

ds4_archaea_limpieza_de_datos

February 1, 2021

Limpieza de datos

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

1 Declaración de variables

```
[2]: organismo = "archaea"
dataset = 4
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo) + " dataset " + str(dataset))
r2 = ("Datos/resultados/" + str(organismo) + "/" + str(nombre) + "/"
      ↪ transformaciones/sin_filtrar")
r3 = ("Datos/resultados/" + str(organismo) + "/" + str(nombre) + "/"
      ↪ transformaciones/sin_atipicos")

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

```

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

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

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

```

2 Composición de aminoácidos (AAC)

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

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

        if etiq == "efectores":
            df=AAC_efec

        if etiq == "no_efectores":
            df=AAC_no_efec

        #del df['X20']
        print (str(titulo) + "Valores del documento csv.\n")
        print (df)
        print ("\n\n" + str(titulo) + "Estadísticas.\n")
        print(df.describe())
        print ("\n\n")

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

efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | \ |
|-----|--------|-------|-------|--------|-------|--------|-------|--------|-------|---|
| 0 | 10.920 | 5.172 | 3.161 | 3.161 | 0.000 | 7.184 | 4.023 | 6.322 | 0.575 | |
| 1 | 12.000 | 6.286 | 0.000 | 5.143 | 0.571 | 3.429 | 1.143 | 9.143 | 2.857 | |
| 2 | 5.988 | 2.994 | 2.395 | 3.593 | 1.198 | 10.778 | 1.796 | 4.192 | 2.395 | |
| 3 | 15.217 | 3.986 | 2.536 | 6.159 | 0.000 | 5.072 | 1.087 | 7.971 | 2.174 | |
| 4 | 15.278 | 5.556 | 0.694 | 6.481 | 0.231 | 4.630 | 2.315 | 12.037 | 0.694 | |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 995 | 9.827 | 8.671 | 2.312 | 2.890 | 0.578 | 15.029 | 1.734 | 7.514 | 1.734 | |
| 996 | 9.350 | 8.537 | 2.439 | 7.724 | 0.813 | 8.943 | 2.846 | 2.846 | 2.439 | |
| 997 | 5.667 | 5.667 | 4.667 | 7.667 | 0.667 | 9.667 | 1.333 | 2.333 | 2.333 | |
| 998 | 6.364 | 7.727 | 1.818 | 3.636 | 0.000 | 8.182 | 4.091 | 6.818 | 1.364 | |
| 999 | 8.726 | 7.783 | 2.358 | 13.443 | 0.943 | 10.613 | 3.538 | 8.726 | 3.066 | |

| | X9 | ... | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | \ |
|-----|--------|-----|-------|-------|-------|-------|-------|-------|-------|-------|---|
| 0 | 7.759 | ... | 1.149 | 1.149 | 5.460 | 3.448 | 7.471 | 7.471 | 1.149 | 2.874 | |
| 1 | 4.000 | ... | 1.143 | 1.714 | 4.000 | 6.857 | 3.429 | 4.571 | 4.000 | 2.857 | |
| 2 | 8.383 | ... | 8.982 | 3.593 | 2.395 | 4.790 | 7.186 | 5.389 | 0.599 | 4.192 | |
| 3 | 6.159 | ... | 0.000 | 2.174 | 5.797 | 3.986 | 4.710 | 4.710 | 1.087 | 2.899 | |
| 4 | 3.472 | ... | 1.157 | 1.389 | 3.935 | 3.704 | 4.398 | 3.935 | 0.926 | 3.472 | |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 995 | 6.936 | ... | 6.358 | 3.468 | 4.046 | 1.156 | 4.624 | 3.468 | 0.578 | 4.624 | |
| 996 | 6.098 | ... | 1.626 | 0.407 | 2.439 | 6.098 | 9.350 | 5.691 | 1.220 | 4.472 | |
| 997 | 6.333 | ... | 9.333 | 5.000 | 4.667 | 1.667 | 8.333 | 4.667 | 0.000 | 4.333 | |
| 998 | 11.364 | ... | 6.364 | 2.727 | 2.273 | 3.636 | 5.455 | 3.636 | 0.000 | 3.636 | |
| 999 | 3.066 | ... | 0.708 | 0.472 | 2.358 | 4.481 | 3.774 | 5.896 | 1.179 | 2.123 | |

| | X19 | X20 |
|-----|--------|-----------|
| 0 | 8.333 | efectores |
| 1 | 13.714 | efectores |
| 2 | 7.784 | efectores |
| 3 | 10.145 | efectores |
| 4 | 12.500 | efectores |
| .. | ... | ... |
| 995 | 3.468 | efectores |
| 996 | 7.724 | efectores |
| 997 | 7.667 | efectores |
| 998 | 7.727 | efectores |
| 999 | 10.142 | efectores |

[1000 rows x 21 columns]

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

Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 9.192813 | 6.029755 | 2.722955 | 6.087815 | 0.633708 | |
| std | 4.302138 | 2.719750 | 2.064523 | 2.747029 | 0.877511 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 5.744750 | 4.046250 | 1.282000 | 3.906000 | 0.000000 | |
| 50% | 8.900000 | 5.795500 | 2.273000 | 5.718500 | 0.351000 | |
| 75% | 12.091000 | 7.856000 | 3.769250 | 7.804750 | 0.887000 | |
| max | 24.183000 | 15.873000 | 14.650000 | 15.385000 | 6.612000 | |

| | X5 | X6 | X7 | X8 | X9 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 7.084187 | 2.299056 | 7.570299 | 1.829916 | 5.903947 | |
| std | 3.779742 | 1.484329 | 2.933412 | 1.255974 | 3.398320 | |

| | | | | | |
|-----|-----------|-----------|-----------|----------|-----------|
| min | 0.000000 | 0.000000 | 0.562000 | 0.000000 | 0.000000 |
| 25% | 4.069000 | 1.299000 | 5.343500 | 0.787000 | 3.359750 |
| 50% | 6.944000 | 2.128000 | 7.385500 | 1.712000 | 5.339500 |
| 75% | 9.762750 | 3.101000 | 9.547750 | 2.644500 | 7.919750 |
| max | 20.787000 | 11.475000 | 19.215000 | 6.667000 | 22.449000 |

| | | | | | |
|-------|-------------|-------------|-------------|-------------|-------------|
| | X10 | X11 | X12 | X13 | X14 \ |
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 11.414362 | 4.436505 | 2.080056 | 3.956146 | 4.101346 |
| std | 3.379396 | 4.364359 | 1.241599 | 2.046636 | 1.854089 |
| min | 0.957000 | 0.000000 | 0.301000 | 0.000000 | 0.000000 |
| 25% | 8.995750 | 1.034000 | 1.128500 | 2.580500 | 2.978250 |
| 50% | 11.212000 | 2.463000 | 1.778000 | 3.846000 | 3.916500 |
| 75% | 13.592000 | 7.193500 | 2.703000 | 4.903500 | 5.128000 |
| max | 26.531000 | 23.494000 | 7.500000 | 13.402000 | 15.979000 |

| | | | | | |
|-------|-------------|-------------|-------------|-------------|-------------|
| | X15 | X16 | X17 | X18 | X19 |
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 5.905036 | 5.482102 | 1.249702 | 3.257625 | 8.762695 |
| std | 2.164733 | 2.085338 | 1.046206 | 1.572519 | 3.426575 |
| min | 0.826000 | 0.000000 | 0.000000 | 0.000000 | 0.645000 |
| 25% | 4.412000 | 4.065250 | 0.549250 | 2.222000 | 6.215250 |
| 50% | 5.729000 | 5.455000 | 1.097000 | 3.125000 | 8.368000 |
| 75% | 7.143000 | 6.684500 | 1.724000 | 4.073000 | 11.153500 |
| max | 15.248000 | 13.661000 | 6.417000 | 10.714000 | 18.156000 |

no_efectores

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

Valores del documento csv.

| | | | | | | | | | |
|-----|--------|--------|-------|--------|-------|--------|-------|--------|-------|
| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 \ |
| 0 | 6.102 | 2.756 | 4.528 | 3.150 | 0.591 | 4.134 | 2.165 | 7.480 | 0.197 |
| 1 | 10.288 | 2.058 | 3.086 | 3.086 | 1.440 | 3.704 | 2.058 | 10.494 | 1.440 |
| 2 | 8.889 | 6.667 | 2.222 | 0.000 | 8.889 | 8.889 | 4.444 | 4.444 | 0.000 |
| 3 | 6.522 | 3.261 | 6.522 | 4.348 | 2.174 | 3.261 | 3.261 | 6.522 | 0.000 |
| 4 | 14.493 | 20.290 | 0.000 | 10.145 | 5.797 | 2.899 | 0.000 | 5.797 | 1.449 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 995 | 10.400 | 4.400 | 2.400 | 0.800 | 0.400 | 3.600 | 2.000 | 10.400 | 1.200 |
| 996 | 6.335 | 8.145 | 4.525 | 1.357 | 0.000 | 12.670 | 1.357 | 3.620 | 0.905 |
| 997 | 6.250 | 6.250 | 2.083 | 0.000 | 0.000 | 2.083 | 2.083 | 6.250 | 0.000 |
| 998 | 4.464 | 8.929 | 2.679 | 8.929 | 0.893 | 12.500 | 4.464 | 8.036 | 1.786 |
| 999 | 9.328 | 8.209 | 2.239 | 7.836 | 0.373 | 10.448 | 1.493 | 6.716 | 3.731 |

| | | | | | | | | | | |
|---|--------|-----|-------|-------|-------|-------|--------|-------|-------|-------|
| | X9 | ... | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 \ |
| 0 | 11.811 | ... | 1.969 | 1.575 | 5.906 | 3.937 | 10.236 | 4.724 | 0.787 | 4.724 |

| | | | | | | | | | | |
|-----|--------|-----|--------|-------|-------|--------|--------|-------|-------|-------|
| 1 | 11.728 | ... | 4.733 | 3.292 | 8.025 | 1.440 | 7.407 | 5.556 | 1.235 | 2.263 |
| 2 | 2.222 | ... | 0.000 | 2.222 | 6.667 | 11.111 | 11.111 | 8.889 | 0.000 | 2.222 |
| 3 | 8.696 | ... | 9.783 | 4.348 | 2.174 | 6.522 | 11.957 | 4.348 | 0.000 | 2.174 |
| 4 | 0.000 | ... | 1.449 | 1.449 | 0.000 | 5.797 | 1.449 | 7.246 | 1.449 | 5.797 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 995 | 8.800 | ... | 0.400 | 0.400 | 7.600 | 6.000 | 5.200 | 7.200 | 0.800 | 1.200 |
| 996 | 9.502 | ... | 11.312 | 4.525 | 3.167 | 2.715 | 5.882 | 4.977 | 0.000 | 2.262 |
| 997 | 2.083 | ... | 2.083 | 2.083 | 4.167 | 10.417 | 6.250 | 8.333 | 0.000 | 0.000 |
| 998 | 4.464 | ... | 1.786 | 2.679 | 1.786 | 3.571 | 6.250 | 1.786 | 2.679 | 3.571 |
| 999 | 2.985 | ... | 2.239 | 2.239 | 4.104 | 7.090 | 5.224 | 4.478 | 0.000 | 2.239 |

| | | |
|-----|--------|--------------|
| | X19 | X20 |
| 0 | 7.874 | no_efectores |
| 1 | 9.671 | no_efectores |
| 2 | 4.444 | no_efectores |
| 3 | 6.522 | no_efectores |
| 4 | 7.246 | no_efectores |
| .. | ... | ... |
| 995 | 11.200 | no_efectores |
| 996 | 4.977 | no_efectores |
| 997 | 6.250 | no_efectores |
| 998 | 8.036 | no_efectores |
| 999 | 8.955 | no_efectores |

[1000 rows x 21 columns]

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

Estadísticas.

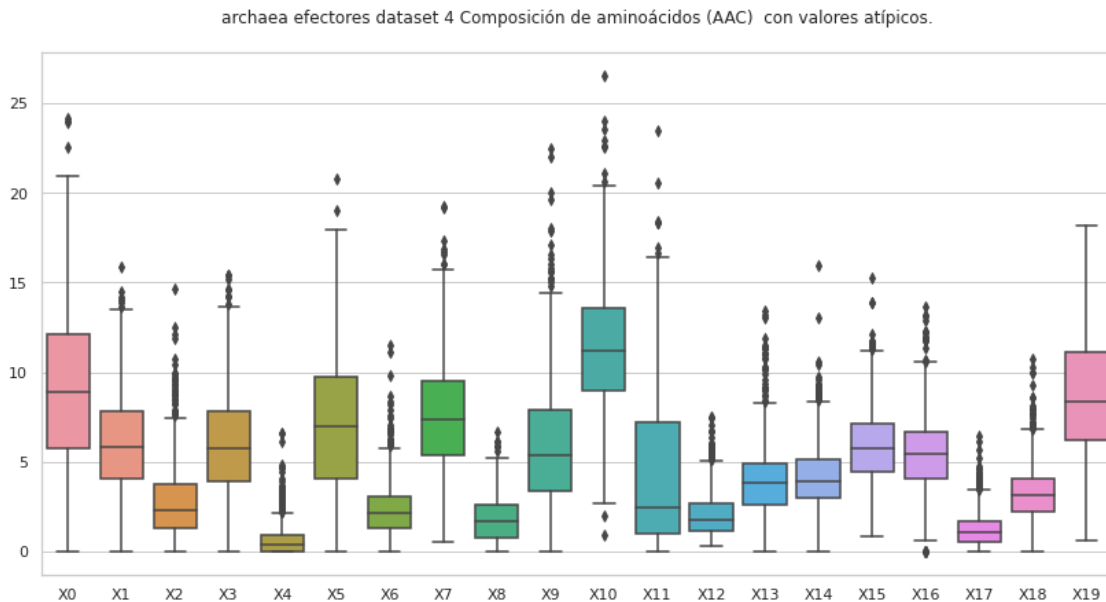
| | | | | | |
|-------|-------------|-------------|-------------|-------------|-------------|
| | X0 | X1 | X2 | X3 | X4 \ |
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 9.607305 | 6.341891 | 2.933450 | 7.580423 | 1.003971 |
| std | 3.932998 | 2.881207 | 2.060948 | 3.496744 | 1.448737 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 6.838000 | 4.396000 | 1.572750 | 5.048750 | 0.000000 |
| 50% | 9.494000 | 6.137500 | 2.500000 | 7.540000 | 0.593500 |
| 75% | 12.305000 | 7.991750 | 3.819000 | 10.041750 | 1.339000 |
| max | 24.242000 | 24.691000 | 12.987000 | 23.410000 | 13.725000 |

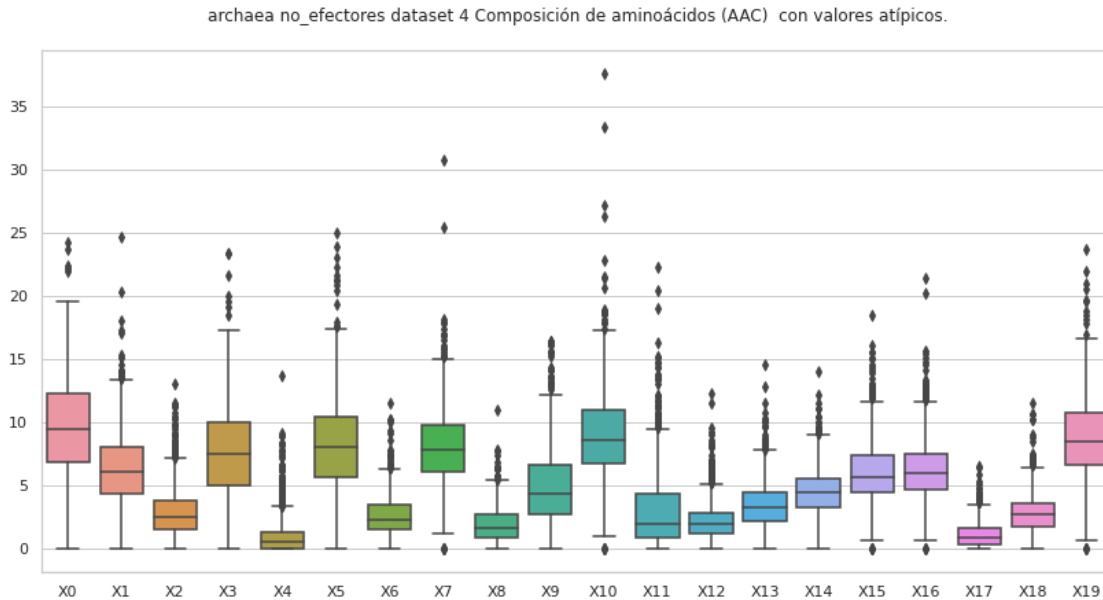
| | | | | | |
|-------|-------------|-------------|-------------|-------------|-------------|
| | X5 | X6 | X7 | X8 | X9 \ |
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 8.055466 | 2.583578 | 7.993374 | 1.875775 | 4.952474 |
| std | 3.793754 | 1.661973 | 3.045818 | 1.377914 | 3.097132 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 5.606000 | 1.486500 | 6.137750 | 0.885000 | 2.728250 |
| 50% | 8.048000 | 2.320500 | 7.843000 | 1.681500 | 4.295000 |

| | | | | | |
|-----|-----------|-----------|-----------|-----------|-----------|
| 75% | 10.400750 | 3.430000 | 9.775500 | 2.703000 | 6.646000 |
| max | 25.000000 | 11.538000 | 30.769000 | 11.000000 | 16.425000 |

| | X10 | X11 | X12 | X13 | X14 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 9.024047 | 3.077586 | 2.218904 | 3.493352 | 4.465421 |
| std | 3.616497 | 3.129549 | 1.436454 | 1.929197 | 1.887414 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 6.688000 | 0.916500 | 1.240000 | 2.155000 | 3.271750 |
| 50% | 8.541000 | 1.981500 | 1.923000 | 3.279000 | 4.416000 |
| 75% | 10.963500 | 4.348000 | 2.809000 | 4.414500 | 5.594000 |
| max | 37.580000 | 22.222000 | 12.308000 | 14.535000 | 14.000000 |

| | X15 | X16 | X17 | X18 | X19 |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 6.012027 | 6.207551 | 1.077021 | 2.812873 | 8.683499 |
| std | 2.488123 | 2.514168 | 1.011161 | 1.623734 | 3.172148 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 4.415000 | 4.695500 | 0.282000 | 1.708250 | 6.667000 |
| 50% | 5.615000 | 5.942000 | 0.905000 | 2.674500 | 8.424000 |
| 75% | 7.347000 | 7.520500 | 1.613000 | 3.614000 | 10.725250 |
| max | 18.471000 | 21.429000 | 6.522000 | 11.538000 | 23.636000 |





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 4, sin valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | \ |
|-----|--------|-------|-------|--------|-------|--------|-------|--------|-------|---|
| 0 | 10.920 | 5.172 | 3.161 | 3.161 | 0.000 | 7.184 | 4.023 | 6.322 | 0.575 | |
| 1 | 12.000 | 6.286 | 0.000 | 5.143 | 0.571 | 3.429 | 1.143 | 9.143 | 2.857 | |
| 2 | 5.988 | 2.994 | 2.395 | 3.593 | 1.198 | 10.778 | 1.796 | 4.192 | 2.395 | |
| 3 | 15.217 | 3.986 | 2.536 | 6.159 | 0.000 | 5.072 | 1.087 | 7.971 | 2.174 | |
| 4 | 15.278 | 5.556 | 0.694 | 6.481 | 0.231 | 4.630 | 2.315 | 12.037 | 0.694 | |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 995 | 9.827 | 8.671 | 2.312 | 2.890 | 0.578 | 15.029 | 1.734 | 7.514 | 1.734 | |
| 996 | 9.350 | 8.537 | 2.439 | 7.724 | 0.813 | 8.943 | 2.846 | 2.846 | 2.439 | |
| 997 | 5.667 | 5.667 | 4.667 | 7.667 | 0.667 | 9.667 | 1.333 | 2.333 | 2.333 | |
| 998 | 6.364 | 7.727 | 1.818 | 3.636 | 0.000 | 8.182 | 4.091 | 6.818 | 1.364 | |
| 999 | 8.726 | 7.783 | 2.358 | 13.443 | 0.943 | 10.613 | 3.538 | 8.726 | 3.066 | |

| | X9 | ... | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | \ |
|-----|-------|-----|-------|-------|-------|-------|-------|-------|-------|-------|---|
| 0 | 7.759 | ... | 1.149 | 1.149 | 5.460 | 3.448 | 7.471 | 7.471 | 1.149 | 2.874 | |
| 1 | 4.000 | ... | 1.143 | 1.714 | 4.000 | 6.857 | 3.429 | 4.571 | 4.000 | 2.857 | |
| 2 | 8.383 | ... | 8.982 | 3.593 | 2.395 | 4.790 | 7.186 | 5.389 | 0.599 | 4.192 | |
| 3 | 6.159 | ... | 0.000 | 2.174 | 5.797 | 3.986 | 4.710 | 4.710 | 1.087 | 2.899 | |
| 4 | 3.472 | ... | 1.157 | 1.389 | 3.935 | 3.704 | 4.398 | 3.935 | 0.926 | 3.472 | |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 995 | 6.936 | ... | 6.358 | 3.468 | 4.046 | 1.156 | 4.624 | 3.468 | 0.578 | 4.624 | |

| | | | | | | | | | | |
|-----|--------|-----|-------|-------|-------|-------|-------|-------|-------|-------|
| 996 | 6.098 | ... | 1.626 | 0.407 | 2.439 | 6.098 | 9.350 | 5.691 | 1.220 | 4.472 |
| 997 | 6.333 | ... | 9.333 | 5.000 | 4.667 | 1.667 | 8.333 | 4.667 | 0.000 | 4.333 |
| 998 | 11.364 | ... | 6.364 | 2.727 | 2.273 | 3.636 | 5.455 | 3.636 | 0.000 | 3.636 |
| 999 | 3.066 | ... | 0.708 | 0.472 | 2.358 | 4.481 | 3.774 | 5.896 | 1.179 | 2.123 |

| | | |
|-----|--------|-----------|
| | X19 | X20 |
| 0 | 8.333 | efectores |
| 1 | 13.714 | efectores |
| 2 | 7.784 | efectores |
| 3 | 10.145 | efectores |
| 4 | 12.500 | efectores |
| .. | ... | ... |
| 995 | 3.468 | efectores |
| 996 | 7.724 | efectores |
| 997 | 7.667 | efectores |
| 998 | 7.727 | efectores |
| 999 | 10.142 | efectores |

[861 rows x 21 columns]

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

Estadísticas.

| | | | | | | | |
|-------|------------|------------|------------|------------|------------|------------|---|
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
| count | 861.000000 | 861.000000 | 861.000000 | 861.000000 | 861.000000 | 861.000000 | |
| mean | 9.540681 | 6.238630 | 2.517122 | 6.169732 | 0.529315 | 7.163204 | |
| std | 4.183061 | 2.599574 | 1.778182 | 2.652858 | 0.675615 | 3.714595 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.752000 | 0.000000 | 0.000000 | |
| 25% | 6.250000 | 4.308000 | 1.195000 | 4.068000 | 0.000000 | 4.128000 | |
| 50% | 9.317000 | 5.970000 | 2.174000 | 5.755000 | 0.291000 | 7.143000 | |
| 75% | 12.422000 | 8.072000 | 3.492000 | 8.000000 | 0.794000 | 9.790000 | |
| max | 20.961000 | 14.062000 | 8.799000 | 14.286000 | 3.261000 | 17.949000 | |

| | | | | | | | |
|-------|------------|------------|------------|------------|------------|------------|---|
| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
| count | 861.000000 | 861.000000 | 861.000000 | 861.000000 | 861.000000 | 861.000000 | |
| mean | 2.231137 | 7.713334 | 1.846046 | 5.611020 | 11.457027 | 4.097503 | |
| std | 1.315407 | 2.818973 | 1.220323 | 3.058827 | 3.140800 | 4.075535 | |
| min | 0.000000 | 0.752000 | 0.000000 | 0.000000 | 1.961000 | 0.000000 | |
| 25% | 1.316000 | 5.556000 | 0.802000 | 3.226000 | 9.112000 | 0.907000 | |
| 50% | 2.105000 | 7.586000 | 1.724000 | 5.128000 | 11.337000 | 2.059000 | |
| 75% | 3.036000 | 9.701000 | 2.678000 | 7.450000 | 13.623000 | 6.818000 | |
| max | 6.725000 | 16.058000 | 5.195000 | 16.000000 | 21.111000 | 16.667000 | |

| | | | | | | | |
|-------|------------|------------|------------|------------|------------|------------|---|
| | X12 | X13 | X14 | X15 | X16 | X17 | \ |
| count | 861.000000 | 861.000000 | 861.000000 | 861.000000 | 861.000000 | 861.000000 | |
| mean | 2.007475 | 3.905904 | 4.131466 | 5.873009 | 5.500892 | 1.266868 | |

| | | | | | | |
|-----|----------|----------|----------|-----------|-----------|----------|
| std | 1.145994 | 1.731459 | 1.735383 | 2.076752 | 1.895299 | 0.999317 |
| min | 0.301000 | 0.000000 | 0.000000 | 0.826000 | 0.000000 | 0.000000 |
| 25% | 1.099000 | 2.655000 | 3.042000 | 4.444000 | 4.144000 | 0.602000 |
| 50% | 1.747000 | 3.906000 | 3.966000 | 5.714000 | 5.523000 | 1.147000 |
| 75% | 2.655000 | 4.895000 | 5.135000 | 7.054000 | 6.723000 | 1.754000 |
| max | 5.769000 | 9.929000 | 9.357000 | 12.097000 | 11.702000 | 4.375000 |

| | | |
|-------|------------|------------|
| | X18 | X19 |
| count | 861.000000 | 861.000000 |
| mean | 3.188093 | 9.011581 |
| std | 1.388830 | 3.386957 |
| min | 0.000000 | 0.926000 |
| 25% | 2.262000 | 6.494000 |
| 50% | 3.125000 | 8.661000 |
| 75% | 4.000000 | 11.392000 |
| max | 7.965000 | 18.018000 |

no_efectores

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

Valores del documento csv.

| | | | | | | | | | | |
|-----|--------|--------|-------|-------|-------|--------|-------|--------|-------|---|
| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | \ |
| 0 | 6.102 | 2.756 | 4.528 | 3.150 | 0.591 | 4.134 | 2.165 | 7.480 | 0.197 | |
| 1 | 10.288 | 2.058 | 3.086 | 3.086 | 1.440 | 3.704 | 2.058 | 10.494 | 1.440 | |
| 3 | 6.522 | 3.261 | 6.522 | 4.348 | 2.174 | 3.261 | 3.261 | 6.522 | 0.000 | |
| 5 | 9.829 | 10.043 | 2.350 | 9.402 | 0.000 | 10.256 | 1.282 | 6.197 | 2.991 | |
| 6 | 9.167 | 3.958 | 3.333 | 6.667 | 0.625 | 6.458 | 2.500 | 11.875 | 3.333 | |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 994 | 6.849 | 2.935 | 4.892 | 8.415 | 0.196 | 7.045 | 3.523 | 9.002 | 3.131 | |
| 995 | 10.400 | 4.400 | 2.400 | 0.800 | 0.400 | 3.600 | 2.000 | 10.400 | 1.200 | |
| 996 | 6.335 | 8.145 | 4.525 | 1.357 | 0.000 | 12.670 | 1.357 | 3.620 | 0.905 | |
| 998 | 4.464 | 8.929 | 2.679 | 8.929 | 0.893 | 12.500 | 4.464 | 8.036 | 1.786 | |
| 999 | 9.328 | 8.209 | 2.239 | 7.836 | 0.373 | 10.448 | 1.493 | 6.716 | 3.731 | |

| | | | | | | | | | | | |
|-----|--------|-----|--------|-------|-------|-------|--------|--------|-------|-------|---|
| | X9 | ... | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | \ |
| 0 | 11.811 | ... | 1.969 | 1.575 | 5.906 | 3.937 | 10.236 | 4.724 | 0.787 | 4.724 | |
| 1 | 11.728 | ... | 4.733 | 3.292 | 8.025 | 1.440 | 7.407 | 5.556 | 1.235 | 2.263 | |
| 3 | 8.696 | ... | 9.783 | 4.348 | 2.174 | 6.522 | 11.957 | 4.348 | 0.000 | 2.174 | |
| 5 | 3.632 | ... | 0.641 | 1.496 | 2.564 | 3.846 | 4.487 | 6.838 | 1.709 | 1.923 | |
| 6 | 3.125 | ... | 1.667 | 3.333 | 3.333 | 7.083 | 6.875 | 8.125 | 0.833 | 3.125 | |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 994 | 6.458 | ... | 1.174 | 0.783 | 1.370 | 4.110 | 11.937 | 12.524 | 0.783 | 3.718 | |
| 995 | 8.800 | ... | 0.400 | 0.400 | 7.600 | 6.000 | 5.200 | 7.200 | 0.800 | 1.200 | |
| 996 | 9.502 | ... | 11.312 | 4.525 | 3.167 | 2.715 | 5.882 | 4.977 | 0.000 | 2.262 | |
| 998 | 4.464 | ... | 1.786 | 2.679 | 1.786 | 3.571 | 6.250 | 1.786 | 2.679 | 3.571 | |

```

999    2.985    ...    2.239    2.239    4.104    7.090    5.224    4.478    0.000    2.239

```

```

          X19          X20
0      7.874    no_efectores
1      9.671    no_efectores
3      6.522    no_efectores
5     11.752    no_efectores
6     10.000    no_efectores
..      ...      ...
994    5.479    no_efectores
995   11.200    no_efectores
996    4.977    no_efectores
998    8.036    no_efectores
999    8.955    no_efectores

```

[824 rows x 21 columns]

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

Estadísticas.

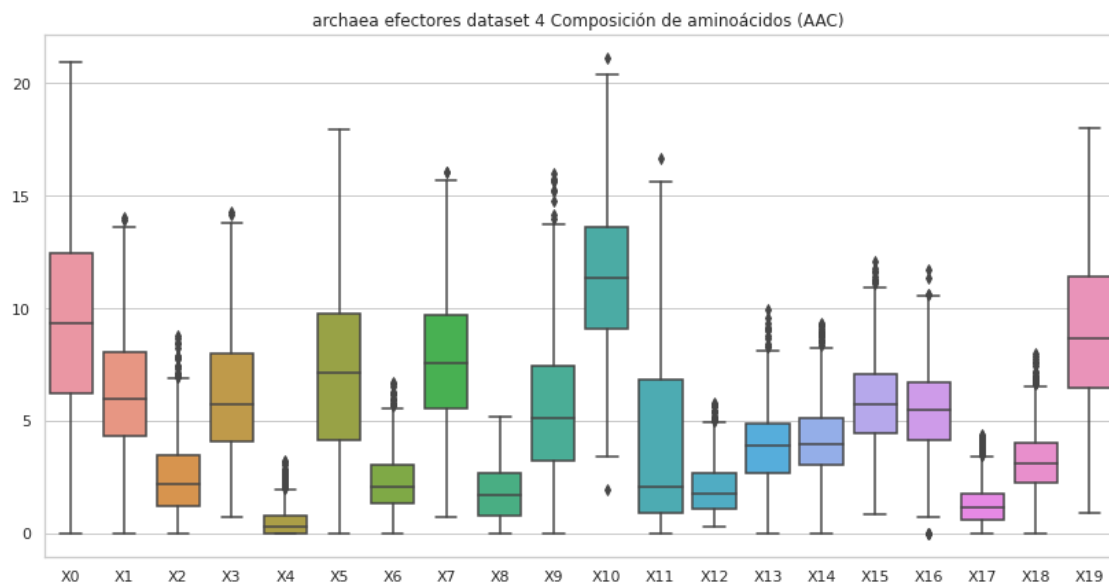
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|-------|------------|------------|------------|------------|------------|------------|---|
| count | 824.000000 | 824.000000 | 824.000000 | 824.000000 | 824.000000 | 824.000000 | |
| mean | 9.958112 | 6.422133 | 2.803417 | 7.824364 | 0.807637 | 8.220227 | |
| std | 3.611473 | 2.544042 | 1.756605 | 3.115689 | 0.888874 | 3.340313 | |
| min | 1.227000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 7.407000 | 4.715750 | 1.608250 | 5.507750 | 0.000000 | 6.170750 | |
| 50% | 9.839500 | 6.257500 | 2.500000 | 7.934500 | 0.589500 | 8.314000 | |
| 75% | 12.500000 | 8.029250 | 3.594500 | 10.120000 | 1.246500 | 10.433750 | |
| max | 19.580000 | 14.583000 | 9.073000 | 17.327000 | 5.063000 | 19.355000 | |

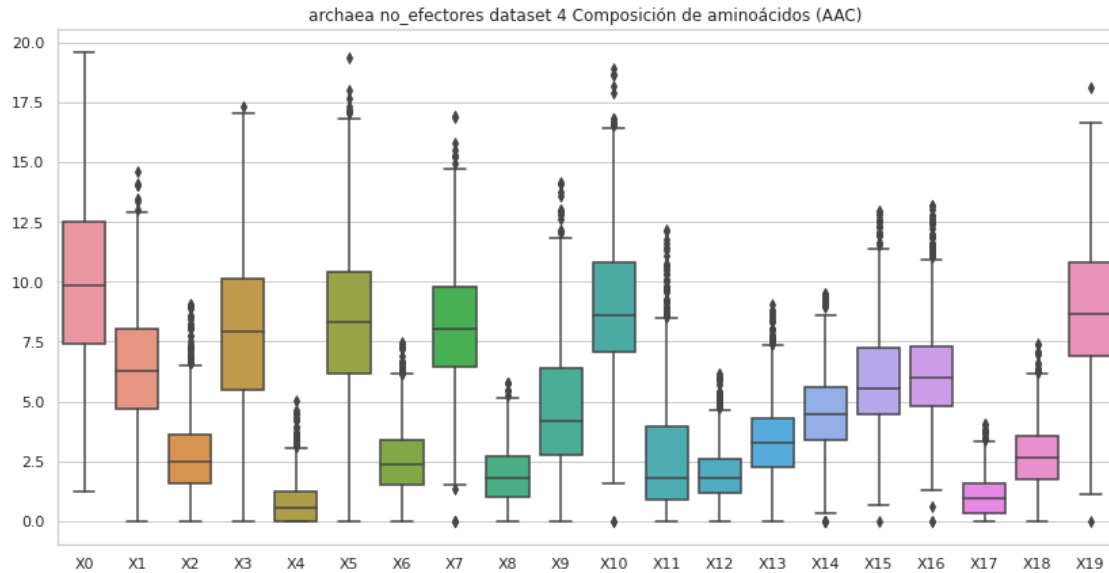
| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|-------|------------|------------|------------|------------|------------|------------|---|
| count | 824.000000 | 824.000000 | 824.000000 | 824.000000 | 824.000000 | 824.000000 | |
| mean | 2.563387 | 8.108751 | 1.915045 | 4.771163 | 9.001772 | 2.785556 | |
| std | 1.475525 | 2.610179 | 1.200688 | 2.745289 | 2.917130 | 2.618075 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 1.530500 | 6.430250 | 1.020000 | 2.755000 | 7.060000 | 0.909000 | |
| 50% | 2.367500 | 8.019500 | 1.800000 | 4.203000 | 8.604500 | 1.826000 | |
| 75% | 3.378000 | 9.784250 | 2.703000 | 6.411750 | 10.837000 | 3.949000 | |
| max | 7.463000 | 16.923000 | 5.797000 | 14.159000 | 18.919000 | 12.195000 | |

| | X12 | X13 | X14 | X15 | X16 | X17 | \ |
|-------|------------|------------|------------|------------|------------|------------|---|
| count | 824.000000 | 824.000000 | 824.000000 | 824.000000 | 824.000000 | 824.000000 | |
| mean | 2.021483 | 3.460603 | 4.511703 | 5.947197 | 6.193448 | 1.046209 | |
| std | 1.113832 | 1.691884 | 1.656888 | 2.186567 | 2.190795 | 0.861112 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 1.191750 | 2.263000 | 3.383250 | 4.471250 | 4.811750 | 0.365500 | |

| | | | | | | |
|-----|----------|----------|----------|-----------|-----------|----------|
| 50% | 1.839500 | 3.280500 | 4.466500 | 5.569500 | 5.975000 | 0.950000 |
| 75% | 2.602500 | 4.319000 | 5.601750 | 7.242000 | 7.293250 | 1.559000 |
| max | 6.195000 | 9.060000 | 9.524000 | 12.984000 | 13.187000 | 4.054000 |

| | X18 | X19 |
|-------|------------|------------|
| count | 824.000000 | 824.000000 |
| mean | 2.770269 | 8.867507 |
| std | 1.369459 | 2.808961 |
| min | 0.000000 | 0.000000 |
| 25% | 1.782000 | 6.912000 |
| 50% | 2.685500 | 8.644500 |
| 75% | 3.548500 | 10.839000 |
| max | 7.407000 | 18.103000 |





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

efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.020607 | 0.000000 | 0.005965 | 0.013557 | 0.010303 | 0.011930 | 0.001085 |
| 1 | 0.034328 | 0.001635 | 0.014712 | 0.009808 | 0.011443 | 0.026155 | 0.008173 |
| 2 | 0.018712 | 0.003742 | 0.011227 | 0.033682 | 0.007485 | 0.013098 | 0.007485 |
| 3 | 0.026948 | 0.000000 | 0.010907 | 0.008983 | 0.010266 | 0.014115 | 0.003850 |
| 4 | 0.018842 | 0.000285 | 0.007993 | 0.005710 | 0.004853 | 0.014845 | 0.000856 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.045000 | 0.002647 | 0.013235 | 0.068824 | 0.018529 | 0.034412 | 0.007941 |
| 996 | 0.062345 | 0.005421 | 0.051502 | 0.059634 | 0.016264 | 0.018975 | 0.016264 |
| 997 | 0.038694 | 0.004552 | 0.052351 | 0.066008 | 0.031866 | 0.015933 | 0.015933 |
| 998 | 0.033451 | 0.000000 | 0.019115 | 0.043008 | 0.011947 | 0.035840 | 0.007168 |
| 999 | 0.029618 | 0.003202 | 0.045628 | 0.036022 | 0.008005 | 0.029618 | 0.010406 |

| | X7 | X8 | X9 ... | X74 | X75 | X76 \ |
|-----|----------|----------|--------------|-----------|-----------|-----------|
| 0 | 0.014641 | 0.002169 | 0.024945 ... | 0.016666 | 0.006193 | 0.013814 |
| 1 | 0.011443 | 0.003269 | 0.037597 ... | 0.012735 | -0.002651 | 0.021100 |
| 2 | 0.026197 | 0.028068 | 0.035553 ... | -0.026432 | 0.000761 | -0.001781 |
| 3 | 0.010907 | 0.000000 | 0.025023 ... | 0.018187 | -0.005497 | 0.021441 |
| 4 | 0.004282 | 0.001427 | 0.016272 ... | 0.016275 | 0.004844 | 0.016310 |
| .. | ... | ... | ... | ... | ... | |
| 995 | 0.031765 | 0.029118 | 0.050294 ... | -0.018238 | 0.001742 | 0.000263 |
| 996 | 0.040660 | 0.010843 | 0.059634 ... | 0.019397 | 0.004715 | 0.007548 |
| 997 | 0.043247 | 0.063732 | 0.054627 ... | 0.001698 | 0.084095 | 0.012937 |
| 998 | 0.059734 | 0.033451 | 0.069291 ... | 0.060295 | 0.031863 | 0.008933 |
| 999 | 0.010406 | 0.002401 | 0.022414 ... | -0.004466 | 0.018404 | 0.011776 |

| | X77 | X78 | X79 | X80 | X81 | X82 | X83 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.014203 | 0.005297 | 0.017436 | 0.012865 | -0.001322 | 0.013447 | efectores |
| 1 | 0.029067 | -0.002316 | 0.004326 | 0.031966 | 0.009912 | -0.001149 | efectores |
| 2 | -0.025964 | -0.005823 | 0.009475 | 0.002180 | 0.018147 | 0.007527 | efectores |
| 3 | 0.023653 | -0.002776 | 0.020193 | 0.014020 | -0.003151 | 0.018563 | efectores |
| 4 | 0.014690 | 0.004084 | 0.015770 | 0.012278 | 0.002981 | 0.017740 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.013304 | 0.016269 | 0.012909 | -0.011333 | -0.017486 | -0.006920 | efectores |
| 996 | 0.014308 | 0.019778 | 0.010480 | -0.005697 | -0.008758 | 0.015877 | efectores |
| 997 | -0.055112 | -0.018319 | -0.012401 | 0.005905 | 0.034171 | 0.013033 | efectores |

```

998  0.008316  0.015329  0.023481 -0.022649 -0.000309  0.026042  efectores
999 -0.001589  0.021363  0.006103  0.012476  0.028520  0.007283  efectores

```

[1000 rows x 84 columns]

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

| | X0 | X1 | X2 | X3 | X4 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.031738 | 0.003414 | 0.027090 | 0.033612 | 0.015723 | |
| std | 0.016094 | 0.006260 | 0.021486 | 0.028415 | 0.012536 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.021111 | 0.000000 | 0.009438 | 0.008021 | 0.007388 | |
| 50% | 0.028288 | 0.000846 | 0.021884 | 0.028013 | 0.012567 | |
| 75% | 0.038847 | 0.004128 | 0.039826 | 0.053881 | 0.020467 | |
| max | 0.134048 | 0.073521 | 0.130122 | 0.195184 | 0.101443 | |

| | X5 | X6 | X7 | X8 | X9 | ... | \ |
|-------|-------------|-------------|-------------|-------------|-------------|-----|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | ... | |
| mean | 0.027137 | 0.008049 | 0.027325 | 0.024517 | 0.044916 | ... | |
| std | 0.013497 | 0.008091 | 0.027242 | 0.031531 | 0.029212 | ... | |
| min | 0.001544 | 0.000000 | 0.000000 | 0.000000 | 0.001579 | ... | |
| 25% | 0.017134 | 0.002045 | 0.007750 | 0.002105 | 0.024522 | ... | |
| 50% | 0.024556 | 0.006179 | 0.017271 | 0.009227 | 0.038070 | ... | |
| 75% | 0.033425 | 0.010999 | 0.038637 | 0.039625 | 0.057245 | ... | |
| max | 0.110281 | 0.061267 | 0.220562 | 0.245069 | 0.325306 | ... | |

| | X73 | X74 | X75 | X76 | X77 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.012450 | 0.004901 | 0.008912 | 0.014298 | 0.006542 | |
| std | 0.019871 | 0.027953 | 0.026688 | 0.017223 | 0.028498 | |
| min | -0.166681 | -0.146645 | -0.124379 | -0.081614 | -0.322585 | |
| 25% | 0.003113 | -0.007561 | -0.001650 | 0.004527 | -0.002968 | |
| 50% | 0.013117 | 0.009181 | 0.004481 | 0.015628 | 0.010856 | |
| 75% | 0.022076 | 0.016891 | 0.016079 | 0.023943 | 0.018872 | |
| max | 0.099982 | 0.152091 | 0.186971 | 0.100503 | 0.176925 | |

| | X78 | X79 | X80 | X81 | X82 |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.009748 | 0.013590 | 0.004679 | 0.009631 | 0.013846 |
| std | 0.025126 | 0.020007 | 0.030340 | 0.025033 | 0.019177 |
| min | -0.201310 | -0.137996 | -0.246811 | -0.134310 | -0.095803 |
| 25% | -0.000183 | 0.005220 | -0.003799 | -0.000295 | 0.004368 |
| 50% | 0.006195 | 0.015543 | 0.009284 | 0.005719 | 0.014775 |
| 75% | 0.019330 | 0.024450 | 0.017978 | 0.018818 | 0.023610 |

max 0.240651 0.080029 0.130458 0.132219 0.126453

[8 rows x 83 columns]

no_efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.014193 | 0.001373 | 0.007325 | 0.009614 | 0.013735 | 0.017397 | 0.000458 |
| 1 | 0.020575 | 0.002881 | 0.006173 | 0.007407 | 0.016049 | 0.020987 | 0.002881 |
| 2 | 0.041756 | 0.041756 | 0.000000 | 0.041756 | 0.031317 | 0.020878 | 0.000000 |
| 3 | 0.023332 | 0.007777 | 0.015555 | 0.011666 | 0.007777 | 0.023332 | 0.000000 |
| 4 | 0.086369 | 0.034548 | 0.060458 | 0.017274 | 0.000000 | 0.034548 | 0.008637 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.016448 | 0.000633 | 0.001265 | 0.005694 | 0.012020 | 0.016448 | 0.001898 |
| 996 | 0.043493 | 0.000000 | 0.009320 | 0.086987 | 0.021747 | 0.024853 | 0.006213 |
| 997 | 0.013115 | 0.000000 | 0.000000 | 0.004372 | 0.008743 | 0.013115 | 0.000000 |
| 998 | 0.022197 | 0.004439 | 0.044393 | 0.062150 | 0.008879 | 0.039954 | 0.008879 |
| 999 | 0.035236 | 0.001409 | 0.029599 | 0.039465 | 0.015504 | 0.025370 | 0.014095 |

| | X7 | X8 | X9 ... | X74 | X75 | X76 \ |
|-----|----------|----------|--------------|-----------|-----------|-----------|
| 0 | 0.027470 | 0.004578 | 0.035710 ... | 0.024001 | 0.005881 | 0.008916 |
| 1 | 0.023456 | 0.009465 | 0.013991 ... | 0.008242 | 0.002609 | 0.017258 |
| 2 | 0.010439 | 0.000000 | 0.031317 ... | 0.005609 | -0.011318 | -0.018691 |
| 3 | 0.031109 | 0.034998 | 0.027221 ... | -0.006271 | -0.003397 | 0.016068 |
| 4 | 0.000000 | 0.008637 | 0.043184 ... | -0.025245 | -0.028102 | -0.101018 |
| .. | ... | ... | ... | ... | ... | ... |
| 995 | 0.013917 | 0.000633 | 0.024672 ... | 0.011205 | 0.001279 | 0.015613 |
| 996 | 0.065240 | 0.077667 | 0.080773 ... | -0.049498 | -0.003355 | -0.010209 |
| 997 | 0.004372 | 0.004372 | 0.069947 ... | -0.017370 | -0.022610 | 0.028489 |
| 998 | 0.022197 | 0.008879 | 0.053272 ... | -0.012515 | 0.025409 | -0.002382 |
| 999 | 0.011276 | 0.008457 | 0.038055 ... | -0.013739 | 0.008740 | 0.026995 |

| | X77 | X78 | X79 | X80 | X81 | X82 | X83 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | 0.027978 | 0.009744 | 0.009256 | 0.015789 | 0.001097 | 0.011361 | no_efectores |
| 1 | 0.012176 | 0.004211 | 0.014271 | 0.009898 | -0.000623 | 0.021883 | no_efectores |
| 2 | 0.045487 | 0.020150 | 0.033154 | -0.063564 | -0.040635 | 0.033594 | no_efectores |
| 3 | -0.013072 | -0.005422 | 0.035689 | -0.034166 | -0.007662 | 0.022184 | no_efectores |
| 4 | -0.025840 | -0.029845 | 0.001138 | -0.020867 | 0.017755 | 0.053877 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.013161 | -0.000907 | 0.027103 | 0.010487 | 0.001049 | 0.020270 | no_efectores |
| 996 | -0.095334 | -0.048757 | 0.032879 | 0.027400 | 0.052060 | -0.002741 | no_efectores |
| 997 | -0.010541 | -0.015429 | 0.023682 | -0.000258 | 0.005112 | -0.006655 | no_efectores |
| 998 | -0.011815 | 0.018126 | -0.015972 | -0.043820 | 0.016302 | -0.069529 | no_efectores |

999 0.003535 0.009679 0.000942 0.023235 0.004619 0.032220 no_efectores

[1000 rows x 84 columns]

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

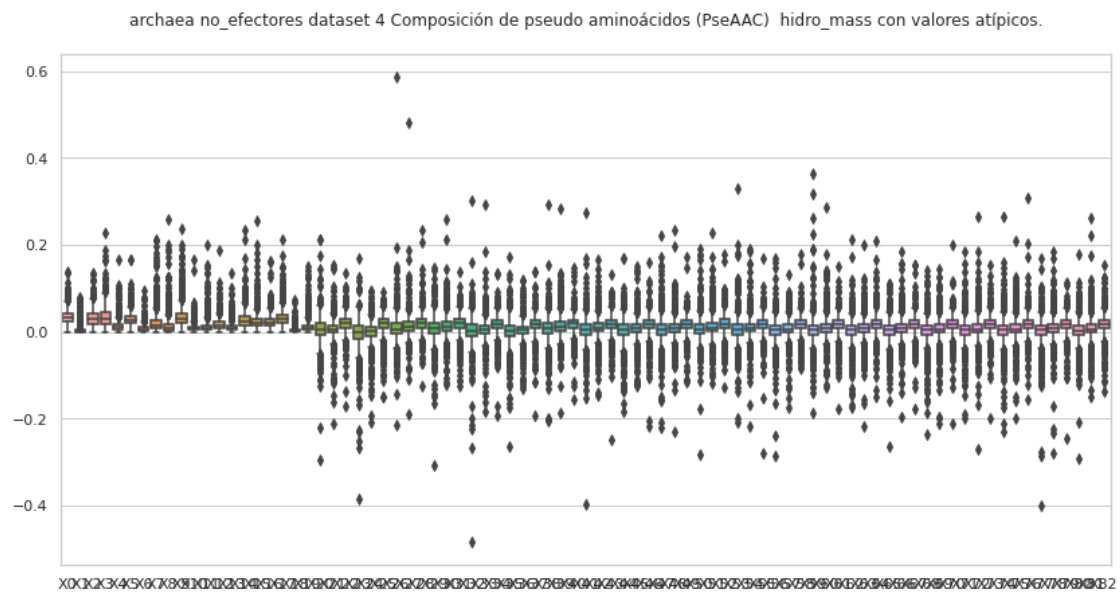
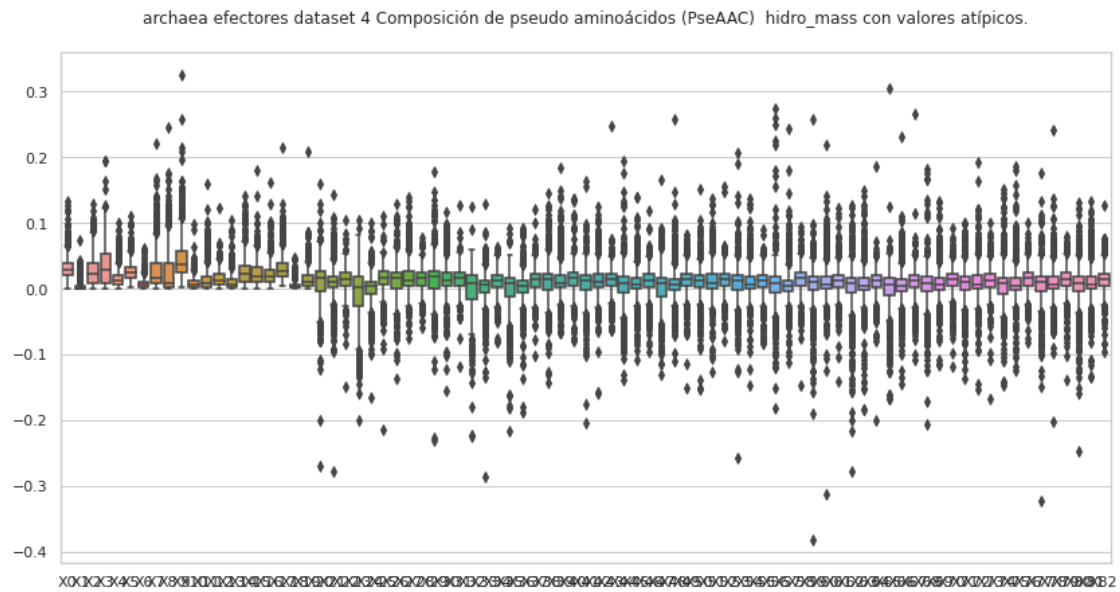
| | X0 | X1 | X2 | X3 | X4 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.033986 | 0.004692 | 0.030284 | 0.033640 | 0.014652 | |
| std | 0.015983 | 0.008699 | 0.018933 | 0.024597 | 0.014170 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.022954 | 0.000000 | 0.017068 | 0.017101 | 0.006533 | |
| 50% | 0.031825 | 0.001926 | 0.029187 | 0.030134 | 0.011108 | |
| 75% | 0.042557 | 0.005478 | 0.040623 | 0.045788 | 0.017580 | |
| max | 0.139107 | 0.079953 | 0.133410 | 0.227438 | 0.165069 | |

| | X5 | X6 | X7 | X8 | X9 | ... | \ |
|-------|-------------|-------------|-------------|-------------|-------------|-----|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | ... | |
| mean | 0.029314 | 0.008385 | 0.021680 | 0.015310 | 0.035346 | ... | |
| std | 0.015441 | 0.009678 | 0.024405 | 0.024015 | 0.024887 | ... | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | ... | |
| 25% | 0.020062 | 0.002295 | 0.007966 | 0.002492 | 0.020378 | ... | |
| 50% | 0.027105 | 0.005999 | 0.014161 | 0.006809 | 0.029066 | ... | |
| 75% | 0.035921 | 0.011180 | 0.026402 | 0.017656 | 0.041796 | ... | |
| max | 0.166763 | 0.094325 | 0.212232 | 0.257172 | 0.236041 | ... | |

| | X73 | X74 | X75 | X76 | X77 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.016643 | 0.002090 | 0.007481 | 0.016715 | 0.001059 | |
| std | 0.022163 | 0.030002 | 0.025197 | 0.023807 | 0.031465 | |
| min | -0.200217 | -0.229256 | -0.199429 | -0.157210 | -0.400319 | |
| 25% | 0.007429 | -0.007427 | -0.002163 | 0.007280 | -0.007132 | |
| 50% | 0.018054 | 0.003650 | 0.005958 | 0.017902 | 0.003518 | |
| 75% | 0.027307 | 0.014220 | 0.017051 | 0.026923 | 0.014137 | |
| max | 0.152348 | 0.265311 | 0.208654 | 0.307374 | 0.169031 | |

| | X78 | X79 | X80 | X81 | X82 |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.008152 | 0.016052 | 0.001027 | 0.008393 | 0.016540 |
| std | 0.027661 | 0.021590 | 0.027178 | 0.026743 | 0.022723 |
| min | -0.280034 | -0.245307 | -0.292333 | -0.150245 | -0.137625 |
| 25% | -0.001478 | 0.007616 | -0.008297 | -0.002173 | 0.007379 |
| 50% | 0.006800 | 0.017465 | 0.002346 | 0.005459 | 0.017591 |
| 75% | 0.017713 | 0.027350 | 0.012937 | 0.016625 | 0.027020 |
| max | 0.183559 | 0.113368 | 0.179035 | 0.261440 | 0.152116 |

```
[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 4, sin valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.020607 | 0.000000 | 0.005965 | 0.013557 | 0.010303 | 0.011930 | 0.001085 |
| 1 | 0.034328 | 0.001635 | 0.014712 | 0.009808 | 0.011443 | 0.026155 | 0.008173 |
| 2 | 0.018712 | 0.003742 | 0.011227 | 0.033682 | 0.007485 | 0.013098 | 0.007485 |
| 3 | 0.026948 | 0.000000 | 0.010907 | 0.008983 | 0.010266 | 0.014115 | 0.003850 |
| 4 | 0.018842 | 0.000285 | 0.007993 | 0.005710 | 0.004853 | 0.014845 | 0.000856 |
| .. | ... | ... | ... | ... | ... | ... | |
| 994 | 0.021092 | 0.007031 | 0.007031 | 0.028123 | 0.000000 | 0.028123 | 0.000000 |
| 995 | 0.045000 | 0.002647 | 0.013235 | 0.068824 | 0.018529 | 0.034412 | 0.007941 |
| 996 | 0.062345 | 0.005421 | 0.051502 | 0.059634 | 0.016264 | 0.018975 | 0.016264 |
| 998 | 0.033451 | 0.000000 | 0.019115 | 0.043008 | 0.011947 | 0.035840 | 0.007168 |
| 999 | 0.029618 | 0.003202 | 0.045628 | 0.036022 | 0.008005 | 0.029618 | 0.010406 |

| | X7 | X8 | X9 ... | X74 | X75 | X76 \ |
|-----|----------|----------|--------------|-----------|-----------|-----------|
| 0 | 0.014641 | 0.002169 | 0.024945 ... | 0.016666 | 0.006193 | 0.013814 |
| 1 | 0.011443 | 0.003269 | 0.037597 ... | 0.012735 | -0.002651 | 0.021100 |
| 2 | 0.026197 | 0.028068 | 0.035553 ... | -0.026432 | 0.000761 | -0.001781 |
| 3 | 0.010907 | 0.000000 | 0.025023 ... | 0.018187 | -0.005497 | 0.021441 |
| 4 | 0.004282 | 0.001427 | 0.016272 ... | 0.016275 | 0.004844 | 0.016310 |
| .. | ... | ... | ... | ... | ... | |
| 994 | 0.042184 | 0.021092 | 0.014061 ... | -0.018947 | 0.005423 | 0.027745 |
| 995 | 0.031765 | 0.029118 | 0.050294 ... | -0.018238 | 0.001742 | 0.000263 |
| 996 | 0.040660 | 0.010843 | 0.059634 ... | 0.019397 | 0.004715 | 0.007548 |
| 998 | 0.059734 | 0.033451 | 0.069291 ... | 0.060295 | 0.031863 | 0.008933 |
| 999 | 0.010406 | 0.002401 | 0.022414 ... | -0.004466 | 0.018404 | 0.011776 |

| | X77 | X78 | X79 | X80 | X81 | X82 | X83 |
|-----|-----------|-----------|----------|-----------|-----------|-----------|-----------|
| 0 | 0.014203 | 0.005297 | 0.017436 | 0.012865 | -0.001322 | 0.013447 | efectores |
| 1 | 0.029067 | -0.002316 | 0.004326 | 0.031966 | 0.009912 | -0.001149 | efectores |
| 2 | -0.025964 | -0.005823 | 0.009475 | 0.002180 | 0.018147 | 0.007527 | efectores |
| 3 | 0.023653 | -0.002776 | 0.020193 | 0.014020 | -0.003151 | 0.018563 | efectores |
| 4 | 0.014690 | 0.004084 | 0.015770 | 0.012278 | 0.002981 | 0.017740 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 994 | 0.009126 | 0.025275 | 0.031921 | 0.063917 | 0.023461 | 0.058974 | efectores |
| 995 | 0.013304 | 0.016269 | 0.012909 | -0.011333 | -0.017486 | -0.006920 | efectores |
| 996 | 0.014308 | 0.019778 | 0.010480 | -0.005697 | -0.008758 | 0.015877 | efectores |
| 998 | 0.008316 | 0.015329 | 0.023481 | -0.022649 | -0.000309 | 0.026042 | efectores |
| 999 | -0.001589 | 0.021363 | 0.006103 | 0.012476 | 0.028520 | 0.007283 | efectores |

[786 rows x 84 columns]

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

| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 786.000000 | 786.000000 | 786.000000 | 786.000000 | 786.000000 | 786.000000 |
| mean | 0.028794 | 0.001827 | 0.020438 | 0.024544 | 0.013122 | 0.024024 |
| std | 0.012078 | 0.002905 | 0.015179 | 0.021109 | 0.008585 | 0.010031 |
| min | 0.002020 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.001544 |
| 25% | 0.020607 | 0.000000 | 0.007991 | 0.006618 | 0.006931 | 0.016252 |
| 50% | 0.026177 | 0.000417 | 0.015567 | 0.015780 | 0.010808 | 0.022418 |
| 75% | 0.035263 | 0.002495 | 0.031407 | 0.041197 | 0.018139 | 0.029736 |
| max | 0.073238 | 0.016428 | 0.074062 | 0.089557 | 0.043231 | 0.059919 |

| | X6 | X7 | X8 | X9 ... | X73 \ |
|-------|------------|------------|------------|------------|------------|
| count | 786.000000 | 786.000000 | 786.000000 | 786.000000 | 786.000000 |
| mean | 0.006135 | 0.019729 | 0.015465 | 0.035961 | 0.014029 |
| std | 0.005407 | 0.017400 | 0.021070 | 0.016886 | 0.014624 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.001579 | -0.032834 |
| 25% | 0.001672 | 0.006557 | 0.001501 | 0.022683 | 0.005899 |
| 50% | 0.005365 | 0.012752 | 0.005111 | 0.033803 | 0.014104 |
| 75% | 0.008980 | 0.029286 | 0.020963 | 0.046187 | 0.021733 |
| max | 0.030166 | 0.093543 | 0.113431 | 0.109285 | 0.068001 |

| | X74 | X75 | X76 | X77 | X78 | X79 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 786.000000 | 786.000000 | 786.000000 | 786.000000 | 786.000000 | 786.000000 |
| mean | 0.006066 | 0.006538 | 0.015578 | 0.008614 | 0.008431 | 0.015250 |
| std | 0.017574 | 0.015085 | 0.012983 | 0.016077 | 0.015232 | 0.013259 |
| min | -0.064007 | -0.041144 | -0.031667 | -0.070324 | -0.058425 | -0.032114 |
| 25% | -0.004526 | -0.001422 | 0.007315 | 0.001277 | 0.000300 | 0.007040 |
| 50% | 0.009846 | 0.003489 | 0.016434 | 0.011678 | 0.005516 | 0.016118 |
| 75% | 0.016488 | 0.011487 | 0.023886 | 0.018231 | 0.014521 | 0.023693 |
| max | 0.063681 | 0.062326 | 0.056817 | 0.061684 | 0.071011 | 0.059411 |

| | X80 | X81 | X82 |
|-------|------------|------------|------------|
| count | 786.000000 | 786.000000 | 786.000000 |
| mean | 0.008237 | 0.008327 | 0.014669 |
| std | 0.017708 | 0.014601 | 0.013662 |
| min | -0.077821 | -0.054481 | -0.026321 |
| 25% | 0.000223 | 0.000166 | 0.006296 |
| 50% | 0.010511 | 0.004869 | 0.015424 |
| 75% | 0.017684 | 0.015651 | 0.023238 |
| max | 0.070228 | 0.084127 | 0.067801 |

[8 rows x 83 columns]

no_efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.014193 | 0.001373 | 0.007325 | 0.009614 | 0.013735 | 0.017397 | 0.000458 |
| 1 | 0.020575 | 0.002881 | 0.006173 | 0.007407 | 0.016049 | 0.020987 | 0.002881 |
| 3 | 0.023332 | 0.007777 | 0.015555 | 0.011666 | 0.007777 | 0.023332 | 0.000000 |
| 5 | 0.046748 | 0.000000 | 0.044715 | 0.048780 | 0.012195 | 0.029471 | 0.014228 |
| 6 | 0.028255 | 0.001926 | 0.020549 | 0.019907 | 0.010274 | 0.036603 | 0.010274 |
| .. | ... | ... | ... | ... | ... | ... | |
| 993 | 0.026485 | 0.002863 | 0.027201 | 0.054401 | 0.012169 | 0.040801 | 0.012169 |
| 994 | 0.022359 | 0.000639 | 0.027470 | 0.022998 | 0.004472 | 0.029386 | 0.010221 |
| 995 | 0.016448 | 0.000633 | 0.001265 | 0.005694 | 0.012020 | 0.016448 | 0.001898 |
| 997 | 0.013115 | 0.000000 | 0.000000 | 0.004372 | 0.008743 | 0.013115 | 0.000000 |
| 999 | 0.035236 | 0.001409 | 0.029599 | 0.039465 | 0.015504 | 0.025370 | 0.014095 |

| | X7 | X8 | X9 | ... | X74 | X75 | X76 \ |
|-----|----------|----------|----------|-----|-----------|-----------|----------|
| 0 | 0.027470 | 0.004578 | 0.035710 | ... | 0.024001 | 0.005881 | 0.008916 |
| 1 | 0.023456 | 0.009465 | 0.013991 | ... | 0.008242 | 0.002609 | 0.017258 |
| 3 | 0.031109 | 0.034998 | 0.027221 | ... | -0.006271 | -0.003397 | 0.016068 |
| 5 | 0.017276 | 0.003049 | 0.041666 | ... | -0.011303 | 0.002359 | 0.011553 |
| 6 | 0.009632 | 0.005137 | 0.014127 | ... | -0.004314 | 0.005906 | 0.036204 |
| .. | ... | ... | ... | ... | ... | ... | |
| 993 | 0.023622 | 0.014316 | 0.032927 | ... | -0.010004 | 0.012568 | 0.014792 |
| 994 | 0.021081 | 0.003833 | 0.018526 | ... | -0.005128 | 0.002490 | 0.022271 |
| 995 | 0.013917 | 0.000633 | 0.024672 | ... | 0.011205 | 0.001279 | 0.015613 |
| 997 | 0.004372 | 0.004372 | 0.069947 | ... | -0.017370 | -0.022610 | 0.028489 |
| 999 | 0.011276 | 0.008457 | 0.038055 | ... | -0.013739 | 0.008740 | 0.026995 |

| | X77 | X78 | X79 | X80 | X81 | X82 | X83 |
|-----|-----------|-----------|----------|-----------|-----------|-----------|--------------|
| 0 | 0.027978 | 0.009744 | 0.009256 | 0.015789 | 0.001097 | 0.011361 | no_efectores |
| 1 | 0.012176 | 0.004211 | 0.014271 | 0.009898 | -0.000623 | 0.021883 | no_efectores |
| 3 | -0.013072 | -0.005422 | 0.035689 | -0.034166 | -0.007662 | 0.022184 | no_efectores |
| 5 | 0.021915 | 0.029524 | 0.006033 | -0.022398 | -0.000444 | 0.000046 | no_efectores |
| 6 | 0.006930 | -0.000030 | 0.029059 | 0.003679 | -0.004786 | 0.032902 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 993 | -0.008128 | 0.019021 | 0.017167 | -0.010308 | 0.011818 | 0.006974 | no_efectores |
| 994 | -0.005132 | -0.004264 | 0.027324 | 0.000754 | 0.008302 | 0.042081 | no_efectores |
| 995 | 0.013161 | -0.000907 | 0.027103 | 0.010487 | 0.001049 | 0.020270 | no_efectores |
| 997 | -0.010541 | -0.015429 | 0.023682 | -0.000258 | 0.005112 | -0.006655 | no_efectores |
| 999 | 0.003535 | 0.009679 | 0.000942 | 0.023235 | 0.004619 | 0.032220 | no_efectores |

[846 rows x 84 columns]

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

| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 846.000000 | 846.000000 | 846.000000 | 846.000000 | 846.000000 | 846.000000 |
| mean | 0.032276 | 0.003128 | 0.026952 | 0.028651 | 0.011923 | 0.026698 |
| std | 0.013117 | 0.004162 | 0.015494 | 0.017763 | 0.008326 | 0.010767 |
| min | 0.004207 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 0.022736 | 0.000000 | 0.015194 | 0.014980 | 0.006111 | 0.019311 |
| 50% | 0.031034 | 0.001843 | 0.027085 | 0.027475 | 0.010183 | 0.025738 |
| 75% | 0.040534 | 0.004664 | 0.037347 | 0.039631 | 0.015582 | 0.033493 |
| max | 0.078077 | 0.026520 | 0.079006 | 0.096475 | 0.056008 | 0.072011 |

| | X6 | X7 | X8 | X9 ... | X73 \ |
|-------|------------|------------|------------|------------|------------|
| count | 846.000000 | 846.000000 | 846.000000 | 846.000000 | 846.000000 |
| mean | 0.006899 | 0.017048 | 0.010692 | 0.030140 | 0.017634 |
| std | 0.006038 | 0.013903 | 0.013296 | 0.015694 | 0.014809 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | -0.038782 |
| 25% | 0.002287 | 0.007466 | 0.002279 | 0.019574 | 0.009185 |
| 50% | 0.005319 | 0.013106 | 0.005811 | 0.027282 | 0.018467 |
| 75% | 0.009973 | 0.021719 | 0.013026 | 0.037341 | 0.026656 |
| max | 0.036885 | 0.082992 | 0.080091 | 0.108351 | 0.069812 |

| | X74 | X75 | X76 | X77 | X78 | X79 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 846.000000 | 846.000000 | 846.000000 | 846.000000 | 846.000000 | 846.000000 |
| mean | 0.003222 | 0.007925 | 0.017504 | 0.003350 | 0.008485 | 0.017834 |
| std | 0.016303 | 0.014939 | 0.014348 | 0.016324 | 0.015486 | 0.014052 |
| min | -0.083954 | -0.055217 | -0.034511 | -0.072319 | -0.057766 | -0.042782 |
| 25% | -0.005123 | -0.001150 | 0.008919 | -0.005401 | -0.000877 | 0.009463 |
| 50% | 0.003911 | 0.005958 | 0.018218 | 0.003939 | 0.006698 | 0.018429 |
| 75% | 0.013480 | 0.015990 | 0.026519 | 0.013269 | 0.015869 | 0.027321 |
| max | 0.067063 | 0.079616 | 0.082224 | 0.056700 | 0.079019 | 0.061603 |

| | X80 | X81 | X82 |
|-------|------------|------------|------------|
| count | 846.000000 | 846.000000 | 846.000000 |
| mean | 0.002449 | 0.007918 | 0.018222 |
| std | 0.016716 | 0.015661 | 0.014356 |
| min | -0.056885 | -0.039922 | -0.037029 |
| 25% | -0.006489 | -0.001460 | 0.009420 |
| 50% | 0.002449 | 0.005272 | 0.018466 |
| 75% | 0.012102 | 0.015321 | 0.027039 |
| max | 0.054575 | 0.084818 | 0.070209 |

[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 4,
con valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.046579 | 0.000000 | 0.013483 | 0.030644 | 0.023290 | 0.026967 | 0.002452 |
| 1 | 0.057973 | 0.002761 | 0.024846 | 0.016564 | 0.019324 | 0.044170 | 0.013803 |
| 2 | 0.039807 | 0.007961 | 0.023884 | 0.071653 | 0.015923 | 0.027865 | 0.015923 |
| 3 | 0.055195 | 0.000000 | 0.022341 | 0.018398 | 0.021027 | 0.028912 | 0.007885 |
| 4 | 0.038296 | 0.000580 | 0.016247 | 0.011605 | 0.009864 | 0.030173 | 0.001741 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.084424 | 0.004966 | 0.024831 | 0.129119 | 0.034763 | 0.064560 | 0.014898 |
| 996 | 0.076709 | 0.006670 | 0.063369 | 0.073374 | 0.020011 | 0.023346 | 0.020011 |
| 997 | 0.055974 | 0.006585 | 0.075730 | 0.095486 | 0.046097 | 0.023048 | 0.023048 |
| 998 | 0.037362 | 0.000000 | 0.021350 | 0.048037 | 0.013344 | 0.040031 | 0.008006 |
| 999 | 0.046876 | 0.005068 | 0.072214 | 0.057011 | 0.012669 | 0.046876 | 0.016470 |

| | X7 | X8 | X9 ... | X32 | X33 | X34 \ |
|---|----------|----------|--------------|----------|----------|----------|
| 0 | 0.033096 | 0.004903 | 0.056385 ... | 0.034216 | 0.023639 | 0.022493 |

| | | | | | | | |
|-----|----------|----------|----------|-----|-----------|----------|-----------|
| 1 | 0.019324 | 0.005521 | 0.063495 | ... | 0.044150 | 0.046488 | 0.037149 |
| 2 | 0.055730 | 0.059710 | 0.075633 | ... | 0.007544 | 0.012877 | 0.035787 |
| 3 | 0.022341 | 0.000000 | 0.051253 | ... | 0.044267 | 0.031211 | 0.017043 |
| 4 | 0.008704 | 0.002901 | 0.033074 | ... | 0.041681 | 0.041562 | 0.023469 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.059594 | 0.054627 | 0.094357 | ... | -0.045545 | 0.037955 | 0.006187 |
| 996 | 0.050028 | 0.013341 | 0.073374 | ... | 0.015526 | 0.025862 | 0.013840 |
| 997 | 0.062560 | 0.092193 | 0.079023 | ... | -0.021800 | 0.016160 | -0.033966 |
| 998 | 0.066719 | 0.037362 | 0.077394 | ... | -0.001604 | 0.037907 | 0.032740 |
| 999 | 0.016470 | 0.003801 | 0.035474 | ... | 0.017255 | 0.024744 | 0.024554 |

| | X35 | X36 | X37 | X38 | X39 | X40 | X41 |
|-----|-----------|----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.001507 | 0.045164 | 0.028675 | 0.031224 | 0.039413 | 0.030396 | efectores |
| 1 | 0.024663 | 0.021956 | 0.028845 | 0.035633 | 0.007307 | -0.001941 | efectores |
| 2 | -0.000630 | 0.015728 | 0.018898 | -0.003788 | 0.020157 | 0.016012 | efectores |
| 3 | 0.005837 | 0.034860 | 0.023530 | 0.043917 | 0.041360 | 0.038021 | efectores |
| 4 | 0.033670 | 0.025594 | 0.029789 | 0.033151 | 0.032052 | 0.036057 | efectores |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.026439 | 0.008540 | 0.027846 | 0.000493 | 0.024219 | -0.012983 | efectores |
| 996 | 0.010002 | 0.000133 | 0.008639 | 0.009287 | 0.012895 | 0.019535 | efectores |
| 997 | -0.011803 | 0.039314 | -0.017662 | 0.018715 | -0.017939 | 0.018854 | efectores |
| 998 | 0.037607 | 0.036380 | 0.028502 | 0.009977 | 0.026227 | 0.029087 | efectores |
| 999 | 0.025456 | 0.029475 | 0.019395 | 0.018638 | 0.009658 | 0.011527 | efectores |

[1000 rows x 42 columns]

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

Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.046098 | 0.004431 | 0.037662 | 0.047359 | 0.023545 | |
| std | 0.016406 | 0.007385 | 0.026183 | 0.038384 | 0.017470 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.035065 | 0.000000 | 0.017649 | 0.016006 | 0.012473 | |
| 50% | 0.043862 | 0.001383 | 0.029998 | 0.036659 | 0.019034 | |
| 75% | 0.054582 | 0.005942 | 0.052343 | 0.071350 | 0.030193 | |
| max | 0.116553 | 0.070593 | 0.167565 | 0.209457 | 0.124294 | |

| | X5 | X6 | X7 | X8 | X9 | ... | \ |
|-------|-------------|-------------|-------------|-------------|-------------|-----|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | ... | |
| mean | 0.039288 | 0.011323 | 0.039163 | 0.033791 | 0.066196 | ... | |
| std | 0.013649 | 0.010038 | 0.032750 | 0.040411 | 0.033206 | ... | |
| min | 0.005107 | 0.000000 | 0.000000 | 0.000000 | 0.002487 | ... | |
| 25% | 0.029777 | 0.003478 | 0.012759 | 0.003464 | 0.042109 | ... | |
| 50% | 0.037428 | 0.008974 | 0.028623 | 0.014028 | 0.059742 | ... | |

| | | | | | | |
|-----|----------|----------|----------|----------|----------|-----|
| 75% | 0.046948 | 0.016302 | 0.055964 | 0.053730 | 0.084728 | ... |
| max | 0.167565 | 0.076030 | 0.188447 | 0.206769 | 0.258459 | ... |

| | X31 | X32 | X33 | X34 | X35 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.015834 | 0.020972 | 0.016603 | 0.016844 | 0.015047 |
| std | 0.025757 | 0.026051 | 0.026586 | 0.027290 | 0.027805 |
| min | -0.175946 | -0.147655 | -0.205833 | -0.289384 | -0.179710 |
| 25% | 0.003370 | 0.005920 | 0.003745 | 0.004102 | 0.002893 |
| 50% | 0.020894 | 0.025334 | 0.021809 | 0.022730 | 0.021123 |
| 75% | 0.032165 | 0.037744 | 0.033382 | 0.033134 | 0.032772 |
| max | 0.098737 | 0.129072 | 0.137814 | 0.108009 | 0.153878 |

| | X36 | X37 | X38 | X39 | X40 |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.019206 | 0.018539 | 0.021273 | 0.020551 | 0.020832 |
| std | 0.025436 | 0.026221 | 0.023524 | 0.025867 | 0.026170 |
| min | -0.172538 | -0.173529 | -0.093852 | -0.165889 | -0.094000 |
| 25% | 0.008379 | 0.005290 | 0.008861 | 0.008534 | 0.006526 |
| 50% | 0.024359 | 0.021767 | 0.025017 | 0.024686 | 0.024939 |
| 75% | 0.034407 | 0.034684 | 0.036749 | 0.035874 | 0.036382 |
| max | 0.102090 | 0.116863 | 0.114128 | 0.109266 | 0.183094 |

[8 rows x 41 columns]

no_efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.029432 | 0.002848 | 0.015191 | 0.019938 | 0.028483 | 0.036078 | 0.000949 |
| 1 | 0.034803 | 0.004872 | 0.010441 | 0.012529 | 0.027147 | 0.035500 | 0.004872 |
| 2 | 0.035941 | 0.035941 | 0.000000 | 0.035941 | 0.026956 | 0.017971 | 0.000000 |
| 3 | 0.023428 | 0.007809 | 0.015619 | 0.011714 | 0.007809 | 0.023428 | 0.000000 |
| 4 | 0.194282 | 0.077713 | 0.135998 | 0.038856 | 0.000000 | 0.077713 | 0.019428 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.032619 | 0.001255 | 0.002509 | 0.011291 | 0.023837 | 0.032619 | 0.003764 |
| 996 | 0.064672 | 0.000000 | 0.013858 | 0.129343 | 0.032336 | 0.036955 | 0.009239 |
| 997 | 0.026084 | 0.000000 | 0.000000 | 0.008695 | 0.017389 | 0.026084 | 0.000000 |
| 998 | 0.047591 | 0.009518 | 0.095183 | 0.133256 | 0.019037 | 0.085665 | 0.019037 |
| 999 | 0.051691 | 0.002068 | 0.043420 | 0.057894 | 0.022744 | 0.037217 | 0.020676 |

| | X7 | X8 | X9 ... | X32 | X33 | X34 \ |
|---|----------|----------|--------------|----------|----------|----------|
| 0 | 0.056966 | 0.009494 | 0.074055 ... | 0.033843 | 0.012361 | 0.020728 |
| 1 | 0.039676 | 0.016010 | 0.023666 ... | 0.040298 | 0.019902 | 0.035659 |

| | | | | | | | |
|-----|----------|----------|----------|-----|-----------|-----------|-----------|
| 2 | 0.008985 | 0.000000 | 0.026956 | ... | 0.005028 | 0.043130 | 0.061577 |
| 3 | 0.031237 | 0.035142 | 0.027332 | ... | 0.033148 | 0.028823 | 0.016893 |
| 4 | 0.000000 | 0.019428 | 0.097141 | ... | -0.182817 | -0.183053 | -0.012636 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.027601 | 0.001255 | 0.048928 | ... | 0.045248 | 0.013821 | 0.013952 |
| 996 | 0.097008 | 0.115485 | 0.120105 | ... | 0.003633 | -0.025999 | 0.003843 |
| 997 | 0.008695 | 0.008695 | 0.139115 | ... | 0.028555 | -0.002692 | 0.041777 |
| 998 | 0.047591 | 0.019037 | 0.114220 | ... | -0.036693 | 0.040132 | 0.046747 |
| 999 | 0.016541 | 0.012406 | 0.055826 | ... | 0.013469 | 0.043790 | -0.003707 |

| | X35 | X36 | X37 | X38 | X39 | X40 | X41 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | 0.017941 | 0.001701 | 0.021919 | 0.018489 | 0.019195 | 0.023560 | no_efectores |
| 1 | 0.031664 | 0.034688 | 0.037143 | 0.029192 | 0.024139 | 0.037016 | no_efectores |
| 2 | 0.035119 | 0.011055 | 0.004903 | -0.016088 | 0.028537 | 0.028916 | no_efectores |
| 3 | 0.037868 | 0.032337 | 0.016353 | 0.016134 | 0.035836 | 0.022275 | no_efectores |
| 4 | -0.040474 | 0.102433 | -0.078567 | -0.227235 | 0.002559 | 0.121192 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.026924 | 0.015931 | 0.029711 | 0.030964 | 0.053751 | 0.040198 | no_efectores |
| 996 | -0.024158 | 0.005870 | 0.011703 | -0.015180 | 0.048889 | -0.004076 | no_efectores |
| 997 | 0.026554 | 0.034554 | 0.052723 | 0.056662 | 0.047100 | -0.013237 | no_efectores |
| 998 | 0.109358 | -0.039526 | 0.073742 | -0.005107 | -0.034245 | -0.149076 | no_efectores |
| 999 | 0.031215 | 0.031538 | 0.001543 | 0.039600 | 0.001382 | 0.047266 | no_efectores |

[1000 rows x 42 columns]

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

Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.045990 | 0.005832 | 0.041045 | 0.046223 | 0.019408 | |
| std | 0.018830 | 0.010072 | 0.026178 | 0.035038 | 0.015666 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.034444 | 0.000000 | 0.022785 | 0.022799 | 0.009242 | |
| 50% | 0.043876 | 0.002543 | 0.037264 | 0.038856 | 0.015750 | |
| 75% | 0.054686 | 0.006928 | 0.054467 | 0.060381 | 0.024786 | |
| max | 0.200635 | 0.082957 | 0.244700 | 0.299919 | 0.109259 | |

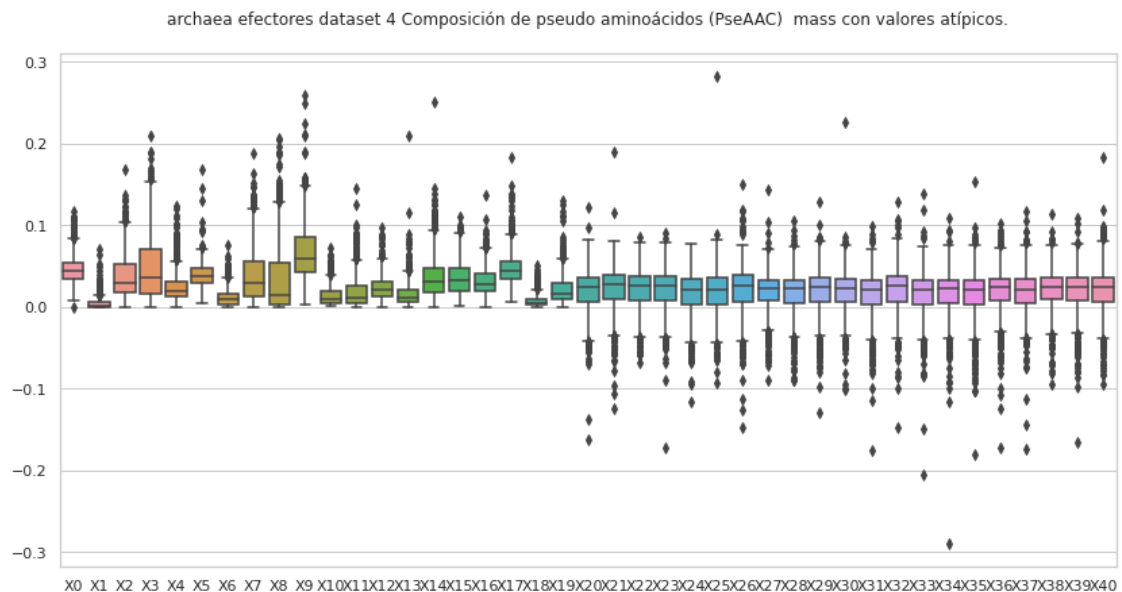
| | X5 | X6 | X7 | X8 | X9 | ... | \ |
|-------|-------------|-------------|-------------|-------------|-------------|-----|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | ... | |
| mean | 0.038715 | 0.010639 | 0.028185 | 0.020060 | 0.047863 | ... | |
| std | 0.014721 | 0.010111 | 0.025266 | 0.027231 | 0.027760 | ... | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | ... | |
| 25% | 0.029618 | 0.003444 | 0.011196 | 0.003542 | 0.029029 | ... | |
| 50% | 0.036459 | 0.008166 | 0.020944 | 0.009674 | 0.041454 | ... | |
| 75% | 0.045946 | 0.014531 | 0.036672 | 0.025049 | 0.058953 | ... | |

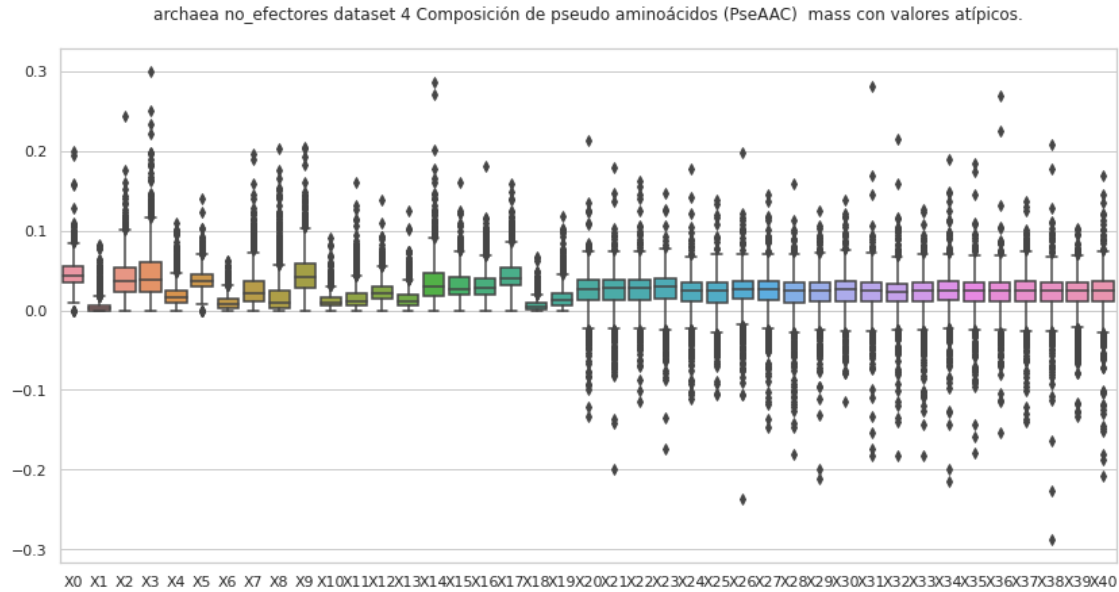
| | | | | | | |
|-----|----------|----------|----------|----------|----------|-----|
| max | 0.140677 | 0.062930 | 0.196156 | 0.204046 | 0.205667 | ... |
|-----|----------|----------|----------|----------|----------|-----|

| | X31 | X32 | X33 | X34 | X35 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.021401 | 0.020976 | 0.020316 | 0.022385 | 0.020540 |
| std | 0.027888 | 0.027317 | 0.027677 | 0.028912 | 0.028330 |
| min | -0.182949 | -0.182817 | -0.183053 | -0.214337 | -0.179244 |
| 25% | 0.010440 | 0.010699 | 0.011086 | 0.012361 | 0.011066 |
| 50% | 0.025235 | 0.023680 | 0.024934 | 0.025633 | 0.024713 |
| 75% | 0.035596 | 0.033785 | 0.035484 | 0.035845 | 0.034866 |
| max | 0.281287 | 0.214619 | 0.127458 | 0.189781 | 0.184237 |

| | X36 | X37 | X38 | X39 | X40 |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.021275 | 0.022055 | 0.020885 | 0.021217 | 0.021103 |
| std | 0.026700 | 0.027357 | 0.029846 | 0.025362 | 0.031268 |
| min | -0.153414 | -0.140642 | -0.287852 | -0.134008 | -0.208300 |
| 25% | 0.010645 | 0.011345 | 0.010836 | 0.011995 | 0.010372 |
| 50% | 0.024942 | 0.025094 | 0.024482 | 0.024941 | 0.025118 |
| 75% | 0.034694 | 0.036691 | 0.035149 | 0.034991 | 0.036162 |
| max | 0.269147 | 0.136360 | 0.208753 | 0.103635 | 0.170019 |

[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) +",\n"
             ' ' + 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 4,
sin valores atípicos.
Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.046579 | 0.000000 | 0.013483 | 0.030644 | 0.023290 | 0.026967 | 0.002452 |
| 1 | 0.057973 | 0.002761 | 0.024846 | 0.016564 | 0.019324 | 0.044170 | 0.013803 |
| 2 | 0.039807 | 0.007961 | 0.023884 | 0.071653 | 0.015923 | 0.027865 | 0.015923 |
| 3 | 0.055195 | 0.000000 | 0.022341 | 0.018398 | 0.021027 | 0.028912 | 0.007885 |
| 4 | 0.038296 | 0.000580 | 0.016247 | 0.011605 | 0.009864 | 0.030173 | 0.001741 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 994 | 0.023042 | 0.007681 | 0.007681 | 0.030723 | 0.000000 | 0.030723 | 0.000000 |
| 995 | 0.084424 | 0.004966 | 0.024831 | 0.129119 | 0.034763 | 0.064560 | 0.014898 |
| 996 | 0.076709 | 0.006670 | 0.063369 | 0.073374 | 0.020011 | 0.023346 | 0.020011 |
| 998 | 0.037362 | 0.000000 | 0.021350 | 0.048037 | 0.013344 | 0.040031 | 0.008006 |
| 999 | 0.046876 | 0.005068 | 0.072214 | 0.057011 | 0.012669 | 0.046876 | 0.016470 |

| | X7 | X8 | X9 ... | X32 | X33 | X34 \ |
|-----|----------|----------|--------------|-----------|----------|----------|
| 0 | 0.033096 | 0.004903 | 0.056385 ... | 0.034216 | 0.023639 | 0.022493 |
| 1 | 0.019324 | 0.005521 | 0.063495 ... | 0.044150 | 0.046488 | 0.037149 |
| 2 | 0.055730 | 0.059710 | 0.075633 ... | 0.007544 | 0.012877 | 0.035787 |
| 3 | 0.022341 | 0.000000 | 0.051253 ... | 0.044267 | 0.031211 | 0.017043 |
| 4 | 0.008704 | 0.002901 | 0.033074 ... | 0.041681 | 0.041562 | 0.023469 |
| .. | ... | ... | ... | ... | ... | ... |
| 994 | 0.046084 | 0.023042 | 0.015361 ... | 0.042825 | 0.051565 | 0.015456 |
| 995 | 0.059594 | 0.054627 | 0.094357 ... | -0.045545 | 0.037955 | 0.006187 |
| 996 | 0.050028 | 0.013341 | 0.073374 ... | 0.015526 | 0.025862 | 0.013840 |
| 998 | 0.066719 | 0.037362 | 0.077394 ... | -0.001604 | 0.037907 | 0.032740 |
| 999 | 0.016470 | 0.003801 | 0.035474 ... | 0.017255 | 0.024744 | 0.024554 |

| | X35 | X36 | X37 | X38 | X39 | X40 | X41 |
|-----|-----------|----------|----------|-----------|----------|-----------|-----------|
| 0 | 0.001507 | 0.045164 | 0.028675 | 0.031224 | 0.039413 | 0.030396 | efectores |
| 1 | 0.024663 | 0.021956 | 0.028845 | 0.035633 | 0.007307 | -0.001941 | efectores |
| 2 | -0.000630 | 0.015728 | 0.018898 | -0.003788 | 0.020157 | 0.016012 | efectores |
| 3 | 0.005837 | 0.034860 | 0.023530 | 0.043917 | 0.041360 | 0.038021 | efectores |
| 4 | 0.033670 | 0.025594 | 0.029789 | 0.033151 | 0.032052 | 0.036057 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 994 | -0.002608 | 0.072808 | 0.023841 | 0.030310 | 0.034872 | 0.064427 | efectores |
| 995 | 0.026439 | 0.008540 | 0.027846 | 0.000493 | 0.024219 | -0.012983 | efectores |
| 996 | 0.010002 | 0.000133 | 0.008639 | 0.009287 | 0.012895 | 0.019535 | efectores |
| 998 | 0.037607 | 0.036380 | 0.028502 | 0.009977 | 0.026227 | 0.029087 | efectores |
| 999 | 0.025456 | 0.029475 | 0.019395 | 0.018638 | 0.009658 | 0.011527 | efectores |

[808 rows x 42 columns]

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

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|-------|------------|------------|------------|------------|------------|------------|---|
| count | 808.000000 | 808.000000 | 808.000000 | 808.000000 | 808.000000 | 808.000000 | |
| mean | 0.044968 | 0.003011 | 0.031720 | 0.038010 | 0.020377 | 0.037494 | |
| std | 0.013536 | 0.004494 | 0.020790 | 0.030732 | 0.013282 | 0.010664 | |
| min | 0.008713 | 0.000000 | 0.003623 | 0.000000 | 0.000000 | 0.007202 | |
| 25% | 0.035475 | 0.000000 | 0.016245 | 0.013220 | 0.011581 | 0.029333 | |
| 50% | 0.042997 | 0.000958 | 0.026263 | 0.027494 | 0.017193 | 0.035898 | |
| 75% | 0.052317 | 0.004554 | 0.043603 | 0.057895 | 0.026404 | 0.044558 | |
| max | 0.090599 | 0.024531 | 0.110556 | 0.148229 | 0.074585 | 0.074159 | |

| | X6 | X7 | X8 | X9 | ... | X31 | \ |
|-------|------------|------------|------------|------------|-----|------------|---|
| count | 808.000000 | 808.000000 | 808.000000 | 808.000000 | ... | 808.000000 | |
| mean | 0.009475 | 0.031497 | 0.025068 | 0.057543 | ... | 0.020626 | |
| std | 0.007691 | 0.026715 | 0.032250 | 0.025131 | ... | 0.019710 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.002487 | ... | -0.058509 | |
| 25% | 0.002905 | 0.011196 | 0.002768 | 0.039749 | ... | 0.009988 | |
| 50% | 0.007634 | 0.022664 | 0.007993 | 0.052972 | ... | 0.024059 | |
| 75% | 0.014514 | 0.045941 | 0.039445 | 0.073999 | ... | 0.033897 | |
| max | 0.039603 | 0.124066 | 0.146833 | 0.158514 | ... | 0.081129 | |

| | X32 | X33 | X34 | X35 | X36 | X37 | \ |
|-------|------------|------------|------------|------------|------------|------------|---|
| count | 808.000000 | 808.000000 | 808.000000 | 808.000000 | 808.000000 | 808.000000 | |
| mean | 0.024553 | 0.019485 | 0.021172 | 0.020731 | 0.023211 | 0.021827 | |
| std | 0.020243 | 0.020858 | 0.020659 | 0.020359 | 0.019432 | 0.021865 | |
| min | -0.048208 | -0.058898 | -0.055411 | -0.067017 | -0.049584 | -0.046286 | |
| 25% | 0.013181 | 0.008338 | 0.010793 | 0.009313 | 0.013763 | 0.010093 | |
| 50% | 0.028244 | 0.023197 | 0.025297 | 0.023953 | 0.026292 | 0.024089 | |
| 75% | 0.038745 | 0.034093 | 0.034425 | 0.034829 | 0.035100 | 0.035406 | |

| | | | | | | |
|-----|----------|----------|----------|----------|----------|----------|
| max | 0.084128 | 0.092703 | 0.080708 | 0.085756 | 0.089706 | 0.092484 |
|-----|----------|----------|----------|----------|----------|----------|

| | | | |
|-------|------------|------------|------------|
| | X38 | X39 | X40 |
| count | 808.000000 | 808.000000 | 808.000000 |
| mean | 0.024303 | 0.024484 | 0.023676 |
| std | 0.019309 | 0.019236 | 0.020868 |
| min | -0.047198 | -0.049900 | -0.053054 |
| 25% | 0.013391 | 0.014206 | 0.010894 |
| 50% | 0.027750 | 0.026717 | 0.027173 |
| 75% | 0.037431 | 0.036221 | 0.036735 |
| max | 0.078941 | 0.092105 | 0.097809 |

[8 rows x 41 columns]

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

| | | | | | | | |
|-----|----------|----------|----------|----------|----------|----------|----------|
| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
| 0 | 0.029432 | 0.002848 | 0.015191 | 0.019938 | 0.028483 | 0.036078 | 0.000949 |
| 1 | 0.034803 | 0.004872 | 0.010441 | 0.012529 | 0.027147 | 0.035500 | 0.004872 |
| 2 | 0.035941 | 0.035941 | 0.000000 | 0.035941 | 0.026956 | 0.017971 | 0.000000 |
| 3 | 0.023428 | 0.007809 | 0.015619 | 0.011714 | 0.007809 | 0.023428 | 0.000000 |
| 5 | 0.062719 | 0.000000 | 0.059992 | 0.065446 | 0.016361 | 0.039540 | 0.019088 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 992 | 0.060160 | 0.008021 | 0.040107 | 0.072192 | 0.012032 | 0.028075 | 0.012032 |
| 993 | 0.039372 | 0.004256 | 0.040436 | 0.080872 | 0.018090 | 0.060654 | 0.018090 |
| 994 | 0.023224 | 0.000664 | 0.028533 | 0.023888 | 0.004645 | 0.030523 | 0.010617 |
| 995 | 0.032619 | 0.001255 | 0.002509 | 0.011291 | 0.023837 | 0.032619 | 0.003764 |
| 999 | 0.051691 | 0.002068 | 0.043420 | 0.057894 | 0.022744 | 0.037217 | 0.020676 |

| | | | | | | | |
|-----|----------|----------|----------|-----|----------|----------|-----------|
| | X7 | X8 | X9 | ... | X32 | X33 | X34 \ |
| 0 | 0.056966 | 0.009494 | 0.074055 | ... | 0.033843 | 0.012361 | 0.020728 |
| 1 | 0.039676 | 0.016010 | 0.023666 | ... | 0.040298 | 0.019902 | 0.035659 |
| 2 | 0.008985 | 0.000000 | 0.026956 | ... | 0.005028 | 0.043130 | 0.061577 |
| 3 | 0.031237 | 0.035142 | 0.027332 | ... | 0.033148 | 0.028823 | 0.016893 |
| 5 | 0.023179 | 0.004090 | 0.055902 | ... | 0.020125 | 0.038867 | 0.010335 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 992 | 0.028075 | 0.020053 | 0.072192 | ... | 0.004587 | 0.063293 | -0.014687 |
| 993 | 0.035115 | 0.021282 | 0.048949 | ... | 0.011979 | 0.021465 | 0.003324 |
| 994 | 0.021897 | 0.003981 | 0.019243 | ... | 0.028216 | 0.041480 | 0.035890 |
| 995 | 0.027601 | 0.001255 | 0.048928 | ... | 0.045248 | 0.013821 | 0.013952 |
| 999 | 0.016541 | 0.012406 | 0.055826 | ... | 0.013469 | 0.043790 | -0.003707 |

| | | | | | | | |
|---|----------|----------|----------|----------|----------|----------|--------------|
| | X35 | X36 | X37 | X38 | X39 | X40 | X41 |
| 0 | 0.017941 | 0.001701 | 0.021919 | 0.018489 | 0.019195 | 0.023560 | no_efectores |

| | | | | | | | |
|-----|----------|-----------|-----------|-----------|----------|----------|--------------|
| 1 | 0.031664 | 0.034688 | 0.037143 | 0.029192 | 0.024139 | 0.037016 | no_efectores |
| 2 | 0.035119 | 0.011055 | 0.004903 | -0.016088 | 0.028537 | 0.028916 | no_efectores |
| 3 | 0.037868 | 0.032337 | 0.016353 | 0.016134 | 0.035836 | 0.022275 | no_efectores |
| 5 | 0.023922 | 0.017129 | -0.007926 | 0.015499 | 0.008093 | 0.000061 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 992 | 0.016906 | -0.001782 | -0.009576 | 0.012616 | 0.054853 | 0.012614 | no_efectores |
| 993 | 0.017297 | 0.031423 | 0.024279 | 0.021989 | 0.025521 | 0.010368 | no_efectores |
| 994 | 0.030321 | 0.025467 | 0.020674 | 0.023133 | 0.028381 | 0.043709 | no_efectores |
| 995 | 0.026924 | 0.015931 | 0.029711 | 0.030964 | 0.053751 | 0.040198 | no_efectores |
| 999 | 0.031215 | 0.031538 | 0.001543 | 0.039600 | 0.001382 | 0.047266 | no_efectores |

[821 rows x 42 columns]

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

Estadísticas.

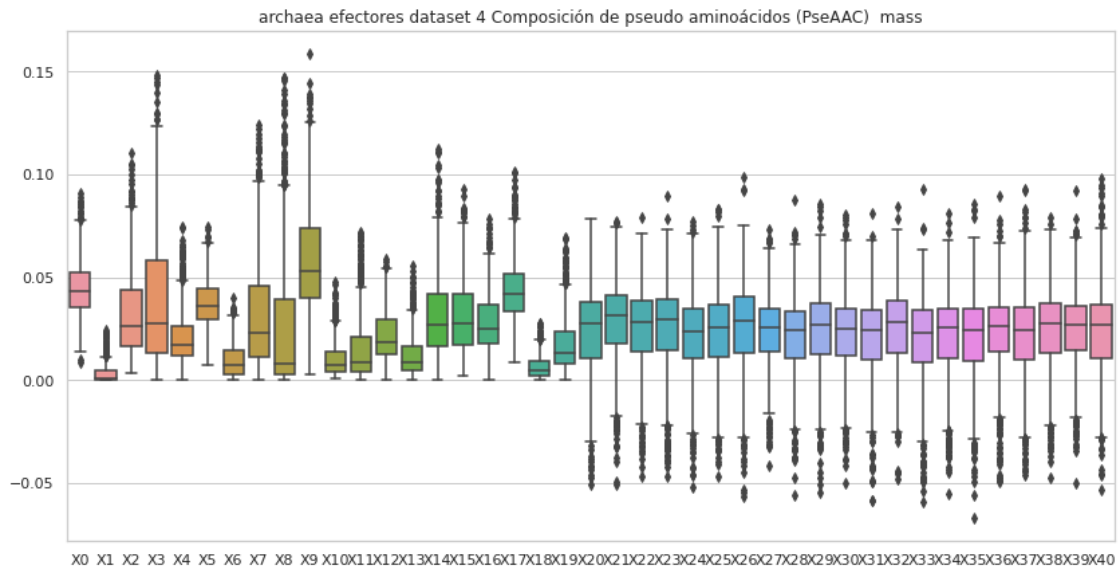
| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 821.000000 | 821.000000 | 821.000000 | 821.000000 | 821.000000 | 821.000000 |
| mean | 0.044705 | 0.004201 | 0.036922 | 0.039153 | 0.016180 | 0.036663 |
| std | 0.014897 | 0.005448 | 0.020782 | 0.024798 | 0.010363 | 0.011294 |
| min | 0.008836 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 0.034738 | 0.000000 | 0.021584 | 0.021000 | 0.008771 | 0.029406 |
| 50% | 0.043142 | 0.002474 | 0.034709 | 0.035864 | 0.013786 | 0.035745 |
| 75% | 0.052971 | 0.006055 | 0.049226 | 0.053322 | 0.021597 | 0.042960 |
| max | 0.100769 | 0.035941 | 0.111443 | 0.136970 | 0.058218 | 0.077889 |

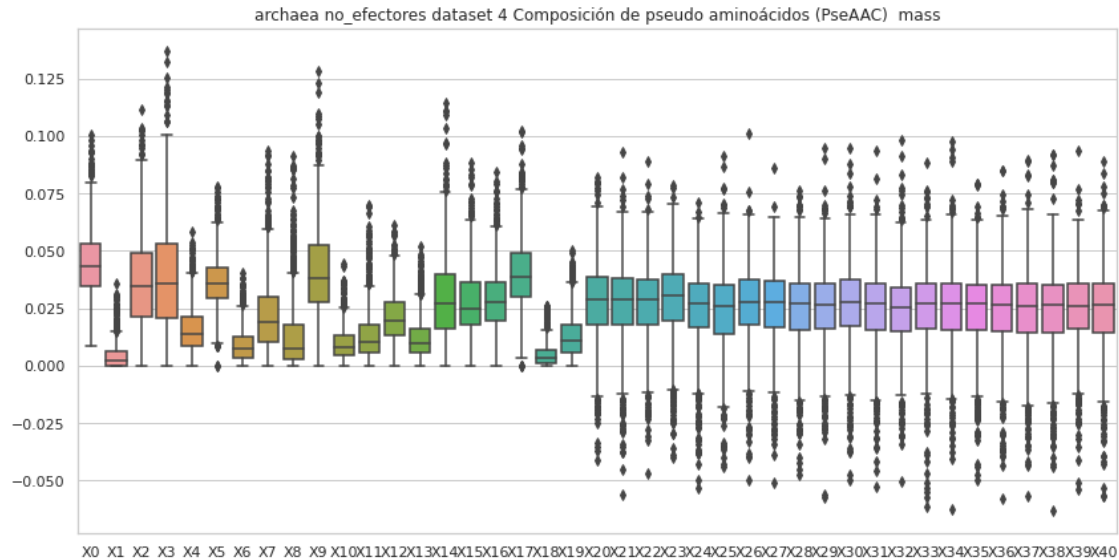
| | X6 | X7 | X8 | X9 ... | X31 \ |
|-------|------------|------------|------------|------------|------------|
| count | 821.000000 | 821.000000 | 821.000000 | 821.000000 | 821.000000 |
| mean | 0.009169 | 0.022873 | 0.013804 | 0.041471 | 0.024889 |
| std | 0.007385 | 0.017056 | 0.016295 | 0.019666 | 0.017682 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | -0.052723 |
| 25% | 0.003658 | 0.010213 | 0.002952 | 0.027895 | 0.015386 |
| 50% | 0.007646 | 0.018816 | 0.007603 | 0.037893 | 0.027172 |
| 75% | 0.012731 | 0.029939 | 0.017996 | 0.052615 | 0.035916 |
| max | 0.040563 | 0.093351 | 0.091324 | 0.128531 | 0.093453 |

| | X32 | X33 | X34 | X35 | X36 | X37 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 821.000000 | 821.000000 | 821.000000 | 821.000000 | 821.000000 | 821.000000 |
| mean | 0.024223 | 0.024954 | 0.025184 | 0.024178 | 0.024439 | 0.023998 |
| std | 0.016818 | 0.017836 | 0.017581 | 0.017490 | 0.017685 | 0.018937 |
| min | -0.050218 | -0.061375 | -0.062739 | -0.049938 | -0.057731 | -0.056631 |
| 25% | 0.015108 | 0.016294 | 0.015615 | 0.015577 | 0.014865 | 0.014543 |
| 50% | 0.025559 | 0.027050 | 0.027001 | 0.026847 | 0.026546 | 0.025935 |
| 75% | 0.034181 | 0.035797 | 0.035679 | 0.034990 | 0.035381 | 0.036050 |
| max | 0.098018 | 0.088374 | 0.097564 | 0.079429 | 0.085070 | 0.089758 |

| | X38 | X39 | X40 |
|-------|------------|------------|------------|
| count | 821.000000 | 821.000000 | 821.000000 |
| mean | 0.024412 | 0.025057 | 0.024740 |
| std | 0.018439 | 0.017192 | 0.018549 |
| min | -0.063134 | -0.053821 | -0.056812 |
| 25% | 0.014389 | 0.016244 | 0.014642 |
| 50% | 0.026363 | 0.026157 | 0.026720 |
| 75% | 0.035177 | 0.035540 | 0.036010 |
| max | 0.091879 | 0.093918 | 0.089127 |

[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 4,
con valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.027611 | 0.000000 | 0.007993 | 0.018165 | 0.013806 | 0.015985 | 0.001453 |
| 1 | 0.049468 | 0.002356 | 0.021201 | 0.014134 | 0.016489 | 0.037690 | 0.011778 |
| 2 | 0.022212 | 0.004442 | 0.013327 | 0.039982 | 0.008885 | 0.015549 | 0.008885 |
| 3 | 0.039119 | 0.000000 | 0.015834 | 0.013040 | 0.014902 | 0.020491 | 0.005588 |
| 4 | 0.029845 | 0.000452 | 0.012661 | 0.009044 | 0.007687 | 0.023514 | 0.001357 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.048653 | 0.002862 | 0.014310 | 0.074411 | 0.020034 | 0.037205 | 0.008586 |
| 996 | 0.072997 | 0.006348 | 0.060302 | 0.069823 | 0.019043 | 0.022217 | 0.019043 |
| 997 | 0.039024 | 0.004591 | 0.052797 | 0.066570 | 0.032137 | 0.016069 | 0.016069 |
| 998 | 0.053067 | 0.000000 | 0.030324 | 0.068229 | 0.018953 | 0.056858 | 0.011372 |
| 999 | 0.041860 | 0.004525 | 0.064487 | 0.050911 | 0.011313 | 0.041860 | 0.014708 |

| | X7 | X8 | X9 ... | X53 | X54 | X55 \ |
|-----|----------|----------|--------------|-----------|-----------|-----------|
| 0 | 0.019618 | 0.002906 | 0.033424 ... | 0.016774 | 0.014382 | 0.010435 |
| 1 | 0.016489 | 0.004711 | 0.054180 ... | -0.004904 | 0.024098 | -0.009235 |
| 2 | 0.031097 | 0.033318 | 0.042203 ... | 0.034902 | 0.016950 | 0.034704 |
| 3 | 0.015834 | 0.000000 | 0.036325 ... | 0.003246 | 0.021194 | -0.000718 |
| 4 | 0.006783 | 0.002261 | 0.025775 ... | 0.012421 | 0.019115 | 0.007264 |
| .. | ... | ... | ... | ... | ... | |
| 995 | 0.034343 | 0.031481 | 0.054377 ... | 0.027628 | 0.026762 | 0.009795 |
| 996 | 0.047607 | 0.012695 | 0.069823 ... | -0.000304 | 0.016518 | 0.047911 |
| 997 | 0.043615 | 0.064275 | 0.055092 ... | 0.016600 | 0.077859 | 0.043262 |
| 998 | 0.094763 | 0.053067 | 0.109925 ... | 0.041817 | -0.008676 | 0.023714 |
| 999 | 0.014708 | 0.003394 | 0.031678 ... | 0.019477 | 0.010458 | 0.036777 |

| | X56 | X57 | X58 | X59 | X60 | X61 | X62 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.022332 | 0.008298 | 0.019030 | 0.007098 | 0.017238 | -0.001771 | efectores |
| 1 | 0.018352 | -0.003821 | 0.041888 | -0.003337 | 0.046064 | 0.014283 | efectores |
| 2 | -0.031376 | 0.000904 | -0.030821 | -0.006912 | 0.002588 | 0.021542 | efectores |
| 3 | 0.026402 | -0.007980 | 0.034337 | -0.004029 | 0.020352 | -0.004574 | efectores |
| 4 | 0.025779 | 0.007673 | 0.023268 | 0.006469 | 0.019448 | 0.004721 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | -0.019719 | 0.001883 | 0.014384 | 0.017590 | -0.012253 | -0.018906 | efectores |
| 996 | 0.022711 | 0.005521 | 0.016752 | 0.023158 | -0.006670 | -0.010255 | efectores |
| 997 | 0.001713 | 0.084811 | -0.055581 | -0.018475 | 0.005956 | 0.034462 | efectores |

```

998 0.095654 0.050548 0.013192 0.024318 -0.035931 -0.000490 efectores
999 -0.006312 0.026010 -0.002245 0.030193 0.017633 0.040308 efectores

```

[1000 rows x 63 columns]

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

| | X0 | X1 | X2 | X3 | X4 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.047118 | 0.004393 | 0.036816 | 0.043998 | 0.020941 | |
| std | 0.028609 | 0.007362 | 0.026230 | 0.032002 | 0.014446 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.028821 | 0.000000 | 0.013686 | 0.011681 | 0.011159 | |
| 50% | 0.041381 | 0.001505 | 0.030886 | 0.044551 | 0.017655 | |
| 75% | 0.059031 | 0.005626 | 0.055170 | 0.067799 | 0.027282 | |
| max | 0.376157 | 0.061576 | 0.198151 | 0.185950 | 0.126096 | |

| | X5 | X6 | X7 | X8 | X9 | ... | \ |
|-------|-------------|-------------|-------------|-------------|-------------|-----|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | ... | |
| mean | 0.039697 | 0.010852 | 0.034893 | 0.030074 | 0.060320 | ... | |
| std | 0.022083 | 0.009907 | 0.030419 | 0.035697 | 0.030366 | ... | |
| min | 0.001587 | 0.000000 | 0.000000 | 0.000000 | 0.002977 | ... | |
| 25% | 0.024692 | 0.003095 | 0.012574 | 0.003272 | 0.039322 | ... | |
| 50% | 0.034614 | 0.008974 | 0.024115 | 0.013025 | 0.054187 | ... | |
| 75% | 0.049818 | 0.014972 | 0.049755 | 0.052075 | 0.074883 | ... | |
| max | 0.217977 | 0.072318 | 0.198151 | 0.234178 | 0.258159 | ... | |

| | X52 | X53 | X54 | X55 | X56 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.007812 | 0.011085 | 0.011901 | 0.015452 | 0.006511 | |
| std | 0.036767 | 0.026762 | 0.033633 | 0.031749 | 0.036559 | |
| min | -0.207928 | -0.132477 | -0.396017 | -0.159899 | -0.269039 | |
| 25% | -0.005858 | -0.001737 | -0.003134 | 0.000356 | -0.010240 | |
| 50% | 0.014734 | 0.007663 | 0.016513 | 0.010546 | 0.013630 | |
| 75% | 0.026821 | 0.022583 | 0.027701 | 0.030386 | 0.025677 | |
| max | 0.182170 | 0.208379 | 0.148285 | 0.204530 | 0.182043 | |

| | X57 | X58 | X59 | X60 | X61 |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.010430 | 0.009597 | 0.012612 | 0.008138 | 0.012789 |
| std | 0.031732 | 0.034506 | 0.029434 | 0.041198 | 0.031471 |
| min | -0.234001 | -0.341253 | -0.212960 | -0.505870 | -0.225521 |
| 25% | -0.002574 | -0.004839 | -0.000302 | -0.005218 | -0.000457 |
| 50% | 0.006557 | 0.015586 | 0.009430 | 0.013505 | 0.008582 |
| 75% | 0.022389 | 0.027681 | 0.026609 | 0.027553 | 0.026085 |

max 0.234285 0.163220 0.190977 0.205921 0.150189

[8 rows x 62 columns]

no_efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.018914 | 0.001830 | 0.009762 | 0.012813 | 0.018304 | 0.023185 | 0.000610 |
| 1 | 0.033796 | 0.004731 | 0.010139 | 0.012166 | 0.026361 | 0.034471 | 0.004731 |
| 2 | 0.135582 | 0.135582 | 0.000000 | 0.135582 | 0.101686 | 0.067791 | 0.000000 |
| 3 | 0.064481 | 0.021494 | 0.042987 | 0.032240 | 0.021494 | 0.064481 | 0.000000 |
| 4 | 0.075013 | 0.030005 | 0.052509 | 0.015003 | 0.000000 | 0.030005 | 0.007501 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.025153 | 0.000967 | 0.001935 | 0.008707 | 0.018381 | 0.025153 | 0.002902 |
| 996 | 0.042891 | 0.000000 | 0.009191 | 0.085781 | 0.021445 | 0.024509 | 0.006127 |
| 997 | 0.018549 | 0.000000 | 0.000000 | 0.006183 | 0.012366 | 0.018549 | 0.000000 |
| 998 | 0.021533 | 0.004307 | 0.043066 | 0.060293 | 0.008613 | 0.038760 | 0.008613 |
| 999 | 0.050623 | 0.002025 | 0.042523 | 0.056698 | 0.022274 | 0.036448 | 0.020249 |

| | X7 | X8 | X9 ... | X53 | X54 | X55 \ |
|-----|----------|----------|--------------|-----------|-----------|-----------|
| 0 | 0.036608 | 0.006101 | 0.047591 ... | 0.000408 | 0.017985 | -0.004583 |
| 1 | 0.038527 | 0.015546 | 0.022981 ... | -0.004850 | 0.006537 | 0.001504 |
| 2 | 0.033895 | 0.000000 | 0.101686 ... | 0.076306 | -0.240220 | -0.081055 |
| 3 | 0.085975 | 0.096721 | 0.075228 ... | 0.059144 | -0.080208 | -0.003245 |
| 4 | 0.000000 | 0.007501 | 0.037506 ... | 0.039277 | 0.023833 | 0.041531 |
| .. | ... | ... | ... | ... | ... | |
| 995 | 0.021283 | 0.000967 | 0.037730 ... | 0.010759 | 0.027087 | 0.011318 |
| 996 | 0.064336 | 0.076591 | 0.079654 ... | 0.029700 | 0.034731 | 0.045942 |
| 997 | 0.006183 | 0.006183 | 0.098929 ... | -0.014401 | 0.020554 | 0.008685 |
| 998 | 0.021533 | 0.008613 | 0.051680 ... | 0.051442 | 0.022868 | 0.071123 |
| 999 | 0.016199 | 0.012149 | 0.054673 ... | 0.025313 | -0.021729 | 0.000695 |

| | X56 | X57 | X58 | X59 | X60 | X61 | X62 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | 0.031985 | 0.007837 | 0.037285 | 0.012985 | 0.021042 | 0.001462 | no_efectores |
| 1 | 0.013538 | 0.004285 | 0.019999 | 0.006917 | 0.016259 | -0.001023 | no_efectores |
| 2 | 0.018213 | -0.036751 | 0.147698 | 0.065429 | -0.206395 | -0.131945 | no_efectores |
| 3 | -0.017329 | -0.009387 | -0.036127 | -0.014984 | -0.094422 | -0.021174 | no_efectores |
| 4 | -0.021926 | -0.024407 | -0.022443 | -0.025921 | -0.018124 | 0.015420 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.017136 | 0.001955 | 0.020126 | -0.001387 | 0.016038 | 0.001604 | no_efectores |
| 996 | -0.048813 | -0.003308 | -0.094014 | -0.048081 | 0.027020 | 0.051339 | no_efectores |
| 997 | -0.024568 | -0.031979 | -0.014909 | -0.021821 | -0.000365 | 0.007231 | no_efectores |
| 998 | -0.012141 | 0.024650 | -0.011462 | 0.017585 | -0.042510 | 0.015815 | no_efectores |

999 -0.019738 0.012556 0.005078 0.013906 0.033382 0.006637 no_efectores

[1000 rows x 63 columns]

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

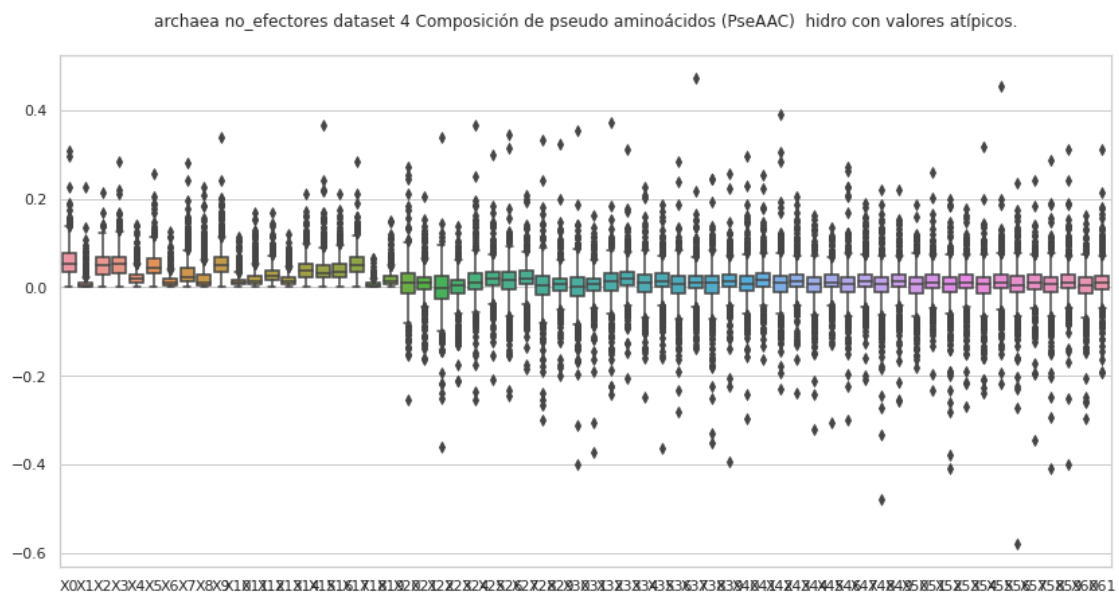
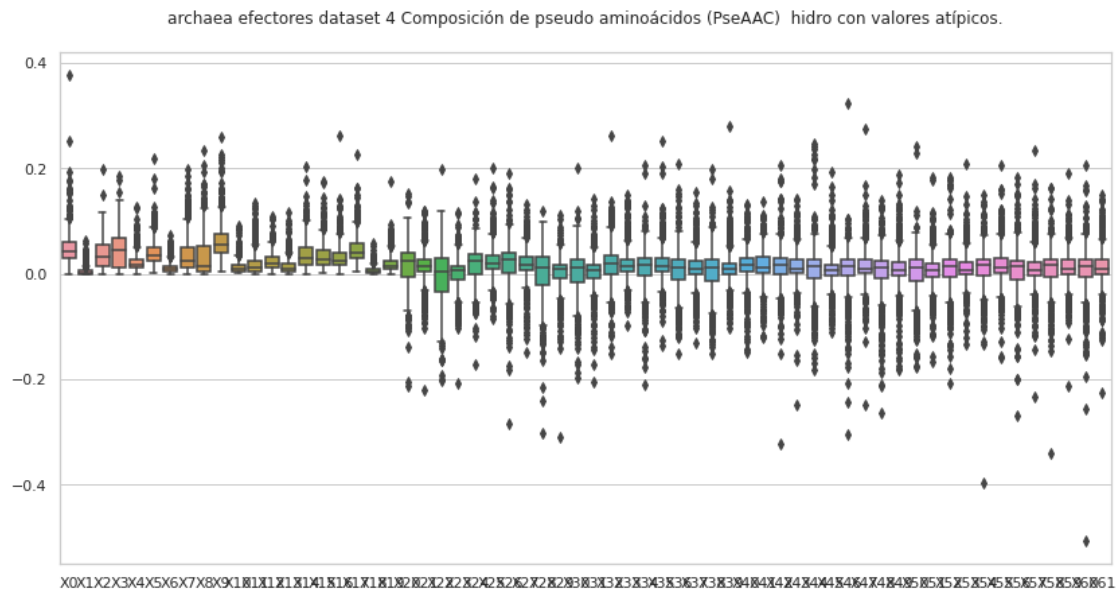
| | X0 | X1 | X2 | X3 | X4 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.058225 | 0.007420 | 0.048310 | 0.051761 | 0.022154 | |
| std | 0.032776 | 0.014143 | 0.026984 | 0.030881 | 0.017006 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.035375 | 0.000000 | 0.027576 | 0.030942 | 0.011187 | |
| 50% | 0.054160 | 0.003387 | 0.050431 | 0.051742 | 0.018633 | |
| 75% | 0.076088 | 0.009381 | 0.066867 | 0.069040 | 0.028244 | |
| max | 0.307732 | 0.226016 | 0.215332 | 0.282521 | 0.140622 | |

| | X5 | X6 | X7 | X8 | X9 | ... | \ |
|-------|-------------|-------------|-------------|-------------|-------------|-----|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | ... | |
| mean | 0.049638 | 0.013005 | 0.032451 | 0.021255 | 0.055096 | ... | |
| std | 0.027773 | 0.013314 | 0.029594 | 0.027278 | 0.031471 | ... | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | ... | |
| 25% | 0.030751 | 0.003769 | 0.013963 | 0.004541 | 0.035631 | ... | |
| 50% | 0.045143 | 0.010262 | 0.023577 | 0.010902 | 0.048887 | ... | |
| 75% | 0.064713 | 0.018719 | 0.042205 | 0.027632 | 0.066925 | ... | |
| max | 0.255568 | 0.126183 | 0.281243 | 0.224995 | 0.337492 | ... | |

| | X52 | X53 | X54 | X55 | X56 | \ |
|-------|-------------|-------------|-------------|-------------|-------------|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.002634 | 0.011165 | 0.003298 | 0.012318 | 0.003157 | |
| std | 0.042913 | 0.035820 | 0.039440 | 0.038305 | 0.044089 | |
| min | -0.408746 | -0.269975 | -0.240220 | -0.217127 | -0.579723 | |
| 25% | -0.011332 | -0.002299 | -0.013731 | -0.003267 | -0.012251 | |
| 50% | 0.006613 | 0.010385 | 0.005866 | 0.010125 | 0.005536 | |
| 75% | 0.022699 | 0.027876 | 0.022214 | 0.029358 | 0.023937 | |
| max | 0.197480 | 0.176400 | 0.318210 | 0.454171 | 0.234814 | |

| | X57 | X58 | X59 | X60 | X61 |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.010789 | 0.002922 | 0.012572 | 0.000960 | 0.011563 |
| std | 0.035013 | 0.041426 | 0.037285 | 0.038892 | 0.035859 |
| min | -0.346299 | -0.408862 | -0.400394 | -0.297016 | -0.192974 |
| 25% | -0.003623 | -0.012469 | -0.002612 | -0.014501 | -0.003778 |
| 50% | 0.009917 | 0.006256 | 0.011151 | 0.004197 | 0.009361 |
| 75% | 0.027624 | 0.023022 | 0.027393 | 0.021827 | 0.025934 |
| max | 0.240099 | 0.285716 | 0.310272 | 0.162786 | 0.310821 |

```
[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 4, sin valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.027611 | 0.000000 | 0.007993 | 0.018165 | 0.013806 | 0.015985 | 0.001453 |
| 1 | 0.049468 | 0.002356 | 0.021201 | 0.014134 | 0.016489 | 0.037690 | 0.011778 |
| 2 | 0.022212 | 0.004442 | 0.013327 | 0.039982 | 0.008885 | 0.015549 | 0.008885 |
| 3 | 0.039119 | 0.000000 | 0.015834 | 0.013040 | 0.014902 | 0.020491 | 0.005588 |
| 4 | 0.029845 | 0.000452 | 0.012661 | 0.009044 | 0.007687 | 0.023514 | 0.001357 |
| .. | ... | ... | ... | ... | ... | ... | |
| 993 | 0.024076 | 0.000000 | 0.010834 | 0.010032 | 0.008427 | 0.021668 | 0.000803 |
| 995 | 0.048653 | 0.002862 | 0.014310 | 0.074411 | 0.020034 | 0.037205 | 0.008586 |
| 996 | 0.072997 | 0.006348 | 0.060302 | 0.069823 | 0.019043 | 0.022217 | 0.019043 |
| 997 | 0.039024 | 0.004591 | 0.052797 | 0.066570 | 0.032137 | 0.016069 | 0.016069 |
| 999 | 0.041860 | 0.004525 | 0.064487 | 0.050911 | 0.011313 | 0.041860 | 0.014708 |

| | X7 | X8 | X9 ... | X53 | X54 | X55 \ |
|-----|----------|----------|--------------|-----------|----------|-----------|
| 0 | 0.019618 | 0.002906 | 0.033424 ... | 0.016774 | 0.014382 | 0.010435 |
| 1 | 0.016489 | 0.004711 | 0.054180 ... | -0.004904 | 0.024098 | -0.009235 |
| 2 | 0.031097 | 0.033318 | 0.042203 ... | 0.034902 | 0.016950 | 0.034704 |
| 3 | 0.015834 | 0.000000 | 0.036325 ... | 0.003246 | 0.021194 | -0.000718 |
| 4 | 0.006783 | 0.002261 | 0.025775 ... | 0.012421 | 0.019115 | 0.007264 |
| .. | ... | ... | ... | ... | ... | |
| 993 | 0.008427 | 0.002408 | 0.024076 ... | 0.017438 | 0.018092 | 0.011518 |
| 995 | 0.034343 | 0.031481 | 0.054377 ... | 0.027628 | 0.026762 | 0.009795 |
| 996 | 0.047607 | 0.012695 | 0.069823 ... | -0.000304 | 0.016518 | 0.047911 |
| 997 | 0.043615 | 0.064275 | 0.055092 ... | 0.016600 | 0.077859 | 0.043262 |
| 999 | 0.014708 | 0.003394 | 0.031678 ... | 0.019477 | 0.010458 | 0.036777 |

| | X56 | X57 | X58 | X59 | X60 | X61 | X62 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.022332 | 0.008298 | 0.019030 | 0.007098 | 0.017238 | -0.001771 | efectores |
| 1 | 0.018352 | -0.003821 | 0.041888 | -0.003337 | 0.046064 | 0.014283 | efectores |
| 2 | -0.031376 | 0.000904 | -0.030821 | -0.006912 | 0.002588 | 0.021542 | efectores |
| 3 | 0.026402 | -0.007980 | 0.034337 | -0.004029 | 0.020352 | -0.004574 | efectores |
| 4 | 0.025779 | 0.007673 | 0.023268 | 0.006469 | 0.019448 | 0.004721 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 993 | 0.019487 | 0.011248 | 0.020299 | 0.009025 | 0.020710 | 0.008374 | efectores |
| 995 | -0.019719 | 0.001883 | 0.014384 | 0.017590 | -0.012253 | -0.018906 | efectores |
| 996 | 0.022711 | 0.005521 | 0.016752 | 0.023158 | -0.006670 | -0.010255 | efectores |
| 997 | 0.001713 | 0.084811 | -0.055581 | -0.018475 | 0.005956 | 0.034462 | efectores |
| 999 | -0.006312 | 0.026010 | -0.002245 | 0.030193 | 0.017633 | 0.040308 | efectores |

[781 rows x 63 columns]

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

| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 781.000000 | 781.000000 | 781.000000 | 781.000000 | 781.000000 | 781.000000 |
| mean | 0.043442 | 0.002546 | 0.029791 | 0.035221 | 0.018327 | 0.035229 |
| std | 0.022210 | 0.003923 | 0.021277 | 0.027700 | 0.010098 | 0.017054 |
| min | 0.002254 | 0.000000 | 0.001524 | 0.000000 | 0.000000 | 0.001587 |
| 25% | 0.027662 | 0.000000 | 0.011379 | 0.009697 | 0.010475 | 0.022851 |
| 50% | 0.039697 | 0.000695 | 0.023067 | 0.025705 | 0.016320 | 0.031376 |
| 75% | 0.054735 | 0.003416 | 0.047090 | 0.058846 | 0.024715 | 0.043787 |
| max | 0.129276 | 0.023363 | 0.097073 | 0.116962 | 0.059699 | 0.101573 |

| | X6 | X7 | X8 | X9 ... | X52 \ |
|-------|------------|------------|------------|----------------|------------|
| count | 781.000000 | 781.000000 | 781.000000 | 781.000000 ... | 781.000000 |
| mean | 0.008674 | 0.026936 | 0.020511 | 0.051597 ... | 0.011991 |
| std | 0.007105 | 0.021051 | 0.025733 | 0.019760 ... | 0.023265 |
| min | 0.000000 | 0.001524 | 0.000000 | 0.002977 ... | -0.100867 |
| 25% | 0.002463 | 0.011267 | 0.002428 | 0.036416 ... | 0.001189 |
| 50% | 0.007800 | 0.019590 | 0.008229 | 0.047692 ... | 0.017377 |
| 75% | 0.012559 | 0.038091 | 0.030729 | 0.064557 ... | 0.026941 |
| max | 0.037926 | 0.109260 | 0.123070 | 0.131825 ... | 0.078813 |

| | X53 | X54 | X55 | X56 | X57 | X58 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 781.000000 | 781.000000 | 781.000000 | 781.000000 | 781.000000 | 781.000000 |
| mean | 0.011343 | 0.014708 | 0.014987 | 0.009132 | 0.009817 | 0.013445 |
| std | 0.019013 | 0.021971 | 0.022436 | 0.025081 | 0.020988 | 0.021913 |
| min | -0.049208 | -0.069150 | -0.070739 | -0.083497 | -0.060849 | -0.069855 |
| 25% | 0.000075 | 0.002149 | 0.001622 | -0.006044 | -0.001295 | 0.002363 |
| 50% | 0.007752 | 0.017585 | 0.009934 | 0.015192 | 0.005952 | 0.017926 |
| 75% | 0.019975 | 0.027121 | 0.026353 | 0.025672 | 0.018436 | 0.027619 |
| max | 0.087991 | 0.094705 | 0.103005 | 0.082860 | 0.097883 | 0.082295 |

| | X59 | X60 | X61 |
|-------|------------|------------|------------|
| count | 781.000000 | 781.000000 | 781.000000 |
| mean | 0.011952 | 0.012120 | 0.011580 |
| std | 0.020783 | 0.025933 | 0.020633 |
| min | -0.062770 | -0.086255 | -0.081479 |
| 25% | 0.000455 | -0.000593 | -0.000036 |
| 50% | 0.008805 | 0.015863 | 0.007513 |
| 75% | 0.021412 | 0.027625 | 0.021867 |
| max | 0.088824 | 0.119143 | 0.103359 |

[8 rows x 62 columns]

no_efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|---------------|-----------|-----------|--------------|
| 0 | 0.018914 | 0.001830 | 0.009762 | 0.012813 | 0.018304 | 0.023185 | 0.000610 |
| 1 | 0.033796 | 0.004731 | 0.010139 | 0.012166 | 0.026361 | 0.034471 | 0.004731 |
| 4 | 0.075013 | 0.030005 | 0.052509 | 0.015003 | 0.000000 | 0.030005 | 0.007501 |
| 5 | 0.064016 | 0.000000 | 0.061233 | 0.066799 | 0.016700 | 0.040358 | 0.019483 |
| 6 | 0.078540 | 0.005355 | 0.057120 | 0.055335 | 0.028560 | 0.101745 | 0.028560 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.025153 | 0.000967 | 0.001935 | 0.008707 | 0.018381 | 0.025153 | 0.002902 |
| 996 | 0.042891 | 0.000000 | 0.009191 | 0.085781 | 0.021445 | 0.024509 | 0.006127 |
| 997 | 0.018549 | 0.000000 | 0.000000 | 0.006183 | 0.012366 | 0.018549 | 0.000000 |
| 998 | 0.021533 | 0.004307 | 0.043066 | 0.060293 | 0.008613 | 0.038760 | 0.008613 |
| 999 | 0.050623 | 0.002025 | 0.042523 | 0.056698 | 0.022274 | 0.036448 | 0.020249 |
| | X7 | X8 | X9 ... | X53 | X54 | X55 \ | |
| 0 | 0.036608 | 0.006101 | 0.047591 | ... 0.000408 | 0.017985 | -0.004583 | |
| 1 | 0.038527 | 0.015546 | 0.022981 | ... -0.004850 | 0.006537 | 0.001504 | |
| 4 | 0.000000 | 0.007501 | 0.037506 | ... 0.039277 | 0.023833 | 0.041531 | |
| 5 | 0.023658 | 0.004175 | 0.057058 | ... 0.005496 | -0.020212 | 0.030153 | |
| 6 | 0.026775 | 0.014280 | 0.039270 | ... 0.001219 | -0.005703 | -0.010563 | |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.021283 | 0.000967 | 0.037730 | ... 0.010759 | 0.027087 | 0.011318 | |
| 996 | 0.064336 | 0.076591 | 0.079654 | ... 0.029700 | 0.034731 | 0.045942 | |
| 997 | 0.006183 | 0.006183 | 0.098929 | ... -0.014401 | 0.020554 | 0.008685 | |
| 998 | 0.021533 | 0.008613 | 0.051680 | ... 0.051442 | 0.022868 | 0.071123 | |
| 999 | 0.016199 | 0.012149 | 0.054673 | ... 0.025313 | -0.021729 | 0.000695 | |
| | X56 | X57 | X58 | X59 | X60 | X61 | X62 |
| 0 | 0.031985 | 0.007837 | 0.037285 | 0.012985 | 0.021042 | 0.001462 | no_efectores |
| 1 | 0.013538 | 0.004285 | 0.019999 | 0.006917 | 0.016259 | -0.001023 | no_efectores |
| 4 | -0.021926 | -0.024407 | -0.022443 | -0.025921 | -0.018124 | 0.015420 | no_efectores |
| 5 | -0.015478 | 0.003230 | 0.030009 | 0.040430 | -0.030672 | -0.000608 | no_efectores |
| 6 | -0.011992 | 0.016417 | 0.019264 | -0.000082 | 0.010227 | -0.013303 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.017136 | 0.001955 | 0.020126 | -0.001387 | 0.016038 | 0.001604 | no_efectores |
| 996 | -0.048813 | -0.003308 | -0.094014 | -0.048081 | 0.027020 | 0.051339 | no_efectores |
| 997 | -0.024568 | -0.031979 | -0.014909 | -0.021821 | -0.000365 | 0.007231 | no_efectores |
| 998 | -0.012141 | 0.024650 | -0.011462 | 0.017585 | -0.042510 | 0.015815 | no_efectores |
| 999 | -0.019738 | 0.012556 | 0.005078 | 0.013906 | 0.033382 | 0.006637 | no_efectores |

[839 rows x 63 columns]

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

Estadísticas.

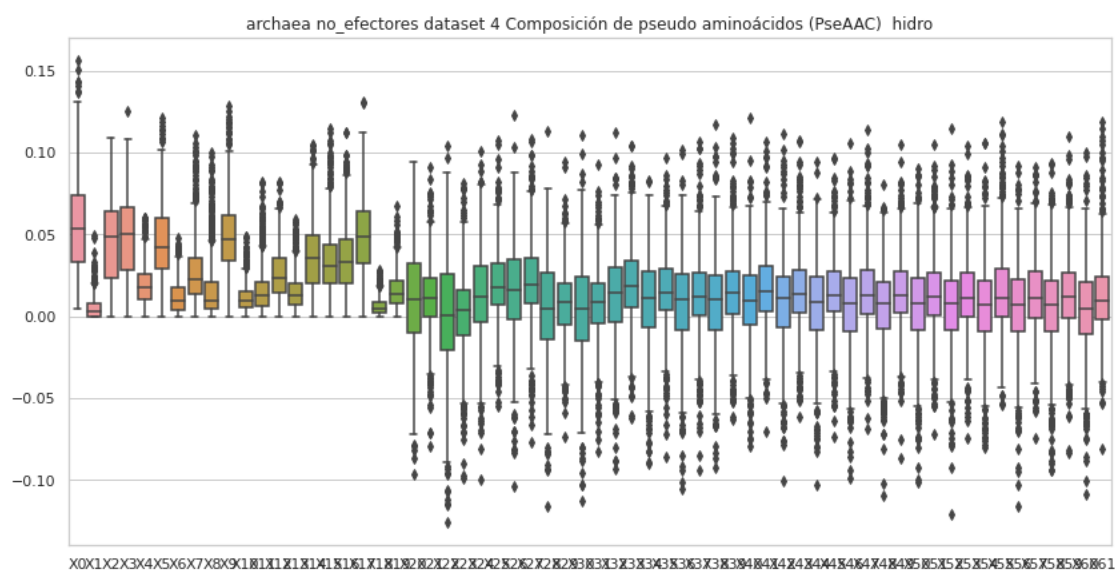
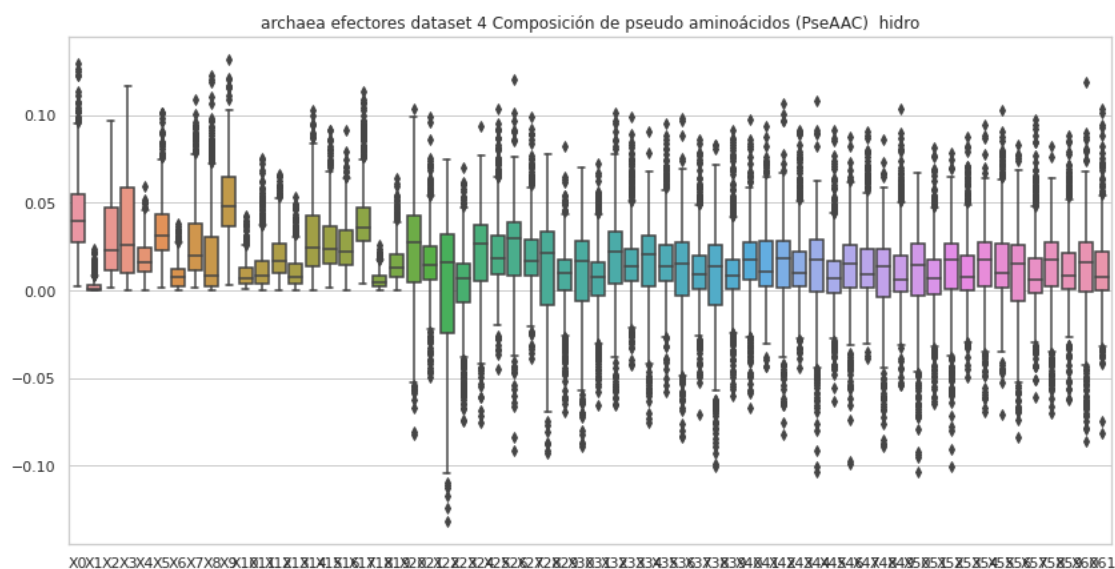
| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 839.000000 | 839.000000 | 839.000000 | 839.000000 | 839.000000 | 839.000000 |
| mean | 0.055889 | 0.005192 | 0.045228 | 0.047258 | 0.018870 | 0.045291 |
| std | 0.028462 | 0.006521 | 0.024749 | 0.025509 | 0.011213 | 0.022210 |
| min | 0.004643 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 0.033420 | 0.000000 | 0.023398 | 0.028335 | 0.010619 | 0.029311 |
| 50% | 0.053056 | 0.003109 | 0.048604 | 0.050063 | 0.017697 | 0.042259 |
| 75% | 0.073664 | 0.007978 | 0.064236 | 0.066239 | 0.025621 | 0.059754 |
| max | 0.156085 | 0.049727 | 0.108647 | 0.125041 | 0.060351 | 0.120900 |

| | X6 | X7 | X8 | X9 ... | X52 \ |
|-------|------------|------------|------------|------------|------------|
| count | 839.000000 | 839.000000 | 839.000000 | 839.000000 | 839.000000 |
| mean | 0.011435 | 0.026992 | 0.016380 | 0.049403 | 0.006228 |
| std | 0.009376 | 0.019814 | 0.018148 | 0.021840 | 0.024722 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | -0.121086 |
| 25% | 0.003925 | 0.013291 | 0.004276 | 0.034193 | -0.008347 |
| 50% | 0.009371 | 0.022281 | 0.009739 | 0.047110 | 0.007572 |
| 75% | 0.017211 | 0.035716 | 0.020920 | 0.061834 | 0.021877 |
| max | 0.047360 | 0.110211 | 0.099966 | 0.128404 | 0.114826 |

| | X53 | X54 | X55 | X56 | X57 | X58 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 839.000000 | 839.000000 | 839.000000 | 839.000000 | 839.000000 | 839.000000 |
| mean | 0.013484 | 0.006176 | 0.014513 | 0.005793 | 0.012940 | 0.004856 |
| std | 0.022638 | 0.026289 | 0.024853 | 0.026161 | 0.022483 | 0.026472 |
| min | -0.074344 | -0.080321 | -0.053997 | -0.116110 | -0.074525 | -0.094434 |
| 25% | -0.000120 | -0.009092 | -0.000427 | -0.008992 | -0.001059 | -0.009102 |
| 50% | 0.011269 | 0.007115 | 0.011038 | 0.006644 | 0.011249 | 0.006816 |
| 75% | 0.026684 | 0.021812 | 0.028750 | 0.022824 | 0.027302 | 0.021909 |
| max | 0.104332 | 0.105783 | 0.118366 | 0.092126 | 0.091073 | 0.093280 |

| | X59 | X60 | X61 |
|-------|------------|------------|------------|
| count | 839.000000 | 839.000000 | 839.000000 |
| mean | 0.013074 | 0.003458 | 0.012387 |
| std | 0.023752 | 0.026806 | 0.023538 |
| min | -0.083599 | -0.109041 | -0.081344 |
| 25% | -0.001023 | -0.010729 | -0.001650 |
| 50% | 0.011564 | 0.004894 | 0.009622 |
| 75% | 0.026419 | 0.020476 | 0.024406 |
| max | 0.109470 | 0.099521 | 0.118608 |

[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 4,
con valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | -0.001255 | 0.024772 | 0.105252 | 0.027670 | 0.004654 | 0.100112 | 0.071641 |
| 1 | 0.056264 | -0.037222 | -0.012612 | -0.002320 | -0.075573 | -0.002453 | -0.112878 |
| 2 | 0.048471 | 0.009440 | 0.039151 | 0.036947 | 0.022486 | 0.029969 | 0.006779 |
| 3 | 0.037604 | 0.038714 | -0.030308 | 0.053671 | -0.083895 | -0.061327 | 0.043566 |
| 4 | 0.038873 | 0.080391 | 0.042054 | 0.061625 | 0.008376 | 0.051868 | 0.068299 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | -0.107209 | -0.047024 | -0.125564 | 0.067282 | -0.027618 | 0.065714 | -0.113395 |
| 996 | 0.040378 | -0.042663 | -0.081736 | -0.017018 | -0.111148 | -0.081607 | -0.044643 |
| 997 | -0.019178 | -0.059432 | -0.016089 | -0.028522 | 0.040618 | -0.041040 | 0.030537 |
| 998 | -0.013216 | 0.084164 | -0.010332 | 0.081858 | 0.013480 | 0.042177 | -0.051119 |
| 999 | -0.114142 | -0.033796 | 0.061763 | 0.004816 | 0.012654 | -0.063315 | 0.050928 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|---|-----------|-----------|-----------|-----------|----------|----------|-----------|
| 0 | -0.039499 | -0.052422 | -0.006391 | -0.008401 | 0.011709 | 0.054551 | efectores |

| | | | | | | | |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.053807 | -0.062937 | -0.007930 | 0.019662 | -0.057114 | 0.071637 | efectores |
| 2 | 0.028147 | -0.007568 | -0.050268 | 0.046678 | -0.044179 | -0.015935 | efectores |
| 3 | -0.083518 | -0.032929 | 0.026655 | -0.021762 | 0.001202 | 0.071403 | efectores |
| 4 | 0.038455 | 0.020836 | 0.119455 | 0.040215 | 0.061706 | 0.088167 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.066609 | 0.077565 | 0.149456 | 0.016210 | -0.058814 | -0.123535 | efectores |
| 996 | 0.070428 | 0.009412 | 0.035696 | 0.007972 | -0.099162 | 0.000970 | efectores |
| 997 | -0.075273 | 0.039889 | 0.053082 | -0.022417 | 0.057374 | -0.050513 | efectores |
| 998 | -0.064832 | 0.065257 | -0.034440 | 0.039394 | -0.047467 | -0.045684 | efectores |
| 999 | 0.022425 | 0.008115 | 0.104509 | 0.007036 | -0.003060 | -0.016903 | efectores |

[1000 rows x 14 columns]

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

Estadísticas.

| | X0 | X1 | X2 | X3 | X4 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.007459 | 0.018213 | 0.011325 | 0.013317 | -0.005331 |
| std | 0.070388 | 0.070291 | 0.066139 | 0.065800 | 0.068501 |
| min | -0.364341 | -0.280487 | -0.258376 | -0.307182 | -0.247753 |
| 25% | -0.033385 | -0.025153 | -0.030579 | -0.028337 | -0.048402 |
| 50% | 0.007369 | 0.017458 | 0.008480 | 0.013613 | -0.002471 |
| 75% | 0.049798 | 0.061457 | 0.055725 | 0.060501 | 0.039118 |
| max | 0.226232 | 0.290879 | 0.242345 | 0.211502 | 0.209231 |

| | X5 | X6 | X7 | X8 | X9 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.000312 | 0.016380 | 0.002073 | -0.004723 | 0.007643 |
| std | 0.068391 | 0.072658 | 0.065792 | 0.069080 | 0.072470 |
| min | -0.203370 | -0.268971 | -0.269360 | -0.243467 | -0.240817 |
| 25% | -0.043956 | -0.027887 | -0.034042 | -0.050060 | -0.034314 |
| 50% | 0.004635 | 0.014954 | 0.002022 | -0.004025 | 0.008281 |
| 75% | 0.041428 | 0.058657 | 0.042244 | 0.040141 | 0.052425 |
| max | 0.334080 | 0.360103 | 0.236864 | 0.204652 | 0.245072 |

| | X10 | X11 | X12 |
|-------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.002253 | -0.005783 | 0.012437 |
| std | 0.069144 | 0.066264 | 0.071493 |
| min | -0.288877 | -0.224103 | -0.223231 |
| 25% | -0.042254 | -0.045211 | -0.036296 |
| 50% | 0.001096 | -0.004689 | 0.011559 |
| 75% | 0.045660 | 0.035199 | 0.061663 |
| max | 0.339901 | 0.229242 | 0.275202 |

no_efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.022156 | 0.011357 | 0.014821 | 0.026827 | 0.034056 | -0.001082 | 0.051075 |
| 1 | 0.036065 | 0.025745 | 0.061728 | 0.043862 | -0.021754 | 0.075709 | -0.017334 |
| 2 | 0.124438 | -0.113069 | 0.021726 | 0.081986 | 0.242298 | 0.071238 | -0.083795 |
| 3 | 0.000457 | -0.069847 | 0.093453 | 0.058773 | -0.108883 | 0.082633 | 0.131747 |
| 4 | 0.289418 | -0.119789 | -0.001436 | -0.005802 | 0.244504 | 0.024128 | -0.105507 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.059785 | -0.027676 | -0.025231 | 0.126949 | 0.062763 | -0.134173 | -0.008077 |
| 996 | 0.041958 | -0.078818 | 0.022183 | 0.028601 | -0.008870 | -0.009356 | -0.052238 |
| 997 | -0.072576 | -0.066887 | -0.165832 | 0.122460 | 0.019564 | 0.040797 | -0.035878 |
| 998 | -0.140859 | 0.012119 | -0.011704 | 0.019137 | -0.123168 | -0.019680 | -0.150390 |
| 999 | -0.068806 | -0.014602 | -0.037819 | 0.021121 | 0.106009 | -0.090962 | 0.077878 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | 0.008119 | 0.024437 | -0.014068 | 0.000133 | 0.017149 | 0.040719 | no_efectores |
| 1 | -0.017676 | 0.054559 | 0.027064 | -0.033452 | 0.031972 | 0.079482 | no_efectores |
| 2 | -0.091966 | 0.083660 | 0.179422 | 0.189599 | 0.063265 | -0.028467 | no_efectores |
| 3 | 0.015062 | 0.051231 | 0.085413 | -0.023376 | -0.024658 | 0.016740 | no_efectores |
| 4 | -0.232342 | -0.282243 | -0.075325 | 0.148232 | 0.151757 | -0.277900 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.070846 | 0.146722 | -0.022102 | 0.029962 | -0.016114 | 0.090253 | no_efectores |
| 996 | -0.091426 | 0.012115 | -0.046669 | -0.018373 | 0.055251 | -0.000757 | no_efectores |
| 997 | 0.055530 | -0.059779 | -0.064707 | 0.181051 | 0.041524 | 0.038964 | no_efectores |
| 998 | 0.205794 | -0.214536 | -0.032255 | 0.071698 | 0.108499 | -0.085263 | no_efectores |
| 999 | -0.031377 | 0.059889 | 0.011886 | -0.100060 | 0.090595 | -0.019082 | no_efectores |

[1000 rows x 14 columns]

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

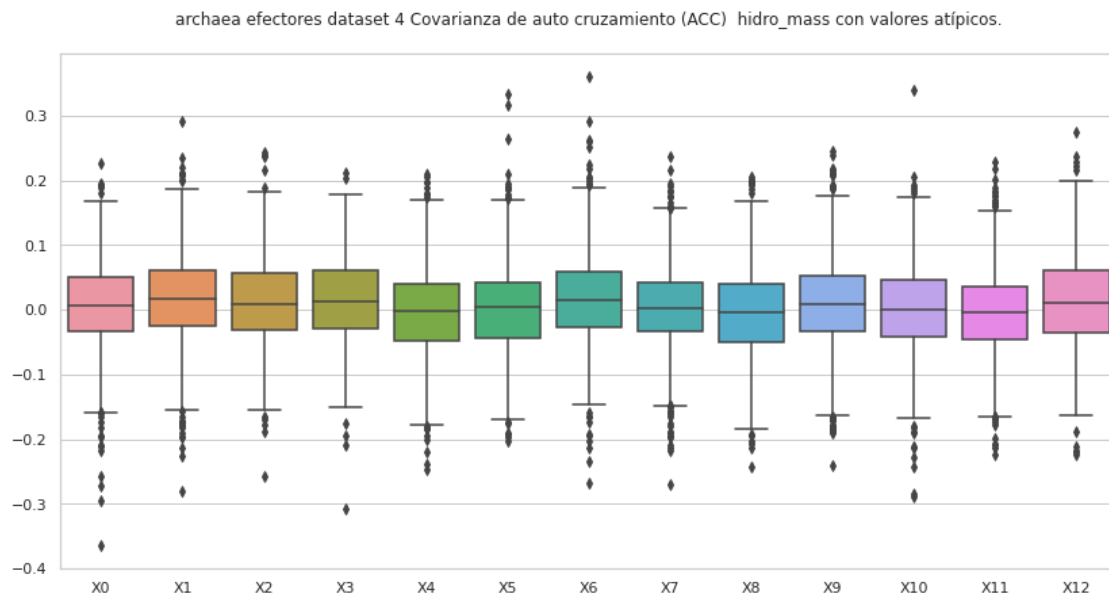
Estadísticas.

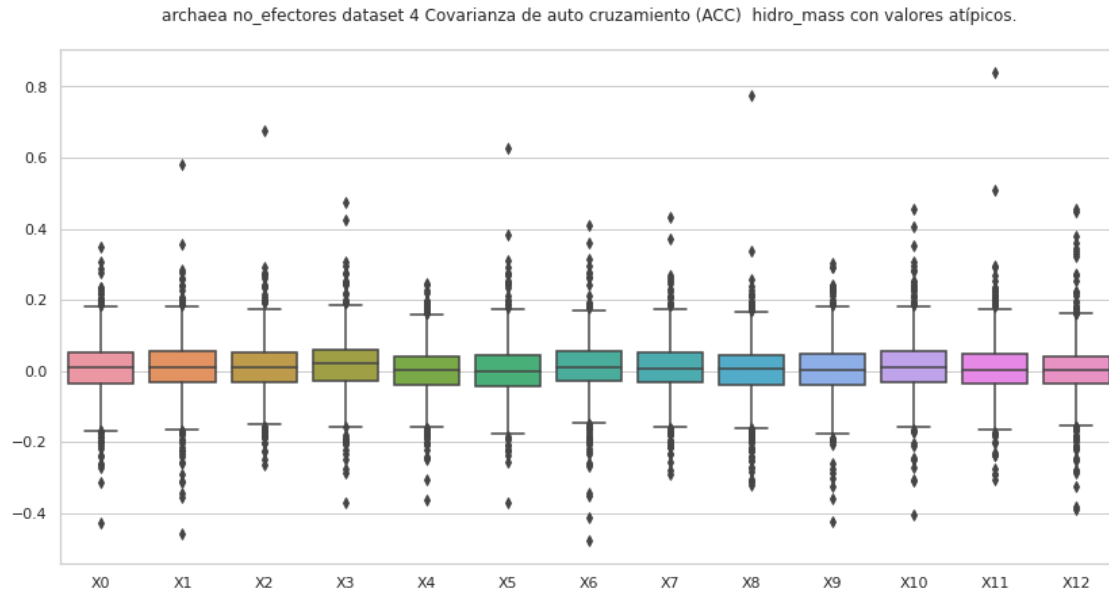
| | X0 | X1 | X2 | X3 | X4 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.007947 | 0.008814 | 0.010001 | 0.017847 | -0.001284 |
| std | 0.078384 | 0.081980 | 0.075870 | 0.078550 | 0.072355 |
| min | -0.426459 | -0.457194 | -0.262757 | -0.371712 | -0.362758 |
| 25% | -0.036988 | -0.033391 | -0.031193 | -0.027152 | -0.040077 |
| 50% | 0.008795 | 0.010182 | 0.009145 | 0.020429 | 0.001722 |
| 75% | 0.051611 | 0.053572 | 0.051003 | 0.059736 | 0.039849 |

| | | | | | |
|-----|----------|----------|----------|----------|----------|
| max | 0.347263 | 0.582129 | 0.676092 | 0.473344 | 0.244504 |
|-----|----------|----------|----------|----------|----------|

| | | | | | |
|-------|-------------|-------------|-------------|-------------|-------------|
| | X5 | X6 | X7 | X8 | X9 \ |
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.000628 | 0.009352 | 0.007033 | 0.001441 | 0.003264 |
| std | 0.078695 | 0.078834 | 0.075733 | 0.081663 | 0.077755 |
| min | -0.371017 | -0.477713 | -0.292932 | -0.322281 | -0.422724 |
| 25% | -0.044784 | -0.026788 | -0.032966 | -0.039433 | -0.041664 |
| 50% | -0.000165 | 0.010245 | 0.006465 | 0.004269 | 0.002672 |
| 75% | 0.043213 | 0.054294 | 0.051003 | 0.043135 | 0.047877 |
| max | 0.626292 | 0.408936 | 0.432952 | 0.774821 | 0.302586 |

| | | | |
|-------|-------------|-------------|-------------|
| | X10 | X11 | X12 |
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.012213 | 0.006756 | 0.003246 |
| std | 0.080101 | 0.080509 | 0.082961 |
| min | -0.404101 | -0.307733 | -0.389975 |
| 25% | -0.032328 | -0.037155 | -0.037302 |
| 50% | 0.008168 | 0.003232 | 0.001337 |
| 75% | 0.054638 | 0.048998 | 0.041797 |
| max | 0.455314 | 0.839554 | 0.453391 |





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 4,
sin valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | -0.001255 | 0.024772 | 0.105252 | 0.027670 | 0.004654 | 0.100112 | 0.071641 |
| 1 | 0.056264 | -0.037222 | -0.012612 | -0.002320 | -0.075573 | -0.002453 | -0.112878 |
| 2 | 0.048471 | 0.009440 | 0.039151 | 0.036947 | 0.022486 | 0.029969 | 0.006779 |
| 3 | 0.037604 | 0.038714 | -0.030308 | 0.053671 | -0.083895 | -0.061327 | 0.043566 |
| 4 | 0.038873 | 0.080391 | 0.042054 | 0.061625 | 0.008376 | 0.051868 | 0.068299 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | -0.107209 | -0.047024 | -0.125564 | 0.067282 | -0.027618 | 0.065714 | -0.113395 |
| 996 | 0.040378 | -0.042663 | -0.081736 | -0.017018 | -0.111148 | -0.081607 | -0.044643 |
| 997 | -0.019178 | -0.059432 | -0.016089 | -0.028522 | 0.040618 | -0.041040 | 0.030537 |
| 998 | -0.013216 | 0.084164 | -0.010332 | 0.081858 | 0.013480 | 0.042177 | -0.051119 |
| 999 | -0.114142 | -0.033796 | 0.061763 | 0.004816 | 0.012654 | -0.063315 | 0.050928 |
| | | | | | | | |
| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
| 0 | -0.039499 | -0.052422 | -0.006391 | -0.008401 | 0.011709 | 0.054551 | efectores |
| 1 | 0.053807 | -0.062937 | -0.007930 | 0.019662 | -0.057114 | 0.071637 | efectores |
| 2 | 0.028147 | -0.007568 | -0.050268 | 0.046678 | -0.044179 | -0.015935 | efectores |
| 3 | -0.083518 | -0.032929 | 0.026655 | -0.021762 | 0.001202 | 0.071403 | efectores |
| 4 | 0.038455 | 0.020836 | 0.119455 | 0.040215 | 0.061706 | 0.088167 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.066609 | 0.077565 | 0.149456 | 0.016210 | -0.058814 | -0.123535 | efectores |

```

996  0.070428  0.009412  0.035696  0.007972 -0.099162  0.000970  efectores
997 -0.075273  0.039889  0.053082 -0.022417  0.057374 -0.050513  efectores
998 -0.064832  0.065257 -0.034440  0.039394 -0.047467 -0.045684  efectores
999  0.022425  0.008115  0.104509  0.007036 -0.003060 -0.016903  efectores

```

[937 rows x 14 columns]

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

| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 937.000000 | 937.000000 | 937.000000 | 937.000000 | 937.000000 | 937.000000 |
| mean | 0.008694 | 0.018280 | 0.011912 | 0.013358 | -0.005806 | -0.000978 |
| std | 0.064333 | 0.065341 | 0.062860 | 0.063290 | 0.065211 | 0.063348 |
| min | -0.197530 | -0.191380 | -0.177268 | -0.175447 | -0.200777 | -0.203370 |
| 25% | -0.030864 | -0.022640 | -0.029052 | -0.027275 | -0.047469 | -0.042791 |
| 50% | 0.007366 | 0.017412 | 0.008497 | 0.012959 | -0.003088 | 0.003987 |
| 75% | 0.048661 | 0.060426 | 0.054827 | 0.059035 | 0.037445 | 0.038489 |
| max | 0.194829 | 0.219450 | 0.189017 | 0.204206 | 0.196978 | 0.191594 |

| | X6 | X7 | X8 | X9 | X10 | X11 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 937.000000 | 937.000000 | 937.000000 | 937.000000 | 937.000000 | 937.000000 |
| mean | 0.016958 | 0.002473 | -0.004763 | 0.008617 | 0.003272 | -0.005148 |
| std | 0.068143 | 0.062191 | 0.066212 | 0.068956 | 0.064684 | 0.061609 |
| min | -0.194633 | -0.193764 | -0.206914 | -0.191389 | -0.190070 | -0.178000 |
| 25% | -0.027334 | -0.032547 | -0.049441 | -0.032916 | -0.040380 | -0.043190 |
| 50% | 0.015647 | 0.002268 | -0.003979 | 0.008475 | 0.001282 | -0.004675 |
| 75% | 0.058703 | 0.041628 | 0.037796 | 0.052034 | 0.044802 | 0.034348 |
| max | 0.225037 | 0.193795 | 0.200822 | 0.217885 | 0.191931 | 0.188673 |

| | X12 |
|-------|------------|
| count | 937.000000 |
| mean | 0.013507 |
| std | 0.067285 |
| min | -0.187702 |
| 25% | -0.034334 |
| 50% | 0.012123 |
| 75% | 0.061529 |
| max | 0.222029 |

no_efectores

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

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.022156 | 0.011357 | 0.014821 | 0.026827 | 0.034056 | -0.001082 | 0.051075 |
| 1 | 0.036065 | 0.025745 | 0.061728 | 0.043862 | -0.021754 | 0.075709 | -0.017334 |
| 3 | 0.000457 | -0.069847 | 0.093453 | 0.058773 | -0.108883 | 0.082633 | 0.131747 |
| 5 | 0.058842 | 0.105379 | -0.021232 | 0.006275 | -0.073154 | -0.019955 | -0.012804 |
| 6 | 0.040239 | 0.030984 | 0.071625 | -0.020042 | -0.071026 | -0.001252 | -0.061482 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.059785 | -0.027676 | -0.025231 | 0.126949 | 0.062763 | -0.134173 | -0.008077 |
| 996 | 0.041958 | -0.078818 | 0.022183 | 0.028601 | -0.008870 | -0.009356 | -0.052238 |
| 997 | -0.072576 | -0.066887 | -0.165832 | 0.122460 | 0.019564 | 0.040797 | -0.035878 |
| 998 | -0.140859 | 0.012119 | -0.011704 | 0.019137 | -0.123168 | -0.019680 | -0.150390 |
| 999 | -0.068806 | -0.014602 | -0.037819 | 0.021121 | 0.106009 | -0.090962 | 0.077878 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | 0.008119 | 0.024437 | -0.014068 | 0.000133 | 0.017149 | 0.040719 | no_efectores |
| 1 | -0.017676 | 0.054559 | 0.027064 | -0.033452 | 0.031972 | 0.079482 | no_efectores |
| 3 | 0.015062 | 0.051231 | 0.085413 | -0.023376 | -0.024658 | 0.016740 | no_efectores |
| 5 | -0.073054 | -0.034731 | 0.033310 | 0.068689 | 0.061954 | 0.012306 | no_efectores |
| 6 | -0.014697 | 0.054025 | 0.051917 | 0.001560 | 0.018100 | -0.018365 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.070846 | 0.146722 | -0.022102 | 0.029962 | -0.016114 | 0.090253 | no_efectores |
| 996 | -0.091426 | 0.012115 | -0.046669 | -0.018373 | 0.055251 | -0.000757 | no_efectores |
| 997 | 0.055530 | -0.059779 | -0.064707 | 0.181051 | 0.041524 | 0.038964 | no_efectores |
| 998 | 0.205794 | -0.214536 | -0.032255 | 0.071698 | 0.108499 | -0.085263 | no_efectores |
| 999 | -0.031377 | 0.059889 | 0.011886 | -0.100060 | 0.090595 | -0.019082 | no_efectores |

[902 rows x 14 columns]

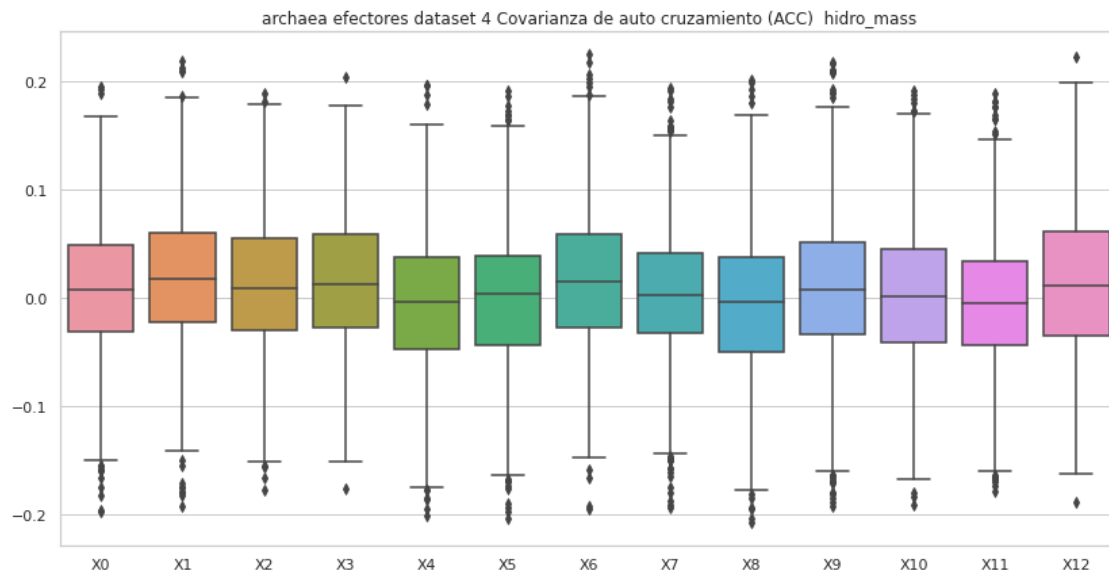
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 4,
sin valores atípicos.
Estadísticas.

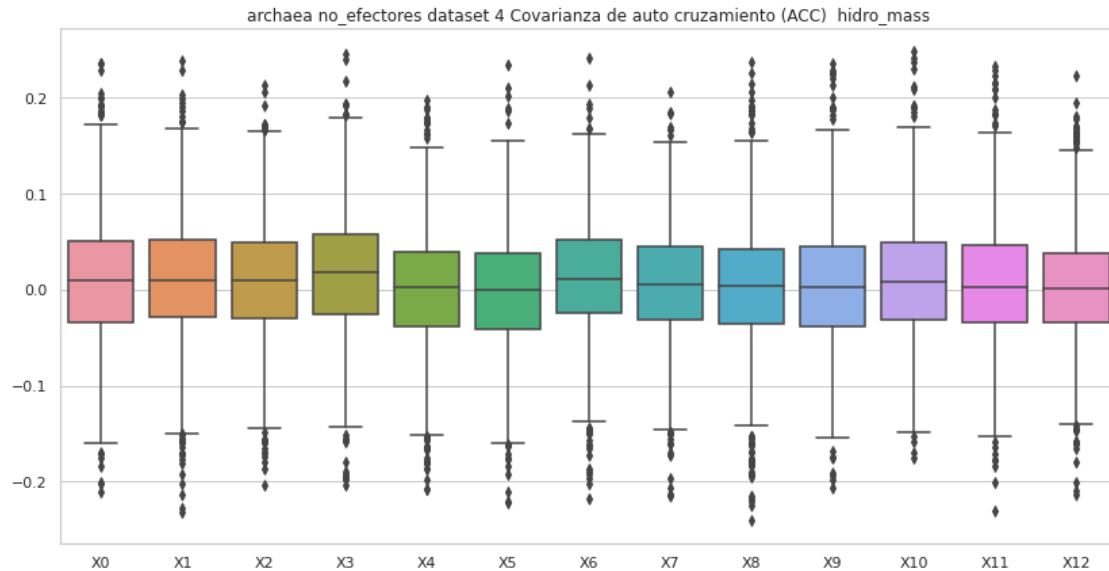
| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 902.000000 | 902.000000 | 902.000000 | 902.000000 | 902.000000 | 902.000000 |
| mean | 0.009660 | 0.009745 | 0.008796 | 0.016481 | 0.000556 | -0.001303 |
| std | 0.068721 | 0.067722 | 0.063897 | 0.067616 | 0.063430 | 0.065625 |
| min | -0.211466 | -0.232113 | -0.204443 | -0.204069 | -0.208351 | -0.222152 |
| 25% | -0.033736 | -0.028875 | -0.029439 | -0.025303 | -0.037768 | -0.041275 |
| 50% | 0.009883 | 0.010296 | 0.009059 | 0.018694 | 0.002978 | -0.000492 |
| 75% | 0.050993 | 0.051842 | 0.048648 | 0.057450 | 0.038943 | 0.038070 |
| max | 0.235989 | 0.238809 | 0.213582 | 0.245622 | 0.197379 | 0.233822 |

| | X6 | X7 | X8 | X9 | X10 | X11 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 902.000000 | 902.000000 | 902.000000 | 902.000000 | 902.000000 | 902.000000 |
| mean | 0.010700 | 0.005814 | 0.002740 | 0.003582 | 0.010701 | 0.005565 |
| std | 0.063956 | 0.063199 | 0.069352 | 0.067146 | 0.065461 | 0.066599 |

| | | | | | | |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|
| min | -0.217428 | -0.215093 | -0.240527 | -0.206498 | -0.174847 | -0.231053 |
| 25% | -0.024104 | -0.031427 | -0.035061 | -0.038231 | -0.031269 | -0.034061 |
| 50% | 0.010424 | 0.006025 | 0.004269 | 0.002672 | 0.008065 | 0.003291 |
| 75% | 0.051687 | 0.045457 | 0.042163 | 0.045350 | 0.049562 | 0.046826 |
| max | 0.241156 | 0.205794 | 0.236932 | 0.235350 | 0.248512 | 0.233203 |

| | X12 |
|-------|------------|
| count | 902.000000 |
| mean | 0.002532 |
| std | 0.063221 |
| min | -0.213234 |
| 25% | -0.033550 |
| 50% | 0.000881 |
| 75% | 0.038248 |
| max | 0.222855 |





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 4, con valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | -0.001255 | 0.024772 | 0.105252 | 0.027670 | 0.004654 | 0.100112 | 0.071641 |
| 1 | 0.056264 | -0.037222 | -0.012612 | -0.002320 | -0.075573 | -0.002453 | -0.112878 |
| 2 | 0.048471 | 0.009440 | 0.039151 | 0.036947 | 0.022486 | 0.029969 | 0.006779 |
| 3 | 0.037604 | 0.038714 | -0.030308 | 0.053671 | -0.083895 | -0.061327 | 0.043566 |
| 4 | 0.038873 | 0.080391 | 0.042054 | 0.061625 | 0.008376 | 0.051868 | 0.068299 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | -0.107209 | -0.047024 | -0.125564 | 0.067282 | -0.027618 | 0.065714 | -0.113395 |
| 996 | 0.040378 | -0.042663 | -0.081736 | -0.017018 | -0.111148 | -0.081607 | -0.044643 |
| 997 | -0.019178 | -0.059432 | -0.016089 | -0.028522 | 0.040618 | -0.041040 | 0.030537 |
| 998 | -0.013216 | 0.084164 | -0.010332 | 0.081858 | 0.013480 | 0.042177 | -0.051119 |
| 999 | -0.114142 | -0.033796 | 0.061763 | 0.004816 | 0.012654 | -0.063315 | 0.050928 |
| | | | | | | | |
| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
| 0 | -0.039499 | -0.052422 | -0.006391 | -0.008401 | 0.011709 | 0.054551 | efectores |
| 1 | 0.053807 | -0.062937 | -0.007930 | 0.019662 | -0.057114 | 0.071637 | efectores |
| 2 | 0.028147 | -0.007568 | -0.050268 | 0.046678 | -0.044179 | -0.015935 | efectores |
| 3 | -0.083518 | -0.032929 | 0.026655 | -0.021762 | 0.001202 | 0.071403 | efectores |
| 4 | 0.038455 | 0.020836 | 0.119455 | 0.040215 | 0.061706 | 0.088167 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.066609 | 0.077565 | 0.149456 | 0.016210 | -0.058814 | -0.123535 | efectores |
| 996 | 0.070428 | 0.009412 | 0.035696 | 0.007972 | -0.099162 | 0.000970 | efectores |
| 997 | -0.075273 | 0.039889 | 0.053082 | -0.022417 | 0.057374 | -0.050513 | efectores |
| 998 | -0.064832 | 0.065257 | -0.034440 | 0.039394 | -0.047467 | -0.045684 | efectores |
| 999 | 0.022425 | 0.008115 | 0.104509 | 0.007036 | -0.003060 | -0.016903 | efectores |

[1000 rows x 14 columns]

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

Estadísticas.

| | X0 | X1 | X2 | X3 | X4 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.007459 | 0.018213 | 0.011325 | 0.013317 | -0.005331 |

| | | | | | |
|-----|-----------|-----------|-----------|-----------|-----------|
| std | 0.070388 | 0.070291 | 0.066139 | 0.065800 | 0.068501 |
| min | -0.364341 | -0.280487 | -0.258376 | -0.307182 | -0.247753 |
| 25% | -0.033385 | -0.025153 | -0.030579 | -0.028337 | -0.048402 |
| 50% | 0.007369 | 0.017458 | 0.008480 | 0.013613 | -0.002471 |
| 75% | 0.049798 | 0.061457 | 0.055725 | 0.060501 | 0.039118 |
| max | 0.226232 | 0.290879 | 0.242345 | 0.211502 | 0.209231 |

| | | | | | |
|-------|-------------|-------------|-------------|-------------|-------------|
| | X5 | X6 | X7 | X8 | X9 \ |
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.000312 | 0.016380 | 0.002073 | -0.004723 | 0.007643 |
| std | 0.068391 | 0.072658 | 0.065792 | 0.069080 | 0.072470 |
| min | -0.203370 | -0.268971 | -0.269360 | -0.243467 | -0.240817 |
| 25% | -0.043956 | -0.027887 | -0.034042 | -0.050060 | -0.034314 |
| 50% | 0.004635 | 0.014954 | 0.002022 | -0.004025 | 0.008281 |
| 75% | 0.041428 | 0.058657 | 0.042244 | 0.040141 | 0.052425 |
| max | 0.334080 | 0.360103 | 0.236864 | 0.204652 | 0.245072 |

| | | | |
|-------|-------------|-------------|-------------|
| | X10 | X11 | X12 |
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.002253 | -0.005783 | 0.012437 |
| std | 0.069144 | 0.066264 | 0.071493 |
| min | -0.288877 | -0.224103 | -0.223231 |
| 25% | -0.042254 | -0.045211 | -0.036296 |
| 50% | 0.001096 | -0.004689 | 0.011559 |
| 75% | 0.045660 | 0.035199 | 0.061663 |
| max | 0.339901 | 0.229242 | 0.275202 |

no_efectores

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

Valores del documento csv.

| | | | | | | | |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
| 0 | 0.022156 | 0.011357 | 0.014821 | 0.026827 | 0.034056 | -0.001082 | 0.051075 |
| 1 | 0.036065 | 0.025745 | 0.061728 | 0.043862 | -0.021754 | 0.075709 | -0.017334 |
| 2 | 0.124438 | -0.113069 | 0.021726 | 0.081986 | 0.242298 | 0.071238 | -0.083795 |
| 3 | 0.000457 | -0.069847 | 0.093453 | 0.058773 | -0.108883 | 0.082633 | 0.131747 |
| 4 | 0.289418 | -0.119789 | -0.001436 | -0.005802 | 0.244504 | 0.024128 | -0.105507 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.059785 | -0.027676 | -0.025231 | 0.126949 | 0.062763 | -0.134173 | -0.008077 |
| 996 | 0.041958 | -0.078818 | 0.022183 | 0.028601 | -0.008870 | -0.009356 | -0.052238 |
| 997 | -0.072576 | -0.066887 | -0.165832 | 0.122460 | 0.019564 | 0.040797 | -0.035878 |
| 998 | -0.140859 | 0.012119 | -0.011704 | 0.019137 | -0.123168 | -0.019680 | -0.150390 |
| 999 | -0.068806 | -0.014602 | -0.037819 | 0.021121 | 0.106009 | -0.090962 | 0.077878 |
| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |

| | | | | | | | |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | 0.008119 | 0.024437 | -0.014068 | 0.000133 | 0.017149 | 0.040719 | no_efectores |
| 1 | -0.017676 | 0.054559 | 0.027064 | -0.033452 | 0.031972 | 0.079482 | no_efectores |
| 2 | -0.091966 | 0.083660 | 0.179422 | 0.189599 | 0.063265 | -0.028467 | no_efectores |
| 3 | 0.015062 | 0.051231 | 0.085413 | -0.023376 | -0.024658 | 0.016740 | no_efectores |
| 4 | -0.232342 | -0.282243 | -0.075325 | 0.148232 | 0.151757 | -0.277900 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.070846 | 0.146722 | -0.022102 | 0.029962 | -0.016114 | 0.090253 | no_efectores |
| 996 | -0.091426 | 0.012115 | -0.046669 | -0.018373 | 0.055251 | -0.000757 | no_efectores |
| 997 | 0.055530 | -0.059779 | -0.064707 | 0.181051 | 0.041524 | 0.038964 | no_efectores |
| 998 | 0.205794 | -0.214536 | -0.032255 | 0.071698 | 0.108499 | -0.085263 | no_efectores |
| 999 | -0.031377 | 0.059889 | 0.011886 | -0.100060 | 0.090595 | -0.019082 | no_efectores |

[1000 rows x 14 columns]

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

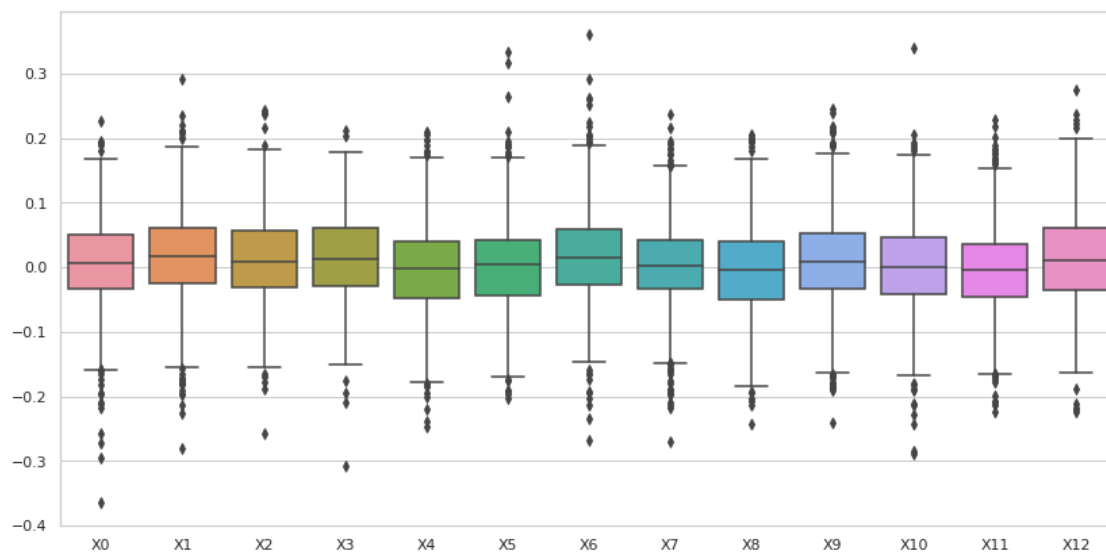
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.007947 | 0.008814 | 0.010001 | 0.017847 | -0.001284 |
| std | 0.078384 | 0.081980 | 0.075870 | 0.078550 | 0.072355 |
| min | -0.426459 | -0.457194 | -0.262757 | -0.371712 | -0.362758 |
| 25% | -0.036988 | -0.033391 | -0.031193 | -0.027152 | -0.040077 |
| 50% | 0.008795 | 0.010182 | 0.009145 | 0.020429 | 0.001722 |
| 75% | 0.051611 | 0.053572 | 0.051003 | 0.059736 | 0.039849 |
| max | 0.347263 | 0.582129 | 0.676092 | 0.473344 | 0.244504 |

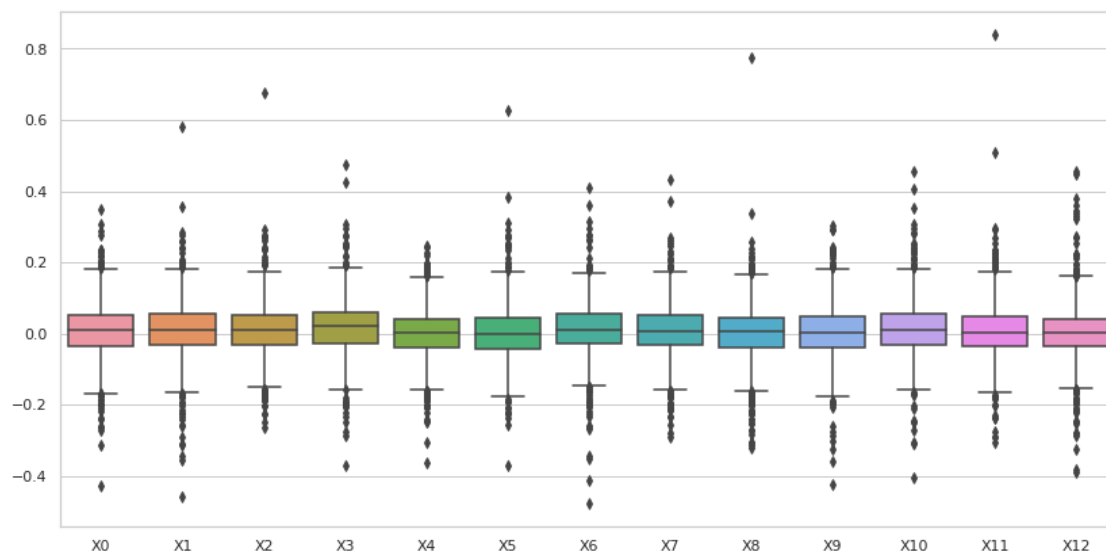
| | X5 | X6 | X7 | X8 | X9 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.000628 | 0.009352 | 0.007033 | 0.001441 | 0.003264 |
| std | 0.078695 | 0.078834 | 0.075733 | 0.081663 | 0.077755 |
| min | -0.371017 | -0.477713 | -0.292932 | -0.322281 | -0.422724 |
| 25% | -0.044784 | -0.026788 | -0.032966 | -0.039433 | -0.041664 |
| 50% | -0.000165 | 0.010245 | 0.006465 | 0.004269 | 0.002672 |
| 75% | 0.043213 | 0.054294 | 0.051003 | 0.043135 | 0.047877 |
| max | 0.626292 | 0.408936 | 0.432952 | 0.774821 | 0.302586 |

| | X10 | X11 | X12 |
|-------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.012213 | 0.006756 | 0.003246 |
| std | 0.080101 | 0.080509 | 0.082961 |
| min | -0.404101 | -0.307733 | -0.389975 |
| 25% | -0.032328 | -0.037155 | -0.037302 |
| 50% | 0.008168 | 0.003232 | 0.001337 |
| 75% | 0.054638 | 0.048998 | 0.041797 |
| max | 0.455314 | 0.839554 | 0.453391 |

archaea efectores dataset 4 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.



archaea_no_efectores dataset 4 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 4, sin valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | -0.001255 | 0.024772 | 0.105252 | 0.027670 | 0.004654 | 0.100112 | 0.071641 |
| 1 | 0.056264 | -0.037222 | -0.012612 | -0.002320 | -0.075573 | -0.002453 | -0.112878 |
| 2 | 0.048471 | 0.009440 | 0.039151 | 0.036947 | 0.022486 | 0.029969 | 0.006779 |
| 3 | 0.037604 | 0.038714 | -0.030308 | 0.053671 | -0.083895 | -0.061327 | 0.043566 |
| 4 | 0.038873 | 0.080391 | 0.042054 | 0.061625 | 0.008376 | 0.051868 | 0.068299 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | -0.107209 | -0.047024 | -0.125564 | 0.067282 | -0.027618 | 0.065714 | -0.113395 |
| 996 | 0.040378 | -0.042663 | -0.081736 | -0.017018 | -0.111148 | -0.081607 | -0.044643 |
| 997 | -0.019178 | -0.059432 | -0.016089 | -0.028522 | 0.040618 | -0.041040 | 0.030537 |
| 998 | -0.013216 | 0.084164 | -0.010332 | 0.081858 | 0.013480 | 0.042177 | -0.051119 |
| 999 | -0.114142 | -0.033796 | 0.061763 | 0.004816 | 0.012654 | -0.063315 | 0.050928 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | -0.039499 | -0.052422 | -0.006391 | -0.008401 | 0.011709 | 0.054551 | efectores |
| 1 | 0.053807 | -0.062937 | -0.007930 | 0.019662 | -0.057114 | 0.071637 | efectores |
| 2 | 0.028147 | -0.007568 | -0.050268 | 0.046678 | -0.044179 | -0.015935 | efectores |
| 3 | -0.083518 | -0.032929 | 0.026655 | -0.021762 | 0.001202 | 0.071403 | efectores |
| 4 | 0.038455 | 0.020836 | 0.119455 | 0.040215 | 0.061706 | 0.088167 | efectores |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.066609 | 0.077565 | 0.149456 | 0.016210 | -0.058814 | -0.123535 | efectores |
| 996 | 0.070428 | 0.009412 | 0.035696 | 0.007972 | -0.099162 | 0.000970 | efectores |
| 997 | -0.075273 | 0.039889 | 0.053082 | -0.022417 | 0.057374 | -0.050513 | efectores |
| 998 | -0.064832 | 0.065257 | -0.034440 | 0.039394 | -0.047467 | -0.045684 | efectores |
| 999 | 0.022425 | 0.008115 | 0.104509 | 0.007036 | -0.003060 | -0.016903 | efectores |

[937 rows x 14 columns]

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

Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 937.000000 | 937.000000 | 937.000000 | 937.000000 | 937.000000 | 937.000000 |
| mean | 0.008694 | 0.018280 | 0.011912 | 0.013358 | -0.005806 | -0.000978 |
| std | 0.064333 | 0.065341 | 0.062860 | 0.063290 | 0.065211 | 0.063348 |
| min | -0.197530 | -0.191380 | -0.177268 | -0.175447 | -0.200777 | -0.203370 |
| 25% | -0.030864 | -0.022640 | -0.029052 | -0.027275 | -0.047469 | -0.042791 |
| 50% | 0.007366 | 0.017412 | 0.008497 | 0.012959 | -0.003088 | 0.003987 |
| 75% | 0.048661 | 0.060426 | 0.054827 | 0.059035 | 0.037445 | 0.038489 |

| | | | | | | |
|-----|----------|----------|----------|----------|----------|----------|
| max | 0.194829 | 0.219450 | 0.189017 | 0.204206 | 0.196978 | 0.191594 |
|-----|----------|----------|----------|----------|----------|----------|

| | X6 | X7 | X8 | X9 | X10 | X11 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 937.000000 | 937.000000 | 937.000000 | 937.000000 | 937.000000 | 937.000000 |
| mean | 0.016958 | 0.002473 | -0.004763 | 0.008617 | 0.003272 | -0.005148 |
| std | 0.068143 | 0.062191 | 0.066212 | 0.068956 | 0.064684 | 0.061609 |
| min | -0.194633 | -0.193764 | -0.206914 | -0.191389 | -0.190070 | -0.178000 |
| 25% | -0.027334 | -0.032547 | -0.049441 | -0.032916 | -0.040380 | -0.043190 |
| 50% | 0.015647 | 0.002268 | -0.003979 | 0.008475 | 0.001282 | -0.004675 |
| 75% | 0.058703 | 0.041628 | 0.037796 | 0.052034 | 0.044802 | 0.034348 |
| max | 0.225037 | 0.193795 | 0.200822 | 0.217885 | 0.191931 | 0.188673 |

| | X12 |
|-------|------------|
| count | 937.000000 |
| mean | 0.013507 |
| std | 0.067285 |
| min | -0.187702 |
| 25% | -0.034334 |
| 50% | 0.012123 |
| 75% | 0.061529 |
| max | 0.222029 |

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.022156 | 0.011357 | 0.014821 | 0.026827 | 0.034056 | -0.001082 | 0.051075 |
| 1 | 0.036065 | 0.025745 | 0.061728 | 0.043862 | -0.021754 | 0.075709 | -0.017334 |
| 3 | 0.000457 | -0.069847 | 0.093453 | 0.058773 | -0.108883 | 0.082633 | 0.131747 |
| 5 | 0.058842 | 0.105379 | -0.021232 | 0.006275 | -0.073154 | -0.019955 | -0.012804 |
| 6 | 0.040239 | 0.030984 | 0.071625 | -0.020042 | -0.071026 | -0.001252 | -0.061482 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.059785 | -0.027676 | -0.025231 | 0.126949 | 0.062763 | -0.134173 | -0.008077 |
| 996 | 0.041958 | -0.078818 | 0.022183 | 0.028601 | -0.008870 | -0.009356 | -0.052238 |
| 997 | -0.072576 | -0.066887 | -0.165832 | 0.122460 | 0.019564 | 0.040797 | -0.035878 |
| 998 | -0.140859 | 0.012119 | -0.011704 | 0.019137 | -0.123168 | -0.019680 | -0.150390 |
| 999 | -0.068806 | -0.014602 | -0.037819 | 0.021121 | 0.106009 | -0.090962 | 0.077878 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | 0.008119 | 0.024437 | -0.014068 | 0.000133 | 0.017149 | 0.040719 | no_efectores |
| 1 | -0.017676 | 0.054559 | 0.027064 | -0.033452 | 0.031972 | 0.079482 | no_efectores |
| 3 | 0.015062 | 0.051231 | 0.085413 | -0.023376 | -0.024658 | 0.016740 | no_efectores |
| 5 | -0.073054 | -0.034731 | 0.033310 | 0.068689 | 0.061954 | 0.012306 | no_efectores |
| 6 | -0.014697 | 0.054025 | 0.051917 | 0.001560 | 0.018100 | -0.018365 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | ... |

```

995  0.070846  0.146722 -0.022102  0.029962 -0.016114  0.090253  no_efectores
996 -0.091426  0.012115 -0.046669 -0.018373  0.055251 -0.000757  no_efectores
997  0.055530 -0.059779 -0.064707  0.181051  0.041524  0.038964  no_efectores
998  0.205794 -0.214536 -0.032255  0.071698  0.108499 -0.085263  no_efectores
999 -0.031377  0.059889  0.011886 -0.100060  0.090595 -0.019082  no_efectores

```

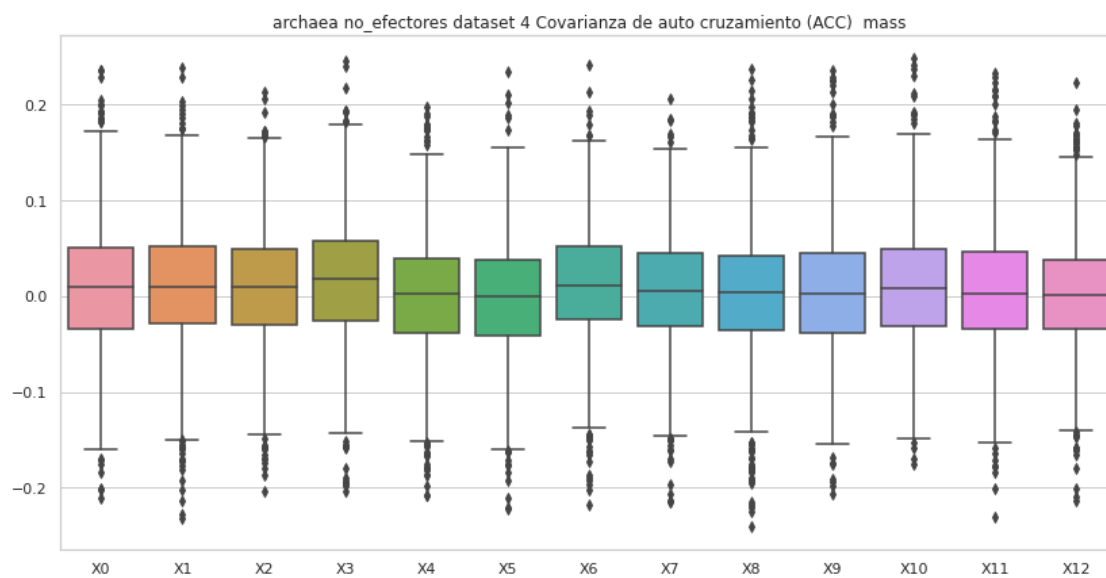
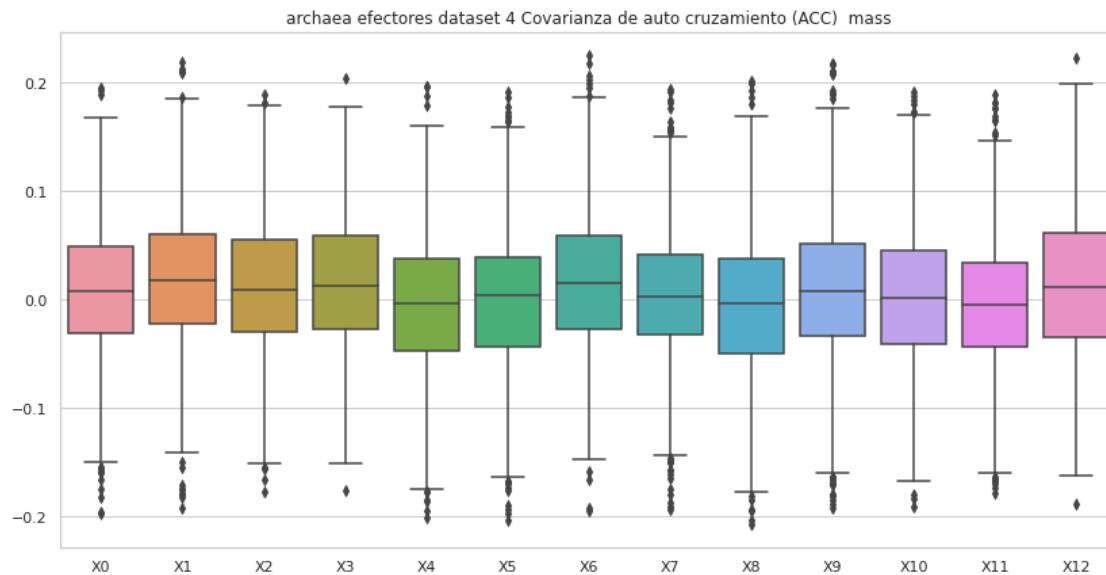
[902 rows x 14 columns]

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

| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 902.000000 | 902.000000 | 902.000000 | 902.000000 | 902.000000 | 902.000000 |
| mean | 0.009660 | 0.009745 | 0.008796 | 0.016481 | 0.000556 | -0.001303 |
| std | 0.068721 | 0.067722 | 0.063897 | 0.067616 | 0.063430 | 0.065625 |
| min | -0.211466 | -0.232113 | -0.204443 | -0.204069 | -0.208351 | -0.222152 |
| 25% | -0.033736 | -0.028875 | -0.029439 | -0.025303 | -0.037768 | -0.041275 |
| 50% | 0.009883 | 0.010296 | 0.009059 | 0.018694 | 0.002978 | -0.000492 |
| 75% | 0.050993 | 0.051842 | 0.048648 | 0.057450 | 0.038943 | 0.038070 |
| max | 0.235989 | 0.238809 | 0.213582 | 0.245622 | 0.197379 | 0.233822 |

| | X6 | X7 | X8 | X9 | X10 | X11 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 902.000000 | 902.000000 | 902.000000 | 902.000000 | 902.000000 | 902.000000 |
| mean | 0.010700 | 0.005814 | 0.002740 | 0.003582 | 0.010701 | 0.005565 |
| std | 0.063956 | 0.063199 | 0.069352 | 0.067146 | 0.065461 | 0.066599 |
| min | -0.217428 | -0.215093 | -0.240527 | -0.206498 | -0.174847 | -0.231053 |
| 25% | -0.024104 | -0.031427 | -0.035061 | -0.038231 | -0.031269 | -0.034061 |
| 50% | 0.010424 | 0.006025 | 0.004269 | 0.002672 | 0.008065 | 0.003291 |
| 75% | 0.051687 | 0.045457 | 0.042163 | 0.045350 | 0.049562 | 0.046826 |
| max | 0.241156 | 0.205794 | 0.236932 | 0.235350 | 0.248512 | 0.233203 |

| | X12 |
|-------|------------|
| count | 902.000000 |
| mean | 0.002532 |
| std | 0.063221 |
| min | -0.213234 |
| 25% | -0.033550 |
| 50% | 0.000881 |
| 75% | 0.038248 |
| max | 0.222855 |



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 4, con valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.073103 | 0.047888 | 0.229624 | 0.133804 | 0.138406 | 0.096016 | 0.133291 |
| 1 | 0.132197 | 0.087201 | 0.173566 | 0.004457 | -0.089030 | -0.081952 | -0.043823 |
| 2 | 0.206988 | 0.038727 | 0.160886 | 0.167273 | 0.158495 | 0.193430 | 0.163025 |
| 3 | 0.190966 | 0.013306 | 0.108509 | 0.120342 | 0.051260 | 0.106575 | 0.200362 |
| 4 | 0.181216 | 0.102647 | 0.220701 | 0.205972 | 0.173005 | 0.118284 | 0.187037 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.086148 | -0.159448 | 0.183191 | 0.088856 | -0.057810 | 0.074775 | -0.007516 |
| 996 | 0.034801 | -0.186235 | -0.061274 | 0.045300 | -0.116021 | -0.095504 | -0.137011 |
| 997 | -0.024582 | -0.284761 | 0.033355 | 0.174170 | -0.137643 | -0.031865 | 0.132479 |
| 998 | 0.038964 | -0.074190 | 0.052857 | -0.020411 | -0.172907 | -0.046421 | -0.100797 |
| 999 | -0.057342 | -0.059894 | 0.007538 | -0.002499 | -0.000530 | -0.064178 | -0.004205 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|---|-----------|----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.132043 | 0.126882 | 0.066562 | 0.051675 | 0.102018 | 0.006840 | efectores |
| 1 | -0.116984 | 0.023273 | -0.074093 | -0.051100 | -0.047509 | -0.041220 | efectores |
| 2 | -0.007763 | 0.067612 | 0.051926 | 0.244052 | 0.177480 | -0.121385 | efectores |

| | | | | | | | |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 3 | 0.097954 | 0.004799 | 0.023843 | 0.026771 | 0.014405 | -0.016163 | efectores |
| 4 | 0.167185 | 0.164737 | 0.139387 | 0.090314 | 0.144300 | 0.110662 | efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | -0.041769 | 0.109429 | -0.034051 | -0.120628 | 0.079967 | 0.201832 | efectores |
| 996 | 0.005455 | 0.106490 | 0.163813 | 0.088847 | -0.103070 | -0.074917 | efectores |
| 997 | -0.083152 | -0.156506 | -0.040095 | -0.019106 | -0.031124 | 0.038410 | efectores |
| 998 | 0.046373 | -0.011373 | -0.084061 | 0.053383 | 0.044857 | -0.033528 | efectores |
| 999 | -0.014105 | -0.008712 | 0.022217 | 0.009260 | 0.034740 | 0.012528 | efectores |

[1000 rows x 14 columns]

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

| | X0 | X1 | X2 | X3 | X4 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.021692 | -0.027724 | 0.048202 | 0.036103 | -0.013778 |
| std | 0.093155 | 0.100098 | 0.086061 | 0.088264 | 0.096437 |
| min | -0.312979 | -0.427764 | -0.189479 | -0.345800 | -0.486959 |
| 25% | -0.034780 | -0.094602 | -0.010066 | -0.016891 | -0.078702 |
| 50% | 0.025678 | -0.018242 | 0.043008 | 0.034673 | -0.006742 |
| 75% | 0.082267 | 0.044366 | 0.102930 | 0.086489 | 0.048804 |
| max | 0.293280 | 0.213932 | 0.455688 | 0.338793 | 0.240639 |

| | X5 | X6 | X7 | X8 | X9 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | -0.018154 | 0.027444 | 0.026488 | 0.001637 | 0.000260 |
| std | 0.090083 | 0.093229 | 0.085363 | 0.090876 | 0.090709 |
| min | -0.387334 | -0.294086 | -0.447834 | -0.353146 | -0.316211 |
| 25% | -0.075875 | -0.027562 | -0.027193 | -0.050200 | -0.050125 |
| 50% | -0.014753 | 0.022657 | 0.023558 | 0.006032 | 0.002760 |
| 75% | 0.038890 | 0.077032 | 0.075658 | 0.050239 | 0.047627 |
| max | 0.249861 | 0.391443 | 0.382895 | 0.347657 | 0.368287 |

| | X10 | X11 | X12 |
|-------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.018039 | 0.014509 | -0.009429 |
| std | 0.077521 | 0.081462 | 0.088262 |
| min | -0.332635 | -0.324915 | -0.406143 |
| 25% | -0.031241 | -0.035323 | -0.062074 |
| 50% | 0.014120 | 0.010301 | -0.007328 |
| 75% | 0.067856 | 0.056774 | 0.039266 |
| max | 0.249589 | 0.304007 | 0.368176 |

no_efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.021389 | -0.023221 | -0.003376 | -0.006759 | 0.030149 | 0.010946 | 0.043874 |
| 1 | 0.086864 | 0.052275 | 0.103618 | 0.078714 | 0.031883 | 0.061350 | 0.052245 |
| 2 | -0.027261 | 0.013393 | 0.113626 | -0.121693 | -0.042602 | -0.016505 | -0.068548 |
| 3 | 0.122311 | -0.035359 | 0.056876 | 0.005380 | -0.004078 | 0.062871 | -0.047718 |
| 4 | -0.132960 | -0.098937 | -0.141043 | -0.162205 | 0.153765 | -0.090614 | 0.100243 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.067231 | 0.021966 | 0.076491 | 0.041262 | 0.085886 | -0.052002 | -0.035470 |
| 996 | -0.097207 | -0.087412 | 0.071586 | -0.056692 | 0.068012 | -0.039669 | 0.071994 |
| 997 | 0.036834 | 0.070139 | 0.059536 | 0.165555 | 0.087895 | 0.095179 | 0.124990 |
| 998 | -0.099871 | 0.047252 | -0.234885 | 0.238080 | -0.094907 | -0.078226 | -0.117305 |
| 999 | 0.034404 | -0.113361 | -0.042875 | 0.041700 | 0.069936 | 0.048365 | -0.003333 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | -0.007754 | -0.028965 | -0.027283 | -0.004521 | 0.015404 | -0.006990 | no_efectores |
| 1 | 0.075153 | 0.034201 | 0.027600 | 0.020005 | -0.035967 | -0.016705 | no_efectores |
| 2 | 0.115860 | 0.168054 | -0.183605 | 0.133668 | 0.047494 | -0.120922 | no_efectores |
| 3 | -0.045343 | 0.089814 | 0.011125 | -0.028297 | -0.084254 | -0.016574 | no_efectores |
| 4 | -0.253715 | 0.025736 | 0.073169 | 0.191246 | 0.104469 | -0.104437 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | -0.005255 | -0.036908 | -0.023992 | -0.030262 | -0.030782 | 0.033767 | no_efectores |
| 996 | -0.126205 | -0.088768 | 0.062843 | 0.041428 | 0.044398 | -0.019511 | no_efectores |
| 997 | -0.016868 | 0.069789 | 0.023366 | 0.025318 | 0.034364 | -0.097421 | no_efectores |
| 998 | 0.087457 | 0.056869 | 0.116679 | -0.062177 | -0.104651 | 0.189224 | no_efectores |
| 999 | 0.044915 | 0.106153 | -0.131960 | -0.004335 | 0.098135 | -0.017842 | no_efectores |

[1000 rows x 14 columns]

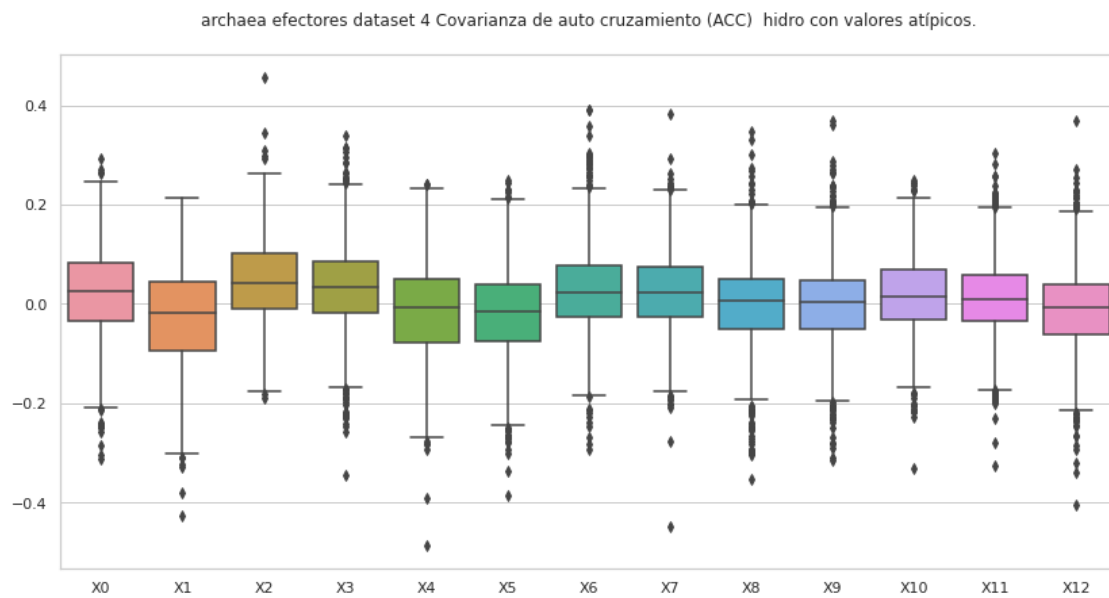
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 4, con valores atípicos.

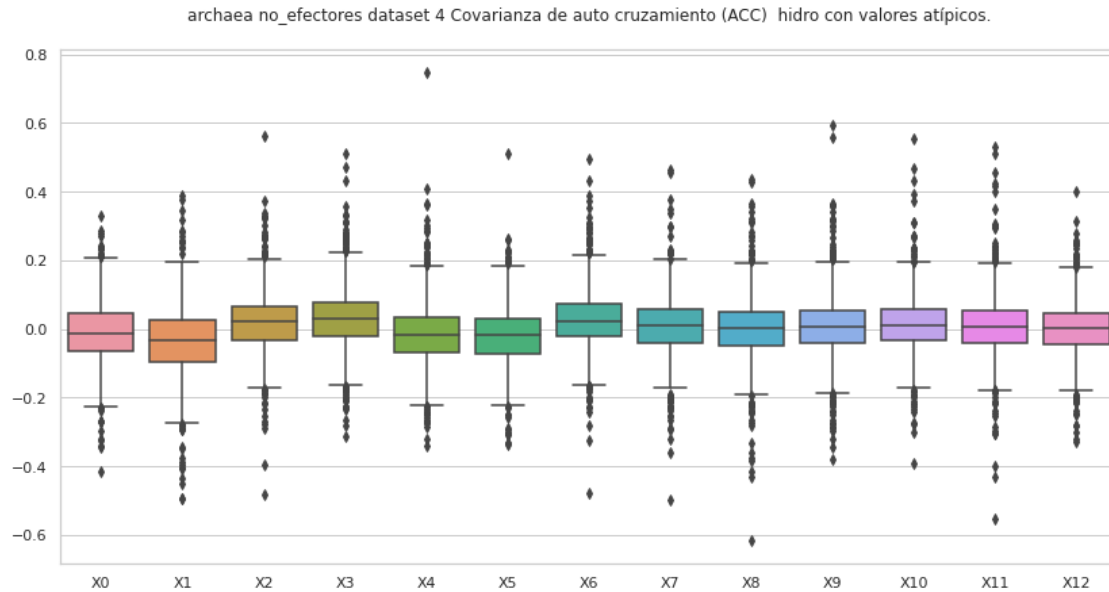
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | -0.009105 | -0.036013 | 0.019300 | 0.030336 | -0.015212 |
| std | 0.088610 | 0.103260 | 0.087002 | 0.093478 | 0.095759 |
| min | -0.415936 | -0.494413 | -0.483153 | -0.313404 | -0.339863 |
| 25% | -0.063915 | -0.094419 | -0.031842 | -0.020538 | -0.070093 |
| 50% | -0.011779 | -0.034456 | 0.020473 | 0.028326 | -0.018625 |
| 75% | 0.045836 | 0.025430 | 0.064009 | 0.077758 | 0.032224 |
| max | 0.331095 | 0.387420 | 0.563856 | 0.512068 | 0.747245 |

| | X5 | X6 | X7 | X8 | X9 \ |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | -0.018976 | 0.026876 | 0.008599 | 0.000870 | 0.006428 |
| std | 0.088596 | 0.090910 | 0.087238 | 0.093896 | 0.092202 |
| min | -0.335986 | -0.479718 | -0.499939 | -0.616275 | -0.381167 |
| 25% | -0.070692 | -0.022477 | -0.039377 | -0.048553 | -0.041684 |
| 50% | -0.015331 | 0.023186 | 0.009001 | 0.002596 | 0.005942 |
| 75% | 0.031950 | 0.072752 | 0.057377 | 0.047957 | 0.054182 |
| max | 0.511724 | 0.493722 | 0.463518 | 0.435126 | 0.594310 |

| | X10 | X11 | X12 |
|-------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.013458 | 0.010395 | 0.002494 |
| std | 0.087265 | 0.093232 | 0.080874 |
| min | -0.391217 | -0.554655 | -0.330246 |
| 25% | -0.033863 | -0.038949 | -0.046262 |
| 50% | 0.012189 | 0.006842 | 0.002831 |
| 75% | 0.057904 | 0.054222 | 0.044715 |
| max | 0.553495 | 0.529428 | 0.402342 |





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 4, sin valores atípicos.

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.073103 | 0.047888 | 0.229624 | 0.133804 | 0.138406 | 0.096016 | 0.133291 |
| 1 | 0.132197 | 0.087201 | 0.173566 | 0.004457 | -0.089030 | -0.081952 | -0.043823 |
| 2 | 0.206988 | 0.038727 | 0.160886 | 0.167273 | 0.158495 | 0.193430 | 0.163025 |
| 3 | 0.190966 | 0.013306 | 0.108509 | 0.120342 | 0.051260 | 0.106575 | 0.200362 |
| 4 | 0.181216 | 0.102647 | 0.220701 | 0.205972 | 0.173005 | 0.118284 | 0.187037 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.086148 | -0.159448 | 0.183191 | 0.088856 | -0.057810 | 0.074775 | -0.007516 |
| 996 | 0.034801 | -0.186235 | -0.061274 | 0.045300 | -0.116021 | -0.095504 | -0.137011 |
| 997 | -0.024582 | -0.284761 | 0.033355 | 0.174170 | -0.137643 | -0.031865 | 0.132479 |
| 998 | 0.038964 | -0.074190 | 0.052857 | -0.020411 | -0.172907 | -0.046421 | -0.100797 |
| 999 | -0.057342 | -0.059894 | 0.007538 | -0.002499 | -0.000530 | -0.064178 | -0.004205 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|-----|-----------|----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.132043 | 0.126882 | 0.066562 | 0.051675 | 0.102018 | 0.006840 | efectores |
| 1 | -0.116984 | 0.023273 | -0.074093 | -0.051100 | -0.047509 | -0.041220 | efectores |
| 2 | -0.007763 | 0.067612 | 0.051926 | 0.244052 | 0.177480 | -0.121385 | efectores |
| 3 | 0.097954 | 0.004799 | 0.023843 | 0.026771 | 0.014405 | -0.016163 | efectores |
| 4 | 0.167185 | 0.164737 | 0.139387 | 0.090314 | 0.144300 | 0.110662 | efectores |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 995 | -0.041769 | 0.109429 | -0.034051 | -0.120628 | 0.079967 | 0.201832 | efectores |

```

996  0.005455  0.106490  0.163813  0.088847 -0.103070 -0.074917  efectores
997 -0.083152 -0.156506 -0.040095 -0.019106 -0.031124  0.038410  efectores
998  0.046373 -0.011373 -0.084061  0.053383  0.044857 -0.033528  efectores
999 -0.014105 -0.008712  0.022217  0.009260  0.034740  0.012528  efectores

```

[936 rows x 14 columns]

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

Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 936.000000 | 936.000000 | 936.000000 | 936.000000 | 936.000000 | 936.000000 |
| mean | 0.023624 | -0.024459 | 0.046158 | 0.038165 | -0.010987 | -0.015752 |
| std | 0.088054 | 0.095524 | 0.082740 | 0.082684 | 0.091383 | 0.085655 |
| min | -0.247711 | -0.310964 | -0.189479 | -0.227881 | -0.293520 | -0.280113 |
| 25% | -0.032942 | -0.090287 | -0.011844 | -0.014870 | -0.076223 | -0.071157 |
| 50% | 0.026185 | -0.015138 | 0.042457 | 0.035469 | -0.005033 | -0.013116 |
| 75% | 0.081898 | 0.044653 | 0.100431 | 0.087928 | 0.049561 | 0.038634 |
| max | 0.293280 | 0.213932 | 0.297843 | 0.294364 | 0.240639 | 0.249861 |

| | X6 | X7 | X8 | X9 | X10 | X11 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 936.000000 | 936.000000 | 936.000000 | 936.000000 | 936.000000 | 936.000000 |
| mean | 0.024810 | 0.027168 | 0.003573 | 0.000528 | 0.017408 | 0.013823 |
| std | 0.084883 | 0.079538 | 0.080990 | 0.081199 | 0.073230 | 0.075288 |
| min | -0.247095 | -0.203365 | -0.266787 | -0.268182 | -0.210591 | -0.200939 |
| 25% | -0.027191 | -0.025837 | -0.045640 | -0.047432 | -0.030895 | -0.033573 |
| 50% | 0.021542 | 0.023845 | 0.006107 | 0.003188 | 0.012283 | 0.009946 |
| 75% | 0.072704 | 0.075129 | 0.049477 | 0.045357 | 0.065433 | 0.054044 |
| max | 0.300658 | 0.263230 | 0.266740 | 0.263118 | 0.244052 | 0.256257 |

| | X12 |
|-------|------------|
| count | 936.000000 |
| mean | -0.008772 |
| std | 0.080694 |
| min | -0.266930 |
| 25% | -0.057623 |
| 50% | -0.007151 |
| 75% | 0.036804 |
| max | 0.253083 |

no_efectores

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

Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 \ |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.021389 | -0.023221 | -0.003376 | -0.006759 | 0.030149 | 0.010946 | 0.043874 |
| 1 | 0.086864 | 0.052275 | 0.103618 | 0.078714 | 0.031883 | 0.061350 | 0.052245 |
| 2 | -0.027261 | 0.013393 | 0.113626 | -0.121693 | -0.042602 | -0.016505 | -0.068548 |
| 3 | 0.122311 | -0.035359 | 0.056876 | 0.005380 | -0.004078 | 0.062871 | -0.047718 |
| 5 | 0.018298 | 0.056027 | -0.065494 | 0.038032 | -0.093607 | -0.069253 | 0.023633 |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | 0.067231 | 0.021966 | 0.076491 | 0.041262 | 0.085886 | -0.052002 | -0.035470 |
| 996 | -0.097207 | -0.087412 | 0.071586 | -0.056692 | 0.068012 | -0.039669 | 0.071994 |
| 997 | 0.036834 | 0.070139 | 0.059536 | 0.165555 | 0.087895 | 0.095179 | 0.124990 |
| 998 | -0.099871 | 0.047252 | -0.234885 | 0.238080 | -0.094907 | -0.078226 | -0.117305 |
| 999 | 0.034404 | -0.113361 | -0.042875 | 0.041700 | 0.069936 | 0.048365 | -0.003333 |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| 0 | -0.007754 | -0.028965 | -0.027283 | -0.004521 | 0.015404 | -0.006990 | no_efectores |
| 1 | 0.075153 | 0.034201 | 0.027600 | 0.020005 | -0.035967 | -0.016705 | no_efectores |
| 2 | 0.115860 | 0.168054 | -0.183605 | 0.133668 | 0.047494 | -0.120922 | no_efectores |
| 3 | -0.045343 | 0.089814 | 0.011125 | -0.028297 | -0.084254 | -0.016574 | no_efectores |
| 5 | 0.022819 | -0.028049 | -0.058511 | 0.056626 | -0.022739 | 0.025485 | no_efectores |
| .. | ... | ... | ... | ... | ... | ... | |
| 995 | -0.005255 | -0.036908 | -0.023992 | -0.030262 | -0.030782 | 0.033767 | no_efectores |
| 996 | -0.126205 | -0.088768 | 0.062843 | 0.041428 | 0.044398 | -0.019511 | no_efectores |
| 997 | -0.016868 | 0.069789 | 0.023366 | 0.025318 | 0.034364 | -0.097421 | no_efectores |
| 998 | 0.087457 | 0.056869 | 0.116679 | -0.062177 | -0.104651 | 0.189224 | no_efectores |
| 999 | 0.044915 | 0.106153 | -0.131960 | -0.004335 | 0.098135 | -0.017842 | no_efectores |

[903 rows x 14 columns]

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

| | X0 | X1 | X2 | X3 | X4 | X5 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 903.000000 | 903.000000 | 903.000000 | 903.000000 | 903.000000 | 903.000000 |
| mean | -0.009200 | -0.035325 | 0.018758 | 0.027518 | -0.016101 | -0.019039 |
| std | 0.077753 | 0.087320 | 0.072297 | 0.080251 | 0.078442 | 0.077003 |
| min | -0.269776 | -0.294729 | -0.234885 | -0.233493 | -0.285459 | -0.248371 |
| 25% | -0.060763 | -0.090691 | -0.030036 | -0.018268 | -0.067361 | -0.069126 |
| 50% | -0.011788 | -0.034613 | 0.020476 | 0.028083 | -0.018601 | -0.015082 |
| 75% | 0.042367 | 0.022297 | 0.062539 | 0.074128 | 0.030392 | 0.028908 |
| max | 0.237613 | 0.250008 | 0.273910 | 0.278880 | 0.237396 | 0.232526 |

| | X6 | X7 | X8 | X9 | X10 | X11 \ |
|-------|------------|------------|------------|------------|------------|------------|
| count | 903.000000 | 903.000000 | 903.000000 | 903.000000 | 903.000000 | 903.000000 |
| mean | 0.022544 | 0.008026 | 0.001033 | 0.003208 | 0.009882 | 0.006652 |
| std | 0.074627 | 0.071784 | 0.073985 | 0.073104 | 0.070891 | 0.072584 |

| | | | | | | |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|
| min | -0.226930 | -0.229025 | -0.247386 | -0.264073 | -0.242306 | -0.254552 |
| 25% | -0.021850 | -0.037568 | -0.046522 | -0.040227 | -0.031643 | -0.038046 |
| 50% | 0.020919 | 0.009141 | 0.003288 | 0.004718 | 0.010815 | 0.005300 |
| 75% | 0.067063 | 0.055230 | 0.046014 | 0.050763 | 0.053155 | 0.049847 |
| max | 0.277533 | 0.231063 | 0.270395 | 0.237814 | 0.235291 | 0.252374 |

| | X12 |
|-------|------------|
| count | 903.000000 |
| mean | 0.002828 |
| std | 0.068774 |
| min | -0.236771 |
| 25% | -0.041010 |
| 50% | 0.002965 |
| 75% | 0.042081 |
| max | 0.236870 |

