

ds3_archaea_limpieza_de_datos

January 19, 2021

Limpieza de datos

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

1 Declaración de variables

```
[2]: organismo = "archaea"
dataset = 3
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo) + " dataset " + str(dataset))
r2 = ("Datos/resultados/" + str(organismo) + "/" + str(nombre) + "/"
      ↪ transformaciones/sin_filtrar")
r3 = ("Datos/resultados/" + str(organismo) + "/" + str(nombre) + "/"
      ↪ transformaciones/sin_atipicos")

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

```

nom8 = ("/ds" + str(dataset) + "_AAC_no_efectores_" + str(organismo) + ".txt")
nom9 = ("/ds" + str(dataset) + "_ACC_hidro_mass_no_efectores_" + str(organismo) +
    ↳ ".txt")
nom10 = ("/ds" + str(dataset) + "_ACC_mass_no_efectores_" + str(organismo) + ".
    ↳ txt")
nom11 = ("/ds" + str(dataset) + "_ACC_hidro_no_efectores_" + str(organismo) + ".
    ↳ txt")
nom12 = ("/ds" + str(dataset) + "_PseAAC_hidro_mass_no_efectores_" +
    ↳ str(organismo) + ".txt")
nom13 = ("/ds" + str(dataset) + "_PseAAC_mass_no_efectores_" + str(organismo) +
    ↳ ".txt")
nom14 = ("/ds" + str(dataset) + "_PseAAC_hidro_no_efectores_" + str(organismo) +
    ↳ ".txt")

#Efectores
AAC_efec= pd.read_csv(str(r2) + str(nom1), header=None,prefix='X',sep=',')
ACC_hidro_mass_efec = pd.read_csv(str(r2) + str(nom2),
    ↳ header=None,prefix='X',sep=',')
ACC_mass_efec = pd.read_csv(str(r2) + str(nom3), header=None,prefix='X',sep=',')
ACC_hidro_efec = pd.read_csv(str(r2) + str(nom4),
    ↳ header=None,prefix='X',sep=',')
PseAAC_hidro_mass_efec = pd.read_csv(str(r2) +str(nom5),
    ↳ header=None,prefix='X',sep=',')
PseAAC_mass_efec = pd.read_csv(str(r2) + str(nom6),
    ↳ header=None,prefix='X',sep=',')
PseAAC_hidro_efec = pd.read_csv(str(r2) + str(nom7),
    ↳ header=None,prefix='X',sep=',')

#No efectores
AAC_no_efec= pd.read_csv(str(r2) + str(nom8), header=None,prefix='X',sep=',')
ACC_hidro_mass_no_efec =pd.read_csv(str(r2) + str(nom9),
    ↳ header=None,prefix='X',sep=',')
ACC_mass_no_efec =pd.read_csv(str(r2) + str(nom10),
    ↳ header=None,prefix='X',sep=',')
ACC_hidro_no_efec =pd.read_csv(str(r2) + str(nom11),
    ↳ header=None,prefix='X',sep=',')
PseAAC_hidro_mass_no_efec =pd.read_csv(str(r2) + str(nom12),
    ↳ header=None,prefix='X',sep=',')
PseAAC_mass_no_efec =pd.read_csv(str(r2) + str(nom13),
    ↳ header=None,prefix='X',sep=',')
PseAAC_hidro_no_efec =pd.read_csv(str(r2) + str(nom14),
    ↳ header=None,prefix='X',sep=',')

```

2 Composición de aminoácidos (AAC)

```
[3]: transf = "Composición de aminoácidos (AAC) "
eti="efectores "
estado = "con valores atípicos.\n"
df=""

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

    if eti == "efectores":
        df=AAC_efec

    if eti == "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)+"\n
    ↪"+str(transf)+" " +str(estado))
```

efectores

Composición de aminoácidos (AAC) efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	3.650	5.839	4.015	5.109	0.000	7.299	2.190	7.299	1.825	8.394	
1	6.250	8.125	1.875	4.375	0.625	13.125	1.250	7.500	0.000	6.250	
2	4.734	5.917	4.142	2.959	0.592	13.609	2.959	5.917	4.142	6.509	
3	13.744	5.687	1.896	5.687	0.000	3.318	1.896	8.057	2.370	1.422	
4	4.000	4.000	2.667	10.667	2.667	4.000	4.000	2.667	0.000	6.667	
..	
495	13.529	6.471	0.588	4.118	0.000	4.118	1.176	8.824	2.353	3.529	
496	3.448	5.517	4.138	4.828	2.759	7.586	5.517	4.828	0.000	4.138	
497	8.989	9.738	1.124	9.363	0.375	11.610	0.749	6.367	2.622	2.622	
498	6.706	4.082	6.997	6.997	0.000	7.289	2.915	4.665	1.166	8.163	
499	4.000	2.857	4.857	5.143	0.286	8.000	2.286	8.857	4.571	7.429	

	...	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	...	6.934	1.095	3.285	3.285	7.299	5.109	1.460	6.204	6.569	
1	...	8.750	3.750	3.125	5.000	3.750	5.625	0.000	1.875	8.125	
2	...	10.059	4.734	3.550	2.959	3.550	4.142	0.592	2.367	5.325	
3	...	0.948	1.422	2.370	6.635	6.161	7.583	2.844	1.896	12.796	
4	...	8.000	5.333	4.000	4.000	4.000	6.667	0.000	4.000	8.000	
..	
495	...	1.176	1.765	4.118	5.882	3.529	3.529	3.529	3.529	17.059	
496	...	9.655	2.069	2.759	3.448	8.966	6.207	0.690	5.517	7.586	
497	...	0.375	1.498	3.745	5.243	6.742	6.742	0.375	1.873	8.614	
498	...	9.038	2.332	4.373	1.749	6.414	8.163	1.166	3.790	4.373	
499	...	3.429	3.714	7.429	5.714	4.857	6.571	0.571	5.429	8.571	

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

[500 rows x 21 columns]

Composición de aminoácidos (AAC) efectores archaea dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	9.348246	5.958586	2.713114	5.902288	0.682458	7.012430	
std	4.516352	2.549451	2.128044	2.749127	0.992053	3.820584	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	5.948000	4.163250	1.136750	3.813750	0.000000	3.864000	
50%	8.513500	5.810000	2.174000	5.430000	0.357000	7.048500	
75%	12.362750	7.760000	3.968500	7.492000	0.862000	9.598500	
max	22.546000	17.460000	12.963000	16.149000	6.796000	17.073000	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	2.303266	7.525722	1.829600	5.930376	11.550512	4.514208	
std	1.627185	2.879360	1.166973	3.471339	3.301073	4.336519	

min	0.000000	1.370000	0.000000	0.000000	3.226000	0.000000
25%	1.131250	5.386000	0.926000	3.295000	9.210250	1.062500
50%	2.046000	7.474000	1.712000	5.115000	11.503000	2.417000
75%	3.148250	9.394000	2.617500	8.128750	13.689500	7.960500
max	11.194000	21.627000	5.505000	20.763000	25.543000	17.252000

	X12	X13	X14	X15	X16	X17 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	2.037300	4.039464	4.053840	5.985384	5.476678	1.249276
std	1.337696	2.040100	1.673869	2.116908	2.048159	1.134337
min	0.000000	0.000000	0.000000	0.826000	0.543000	0.000000
25%	0.988750	2.653000	2.905750	4.648000	3.998000	0.513250
50%	1.709500	3.877000	3.922000	5.716000	5.385500	1.017500
75%	2.748250	5.137000	5.019000	7.037500	6.587750	1.703000
max	7.500000	12.712000	12.883000	16.883000	15.054000	10.326000

	X18	X19
count	500.000000	500.000000
mean	3.306446	8.580622
std	1.593230	3.408375
min	0.000000	0.671000
25%	2.273000	5.875000
50%	3.188500	8.127500
75%	4.180000	11.008250
max	10.870000	19.288000

no_efectores

Composición de aminoácidos (AAC) no_efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8 \
0	12.727	6.182	0.727	9.455	0.000	6.909	2.182	9.091	1.091
1	5.660	11.321	1.887	3.774	15.094	3.774	1.887	7.547	5.660
2	16.906	9.729	0.797	5.263	0.000	4.306	1.595	7.656	1.754
3	11.282	13.333	0.513	11.282	0.000	8.205	1.026	9.231	2.051
4	8.929	11.012	1.488	2.976	0.595	3.571	1.190	7.143	0.595
..
495	9.906	3.774	0.943	1.887	0.000	2.358	3.302	9.434	2.830
496	15.625	7.031	0.781	1.562	0.781	1.562	1.562	11.719	1.562
497	8.455	5.539	3.499	10.787	0.292	9.038	4.373	9.329	2.624
498	11.951	2.683	3.902	7.805	0.244	9.756	3.902	11.707	1.463
499	10.270	5.946	1.622	7.568	1.081	10.811	2.162	8.108	3.784

	X9	...	X11	X12	X13	X14	X15	X16	X17	X18 \
0	2.909	...	0.727	1.818	1.455	5.818	7.273	9.091	0.000	1.818

1	7.547	...	3.774	3.774	1.887	0.000	11.321	1.887	0.000	1.887
2	1.116	...	0.957	1.754	3.349	4.625	5.423	6.220	0.957	2.711
3	1.538	...	0.000	0.513	2.051	6.667	4.103	11.795	0.000	0.513
4	6.250	...	1.190	2.083	3.571	3.571	7.143	3.571	2.679	2.083
..
495	5.189	...	0.472	3.302	3.302	5.660	5.660	5.660	4.245	3.774
496	4.688	...	0.000	0.781	3.906	5.469	3.906	6.250	1.562	0.781
497	5.539	...	3.499	2.332	3.207	3.207	5.831	4.373	0.583	1.749
498	6.585	...	0.244	1.220	3.902	5.854	7.805	5.610	0.732	3.902
499	2.703	...	1.081	1.081	3.784	4.324	1.081	4.324	1.622	5.405

	X19	X20
0	11.273	no_efectores
1	3.774	no_efectores
2	13.078	no_efectores
3	8.718	no_efectores
4	12.798	no_efectores
..
495	15.094	no_efectores
496	15.625	no_efectores
497	6.997	no_efectores
498	5.610	no_efectores
499	10.270	no_efectores

[500 rows x 21 columns]

Composición de aminoácidos (AAC) no_efectores archaea dataset 3, con valores atípicos.

Estadísticas.

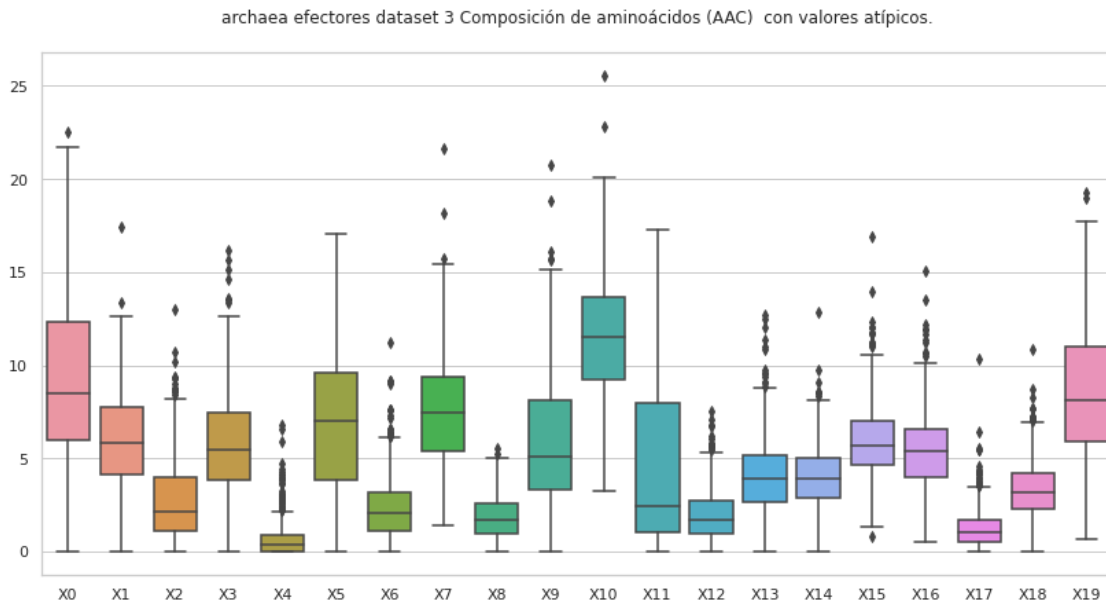
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	9.936334	6.420608	2.775802	7.830796	1.020508	8.298090	
std	3.991964	3.054691	2.047860	3.507036	1.740918	3.866161	
min	1.020000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	7.246000	4.557750	1.478500	5.263000	0.000000	5.791000	
50%	9.714000	6.067000	2.445000	7.878000	0.547500	8.333000	
75%	12.404500	8.116750	3.571000	9.955250	1.214000	10.554500	
max	26.389000	22.449000	12.450000	22.222000	15.094000	23.404000	

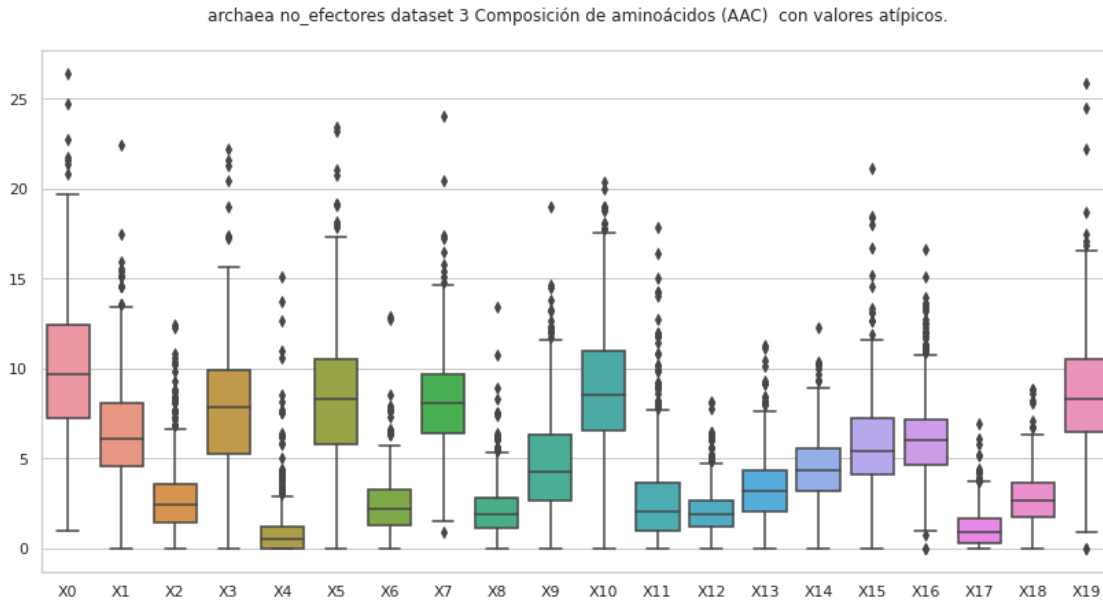
	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	2.416804	8.175994	2.147066	4.779728	8.955216	2.802592	
std	1.653418	2.835019	1.585133	2.826481	3.385651	2.828402	
min	0.000000	0.901000	0.000000	0.000000	0.000000	0.000000	
25%	1.326500	6.382500	1.121250	2.703000	6.548250	0.955750	
50%	2.182000	8.082000	1.923000	4.310000	8.578000	2.027000	

75%	3.284000	9.706250	2.815750	6.323250	10.994000	3.670250
max	12.857000	24.000000	13.462000	18.987000	20.339000	17.857000

	X12	X13	X14	X15	X16	X17 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	2.127840	3.379694	4.423932	5.866280	6.123758	1.158460
std	1.277302	1.894418	1.927234	2.679329	2.389128	1.085067
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.239500	2.044750	3.215000	4.119500	4.672000	0.336250
50%	1.887000	3.222000	4.348000	5.418000	6.040000	0.952000
75%	2.685000	4.374000	5.556000	7.222750	7.157000	1.698000
max	8.163000	11.321000	12.245000	21.127000	16.667000	6.944000

	X18	X19
count	500.000000	500.000000
mean	2.758168	8.602396
std	1.572551	3.314335
min	0.000000	0.000000
25%	1.760000	6.464000
50%	2.663500	8.333000
75%	3.647250	10.553250
max	8.861000	25.862000





2.1 Composición de aminoácidos (AAC), sin valores atípicos

```
[4]: transf = "Composición de aminoácidos (AAC) "
estado = "sin valores atípicos.\n"
transf2="AAC"

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' +
      str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df=""
df_out = pd.DataFrame()

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

    del df['X20']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de
    sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
```



```

df['X20'] = etiq
df_out = pd.concat([df_out,df])

#Guarda la lista csv sin valores atípicos.
df_out.to_csv(str(out), index=False, header=False)

print (str(titulo) + "Valores del documento csv.\n")
print (df)
print ("\n\n" + str(titulo) + "Estadísticas.\n")
print(df.describe())
print ("\n\n")

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

```

efectores

Composición de aminoácidos (AAC) efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	3.650	5.839	4.015	5.109	0.000	7.299	2.190	7.299	1.825	8.394	
1	6.250	8.125	1.875	4.375	0.625	13.125	1.250	7.500	0.000	6.250	
2	4.734	5.917	4.142	2.959	0.592	13.609	2.959	5.917	4.142	6.509	
3	13.744	5.687	1.896	5.687	0.000	3.318	1.896	8.057	2.370	1.422	
4	4.000	4.000	2.667	10.667	2.667	4.000	4.000	2.667	0.000	6.667	
..	
495	13.529	6.471	0.588	4.118	0.000	4.118	1.176	8.824	2.353	3.529	
496	3.448	5.517	4.138	4.828	2.759	7.586	5.517	4.828	0.000	4.138	
497	8.989	9.738	1.124	9.363	0.375	11.610	0.749	6.367	2.622	2.622	
498	6.706	4.082	6.997	6.997	0.000	7.289	2.915	4.665	1.166	8.163	
499	4.000	2.857	4.857	5.143	0.286	8.000	2.286	8.857	4.571	7.429	
...	
	X11	X12	X13	X14	X15	X16	X17	X18	X19	\	
0	6.934	1.095	3.285	3.285	7.299	5.109	1.460	6.204	6.569		
1	8.750	3.750	3.125	5.000	3.750	5.625	0.000	1.875	8.125		
2	10.059	4.734	3.550	2.959	3.550	4.142	0.592	2.367	5.325		
3	0.948	1.422	2.370	6.635	6.161	7.583	2.844	1.896	12.796		
4	8.000	5.333	4.000	4.000	4.000	6.667	0.000	4.000	8.000		
..		
495	1.176	1.765	4.118	5.882	3.529	3.529	3.529	3.529	17.059		

496	...	9.655	2.069	2.759	3.448	8.966	6.207	0.690	5.517	7.586
497	...	0.375	1.498	3.745	5.243	6.742	6.742	0.375	1.873	8.614
498	...	9.038	2.332	4.373	1.749	6.414	8.163	1.166	3.790	4.373
499	...	3.429	3.714	7.429	5.714	4.857	6.571	0.571	5.429	8.571

```

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

```

[435 rows x 21 columns]

Composición de aminoácidos (AAC) efectores archaea dataset 3, sin valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	435.000000	435.000000	435.000000	435.000000	435.000000	435.000000	
mean	9.589855	6.066092	2.647005	5.902306	0.563007	7.054349	
std	4.541112	2.410589	1.970806	2.624514	0.734000	3.704117	
min	0.000000	0.467000	0.000000	0.000000	0.000000	0.000000	
25%	6.114500	4.279000	1.112000	3.829500	0.000000	3.984000	
50%	9.091000	5.921000	2.174000	5.386000	0.307000	7.099000	
75%	12.530000	7.933500	3.927000	7.471500	0.803000	9.574000	
max	22.546000	13.372000	8.974000	13.580000	3.604000	17.073000	

	X6	X7	X8	X9	X10	X11	\
count	435.000000	435.000000	435.000000	435.000000	435.000000	435.000000	
mean	2.261772	7.589501	1.837722	5.721492	11.664763	4.335352	
std	1.416131	2.718641	1.137747	3.318045	3.115375	4.226037	
min	0.000000	1.370000	0.000000	0.000000	4.167000	0.000000	
25%	1.211000	5.540500	0.944500	3.187500	9.515000	0.962500	
50%	2.108000	7.524000	1.754000	4.895000	11.707000	2.336000	
75%	3.096000	9.489000	2.627000	8.082000	13.824000	7.735500	
max	7.143000	15.714000	5.208000	16.129000	20.118000	17.252000	

	X12	X13	X14	X15	X16	X17	\
count	435.000000	435.000000	435.000000	435.000000	435.000000	435.000000	
mean	1.955257	4.041559	4.065984	5.938885	5.468607	1.227453	

std	1.245118	1.811097	1.622816	1.950585	1.812598	0.993868
min	0.000000	0.000000	0.000000	1.449000	0.901000	0.000000
25%	0.979000	2.821000	2.939500	4.636000	4.183000	0.532500
50%	1.633000	3.968000	3.955000	5.729000	5.405000	1.042000
75%	2.586000	5.098000	5.057000	6.965000	6.571000	1.716500
max	6.024000	9.732000	8.571000	12.069000	11.364000	4.615000

	X18	X19
count	435.000000	435.000000
mean	3.328064	8.740818
std	1.500531	3.298998
min	0.000000	0.671000
25%	2.361000	6.166500
50%	3.235000	8.475000
75%	4.184000	11.100000
max	7.692000	17.718000

no_efectores

Composición de aminoácidos (AAC) no_efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	\
0	12.727	6.182	0.727	9.455	0.000	6.909	2.182	9.091	1.091	
2	16.906	9.729	0.797	5.263	0.000	4.306	1.595	7.656	1.754	
3	11.282	13.333	0.513	11.282	0.000	8.205	1.026	9.231	2.051	
4	8.929	11.012	1.488	2.976	0.595	3.571	1.190	7.143	0.595	
5	6.826	4.096	5.802	7.509	0.341	7.509	2.048	4.778	2.389	
..	
495	9.906	3.774	0.943	1.887	0.000	2.358	3.302	9.434	2.830	
496	15.625	7.031	0.781	1.562	0.781	1.562	1.562	11.719	1.562	
497	8.455	5.539	3.499	10.787	0.292	9.038	4.373	9.329	2.624	
498	11.951	2.683	3.902	7.805	0.244	9.756	3.902	11.707	1.463	
499	10.270	5.946	1.622	7.568	1.081	10.811	2.162	8.108	3.784	

	X9	...	X11	X12	X13	X14	X15	X16	X17	X18	\
0	2.909	...	0.727	1.818	1.455	5.818	7.273	9.091	0.000	1.818	
2	1.116	...	0.957	1.754	3.349	4.625	5.423	6.220	0.957	2.711	
3	1.538	...	0.000	0.513	2.051	6.667	4.103	11.795	0.000	0.513	
4	6.250	...	1.190	2.083	3.571	3.571	7.143	3.571	2.679	2.083	
5	8.532	...	8.191	2.730	3.072	1.706	10.239	3.072	0.683	3.754	
..	
495	5.189	...	0.472	3.302	3.302	5.660	5.660	5.660	4.245	3.774	
496	4.688	...	0.000	0.781	3.906	5.469	3.906	6.250	1.562	0.781	
497	5.539	...	3.499	2.332	3.207	3.207	5.831	4.373	0.583	1.749	
498	6.585	...	0.244	1.220	3.902	5.854	7.805	5.610	0.732	3.902	

499 2.703 ... 1.081 1.081 3.784 4.324 1.081 4.324 1.622 5.405

	X19	X20
0	11.273	no_efectores
2	13.078	no_efectores
3	8.718	no_efectores
4	12.798	no_efectores
5	3.754	no_efectores
..
495	15.094	no_efectores
496	15.625	no_efectores
497	6.997	no_efectores
498	5.610	no_efectores
499	10.270	no_efectores

[404 rows x 21 columns]

Composición de aminoácidos (AAC) no_efectores archaea dataset 3, sin valores atípicos.

Estadísticas.

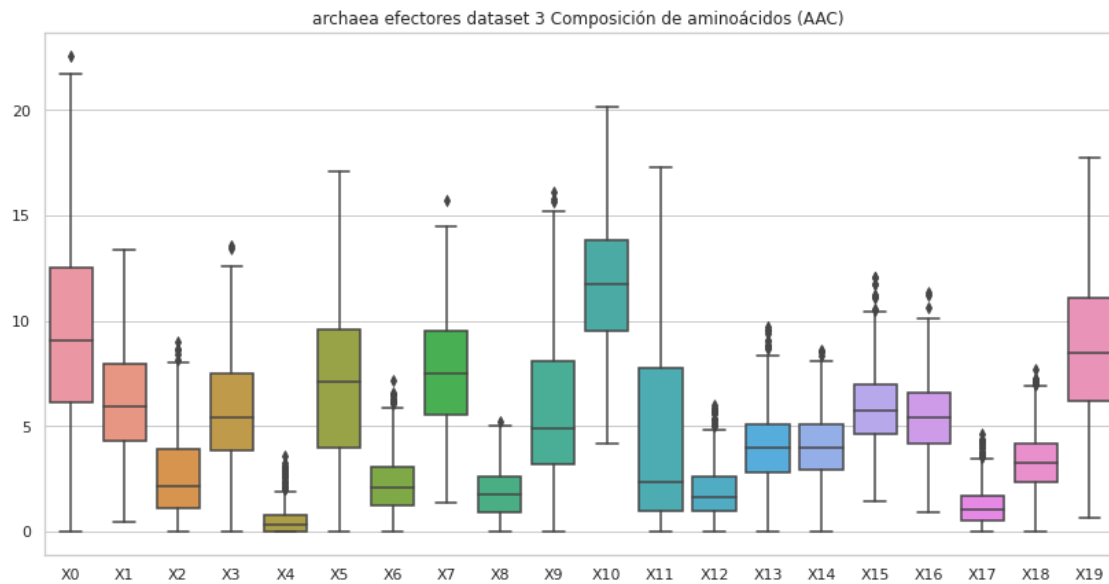
	X0	X1	X2	X3	X4	X5	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	10.133683	6.499500	2.673116	7.925428	0.798141	8.350428	
std	3.510460	2.657614	1.699534	3.060938	0.934261	3.429493	
min	1.639000	0.000000	0.000000	0.000000	0.000000	0.508000	
25%	7.732500	4.698750	1.523000	5.882000	0.000000	6.234000	
50%	10.045000	6.258000	2.416000	8.082500	0.540500	8.417000	
75%	12.491000	8.098250	3.519250	9.955250	1.121750	10.249500	
max	21.384000	15.385000	8.750000	17.308000	6.154000	19.178000	

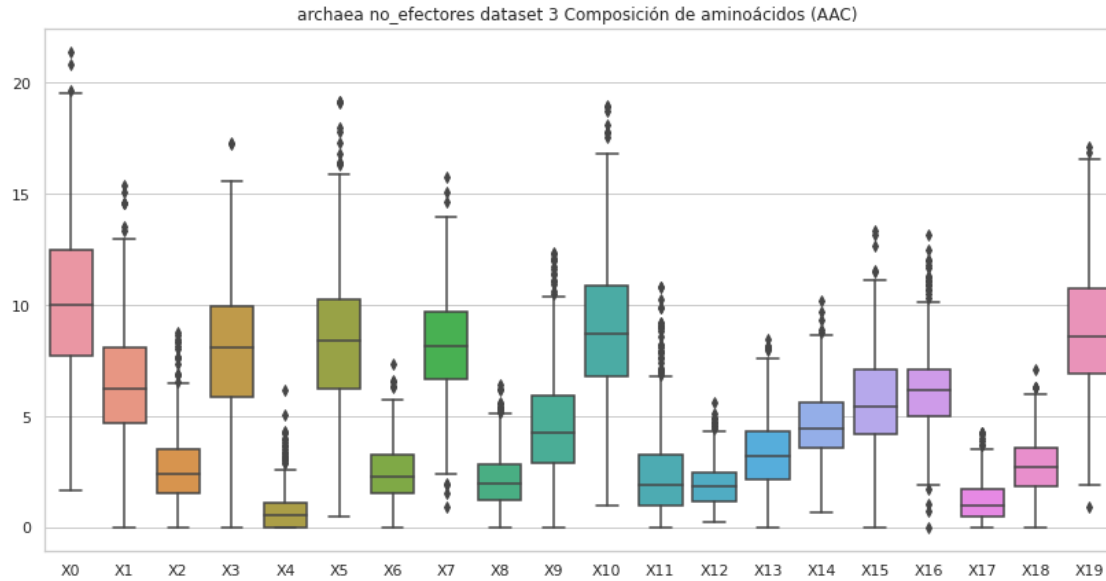
	X6	X7	X8	X9	X10	X11	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	2.446827	8.249540	2.132847	4.666074	8.998653	2.470557	
std	1.326419	2.485961	1.244107	2.536047	3.126522	2.190023	
min	0.000000	0.901000	0.000000	0.000000	1.020000	0.000000	
25%	1.531750	6.667000	1.235500	2.896750	6.829000	0.962000	
50%	2.312000	8.160500	1.983500	4.278500	8.700500	1.928500	
75%	3.307250	9.730500	2.815750	5.945500	10.868500	3.292500	
max	7.353000	15.789000	6.404000	12.368000	18.981000	10.821000	

	X12	X13	X14	X15	X16	X17	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	1.975458	3.348111	4.567564	5.735673	6.192022	1.177948	
std	1.034666	1.668612	1.685052	2.084348	1.994315	0.954898	
min	0.225000	0.000000	0.709000	0.000000	0.000000	0.000000	
25%	1.176000	2.159500	3.568000	4.214750	4.981750	0.473250	

50%	1.826500	3.222000	4.478500	5.418000	6.166500	0.963000
75%	2.481000	4.341250	5.646250	7.092750	7.103250	1.712750
max	5.607000	8.456000	10.194000	13.371000	13.171000	4.255000

	X18	X19
count	404.000000	404.000000
mean	2.778129	8.880369
std	1.356061	2.865315
min	0.000000	0.943000
25%	1.868250	6.898250
50%	2.741500	8.587500
75%	3.583750	10.782500
max	7.123000	17.105000





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) +",\n"
    ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

    #del df['X83']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

```

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

```

efectores

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.021378	0.000000	0.029929	0.042756	0.019240	0.042756	0.010689
1	0.028068	0.002807	0.019648	0.058943	0.014034	0.033682	0.000000
2	0.025200	0.003150	0.015750	0.072451	0.018900	0.031501	0.022050
3	0.033746	0.000000	0.013964	0.008145	0.005818	0.019782	0.005818
4	0.021105	0.014070	0.056279	0.021105	0.021105	0.014070	0.000000
..
495	0.036886	0.000000	0.011226	0.011226	0.011226	0.024056	0.006415
496	0.021883	0.017506	0.030636	0.048142	0.017506	0.030636	0.000000
497	0.041233	0.001718	0.042951	0.053259	0.017180	0.029206	0.012026
498	0.050157	0.000000	0.052337	0.054518	0.032711	0.034892	0.008723
499	0.028460	0.002033	0.036592	0.056920	0.052854	0.063019	0.032526

	X7	X8	X9 ...	X74	X75	X76 \
0	0.049169	0.040618	0.076961 ...	0.030891	0.031710	0.009396
1	0.028068	0.039295	0.047716 ...	-0.003517	0.021302	0.012313
2	0.034651	0.053551	0.059851 ...	-0.029535	-0.017982	-0.000289
3	0.003491	0.002327	0.032582 ...	0.009233	-0.000952	0.012488
4	0.035175	0.042209	0.077384 ...	0.021736	0.044068	0.021114
..
495	0.009622	0.003207	0.030471 ...	0.011786	-0.012539	0.027902
496	0.026259	0.061271	0.065648 ...	0.005103	0.007290	0.030647
497	0.012026	0.001718	0.051541 ...	-0.008408	0.020888	0.000457
498	0.061060	0.067602	0.071964 ...	0.043106	0.017188	-0.005510
499	0.052854	0.024394	0.038624 ...	0.008581	-0.011754	-0.009309

	X77	X78	X79	X80	X81	X82	X83
0	-0.038523	-0.023506	0.004674	0.014311	0.008431	0.007089	efectores
1	0.024241	0.049406	-0.001244	0.018841	0.016739	0.000462	efectores
2	-0.012465	0.015645	0.039493	0.006286	0.044886	-0.009803	efectores
3	0.007820	0.003345	0.017707	0.004287	0.000771	0.029113	efectores
4	0.023565	-0.001891	0.005793	0.063587	0.056119	-0.015616	efectores
..
495	0.029831	-0.004224	0.017067	0.025519	0.003528	0.039420	efectores
496	0.071019	0.078740	-0.034212	-0.041836	-0.019338	-0.000498	efectores

```

497 -0.003679  0.027708  0.019038  0.017227  0.024151  0.017476  efectores
498 -0.021872  0.011002  0.013359  0.001139  0.020720  0.026122  efectores
499 -0.008931 -0.006949 -0.021392 -0.005964 -0.000670 -0.000728  efectores

```

[500 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.032731	0.003622	0.026647	0.034797	0.016676	0.027721
std	0.016583	0.006021	0.020982	0.031488	0.013620	0.013927
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.003555
25%	0.021548	0.000000	0.008993	0.008091	0.007285	0.017770
50%	0.030084	0.000965	0.021816	0.027232	0.013556	0.025040
75%	0.039579	0.004958	0.038199	0.055779	0.021599	0.033700
max	0.168769	0.042541	0.135015	0.337539	0.101006	0.135015

	X6	X7	X8	X9 ...	X73 \
count	500.000000	500.000000	500.000000	500.000000 ...	500.000000
mean	0.008318	0.028686	0.025629	0.047263 ...	0.014153
std	0.008538	0.029638	0.032609	0.032516 ...	0.018435
min	0.000000	0.000000	0.000000	0.005264 ...	-0.098819
25%	0.002474	0.008172	0.002253	0.024743 ...	0.005343
50%	0.006272	0.017730	0.009153	0.040085 ...	0.015409
75%	0.011514	0.039159	0.042783	0.061732 ...	0.023854
max	0.101262	0.265384	0.193044	0.405046 ...	0.094551

	X74	X75	X76	X77	X78	X79 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.005807	0.008951	0.013542	0.004112	0.007734	0.013311
std	0.023269	0.024996	0.017185	0.028947	0.023837	0.021432
min	-0.140793	-0.246148	-0.061740	-0.235348	-0.093121	-0.242156
25%	-0.004098	-0.002070	0.004728	-0.006338	-0.002036	0.004710
50%	0.008644	0.004157	0.015747	0.009501	0.005398	0.014674
75%	0.016728	0.019672	0.024116	0.018168	0.016532	0.024531
max	0.137971	0.129962	0.063801	0.079081	0.151038	0.087782

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.005151	0.008601	0.014754
std	0.027314	0.024880	0.019526
min	-0.185458	-0.176735	-0.078956
25%	-0.003875	-0.001141	0.004032
50%	0.010424	0.004884	0.015724

75%	0.017678	0.020674	0.024578
max	0.100394	0.122977	0.087735

[8 rows x 83 columns]

no_efectores

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.028805	0.000000	0.021398	0.015637	0.003292	0.020575	0.002469
1	0.036565	0.097508	0.024377	0.024377	0.012188	0.048754	0.036565
2	0.031255	0.000000	0.009730	0.007961	0.006192	0.014153	0.003243
3	0.029581	0.000000	0.029581	0.021514	0.005378	0.024203	0.005378
4	0.023182	0.001545	0.007727	0.009273	0.009273	0.018546	0.001545
..
495	0.019160	0.000000	0.003650	0.004562	0.006387	0.018248	0.005474
496	0.026139	0.001307	0.002614	0.002614	0.006535	0.019604	0.002614
497	0.029901	0.001031	0.038149	0.031963	0.011342	0.032994	0.009280
498	0.032365	0.000661	0.021136	0.026421	0.010568	0.031705	0.003963
499	0.051049	0.005374	0.037615	0.053736	0.018808	0.040302	0.018808

	X7	X8	X9	...	X74	X75	X76 \
0	0.006584	0.001646	0.021398	...	0.003023	0.002166	0.037563
1	0.048754	0.024377	0.048754	...	0.043920	-0.018776	0.056074
2	0.002064	0.001769	0.021819	...	-0.001396	-0.003479	0.014636
3	0.004034	0.000000	0.018825	...	-0.011923	0.016642	-0.004218
4	0.016227	0.003091	0.045591	...	-0.003725	-0.001325	0.012610
..
495	0.010036	0.000912	0.025547	...	0.019707	0.005438	0.020553
496	0.007842	0.000000	0.024832	...	0.020913	0.000177	0.025730
497	0.019590	0.012373	0.030932	...	0.021024	0.010763	0.019472
498	0.017834	0.000661	0.013871	...	0.008917	0.005644	0.043505
499	0.013434	0.005374	0.064483	...	0.022434	0.004069	0.020315

	X77	X78	X79	X80	X81	X82	X83
0	0.013964	0.011332	0.038735	-0.000353	-0.005008	0.021565	no_efectores
1	-0.099668	-0.037803	0.048541	-0.022254	0.002988	-0.037594	no_efectores
2	0.004718	0.001055	0.015117	0.008095	0.002599	0.019742	no_efectores
3	-0.020803	0.020905	0.014162	0.005775	0.011591	0.009711	no_efectores
4	0.045154	0.010148	0.020662	0.024527	0.000184	0.017087	no_efectores
..
495	0.013840	0.005404	0.016717	0.011613	0.004589	0.010401	no_efectores
496	0.030932	0.010274	0.017748	0.023375	0.005875	0.022163	no_efectores
497	-0.000733	0.011559	0.014669	-0.020268	0.010403	0.023157	no_efectores

```

498  0.003857 -0.005458  0.037807  0.000621 -0.003338  0.033398  no_efectores
499 -0.002452 -0.037104  0.014303  0.026715 -0.007469  0.002963  no_efectores

```

[500 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.036195	0.005282	0.032115	0.036792	0.014758	0.031046
std	0.021212	0.017395	0.023985	0.035568	0.022812	0.020230
min	0.001724	0.000000	0.000000	0.000000	0.000000	0.002040
25%	0.024561	0.000000	0.017695	0.017879	0.006058	0.020289
50%	0.031589	0.001883	0.029606	0.030646	0.010736	0.027576
75%	0.042523	0.004564	0.041109	0.046093	0.017860	0.035538
max	0.224205	0.292862	0.292862	0.439293	0.439293	0.207300

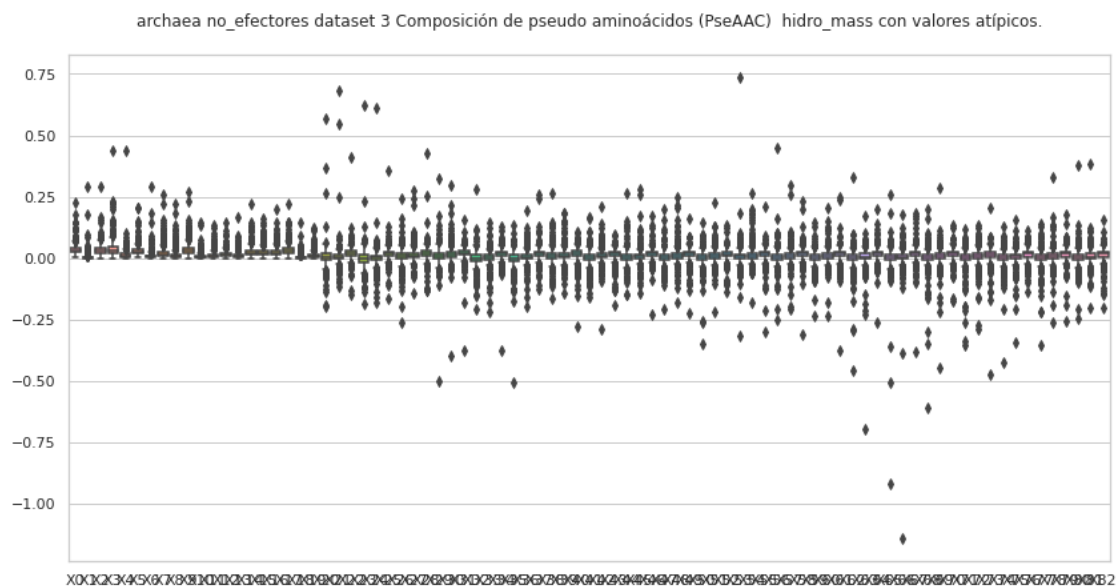
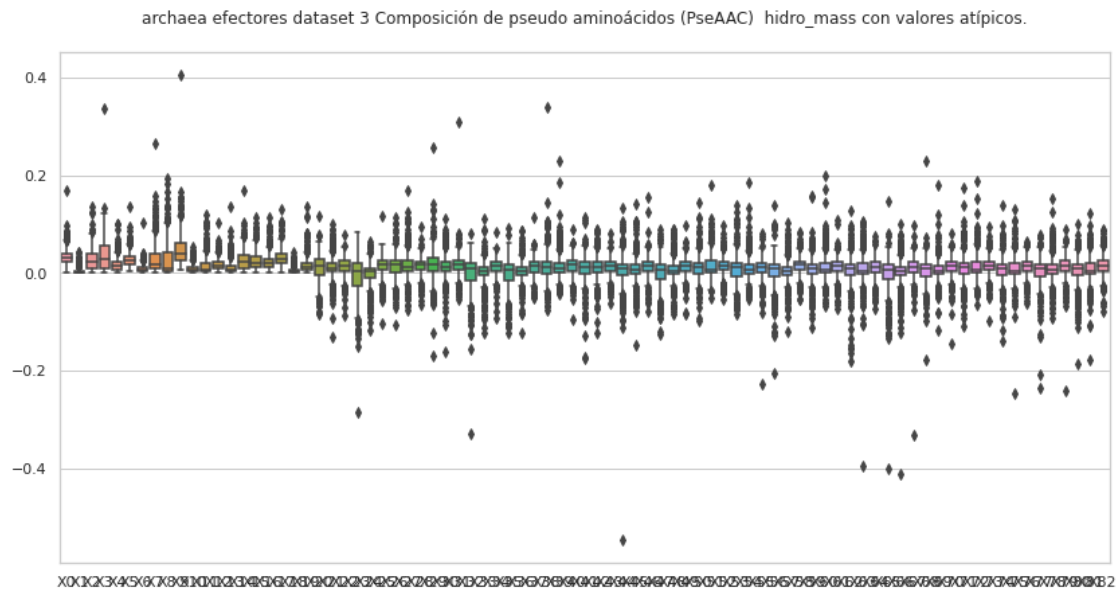
	X6	X7	X8	X9 ...	X73 \
count	500.000000	500.000000	500.000000	500.000000 ...	500.000000
mean	0.010514	0.021484	0.014085	0.035720 ...	0.013938
std	0.019247	0.025058	0.022562	0.026773 ...	0.032912
min	0.000000	0.000000	0.000000	0.000000 ...	-0.472014
25%	0.003144	0.008234	0.002655	0.020719 ...	0.006224
50%	0.006858	0.014825	0.006783	0.029830 ...	0.017018
75%	0.011474	0.025460	0.014813	0.041701 ...	0.027290
max	0.292862	0.259126	0.219646	0.269045 ...	0.204270

	X74	X75	X76	X77	X78	X79 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.000633	0.006554	0.015385	0.001047	0.007809	0.015608
std	0.032953	0.030769	0.022340	0.033250	0.031767	0.026971
min	-0.426181	-0.341878	-0.135159	-0.352072	-0.262199	-0.254713
25%	-0.006591	-0.001360	0.006137	-0.006878	-0.001296	0.006237
50%	0.003592	0.006792	0.017676	0.004451	0.006791	0.017753
75%	0.012905	0.016988	0.027084	0.012346	0.017603	0.026057
max	0.130589	0.153117	0.113964	0.138357	0.328266	0.176277

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.002131	0.008286	0.014102
std	0.033750	0.030654	0.026732
min	-0.247907	-0.199923	-0.199986
25%	-0.008271	-0.002257	0.006288
50%	0.004255	0.005935	0.017867
75%	0.013575	0.018749	0.026252

max 0.380875 0.382561 0.154920

[8 rows x 83 columns]



3.1 Composición de pseudo aminoácidos (PseAAC) hidro_mass, sin valores atípicos

```
[6]: #hidro_mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "hidro_mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +
      ↪ '_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

    del df['X83']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de
    ↪ sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X83'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
```

```
ax = sns.boxplot(data=df)
ax.set_title(organismo + ' ' + str(etiq) + " dataset " + str(dataset) + "
↳ " + str(transf) + " " + str(comp))
```

efectores

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.021378	0.000000	0.029929	0.042756	0.019240	0.042756	0.010689
1	0.028068	0.002807	0.019648	0.058943	0.014034	0.033682	0.000000
2	0.025200	0.003150	0.015750	0.072451	0.018900	0.031501	0.022050
3	0.033746	0.000000	0.013964	0.008145	0.005818	0.019782	0.005818
5	0.033039	0.009440	0.029892	0.020453	0.018879	0.036185	0.014159
..	
495	0.036886	0.000000	0.011226	0.011226	0.011226	0.024056	0.006415
496	0.021883	0.017506	0.030636	0.048142	0.017506	0.030636	0.000000
497	0.041233	0.001718	0.042951	0.053259	0.017180	0.029206	0.012026
498	0.050157	0.000000	0.052337	0.054518	0.032711	0.034892	0.008723
499	0.028460	0.002033	0.036592	0.056920	0.052854	0.063019	0.032526

	X7	X8	X9 ...	X74	X75	X76 \
0	0.049169	0.040618	0.076961 ...	0.030891	0.031710	0.009396
1	0.028068	0.039295	0.047716 ...	-0.003517	0.021302	0.012313
2	0.034651	0.053551	0.059851 ...	-0.029535	-0.017982	-0.000289
3	0.003491	0.002327	0.032582 ...	0.009233	-0.000952	0.012488
5	0.018879	0.012586	0.050345 ...	0.009158	0.007607	0.025471
..	
495	0.009622	0.003207	0.030471 ...	0.011786	-0.012539	0.027902
496	0.026259	0.061271	0.065648 ...	0.005103	0.007290	0.030647
497	0.012026	0.001718	0.051541 ...	-0.008408	0.020888	0.000457
498	0.061060	0.067602	0.071964 ...	0.043106	0.017188	-0.005510
499	0.052854	0.024394	0.038624 ...	0.008581	-0.011754	-0.009309

	X77	X78	X79	X80	X81	X82	X83
0	-0.038523	-0.023506	0.004674	0.014311	0.008431	0.007089	efectores
1	0.024241	0.049406	-0.001244	0.018841	0.016739	0.000462	efectores
2	-0.012465	0.015645	0.039493	0.006286	0.044886	-0.009803	efectores
3	0.007820	0.003345	0.017707	0.004287	0.000771	0.029113	efectores
5	0.001431	0.007965	0.027851	0.017537	0.009421	0.047842	efectores
..	
495	0.029831	-0.004224	0.017067	0.025519	0.003528	0.039420	efectores
496	0.071019	0.078740	-0.034212	-0.041836	-0.019338	-0.000498	efectores
497	-0.003679	0.027708	0.019038	0.017227	0.024151	0.017476	efectores
498	-0.021872	0.011002	0.013359	0.001139	0.020720	0.026122	efectores
499	-0.008931	-0.006949	-0.021392	-0.005964	-0.000670	-0.000728	efectores

[389 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5 \
count	389.000000	389.000000	389.000000	389.000000	389.000000	389.000000
mean	0.029693	0.002096	0.020686	0.024673	0.013728	0.024533
std	0.011945	0.003656	0.015376	0.021539	0.009473	0.010127
min	0.004203	0.000000	0.000522	0.000000	0.000000	0.003555
25%	0.021007	0.000000	0.007905	0.006172	0.006635	0.016595
50%	0.027641	0.000386	0.016236	0.016332	0.011497	0.022616
75%	0.036023	0.002668	0.030876	0.043140	0.018879	0.030221
max	0.077638	0.020296	0.081951	0.088428	0.055941	0.063019

	X6	X7	X8	X9 ...	X73 \
count	389.000000	389.000000	389.000000	389.000000	389.000000
mean	0.006559	0.020215	0.015860	0.037057	0.015431
std	0.005526	0.018755	0.021637	0.017789	0.014246
min	0.000000	0.000000	0.000000	0.005264	-0.024111
25%	0.002289	0.006967	0.001592	0.022536	0.007313
50%	0.005494	0.013338	0.005390	0.035220	0.016383
75%	0.009637	0.028013	0.022162	0.046569	0.023819
max	0.032526	0.093543	0.106121	0.106202	0.066993

	X74	X75	X76	X77	X78	X79 \
count	389.000000	389.000000	389.000000	389.000000	389.000000	389.000000
mean	0.006785	0.006991	0.015480	0.008203	0.007941	0.015560
std	0.015618	0.015343	0.012535	0.017858	0.015972	0.013677
min	-0.048487	-0.040268	-0.022168	-0.059300	-0.049904	-0.034212
25%	-0.001914	-0.001307	0.007620	-0.000600	-0.000433	0.007281
50%	0.009369	0.003088	0.016111	0.010633	0.005408	0.015499
75%	0.015908	0.012800	0.024098	0.018155	0.015016	0.024320
max	0.073440	0.079413	0.048468	0.072137	0.078740	0.068802

	X80	X81	X82
count	389.000000	389.000000	389.000000
mean	0.008566	0.008217	0.015427
std	0.018451	0.016130	0.014906
min	-0.071591	-0.054948	-0.038553
25%	0.000397	-0.000670	0.006707
50%	0.011563	0.004169	0.016626
75%	0.017598	0.015733	0.024410
max	0.066696	0.080542	0.072174

[8 rows x 83 columns]

no_efectores

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.028805	0.000000	0.021398	0.015637	0.003292	0.020575	0.002469
2	0.031255	0.000000	0.009730	0.007961	0.006192	0.014153	0.003243
3	0.029581	0.000000	0.029581	0.021514	0.005378	0.024203	0.005378
4	0.023182	0.001545	0.007727	0.009273	0.009273	0.018546	0.001545
5	0.037359	0.001868	0.041095	0.041095	0.016812	0.026152	0.013076
..
495	0.019160	0.000000	0.003650	0.004562	0.006387	0.018248	0.005474
496	0.026139	0.001307	0.002614	0.002614	0.006535	0.019604	0.002614
497	0.029901	0.001031	0.038149	0.031963	0.011342	0.032994	0.009280
498	0.032365	0.000661	0.021136	0.026421	0.010568	0.031705	0.003963
499	0.051049	0.005374	0.037615	0.053736	0.018808	0.040302	0.018808
	X7	X8	X9 ...	X74	X75	X76 \	
0	0.006584	0.001646	0.021398	...	0.003023	0.002166	0.037563
2	0.002064	0.001769	0.021819	...	-0.001396	-0.003479	0.014636
3	0.004034	0.000000	0.018825	...	-0.011923	0.016642	-0.004218
4	0.016227	0.003091	0.045591	...	-0.003725	-0.001325	0.012610
5	0.046699	0.044831	0.070983	...	0.006257	0.032799	-0.014738
..
495	0.010036	0.000912	0.025547	...	0.019707	0.005438	0.020553
496	0.007842	0.000000	0.024832	...	0.020913	0.000177	0.025730
497	0.019590	0.012373	0.030932	...	0.021024	0.010763	0.019472
498	0.017834	0.000661	0.013871	...	0.008917	0.005644	0.043505
499	0.013434	0.005374	0.064483	...	0.022434	0.004069	0.020315
	X77	X78	X79	X80	X81	X82	X83
0	0.013964	0.011332	0.038735	-0.000353	-0.005008	0.021565	no_efectores
2	0.004718	0.001055	0.015117	0.008095	0.002599	0.019742	no_efectores
3	-0.020803	0.020905	0.014162	0.005775	0.011591	0.009711	no_efectores
4	0.045154	0.010148	0.020662	0.024527	0.000184	0.017087	no_efectores
5	-0.009334	0.003795	0.024556	-0.020546	0.010891	0.015530	no_efectores
..
495	0.013840	0.005404	0.016717	0.011613	0.004589	0.010401	no_efectores
496	0.030932	0.010274	0.017748	0.023375	0.005875	0.022163	no_efectores
497	-0.000733	0.011559	0.014669	-0.020268	0.010403	0.023157	no_efectores
498	0.003857	-0.005458	0.037807	0.000621	-0.003338	0.033398	no_efectores
499	-0.002452	-0.037104	0.014303	0.026715	-0.007469	0.002963	no_efectores

[434 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	434.000000	434.000000	434.000000	434.000000	434.000000	434.000000
mean	0.032671	0.003223	0.028196	0.029389	0.011652	0.027187
std	0.012606	0.005376	0.015911	0.018731	0.008165	0.010731
min	0.001724	0.000000	0.000000	0.000000	0.000000	0.002040
25%	0.024134	0.000000	0.016055	0.015458	0.005904	0.019866
50%	0.030988	0.001816	0.027729	0.027752	0.009823	0.026664
75%	0.039867	0.003820	0.038300	0.040822	0.015810	0.032974
max	0.079628	0.054225	0.079316	0.087771	0.062319	0.071751

	X6	X7	X8	X9 ...	X73 \
count	434.000000	434.000000	434.000000	434.000000 ...	434.000000
mean	0.007403	0.016323	0.009538	0.029950 ...	0.017177
std	0.006346	0.012656	0.012075	0.014337 ...	0.013852
min	0.000000	0.000000	0.000000	0.000000 ...	-0.034332
25%	0.002858	0.007503	0.002361	0.019781 ...	0.009114
50%	0.006127	0.013535	0.005793	0.028016 ...	0.017507
75%	0.010343	0.021408	0.011676	0.037582 ...	0.027276
max	0.045125	0.078167	0.080030	0.080080 ...	0.055410

	X74	X75	X76	X77	X78	X79 \
count	434.000000	434.000000	434.000000	434.000000	434.000000	434.000000
mean	0.004034	0.009059	0.017188	0.002915	0.008132	0.017044
std	0.016202	0.014708	0.014712	0.016317	0.014072	0.014343
min	-0.059203	-0.058250	-0.044050	-0.082405	-0.042332	-0.024314
25%	-0.005090	0.000090	0.008195	-0.004950	-0.000718	0.009086
50%	0.004707	0.007012	0.018463	0.004602	0.006648	0.018108
75%	0.012707	0.016452	0.026839	0.011834	0.016349	0.025804
max	0.069862	0.058535	0.066337	0.082180	0.062366	0.073328

	X80	X81	X82
count	434.000000	434.000000	434.000000
mean	0.003710	0.008248	0.017247
std	0.018925	0.017055	0.014585
min	-0.067094	-0.049326	-0.044177
25%	-0.006188	-0.001920	0.009740
50%	0.004978	0.006021	0.018790
75%	0.013383	0.016983	0.026390
max	0.083989	0.070252	0.065072

[8 rows x 83 columns]


```

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 3,
con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.028365	0.000000	0.039710	0.056729	0.025528	0.056729	0.014182
1	0.048566	0.004857	0.033996	0.101988	0.024283	0.058279	0.000000
2	0.035954	0.004494	0.022471	0.103367	0.026965	0.044942	0.031459
3	0.044550	0.000000	0.018435	0.010754	0.007681	0.026116	0.007681
4	0.032878	0.021919	0.087676	0.032878	0.032878	0.021919	0.000000
..
495	0.058034	0.000000	0.017663	0.017663	0.017663	0.037848	0.010093
496	0.031065	0.024852	0.043491	0.068344	0.024852	0.043491	0.000000
497	0.056822	0.002368	0.059190	0.073396	0.023676	0.040249	0.016573
498	0.054515	0.000000	0.056886	0.059256	0.035554	0.037924	0.009481
499	0.030767	0.002198	0.039558	0.061534	0.057139	0.068127	0.035163

	X7	X8	X9 ...	X32	X33	X34 \
0	0.065238	0.053893	0.102112 ...	0.017060	0.000578	0.023698

1	0.048566	0.067992	0.082562	...	-0.001135	0.029328	0.059494
2	0.049436	0.076401	0.085390	...	0.030143	-0.014128	0.026419
3	0.004609	0.003072	0.043014	...	0.030328	0.012445	0.027455
4	0.054797	0.065757	0.120554	...	0.031265	0.020577	-0.049017
..
495	0.015139	0.005046	0.047941	...	0.023294	0.017113	0.027645
496	0.037278	0.086983	0.093196	...	-0.040239	0.014119	-0.006884
497	0.016573	0.002368	0.071028	...	0.031730	-0.013815	0.019880
498	0.066367	0.073477	0.078218	...	-0.010496	-0.006208	-0.014621
499	0.057139	0.026372	0.041756	...	0.046370	0.016607	0.035145

	X35	X36	X37	X38	X39	X40	X41
0	0.010700	-0.012610	-0.001292	0.012467	0.006201	0.009406	efectores
1	-0.024098	0.035541	-0.031706	0.021306	-0.002152	0.000800	efectores
2	-0.012108	0.036806	-0.012355	-0.000412	0.056345	-0.013986	efectores
3	0.022504	0.034217	0.042594	0.016486	0.023376	0.038434	efectores
4	0.020263	0.038868	0.013455	0.032892	0.009024	-0.024328	efectores
..
495	0.035289	0.004501	0.058854	0.043899	0.026852	0.062022	efectores
496	-0.004353	0.011871	0.019981	0.043508	-0.048568	-0.000707	efectores
497	0.020295	0.020882	0.032749	0.000630	0.026236	0.024083	efectores
498	-0.008238	0.032307	-0.023013	-0.005989	0.014520	0.028392	efectores
499	0.017869	-0.022622	0.032106	-0.010064	-0.023126	-0.000787	efectores

[500 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 3, con valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.046522	0.004745	0.035770	0.046454	0.023906	0.038964
std	0.016650	0.007637	0.024158	0.037394	0.017227	0.012926
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.012156
25%	0.036082	0.000000	0.016885	0.014847	0.012322	0.030311
50%	0.045027	0.001532	0.030390	0.037045	0.019757	0.037731
75%	0.054843	0.006207	0.048405	0.070967	0.031533	0.045569
max	0.114148	0.049706	0.138996	0.228296	0.108486	0.112076

	X6	X7	X8	X9	...	X31 \
count	500.000000	500.000000	500.000000	500.000000	...	500.000000
mean	0.011349	0.039486	0.033956	0.066911	...	0.014469
std	0.009580	0.034460	0.039509	0.033925	...	0.025239
min	0.000000	0.000000	0.000000	0.005320	...	-0.153723
25%	0.003977	0.013550	0.003641	0.043333	...	0.001558
50%	0.009432	0.028405	0.014281	0.060002	...	0.019279

75%	0.016153	0.056588	0.057651	0.086216	...	0.032590
max	0.068489	0.245316	0.186144	0.273955	...	0.092772

	X32	X33	X34	X35	X36	X37 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.021590	0.018510	0.017355	0.015860	0.018465	0.019534
std	0.024259	0.036725	0.024842	0.027162	0.028588	0.026900
min	-0.069306	-0.104445	-0.073356	-0.225362	-0.300764	-0.254841
25%	0.008480	0.005922	0.003219	0.004312	0.006037	0.009009
50%	0.025637	0.021358	0.021380	0.020239	0.025304	0.023258
75%	0.037455	0.033447	0.033765	0.032822	0.035593	0.035701
max	0.104675	0.652221	0.085517	0.113945	0.076173	0.091772

	X38	X39	X40
count	500.000000	500.000000	500.000000
mean	0.020190	0.020150	0.022255
std	0.023874	0.025609	0.038374
min	-0.119362	-0.163783	-0.081095
25%	0.006775	0.008287	0.006078
50%	0.024839	0.023096	0.024648
75%	0.035214	0.035027	0.036097
max	0.080617	0.086622	0.656568

[8 rows x 41 columns]

no_efectores

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.033975	0.000000	0.025239	0.018444	0.003883	0.024268	0.002912
1	0.034552	0.092140	0.023035	0.023035	0.011517	0.046070	0.034552
2	0.049191	0.000000	0.015314	0.012530	0.009745	0.022275	0.005105
3	0.046155	0.000000	0.046155	0.033568	0.008392	0.037764	0.008392
4	0.042611	0.002841	0.014204	0.017044	0.017044	0.034089	0.002841
..
495	0.040255	0.000000	0.007668	0.009584	0.013418	0.038338	0.011501
496	0.037742	0.001887	0.003774	0.003774	0.009436	0.028307	0.003774
497	0.038273	0.001320	0.048831	0.040913	0.014517	0.042232	0.011878
498	0.033978	0.000693	0.022189	0.027737	0.011095	0.033284	0.004161
499	0.064111	0.006749	0.047240	0.067486	0.023620	0.050614	0.023620

	X7	X8	X9	...	X32	X33	X34 \
0	0.007766	0.001941	0.025239	...	0.031453	0.032947	0.044014
1	0.046070	0.023035	0.046070	...	0.080080	0.007990	0.034117

2	0.003248	0.002784	0.034341	...	0.036772	0.044725	0.033724
3	0.006294	0.000000	0.029372	...	0.020173	0.034002	0.024562
4	0.029828	0.005681	0.083802	...	0.013775	0.019952	0.010516
..
495	0.021086	0.001917	0.053673	...	0.006279	0.016688	0.036326
496	0.011323	0.000000	0.035855	...	0.032654	0.047708	0.034922
497	0.025075	0.015837	0.039593	...	0.030940	0.030826	0.018945
498	0.018722	0.000693	0.014562	...	0.034248	0.035489	0.025459
499	0.016871	0.006749	0.080983	...	0.033337	0.027406	0.036148

	X35	X36	X37	X38	X39	X40	X41
0	0.036630	0.039320	0.024901	0.044305	0.045688	0.025437	no_efectores
1	0.017184	0.053710	-0.017651	0.052987	0.045869	-0.035524	no_efectores
2	0.031979	0.019132	0.033635	0.023035	0.023792	0.031071	no_efectores
3	0.050282	0.021645	0.026889	-0.006582	0.022096	0.015151	no_efectores
4	0.028435	0.050716	0.013637	0.023178	0.037979	0.031408	no_efectores
..
495	0.022667	0.023278	0.031876	0.043181	0.035122	0.021851	no_efectores
496	0.030192	0.050296	0.037511	0.037151	0.025627	0.032002	no_efectores
497	0.032895	0.026452	0.020637	0.024924	0.018776	0.029641	no_efectores
498	0.031656	0.036361	0.030804	0.045672	0.039691	0.035062	no_efectores
499	0.031190	0.027419	0.039987	0.025513	0.017962	0.003721	no_efectores

[500 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 3, con valores atípicos.
Estadísticas.

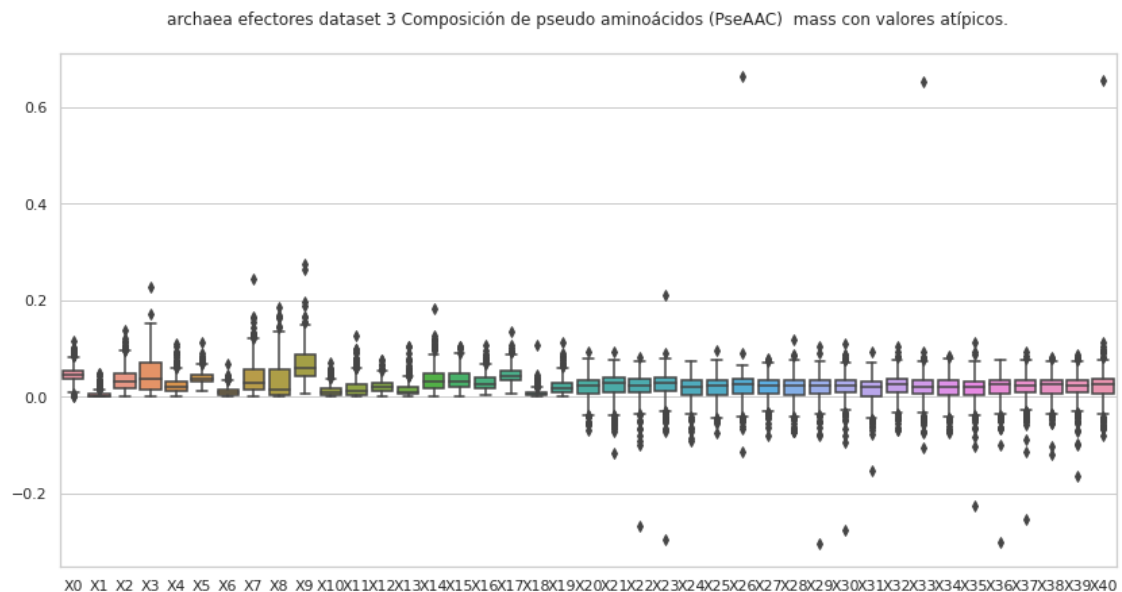
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.048476	0.006104	0.042790	0.049387	0.018795	0.040507	
std	0.023212	0.013677	0.028151	0.041203	0.015446	0.017415	
min	0.005899	0.000000	0.000000	0.000000	0.000000	0.006221	
25%	0.035943	0.000000	0.025146	0.022598	0.008414	0.030900	
50%	0.044270	0.002573	0.038068	0.039305	0.014884	0.037227	
75%	0.056151	0.005789	0.055639	0.063619	0.024200	0.046152	
max	0.265705	0.134999	0.232415	0.342074	0.141296	0.159423	

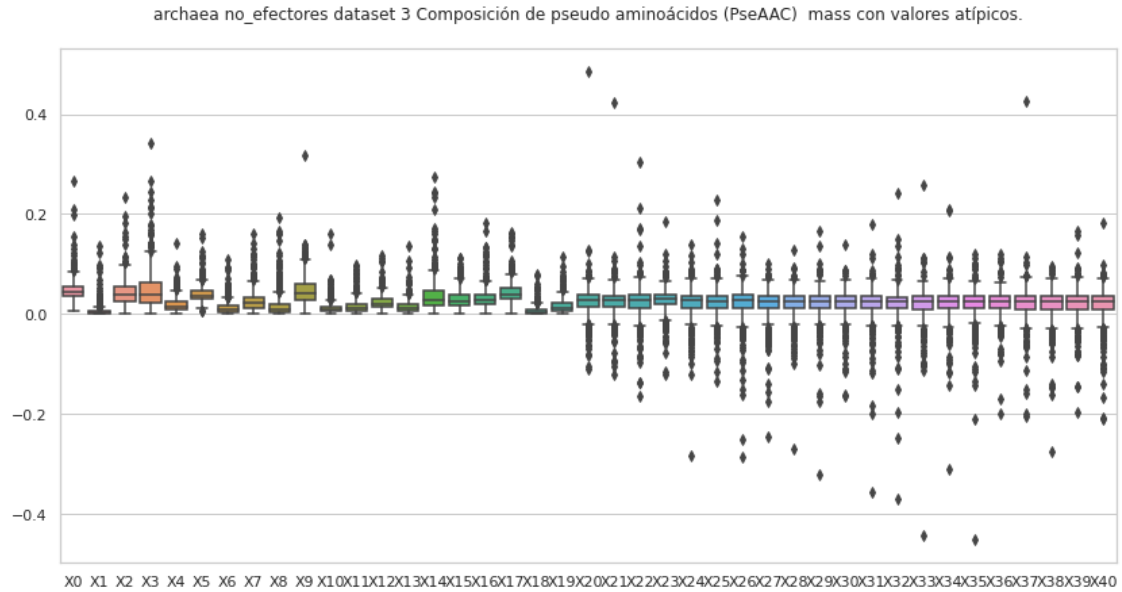
	X6	X7	X8	X9	...	X31	\
count	500.000000	500.000000	500.000000	500.000000	...	500.000000	
mean	0.012561	0.027296	0.018270	0.048017	...	0.020520	
std	0.013446	0.023305	0.025839	0.028733	...	0.033935	
min	0.000000	0.000000	0.000000	0.000000	...	-0.355346	
25%	0.004430	0.011718	0.003864	0.028400	...	0.012334	
50%	0.009058	0.021603	0.009176	0.041794	...	0.026000	
75%	0.016168	0.034271	0.020248	0.061493	...	0.035966	

max	0.110427	0.159423	0.191819	0.318846	...	0.179857
	X32	X33	X34	X35	X36	X37 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.020457	0.020311	0.021483	0.020828	0.021475	0.020147
std	0.036690	0.035218	0.033263	0.036333	0.027368	0.035554
min	-0.370492	-0.442716	-0.310994	-0.450821	-0.199475	-0.203303
25%	0.011252	0.010421	0.010615	0.012768	0.012064	0.009438
50%	0.025158	0.024843	0.026495	0.025645	0.025408	0.024505
75%	0.033979	0.035455	0.036439	0.035825	0.035578	0.035093
max	0.241549	0.259241	0.208260	0.119210	0.119811	0.425374

	X38	X39	X40
count	500.000000	500.000000	500.000000
mean	0.018902	0.020732	0.019170
std	0.031960	0.029996	0.033210
min	-0.273794	-0.197193	-0.208786
25%	0.009363	0.009730	0.009625
50%	0.024578	0.024604	0.025407
75%	0.036145	0.035877	0.034949
max	0.095469	0.165132	0.183596

[8 rows x 41 columns]





4.1 Composición de pseudo aminoácidos (PseAAC) mass, sin valores atípicos

```
[8]: #mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +
      ' ' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

del df['X41']
df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
df['X41'] = etiq
```

```

df_out = pd.concat([df_out,df])

#Guarda la lista csv sin valores atípicos.
df_out.to_csv(str(out), index=False, header=False)

print (str(titulo) + "Valores del documento csv.\n")
print (df)
print ("\n\n" + str(titulo) + "Estadísticas.\n")
print(df.describe())
print ("\n\n")

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

```

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 3,
sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.028365	0.000000	0.039710	0.056729	0.025528	0.056729	0.014182
1	0.048566	0.004857	0.033996	0.101988	0.024283	0.058279	0.000000
2	0.035954	0.004494	0.022471	0.103367	0.026965	0.044942	0.031459
3	0.044550	0.000000	0.018435	0.010754	0.007681	0.026116	0.007681
4	0.032878	0.021919	0.087676	0.032878	0.032878	0.021919	0.000000
..
495	0.058034	0.000000	0.017663	0.017663	0.017663	0.037848	0.010093
496	0.031065	0.024852	0.043491	0.068344	0.024852	0.043491	0.000000
497	0.056822	0.002368	0.059190	0.073396	0.023676	0.040249	0.016573
498	0.054515	0.000000	0.056886	0.059256	0.035554	0.037924	0.009481
499	0.030767	0.002198	0.039558	0.061534	0.057139	0.068127	0.035163

	X7	X8	X9 ...	X32	X33	X34 \
0	0.065238	0.053893	0.102112 ...	0.017060	0.000578	0.023698
1	0.048566	0.067992	0.082562 ...	-0.001135	0.029328	0.059494
2	0.049436	0.076401	0.085390 ...	0.030143	-0.014128	0.026419
3	0.004609	0.003072	0.043014 ...	0.030328	0.012445	0.027455
4	0.054797	0.065757	0.120554 ...	0.031265	0.020577	-0.049017
..
495	0.015139	0.005046	0.047941 ...	0.023294	0.017113	0.027645
496	0.037278	0.086983	0.093196 ...	-0.040239	0.014119	-0.006884
497	0.016573	0.002368	0.071028 ...	0.031730	-0.013815	0.019880
498	0.066367	0.073477	0.078218 ...	-0.010496	-0.006208	-0.014621
499	0.057139	0.026372	0.041756 ...	0.046370	0.016607	0.035145

	X35	X36	X37	X38	X39	X40	X41
0	0.010700	-0.012610	-0.001292	0.012467	0.006201	0.009406	efectores
1	-0.024098	0.035541	-0.031706	0.021306	-0.002152	0.000800	efectores
2	-0.012108	0.036806	-0.012355	-0.000412	0.056345	-0.013986	efectores
3	0.022504	0.034217	0.042594	0.016486	0.023376	0.038434	efectores
4	0.020263	0.038868	0.013455	0.032892	0.009024	-0.024328	efectores
..	
495	0.035289	0.004501	0.058854	0.043899	0.026852	0.062022	efectores
496	-0.004353	0.011871	0.019981	0.043508	-0.048568	-0.000707	efectores
497	0.020295	0.020882	0.032749	0.000630	0.026236	0.024083	efectores
498	-0.008238	0.032307	-0.023013	-0.005989	0.014520	0.028392	efectores
499	0.017869	-0.022622	0.032106	-0.010064	-0.023126	-0.000787	efectores

[411 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	411.000000	411.000000	411.000000	411.000000	411.000000	411.000000	
mean	0.044948	0.003258	0.031329	0.038933	0.021519	0.037370	
std	0.014048	0.005091	0.020330	0.031126	0.014213	0.010940	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.012156	
25%	0.036249	0.000000	0.015332	0.013712	0.011539	0.029429	
50%	0.043867	0.000885	0.026613	0.029231	0.018294	0.036635	
75%	0.051944	0.004703	0.043386	0.060541	0.027736	0.044004	
max	0.089513	0.027417	0.102115	0.143823	0.071837	0.069631	

	X6	X7	X8	X9	...	X31	\
count	411.000000	411.000000	411.000000	411.000000	...	411.000000	
mean	0.009749	0.032327	0.026425	0.059168	...	0.019142	
std	0.007890	0.027504	0.032723	0.026388	...	0.019985	
min	0.000000	0.000000	0.000000	0.005320	...	-0.055617	
25%	0.003610	0.011806	0.002809	0.040650	...	0.006481	
50%	0.007829	0.023485	0.010068	0.054599	...	0.022695	
75%	0.014102	0.044942	0.042840	0.077069	...	0.033384	
max	0.038356	0.133225	0.132740	0.146360	...	0.069642	

	X32	X33	X34	X35	X36	X37	\
count	411.000000	411.000000	411.000000	411.000000	411.000000	411.000000	
mean	0.024012	0.018762	0.020513	0.020158	0.022888	0.023076	
std	0.020118	0.020165	0.020700	0.020413	0.020979	0.021090	
min	-0.045458	-0.072242	-0.052089	-0.047142	-0.050612	-0.051776	
25%	0.011925	0.008238	0.009829	0.009353	0.012642	0.012128	
50%	0.027823	0.021746	0.024609	0.023683	0.027192	0.025630	
75%	0.037559	0.033832	0.034547	0.033350	0.036837	0.036249	

max	0.084997	0.092703	0.067406	0.093812	0.074521	0.091772
-----	----------	----------	----------	----------	----------	----------

	X38	X39	X40
count	411.000000	411.000000	411.000000
mean	0.023454	0.023069	0.024061
std	0.019583	0.019415	0.021597
min	-0.037704	-0.054119	-0.053006
25%	0.011396	0.012582	0.011338
50%	0.026957	0.024772	0.026431
75%	0.036520	0.035080	0.036938
max	0.080617	0.081605	0.098670

[8 rows x 41 columns]

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 3,
sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.033975	0.000000	0.025239	0.018444	0.003883	0.024268	0.002912
2	0.049191	0.000000	0.015314	0.012530	0.009745	0.022275	0.005105
3	0.046155	0.000000	0.046155	0.033568	0.008392	0.037764	0.008392
4	0.042611	0.002841	0.014204	0.017044	0.017044	0.034089	0.002841
5	0.044154	0.002208	0.048569	0.048569	0.019869	0.030908	0.015454
..
495	0.040255	0.000000	0.007668	0.009584	0.013418	0.038338	0.011501
496	0.037742	0.001887	0.003774	0.003774	0.009436	0.028307	0.003774
497	0.038273	0.001320	0.048831	0.040913	0.014517	0.042232	0.011878
498	0.033978	0.000693	0.022189	0.027737	0.011095	0.033284	0.004161
499	0.064111	0.006749	0.047240	0.067486	0.023620	0.050614	0.023620

	X7	X8	X9	...	X32	X33	X34 \
0	0.007766	0.001941	0.025239	...	0.031453	0.032947	0.044014
2	0.003248	0.002784	0.034341	...	0.036772	0.044725	0.033724
3	0.006294	0.000000	0.029372	...	0.020173	0.034002	0.024562
4	0.029828	0.005681	0.083802	...	0.013775	0.019952	0.010516
5	0.055192	0.052985	0.083893	...	0.016671	0.020693	0.031757
..
495	0.021086	0.001917	0.053673	...	0.006279	0.016688	0.036326
496	0.011323	0.000000	0.035855	...	0.032654	0.047708	0.034922
497	0.025075	0.015837	0.039593	...	0.030940	0.030826	0.018945
498	0.018722	0.000693	0.014562	...	0.034248	0.035489	0.025459
499	0.016871	0.006749	0.080983	...	0.033337	0.027406	0.036148

	X35	X36	X37	X38	X39	X40	X41
0	0.036630	0.039320	0.024901	0.044305	0.045688	0.025437	no_efectores

2	0.031979	0.019132	0.033635	0.023035	0.023792	0.031071	no_efectores
3	0.050282	0.021645	0.026889	-0.006582	0.022096	0.015151	no_efectores
4	0.028435	0.050716	0.013637	0.023178	0.037979	0.031408	no_efectores
5	0.020252	0.003711	0.018026	-0.017418	0.029022	0.018354	no_efectores
..	
495	0.022667	0.023278	0.031876	0.043181	0.035122	0.021851	no_efectores
496	0.030192	0.050296	0.037511	0.037151	0.025627	0.032002	no_efectores
497	0.032895	0.026452	0.020637	0.024924	0.018776	0.029641	no_efectores
498	0.031656	0.036361	0.030804	0.045672	0.039691	0.035062	no_efectores
499	0.031190	0.027419	0.039987	0.025513	0.017962	0.003721	no_efectores

[426 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 3,
sin valores atípicos.

Estadísticas.

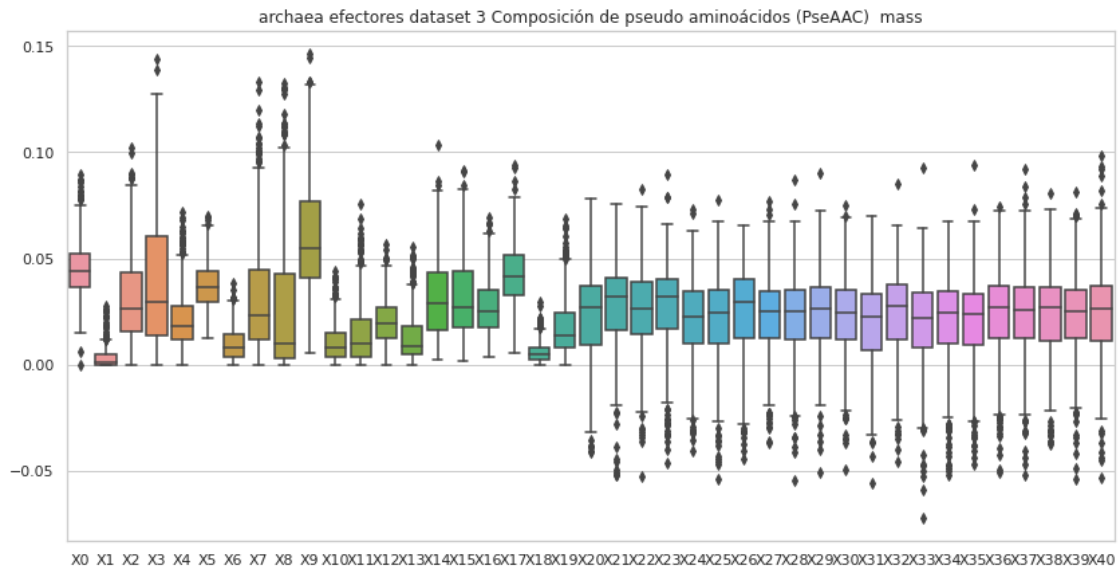
	X0	X1	X2	X3	X4	X5 \
count	426.000000	426.000000	426.000000	426.000000	426.000000	426.000000
mean	0.046137	0.004054	0.037996	0.040542	0.016214	0.037816
std	0.014869	0.005804	0.021129	0.027640	0.011170	0.012016
min	0.005899	0.000000	0.000000	0.000000	0.000000	0.006221
25%	0.036305	0.000000	0.022088	0.019855	0.008034	0.030402
50%	0.043983	0.002528	0.036259	0.035575	0.013845	0.036114
75%	0.055146	0.005099	0.050316	0.054533	0.022145	0.043141
max	0.104310	0.041860	0.114272	0.168908	0.058902	0.085795

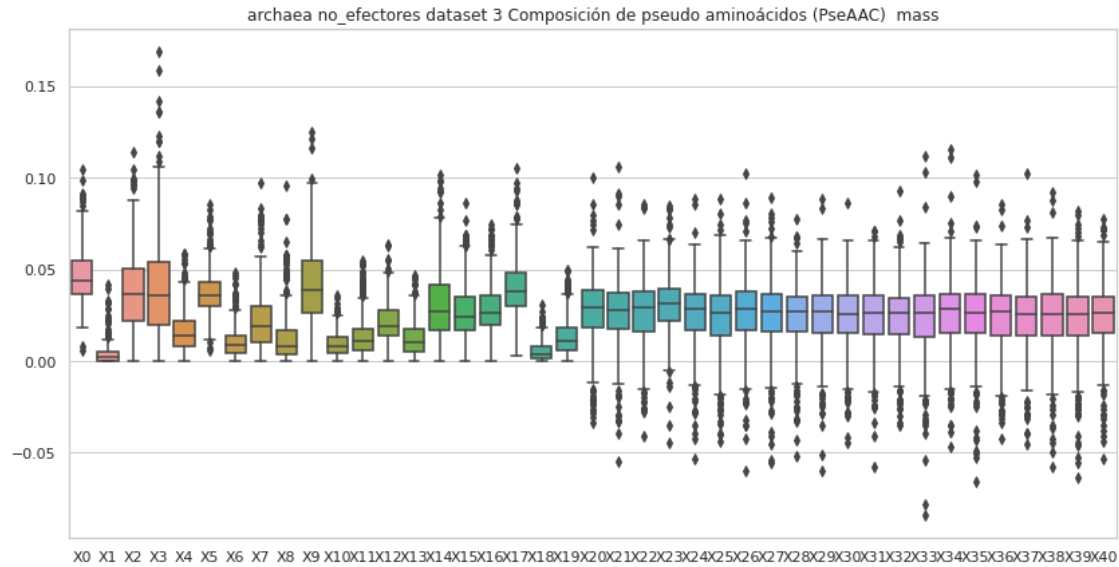
	X6	X7	X8	X9 ...	X31 \
count	426.000000	426.000000	426.000000	426.000000 ...	426.000000
mean	0.010326	0.022505	0.012514	0.042680 ...	0.024093
std	0.008421	0.016300	0.014102	0.021537 ...	0.017270
min	0.000000	0.000000	0.000000	0.000000 ...	-0.057916
25%	0.004374	0.010559	0.003391	0.026564 ...	0.014558
50%	0.008553	0.019424	0.008141	0.038812 ...	0.026384
75%	0.013885	0.029992	0.016644	0.054890 ...	0.035820
max	0.048231	0.097019	0.095695	0.125302 ...	0.071230

	X32	X33	X34	X35	X36	X37 \
count	426.000000	426.000000	426.000000	426.000000	426.000000	426.000000
mean	0.024316	0.023556	0.025551	0.025116	0.024087	0.023250
std	0.017818	0.020481	0.019363	0.019108	0.018364	0.018577
min	-0.034997	-0.084029	-0.047060	-0.065409	-0.042235	-0.045305
25%	0.014438	0.013328	0.015647	0.015578	0.013673	0.013710
50%	0.026229	0.026640	0.028263	0.026505	0.026927	0.025843
75%	0.034073	0.035650	0.036714	0.036291	0.035747	0.035079
max	0.092686	0.112070	0.115429	0.101521	0.085921	0.102647

	X38	X39	X40
count	426.000000	426.000000	426.000000
mean	0.023876	0.022904	0.024296
std	0.019279	0.019881	0.018310
min	-0.057525	-0.063567	-0.053105
25%	0.013736	0.013649	0.015240
50%	0.025952	0.025696	0.026559
75%	0.036579	0.035109	0.035448
max	0.092168	0.081972	0.077911

[8 rows x 41 columns]





5 Composición de pseudo aminoácidos (PseAAC) hidro

```
[9]: #hidro
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
comp = "hidro"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",\n"
    ↪ " + 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)+"\n
↪ "+str(transf)+" "+str(comp)+" "+str(estado))
```

efectores

Composición de pseudo aminoácidos (PseAAC) hidro efectores archaea dataset 3,
con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.025692	0.000000	0.035969	0.051385	0.023123	0.051385	0.012846
1	0.032220	0.003222	0.022554	0.067661	0.016110	0.038664	0.000000
2	0.030309	0.003789	0.018943	0.087139	0.022732	0.037887	0.026521
3	0.069143	0.000000	0.028611	0.016690	0.011921	0.040532	0.011921
4	0.023828	0.015885	0.063541	0.023828	0.023828	0.015885	0.000000
..
495	0.057901	0.000000	0.017622	0.017622	0.017622	0.037761	0.010070
496	0.023525	0.018820	0.032935	0.051755	0.018820	0.032935	0.000000
497	0.056246	0.002344	0.058590	0.072652	0.023436	0.039841	0.016405
498	0.060580	0.000000	0.063214	0.065848	0.039509	0.042143	0.010536
499	0.036186	0.002585	0.046525	0.072372	0.067203	0.080127	0.041356

	X7	X8	X9 ...	X53	X54	X55 \
0	0.059093	0.048816	0.092493 ...	0.008185	0.031108	0.034273
1	0.032220	0.045108	0.054773 ...	0.033942	-0.015997	0.017189
2	0.041675	0.064407	0.071984 ...	-0.019889	0.063133	0.056836
3	0.007153	0.004768	0.066759 ...	-0.013386	-0.019174	-0.021747
4	0.039713	0.047656	0.087369 ...	0.046116	0.027693	0.036520
..
495	0.015105	0.005035	0.047831 ...	-0.014936	0.063278	0.005675
496	0.028230	0.065870	0.070575 ...	0.007990	-0.002873	-0.002299
497	0.016405	0.002344	0.070308 ...	0.036191	0.032547	0.034899
498	0.073749	0.081651	0.086919 ...	-0.020986	0.008568	0.016317
499	0.067203	0.031017	0.049110 ...	0.034063	0.005738	0.035180

	X56	X57	X58	X59	X60	X61	X62
0	0.037126	0.038109	-0.046298	-0.028250	0.017200	0.010133	efectores
1	-0.004037	0.024453	0.027827	0.056714	0.021628	0.019215	efectores
2	-0.035522	-0.021627	-0.014993	0.018817	0.007560	0.053986	efectores
3	0.018918	-0.001951	0.016023	0.006854	0.008783	0.001580	efectores
4	0.024541	0.049754	0.026605	-0.002135	0.071792	0.063360	efectores
..
495	0.018500	-0.019683	0.046828	-0.006631	0.040057	0.005538	efectores
496	0.005486	0.007837	0.076349	0.084649	-0.044976	-0.020789	efectores
497	-0.011469	0.028494	-0.005018	0.037797	0.023499	0.032945	efectores

```

498  0.052064  0.020760 -0.026418  0.013288  0.001376  0.025025  efectores
499  0.010910 -0.014945 -0.011356 -0.008835 -0.007583 -0.000852  efectores

```

[500 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) hidro efectores archaea dataset 3,
con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.048794	0.004769	0.036522	0.045650	0.022430	0.040823
std	0.027539	0.008021	0.025026	0.033641	0.015414	0.022972
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.003623
25%	0.028745	0.000000	0.013206	0.011691	0.011238	0.026092
50%	0.044241	0.001364	0.032714	0.048199	0.018579	0.035517
75%	0.062296	0.005906	0.054970	0.070096	0.029518	0.049243
max	0.268915	0.080021	0.126975	0.179136	0.109623	0.201630

	X6	X7	X8	X9 ...	X52 \
count	500.000000	500.000000	500.000000	500.000000 ...	500.000000
mean	0.011143	0.036682	0.031545	0.063675 ...	0.005322
std	0.009309	0.031861	0.036343	0.031664 ...	0.034188
min	0.000000	0.000000	0.000000	0.014548 ...	-0.269425
25%	0.003764	0.013861	0.003567	0.039299 ...	-0.011280
50%	0.009443	0.024968	0.013688	0.056797 ...	0.013959
75%	0.015626	0.052117	0.055623	0.082257 ...	0.025107
max	0.053741	0.200023	0.180159	0.214963 ...	0.122056

	X53	X54	X55	X56	X57	X58 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.009076	0.015116	0.016910	0.008473	0.011389	0.007065
std	0.027444	0.031103	0.030797	0.031711	0.029339	0.036507
min	-0.242006	-0.112544	-0.079376	-0.204295	-0.185524	-0.271649
25%	-0.002241	-0.001478	-0.000178	-0.006060	-0.003086	-0.008622
50%	0.007813	0.017908	0.010638	0.013612	0.006261	0.014105
75%	0.020914	0.030842	0.032727	0.024224	0.025146	0.027205
max	0.123767	0.156066	0.166768	0.165019	0.178743	0.172815

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.010505	0.008740	0.011492
std	0.029883	0.035358	0.031215
min	-0.135071	-0.201543	-0.180516
25%	-0.003285	-0.005449	-0.001888
50%	0.008736	0.015807	0.007614
75%	0.022134	0.027861	0.026921

max 0.154269 0.140950 0.137795

[8 rows x 62 columns]

no_efectores

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.076099	0.000000	0.056531	0.041311	0.008697	0.054357	0.006523
1	0.062215	0.165906	0.041477	0.041477	0.020738	0.082953	0.062215
2	0.056879	0.000000	0.017708	0.014488	0.011269	0.025757	0.005903
3	0.047613	0.000000	0.047613	0.034628	0.008657	0.038956	0.008657
4	0.032395	0.002160	0.010798	0.012958	0.012958	0.025916	0.002160
..	
495	0.026706	0.000000	0.005087	0.006359	0.008902	0.025434	0.007630
496	0.055061	0.002753	0.005506	0.005506	0.013765	0.041295	0.005506
497	0.052237	0.001801	0.066647	0.055839	0.019814	0.057641	0.016211
498	0.101693	0.002075	0.066412	0.083015	0.033206	0.099618	0.012452
499	0.072844	0.007668	0.053675	0.076678	0.026837	0.057509	0.026837

	X7	X8	X9 ...	X53	X54	X55 \
0	0.017394	0.004349	0.056531 ...	0.014477	0.024466	0.018341
1	0.082953	0.041477	0.082953 ...	-0.005951	-0.070726	-0.078611
2	0.003756	0.003220	0.039708 ...	0.004766	0.003733	-0.000632
3	0.006493	0.000000	0.030299 ...	0.020491	0.007814	0.017681
4	0.022676	0.004319	0.063710 ...	0.001998	0.005735	-0.009084
..	
495	0.013989	0.001272	0.035608 ...	0.000787	0.016234	0.002064
496	0.016518	0.000000	0.052308 ...	-0.000812	0.010673	-0.015158
497	0.034224	0.021615	0.054038 ...	0.023254	-0.011829	0.031306
498	0.056035	0.002075	0.043583 ...	0.001141	0.011456	0.028561
499	0.019170	0.007668	0.092014 ...	0.015406	-0.012750	-0.017090

	X56	X57	X58	X59	X60	X61	X62
0	0.007986	0.005723	0.036892	0.029937	-0.000932	-0.013230	no_efectores
1	0.074728	-0.031948	-0.169582	-0.064321	-0.037865	0.005083	no_efectores
2	-0.002540	-0.006331	0.008586	0.001920	0.014732	0.004730	no_efectores
3	-0.019191	0.026786	-0.033483	0.033648	0.009294	0.018657	no_efectores
4	-0.005206	-0.001851	0.063099	0.014181	0.034274	0.000257	no_efectores
..	
495	0.027467	0.007580	0.019291	0.007532	0.016187	0.006396	no_efectores
496	0.044053	0.000372	0.065157	0.021641	0.049238	0.012377	no_efectores
497	0.036729	0.018803	-0.001281	0.020194	-0.035409	0.018174	no_efectores
498	0.028018	0.017732	0.012119	-0.017149	0.001951	-0.010490	no_efectores

499 0.032011 0.005806 -0.003499 -0.052945 0.038121 -0.010657 no_efectores

[500 rows x 63 columns]

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

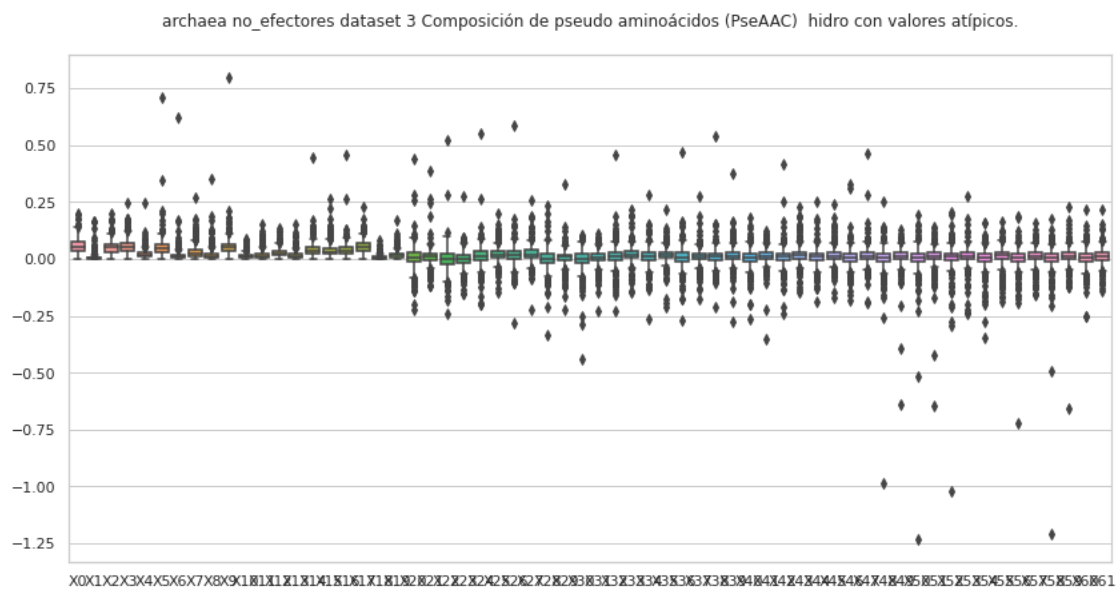
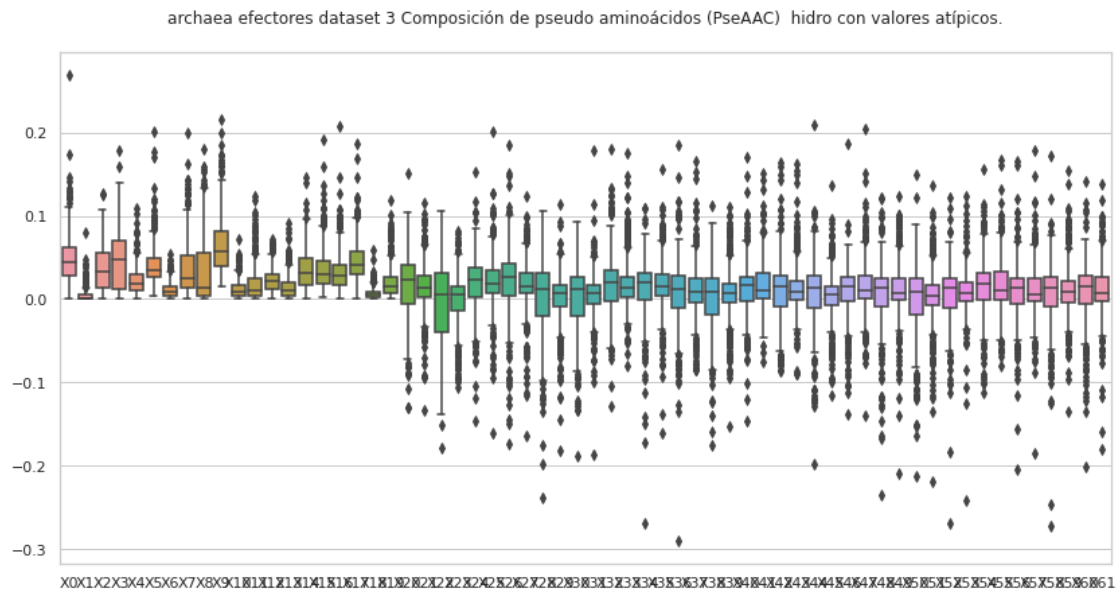
	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.059206	0.007438	0.049895	0.053553	0.021385	0.051659
std	0.029577	0.016779	0.028325	0.032627	0.018385	0.041335
min	0.001745	0.000000	0.000000	0.000000	0.000000	0.002177
25%	0.036965	0.000000	0.031570	0.032566	0.011049	0.031563
50%	0.054678	0.003152	0.051149	0.054539	0.018106	0.045953
75%	0.077875	0.007906	0.066233	0.069385	0.026653	0.064010
max	0.197316	0.165906	0.197316	0.247992	0.247992	0.708303

	X6	X7	X8	X9 ...	X52 \
count	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.015913	0.031944	0.020000	0.055230	0.004291
std	0.031736	0.028966	0.028835	0.043511	0.061303
min	0.000000	0.000000	0.000000	0.000000	-1.021464
25%	0.005190	0.013334	0.004800	0.034994	-0.008562
50%	0.011271	0.024022	0.010845	0.049479	0.009664
75%	0.019140	0.040340	0.023964	0.066455	0.022860
max	0.619765	0.266759	0.354152	0.796841	0.205690

	X53	X54	X55	X56	X57	X58 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.012140	0.002182	0.012464	0.001830	0.010841	0.000096
std	0.037808	0.042329	0.037144	0.049188	0.033667	0.068119
min	-0.242368	-0.347873	-0.190004	-0.719740	-0.160979	-1.209794
25%	-0.000609	-0.010162	-0.000678	-0.012087	-0.002780	-0.011281
50%	0.012114	0.006967	0.014064	0.007067	0.011313	0.007165
75%	0.029406	0.021754	0.031327	0.020615	0.027694	0.020704
max	0.275590	0.158698	0.164247	0.185252	0.155626	0.173233

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.009998	0.002562	0.011078
std	0.044234	0.039790	0.033685
min	-0.659605	-0.253630	-0.144304
25%	-0.002484	-0.014361	-0.003904
50%	0.010366	0.007597	0.009877
75%	0.027578	0.021671	0.029221
max	0.229922	0.215014	0.215965

[8 rows x 62 columns]



5.1 Composición de pseudo aminoácidos (PseAAC) hidro, sin valores atípicos

```
[10]: #hidro
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "hidro"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +
      ' ' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

    del df['X62']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de
    #sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X62'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
```

```
ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"  
↪ "+str(transf)+" "+str(comp))
```

efectores

Composición de pseudo aminoácidos (PseAAC) efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.025692	0.000000	0.035969	0.051385	0.023123	0.051385	0.012846
1	0.032220	0.003222	0.022554	0.067661	0.016110	0.038664	0.000000
2	0.030309	0.003789	0.018943	0.087139	0.022732	0.037887	0.026521
3	0.069143	0.000000	0.028611	0.016690	0.011921	0.040532	0.011921
4	0.023828	0.015885	0.063541	0.023828	0.023828	0.015885	0.000000
..	
494	0.039364	0.017495	0.026242	0.078727	0.021869	0.030616	0.008747
495	0.057901	0.000000	0.017622	0.017622	0.017622	0.037761	0.010070
496	0.023525	0.018820	0.032935	0.051755	0.018820	0.032935	0.000000
497	0.056246	0.002344	0.058590	0.072652	0.023436	0.039841	0.016405
498	0.060580	0.000000	0.063214	0.065848	0.039509	0.042143	0.010536

	X7	X8	X9	...	X53	X54	X55 \
0	0.059093	0.048816	0.092493	...	0.008185	0.031108	0.034273
1	0.032220	0.045108	0.054773	...	0.033942	-0.015997	0.017189
2	0.041675	0.064407	0.071984	...	-0.019889	0.063133	0.056836
3	0.007153	0.004768	0.066759	...	-0.013386	-0.019174	-0.021747
4	0.039713	0.047656	0.087369	...	0.046116	0.027693	0.036520
..	
494	0.056859	0.074354	0.074354	...	0.009142	-0.009182	0.053055
495	0.015105	0.005035	0.047831	...	-0.014936	0.063278	0.005675
496	0.028230	0.065870	0.070575	...	0.007990	-0.002873	-0.002299
497	0.016405	0.002344	0.070308	...	0.036191	0.032547	0.034899
498	0.073749	0.081651	0.086919	...	-0.020986	0.008568	0.016317

	X56	X57	X58	X59	X60	X61	X62
0	0.037126	0.038109	-0.046298	-0.028250	0.017200	0.010133	efectores
1	-0.004037	0.024453	0.027827	0.056714	0.021628	0.019215	efectores
2	-0.035522	-0.021627	-0.014993	0.018817	0.007560	0.053986	efectores
3	0.018918	-0.001951	0.016023	0.006854	0.008783	0.001580	efectores
4	0.024541	0.049754	0.026605	-0.002135	0.071792	0.063360	efectores
..	
494	0.041608	0.047286	0.054892	0.091223	-0.029320	-0.001275	efectores
495	0.018500	-0.019683	0.046828	-0.006631	0.040057	0.005538	efectores
496	0.005486	0.007837	0.076349	0.084649	-0.044976	-0.020789	efectores
497	-0.011469	0.028494	-0.005018	0.037797	0.023499	0.032945	efectores
498	0.052064	0.020760	-0.026418	0.013288	0.001376	0.025025	efectores

[394 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	394.000000	394.000000	394.000000	394.000000	394.000000	394.000000
mean	0.045926	0.003044	0.031046	0.035872	0.019501	0.035923
std	0.023115	0.005215	0.022191	0.028222	0.011265	0.016499
min	0.000000	0.000000	0.000520	0.000000	0.000000	0.003623
25%	0.028070	0.000000	0.011255	0.009431	0.010993	0.024975
50%	0.042067	0.000610	0.024467	0.027151	0.017554	0.032877
75%	0.059967	0.003674	0.049086	0.060015	0.025399	0.043938
max	0.129276	0.028764	0.098717	0.113937	0.057993	0.107795

	X6	X7	X8	X9 ...	X52 \
count	394.000000	394.000000	394.000000	394.000000 ...	394.000000
mean	0.009339	0.028302	0.022198	0.055124 ...	0.010687
std	0.007322	0.023586	0.027941	0.023040 ...	0.023901
min	0.000000	0.000000	0.000000	0.014548 ...	-0.084176
25%	0.003125	0.011935	0.002604	0.036465 ...	-0.003123
50%	0.008558	0.019865	0.009245	0.050767 ...	0.017391
75%	0.012988	0.039079	0.031367	0.070287 ...	0.025807
max	0.038299	0.125086	0.126537	0.153913 ...	0.095565

	X53	X54	X55	X56	X57	X58 \
count	394.000000	394.000000	394.000000	394.000000	394.000000	394.000000
mean	0.010163	0.016866	0.016083	0.010317	0.010722	0.012083
std	0.018503	0.024456	0.024828	0.022680	0.020140	0.024613
min	-0.039490	-0.069150	-0.062023	-0.068777	-0.038889	-0.081132
25%	-0.000717	0.001602	0.000494	-0.002162	-0.001532	0.000224
50%	0.008060	0.018810	0.010236	0.014580	0.006172	0.016689
75%	0.019073	0.029563	0.028528	0.024256	0.020714	0.027442
max	0.081987	0.105394	0.095691	0.096322	0.084112	0.082349

	X59	X60	X61
count	394.000000	394.000000	394.000000
mean	0.011037	0.011452	0.011032
std	0.021892	0.027173	0.021537
min	-0.062770	-0.097124	-0.060663
25%	-0.000700	-0.001044	-0.001104
50%	0.009018	0.017000	0.007104
75%	0.020928	0.027023	0.023809
max	0.096668	0.085630	0.088161

[8 rows x 62 columns]

no_efectores

Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.076099	0.000000	0.056531	0.041311	0.008697	0.054357	0.006523
2	0.056879	0.000000	0.017708	0.014488	0.011269	0.025757	0.005903
3	0.047613	0.000000	0.047613	0.034628	0.008657	0.038956	0.008657
4	0.032395	0.002160	0.010798	0.012958	0.012958	0.025916	0.002160
5	0.053279	0.002664	0.058607	0.058607	0.023976	0.037296	0.018648
..	
495	0.026706	0.000000	0.005087	0.006359	0.008902	0.025434	0.007630
496	0.055061	0.002753	0.005506	0.005506	0.013765	0.041295	0.005506
497	0.052237	0.001801	0.066647	0.055839	0.019814	0.057641	0.016211
498	0.101693	0.002075	0.066412	0.083015	0.033206	0.099618	0.012452
499	0.072844	0.007668	0.053675	0.076678	0.026837	0.057509	0.026837

	X7	X8	X9	...	X53	X54	X55 \
0	0.017394	0.004349	0.056531	...	0.014477	0.024466	0.018341
2	0.003756	0.003220	0.039708	...	0.004766	0.003733	-0.000632
3	0.006493	0.000000	0.030299	...	0.020491	0.007814	0.017681
4	0.022676	0.004319	0.063710	...	0.001998	0.005735	-0.009084
5	0.066599	0.063935	0.101231	...	0.019074	0.006629	-0.000092
..	
495	0.013989	0.001272	0.035608	...	0.000787	0.016234	0.002064
496	0.016518	0.000000	0.052308	...	-0.000812	0.010673	-0.015158
497	0.034224	0.021615	0.054038	...	0.023254	-0.011829	0.031306
498	0.056035	0.002075	0.043583	...	0.001141	0.011456	0.028561
499	0.019170	0.007668	0.092014	...	0.015406	-0.012750	-0.017090

	X56	X57	X58	X59	X60	X61	X62
0	0.007986	0.005723	0.036892	0.029937	-0.000932	-0.013230	no_efectores
2	-0.002540	-0.006331	0.008586	0.001920	0.014732	0.004730	no_efectores
3	-0.019191	0.026786	-0.033483	0.033648	0.009294	0.018657	no_efectores
4	-0.005206	-0.001851	0.063099	0.014181	0.034274	0.000257	no_efectores
5	0.008923	0.046775	-0.013311	0.005412	-0.029301	0.015532	no_efectores
..	
495	0.027467	0.007580	0.019291	0.007532	0.016187	0.006396	no_efectores
496	0.044053	0.000372	0.065157	0.021641	0.049238	0.012377	no_efectores
497	0.036729	0.018803	-0.001281	0.020194	-0.035409	0.018174	no_efectores
498	0.028018	0.017732	0.012119	-0.017149	0.001951	-0.010490	no_efectores
499	0.032011	0.005806	-0.003499	-0.052945	0.038121	-0.010657	no_efectores

[420 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 3, sin valores atípicos.

Estadísticas.

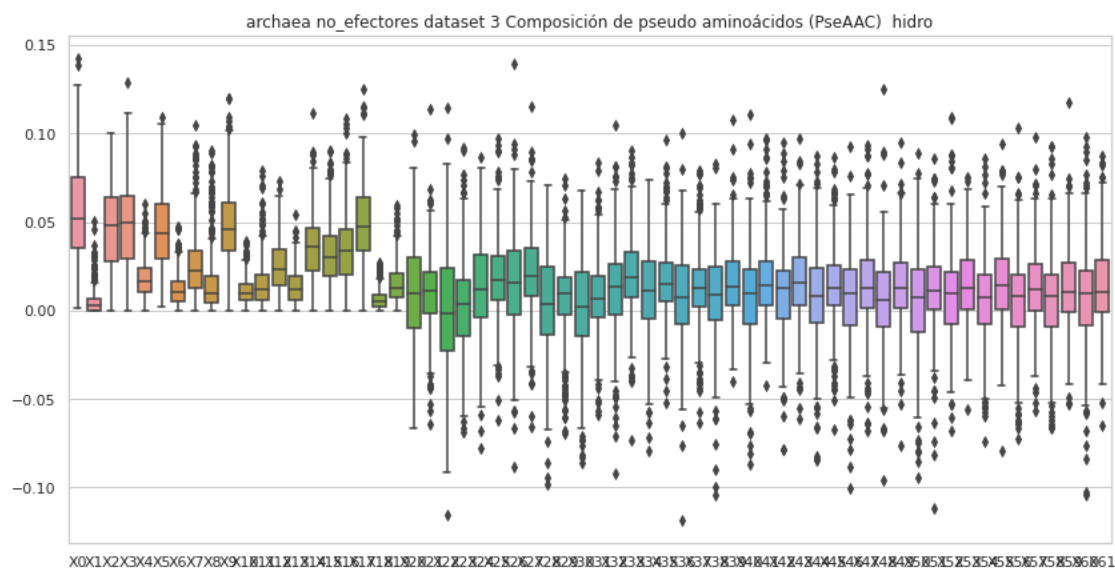
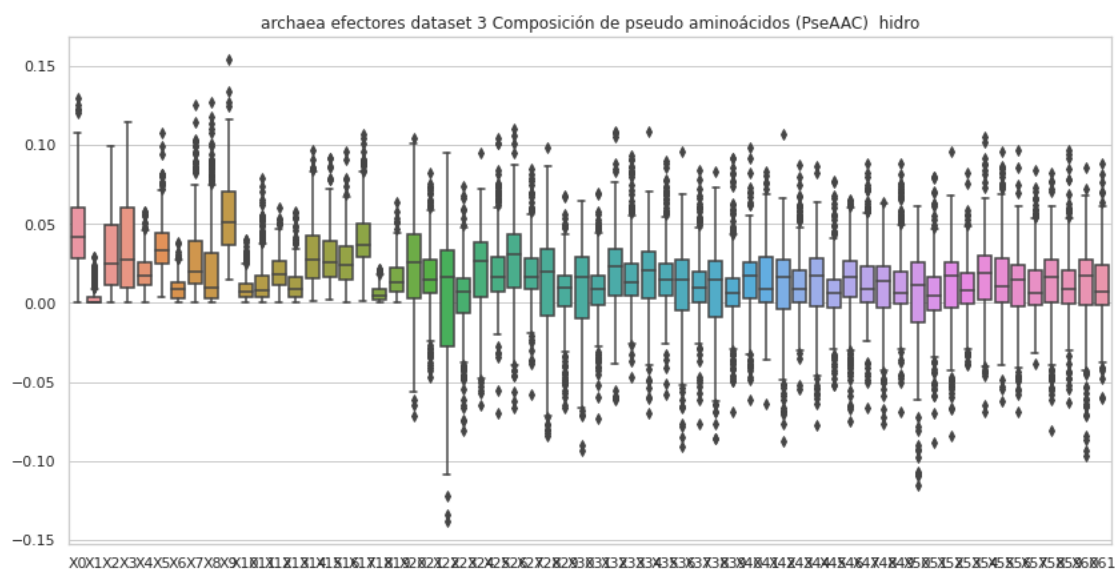
	X0	X1	X2	X3	X4	X5 \
count	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000
mean	0.055832	0.004864	0.045889	0.047002	0.018217	0.045936
std	0.025879	0.006743	0.024145	0.025119	0.010996	0.021120
min	0.001745	0.000000	0.000000	0.000000	0.000000	0.002177
25%	0.035634	0.000000	0.027420	0.029126	0.010480	0.029209
50%	0.051713	0.002924	0.048414	0.049396	0.016786	0.043278
75%	0.075239	0.006502	0.063911	0.064649	0.023806	0.060545
max	0.142322	0.050625	0.100490	0.128896	0.060360	0.109027

	X6	X7	X8	X9 ...	X52 \
count	420.000000	420.000000	420.000000	420.000000	420.000000
mean	0.011863	0.025675	0.014651	0.048330	0.007931
std	0.009095	0.018129	0.016434	0.020567	0.024134
min	0.000000	0.000000	0.000000	0.000000	-0.067944
25%	0.004993	0.012356	0.004431	0.033951	-0.007769
50%	0.010617	0.022811	0.009823	0.046249	0.009586
75%	0.016339	0.034065	0.019166	0.061258	0.021370
max	0.047280	0.104835	0.090168	0.120000	0.109148

	X53	X54	X55	X56	X57	X58 \
count	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000
mean	0.014421	0.005775	0.015102	0.006145	0.013680	0.005327
std	0.019763	0.023262	0.021356	0.024011	0.021139	0.024412
min	-0.055681	-0.074044	-0.079392	-0.072307	-0.056534	-0.066103
25%	0.000956	-0.007692	0.000613	-0.009261	0.000234	-0.009083
50%	0.012579	0.007781	0.014221	0.007946	0.011752	0.008165
75%	0.028175	0.020446	0.029437	0.020566	0.026514	0.020397
max	0.075285	0.085970	0.094170	0.102992	0.097529	0.092873

	X59	X60	X61
count	420.000000	420.000000	420.000000
mean	0.012649	0.006735	0.013673
std	0.021141	0.027262	0.023952
min	-0.052945	-0.104184	-0.065260
25%	-0.000834	-0.008496	-0.000749
50%	0.010704	0.009422	0.010484
75%	0.026760	0.022250	0.028814
max	0.117668	0.097967	0.087406

[8 rows x 62 columns]



6 Covarianza de auto cruzamiento (ACC) hidro_mass

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



```

comp = "hidro_mass"
df=""

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

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

```

efectores

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 3,
con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.090577	0.116321	-0.044794	0.101101	-0.033592	0.007461	-0.097244
1	-0.100065	-0.052556	-0.061478	0.027026	0.086400	-0.071144	-0.105030
2	0.018098	0.008925	0.032485	0.022552	0.118641	-0.050163	-0.005134
3	0.117153	0.083594	0.056243	0.062774	-0.038136	-0.015458	-0.014148
4	0.078087	-0.015119	-0.006791	-0.070146	-0.031690	-0.000582	-0.014888
..
495	-0.070786	0.098400	-0.045124	0.071631	0.020758	-0.115568	0.012798
496	0.011383	-0.031248	-0.036010	0.032164	-0.107343	-0.045003	0.070823
497	-0.044145	-0.022006	-0.038821	-0.119032	0.070035	-0.016466	-0.017405
498	-0.057032	0.042631	-0.026448	-0.003339	0.052676	-0.078402	0.036407
499	0.030194	0.006911	0.086152	-0.073459	-0.062670	-0.048895	-0.117928
	X7	X8	X9	X10	X11	X12	X13
0	-0.053554	-0.041735	0.039170	0.059882	0.042448	0.025450	efectores

1	-0.013180	-0.014414	0.075782	0.012873	0.028094	-0.048032	efectores
2	0.049338	0.018123	-0.040095	0.041548	0.055932	0.067426	efectores
3	-0.048188	0.084097	0.053476	-0.050901	-0.037147	0.000811	efectores
4	0.024522	-0.058602	0.004804	0.006654	-0.006080	0.057311	efectores
..	
495	-0.048309	-0.016352	-0.190352	0.122769	-0.135269	-0.025126	efectores
496	0.073061	-0.057323	0.034942	0.008197	0.085763	-0.104814	efectores
497	-0.031513	-0.002641	0.078650	-0.007854	-0.032417	0.033197	efectores
498	0.017665	0.116583	0.017165	0.004276	0.001805	-0.052195	efectores
499	-0.035858	0.043065	0.071240	0.028677	0.023344	0.087320	efectores

[500 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 3,
con valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.005211	0.020576	0.010491	0.020006	-0.005296	-0.001076	
std	0.066745	0.072385	0.071983	0.069776	0.068676	0.068854	
min	-0.274871	-0.230505	-0.524088	-0.579639	-0.219291	-0.247234	
25%	-0.033203	-0.022107	-0.029346	-0.016286	-0.047750	-0.044824	
50%	0.008282	0.019786	0.010159	0.024107	-0.007034	-0.000023	
75%	0.045303	0.066767	0.054404	0.063312	0.038549	0.041546	
max	0.226232	0.220448	0.263949	0.251551	0.210008	0.208636	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.016566	0.000710	-0.002518	0.002186	0.003222	-0.010351	
std	0.089162	0.068726	0.072914	0.074536	0.072436	0.067376	
min	-0.268971	-0.269360	-0.278363	-0.592968	-0.538224	-0.305499	
25%	-0.029878	-0.035280	-0.048262	-0.036789	-0.036628	-0.051687	
50%	0.013406	0.003207	-0.003176	0.002250	0.005041	-0.005934	
75%	0.058197	0.044514	0.044018	0.046234	0.045970	0.029660	
max	1.283083	0.236326	0.254521	0.237995	0.215082	0.218471	

	X12
count	500.000000
mean	0.013938
std	0.067940
min	-0.212499
25%	-0.031573
50%	0.011853
75%	0.058340
max	0.266121

no_efectores

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 3,
con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.025169	-0.017296	-0.042496	0.049787	-0.085673	-0.017847	0.056794
1	-0.084197	-0.062985	0.042907	-0.083224	0.198237	-0.126748	-0.220461
2	0.075630	0.086890	0.062427	0.047605	0.112470	0.108294	0.025755
3	0.108840	0.148543	-0.054125	0.048530	-0.048280	0.021827	-0.123771
4	0.099334	0.088008	0.081632	0.117466	0.032784	0.050273	-0.015217
..	
495	0.168900	0.138382	0.035177	-0.021085	0.089214	0.050359	0.069464
496	0.153612	0.137255	-0.117263	-0.065715	0.057052	-0.050742	0.136148
497	0.046857	-0.013265	0.105587	0.058155	0.001258	-0.019465	0.018188
498	0.053520	-0.005163	0.022098	0.084011	0.001717	0.049541	0.076803
499	-0.032584	0.089332	-0.056060	0.050268	-0.062538	0.027189	0.006430

	X7	X8	X9	X10	X11	X12	X13
0	0.031925	-0.026479	0.015174	0.054948	0.003791	-0.008131	no_efectores
1	0.098511	-0.167152	0.084196	0.002181	0.057468	0.214752	no_efectores
2	0.048789	0.051203	0.103743	0.043385	0.075028	0.073661	no_efectores
3	0.119873	-0.005985	-0.010499	-0.129302	0.022737	-0.047830	no_efectores
4	0.058111	-0.042058	-0.032770	0.028078	-0.054996	-0.006981	no_efectores
..	
495	0.005489	-0.026977	0.008763	-0.090305	-0.030215	-0.081442	no_efectores
496	-0.022779	0.037531	-0.095281	-0.032210	-0.015361	0.011746	no_efectores
497	-0.020544	0.046007	0.055894	0.012101	-0.061160	0.045674	no_efectores
498	0.039555	0.027703	-0.037534	-0.006177	0.019143	0.039990	no_efectores
499	0.098874	-0.105017	-0.035898	-0.009260	-0.081068	0.061638	no_efectores

[500 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 3,
con valores atípicos.

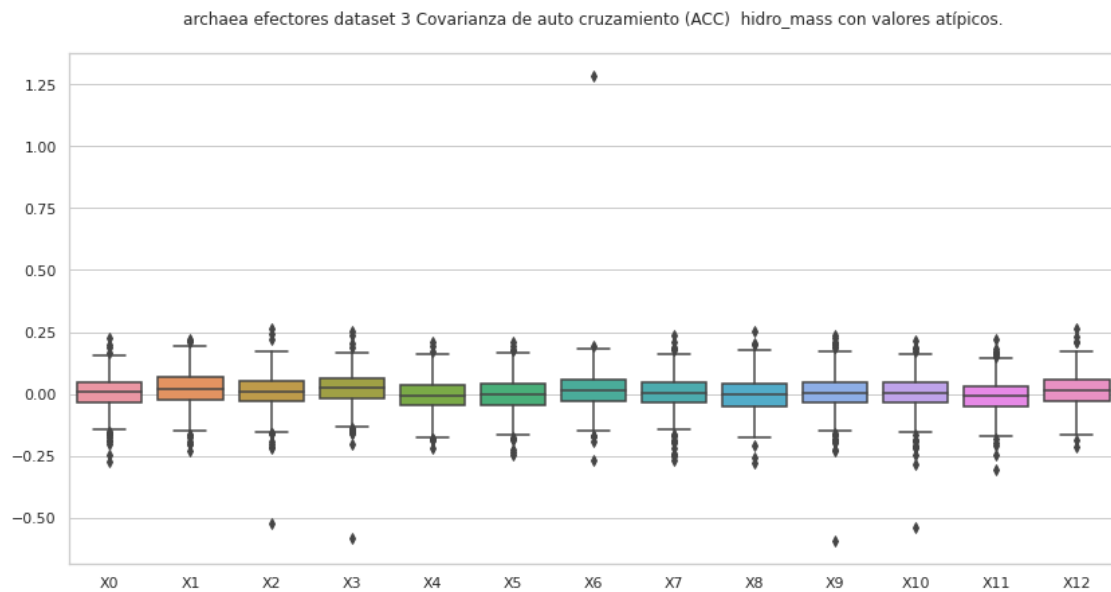
Estadísticas.

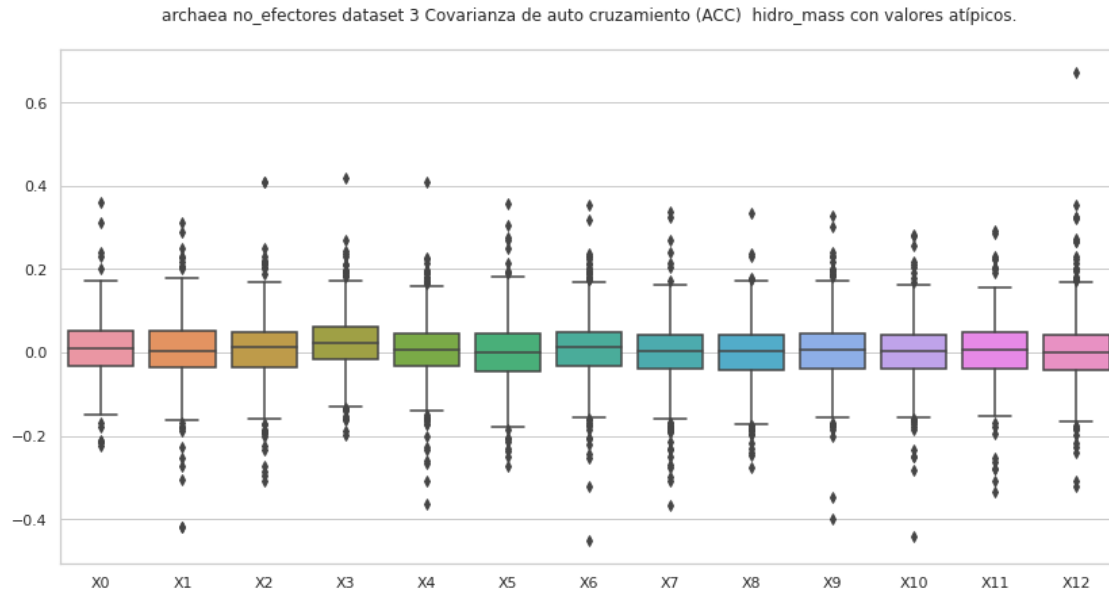
	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.009214	0.004362	0.008578	0.025144	0.006156	-0.000172
std	0.072435	0.081138	0.082251	0.070107	0.078451	0.079318
min	-0.222567	-0.419678	-0.309233	-0.199495	-0.365167	-0.272161
25%	-0.031210	-0.037073	-0.036728	-0.016244	-0.031181	-0.045405
50%	0.010361	0.004042	0.012114	0.023713	0.006262	-0.000134
75%	0.052444	0.050645	0.048919	0.061299	0.045743	0.045776

max	0.360121	0.312745	0.410104	0.419401	0.408482	0.356965
-----	----------	----------	----------	----------	----------	----------

	X6	X7	X8	X9	X10	X11 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.009745	-0.001283	0.000349	0.005702	0.001661	0.003741
std	0.082907	0.080862	0.075703	0.076561	0.074342	0.078836
min	-0.451340	-0.367058	-0.275858	-0.399014	-0.441894	-0.334866
25%	-0.032971	-0.040060	-0.042459	-0.039949	-0.037290	-0.037292
50%	0.013752	0.004002	0.002951	0.006081	0.002218	0.005944
75%	0.049845	0.042591	0.043602	0.045665	0.043563	0.049010
max	0.353556	0.337043	0.333899	0.327291	0.281982	0.291854

	X12
count	500.000000
mean	0.004942
std	0.088967
min	-0.321513
25%	-0.041185
50%	0.001044
75%	0.042686
max	0.672057





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) +",\n"
      '↪' + 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)+"\n
→"+str(transf)+" "+str(comp))

```

efectores

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 3,
sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.090577	0.116321	-0.044794	0.101101	-0.033592	0.007461	-0.097244
1	-0.100065	-0.052556	-0.061478	0.027026	0.086400	-0.071144	-0.105030
2	0.018098	0.008925	0.032485	0.022552	0.118641	-0.050163	-0.005134
3	0.117153	0.083594	0.056243	0.062774	-0.038136	-0.015458	-0.014148
4	0.078087	-0.015119	-0.006791	-0.070146	-0.031690	-0.000582	-0.014888
..
495	-0.070786	0.098400	-0.045124	0.071631	0.020758	-0.115568	0.012798
496	0.011383	-0.031248	-0.036010	0.032164	-0.107343	-0.045003	0.070823
497	-0.044145	-0.022006	-0.038821	-0.119032	0.070035	-0.016466	-0.017405
498	-0.057032	0.042631	-0.026448	-0.003339	0.052676	-0.078402	0.036407
499	0.030194	0.006911	0.086152	-0.073459	-0.062670	-0.048895	-0.117928

	X7	X8	X9	X10	X11	X12	X13
0	-0.053554	-0.041735	0.039170	0.059882	0.042448	0.025450	efectores
1	-0.013180	-0.014414	0.075782	0.012873	0.028094	-0.048032	efectores
2	0.049338	0.018123	-0.040095	0.041548	0.055932	0.067426	efectores
3	-0.048188	0.084097	0.053476	-0.050901	-0.037147	0.000811	efectores
4	0.024522	-0.058602	0.004804	0.006654	-0.006080	0.057311	efectores
..
495	-0.048309	-0.016352	-0.190352	0.122769	-0.135269	-0.025126	efectores

```

496  0.073061 -0.057323  0.034942  0.008197  0.085763 -0.104814  efectores
497 -0.031513 -0.002641  0.078650 -0.007854 -0.032417  0.033197  efectores
498  0.017665  0.116583  0.017165  0.004276  0.001805 -0.052195  efectores
499 -0.035858  0.043065  0.071240  0.028677  0.023344  0.087320  efectores

```

[459 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 3,
sin valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000	
mean	0.005936	0.020190	0.012501	0.019932	-0.005535	0.000615	
std	0.061918	0.068677	0.063820	0.061326	0.065117	0.064214	
min	-0.189801	-0.191380	-0.203654	-0.160931	-0.185612	-0.186428	
25%	-0.031145	-0.021819	-0.028587	-0.016690	-0.046846	-0.041980	
50%	0.009043	0.018140	0.010403	0.023328	-0.006678	0.001879	
75%	0.043818	0.065747	0.053940	0.062961	0.035059	0.040311	
max	0.187354	0.220448	0.218388	0.186052	0.195969	0.195392	

	X6	X7	X8	X9	X10	X11	\
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000	
mean	0.015317	0.002738	-0.000647	0.003221	0.006487	-0.008057	
std	0.065936	0.061747	0.067429	0.066212	0.062442	0.061534	
min	-0.171652	-0.196812	-0.205515	-0.195717	-0.205868	-0.178529	
25%	-0.028883	-0.034375	-0.044707	-0.036622	-0.033161	-0.048210	
50%	0.013546	0.003131	-0.001346	0.001404	0.006350	-0.005488	
75%	0.057608	0.042570	0.043492	0.045487	0.046337	0.028773	
max	0.191644	0.190355	0.204652	0.207193	0.183591	0.181912	

	X12
count	459.000000
mean	0.013220
std	0.063657
min	-0.187702
25%	-0.031263
50%	0.011827
75%	0.056628
max	0.173780

no_efectores

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 3,
sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.025169	-0.017296	-0.042496	0.049787	-0.085673	-0.017847	0.056794
1	-0.084197	-0.062985	0.042907	-0.083224	0.198237	-0.126748	-0.220461
2	0.075630	0.086890	0.062427	0.047605	0.112470	0.108294	0.025755
3	0.108840	0.148543	-0.054125	0.048530	-0.048280	0.021827	-0.123771
4	0.099334	0.088008	0.081632	0.117466	0.032784	0.050273	-0.015217
..	
495	0.168900	0.138382	0.035177	-0.021085	0.089214	0.050359	0.069464
496	0.153612	0.137255	-0.117263	-0.065715	0.057052	-0.050742	0.136148
497	0.046857	-0.013265	0.105587	0.058155	0.001258	-0.019465	0.018188
498	0.053520	-0.005163	0.022098	0.084011	0.001717	0.049541	0.076803
499	-0.032584	0.089332	-0.056060	0.050268	-0.062538	0.027189	0.006430

	X7	X8	X9	X10	X11	X12	X13
0	0.031925	-0.026479	0.015174	0.054948	0.003791	-0.008131	no_efectores
1	0.098511	-0.167152	0.084196	0.002181	0.057468	0.214752	no_efectores
2	0.048789	0.051203	0.103743	0.043385	0.075028	0.073661	no_efectores
3	0.119873	-0.005985	-0.010499	-0.129302	0.022737	-0.047830	no_efectores
4	0.058111	-0.042058	-0.032770	0.028078	-0.054996	-0.006981	no_efectores
..	
495	0.005489	-0.026977	0.008763	-0.090305	-0.030215	-0.081442	no_efectores
496	-0.022779	0.037531	-0.095281	-0.032210	-0.015361	0.011746	no_efectores
497	-0.020544	0.046007	0.055894	0.012101	-0.061160	0.045674	no_efectores
498	0.039555	0.027703	-0.037534	-0.006177	0.019143	0.039990	no_efectores
499	0.098874	-0.105017	-0.035898	-0.009260	-0.081068	0.061638	no_efectores

[452 rows x 14 columns]

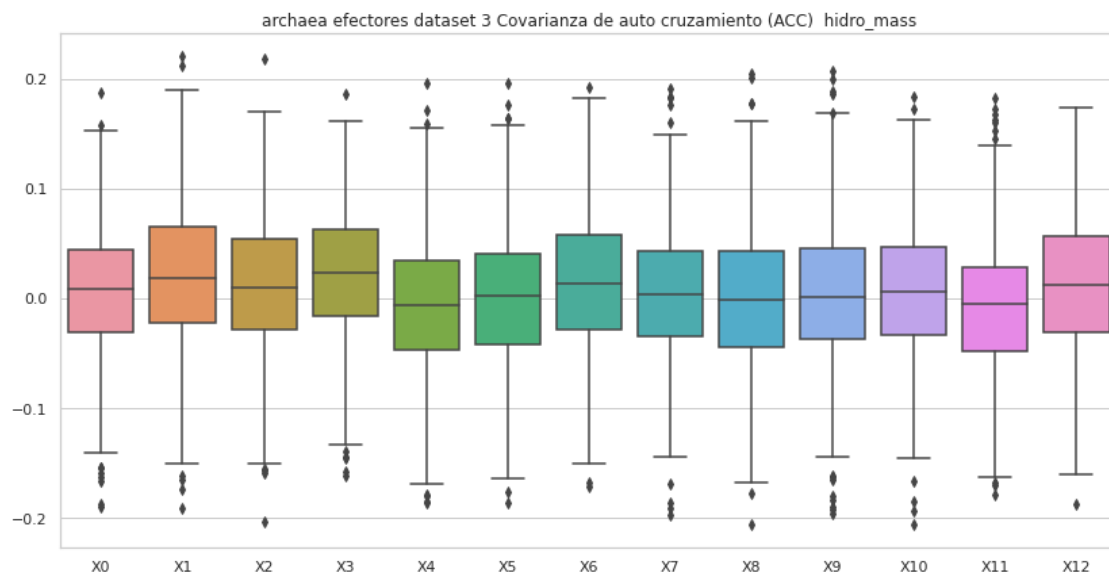
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 3,
sin valores atípicos.
Estadísticas.

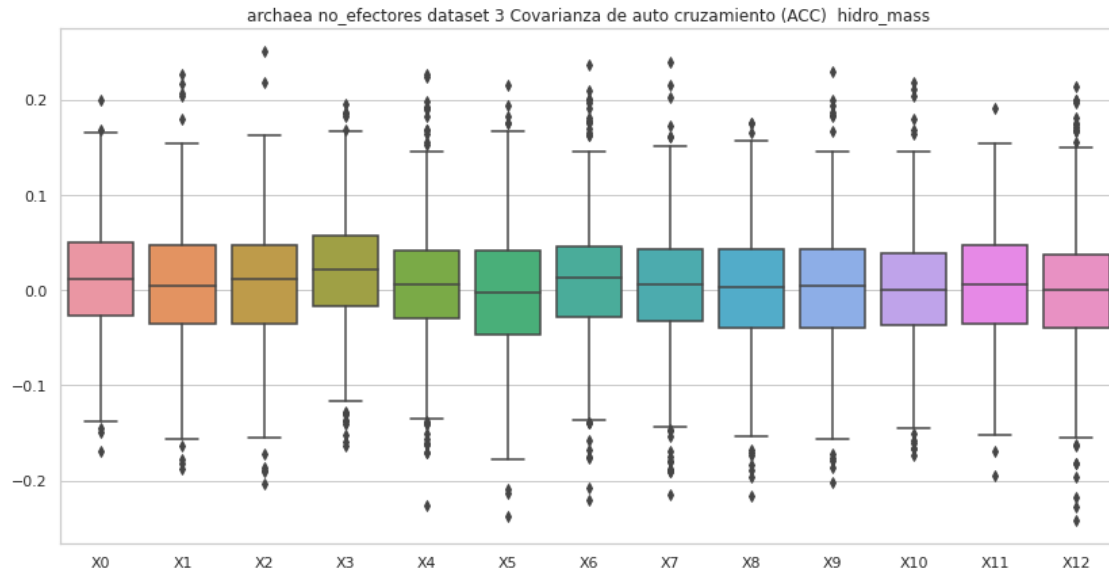
	X0	X1	X2	X3	X4	X5 \
count	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000
mean	0.010903	0.005777	0.008264	0.021231	0.006277	-0.003670
std	0.060097	0.066636	0.066186	0.059707	0.066549	0.068617
min	-0.168802	-0.187604	-0.202767	-0.163079	-0.226212	-0.236889
25%	-0.027143	-0.034794	-0.035291	-0.016244	-0.029643	-0.046834
50%	0.011133	0.005002	0.011582	0.021220	0.005985	-0.001989
75%	0.050895	0.048043	0.047347	0.057670	0.041726	0.041805
max	0.200352	0.227570	0.251125	0.195907	0.226392	0.214951

	X6	X7	X8	X9	X10	X11 \
count	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000
mean	0.009332	0.003813	0.000408	0.003177	0.001577	0.002877
std	0.067531	0.066938	0.065739	0.064207	0.061899	0.065400

min	-0.220461	-0.214794	-0.216761	-0.201672	-0.173166	-0.195505
25%	-0.027599	-0.032018	-0.038899	-0.039409	-0.036218	-0.035342
50%	0.012634	0.005539	0.002731	0.005219	0.000764	0.005440
75%	0.045875	0.043556	0.042714	0.042665	0.038730	0.047526
max	0.236571	0.240307	0.175653	0.230281	0.218007	0.190981

	X12
count	452.000000
mean	-0.000364
std	0.069390
min	-0.241842
25%	-0.039695
50%	0.000521
75%	0.037843
max	0.214752





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) +",\n"
↪ " + 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)+"\n
↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

efectores

Covarianza de auto cruzamiento (ACC) mass efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.090577	0.116321	-0.044794	0.101101	-0.033592	0.007461	-0.097244
1	-0.100065	-0.052556	-0.061478	0.027026	0.086400	-0.071144	-0.105030
2	0.018098	0.008925	0.032485	0.022552	0.118641	-0.050163	-0.005134
3	0.117153	0.083594	0.056243	0.062774	-0.038136	-0.015458	-0.014148
4	0.078087	-0.015119	-0.006791	-0.070146	-0.031690	-0.000582	-0.014888
..	
495	-0.070786	0.098400	-0.045124	0.071631	0.020758	-0.115568	0.012798
496	0.011383	-0.031248	-0.036010	0.032164	-0.107343	-0.045003	0.070823
497	-0.044145	-0.022006	-0.038821	-0.119032	0.070035	-0.016466	-0.017405
498	-0.057032	0.042631	-0.026448	-0.003339	0.052676	-0.078402	0.036407
499	0.030194	0.006911	0.086152	-0.073459	-0.062670	-0.048895	-0.117928
	X7	X8	X9	X10	X11	X12	X13
0	-0.053554	-0.041735	0.039170	0.059882	0.042448	0.025450	efectores
1	-0.013180	-0.014414	0.075782	0.012873	0.028094	-0.048032	efectores
2	0.049338	0.018123	-0.040095	0.041548	0.055932	0.067426	efectores
3	-0.048188	0.084097	0.053476	-0.050901	-0.037147	0.000811	efectores
4	0.024522	-0.058602	0.004804	0.006654	-0.006080	0.057311	efectores
..	
495	-0.048309	-0.016352	-0.190352	0.122769	-0.135269	-0.025126	efectores
496	0.073061	-0.057323	0.034942	0.008197	0.085763	-0.104814	efectores
497	-0.031513	-0.002641	0.078650	-0.007854	-0.032417	0.033197	efectores
498	0.017665	0.116583	0.017165	0.004276	0.001805	-0.052195	efectores
499	-0.035858	0.043065	0.071240	0.028677	0.023344	0.087320	efectores

[500 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) mass efectores archaea dataset 3, con valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.005211	0.020576	0.010491	0.020006	-0.005296	-0.001076

std	0.066745	0.072385	0.071983	0.069776	0.068676	0.068854
min	-0.274871	-0.230505	-0.524088	-0.579639	-0.219291	-0.247234
25%	-0.033203	-0.022107	-0.029346	-0.016286	-0.047750	-0.044824
50%	0.008282	0.019786	0.010159	0.024107	-0.007034	-0.000023
75%	0.045303	0.066767	0.054404	0.063312	0.038549	0.041546
max	0.226232	0.220448	0.263949	0.251551	0.210008	0.208636

	X6	X7	X8	X9	X10	X11 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.016566	0.000710	-0.002518	0.002186	0.003222	-0.010351
std	0.089162	0.068726	0.072914	0.074536	0.072436	0.067376
min	-0.268971	-0.269360	-0.278363	-0.592968	-0.538224	-0.305499
25%	-0.029878	-0.035280	-0.048262	-0.036789	-0.036628	-0.051687
50%	0.013406	0.003207	-0.003176	0.002250	0.005041	-0.005934
75%	0.058197	0.044514	0.044018	0.046234	0.045970	0.029660
max	1.283083	0.236326	0.254521	0.237995	0.215082	0.218471

	X12
count	500.000000
mean	0.013938
std	0.067940
min	-0.212499
25%	-0.031573
50%	0.011853
75%	0.058340
max	0.266121

no_efectores

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.025169	-0.017296	-0.042496	0.049787	-0.085673	-0.017847	0.056794
1	-0.084197	-0.062985	0.042907	-0.083224	0.198237	-0.126748	-0.220461
2	0.075630	0.086890	0.062427	0.047605	0.112470	0.108294	0.025755
3	0.108840	0.148543	-0.054125	0.048530	-0.048280	0.021827	-0.123771
4	0.099334	0.088008	0.081632	0.117466	0.032784	0.050273	-0.015217
..
495	0.168900	0.138382	0.035177	-0.021085	0.089214	0.050359	0.069464
496	0.153612	0.137255	-0.117263	-0.065715	0.057052	-0.050742	0.136148
497	0.046857	-0.013265	0.105587	0.058155	0.001258	-0.019465	0.018188
498	0.053520	-0.005163	0.022098	0.084011	0.001717	0.049541	0.076803
499	-0.032584	0.089332	-0.056060	0.050268	-0.062538	0.027189	0.006430
	X7	X8	X9	X10	X11	X12	X13

0	0.031925	-0.026479	0.015174	0.054948	0.003791	-0.008131	no_efectores
1	0.098511	-0.167152	0.084196	0.002181	0.057468	0.214752	no_efectores
2	0.048789	0.051203	0.103743	0.043385	0.075028	0.073661	no_efectores
3	0.119873	-0.005985	-0.010499	-0.129302	0.022737	-0.047830	no_efectores
4	0.058111	-0.042058	-0.032770	0.028078	-0.054996	-0.006981	no_efectores
..	
495	0.005489	-0.026977	0.008763	-0.090305	-0.030215	-0.081442	no_efectores
496	-0.022779	0.037531	-0.095281	-0.032210	-0.015361	0.011746	no_efectores
497	-0.020544	0.046007	0.055894	0.012101	-0.061160	0.045674	no_efectores
498	0.039555	0.027703	-0.037534	-0.006177	0.019143	0.039990	no_efectores
499	0.098874	-0.105017	-0.035898	-0.009260	-0.081068	0.061638	no_efectores

[500 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 3, con valores atípicos.

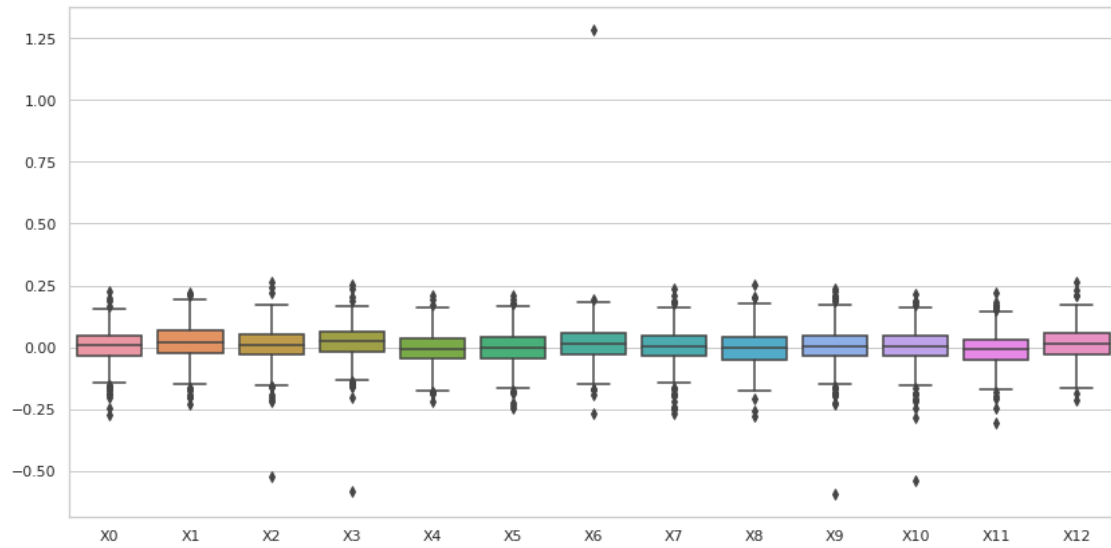
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.009214	0.004362	0.008578	0.025144	0.006156	-0.000172
std	0.072435	0.081138	0.082251	0.070107	0.078451	0.079318
min	-0.222567	-0.419678	-0.309233	-0.199495	-0.365167	-0.272161
25%	-0.031210	-0.037073	-0.036728	-0.016244	-0.031181	-0.045405
50%	0.010361	0.004042	0.012114	0.023713	0.006262	-0.000134
75%	0.052444	0.050645	0.048919	0.061299	0.045743	0.045776
max	0.360121	0.312745	0.410104	0.419401	0.408482	0.356965

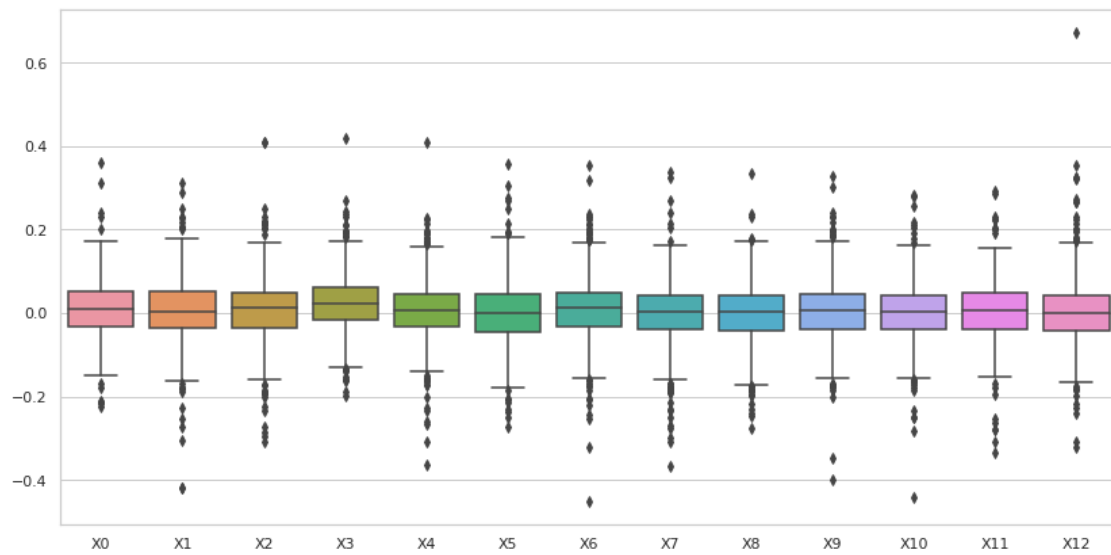
	X6	X7	X8	X9	X10	X11 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.009745	-0.001283	0.000349	0.005702	0.001661	0.003741
std	0.082907	0.080862	0.075703	0.076561	0.074342	0.078836
min	-0.451340	-0.367058	-0.275858	-0.399014	-0.441894	-0.334866
25%	-0.032971	-0.040060	-0.042459	-0.039949	-0.037290	-0.037292
50%	0.013752	0.004002	0.002951	0.006081	0.002218	0.005944
75%	0.049845	0.042591	0.043602	0.045665	0.043563	0.049010
max	0.353556	0.337043	0.333899	0.327291	0.281982	0.291854

	X12
count	500.000000
mean	0.004942
std	0.088967
min	-0.321513
25%	-0.041185
50%	0.001044
75%	0.042686
max	0.672057

archaea efectores dataset 3 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.



archaea no_efectores dataset 3 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.



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

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

#Se eliminan todas las filas que tengan valores atípicos en al menos una de sus
→columnas.
out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +
→ '_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df=""
df_out = pd.DataFrame()

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

    del df['X13']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de
→sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X13'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
```

```
ax.set_title(organismo + ' '+str(etiq)+" dataset "+str(dataset)+"\n
↪ "+str(transf)+" "+str(comp))
```

Covarianza de auto cruzamiento (ACC) mass efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.090577	0.116321	-0.044794	0.101101	-0.033592	0.007461	-0.097244
1	-0.100065	-0.052556	-0.061478	0.027026	0.086400	-0.071144	-0.105030
2	0.018098	0.008925	0.032485	0.022552	0.118641	-0.050163	-0.005134
3	0.117153	0.083594	0.056243	0.062774	-0.038136	-0.015458	-0.014148
4	0.078087	-0.015119	-0.006791	-0.070146	-0.031690	-0.000582	-0.014888
..
495	-0.070786	0.098400	-0.045124	0.071631	0.020758	-0.115568	0.012798
496	0.011383	-0.031248	-0.036010	0.032164	-0.107343	-0.045003	0.070823
497	-0.044145	-0.022006	-0.038821	-0.119032	0.070035	-0.016466	-0.017405
498	-0.057032	0.042631	-0.026448	-0.003339	0.052676	-0.078402	0.036407
499	0.030194	0.006911	0.086152	-0.073459	-0.062670	-0.048895	-0.117928

	X7	X8	X9	X10	X11	X12	X13
0	-0.053554	-0.041735	0.039170	0.059882	0.042448	0.025450	efectores
1	-0.013180	-0.014414	0.075782	0.012873	0.028094	-0.048032	efectores
2	0.049338	0.018123	-0.040095	0.041548	0.055932	0.067426	efectores
3	-0.048188	0.084097	0.053476	-0.050901	-0.037147	0.000811	efectores
4	0.024522	-0.058602	0.004804	0.006654	-0.006080	0.057311	efectores
..
495	-0.048309	-0.016352	-0.190352	0.122769	-0.135269	-0.025126	efectores
496	0.073061	-0.057323	0.034942	0.008197	0.085763	-0.104814	efectores
497	-0.031513	-0.002641	0.078650	-0.007854	-0.032417	0.033197	efectores
498	0.017665	0.116583	0.017165	0.004276	0.001805	-0.052195	efectores
499	-0.035858	0.043065	0.071240	0.028677	0.023344	0.087320	efectores

[459 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) mass efectores archaea dataset 3, sin valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000
mean	0.005936	0.020190	0.012501	0.019932	-0.005535	0.000615
std	0.061918	0.068677	0.063820	0.061326	0.065117	0.064214
min	-0.189801	-0.191380	-0.203654	-0.160931	-0.185612	-0.186428
25%	-0.031145	-0.021819	-0.028587	-0.016690	-0.046846	-0.041980
50%	0.009043	0.018140	0.010403	0.023328	-0.006678	0.001879
75%	0.043818	0.065747	0.053940	0.062961	0.035059	0.040311

max	0.187354	0.220448	0.218388	0.186052	0.195969	0.195392
-----	----------	----------	----------	----------	----------	----------

	X6	X7	X8	X9	X10	X11 \
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000
mean	0.015317	0.002738	-0.000647	0.003221	0.006487	-0.008057
std	0.065936	0.061747	0.067429	0.066212	0.062442	0.061534
min	-0.171652	-0.196812	-0.205515	-0.195717	-0.205868	-0.178529
25%	-0.028883	-0.034375	-0.044707	-0.036622	-0.033161	-0.048210
50%	0.013546	0.003131	-0.001346	0.001404	0.006350	-0.005488
75%	0.057608	0.042570	0.043492	0.045487	0.046337	0.028773
max	0.191644	0.190355	0.204652	0.207193	0.183591	0.181912

	X12
count	459.000000
mean	0.013220
std	0.063657
min	-0.187702
25%	-0.031263
50%	0.011827
75%	0.056628
max	0.173780

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 3, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.025169	-0.017296	-0.042496	0.049787	-0.085673	-0.017847	0.056794
1	-0.084197	-0.062985	0.042907	-0.083224	0.198237	-0.126748	-0.220461
2	0.075630	0.086890	0.062427	0.047605	0.112470	0.108294	0.025755
3	0.108840	0.148543	-0.054125	0.048530	-0.048280	0.021827	-0.123771
4	0.099334	0.088008	0.081632	0.117466	0.032784	0.050273	-0.015217
..
495	0.168900	0.138382	0.035177	-0.021085	0.089214	0.050359	0.069464
496	0.153612	0.137255	-0.117263	-0.065715	0.057052	-0.050742	0.136148
497	0.046857	-0.013265	0.105587	0.058155	0.001258	-0.019465	0.018188
498	0.053520	-0.005163	0.022098	0.084011	0.001717	0.049541	0.076803
499	-0.032584	0.089332	-0.056060	0.050268	-0.062538	0.027189	0.006430

	X7	X8	X9	X10	X11	X12	X13
0	0.031925	-0.026479	0.015174	0.054948	0.003791	-0.008131	no_efectores
1	0.098511	-0.167152	0.084196	0.002181	0.057468	0.214752	no_efectores
2	0.048789	0.051203	0.103743	0.043385	0.075028	0.073661	no_efectores
3	0.119873	-0.005985	-0.010499	-0.129302	0.022737	-0.047830	no_efectores
4	0.058111	-0.042058	-0.032770	0.028078	-0.054996	-0.006981	no_efectores
..

```

495  0.005489 -0.026977  0.008763 -0.090305 -0.030215 -0.081442 no_efectores
496 -0.022779  0.037531 -0.095281 -0.032210 -0.015361  0.011746 no_efectores
497 -0.020544  0.046007  0.055894  0.012101 -0.061160  0.045674 no_efectores
498  0.039555  0.027703 -0.037534 -0.006177  0.019143  0.039990 no_efectores
499  0.098874 -0.105017 -0.035898 -0.009260 -0.081068  0.061638 no_efectores

```

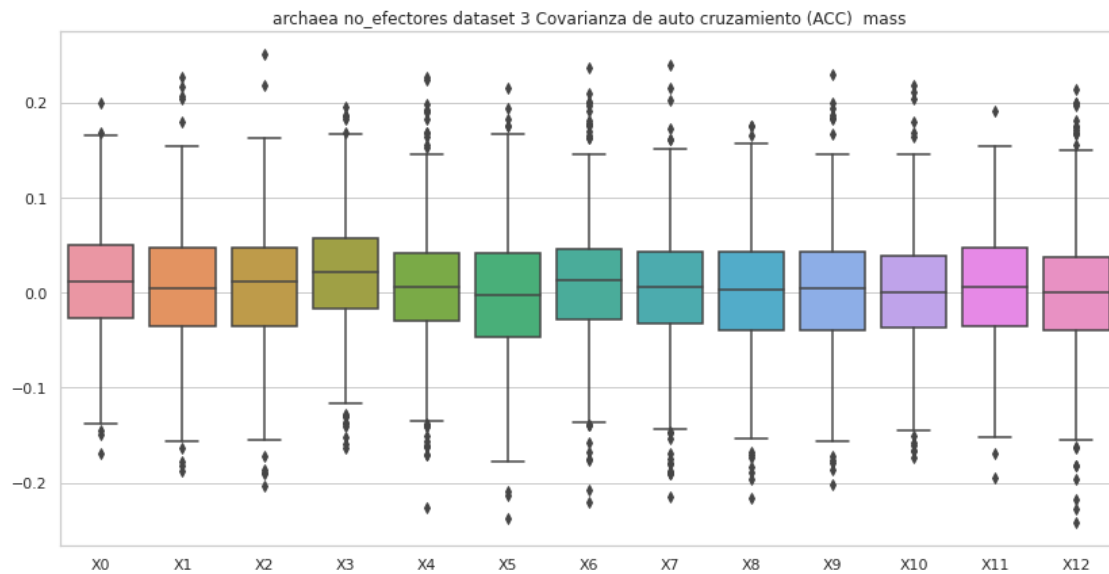
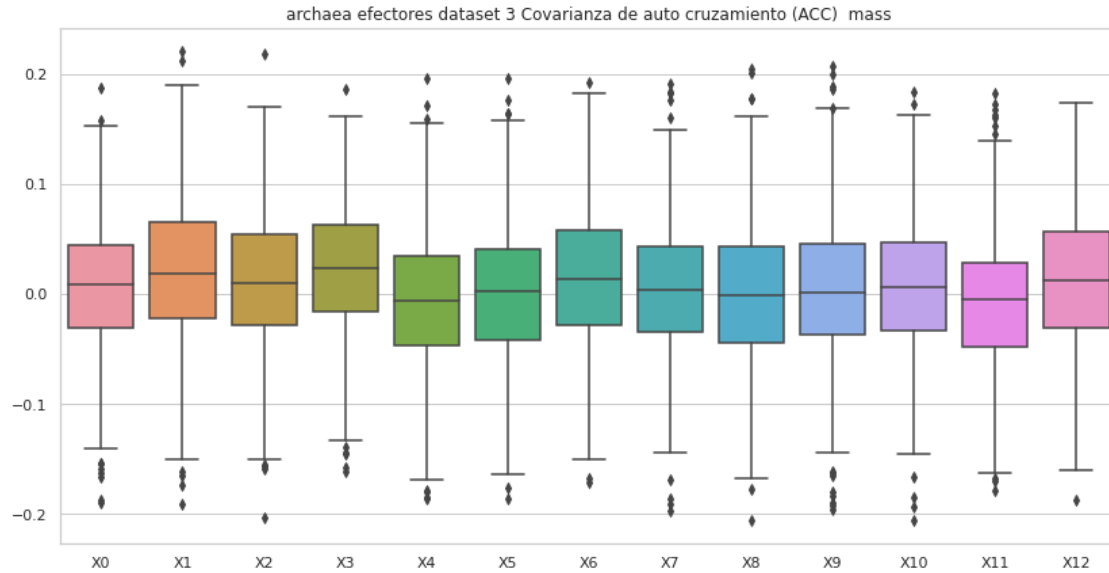
[452 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000
mean	0.010903	0.005777	0.008264	0.021231	0.006277	-0.003670
std	0.060097	0.066636	0.066186	0.059707	0.066549	0.068617
min	-0.168802	-0.187604	-0.202767	-0.163079	-0.226212	-0.236889
25%	-0.027143	-0.034794	-0.035291	-0.016244	-0.029643	-0.046834
50%	0.011133	0.005002	0.011582	0.021220	0.005985	-0.001989
75%	0.050895	0.048043	0.047347	0.057670	0.041726	0.041805
max	0.200352	0.227570	0.251125	0.195907	0.226392	0.214951

	X6	X7	X8	X9	X10	X11 \
count	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000
mean	0.009332	0.003813	0.000408	0.003177	0.001577	0.002877
std	0.067531	0.066938	0.065739	0.064207	0.061899	0.065400
min	-0.220461	-0.214794	-0.216761	-0.201672	-0.173166	-0.195505
25%	-0.027599	-0.032018	-0.038899	-0.039409	-0.036218	-0.035342
50%	0.012634	0.005539	0.002731	0.005219	0.000764	0.005440
75%	0.045875	0.043556	0.042714	0.042665	0.038730	0.047526
max	0.236571	0.240307	0.175653	0.230281	0.218007	0.190981

	X12
count	452.000000
mean	-0.000364
std	0.069390
min	-0.241842
25%	-0.039695
50%	0.000521
75%	0.037843
max	0.214752



8 Covarianza de auto cruzamiento (ACC) hidro

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

```

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

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

```

efectores

Covarianza de auto cruzamiento (ACC) hidro efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.001334	-0.078115	0.065274	0.059054	-0.148289	0.020244	-0.066984
1	-0.209555	-0.165094	0.147335	0.109168	-0.132247	-0.013569	0.053576
2	0.008769	-0.213234	-0.002806	0.128194	-0.084395	-0.096087	0.141392
3	0.031119	0.011840	0.002290	0.084728	0.090296	0.085032	-0.005594
4	-0.028848	0.118231	-0.197788	-0.077965	-0.162772	0.219492	-0.080919
..
495	0.086541	0.105521	0.169713	0.030004	-0.054771	-0.017605	-0.121119
496	-0.110753	-0.126949	-0.009137	0.036818	-0.060854	-0.001347	0.112616
497	-0.010647	-0.285446	0.014826	0.054012	-0.170633	0.002087	0.030357
498	-0.060631	-0.073685	-0.017376	0.043008	-0.003854	-0.093078	0.072656
499	0.004729	0.060127	0.024806	0.011272	0.004416	-0.047722	-0.019214

	X7	X8	X9	X10	X11	X12	X13
0	0.106300	0.023244	0.102153	0.042453	-0.025367	-0.076734	efectores
1	-0.080196	0.143688	-0.049210	0.030079	0.054003	0.136884	efectores
2	0.091584	-0.060633	-0.053605	0.058071	0.021578	0.006033	efectores

```

3      0.003045 -0.050264 -0.036209 -0.042310 -0.030219 -0.067083  efectores
4      0.099835  0.050105 -0.006482 -0.005081  0.121121 -0.073096  efectores
..      ...      ...      ...      ...      ...      ...
495 -0.037521 -0.089880 -0.090613 -0.131195 -0.036657 -0.019210  efectores
496  0.026639  0.080214  0.021824 -0.067773 -0.043500 -0.031318  efectores
497 -0.018544 -0.043215  0.183990  0.048264 -0.090919 -0.012413  efectores
498  0.037365 -0.141983  0.064678  0.058750 -0.044585 -0.108857  efectores
499 -0.018877 -0.060200 -0.119911 -0.093673 -0.019868 -0.004780  efectores

```

[500 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro efectores archaea dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.021401	-0.026926	0.041916	0.033994	-0.019726	-0.012588
std	0.089031	0.098723	0.083975	0.090308	0.098106	0.079888
min	-0.292209	-0.290592	-0.240157	-0.312581	-0.403740	-0.282591
25%	-0.033930	-0.094285	-0.013188	-0.019469	-0.084609	-0.060709
50%	0.026522	-0.019985	0.035865	0.035725	-0.010409	-0.014515
75%	0.077309	0.043629	0.098666	0.084484	0.047129	0.037607
max	0.262426	0.326408	0.344399	0.330617	0.262869	0.258743

	X6	X7	X8	X9	X10	X11 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.028898	0.025645	0.000720	-0.007551	0.015198	0.009653
std	0.091533	0.085768	0.086463	0.089104	0.077240	0.080689
min	-0.214626	-0.334885	-0.312991	-0.313329	-0.152389	-0.240758
25%	-0.027674	-0.017730	-0.051248	-0.054274	-0.035720	-0.041865
50%	0.018916	0.025464	0.005452	-0.006572	0.008970	0.000679
75%	0.072981	0.073035	0.049376	0.036614	0.058929	0.051986
max	0.364051	0.274499	0.234573	0.254200	0.352269	0.329906

	X12
count	500.000000
mean	-0.016581
std	0.082321
min	-0.314152
25%	-0.068389
50%	-0.015795
75%	0.033326
max	0.348730

no_efectores

Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 3, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.049151	-0.100241	0.010074	-0.013898	0.058196	0.002106	0.055131
1	-0.075482	-0.092486	-0.009951	-0.002864	0.160777	0.053473	-0.084158
2	0.120620	0.060101	0.119863	0.112958	0.071121	0.013496	0.157675
3	-0.201170	0.031225	-0.147973	0.159545	-0.027315	-0.000714	-0.024387
4	0.083268	-0.046766	0.105757	0.129276	-0.039703	-0.059687	0.087720
..	
495	-0.002985	-0.067187	0.012568	-0.027409	0.092101	0.020883	-0.018056
496	0.079993	-0.075394	-0.045645	0.081878	0.114053	-0.069825	-0.048182
497	-0.029289	-0.108074	0.078426	0.073549	-0.064011	0.018394	0.092740
498	-0.065037	-0.060497	-0.038701	-0.008915	0.012496	0.022432	0.026583
499	-0.039762	-0.120202	0.058116	-0.107267	-0.054090	-0.063133	-0.046807

	X7	X8	X9	X10	X11	X12	X13
0	-0.019627	0.010769	0.018364	0.043057	0.111761	-0.000999	no_efectores
1	0.217777	0.095803	0.099665	-0.165401	-0.101683	-0.057392	no_efectores
2	0.059579	0.073254	0.087826	0.047967	0.044249	0.027650	no_efectores
3	0.136956	-0.034894	-0.082392	0.005693	0.005521	-0.100569	no_efectores
4	0.035379	-0.021442	-0.033629	0.120359	0.075972	-0.012803	no_efectores
..	
495	0.012292	0.014216	0.016858	-0.009087	-0.096624	-0.015924	no_efectores
496	-0.042782	-0.061544	-0.024841	0.032116	-0.007212	-0.127265	no_efectores
497	-0.088081	0.082774	0.047387	-0.031595	-0.001052	0.050766	no_efectores
498	0.022621	-0.108113	-0.034716	0.048208	0.054849	0.037530	no_efectores
499	-0.024912	-0.033091	0.068117	0.037507	-0.095027	0.038437	no_efectores

[500 rows x 14 columns]

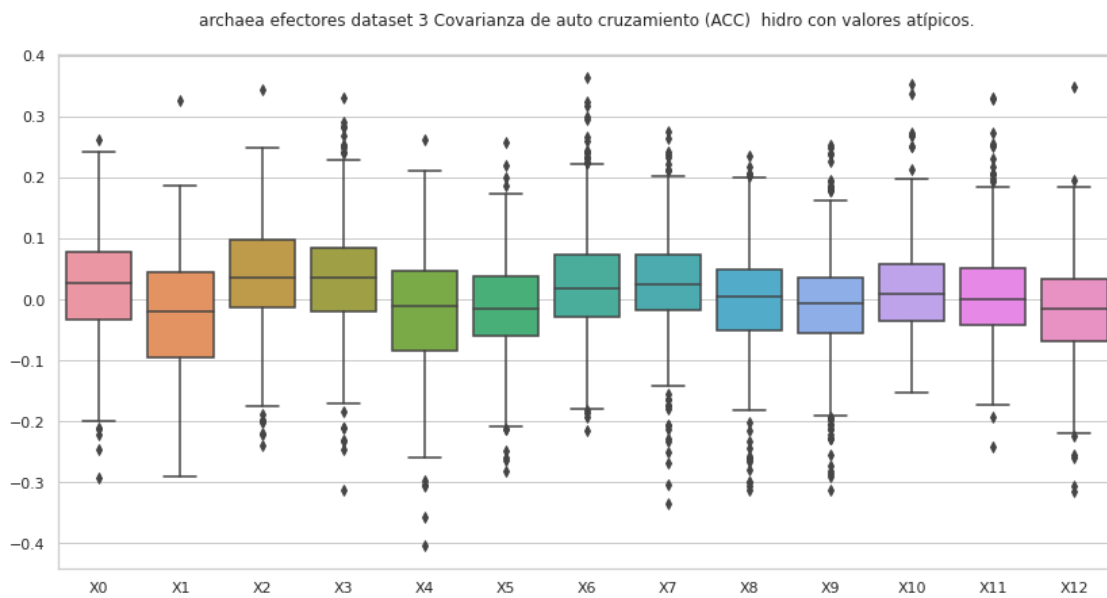
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 3, con valores atípicos.

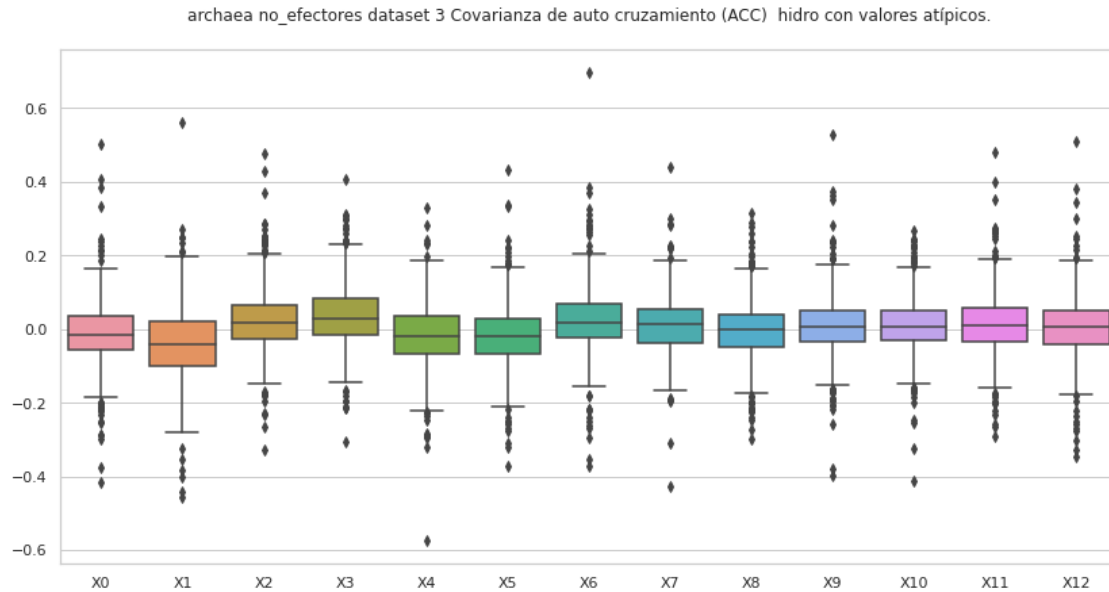
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	-0.013742	-0.036812	0.023240	0.032558	-0.021451	-0.018080
std	0.093007	0.103915	0.088464	0.085782	0.091950	0.091924
min	-0.416409	-0.456392	-0.327424	-0.306730	-0.574279	-0.372412
25%	-0.057767	-0.102168	-0.028766	-0.016730	-0.069039	-0.066485
50%	-0.014454	-0.040637	0.015883	0.028316	-0.019045	-0.019511
75%	0.035357	0.020175	0.064020	0.082574	0.033769	0.029657
max	0.500234	0.559615	0.475067	0.407495	0.330994	0.433255

	X6	X7	X8	X9	X10	X11 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.024012	0.011414	-0.003206	0.009424	0.010159	0.013148
std	0.096193	0.081884	0.084367	0.085851	0.079836	0.088731
min	-0.371928	-0.425739	-0.299911	-0.399504	-0.410964	-0.289752
25%	-0.021789	-0.036309	-0.049030	-0.034136	-0.032348	-0.032928
50%	0.018188	0.012695	-0.002301	0.005948	0.007298	0.010733
75%	0.069558	0.054583	0.038097	0.051628	0.048341	0.056509
max	0.696166	0.438905	0.313150	0.526176	0.266139	0.480935

	X12
count	500.000000
mean	0.004293
std	0.087090
min	-0.347745
25%	-0.040137
50%	0.004912
75%	0.050569
max	0.509297





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) +
      '\n' + 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)+"\n
→"+str(transf)+" "+str(comp))

```

efectores

Covarianza de auto cruzamiento (ACC) efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.001334	-0.078115	0.065274	0.059054	-0.148289	0.020244	-0.066984
1	-0.209555	-0.165094	0.147335	0.109168	-0.132247	-0.013569	0.053576
2	0.008769	-0.213234	-0.002806	0.128194	-0.084395	-0.096087	0.141392
3	0.031119	0.011840	0.002290	0.084728	0.090296	0.085032	-0.005594
4	-0.028848	0.118231	-0.197788	-0.077965	-0.162772	0.219492	-0.080919
..
495	0.086541	0.105521	0.169713	0.030004	-0.054771	-0.017605	-0.121119
496	-0.110753	-0.126949	-0.009137	0.036818	-0.060854	-0.001347	0.112616
497	-0.010647	-0.285446	0.014826	0.054012	-0.170633	0.002087	0.030357
498	-0.060631	-0.073685	-0.017376	0.043008	-0.003854	-0.093078	0.072656
499	0.004729	0.060127	0.024806	0.011272	0.004416	-0.047722	-0.019214
	X7	X8	X9	X10	X11	X12	X13
0	0.106300	0.023244	0.102153	0.042453	-0.025367	-0.076734	efectores
1	-0.080196	0.143688	-0.049210	0.030079	0.054003	0.136884	efectores
2	0.091584	-0.060633	-0.053605	0.058071	0.021578	0.006033	efectores
3	0.003045	-0.050264	-0.036209	-0.042310	-0.030219	-0.067083	efectores
4	0.099835	0.050105	-0.006482	-0.005081	0.121121	-0.073096	efectores
..
495	-0.037521	-0.089880	-0.090613	-0.131195	-0.036657	-0.019210	efectores

```

496  0.026639  0.080214  0.021824 -0.067773 -0.043500 -0.031318  efectores
497 -0.018544 -0.043215  0.183990  0.048264 -0.090919 -0.012413  efectores
498  0.037365 -0.141983  0.064678  0.058750 -0.044585 -0.108857  efectores
499 -0.018877 -0.060200 -0.119911 -0.093673 -0.019868 -0.004780  efectores

```

[459 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) efectores archaea dataset 3, sin valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000	
mean	0.020854	-0.024513	0.040349	0.033874	-0.013804	-0.010133	
std	0.085718	0.093913	0.077879	0.086087	0.091382	0.073551	
min	-0.245063	-0.290592	-0.202350	-0.232087	-0.306600	-0.248520	
25%	-0.033696	-0.089430	-0.013140	-0.017926	-0.076472	-0.056390	
50%	0.025840	-0.015173	0.033361	0.035717	-0.006079	-0.011412	
75%	0.070053	0.043673	0.093412	0.083869	0.047984	0.037819	
max	0.262426	0.179859	0.244339	0.291520	0.262869	0.219492	

	X6	X7	X8	X9	X10	X11	\
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000	
mean	0.024544	0.025371	0.005397	-0.006415	0.010218	0.007375	
std	0.083992	0.077535	0.073623	0.081130	0.068094	0.070931	
min	-0.214626	-0.229219	-0.256282	-0.271892	-0.152389	-0.173090	
25%	-0.027023	-0.017875	-0.046321	-0.053775	-0.037044	-0.040049	
50%	0.017992	0.024209	0.007086	-0.005091	0.005622	0.000551	
75%	0.065546	0.065669	0.048094	0.035171	0.049025	0.049346	
max	0.299415	0.242999	0.217579	0.239017	0.212576	0.231501	

	X12
count	459.000000
mean	-0.015088
std	0.074771
min	-0.223940
25%	-0.065277
50%	-0.013317
75%	0.032185
max	0.184597

no_efectores

Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.049151	-0.100241	0.010074	-0.013898	0.058196	0.002106	0.055131
1	-0.075482	-0.092486	-0.009951	-0.002864	0.160777	0.053473	-0.084158
2	0.120620	0.060101	0.119863	0.112958	0.071121	0.013496	0.157675
3	-0.201170	0.031225	-0.147973	0.159545	-0.027315	-0.000714	-0.024387
4	0.083268	-0.046766	0.105757	0.129276	-0.039703	-0.059687	0.087720
..	
495	-0.002985	-0.067187	0.012568	-0.027409	0.092101	0.020883	-0.018056
496	0.079993	-0.075394	-0.045645	0.081878	0.114053	-0.069825	-0.048182
497	-0.029289	-0.108074	0.078426	0.073549	-0.064011	0.018394	0.092740
498	-0.065037	-0.060497	-0.038701	-0.008915	0.012496	0.022432	0.026583
499	-0.039762	-0.120202	0.058116	-0.107267	-0.054090	-0.063133	-0.046807

	X7	X8	X9	X10	X11	X12	X13
0	-0.019627	0.010769	0.018364	0.043057	0.111761	-0.000999	no_efectores
1	0.217777	0.095803	0.099665	-0.165401	-0.101683	-0.057392	no_efectores
2	0.059579	0.073254	0.087826	0.047967	0.044249	0.027650	no_efectores
3	0.136956	-0.034894	-0.082392	0.005693	0.005521	-0.100569	no_efectores
4	0.035379	-0.021442	-0.033629	0.120359	0.075972	-0.012803	no_efectores
..	
495	0.012292	0.014216	0.016858	-0.009087	-0.096624	-0.015924	no_efectores
496	-0.042782	-0.061544	-0.024841	0.032116	-0.007212	-0.127265	no_efectores
497	-0.088081	0.082774	0.047387	-0.031595	-0.001052	0.050766	no_efectores
498	0.022621	-0.108113	-0.034716	0.048208	0.054849	0.037530	no_efectores
499	-0.024912	-0.033091	0.068117	0.037507	-0.095027	0.038437	no_efectores

[460 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	460.000000	460.000000	460.000000	460.000000	460.000000	460.000000
mean	-0.012059	-0.036863	0.019777	0.027822	-0.018397	-0.019377
std	0.077238	0.088847	0.076545	0.075849	0.079612	0.078444
min	-0.289243	-0.323739	-0.231135	-0.215665	-0.291030	-0.270981
25%	-0.051833	-0.100383	-0.028766	-0.016992	-0.065067	-0.065997
50%	-0.011078	-0.040637	0.014772	0.026373	-0.017512	-0.019616
75%	0.035184	0.017039	0.058780	0.079261	0.033249	0.029185
max	0.244612	0.268785	0.284829	0.280446	0.239957	0.242219

	X6	X7	X8	X9	X10	X11 \
count	460.000000	460.000000	460.000000	460.000000	460.000000	460.000000
mean	0.020319	0.009053	-0.003756	0.006236	0.011323	0.009375
std	0.074408	0.068010	0.071375	0.066122	0.068831	0.073433

min	-0.262414	-0.189537	-0.246474	-0.218191	-0.174614	-0.222724
25%	-0.020666	-0.035668	-0.046897	-0.033056	-0.030448	-0.032273
50%	0.017655	0.012474	-0.002301	0.005426	0.006805	0.008872
75%	0.060378	0.051719	0.037507	0.043864	0.045148	0.052616
max	0.294333	0.223943	0.237298	0.223307	0.244237	0.272724

	X12
count	460.000000
mean	0.005751
std	0.071549
min	-0.250354
25%	-0.037296
50%	0.005263
75%	0.048861
max	0.251428

