



Componentes Principales Análisis

Presentado en la fecha: 7/09/2019

Hecho por: Andrada Alexander

Encina Guadalupe

Huarca Brian

Contenido

| | |
|---|-----------|
| Introducción | 2 |
| Objetivo | 4 |
| Componentes Principales | 5 |
| Datos Iniciales | 5 |
| Tabla de componentes principales | 6 |
| Gráfico de Sedimentacion para determinar CP | 7 |
| Gráfico de Correlación sin clase | 8 |
| Gráfico de Dispersion Ozonoe | 9 |
| Gráfico de relación de componentes 1 y 2 | 10 |
| Tabla de Componentes principales para un dataset con Contaminacion | 11 |
| Gráfico de Sedimentacion para determinar CP | 12 |
| Gráfico de Correlación con Contaminacion | 13 |
| Gráfico de Dispersión con Contaminación | 14 |
| Gráfico de Dispersión con Contaminación | 15 |
| Gráfico de relación de componentes 1 y 2 | 16 |
| Tabla de Componentes principales para un dataset Sin Contaminacion | 17 |
| Gráfico de Sedimentacion para determinar CP | 18 |
| Círculo de Correlación sin contaminación | 19 |
| Gráfico de Dispersión sin Contaminación | 20 |
| Gráfico de Dispersión sin Contaminación | 21 |
| Gráfico de relación de componentes 1 y 2 | 22 |
| Conclusión | 23 |
| Anexo | 24 |

Introducción

En el presente trabajo desarrollaremos el proceso llevado a cabo para la realización del Análisis de Componentes Principales de un conjunto de datos dado o dataset. En primera medida se realizará un proceso de ETL, bien conocido y aplicado en el ambiente de BI y Big Data, este proceso nos permitirá una mejor manipulación de datos puros sin valores null, simbolo desconocido, cualquier otra anomalia, o simplemente para una selección de datos de interes. Despues del filtrado de datos se trabajará sobre el dataset resultante para analisis, explorar y obtener predicciones y aproximaciones sobre los datos.

El conjunto en cuestión se trata de una representación de las variaciones en la concentración de Ozono, dadas diferentes variables, a saber:

- **T:** Temperatura
- **T-PK:** continuo. Pico T
- **WSR-PK:** continuo. Pico de velocidad del viento - resultante (promedio del vector del viento)
- **WSR-AV:** continuo. Velocidad media del viento
- **T-AV:** continuo. Promedio T
- **T85:** continuo. T a nivel de 850 hpa o aproximadamente 1500m de altura.
- **RH85:** continuo. Humedad relativa a 850 hpa
- **U85:** continuo. (Viento U - viento de dirección este-oeste a 850 hpa)
- **V85:** continuo. Viento V - viento de dirección N-S a 850
- **HT85:** continuo. Altura geopotencial a 850 hpa, es casi lo mismo que la altura a baja altitud

- **T70:** continuo. T a nivel de 700 hpa (aproximadamente 3100 m de altura)
- **RH70:** continuo.
- **U70:** continuo.
- **V70:** continuo.
- **HT70:** continuo.
- **T50:** continuo. T a nivel de 500 hpa (aproximadamente a 5500 m de altura)
- **RH50:** continuo.
- **U50:** continuo.
- **V50:** continuo.
- **HT50:** continuo.
- **KI:** continuo. Índice K [Enlace web]
- **TT:** continuo. T-Totales [Enlace web]
- **SLP:** continuo. Presión a nivel del mar
- **SLP-:** continuo. Cambio de SLP del día anterior
- **Precp:** continuo. - precipitación

La gran dimension del conjunto de datos, incluso habiendo separado las componentes principales, nos presentó algunas dificultades al momento de realizar una correcta o eficiente visualización de los datos.

Dicho esto, para el desarrollo del análisis pedido utilizamos la herramienta InfoStat, habiéndonos brindado ésta una mejor manera de representar tantas componentes principales y puntos.

Objetivo

Se trata de, a partir del conjunto de datos descrito en la introducción, representar un dataset con muchas variables en un número de ejes reducido, dado por las componentes principales.

Una de las principales premisas de este análisis es la construcción de una línea de mejor ajuste, la cual se puede obtener de la suma de los cuadrados de la perpendiculares del sistema de puntos de un plano determinado.

Partiendo de esta base, buscamos realizar una visualización lo más clara y representativa posible, tomando para esto a las componentes principales que cumplan con esto.

Componentes Principales

Datos Iniciales

La primera figura corresponde a los datos que se estara trabajando en el presente documento. El dataset pertece a datos de Ozono donde muestra variaciones entre temperatura y vientos. Notese que para esta imagen se ignora la columna fecha ya que no es de tipo entero y no importa en el analisis, en palabras simples, no es relevante.

| | WSR0 | WSR1 | WSR2 | WSR3 | WSR4 | WSR5 | WSR6 | WSR7 | WSR8 | WSR9 | WSR10 | WSR11 | WSR12 | WSR13 |
|----|------|------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|
| 1 | 0.8 | 1.8 | 2.4 | 2.1 | 2.0 | 2.1 | 1.5 | 1.7 | 1.9 | 2.3 | 3.7 | 5.5 | 5.1 | 5.4 |
| 2 | 2.8 | 3.2 | 3.3 | 2.7 | 3.3 | 3.2 | 2.9 | 2.8 | 3.1 | 3.4 | 4.2 | 4.5 | 4.5 | 4.3 |
| 3 | 2.9 | 2.8 | 2.6 | 2.1 | 2.2 | 2.5 | 2.5 | 2.7 | 2.2 | 2.5 | 3.1 | 4.0 | 4.4 | 4.6 |
| 4 | 4.7 | 3.8 | 3.7 | 3.8 | 2.9 | 3.1 | 2.8 | 2.5 | 2.4 | 3.1 | 3.3 | 3.1 | 2.3 | 2.1 |
| 5 | 3.7 | 3.2 | 3.8 | 5.1 | 6.0 | 7.0 | 6.3 | 6.4 | 6.3 | 5.4 | 6.1 | 6.4 | 6.1 | 5.8 |
| 6 | 2.2 | 2.9 | 3.4 | 4.2 | 4.7 | 4.7 | 5.3 | 4.9 | 5.2 | 6.0 | 5.9 | 6.1 | 6.8 | 6.3 |
| 7 | 0.9 | 0.6 | 0.5 | 0.5 | 0.6 | 0.4 | 0.4 | 0.6 | 1.3 | 1.5 | 2.8 | 3.3 | 3.1 | 2.7 |
| 8 | 1.1 | 1.7 | 1.4 | 1.5 | 0.9 | 1.5 | 1.4 | 1.6 | 1.9 | 1.9 | 2.7 | 4.2 | 4.3 | 3.8 |
| 9 | 3.7 | 4.2 | 3.1 | 2.6 | 2.3 | 2.3 | 1.7 | 1.0 | 1.3 | 1.9 | 2.3 | 2.6 | 3.4 | 3.4 |
| 10 | 1.0 | 0.6 | 0.3 | 1.1 | 1.3 | 1.2 | 1.0 | 1.3 | 3.0 | 2.7 | 3.0 | 4.3 | 4.2 | 4.2 |
| 11 | 1.2 | 1.2 | 1.6 | 1.7 | 1.4 | 1.2 | 1.4 | 1.4 | 1.1 | 2.2 | 2.2 | 2.1 | 1.8 | 1.7 |

Figure 1: Datos Iniciales

Tabla de componentes principales

A continuación se muestra la tabla del análisis de los componentes principales, cabe destacar que esta no contiene la clase.

```
> summary(acp2)
```

Importance of components:

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 | PC11 | PC12 |
|------------------------|---------|---------|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Standard deviation | 5.4941 | 3.8989 | 2.33886 | 1.78038 | 1.61852 | 1.38762 | 1.28712 | 1.08414 | 1.02763 | 0.97393 | 0.9295 | 0.88753 |
| Proportion of Variance | 0.4192 | 0.2111 | 0.07598 | 0.04402 | 0.03638 | 0.02674 | 0.02301 | 0.01632 | 0.01467 | 0.01317 | 0.0120 | 0.01094 |
| Cumulative Proportion | 0.4192 | 0.6304 | 0.70635 | 0.75037 | 0.78676 | 0.81350 | 0.83651 | 0.85283 | 0.86750 | 0.88067 | 0.8927 | 0.90361 |
| | PC13 | PC14 | PC15 | PC16 | PC17 | PC18 | PC19 | PC20 | PC21 | PC22 | PC23 | PC24 |
| Standard deviation | 0.83298 | 0.79731 | 0.75125 | 0.7249 | 0.65248 | 0.60336 | 0.56149 | 0.53762 | 0.51679 | 0.48860 | 0.4798 | 0.45328 |
| Proportion of Variance | 0.00964 | 0.00883 | 0.00784 | 0.0073 | 0.00591 | 0.00506 | 0.00438 | 0.00401 | 0.00371 | 0.00332 | 0.0032 | 0.00285 |
| Cumulative Proportion | 0.91325 | 0.92208 | 0.92992 | 0.9372 | 0.94313 | 0.94819 | 0.95256 | 0.95658 | 0.96029 | 0.96360 | 0.9668 | 0.96965 |
| | PC25 | PC26 | PC27 | PC28 | PC29 | PC30 | PC31 | PC32 | PC33 | PC34 | PC35 | PC36 |
| Standard deviation | 0.42996 | 0.41355 | 0.39410 | 0.3791 | 0.35346 | 0.33683 | 0.33350 | 0.32212 | 0.30529 | 0.29899 | 0.29261 | 0.28381 |
| Proportion of Variance | 0.00257 | 0.00238 | 0.00216 | 0.0020 | 0.00174 | 0.00158 | 0.00154 | 0.00144 | 0.00129 | 0.00124 | 0.00119 | 0.00112 |
| Cumulative Proportion | 0.97222 | 0.97460 | 0.97675 | 0.9788 | 0.98049 | 0.98206 | 0.98361 | 0.98505 | 0.98634 | 0.98758 | 0.98877 | 0.98989 |
| | PC37 | PC38 | PC39 | PC40 | PC41 | PC42 | PC43 | PC44 | PC45 | PC46 | PC47 | |
| Standard deviation | 0.27003 | 0.26409 | 0.26012 | 0.24891 | 0.24467 | 0.23657 | 0.22635 | 0.21915 | 0.20960 | 0.20181 | 0.17985 | |
| Proportion of Variance | 0.00101 | 0.00097 | 0.00094 | 0.00086 | 0.00083 | 0.00078 | 0.00071 | 0.00067 | 0.00061 | 0.00057 | 0.00045 | |
| Cumulative Proportion | 0.99090 | 0.99187 | 0.99281 | 0.99367 | 0.99450 | 0.99528 | 0.99599 | 0.99666 | 0.99727 | 0.99784 | 0.99829 | |
| | PC48 | PC49 | PC50 | PC51 | PC52 | PC53 | PC54 | PC55 | PC56 | PC57 | PC58 | |
| Standard deviation | 0.14173 | 0.12687 | 0.12412 | 0.11717 | 0.09662 | 0.09552 | 0.08146 | 0.07356 | 0.06470 | 0.06222 | 0.05705 | |
| Proportion of Variance | 0.00028 | 0.00022 | 0.00021 | 0.00019 | 0.00013 | 0.00013 | 0.00009 | 0.00008 | 0.00006 | 0.00005 | 0.00005 | |
| Cumulative Proportion | 0.99856 | 0.99879 | 0.99900 | 0.99919 | 0.99932 | 0.99945 | 0.99954 | 0.99962 | 0.99967 | 0.99973 | 0.99977 | |
| | PC59 | PC60 | PC61 | PC62 | PC63 | PC64 | PC65 | PC66 | PC67 | PC68 | PC69 | |
| Standard deviation | 0.05086 | 0.04831 | 0.04370 | 0.04250 | 0.03750 | 0.03412 | 0.03325 | 0.02997 | 0.02884 | 0.02751 | 0.02461 | |
| Proportion of Variance | 0.00004 | 0.00003 | 0.00003 | 0.00003 | 0.00002 | 0.00002 | 0.00002 | 0.00001 | 0.00001 | 0.00001 | 0.00001 | |
| Cumulative Proportion | 0.99981 | 0.99984 | 0.99987 | 0.99989 | 0.99991 | 0.99993 | 0.99994 | 0.99996 | 0.99997 | 0.99998 | 0.99999 | |
| | PC70 | PC71 | PC72 | | | | | | | | | |
| Standard deviation | 0.02234 | 0.01990 | 0.004221 | | | | | | | | | |
| Proportion of Variance | 0.00001 | 0.00001 | 0.000000 | | | | | | | | | |
| Cumulative Proportion | 0.99999 | 1.00000 | 1.000000 | | | | | | | | | |

```
> |
```

Figure 2: Tabla de componentes principales

Gráfico de Sedimentacion para determinar CP

Este grafico nos permitira determinar que CP debemos seleccionar de acuerdo a la significancia del mismo, dicha relevancia se deducira en base a la tabla anterior buscando un gran porcentaje de proporcion acumulada (0.78 hasta la CP 5)

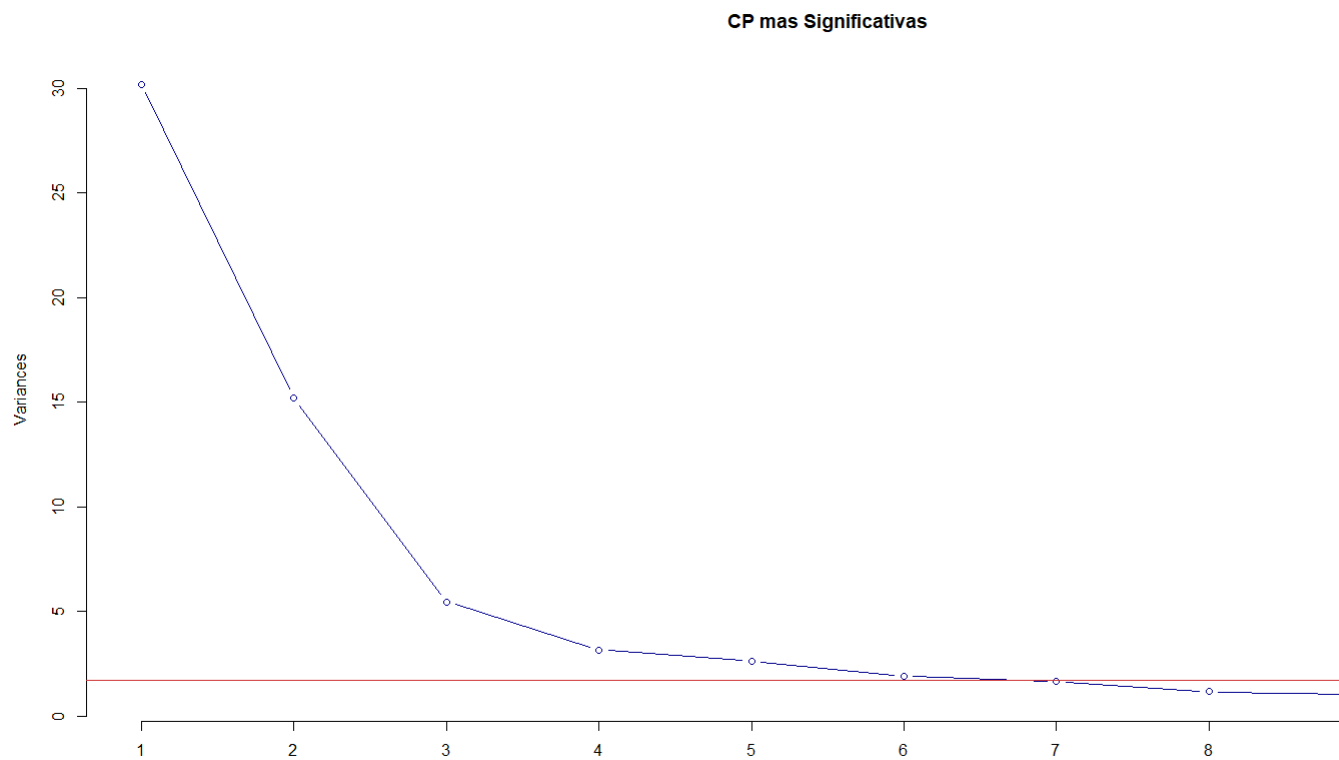


Figure 3: Sedimentacion Ozono

Gráfico de Correlación sin clase

A continuación se muestra el gráfico de correlación sin clase.

CP1-CP2

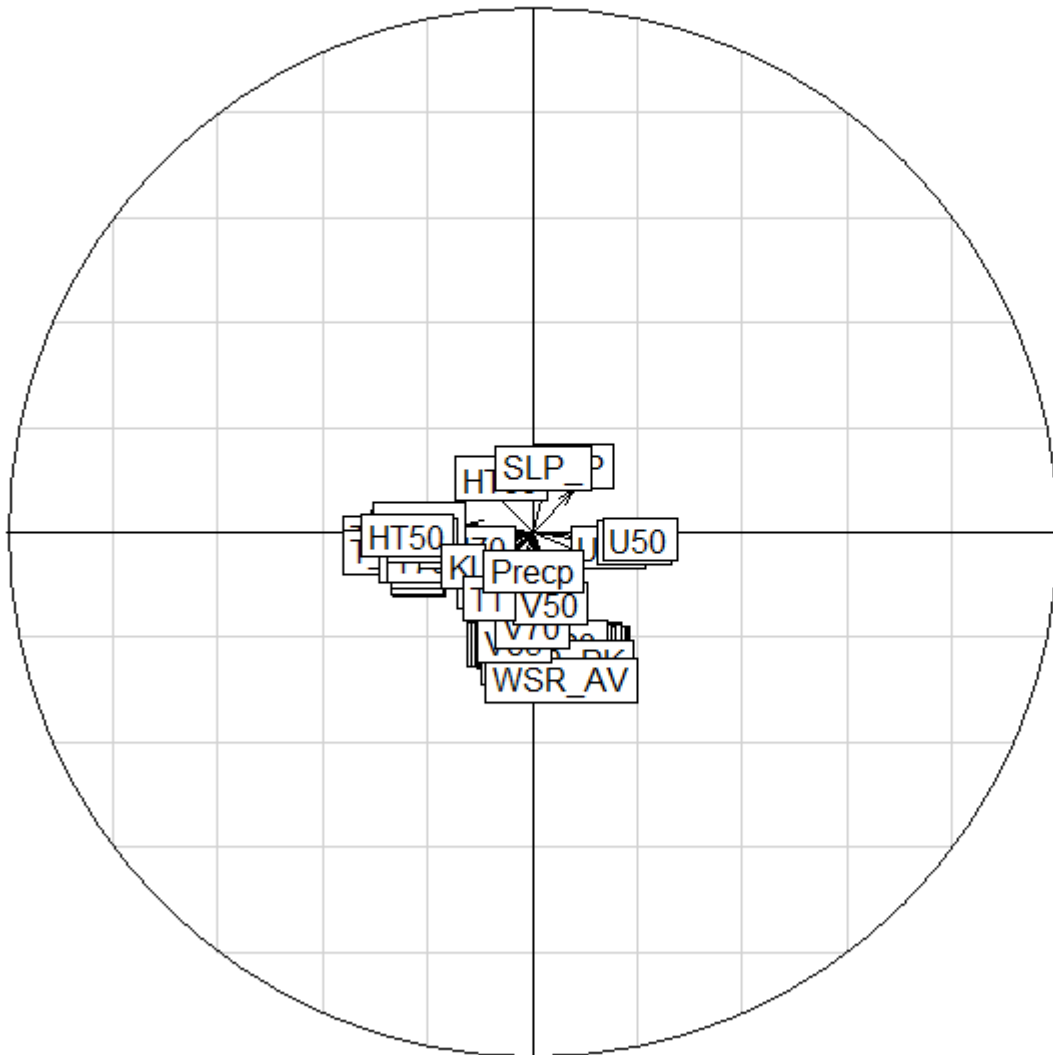


Figure 4: Gráfico de correlación sin clase

Gráfico de Dispersion Ozono

A continuación se muestra el gráfico de correlación sin clase mostrando una dispersion de todas las variables en el plano.

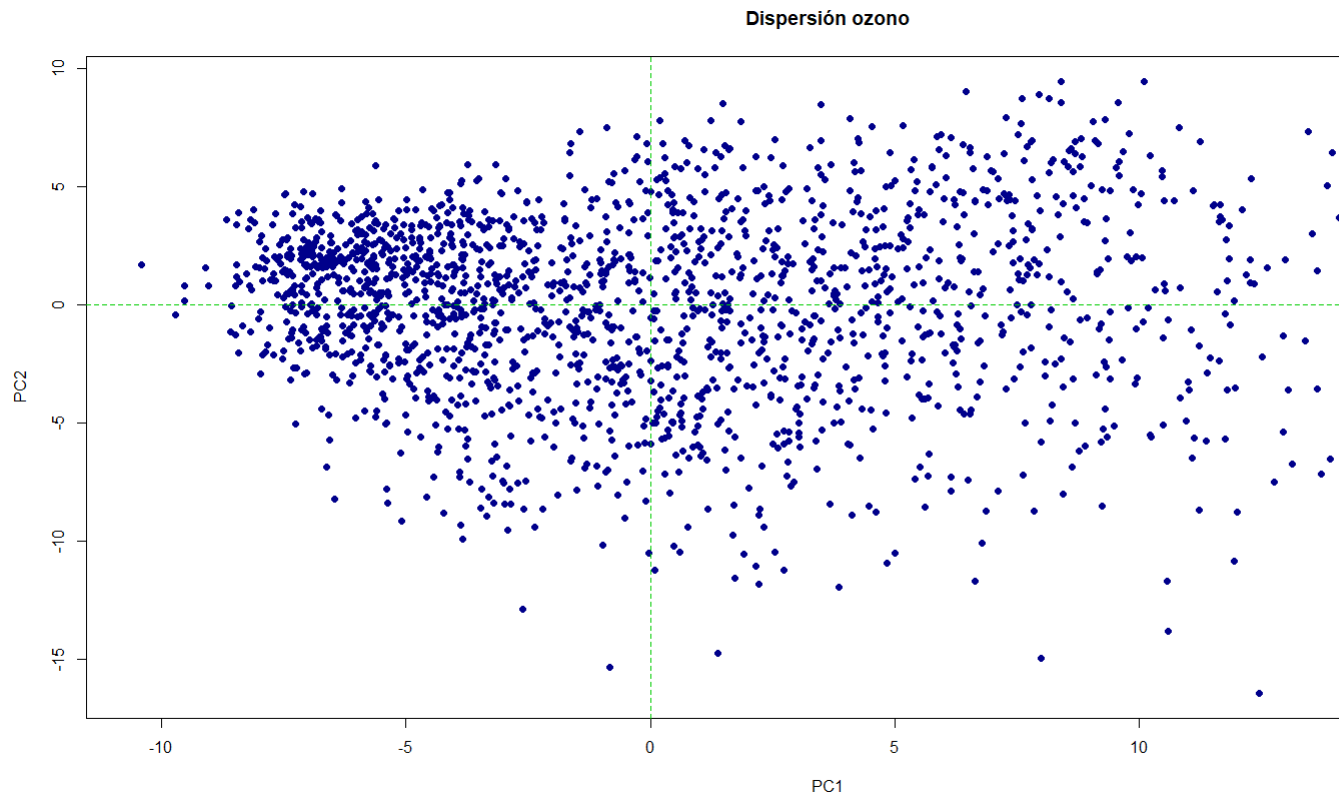


Figure 5: Gráfico de correlación sin clase

Tabla de componentes principales para un dataset con Contaminacion

A continuación se muestra la tabla del analisis de los componentes principales,

```
> summary(acp3)
Importance of components:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10     PC11     PC12
Standard deviation 5.3989 3.3011 2.38123 1.92161 1.7636 1.66621 1.49005 1.28601 1.16828 1.15038 1.06576 0.98347
Proportion of Variance 0.4048 0.1514 0.07875 0.05129 0.0432 0.03856 0.03084 0.02297 0.01896 0.01838 0.01578 0.01343
Cumulative Proportion 0.4048 0.5562 0.63495 0.68624 0.7294 0.76799 0.79883 0.82180 0.84075 0.85913 0.87491 0.88834
      PC13      PC14      PC15      PC16      PC17      PC18      PC19      PC20      PC21      PC22      PC23      PC24
Standard deviation 0.87387 0.83758 0.8004 0.7869 0.74879 0.69304 0.63195 0.59252 0.57693 0.56183 0.51733 0.48941
Proportion of Variance 0.01061 0.00974 0.0089 0.0086 0.00779 0.00667 0.00555 0.00488 0.00462 0.00438 0.00372 0.00333
Cumulative Proportion 0.89895 0.90869 0.9176 0.9262 0.93398 0.94065 0.94620 0.95107 0.95570 0.96008 0.96380 0.96712
      PC25      PC26      PC27      PC28      PC29      PC30      PC31      PC32      PC33      PC34      PC35      PC36
Standard deviation 0.4648 0.45912 0.43040 0.41671 0.41248 0.39171 0.36544 0.35419 0.33298 0.3176 0.31054 0.29107
Proportion of Variance 0.0030 0.00293 0.00257 0.00241 0.00236 0.00213 0.00185 0.00174 0.00154 0.0014 0.00134 0.00118
Cumulative Proportion 0.9701 0.97305 0.97562 0.97804 0.98040 0.98253 0.98439 0.98613 0.98767 0.9891 0.99041 0.99158
      PC37      PC38      PC39      PC40      PC41      PC42      PC43      PC44      PC45      PC46      PC47
Standard deviation 0.26104 0.26078 0.24186 0.23455 0.22722 0.21884 0.20884 0.19416 0.17937 0.16171 0.15199
Proportion of Variance 0.00095 0.00094 0.00081 0.00076 0.00072 0.00067 0.00061 0.00052 0.00045 0.00036 0.00032
Cumulative Proportion 0.99253 0.99348 0.99429 0.99505 0.99577 0.99643 0.99704 0.99756 0.99801 0.99837 0.99869
      PC48      PC49      PC50      PC51      PC52      PC53      PC54      PC55      PC56      PC57      PC58
Standard deviation 0.13957 0.11511 0.10840 0.10397 0.07879 0.07094 0.06630 0.06197 0.05395 0.05063 0.04532
Proportion of Variance 0.00027 0.00018 0.00016 0.00015 0.00009 0.00007 0.00006 0.00005 0.00004 0.00004 0.00003
Cumulative Proportion 0.99897 0.99915 0.99931 0.99946 0.99955 0.99962 0.99968 0.99973 0.99977 0.99981 0.99984
      PC59      PC60      PC61      PC62      PC63      PC64      PC65      PC66      PC67      PC68      PC69      PC70
Standard deviation 0.04503 0.04003 0.03720 0.03515 0.03259 0.03108 0.02778 0.02536 0.02410 0.02067 0.01892 0.018
Proportion of Variance 0.00003 0.00002 0.00002 0.00002 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001 0.00000 0.000
Cumulative Proportion 0.99987 0.99989 0.99991 0.99992 0.99994 0.99995 0.99996 0.99997 0.99998 0.99999 0.99999 1.000
      PC71      PC72
Standard deviation 0.01702 0.005154
Proportion of Variance 0.00000 0.000000
Cumulative Proportion 1.00000 1.000000
> |
```

Figure 7: Tabla de componentes principales

Gráfico de Sedimentacion para determinar CP

Este grafico nos permitira determinar que CP debemos seleccionar de acuerdo a la significancia del mismo, dicha relevancia se deducira en base a la tabla anterior buscando un gran porcentaje de proporcion acumulada (0.72 hasta la CP 5)

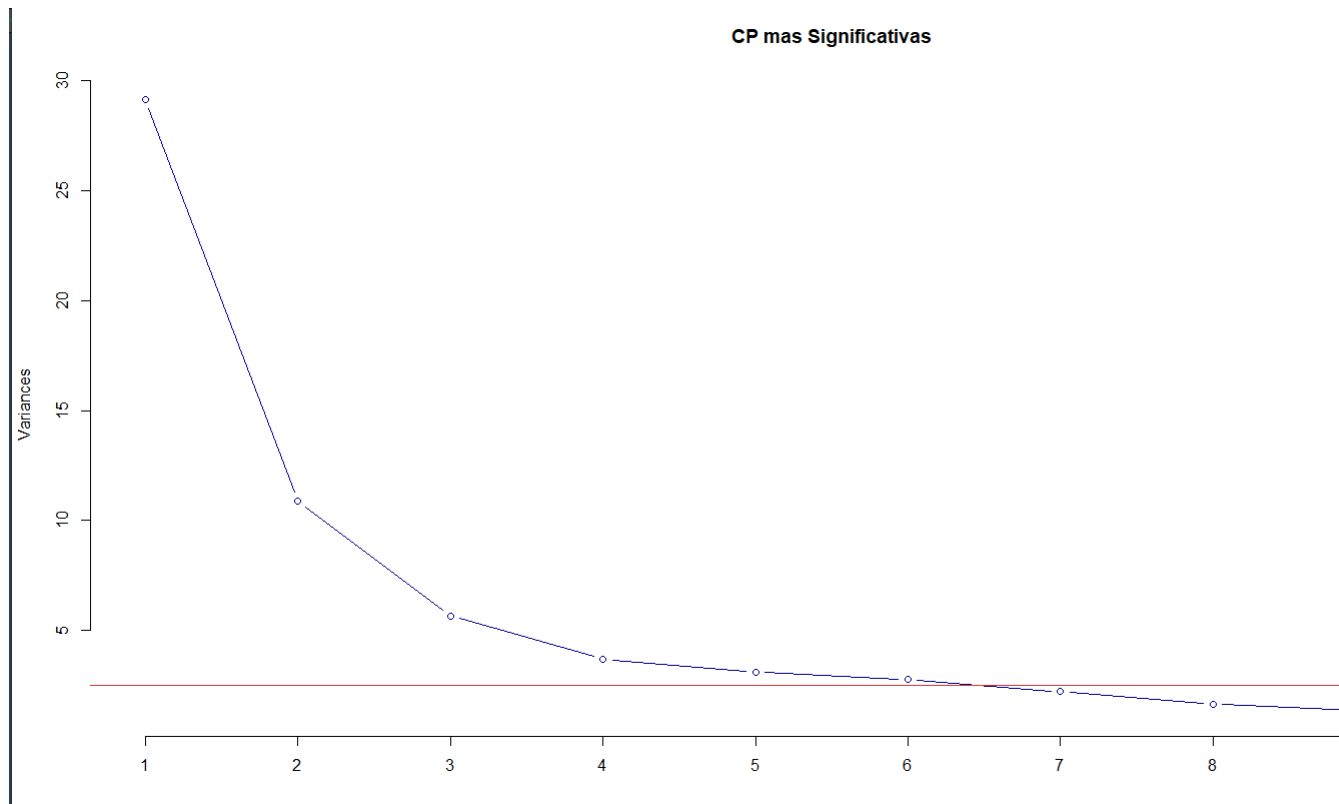


Figure 8: Sedimentacion Ozono

Gráfico de Correlación con Contaminacion

A continuación se muestra el gráfico de correlación con Contaminacion.

CP1-CP2

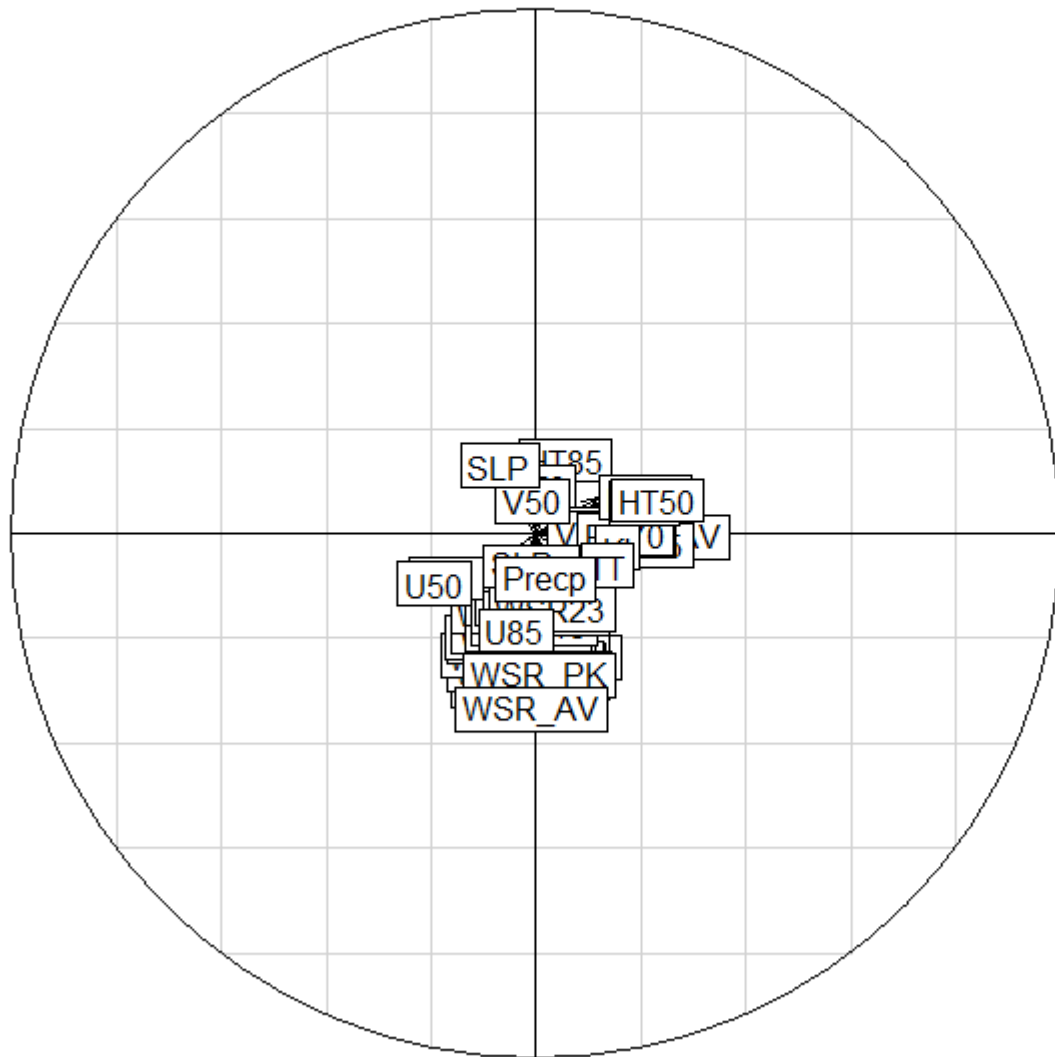


Figure 9: Gráfico de correlación sin clase

Gráfico de Dispersión con Contaminación

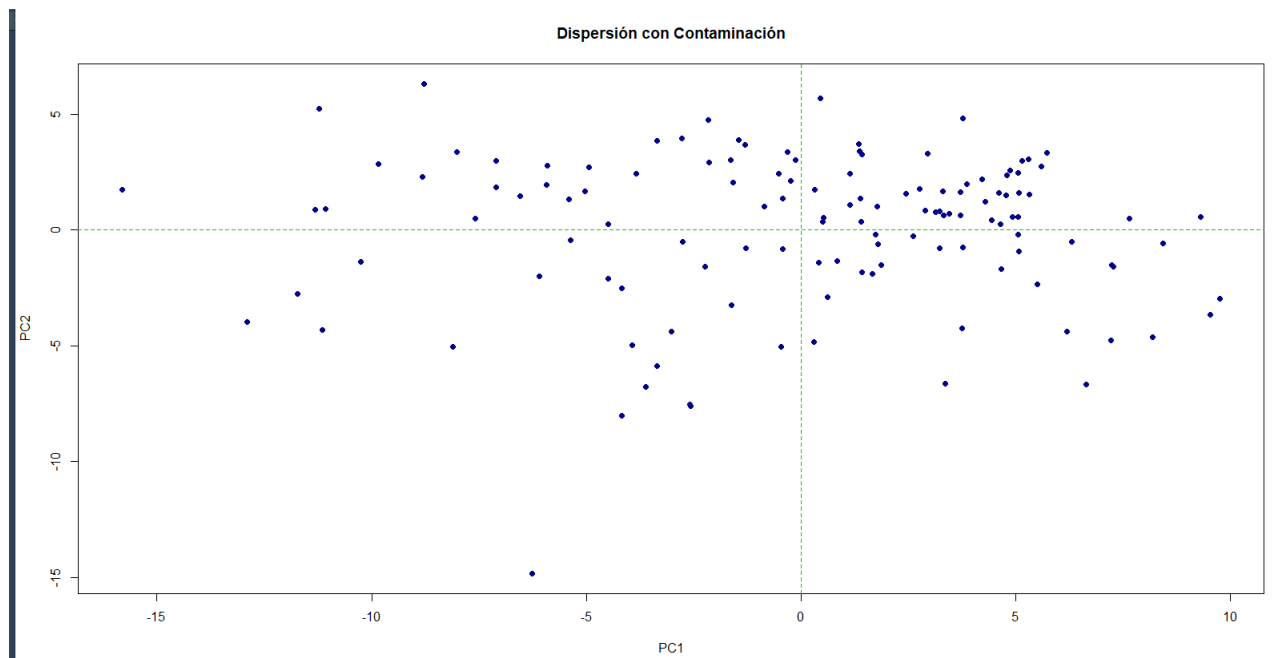


Figure 10: Gráfico de Dispersión que muestra la contaminación

Este Grafico fue realizado en Infostat con el proposito de mostrar una mejor perspectiva en la visualizacion de CP



Este gráfico muestra la relación entre las temperaturas y el viento, donde a mayor temperatura mayor contaminación, o se puede decir que la contaminación es proporcional a la temperatura, marcando sobre esta cierta incidencia muy notoria. Para este caso, se toma la relación entre los componentes 1 y 2.

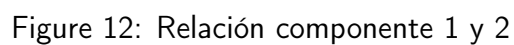


Tabla de componentes principales para un dataset Sin Contaminacion

A continuación se muestra la tabla del analisis de los componentes principales,

```
> summary(acp4)
```

Importance of components:

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 | PC11 | PC12 |
|------------------------|---------|---------|---------|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| Standard deviation | 5.4689 | 3.9199 | 2.35369 | 1.79833 | 1.62226 | 1.39005 | 1.27249 | 1.07364 | 1.0358 | 0.97333 | 0.93189 | 0.88780 |
| Proportion of Variance | 0.4154 | 0.2134 | 0.07694 | 0.04492 | 0.03655 | 0.02684 | 0.02249 | 0.01601 | 0.0149 | 0.01316 | 0.01206 | 0.01095 |
| Cumulative Proportion | 0.4154 | 0.6288 | 0.70575 | 0.75067 | 0.78722 | 0.81405 | 0.83654 | 0.85255 | 0.8675 | 0.88061 | 0.89267 | 0.90362 |
| | PC13 | PC14 | PC15 | PC16 | PC17 | PC18 | PC19 | PC20 | PC21 | PC22 | PC23 | |
| Standard deviation | 0.82792 | 0.79971 | 0.75556 | 0.72584 | 0.65528 | 0.59807 | 0.55882 | 0.53549 | 0.51205 | 0.48797 | 0.48069 | |
| Proportion of Variance | 0.00952 | 0.00888 | 0.00793 | 0.00732 | 0.00596 | 0.00497 | 0.00434 | 0.00398 | 0.00364 | 0.00331 | 0.00321 | |
| Cumulative Proportion | 0.91314 | 0.92202 | 0.92995 | 0.93727 | 0.94323 | 0.94820 | 0.95254 | 0.95652 | 0.96016 | 0.96347 | 0.96668 | |
| | PC24 | PC25 | PC26 | PC27 | PC28 | PC29 | PC30 | PC31 | PC32 | PC33 | PC34 | |
| Standard deviation | 0.45860 | 0.43103 | 0.41320 | 0.39506 | 0.37732 | 0.35787 | 0.33846 | 0.33429 | 0.32133 | 0.30505 | 0.29979 | |
| Proportion of Variance | 0.00292 | 0.00258 | 0.00237 | 0.00217 | 0.00198 | 0.00178 | 0.00159 | 0.00155 | 0.00143 | 0.00129 | 0.00125 | |
| Cumulative Proportion | 0.96960 | 0.97218 | 0.97455 | 0.97672 | 0.97870 | 0.98047 | 0.98207 | 0.98362 | 0.98505 | 0.98634 | 0.98759 | |
| | PC35 | PC36 | PC37 | PC38 | PC39 | PC40 | PC41 | PC42 | PC43 | PC44 | PC45 | PC46 |
| Standard deviation | 0.29312 | 0.27988 | 0.27016 | 0.26450 | 0.26028 | 0.25133 | 0.24436 | 0.23502 | 0.2248 | 0.21899 | 0.20983 | 0.20242 |
| Proportion of Variance | 0.00119 | 0.00109 | 0.00101 | 0.00097 | 0.00094 | 0.00088 | 0.00083 | 0.00077 | 0.0007 | 0.00067 | 0.00061 | 0.00057 |
| Cumulative Proportion | 0.98879 | 0.98987 | 0.99089 | 0.99186 | 0.99280 | 0.99368 | 0.99451 | 0.99527 | 0.9960 | 0.99664 | 0.99725 | 0.99782 |
| | PC47 | PC48 | PC49 | PC50 | PC51 | PC52 | PC53 | PC54 | PC55 | PC56 | PC57 | |
| Standard deviation | 0.17975 | 0.14351 | 0.12916 | 0.12350 | 0.11750 | 0.09696 | 0.09562 | 0.08106 | 0.07433 | 0.06454 | 0.06249 | |
| Proportion of Variance | 0.00045 | 0.00029 | 0.00023 | 0.00021 | 0.00019 | 0.00013 | 0.00013 | 0.00009 | 0.00008 | 0.00006 | 0.00005 | |
| Cumulative Proportion | 0.99827 | 0.99856 | 0.99879 | 0.99900 | 0.99919 | 0.99932 | 0.99945 | 0.99954 | 0.99962 | 0.99968 | 0.99973 | |
| | PC58 | PC59 | PC60 | PC61 | PC62 | PC63 | PC64 | PC65 | PC66 | PC67 | PC68 | |
| Standard deviation | 0.05635 | 0.05095 | 0.04827 | 0.04364 | 0.04266 | 0.03720 | 0.03404 | 0.03330 | 0.02993 | 0.02834 | 0.02739 | |
| Proportion of Variance | 0.00004 | 0.00004 | 0.00003 | 0.00003 | 0.00003 | 0.00002 | 0.00002 | 0.00002 | 0.00001 | 0.00001 | 0.00001 | |
| Cumulative Proportion | 0.99977 | 0.99981 | 0.99984 | 0.99987 | 0.99989 | 0.99991 | 0.99993 | 0.99995 | 0.99996 | 0.99997 | 0.99998 | |
| | PC69 | PC70 | PC71 | PC72 | | | | | | | | |
| Standard deviation | 0.02437 | 0.02205 | 0.01960 | 0.004215 | | | | | | | | |
| Proportion of Variance | 0.00001 | 0.00001 | 0.00001 | 0.000000 | | | | | | | | |
| Cumulative Proportion | 0.99999 | 0.99999 | 1.00000 | 1.000000 | | | | | | | | |

Figure 13: Tabla de componentes principales

Gráfico de Sedimentacion para determinar CP

Este grafico nos permitira determinar que CP debemos seleccionar de acuerdo a la significancia del mismo, dicha relevancia se deducira en base a la tabla anterior buscando un gran porcentaje de proporcion acumulada (0.78 hasta la CP 5)

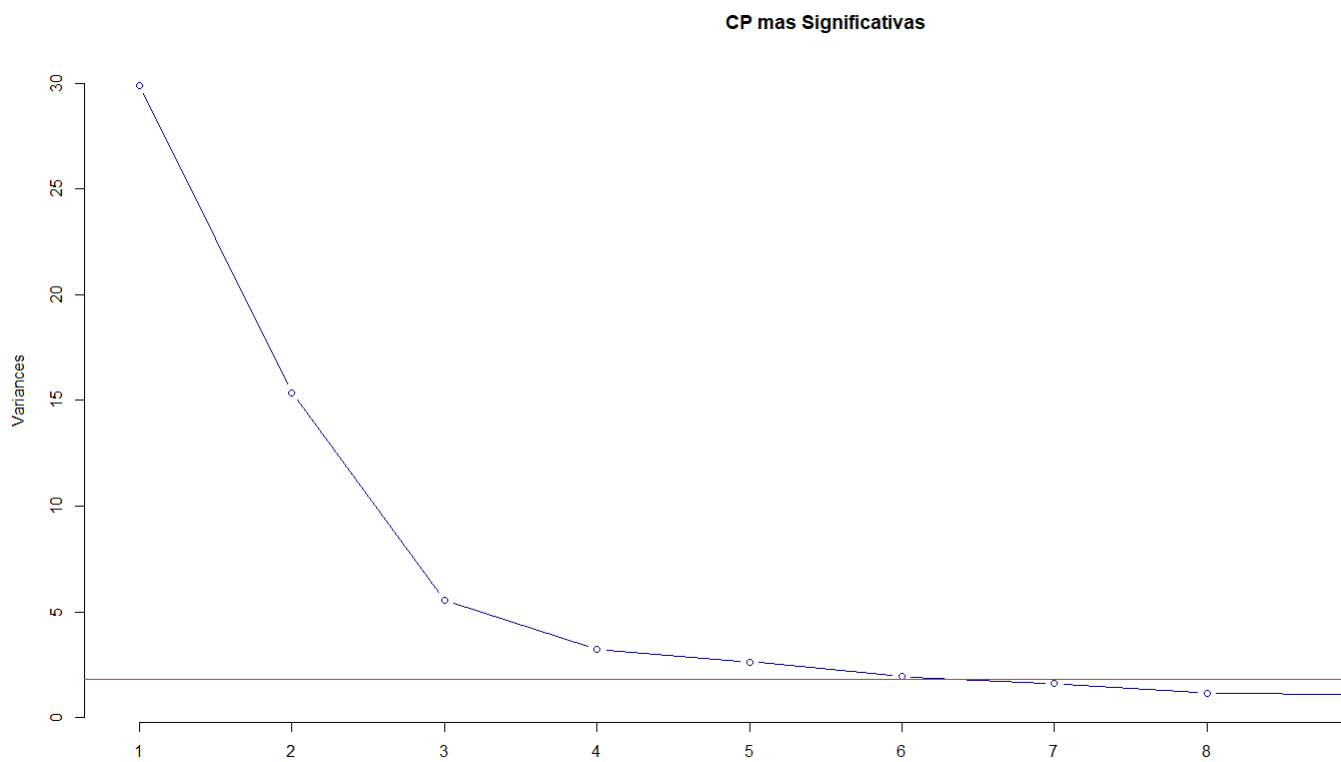


Figure 14: Sedimentacion Ozono

Círculo de Correlación sin contaminación

CP1-CP2

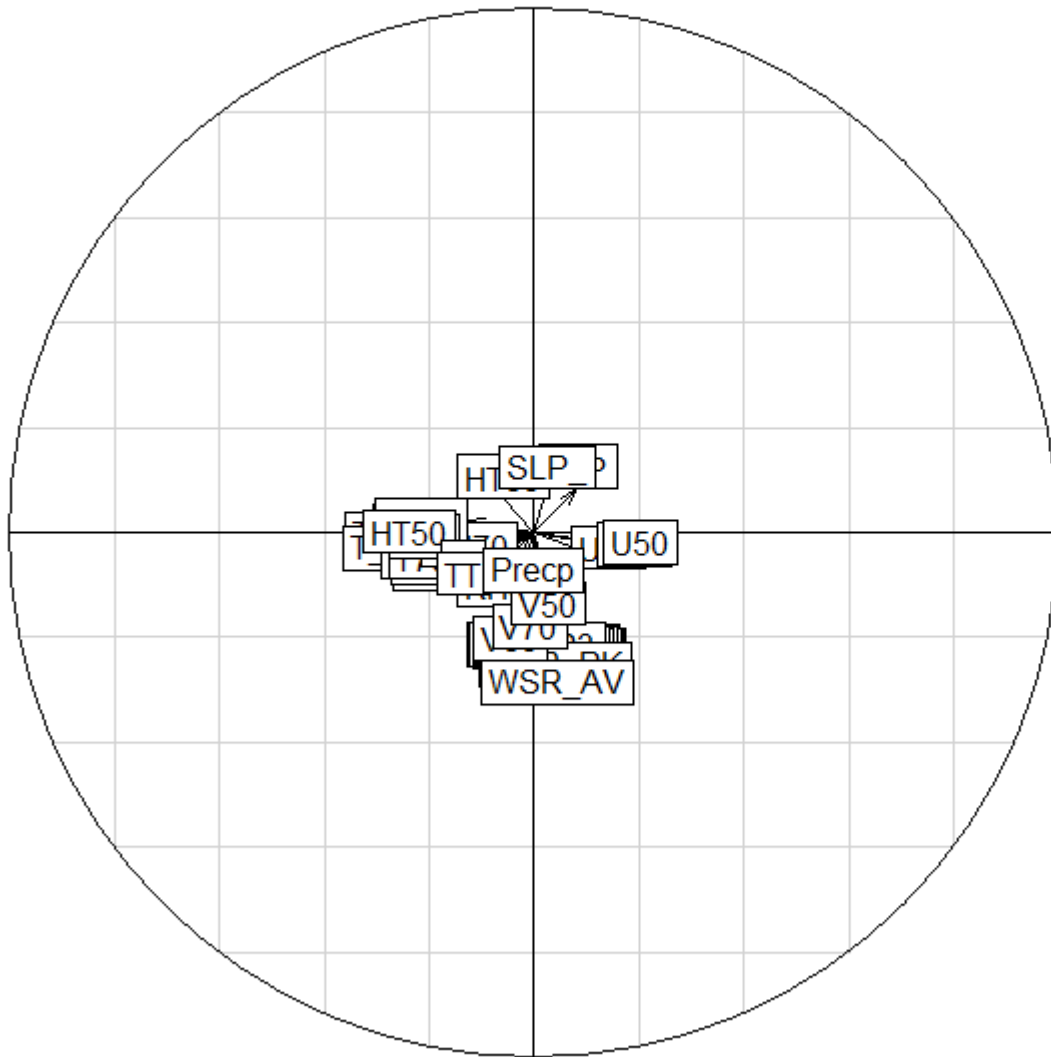


Figure 15: Círculo de Correlación sin contaminación

Gráfico de Dispersión sin Contaminación

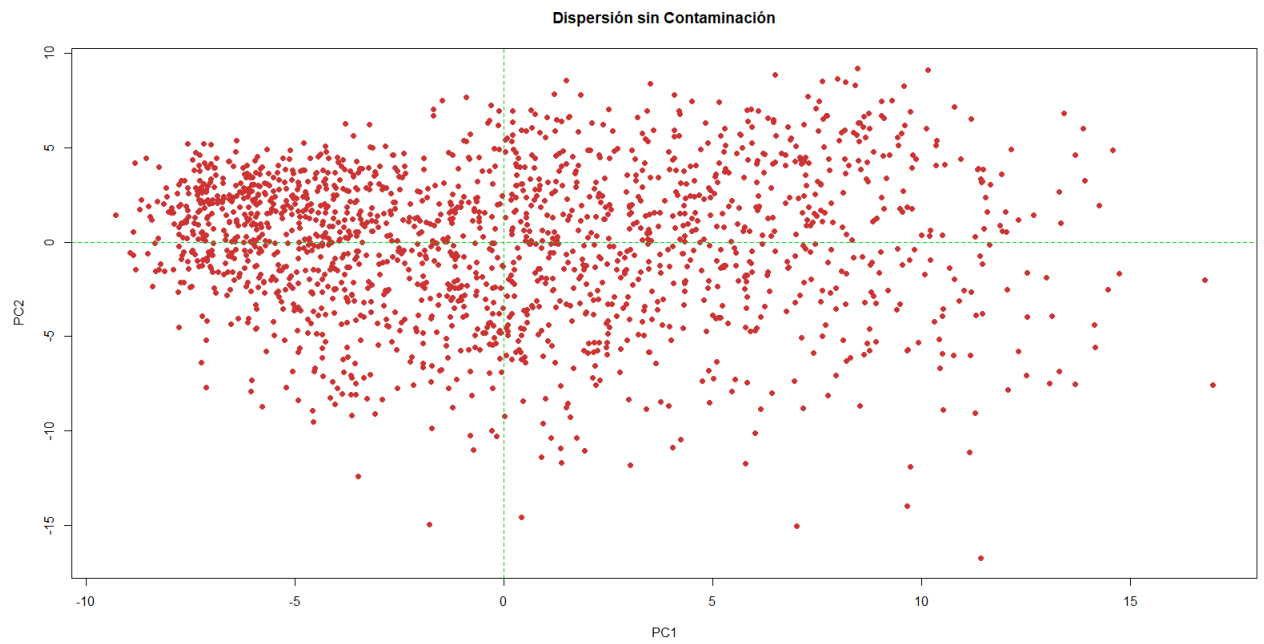


Figure 16: Gráfico de Dispersión sin contaminación

Gráfico de relación de componentes 1 y 2

Este gráfico muestra la relación entre las temperaturas y el viento (en mayor medida de exploración), donde se puede decir que la contaminación marca cierta incidencia desde la densidad de vientos. Para este caso, se toma la relación entre los componentes 1 y 2.

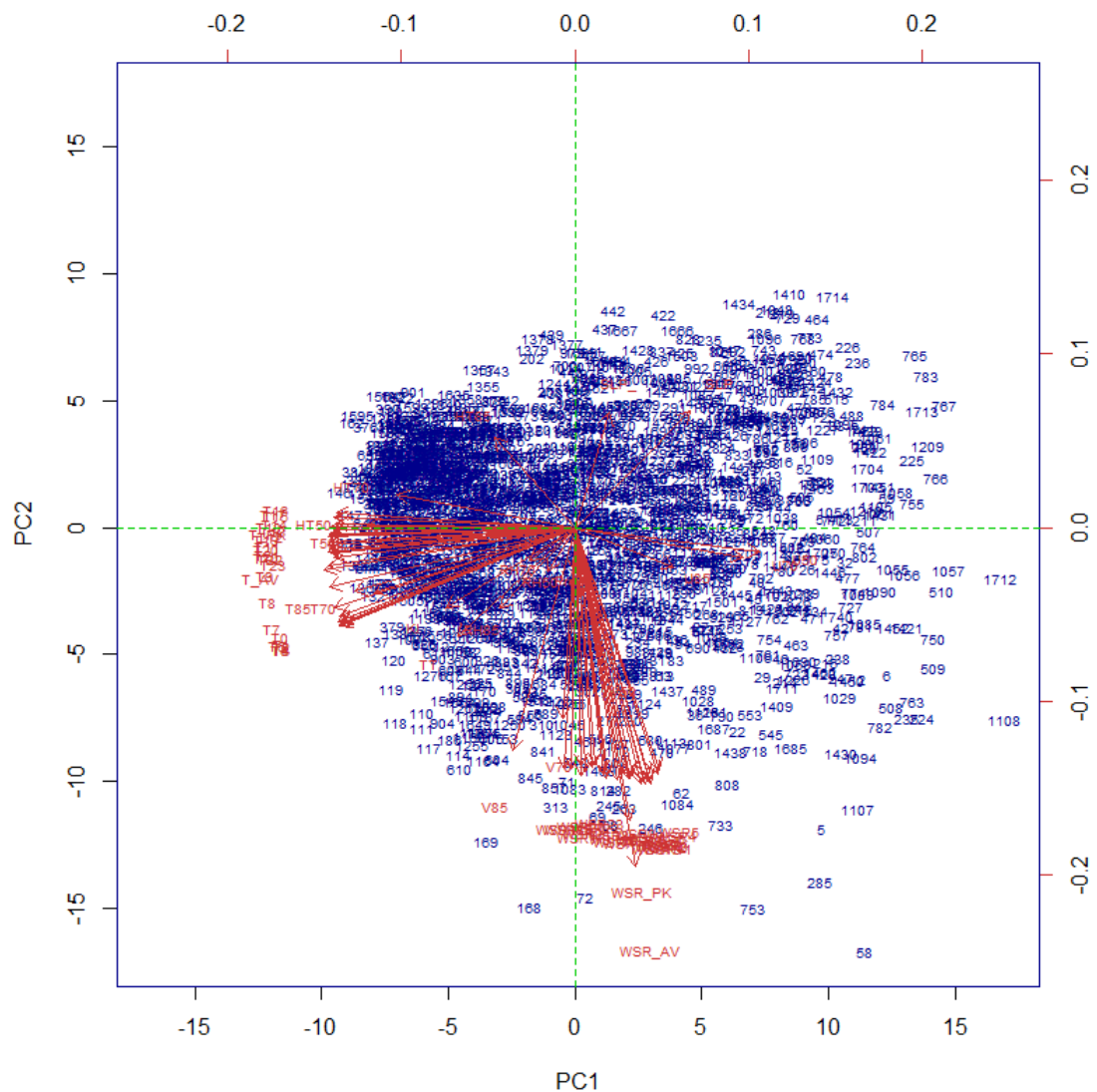


Figure 18: Relación componente 1 y 2

Conclusión

Luego de desarrollar el análisis, y haciendo énfasis en el uso de herramientas estadísticas, concluimos que el Análisis de Componentes Principales nos ayuda a reaalizar una mejor visualización y representación del problema ante un conjunto de datos de grandes dimensiones, pudiendo definir cuáles son las variables que tienen una influencia más considerable y descartando aquellas que no llegan a influir demasiado o incluso entorpezcan el análisis.

Al realizar los distintos gráficos pudimos inferir que las variables más significativas en este estudio fueron la temperatura y el viento .

Anexo

Código del Primer Gráfico - Sin tipo de Clase

```
1 library(ade4)
2 library(dplyr)
3 library(readr)
4 library(readxl)
5 library(knitr)
6 library(xtable)
7 library(outliers)
8 library(openxlsx)
9
10 #Exporto el archivo en xls para categorizar datos y usarlos en Infostat
11 write.xlsx(ozon2, 'dfiltrado.xlsx')
12
13 ozon1 <- read_excel("C:/Users/brian/Documents/ozo/ozono.xls", col_types = c("numeric"))
14 #Debug de datos
15 View(ozon1)
16 names(ozon1)
17 length(ozon1$WSR0)
18 sapply(ozon1, function(x) sum(is.na(x)))
19
20 #Creo una nueva DF desde los atributos 2 a 74
21 ozon2 = ozon1[,2:73]
22 #Visualizo los datos
23 View(ozon2)
24 names(ozon2)
25 sapply(ozon2, function(x) sum(is.na(x)))
26
27 #Funcion "delete.na"
28 delete.na <- function(df, n=0) {
29   df[rowSums(is.na(df)) <= n,]
30 }
31
32 #Elimino los datos
33 ozon2 <- delete.na(ozon2)
34 View(ozon2)
35 names(ozon2)
36 length(ozon2$WSR0)
37 sapply(ozon2, function(x) sum(is.na(x)))
38
39 #Funcion prcomp: Por defecto, prcomp() centra las variables para que tengan media cero,
40 #pero si se quiere adem s que su desviaci n est ndar sea de uno, hay que indicar:
41   scale = TRUE.
```

```

41
42 #CP
43 acp2 = prcomp(ozon2, scale=TRUE)
44
45 #Se puede ver la proporcion de varianza acumulada
46 #Set de todas las CP
47 summary(acp2)
48
49 #Grafico de Sedimentacion de todas las CP
50 plot(acp2, type="l", main="CP mas Significativas", col=c("blue4"))
51 abline(1.7,0,col=c("brown3"))
52
53 #####Selecciono los Datos#####
54 aCP1 = acp2[[2]][,1]
55 aCP2 = acp2[[2]][,2]
56 aCP3 = acp2[[2]][,3]
57 aCP4 = acp2[[2]][,4]
58 aCP5 = acp2[[2]][,5]
59
60 acp = cbind(aCP1,aCP2,aCP3,aCP4,aCP5)
61 #Proporcion peso de variables para cada CP
62 acp
63
64 #Biplot para todos los CP!!!!
65 x11(15)
66 biplot(x=acp2, scale=0, cex=0.6, col=c("blue4", "brown3"))
67 abline(h = 0, v = 0, lty = 2, col = 3)
68
69 indiv1=acp2$x[,1:5]
70 indiv1
71
72 x11()
73 plot(indiv1[,1:5], pch = 19, main="Dispersi n ozono",col=c("blue4"))
74 abline(h = 0, v = 0, lty = 2, col = 3)
75
76 x11(15)
77 s.corcircle(acp[,1:2], sub = "CP1-CP2", possub = "topright")
78
79 #Ver datos correlacionados
80 cor(ozon2)

```

Código del Segundo Gráfico - Con Contaminación

```

1 library(ade4)
2 library(dplyr)
3 library(readr)
4 library(readxl)
5 library(knitr)
6 library(xtable)
7 library(outliers)
8
9 #Categorizo los datos para llevarlos a infostat
10 write.xlsx(ozon3, 'cpclase1.xlsx')
11
12 ozon1 <- read_excel("C:/Users/brian/Documents/ozo/ozono.xls", col_types = c("numeric"))
13 names(ozon1)

```

```
14 str(ozon1)
15 View(ozon1)
16 sapply(ozon1, function(x) sum(is.na(x)))
17 length(ozon1$WSR0)
18
19 #Creo una nueva DF desde los atributos 2 a 74
20 ozon3 = ozon1[,2:74]
21 #Visualizo los datos
22 View(ozon3)
23 names(ozon3)
24 sapply(ozon3, function(x) sum(is.na(x)))
25
26 #Funcion "delete.na"
27 delete.na <- function(df, n=0) {
28   df[rowSums(is.na(df)) <= n,]
29 }
30
31 #Elimino los datos
32 ozon3 <- delete.na(ozon3)
33 View(ozon3)
34 names(ozon3)
35 length(ozon3$WSR0)
36 sapply(ozon3, function(x) sum(is.na(x)))
37
38 ozon3 <- subset(ozon3, clase == 1)
39 str(ozon3)
40 names(ozon3)
41 length(ozon3$WSR0)
42 View(ozon3)
43
44 #DF solo con Clase de tipo 1
45 ozon3 <- select(ozon3, -clase) #Elimino la columna clase para continuar con el ACP.
46 str(ozon3)
47 names(ozon3)
48 length(ozon3$WSR0)
49 View(ozon3)
50
51 acp3 <- prcomp(ozon3, scale = TRUE)
52 summary(acp3)
53
54 ##grafico de sedimentacion.
55 plot(acp3, type="l", main="CP mas Significativas", col=c("blue4"))
56 abline(2.5,0, col=c("brown3"))
57
58 #Sedimentacion y barras
59 plot(acp3)
60 screeplot(acp3)
61 screeplot(acp3, type="lines")
62
63 #Biplot de todos los datos
64 x11(15)
65 biplot(x=acp3, scale=0, cex=0.6, col=c("blue4", "brown3"))
66 abline(h = 0, v = 0, lty = 2, col = 3)
67
68 CP1 = acp3[[2]][,1]
```

```
69 CP2 = acp3[[2]][,2]
70 CP3 = acp3[[2]][,3]
71 CP4 = acp3[[2]][,4]
72 CP5 = acp3[[2]][,5]
73
74 cpp1 = cbind(CP1,CP2,CP3,CP4,CP5)
75 #Proporcion peso de variables para cada CP
76 cpp1
77 print(cpp1)
78
79 ind=acp3$x[,1:5]##Guardo los individuos, que seria el valor x. y solo los que estoy
    trabajando, en este caso del 1 al 4
80 ind
81
82 plot(ind[,1:5], pch = 19, main="Dispersin con Contaminacin",col=c("blue4"))
83 abline(h = 0, v = 0, lty = 2, col = 3)
84
85 x11(15)
86 s.corcircle(cpp1[,1:2], sub = "CP1-CP2", possub = "topright")
87
88 #Ver datos correlacionados
89 cor(ozon3)
90 summary(acp3)
```

Código del Tercer Gráfico - Sin Contaminación

```
1 library(ade4)
2 library(dplyr)
3 library(readr)
4 library(readxl)
5 library(knitr)
6 library(xtable)
7 library(outliers)
8
9 #Categorizar datos para infostat
10 write.xlsx(ozon4, 'cpclase0.xlsx')
11
12 ozon1 <- read_excel("C:/Users/brian/Documents/ozo/ozono.xls", col_types = c("numeric"))
13 #Verificamos la existencia y tipo de datos importados
14 View(ozon1)
15 names(ozon1)
16 sapply(ozon1, function(x) sum(is.na(x)))
17
18 #Creo una nueva DF desde los atributos 2 a 74
19 ozon4 = ozon1[,2:74]
20 #Visualizo los datos
21 View(ozon4)
22 names(ozon4)
23 sapply(ozon4, function(x) sum(is.na(x)))
24
25 #Funcion "delete.na"
26 delete.na <- function(df, n=0) {
27   df[rowSums(is.na(df)) <= n,]
28 }
29
30 #Elimino los datos
```

```
31 ozon4 <- delete.na(ozon4)
32 View(ozon4)
33 names(ozon4)
34 sapply(ozon4, function(x) sum(is.na(x)))
35
36 #DF solo con Clase de tipo 0
37 ozon4<- ozon4[ozon4$clase == 0, ]
38
39 str(ozon4)
40 length(ozon4$WSR0)
41 View(ozon4)
42 names(ozon4)
43
44 ozon4 <- select(ozon4, -clase)
45 str(ozon4)
46 length(ozon4$WSR0)
47 View(ozon4)
48 names(ozon4)
49
50 acp4 = prcomp(ozon4, scale=TRUE)
51 summary(acp4)
52 #####Muestreo de datos de CP generales---Matematicamente
53
54 ##grafico de sedimentacion.
55 plot(acp4, type="l", main="CP mas Significativas",col=c("blue4"))
56 abline(1.8,0,col=c("brown3"))
57
58 x11(15)
59 biplot(x=acp4, scale=0, cex=0.6, col=c("blue4", "brown3"))
60 abline(h = 0, v = 0, lty = 2, col = 3)
61
62 CP41 = acp4[[2]][,1]
63 CP42 = acp4[[2]][,2]
64 CP43 = acp4[[2]][,3]
65 CP44 = acp4[[2]][,4]
66 CP45 = acp4[[2]][,5]
67
68 cpp4 = cbind(CP41,CP42,CP43,CP44,CP45)
69 cpp4
70
71 ind4=acp4$x[,1:5]##Guardo los individuos, que seria el valor x. y solo los que estoy
    trabajando, en este caso del 1 al 4
72 ind4
73
74 plot(ind4[,1:5], pch = 19, main="Dispersin sin Contaminacin", col=c("brown3"))
75 abline(h = 0, v = 0, lty = 2, col = 3)
76
77 x11(15)
78 s.corcircle(cpp4[,1:2], sub = "CP1-CP2", possub = "topright")
```