011817
4    0.010885  0.019049  0.038098  …  0.058136  0.022805  0.022768
..        …         …         …   …        …         …         …
493  0.005902  0.005902  0.043279  …  0.038329  0.006332  0.038735
494  0.067293  0.025694  0.047717  … -0.011558  0.020376 -0.004925
496  0.012146  0.003470  0.031233  … -0.016630  0.010131  0.026529
498  0.027405  0.002741  0.038367  …  0.004165  0.008756 -0.015071
499  0.064859  0.063196  0.051554  …  0.025399  0.002002  0.006436

           X56       X57       X58       X59       X60       X61        X62
0    0.033408  0.044722 -0.003016 -0.020466  0.075692  0.073601  efectores
1   -0.016895  0.034903  0.035771  0.030136  0.048186  0.071167  efectores
2   -0.025428 -0.026975 -0.075861  0.032401  0.066171  0.014149  efectores
3    0.010110  0.007648 -0.000538 -0.037871  0.030185 -0.011167  efectores
4    0.001584  0.022602  0.015820  0.015794  0.000487  0.031097  efectores
..        …         …         …         …         …         …
493 -0.023303 -0.001206  0.008373  0.018273  0.011245  0.027292  efectores
494  0.005424 -0.007186  0.049925  0.016349  0.023706 -0.006442  efectores
496  0.018749  0.008178  0.006188  0.014914  0.016242  0.015540  efectores
498  0.040610  0.004366  0.029780  0.001325  0.033689  0.006250  efectores
499  0.019481  0.026406 -0.001804  0.006512  0.020029  0.031788  efectores

[395 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	395.000000	395.000000	395.000000	395.000000	395.000000	395.000000	
mean	0.042791	0.003152	0.030605	0.035491	0.018246	0.035791	
std	0.020153	0.005074	0.021689	0.026437	0.011423	0.016839	
min	0.005081	0.000000	0.002058	0.000000	0.000000	0.002523	
25%	0.027172	0.000000	0.011996	0.009509	0.010097	0.024451	
50%	0.040495	0.000775	0.022892	0.032390	0.016204	0.031815	
75%	0.053227	0.004142	0.048289	0.057213	0.024016	0.043570	
max	0.109454	0.027661	0.092188	0.100559	0.072704	0.102817	

	X6	X7	X8	X9	…	X52	\
count	395.000000	395.000000	395.000000	395.000000	…	395.000000	
mean	0.008608	0.028101	0.022376	0.052241	…	0.009833	
std	0.006636	0.022784	0.026081	0.019900	…	0.024343	
min	0.000000	0.000000	0.000000	0.007326	…	-0.084921	
25%	0.003072	0.011311	0.002914	0.037899	…	-0.001066	
50%	0.007722	0.020038	0.010548	0.048007	…	0.015386	
75%	0.012537	0.039714	0.036657	0.064561	…	0.025821	
max	0.036603	0.129234	0.106633	0.152874	…	0.085825	

	X53	X54	X55	X56	X57	X58	\
count	395.000000	395.000000	395.000000	395.000000	395.000000	395.000000	
mean	0.012132	0.015650	0.015574	0.009884	0.011857	0.011820	
std	0.019284	0.023632	0.022811	0.022714	0.021015	0.025242	
min	-0.062952	-0.097921	-0.056082	-0.071465	-0.044097	-0.105503	
25%	0.000425	0.006216	0.002020	-0.006238	-0.000804	-0.002542	
50%	0.008793	0.018424	0.010158	0.013953	0.007601	0.015689	
75%	0.021194	0.028412	0.025059	0.025738	0.020841	0.026923	
max	0.089089	0.111519	0.110814	0.087203	0.094306	0.092611	

	X59	X60	X61
count	395.000000	395.000000	395.000000
mean	0.011460	0.012812	0.012408
std	0.020654	0.024850	0.021192
min	-0.058079	-0.103343	-0.081479
25%	0.000274	0.000067	-0.000875
50%	0.009029	0.017469	0.008367
75%	0.020244	0.027725	0.023731
max	0.099019	0.079131	0.097772

[8 rows x 62 columns]

no_efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6
0	0.054688	0.000000	0.030762	0.061523	0.017090	0.034180	0.017090
1	0.024051	0.000000	0.000000	0.012026	0.018039	0.048103	0.006013
3	0.060476	0.000000	0.047201	0.050151	0.007375	0.030976	0.010325
4	0.043371	0.000000	0.040273	0.018587	0.037175	0.049567	0.015490
5	0.089123	0.008695	0.073907	0.073907	0.028258	0.065212	0.019564
..	…	…	…	…	…	…	
494	0.075921	0.004218	0.065376	0.061159	0.018980	0.078030	0.008436
495	0.012922	0.000000	0.019382	0.029074	0.009691	0.032304	0.009691
496	0.069883	0.000000	0.009983	0.009983	0.014420	0.038824	0.002218
498	0.027608	0.000000	0.005522	0.003681	0.012884	0.034970	0.003681
499	0.027673	0.000000	0.008071	0.014989	0.013260	0.026520	0.004612

	X7	X8	X9	…	X53	X54	X55
0	0.030762	0.003418	0.037598	…	-0.018116	0.036310	0.017863
1	0.036077	0.000000	0.054116	…	0.010452	0.012281	0.008471
3	0.007375	0.002950	0.038351	…	0.050795	0.020742	0.036178
4	0.049567	0.018587	0.052665	…	0.009193	0.000004	-0.005006
5	0.015216	0.019564	0.052169	…	0.027832	-0.036310	-0.014271
..	…	…	… …	…	…	…	
494	0.023198	0.018980	0.054832	…	-0.009181	0.022844	0.019234
495	0.009691	0.025843	0.019382	…	0.038910	-0.020723	0.013214
496	0.014420	0.002218	0.048807	…	-0.017378	0.021845	0.001034
498	0.003681	0.001841	0.040492	…	0.002275	0.018484	0.017380
499	0.014989	0.002306	0.032861	…	0.006385	0.024840	-0.000655

	X56	X57	X58	X59	X60	X61	X62
0	-0.002944	0.018008	-0.038422	-0.003059	-0.025216	0.016707	no_efectores
1	0.047753	0.003766	0.077156	0.024743	0.093333	0.049951	no_efectores
3	-0.013067	0.018318	0.039017	0.050300	-0.006009	0.011244	no_efectores
4	0.029209	0.040418	0.037672	0.008329	0.020642	-0.007011	no_efectores
5	0.013070	0.019618	-0.006460	-0.000256	-0.014982	0.043135	no_efectores
..	…	…	…	…	…	…	
494	-0.019298	0.008172	0.004814	0.009007	0.008163	0.011862	no_efectores
495	-0.016116	0.023350	-0.007623	0.016456	0.013010	0.022560	no_efectores
496	0.028954	0.006134	0.007126	-0.006694	0.023699	-0.002590	no_efectores
498	0.013337	0.002665	0.020833	0.004500	0.027447	0.009701	no_efectores
499	0.023098	0.001819	0.017925	-0.004459	0.024738	0.005969	no_efectores

[418 rows x 63 columns]

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

```
             X0          X1          X2          X3          X4          X5   \
count  418.000000  418.000000  418.000000  418.000000  418.000000  418.000000
mean     0.055852    0.005560    0.045682    0.045928    0.018442    0.045698
std      0.027340    0.007416    0.026099    0.025088    0.010561    0.021409
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.002315
25%      0.036328    0.000000    0.023557    0.024920    0.010720    0.030121
50%      0.052728    0.003105    0.047021    0.048107    0.016897    0.043447
75%      0.073121    0.007885    0.065145    0.064222    0.024469    0.059455
max      0.136488    0.040247    0.124807    0.121337    0.058508    0.123009

             X6          X7          X8          X9    …         X52   \
count  418.000000  418.000000  418.000000  418.000000  …   418.000000
mean     0.011291    0.027343    0.016313    0.048779  …      0.006776
std      0.009164    0.020998    0.018438    0.021444  …      0.024080
min      0.000000    0.000000    0.000000    0.000000  …     -0.077834
25%      0.004189    0.011964    0.003958    0.033901  …     -0.006390
50%      0.009465    0.023058    0.010629    0.047008  …      0.008010
75%      0.016278    0.036042    0.020934    0.062091  …      0.021656
max      0.043729    0.115437    0.099063    0.122116  …      0.083580

             X53         X54         X55         X56         X57         X58   \
count  418.000000  418.000000  418.000000  418.000000  418.000000  418.000000
mean     0.014701    0.006781    0.014613    0.007240    0.013091    0.006577
std      0.021439    0.025137    0.024499    0.025622    0.022342    0.026368
min     -0.053588   -0.093226   -0.065571   -0.085179   -0.056010   -0.075680
25%      0.001935   -0.007373   -0.000495   -0.007502   -0.000532   -0.008636
50%      0.012957    0.008316    0.012670    0.009769    0.011164    0.007974
75%      0.027566    0.022888    0.029329    0.023904    0.026668    0.022540
max      0.107936    0.105020    0.096499    0.065262    0.101706    0.093841

             X59         X60         X61
count  418.000000  418.000000  418.000000
mean     0.011680    0.008354    0.014138
std      0.023262    0.025857    0.023858
min     -0.060286   -0.083033   -0.093557
25%     -0.002314   -0.007782   -0.000549
50%      0.009338    0.008475    0.012040
75%      0.025601    0.023225    0.027854
max      0.089212    0.099464    0.097503


[8 rows x 62 columns]
```
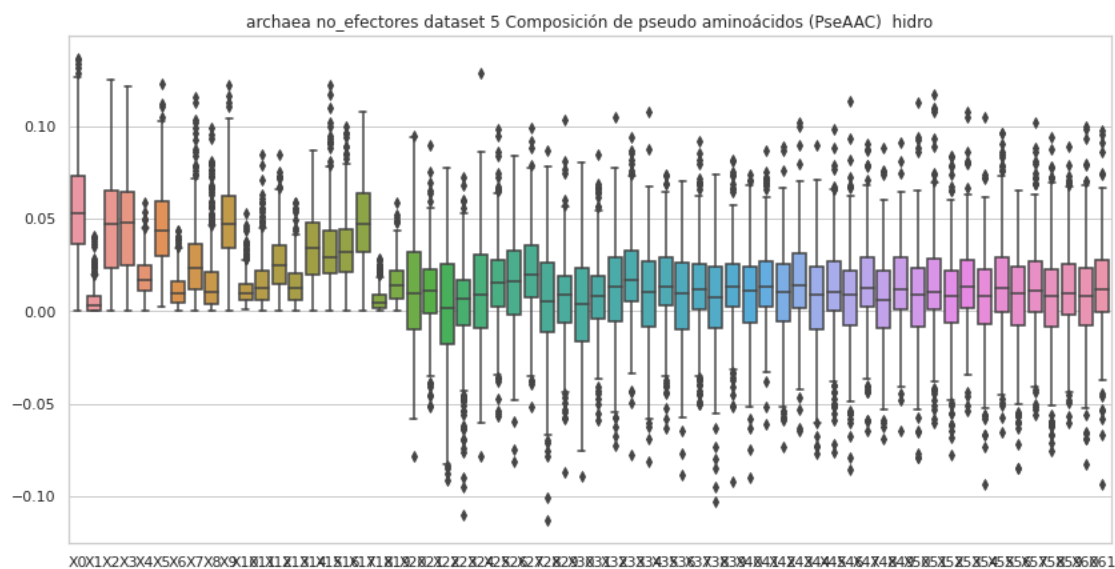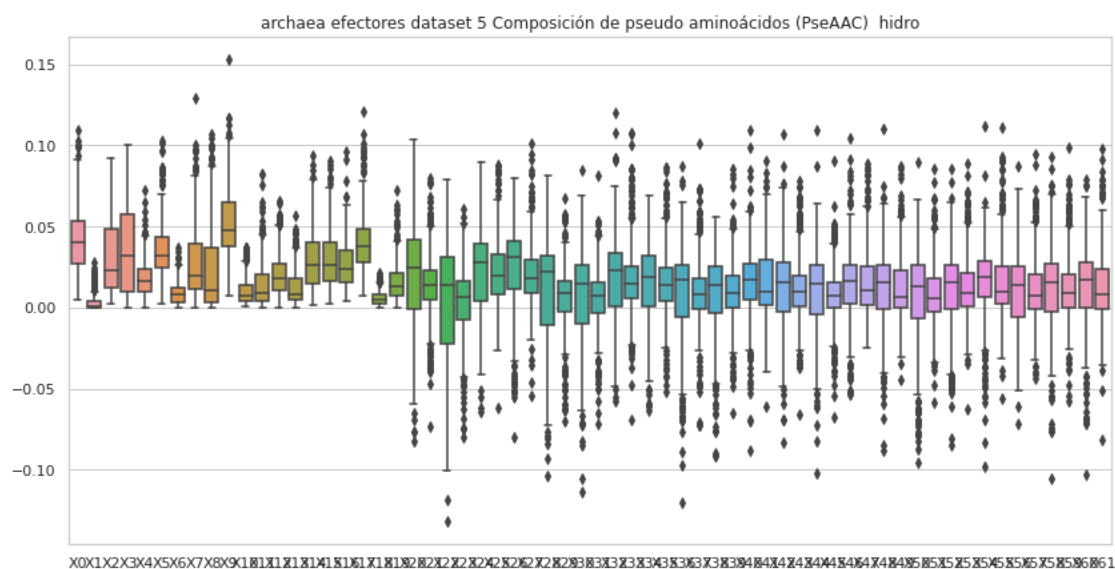
archaea efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro

archaea no_efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro

6 Covarianza de auto cruzamiento (ACC) hidro_mass

```
[11]: #hidro_mass
      transf = "Covarianza de auto cruzamiento (ACC) "
      transf2 = "ACC"
      estado = "con valores atípicos.\n"
```

```python
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 archaea dataset 5,
con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0     0.068499  0.107372 -0.178466 -0.105813 -0.104190 -0.079042 -0.023018
1     0.016127  0.057621 -0.090205 -0.135102 -0.101268 -0.026928 -0.027177
2    -0.064370  0.019903 -0.113168 -0.112980 -0.147626 -0.056355  0.037008
3    -0.003813  0.007852 -0.010343  0.051148 -0.010351 -0.014167 -0.020029
4    -0.013156 -0.009046 -0.169310 -0.000608 -0.071155 -0.107506 -0.017740

..        …         …         …         …         …         …         …
495  -0.063990  0.072973  0.051820  0.175759  0.069077  0.080243  0.115945
496   0.055211  0.003064 -0.052672  0.106847  0.064771  0.102201  0.009929
497  -0.081476  0.046156 -0.038680  0.093794 -0.014769 -0.091086 -0.026024
498  -0.028050  0.073232  0.096939 -0.019894  0.010355  0.114560 -0.000925
499  -0.061669  0.062469  0.010104  0.054898  0.062286 -0.013364  0.041625

           X7        X8        X9       X10       X11       X12       X13
0    -0.067292  0.034984  0.057295  0.135773  0.043677 -0.005830  efectores
```

```
1    0.024964 -0.056175 -0.035597  0.147024  0.061575  0.100781  efectores
2    0.040063 -0.002298  0.110706  0.004580  0.067847  0.040649  efectores
3    0.061273 -0.010972 -0.012717 -0.010953  0.021089  0.009181  efectores
4   -0.025899  0.188517 -0.147207  0.082376 -0.043421  0.107928  efectores
..       …         …         …         …         …         …
495  0.024596 -0.004279  0.096415 -0.059640  0.128476 -0.077162  efectores
496  0.104859  0.062971  0.102625  0.003787 -0.003867 -0.069779  efectores
497 -0.060197  0.059830  0.010803 -0.086266 -0.145322  0.051021  efectores
498 -0.048277 -0.006971  0.021771  0.010893  0.045474  0.087363  efectores
499 -0.044397  0.048455 -0.036730  0.011376  0.013997  0.026100  efectores

[500 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4           X5  \
count  500.000000   500.000000   500.000000   500.000000   500.000000   500.000000
mean     0.004686     0.019443     0.006366     0.012882    -0.003335     0.000429
std      0.068388     0.067242     0.068452     0.071165     0.069201     0.067680
min     -0.228927    -0.238098    -0.256714    -0.198361    -0.261174    -0.219974
25%     -0.038149    -0.018739    -0.037028    -0.029823    -0.047300    -0.040613
50%      0.006632     0.018324    -0.000204     0.019779    -0.003333     0.002533
75%      0.047288     0.060533     0.049812     0.061666     0.045616     0.044477
max      0.206642     0.236765     0.238997     0.175759     0.285489     0.218508

                X6           X7           X8           X9          X10          X11  \
count  500.000000   500.000000   500.000000   500.000000   500.000000   500.000000
mean     0.014125     0.001228    -0.001182     0.003593     0.002251    -0.010216
std      0.068203     0.064911     0.073297     0.070919     0.067539     0.072233
min     -0.212350    -0.266405    -0.261605    -0.294274    -0.254213    -0.257899
25%     -0.026883    -0.034410    -0.050175    -0.035483    -0.039186    -0.050256
50%      0.009027    -0.000754     0.004214     0.003678     0.001396    -0.006946
75%      0.060594     0.046850     0.049137     0.049204     0.047463     0.037677
max      0.218983     0.238644     0.267867     0.198706     0.234874     0.241133

               X12
count  500.000000
mean     0.010351
std      0.071877
min     -0.247439
25%     -0.034986
50%      0.011396
75%      0.058693
max      0.237507
```

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

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.048139	0.023891	-0.115901	0.057774	0.032071	0.063435	-0.184426
1	0.143509	0.082342	-0.131607	-0.059306	-0.041152	-0.013132	0.189856
2	0.120228	-0.038323	-0.017979	0.019364	-0.015202	0.066651	0.075608
3	-0.019050	0.098471	-0.011074	-0.048137	-0.058446	-0.109650	0.007446
4	0.008016	-0.059660	-0.092342	-0.050158	0.103892	0.017644	0.001662
..	…	…	…	…	…	…	…
495	0.042510	0.038375	0.025392	0.100591	-0.015054	-0.192800	0.010897
496	0.086749	0.038354	-0.072797	-0.003569	-0.052947	0.051965	0.085290
497	-0.112976	0.071173	-0.098027	-0.047056	-0.059595	0.125019	-0.072861
498	0.214405	-0.079456	0.031113	0.169249	0.023345	-0.093321	-0.082705
499	-0.051489	0.010894	0.033417	0.034248	0.035250	-0.025396	0.023340

	X7	X8	X9	X10	X11	X12	X13
0	0.053476	-0.006586	0.049907	-0.019029	0.049160	0.028856	no_efectores
1	-0.085710	-0.049765	-0.286697	-0.160311	-0.190254	0.146958	no_efectores
2	0.021937	-0.104300	0.026756	0.137558	-0.035878	-0.022094	no_efectores
3	-0.071513	0.085593	-0.084771	0.001188	0.013352	0.057126	no_efectores
4	0.024062	0.011749	0.056943	-0.104491	-0.107386	0.042677	no_efectores
..	…	…	…	…	…	…	
495	0.055095	-0.089681	-0.079964	0.057654	-0.078424	0.014737	no_efectores
496	0.053648	-0.060742	0.020493	-0.052042	0.023949	-0.104392	no_efectores
497	0.133627	0.007436	0.053858	-0.056722	-0.011216	-0.015837	no_efectores
498	0.226316	-0.064895	0.015656	0.201966	0.059112	0.053016	no_efectores
499	0.002935	-0.020942	0.034431	0.016460	0.042444	-0.070771	no_efectores

[500 rows x 14 columns]

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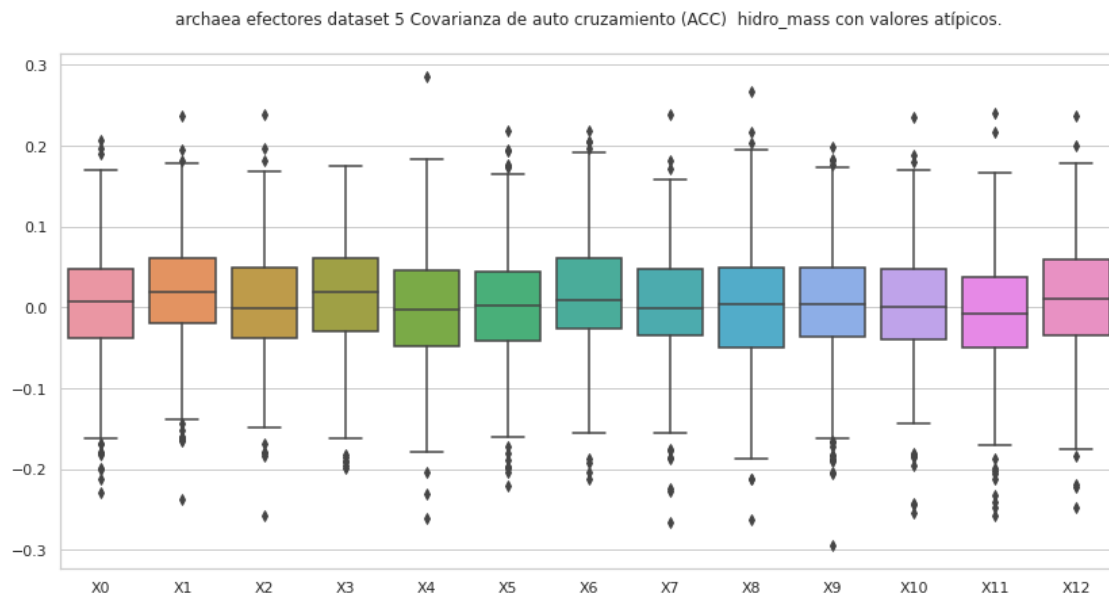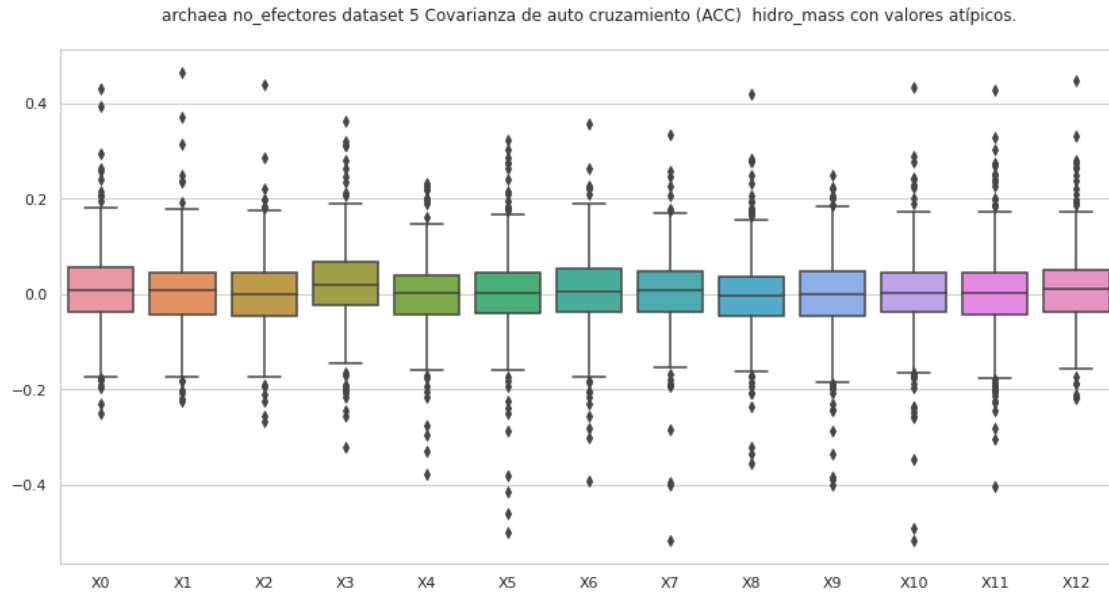con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.012109	0.003867	0.002089	0.021783	-0.001476	0.000593
std	0.078780	0.078712	0.075889	0.081485	0.074803	0.086175
min	-0.248988	-0.225141	-0.268128	-0.320227	-0.378918	-0.501157
25%	-0.036597	-0.043747	-0.045124	-0.021818	-0.042399	-0.039860
50%	0.009497	0.008305	-0.000183	0.019198	0.002477	0.002454
75%	0.055105	0.046172	0.045564	0.067872	0.038706	0.044546

```
max         0.430307        0.465249        0.441076        0.362052        0.233730        0.324680


                   X6              X7              X8              X9             X10             X11    \
count     500.000000      500.000000      500.000000      500.000000      500.000000      500.000000
mean        0.005788        0.005003       -0.002200       -0.001916        0.000901        0.000368
std         0.083560        0.080759        0.081018        0.082513        0.086986        0.087428
min        -0.393276       -0.517270       -0.355367       -0.399693       -0.517175       -0.403017
25%        -0.038643       -0.036374       -0.046123       -0.044954       -0.038545       -0.043917
50%         0.004078        0.007814       -0.004344       -0.000443        0.002162        0.001473
75%         0.054093        0.047790        0.036504        0.047868        0.046292        0.045048
max         0.356660        0.333427        0.418887        0.248618        0.433093        0.427972


                  X12
count     500.000000
mean        0.012156
std         0.077348
min        -0.218254
25%        -0.037047
50%         0.009869
75%         0.049863
max         0.447849
```


archaea efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.

archaea no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.

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 archaea dataset 5,
sin valores atípicos.
Valores del documento csv.


           X0        X1        X2        X3        X4        X5        X6  \
0     0.068499  0.107372 -0.178466 -0.105813 -0.104190 -0.079042 -0.023018
1     0.016127  0.057621 -0.090205 -0.135102 -0.101268 -0.026928 -0.027177
2    -0.064370  0.019903 -0.113168 -0.112980 -0.147626 -0.056355  0.037008
3    -0.003813  0.007852 -0.010343  0.051148 -0.010351 -0.014167 -0.020029
4    -0.013156 -0.009046 -0.169310 -0.000608 -0.071155 -0.107506 -0.017740
..         …         …         …         …         …         …         …
495  -0.063990  0.072973  0.051820  0.175759  0.069077  0.080243  0.115945
496   0.055211  0.003064 -0.052672  0.106847  0.064771  0.102201  0.009929
497  -0.081476  0.046156 -0.038680  0.093794 -0.014769 -0.091086 -0.026024
498  -0.028050  0.073232  0.096939 -0.019894  0.010355  0.114560 -0.000925
499  -0.061669  0.062469  0.010104  0.054898  0.062286 -0.013364  0.041625


           X7        X8        X9       X10       X11       X12       X13
0    -0.067292  0.034984  0.057295  0.135773  0.043677 -0.005830  efectores
1     0.024964 -0.056175 -0.035597  0.147024  0.061575  0.100781  efectores
2     0.040063 -0.002298  0.110706  0.004580  0.067847  0.040649  efectores
3     0.061273 -0.010972 -0.012717 -0.010953  0.021089  0.009181  efectores
4    -0.025899  0.188517 -0.147207  0.082376 -0.043421  0.107928  efectores
..         …         …         …         …         …         …         …
495   0.024596 -0.004279  0.096415 -0.059640  0.128476 -0.077162  efectores
```

```
496  0.104859   0.062971   0.102625   0.003787 -0.003867 -0.069779   efectores
497 -0.060197   0.059830   0.010803 -0.086266 -0.145322  0.051021   efectores
498 -0.048277 -0.006971   0.021771   0.010893  0.045474  0.087363   efectores
499 -0.044397   0.048455 -0.036730   0.011376  0.013997  0.026100   efectores

[466 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  466.000000  466.000000  466.000000  466.000000  466.000000  466.000000
mean     0.005248    0.021037    0.008234    0.013402   -0.004776    0.001811
std      0.063471    0.063015    0.066515    0.070671    0.065079    0.062894
min     -0.198197   -0.162904   -0.178466   -0.196046   -0.204237   -0.199454
25%     -0.036146   -0.016249   -0.034262   -0.028269   -0.047194   -0.038447
50%      0.006505    0.018555    0.000865    0.022314   -0.004390    0.002571
75%      0.044801    0.060682    0.050766    0.061243    0.040586    0.041345
max      0.206642    0.194479    0.196020    0.175759    0.183584    0.193606

                X6          X7          X8          X9         X10         X11  \
count  466.000000  466.000000  466.000000  466.000000  466.000000  466.000000
mean     0.015261    0.002141   -0.000992    0.004903    0.003699   -0.008811
std      0.064530    0.060909    0.071050    0.066901    0.062451    0.067147
min     -0.186369   -0.186346   -0.212612   -0.204897   -0.194766   -0.211410
25%     -0.024987   -0.033274   -0.048490   -0.033241   -0.037731   -0.046755
50%      0.010528   -0.000152    0.004214    0.003923    0.001396   -0.006582
75%      0.060505    0.047064    0.048784    0.049116    0.046760    0.037473
max      0.205978    0.181051    0.216782    0.198706    0.189000    0.166311

                X12
count  466.000000
mean     0.012900
std      0.067259
min     -0.174776
25%     -0.031809
50%      0.012721
75%      0.058772
max      0.199475



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

```
             X0        X1        X2        X3        X4        X5        X6   \
0     -0.048139  0.023891 -0.115901  0.057774  0.032071  0.063435 -0.184426
2      0.120228 -0.038323 -0.017979  0.019364 -0.015202  0.066651  0.075608
3     -0.019050  0.098471 -0.011074 -0.048137 -0.058446 -0.109650  0.007446
4      0.008016 -0.059660 -0.092342 -0.050158  0.103892  0.017644  0.001662
5      0.016298 -0.025893 -0.033229 -0.035516  0.014706 -0.027842 -0.055578
..          ...       ...       ...       ...       ...       ...
495    0.042510  0.038375  0.025392  0.100591 -0.015054 -0.192800  0.010897
496    0.086749  0.038354 -0.072797 -0.003569 -0.052947  0.051965  0.085290
497   -0.112976  0.071173 -0.098027 -0.047056 -0.059595  0.125019 -0.072861
498    0.214405 -0.079456  0.031113  0.169249  0.023345 -0.093321 -0.082705
499   -0.051489  0.010894  0.033417  0.034248  0.035250 -0.025396  0.023340

             X7        X8        X9       X10       X11       X12          X13
0      0.053476 -0.006586  0.049907 -0.019029  0.049160  0.028856  no_efectores
2      0.021937 -0.104300  0.026756  0.137558 -0.035878 -0.022094  no_efectores
3     -0.071513  0.085593 -0.084771  0.001188  0.013352  0.057126  no_efectores
4      0.024062  0.011749  0.056943 -0.104491 -0.107386  0.042677  no_efectores
5     -0.006026  0.035460  0.008538 -0.086719 -0.014035 -0.058483  no_efectores
..          ...       ...       ...       ...       ...       ...
495    0.055095 -0.089681 -0.079964  0.057654 -0.078424  0.014737  no_efectores
496    0.053648 -0.060742  0.020493 -0.052042  0.023949 -0.104392  no_efectores
497    0.133627  0.007436  0.053858 -0.056722 -0.011216 -0.015837  no_efectores
498    0.226316 -0.064895  0.015656  0.201966  0.059112  0.053016  no_efectores
499    0.002935 -0.020942  0.034431  0.016460  0.042444 -0.070771  no_efectores

[459 rows x 14 columns]
```

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

```
               X0          X1          X2          X3          X4          X5   \
count  459.000000  459.000000  459.000000  459.000000  459.000000  459.000000
mean     0.009558    0.003412    0.002798    0.018100   -0.001171    0.000444
std      0.066760    0.066325    0.065009    0.069054    0.064208    0.064575
min     -0.180572   -0.200821   -0.209648   -0.215713   -0.216231   -0.239224
25%     -0.035250   -0.039789   -0.040816   -0.021809   -0.041233   -0.037349
50%      0.008974    0.009393    0.002399    0.017930    0.002317    0.001840
75%      0.050898    0.043463    0.045114    0.059468    0.036563    0.040196
max      0.242271    0.237388    0.181767    0.212763    0.202340    0.240555

               X6          X7          X8          X9         X10         X11   \
count  459.000000  459.000000  459.000000  459.000000  459.000000  459.000000
mean     0.006616    0.006808   -0.001949   -0.000065    0.001553   -0.001439
std      0.070959    0.066106    0.064862    0.069486    0.067825    0.067652
```
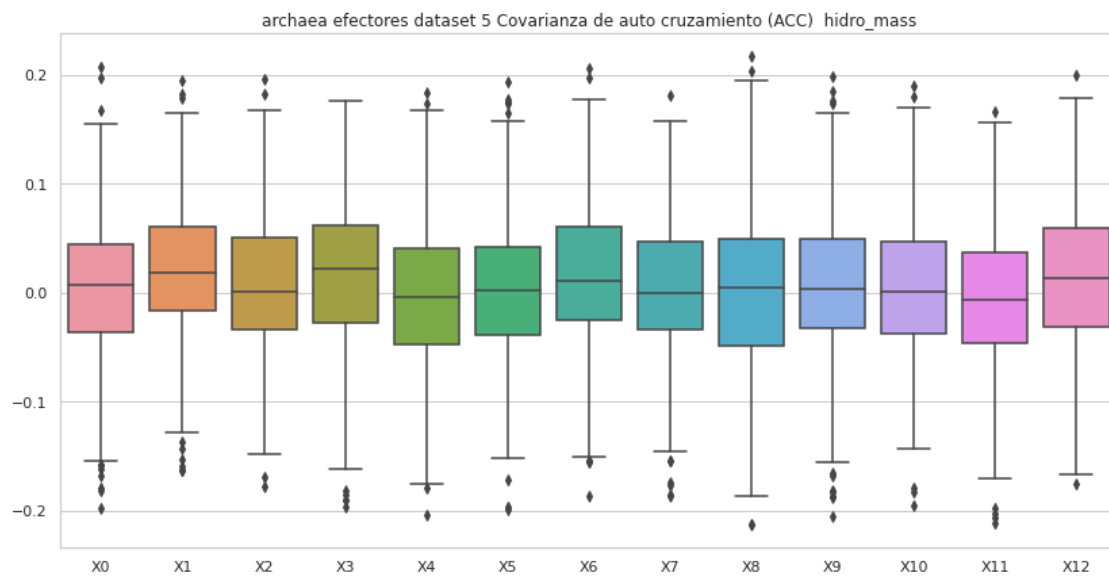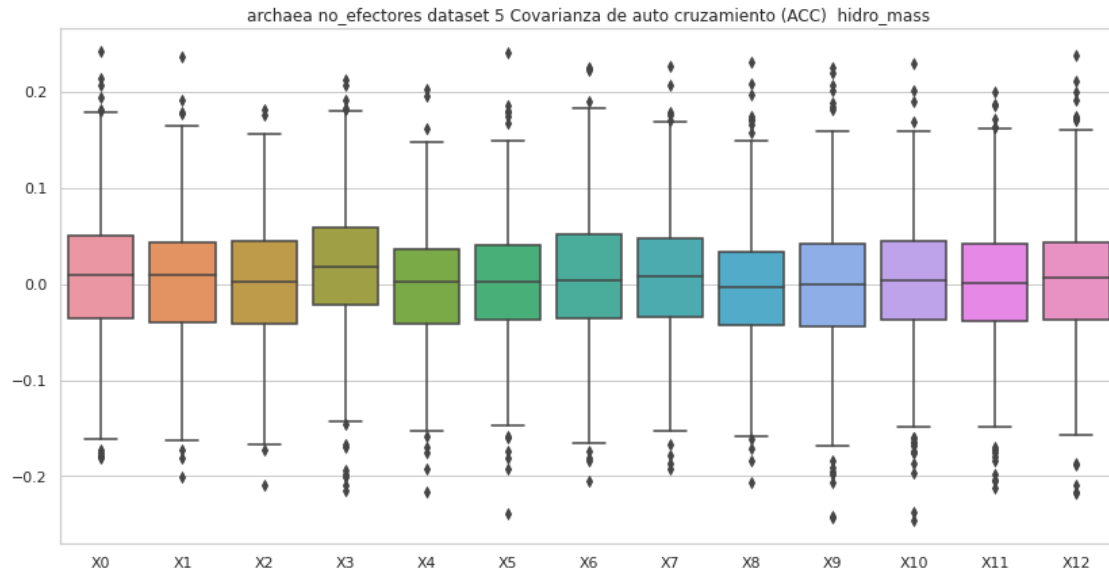
min	-0.205749	-0.191818	-0.207025	-0.243338	-0.246122	-0.212359
25%	-0.035645	-0.034537	-0.043265	-0.043730	-0.037068	-0.038730
50%	0.003541	0.007877	-0.002879	-0.000651	0.003217	0.000835
75%	0.052084	0.046913	0.033999	0.042362	0.044183	0.041906
max	0.225469	0.226316	0.231333	0.224799	0.230170	0.200571

	X12
count	459.000000
mean	0.005806
std	0.065855
min	-0.218254
25%	-0.037072
50%	0.007276
75%	0.043040
max	0.238349

archaea efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro_mass

archaea no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro_mass

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 archaea dataset 5, con
valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.068499	0.107372	-0.178466	-0.105813	-0.104190	-0.079042	-0.023018
1	0.016127	0.057621	-0.090205	-0.135102	-0.101268	-0.026928	-0.027177
2	-0.064370	0.019903	-0.113168	-0.112980	-0.147626	-0.056355	0.037008
3	-0.003813	0.007852	-0.010343	0.051148	-0.010351	-0.014167	-0.020029
4	-0.013156	-0.009046	-0.169310	-0.000608	-0.071155	-0.107506	-0.017740
..	…	…	…	…	…	…	…
495	-0.063990	0.072973	0.051820	0.175759	0.069077	0.080243	0.115945
496	0.055211	0.003064	-0.052672	0.106847	0.064771	0.102201	0.009929
497	-0.081476	0.046156	-0.038680	0.093794	-0.014769	-0.091086	-0.026024
498	-0.028050	0.073232	0.096939	-0.019894	0.010355	0.114560	-0.000925
499	-0.061669	0.062469	0.010104	0.054898	0.062286	-0.013364	0.041625

	X7	X8	X9	X10	X11	X12	X13
0	-0.067292	0.034984	0.057295	0.135773	0.043677	-0.005830	efectores
1	0.024964	-0.056175	-0.035597	0.147024	0.061575	0.100781	efectores
2	0.040063	-0.002298	0.110706	0.004580	0.067847	0.040649	efectores
3	0.061273	-0.010972	-0.012717	-0.010953	0.021089	0.009181	efectores
4	-0.025899	0.188517	-0.147207	0.082376	-0.043421	0.107928	efectores
..	…	…	…	…	…	…	…
495	0.024596	-0.004279	0.096415	-0.059640	0.128476	-0.077162	efectores
496	0.104859	0.062971	0.102625	0.003787	-0.003867	-0.069779	efectores
497	-0.060197	0.059830	0.010803	-0.086266	-0.145322	0.051021	efectores
498	-0.048277	-0.006971	0.021771	0.010893	0.045474	0.087363	efectores
499	-0.044397	0.048455	-0.036730	0.011376	0.013997	0.026100	efectores

[500 rows x 14 columns]

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

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.004686	0.019443	0.006366	0.012882	-0.003335	0.000429

```
std      0.068388     0.067242     0.068452     0.071165     0.069201     0.067680
min     -0.228927    -0.238098    -0.256714    -0.198361    -0.261174    -0.219974
25%     -0.038149    -0.018739    -0.037028    -0.029823    -0.047300    -0.040613
50%      0.006632     0.018324    -0.000204     0.019779    -0.003333     0.002533
75%      0.047288     0.060533     0.049812     0.061666     0.045616     0.044477
max      0.206642     0.236765     0.238997     0.175759     0.285489     0.218508

                   X6           X7           X8           X9          X10          X11   \
count  500.000000   500.000000   500.000000   500.000000   500.000000   500.000000
mean     0.014125     0.001228    -0.001182     0.003593     0.002251    -0.010216
std      0.068203     0.064911     0.073297     0.070919     0.067539     0.072233
min     -0.212350    -0.266405    -0.261605    -0.294274    -0.254213    -0.257899
25%     -0.026883    -0.034410    -0.050175    -0.035483    -0.039186    -0.050256
50%      0.009027    -0.000754     0.004214     0.003678     0.001396    -0.006946
75%      0.060594     0.046850     0.049137     0.049204     0.047463     0.037677
max      0.218983     0.238644     0.267867     0.198706     0.234874     0.241133

                  X12
count  500.000000
mean     0.010351
std      0.071877
min     -0.247439
25%     -0.034986
50%      0.011396
75%      0.058693
max      0.237507




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

            X0           X1           X2           X3           X4           X5           X6   \
0   -0.048139   0.023891  -0.115901   0.057774   0.032071   0.063435  -0.184426
1    0.143509   0.082342  -0.131607  -0.059306  -0.041152  -0.013132   0.189856
2    0.120228  -0.038323  -0.017979   0.019364  -0.015202   0.066651   0.075608
3   -0.019050   0.098471  -0.011074  -0.048137  -0.058446  -0.109650   0.007446
4    0.008016  -0.059660  -0.092342  -0.050158   0.103892   0.017644   0.001662
..        …           …           …           …           …           …           …
495  0.042510   0.038375   0.025392   0.100591  -0.015054  -0.192800   0.010897
496  0.086749   0.038354  -0.072797  -0.003569  -0.052947   0.051965   0.085290
497 -0.112976   0.071173  -0.098027  -0.047056  -0.059595   0.125019  -0.072861
498  0.214405  -0.079456   0.031113   0.169249   0.023345  -0.093321  -0.082705
499 -0.051489   0.010894   0.033417   0.034248   0.035250  -0.025396   0.023340

            X7           X8           X9          X10          X11          X12          X13
```
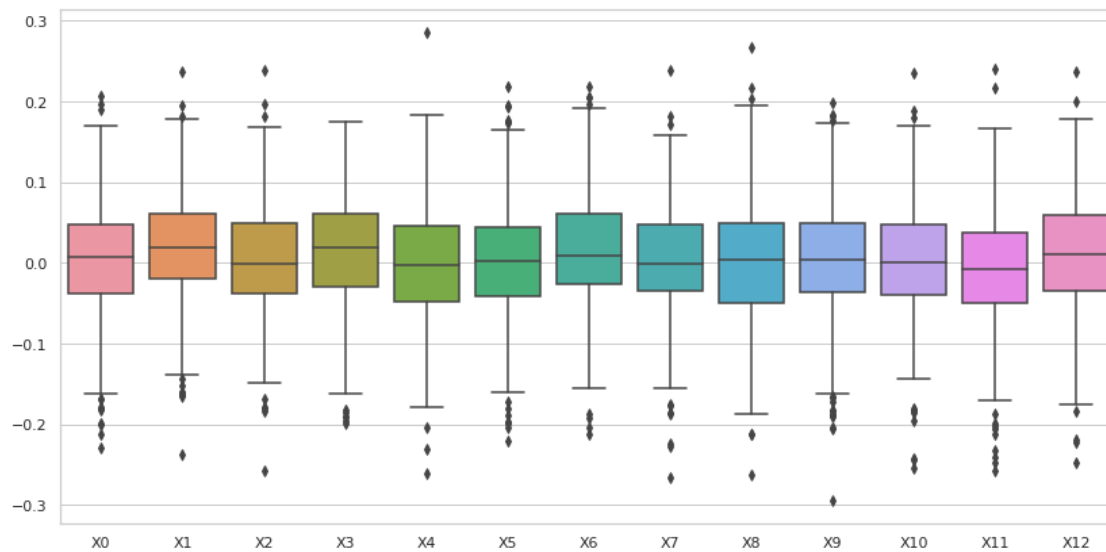
```
0      0.053476 -0.006586  0.049907 -0.019029  0.049160  0.028856  no_efectores
1     -0.085710 -0.049765 -0.286697 -0.160311 -0.190254  0.146958  no_efectores
2      0.021937 -0.104300  0.026756  0.137558 -0.035878 -0.022094  no_efectores
3     -0.071513  0.085593 -0.084771  0.001188  0.013352  0.057126  no_efectores
4      0.024062  0.011749  0.056943 -0.104491 -0.107386  0.042677  no_efectores

..          …         …         …         …         …         …
495    0.055095 -0.089681 -0.079964  0.057654 -0.078424  0.014737  no_efectores
496    0.053648 -0.060742  0.020493 -0.052042  0.023949 -0.104392  no_efectores
497    0.133627  0.007436  0.053858 -0.056722 -0.011216 -0.015837  no_efectores
498    0.226316 -0.064895  0.015656  0.201966  0.059112  0.053016  no_efectores
499    0.002935 -0.020942  0.034431  0.016460  0.042444 -0.070771  no_efectores

[500 rows x 14 columns]


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

              X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.012109    0.003867    0.002089    0.021783   -0.001476    0.000593
std      0.078780    0.078712    0.075889    0.081485    0.074803    0.086175
min     -0.248988   -0.225141   -0.268128   -0.320227   -0.378918   -0.501157
25%     -0.036597   -0.043747   -0.045124   -0.021818   -0.042399   -0.039860
50%      0.009497    0.008305   -0.000183    0.019198    0.002477    0.002454
75%      0.055105    0.046172    0.045564    0.067872    0.038706    0.044546
max      0.430307    0.465249    0.441076    0.362052    0.233730    0.324680

              X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.005788    0.005003   -0.002200   -0.001916    0.000901    0.000368
std      0.083560    0.080759    0.081018    0.082513    0.086986    0.087428
min     -0.393276   -0.517270   -0.355367   -0.399693   -0.517175   -0.403017
25%     -0.038643   -0.036374   -0.046123   -0.044954   -0.038545   -0.043917
50%      0.004078    0.007814   -0.004344   -0.000443    0.002162    0.001473
75%      0.054093    0.047790    0.036504    0.047868    0.046292    0.045048
max      0.356660    0.333427    0.418887    0.248618    0.433093    0.427972

             X12
count  500.000000
mean     0.012156
std      0.077348
min     -0.218254
25%     -0.037047
50%      0.009869
75%      0.049863
max      0.447849
```
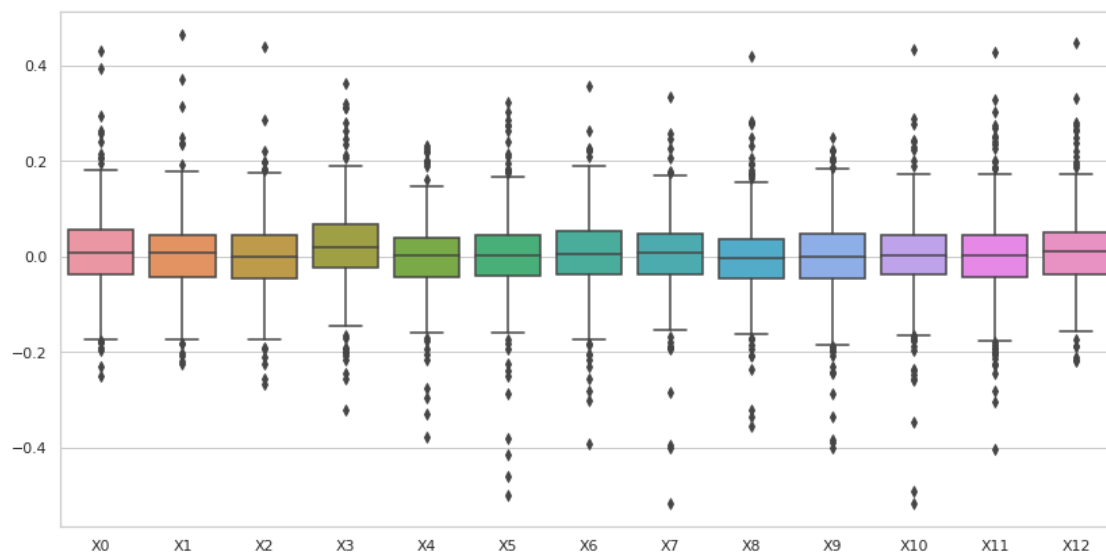
archaea efectores dataset 5 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.

archaea no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.

7.1 Covarianza de auto cruzamiento (ACC) mass, sin valores atípicos

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

```
          X0        X1        X2        X3        X4        X5        X6  \
0     0.068499  0.107372 -0.178466 -0.105813 -0.104190 -0.079042 -0.023018
1     0.016127  0.057621 -0.090205 -0.135102 -0.101268 -0.026928 -0.027177
2    -0.064370  0.019903 -0.113168 -0.112980 -0.147626 -0.056355  0.037008
3    -0.003813  0.007852 -0.010343  0.051148 -0.010351 -0.014167 -0.020029
4    -0.013156 -0.009046 -0.169310 -0.000608 -0.071155 -0.107506 -0.017740
..       …         …         …         …         …         …         …
495  -0.063990  0.072973  0.051820  0.175759  0.069077  0.080243  0.115945
496   0.055211  0.003064 -0.052672  0.106847  0.064771  0.102201  0.009929
497  -0.081476  0.046156 -0.038680  0.093794 -0.014769 -0.091086 -0.026024
498  -0.028050  0.073232  0.096939 -0.019894  0.010355  0.114560 -0.000925
499  -0.061669  0.062469  0.010104  0.054898  0.062286 -0.013364  0.041625

          X7        X8        X9       X10       X11       X12       X13
0    -0.067292  0.034984  0.057295  0.135773  0.043677 -0.005830  efectores
1     0.024964 -0.056175 -0.035597  0.147024  0.061575  0.100781  efectores
2     0.040063 -0.002298  0.110706  0.004580  0.067847  0.040649  efectores
3     0.061273 -0.010972 -0.012717 -0.010953  0.021089  0.009181  efectores
4    -0.025899  0.188517 -0.147207  0.082376 -0.043421  0.107928  efectores
..       …         …         …         …         …         …
495   0.024596 -0.004279  0.096415 -0.059640  0.128476 -0.077162  efectores
496   0.104859  0.062971  0.102625  0.003787 -0.003867 -0.069779  efectores
497  -0.060197  0.059830  0.010803 -0.086266 -0.145322  0.051021  efectores
498  -0.048277 -0.006971  0.021771  0.010893  0.045474  0.087363  efectores
499  -0.044397  0.048455 -0.036730  0.011376  0.013997  0.026100  efectores

[466 rows x 14 columns]
```

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

```
              X0          X1          X2          X3          X4          X5  \
count  466.000000  466.000000  466.000000  466.000000  466.000000  466.000000
mean     0.005248    0.021037    0.008234    0.013402   -0.004776    0.001811
std      0.063471    0.063015    0.066515    0.070671    0.065079    0.062894
min     -0.198197   -0.162904   -0.178466   -0.196046   -0.204237   -0.199454
25%     -0.036146   -0.016249   -0.034262   -0.028269   -0.047194   -0.038447
50%      0.006505    0.018555    0.000865    0.022314   -0.004390    0.002571
75%      0.044801    0.060682    0.050766    0.061243    0.040586    0.041345
```

```
max       0.206642     0.194479     0.196020     0.175759     0.183584     0.193606

                   X6           X7           X8           X9          X10          X11  \
count   466.000000   466.000000   466.000000   466.000000   466.000000   466.000000
mean      0.015261     0.002141    -0.000992     0.004903     0.003699    -0.008811
std       0.064530     0.060909     0.071050     0.066901     0.062451     0.067147
min      -0.186369    -0.186346    -0.212612    -0.204897    -0.194766    -0.211410
25%      -0.024987    -0.033274    -0.048490    -0.033241    -0.037731    -0.046755
50%       0.010528    -0.000152     0.004214     0.003923     0.001396    -0.006582
75%       0.060505     0.047064     0.048784     0.049116     0.046760     0.037473
max       0.205978     0.181051     0.216782     0.198706     0.189000     0.166311

                  X12
count   466.000000
mean      0.012900
std       0.067259
min      -0.174776
25%      -0.031809
50%       0.012721
75%       0.058772
max       0.199475
```

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

```
             X0          X1          X2          X3          X4          X5          X6  \
0     -0.048139    0.023891   -0.115901    0.057774    0.032071    0.063435   -0.184426
2      0.120228   -0.038323   -0.017979    0.019364   -0.015202    0.066651    0.075608
3     -0.019050    0.098471   -0.011074   -0.048137   -0.058446   -0.109650    0.007446
4      0.008016   -0.059660   -0.092342   -0.050158    0.103892    0.017644    0.001662
5      0.016298   -0.025893   -0.033229   -0.035516    0.014706   -0.027842   -0.055578
..          …           …           …           …           …           …
495    0.042510    0.038375    0.025392    0.100591   -0.015054   -0.192800    0.010897
496    0.086749    0.038354   -0.072797   -0.003569   -0.052947    0.051965    0.085290
497   -0.112976    0.071173   -0.098027   -0.047056   -0.059595    0.125019   -0.072861
498    0.214405   -0.079456    0.031113    0.169249    0.023345   -0.093321   -0.082705
499   -0.051489    0.010894    0.033417    0.034248    0.035250   -0.025396    0.023340

             X7          X8          X9         X10         X11         X12          X13
0      0.053476   -0.006586    0.049907   -0.019029    0.049160    0.028856   no_efectores
2      0.021937   -0.104300    0.026756    0.137558   -0.035878   -0.022094   no_efectores
3     -0.071513    0.085593   -0.084771    0.001188    0.013352    0.057126   no_efectores
4      0.024062    0.011749    0.056943   -0.104491   -0.107386    0.042677   no_efectores
5     -0.006026    0.035460    0.008538   -0.086719   -0.014035   -0.058483   no_efectores
..          …           …           …           …           …           …
```

```
495  0.055095 -0.089681 -0.079964  0.057654 -0.078424  0.014737  no_efectores
496  0.053648 -0.060742  0.020493 -0.052042  0.023949 -0.104392  no_efectores
497  0.133627  0.007436  0.053858 -0.056722 -0.011216 -0.015837  no_efectores
498  0.226316 -0.064895  0.015656  0.201966  0.059112  0.053016  no_efectores
499  0.002935 -0.020942  0.034431  0.016460  0.042444 -0.070771  no_efectores

[459 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5 \
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000
mean	0.009558	0.003412	0.002798	0.018100	-0.001171	0.000444
std	0.066760	0.066325	0.065009	0.069054	0.064208	0.064575
min	-0.180572	-0.200821	-0.209648	-0.215713	-0.216231	-0.239224
25%	-0.035250	-0.039789	-0.040816	-0.021809	-0.041233	-0.037349
50%	0.008974	0.009393	0.002399	0.017930	0.002317	0.001840
75%	0.050898	0.043463	0.045114	0.059468	0.036563	0.040196
max	0.242271	0.237388	0.181767	0.212763	0.202340	0.240555

	X6	X7	X8	X9	X10	X11 \
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000
mean	0.006616	0.006808	-0.001949	-0.000065	0.001553	-0.001439
std	0.070959	0.066106	0.064862	0.069486	0.067825	0.067652
min	-0.205749	-0.191818	-0.207025	-0.243338	-0.246122	-0.212359
25%	-0.035645	-0.034537	-0.043265	-0.043730	-0.037068	-0.038730
50%	0.003541	0.007877	-0.002879	-0.000651	0.003217	0.000835
75%	0.052084	0.046913	0.033999	0.042362	0.044183	0.041906
max	0.225469	0.226316	0.231333	0.224799	0.230170	0.200571

	X12
count	459.000000
mean	0.005806
std	0.065855
min	-0.218254
25%	-0.037072
50%	0.007276
75%	0.043040
max	0.238349

archaea efectores dataset 5 Covarianza de auto cruzamiento (ACC) mass

archaea no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) mass

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=""
```

```python
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 archaea dataset 5, con
valores atípicos.
Valores del documento csv.

```
           X0         X1         X2         X3         X4         X5         X6  \
0    0.089182  -0.174278   0.046602   0.033341  -0.082182  -0.120752   0.219370
1    0.138356  -0.111038   0.114225   0.018389  -0.106915  -0.016479   0.164796
2   -0.121338   0.063207  -0.020310  -0.091511  -0.015375  -0.023094  -0.080522
3   -0.066724  -0.070134  -0.010671  -0.014001  -0.021875  -0.032418   0.001706
4   -0.057204  -0.205386  -0.047229  -0.002657  -0.062139  -0.123377   0.025055
..        …          …          …          …          …          …          …
495 -0.129926  -0.088610   0.087587   0.058702  -0.046825  -0.123631   0.152195
496 -0.074801  -0.017113   0.019233   0.079899  -0.097704  -0.049722  -0.001414
497  0.014260  -0.090252   0.100970   0.098423  -0.019913   0.072377  -0.006211
498  0.014558  -0.064813   0.111642   0.097768   0.041359  -0.065428   0.045989
499 -0.009802   0.003879   0.089490  -0.052685   0.051484  -0.061831  -0.037548


           X7         X8         X9        X10        X11        X12        X13
0    0.119642  -0.207816  -0.071296   0.135222  -0.148161  -0.184044   efectores
1    0.036070  -0.250797  -0.175054   0.041686  -0.208852  -0.077793   efectores
2    0.144137  -0.116426   0.107286  -0.048817   0.051964  -0.080031   efectores
```

```
3   -0.089313  0.015779  0.037114 -0.010818 -0.058484  0.045104  efectores
4    0.135697  0.212670 -0.010802 -0.026986 -0.013582 -0.126521  efectores
..        …         …         …         …         …         …
495 -0.113882 -0.098616  0.112847  0.064770  0.016111 -0.106549  efectores
496  0.083770  0.031161  0.000394  0.138276 -0.018115 -0.057235  efectores
497 -0.092420 -0.017312  0.051517 -0.098384 -0.108518  0.135634  efectores
498  0.070769  0.031793 -0.065109 -0.009433  0.046870 -0.001291  efectores
499  0.030382 -0.068497  0.040019 -0.022607 -0.082964  0.003571  efectores

[500 rows x 14 columns]
```

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

```
                 X0          X1          X2          X3          X4          X5  \
count    500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean       0.018141   -0.029878    0.046716    0.040358   -0.015825   -0.018196
std        0.094314    0.097501    0.085078    0.086977    0.093483    0.093953
min       -0.317312   -0.385520   -0.258043   -0.227000   -0.449435   -0.471736
25%       -0.035890   -0.091165   -0.007800   -0.010210   -0.074594   -0.070552
50%        0.021355   -0.018896    0.039793    0.037988   -0.007382   -0.017914
75%        0.074189    0.038241    0.095117    0.093591    0.047697    0.039327
max        0.378712    0.299185    0.471753    0.338273    0.322931    0.319221

                 X6          X7          X8          X9         X10         X11  \
count    500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean       0.031763    0.022032   -0.006254   -0.001127    0.021722    0.004634
std        0.092079    0.084227    0.096588    0.086142    0.084276    0.080921
min       -0.335820   -0.303149   -0.350234   -0.343687   -0.342963   -0.311324
25%       -0.026073   -0.025368   -0.051440   -0.044648   -0.027941   -0.043923
50%        0.027930    0.014248   -0.002416    0.002879    0.013504    0.006744
75%        0.083497    0.071738    0.044735    0.049114    0.071717    0.053549
max        0.412053    0.364225    0.285005    0.278366    0.287297    0.282216

                X12
count    500.000000
mean      -0.010732
std        0.084431
min       -0.436184
25%       -0.056343
50%       -0.002772
75%        0.036446
max        0.297516
```

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

```
            X0         X1         X2         X3         X4         X5         X6  \
0    -0.125232  -0.038040   0.005123   0.082253   0.017242  -0.015029   0.112619
1    -0.013821   0.054194   0.095463   0.023881  -0.021312  -0.054269   0.024202
2    -0.011325   0.016204  -0.075378  -0.007041  -0.105749  -0.091757   0.011913
3     0.037439  -0.069923   0.014851  -0.097595  -0.027776   0.011152  -0.051213
4    -0.023366   0.042905   0.045191  -0.024311   0.029666  -0.022818  -0.074880
..         …          …          …          …          …          …
495  -0.092752   0.087656  -0.050109   0.014439  -0.004154   0.104220   0.122174
496  -0.018171   0.016649   0.092600   0.043949  -0.027676  -0.018707   0.065040
497  -0.154654   0.200545  -0.088569   0.097168   0.022947   0.192665  -0.029478
498   0.081605   0.037719   0.060628  -0.014753   0.038763   0.073643   0.073367
499   0.063300   0.043111   0.123765   0.119925   0.054601   0.084780   0.061398


            X7         X8         X9        X10        X11        X12          X13
0    -0.068698  -0.024696   0.130475   0.085890  -0.127741   0.064292  no_efectores
1    -0.115146  -0.134892   0.039608   0.027233  -0.280978   0.006459  no_efectores
2     0.036224  -0.033645   0.152743  -0.114465  -0.002872  -0.031007  no_efectores
3     0.030127  -0.032373  -0.073340  -0.090690   0.025216   0.040694  no_efectores
4    -0.031535  -0.046741  -0.039711  -0.068718  -0.018166   0.028707  no_efectores
..         …          …          …          …          …          …
495   0.063123  -0.016370  -0.048066   0.056469  -0.100316   0.122003  no_efectores
496   0.032224   0.029022  -0.000252   0.001392  -0.002465  -0.017014  no_efectores
497   0.220080   0.030072  -0.146169   0.098218  -0.111060   0.141097  no_efectores
498   0.045005   0.049532  -0.056570  -0.009485   0.014066   0.014227  no_efectores
499  -0.028222   0.036358   0.041051  -0.023485  -0.030115   0.096599  no_efectores


[500 rows x 14 columns]
```
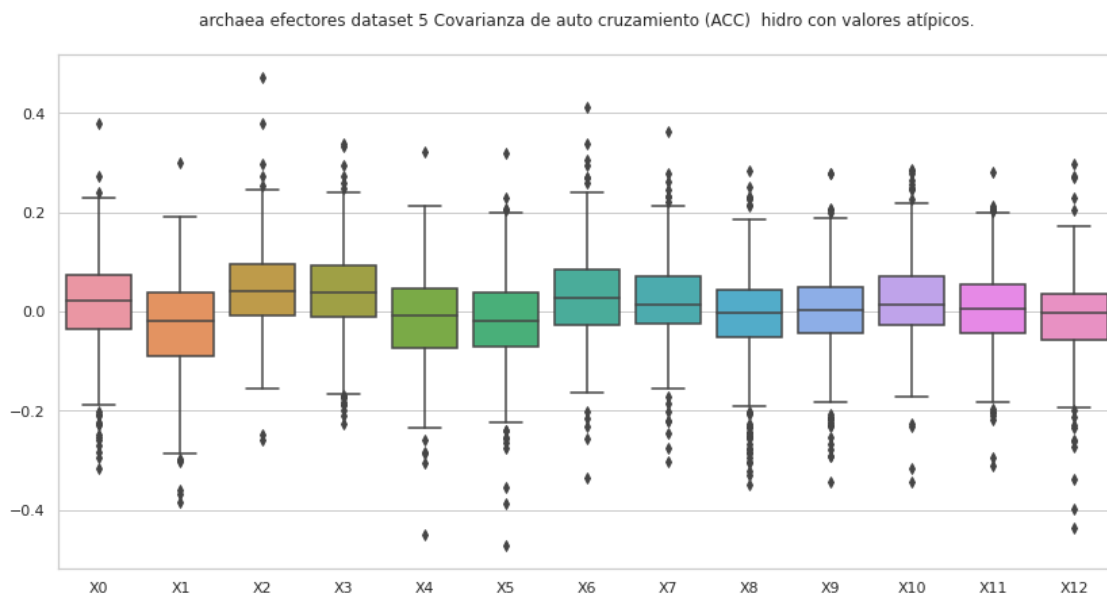
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 5, con
valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean    -0.011459   -0.033609    0.010999    0.032321   -0.021461   -0.019881
std      0.090281    0.105733    0.091480    0.092174    0.092410    0.087450
min     -0.368536   -0.586050   -0.304335   -0.330369   -0.399798   -0.308088
25%     -0.064826   -0.088672   -0.041199   -0.022568   -0.077070   -0.071096
50%     -0.010917   -0.024311    0.008040    0.031574   -0.021761   -0.022617
75%      0.042102    0.029092    0.060245    0.081810    0.033506    0.025901
max      0.489602    0.318450    0.394332    0.403058    0.355171    0.487786
```
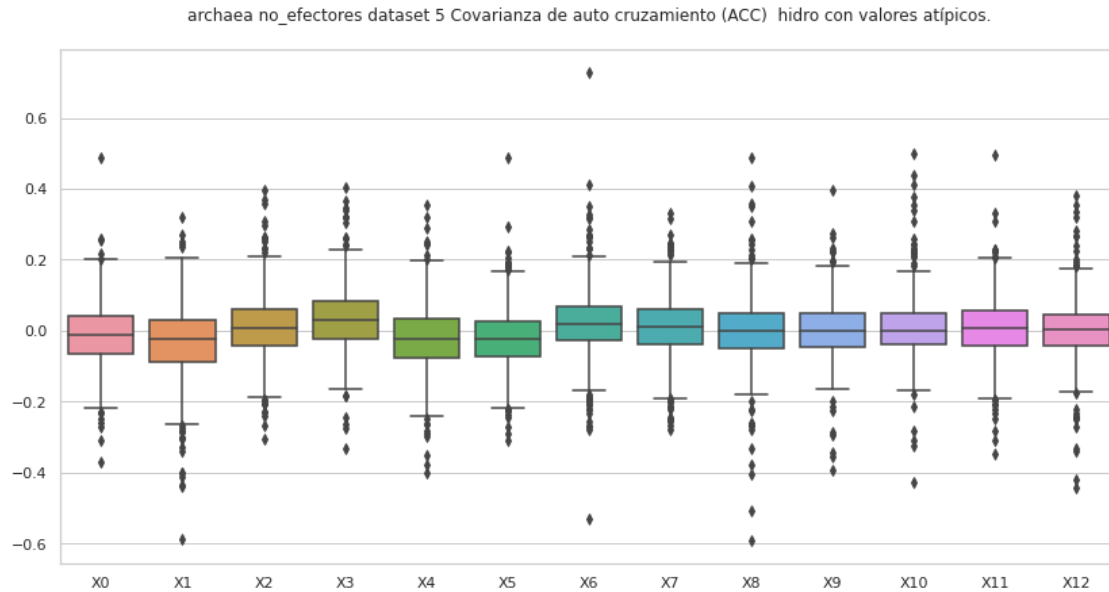
```
              X6          X7          X8          X9         X10         X11    \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.020832    0.009813   -0.001871    0.001837    0.009734    0.007149
std      0.101753    0.087090    0.099632    0.083938    0.092684    0.089704
min     -0.531214   -0.278428   -0.591300   -0.394478   -0.426880   -0.347929
25%     -0.028092   -0.039737   -0.050942   -0.046605   -0.039551   -0.043810
50%      0.019600    0.012930    0.000739   -0.000180    0.001044    0.009196
75%      0.068274    0.060635    0.047424    0.047811    0.048753    0.056326
max      0.727644    0.330108    0.488435    0.397984    0.499322    0.496839

             X12
count  500.000000
mean     0.001387
std      0.091095
min     -0.441250
25%     -0.043113
50%      0.001831
75%      0.044031
max      0.379894
```


archaea efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.

archaea no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.

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 archaea dataset 5, sin valores
atípicos.
Valores del documento csv.

```
           X0         X1         X2         X3         X4         X5         X6  \
0     0.089182 -0.174278   0.046602   0.033341 -0.082182 -0.120752   0.219370
1     0.138356 -0.111038   0.114225   0.018389 -0.106915 -0.016479   0.164796
2    -0.121338   0.063207 -0.020310 -0.091511 -0.015375 -0.023094 -0.080522
3    -0.066724 -0.070134 -0.010671 -0.014001 -0.021875 -0.032418   0.001706
4    -0.057204 -0.205386 -0.047229 -0.002657 -0.062139 -0.123377   0.025055
..         …          …          …          …          …          …          …
495  -0.129926 -0.088610   0.087587   0.058702 -0.046825 -0.123631   0.152195
496  -0.074801 -0.017113   0.019233   0.079899 -0.097704 -0.049722 -0.001414
497   0.014260 -0.090252   0.100970   0.098423 -0.019913   0.072377 -0.006211
498   0.014558 -0.064813   0.111642   0.097768   0.041359 -0.065428   0.045989
499  -0.009802   0.003879   0.089490 -0.052685   0.051484 -0.061831 -0.037548

           X7         X8         X9        X10        X11        X12        X13
0     0.119642 -0.207816 -0.071296   0.135222 -0.148161 -0.184044   efectores
1     0.036070 -0.250797 -0.175054   0.041686 -0.208852 -0.077793   efectores
2     0.144137 -0.116426   0.107286 -0.048817   0.051964 -0.080031   efectores
3    -0.089313   0.015779   0.037114 -0.010818 -0.058484   0.045104   efectores
4     0.135697   0.212670 -0.010802 -0.026986 -0.013582 -0.126521   efectores
..         …          …          …          …          …          …          …
495  -0.113882 -0.098616   0.112847   0.064770   0.016111 -0.106549   efectores
```

```
496  0.083770  0.031161  0.000394  0.138276 -0.018115 -0.057235  efectores
497 -0.092420 -0.017312  0.051517 -0.098384 -0.108518  0.135634  efectores
498  0.070769  0.031793 -0.065109 -0.009433  0.046870 -0.001291  efectores
499  0.030382 -0.068497  0.040019 -0.022607 -0.082964  0.003571  efectores

[463 rows x 14 columns]


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

              X0          X1          X2          X3          X4          X5  \
count  463.000000  463.000000  463.000000  463.000000  463.000000  463.000000
mean     0.020901   -0.026535    0.043949    0.042945   -0.012389   -0.014566
std      0.085543    0.092141    0.079202    0.080781    0.086765    0.084907
min     -0.258150   -0.303593   -0.155146   -0.198804   -0.258643   -0.274710
25%     -0.030363   -0.088453   -0.008847   -0.008351   -0.069221   -0.067133
50%      0.023113   -0.017693    0.035116    0.038378   -0.002967   -0.015171
75%      0.071355    0.039473    0.092550    0.093739    0.047724    0.040826
max      0.271874    0.190216    0.273258    0.294365    0.213410    0.205809

              X6          X7          X8          X9         X10         X11  \
count  463.000000  463.000000  463.000000  463.000000  463.000000  463.000000
mean     0.032748    0.023102   -0.004392    0.000432    0.019298    0.005858
std      0.084512    0.076869    0.086280    0.076315    0.075185    0.075211
min     -0.230878   -0.221205   -0.293547   -0.232339   -0.226890   -0.216718
25%     -0.023997   -0.022583   -0.046908   -0.043653   -0.027618   -0.041629
50%      0.028502    0.013244   -0.002126    0.003199    0.011018    0.007586
75%      0.082337    0.071262    0.043357    0.045930    0.066854    0.051955
max      0.305363    0.263230    0.229358    0.203648    0.257069    0.213794

             X12
count  463.000000
mean    -0.007527
std      0.072539
min     -0.260800
25%     -0.052533
50%     -0.001085
75%      0.036142
max      0.229786



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

```
            X0        X1        X2        X3        X4        X5        X6   \
0    -0.125232 -0.038040  0.005123  0.082253  0.017242 -0.015029  0.112619
2    -0.011325  0.016204 -0.075378 -0.007041 -0.105749 -0.091757  0.011913
3     0.037439 -0.069923  0.014851 -0.097595 -0.027776  0.011152 -0.051213
4    -0.023366  0.042905  0.045191 -0.024311  0.029666 -0.022818 -0.074880
5    -0.135607 -0.119306 -0.038440  0.033921 -0.019414  0.024013  0.041417
..         …         …         …         …         …         …
495  -0.092752  0.087656 -0.050109  0.014439 -0.004154  0.104220  0.122174
496  -0.018171  0.016649  0.092600  0.043949 -0.027676 -0.018707  0.065040
497  -0.154654  0.200545 -0.088569  0.097168  0.022947  0.192665 -0.029478
498   0.081605  0.037719  0.060628 -0.014753  0.038763  0.073643  0.073367
499   0.063300  0.043111  0.123765  0.119925  0.054601  0.084780  0.061398

            X7        X8        X9       X10       X11       X12          X13
0    -0.068698 -0.024696  0.130475  0.085890 -0.127741  0.064292  no_efectores
2     0.036224 -0.033645  0.152743 -0.114465 -0.002872 -0.031007  no_efectores
3     0.030127 -0.032373 -0.073340 -0.090690  0.025216  0.040694  no_efectores
4    -0.031535 -0.046741 -0.039711 -0.068718 -0.018166  0.028707  no_efectores
5    -0.035031 -0.061368 -0.006950  0.032606  0.053671 -0.060363  no_efectores
..         …         …         …         …         …         …
495   0.063123 -0.016370 -0.048066  0.056469 -0.100316  0.122003  no_efectores
496   0.032224  0.029022 -0.000252  0.001392 -0.002465 -0.017014  no_efectores
497   0.220080  0.030072 -0.146169  0.098218 -0.111060  0.141097  no_efectores
498   0.045005  0.049532 -0.056570 -0.009485  0.014066  0.014227  no_efectores
499  -0.028222  0.036358  0.041051 -0.023485 -0.030115  0.096599  no_efectores

[452 rows x 14 columns]
```

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

```
               X0          X1          X2          X3          X4          X5   \
count  452.000000  452.000000  452.000000  452.000000  452.000000  452.000000
mean    -0.011809   -0.027066    0.007746    0.027639   -0.021407   -0.019043
std      0.080782    0.084753    0.076409    0.075793    0.078987    0.075392
min     -0.270045   -0.338999   -0.230250   -0.184715   -0.284341   -0.270038
25%     -0.062327   -0.081924   -0.039633   -0.021872   -0.074598   -0.066959
50%     -0.010337   -0.021051    0.007235    0.028817   -0.022559   -0.022004
75%      0.037744    0.027248    0.055574    0.076605    0.030987    0.024131
max      0.254214    0.243426    0.262391    0.262218    0.243272    0.224437

               X6          X7          X8          X9         X10         X11   \
count  452.000000  452.000000  452.000000  452.000000  452.000000  452.000000
mean     0.017557    0.008531    0.000361    0.003169    0.005975    0.007268
std      0.081705    0.076811    0.073809    0.068432    0.069281    0.076526
```

```
min     -0.276928   -0.250099   -0.280223   -0.226412   -0.213873   -0.232643
25%     -0.027237   -0.036343   -0.046495   -0.043758   -0.037307   -0.037651
50%      0.018399    0.012039    0.000739   -0.000180    0.000433    0.009315
75%      0.064954    0.058389    0.045585    0.042353    0.047099    0.052320
max      0.319408    0.243626    0.254570    0.234051    0.243453    0.228831

                X12
count   452.000000
mean      0.002184
std       0.070516
min      -0.246576
25%      -0.038283
50%       0.003189
75%       0.041881
max       0.265398
```


archaea efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro

archaea no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro