

# 1. Análisis de ordenación

| Introducción  | <br>1  |
|---|--------|
| 1.0 Análisis de componentes principales (ACP o PCA)       | <br>2  |
| 1.1 Elementos principales de un ACP                       |        |
| 1.2 Ejercicios propuestos                                 | <br>11 |
| 2. Análisis de Factores (AF)                              | <br>12 |
| 2.1 Ejercicios propuestos                                 |        |
| 3. Análisis de escalamiento multidimensional (MDS y nMDS) |        |
| 3.1 Ejercicios propuestos                                 | <br>12 |
| 4. Análisis de correspondencia (AC o CA)                  |        |
| 4.1. Ejercicios propuestos                                | <br>12 |
| 5. Análisis de correspondencia dirigido (ACD o DCA)       | <br>12 |
| 5.1 Análisis canónicos                                    | <br>12 |
| 7.2 Ejercicios propuestos                                 | <br>12 |
|   |        |

#### Introducción

La ordenación busca ubicar a un conjunto de ítems en una escala (ejes) o múltiples ejes. El propósito de los ejes consiste en resumir relaciones complejas multivariadas. Se busca extraer uno o pocos patrones dominantes de un número infinito de patrones. El proceso de extraer estos ejes se denomina "ordenación", ubicando a las observaciones en planos cartesianos. El nivel de correlación entre las variables es lo que hace posible la ordenación de las observaciones en gradientes (sitios, visitas, individuos, etc.).

En el análisis directo de gradientes, las observaciones son posicionadas (ordenadas) de acuerdo a medidas de variables ambientales, se suele usar este análisis para entender como las especies se distribuyen a lo largo de gradientes de interés. En el análisis indirecto de gradientes, la posición de las observaciones se da de acuerdo a la correlación o covarianza entre las variables.

Tabla. Aplicación de los métodos de ordenación comparación indirecta, más utilizados (tomado de Legendre & Legendre, 1998).

| Método   | Distancia de ordenación     | Variables de estudio  |
|--|-----------------------------|---|
| 1. Análisis de componentes principales (PCA)   | Distancia euclídea          | Cuantitativas, relacionadas linealmente sin doble ceros             |
| 2. Análisis Factorial (AF)   | Distancia euclídea          | Cuantitativas, relacionadas linealmente sin doble ceros             |
| Escalamiento multidimensional métrico o Análisis de coordenadas principales (MDS = PCoA) | Cualquier distancia métrica | Cuantitativas, semicuantitativas, cualitativas o mixtas             |
| 4. Escalamiento multidimensional no métrico (nMDS)                                       | Cualquier distancia         | Cuantitativas, semicuantitativas, cualitativas o mixtas             |
| 5. Análisis de correspondencia (CA)  | Distancia Chi-Cuadrado      | Cuantitativas, binarias, no negativas y dimensionalmente homogéneas |
| 6. Análisis de Redundancia (RDA)   | Distancia Euclídea          | Cuantitativas, relacionadas linealmente sin doble ceros             |
| 7. Análisis de Correspondencia dirigido (DCA) y correspondencia canónica (CCA)           | Distancia Chi-Cuadrado      | Cuantitativas, binarias, no negativas y dimensionalmente homogéneas |





# 1. Análisis de componentes principales (ACP o PCA)

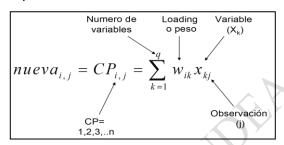
Es la forma más básica de ordenar a los datos. Fue desarrollado por Pearson en 1901. La utilidad principal del ACP, consiste en reducir la multidimensionalidad de datos multivariados, creando pocas variables sintéticas (componentes), que caractericen de la mejor forma posible la variación de datos multivariados (observaciones o variables). El atributo geométrico más importante del ACP es que estas nuevas variables no muestran correlación entre ellas (son ortogonales), pero muestran la mayor correlación con los datos originales.

En este sentido, el ACP es una técnica que permite clasificar o formar grupos de observaciones a partir de múltiples variables, realizando únicamente trasformaciones lineales de los datos originales (únicamente rota la "nube" de puntos u observaciones). En el PCA la el análisis de relación entre los datos se realiza únicamente de forma lineal (por medio de la distancia euclídea). Para mejorar la ordenación de conjunto de datos no lineales (que se presentan en arco), se utilizan métodos como el DCA (análisis de correspondencia sin tendencias), que trabaja con la distancia Chi-Cuadrado.

### 1.1 Elementos principales de un ACP

Al realizar un ACP, se generan cuatro elementos básicos, descritos a continuación:

(1) los "componentes principales" corresponden a los nuevos ejes generados a partir del nivel de relación entre los objetos u observaciones, la ubicación de ellos en el espacio multivariado, intenta representar la máxima correlación con los objetos pero la mínima correlación entre ellos.



Donde, q representa a las nuevas dimensiones de las variables, k varía de acuerdo a los valores de las variables, w es la carga o el peso de las variables (autovector), x es cada una de las variables e i es cada componente generado.

El procedimiento general, para hallar las componentes principales consiste en determinar el producto entre matrices de covarianza o correlación (centradas y estandarizadas), con matrices de autovectores (figura).



Figura 1. Procedimiento resumido para hallar las componentes principales a partir de una tabla de datos de campo o matriz de datos crudos (tomado de Herrándo, 2005).

(2) Los "autovalores" (eigenvalues " $\lambda$ "), tienen la propiedad de dar información de la matriz de datos crudos mediante un único valor. Su función consiste en informar sobre el nivel de ordenación o ajuste (variación explicada) que se obtiene con cada componente principal del análisis.



- (3) Los "autovectores" (eigenvectors " $\mu$ "), son los que determinan los pesos, las cargas o el nivel de relación de las variables (vectores) con cada componente principal generado. En este sentido cada autovalor tendrá un peso que varía entre los componentes en que se encuentre.
- (4) Los "scores", corresponden a las cargas o a la posición de los objetos (ej. sitios) en los componentes principales que han sido generados. Por ejemplo, si en un ACP se traba con 5 variables, podrá haber hasta 5 componentes principales y cada observación tendrá una ubicación con 5 dimensiones o coordenadas.

# **Ejemplo 1. ACP CON VARIABLES AMBIENTALES**

Se tienen 7 localidades en playas de la ciudad de Santa Marta (observaciones) y en cada una de ellas se midió a 7 variables ambientales (descriptores), se intenta saber cuales variables ambientales estructuran o caracterizan a las localidades estudiadas. Los archivos que se trabajan son: **FQmarino.csv** y **Comprin1.r.** 

Al realizar el ACP, los pasos a seguir son: (1) llamar la base de datos "datos", mediante el siguiente comando en R:

datos<-read.csv2("FQmarino.csv",row.names=1)
datos</pre>

|              |       |        |          |        | /         |            |         |
|--------------|-------|--------|----------|--------|-----------|------------|---------|
|              | рН    | Cond   | Turbidez | Temp   | Salinidad | CapaFotica | Oxigeno |
| BahíaTaganga | 8.421 | 37.982 | 1.364    | 29.5   | 2.422     | 19.72      | 0.097   |
| PuntaBetín   | 8.49  | 38.073 | 0.545    | 29.545 | 2.431     | 22.1       | 0.147   |
| Monohuaca    | 8.505 | 37.836 | 1.273    | 29.6   | 2.416     | 22.1       | 0.331   |
| Granate      | 8.562 | 37.336 | 1.273    | 29.255 | 2.382     | 10.8       | 0.17    |
| PlayaGrande  | 8.608 | 37.255 | 0.636    | 29.291 | 2.375     | 9          | 0.098   |
| Rodadero     | 8.808 | 38.063 | 1.273    | 29.31  | 2.38      | 8.8        | 0.098   |
| Aeropuerto   | 8.808 | 37.355 | 0.936    | 29.291 | 2.575     | 10.5       | 0.098   |

El número de variables debe ser ≤ al número de observaciones.

(2) Realizar el ACP con una matriz de correlación, debido a que las variables ambientales se encuentran en unidades distintas y se necesita una estrategia que estandarice o vuelva más comparativa la relación entre las variables que agrupan a las localidades:

#### acp<-princomp(datos,cor=T)</pre>

acp representa al nombre del primer análisis que se hace, prcomp corresponde al comando que ejecuta el ACP, datos es el nombre de nuestra base de datos y cor=T indica que se utilizó una matriz de correlación (si no se coloca, se realizará el análisis con la matriz de covarianza).

(3) Calcular la tabla de autovalores (eigenvaues " $\lambda$ "), de la siguiente forma: summary(acp)

| Importance of compone  | nts:   |        |        |        |        |        |         |
|------------------------|--------|--------|--------|--------|--------|--------|---------|
|                        | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7  |
| Standard deviation     | 1.85   | 1.11   | 0.99   | 0.95   | 0.68   | 0.16   | 1.4E-08 |
| Proportion of Variance | 0.49   | 0.17   | 0.14   | 0.13   | 0.07   | 0.00   | 2.8E-17 |
| Cumulative Proportion  | 0.49   | 0.66   | 0.80   | 0.93   | 1.00   | 1.00   | 1.0E+00 |

Ejemplo 1. ACP CON VARIABLES AMBIENTALES (Continuación)





Se presenta una tabla que indica que se requiere de 7 componentes principales (CP) para explicar el 100% de la ordenación (son 7 componentes por tratarse de 7 variables). Se muestra la variación que explica cada componente principal, presentándose un 0.66 (66%) de variación (ajuste) en las dos primeras componentes principales (CP).

Para determinar el resto de elementos del ACP de escriben los siguientes comandos: names(acp) acp\$loadings

| Loadings:  |        |        |        |        |        |        |        |
|------------|--------|--------|--------|--------|--------|--------|--------|
|            | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 |
| pН         | 0.428  | -0.415 | -0.184 | 0.608  | 0.197  | 0.446  |        |
| Cond       | -0.361 | -0.163 | -0.741 | 0.268  | -0.347 | -0.318 |        |
| Turbidez   | 0.649  | -0.619 | -0.407 | 0.107  |        |        |        |
| Temp       | -0.52  | -0.206 | 0.814  |        |        |        | X      |
| Salinidad  | 0.125  | -0.669 | -0.596 | 0.108  | -0.288 | -0.134 | -0.263 |
| CapaFotica | -0.518 | -0.245 | -0.134 | -0.292 | 0.753  |        |        |
| Oxigeno    | -0.352 | 0.154  | -0.237 | 0.635  | 0.532  | -0.253 | -0.215 |

En donde *names(acp)*, genera los diferentes tipos de información adicional para el ACP. *Acp\$loadings*, muestra a la matriz de autovectores los cuales, en los cuales se muestra el peso de cada variable en la ordenación, el cual varía entre componentes. Para el Componente 1 la variable más correlacionada positivamente es la *Turbidez* (0,649) y negativamente es *CapaFotica* (-0,518). El signo (+ -) de la correlación muestra que estas variables pueden estar relacionadas inversamente (relación negativa).

acp\$scores

|              | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7   |
|--------------|--------|--------|--------|--------|--------|--------|----------|
| BahíaTaganga | -1.55  | 0.34   | -0.14  | -0.85  | -1.34  | 0.11   | -5.4E-15 |
| PuntaBetín   | -1.80  | -1.54  | 0.97   | -0.67  | 0.36   | -0.19  | -5.4E-15 |
| Monohuaca    | -2.77  | 0.36   | -0.84  | 1.20   | 0.63   | 0.11   | -5.1E-15 |
| Granate      | 0.93   | 1.34   | 0.27   | 0.96   | -0.41  | -0.27  | -5.6E-15 |
| PlayaGrande  | 1.67   | -0.05  | 1.70   | 0.52   | 0.16   | 0.22   | -5.3E-15 |
| Rodadero     | 1.19   | 1.20   | -0.58  | -1.53  | 0.82   | 0.01   | -6.4E-15 |
| Aeropuerto   | 2.33   | -1.64  | -1.38  | 0.36   | -0.22  | 0.01   | -6.1E-15 |

En esta matriz se muestra las coordenadas de cada observación (bahía) en los nuevos ejes o componentes. Cada sitio tendrá 7 coordenadas debido a la generación de 7 componentes principales. Este insumo es importante para la construcción del gráfico de componentes principales.

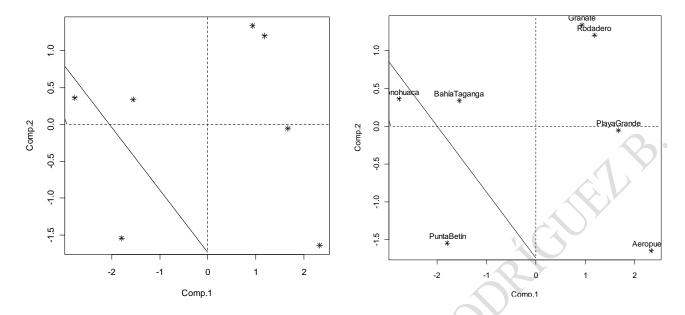
(4) Figuras del PCA. Primero se grafican las observaciones en el plano cartesiano de las componentes:

```
with(acp,plot(acp$scores,pch=8))
abline(v=0,lty=2)
abline(h=0,lty=2)
with(acp,text(acp$scores,row.names(acp$scores),pos=3))
abline(v=0,lty=2)
abline(h=0,lty=2)
```

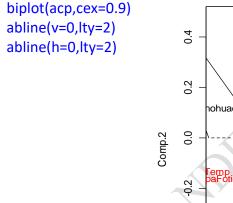
Ejemplo 1. ACP CON VARIABLES AMBIENTALES (Continuación)

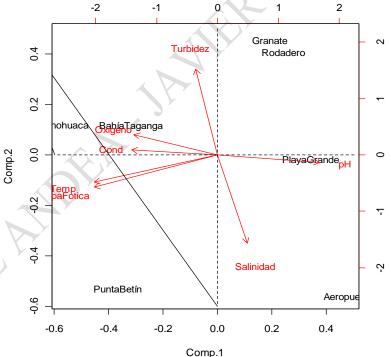






Se observa un gradiente en los sitios de muestreo, realizado con base en las similitudes lineales entre sus variables ambientales.





La figura integra la ordenación de las variables y las observaciones (grafico biplot). Una forma de analizar esta figura es analizando el plano cartesiano, cuyo "eje x" representa el CP1, en orden jerárquico este es el más importante pues resume la mayor cantidad de varianza (mejor ordenamiento), luego se analiza el "eje y" (CP2) y así sucesivamente.

Ejemplo 1. ACP CON VARIABLES AMBIENTALES (Continuación)





Al analizar el eje positivo del CP1, se encuentra que *Playa Grande* está asociada con la variable *pH* y cercano al eje negativo del CP1, se presentan las estaciones *Bahía Taganga* y *Monohuaca*, asociadas a altos valores de Oxigeno, Conductividad y en cierta relación con *Temp*. Y *Capa Fótica*. Posteriormente al analizar el eje positivo del CP2, se presenta una asociación entre la estación *Granate* y *Rodadero* con valores importantes de *Turbidez*, así como en el eje negativo de este componente se muestra una relación entre las estaciones *Aeropuerto* y *Punta Betín* con valores de Salinidad (dicha relación es más débil que las anteriores, debido a que CP2, captura menor ajuste que CP1).

**Modelo Lineal y selección de variables a partir del PCA.** De manera cualitativa se escogerán las variables más correlacionadas con el primer eje y se construirá un modelo lineal a partir de estas.

#### library(FactoMineR)

Esta librería permite desarrollar figuras de ordenación más elaboradas de componentes principales, principalmente cuando tenemos muchas observaciones (minería de datos). datos.PCA<-datos[, c("pH", "Cond", "Turbidez", "Temp", "Salinidad", "CapaFotica", "Oxigeno")] datosPCA agrupará a las 7 variables en estudio.

### pca<-PCA(datos.PCA, scale.unit=TRUE, ncp=7, graph = FALSE)

PCA representa el comando para realizar el análisis de componentes principales requerido, en algunos casos permite incluir variables suplementarias o factores en el análisis. El comando scale.unit, corresponde a la estandarización de las variables (similar al comando cor), ncp representa el número de ejes o componentes que puede incluir. Por defecto incluye a cinco ejes. Graph=FALSE, impide que se realicen dos figuras del análisis que a continuación se elaborarán de una manera más detallada.

### plot.PCA(pca, axes=c(1, 2), choix="ind", col.ind="black", col.ind.sup= "blue)

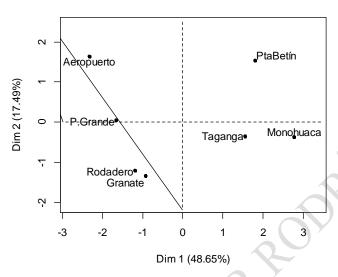
plot.PCA, es el comando utilizado para la realización de figuras de componentes principales. PCA, representa al análisis de ordenación que se ha realizado, axes=c(1, 2), indica los ejes que se seleccionarán. Choix="ind", indica que la ordenación que se muestra es de las observaciones (bahías). El resto de posibilidades para realizar este tipo de figuras se puede encontrar ingresando a la ayuda: help=(plot.PCA).





Ejemplo 1. ACP CON VARIABLES AMBIENTALES (Continuación)

### Individuals factor map (PCA)

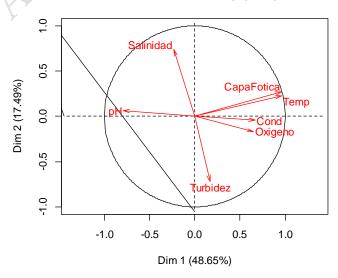


La anterior figura muestra en los ejes, el porcentaje de varianza de cada uno de los ejes sintéticos de la ordenación (autovalores) y la relación de las observaciones (bahías), medida en términos de su distancia euclideana.

plot.PCA(pca, axes=c(1, 2), choix="var", col.var="#ff0000", new.plot=T, label=c("var", "quanti.sup"))

Con el anterior comando, se realizará una ordenación de las variables que se encuentran caracterizando a las bahías (choix="var"). new.plot=T, corresponde a la realización de un panel gráfico adicional para presentar la figura, col.var="#ff0000", permite que las variables sea coloreadas en rojo. label=c("var", "quanti.sup"), permite incluir los rótulos a las variables (var) y para el caso en que haya variables suplementarias (quanti.var).

### Variables factor map (PCA)







### Ejemplo 1. ACP CON VARIABLES AMBIENTALES (Continuación)

La anterior figura muestra la ordenación de las variables (loadings) y su relación suele medirse con respecto al coseno del ángulo entre parejas de variables. Ej. El ángulo entre la turbidez y la salinidad es cercano a 180º, el coseno de ese ángulo es de -1, lo cual indica que existe una relación inversa entre esas variables. El círculo de la figura es llamado "círculo de contribución al equilibrio" (Borcard 2011), lo cual relaciona a la contribución de las variables. Aquellas variables con una magnitud superior al radio del círculo serán las de mayor contribución. En este caso Capa Fótica y Temperatura al menos alcanzan la longitud radio y serán las de mayor contribución, seguidas del pH.

Existen algunos insumos de este análisis que pueden encontrarse en names(pca), los cuales pueden ser: "eig" "var" "ind" "svd" "call". En donde pca\$eig, muestra a la matriz de autovalores, la cual define la varianza que captura cada eje de la ordenación. pca\$var, muestra varias matrices que asocian a las variables con los ejes del análisis, como son las coordenadas (coord), la correlación (cor), el coseno cuadrado (cos2) y las contribuaciones (contrib).

#### dimdesc(pca)

El anterior comando es muy importante, para el caso en que haya una marcada relación lineal de las variables, debido a que permitirá realizar propuestas de modelo lineales.

| \$Dim.1         |             |         |  |  |  |
|-----------------|-------------|---------|--|--|--|
| \$Dim.1\$quanti |             |         |  |  |  |
|                 | correlation | p.value |  |  |  |
| Temp            | 0,959       | 0,0006  |  |  |  |
| CapaFotica      | 0,956       | 0,0007  |  |  |  |
| рН              | -0,790      | 0,0346  |  |  |  |

\$Dim.1, relaciona a las variables con el primer eje, \$Dim.1\$quanti, indica que la relación se hace con variables cuantitativas, correlation, define el coeficiente de correlación de Pearson para cada variable que muestra relación con las observaciones ordenadas (bahías). Para este caso son 3 de las 7 variables las que muestran mayor significancia (p.value) en la relación con las bahías (Temperatura, Capa Fótica y el pH). En este sentido la propuesta de modelo lineal podría ser la siguiente:

Físico-química de las bahías estudiadas (Y) = 0.959.Temperatura (X1) + 0.956.Capa Fótica (X2) – 0.790.pH (X3).

Funciones Scaling1 y Scaling2 para ordenar y elección del número de ejes importantes. Para la siguiente ordenación se llamará en su orden, a las siguientes liberarías:

library(vegan) library(gclus) library(ape)

El análisis de componentes principales, se realizará con la función "rda", que representa a un análisis de redundancia, que para este caso permitirá analizar la relación lineal entre las variables y las observaciones, con el siguiente comando.





# Ejemplo 1. ACP CON VARIABLES AMBIENTALES (Continuación)

#### pca1 <- rda(datos, scale=TRUE)</pre>

pca1, representan en adelante el análisis de componentes principales realizado, datos, corresponde al nombre que presenta la base de datos originales y scale=TRUE, relaciona a la estandarización (scale) que se aplicó a las variables, por estar en escalas disímiles.

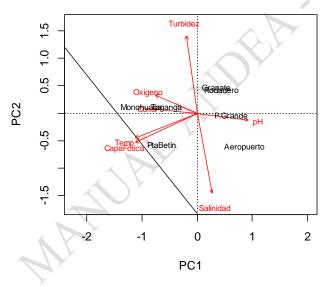
### summary(pca1)

El resumen estadístico del análisis realizado (summary), muestra una serie de insumos numéricos que serán detallados más adelante en el análisis de redundancia. Se destacan algunos resultados como Importance of components, en donde se muestran los autovalores y la proporción de varianza que explica cada componente (descrito anteriormente). Posteriormente se presentan las coordenadas de las variables y de las localidades u observaciones (scores).

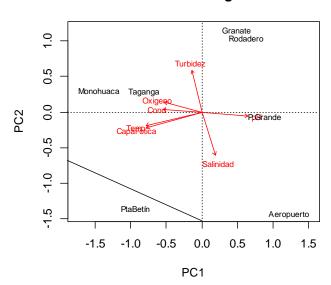
windows(title="Biplot de datos ambientales", 12, 6) par(mfrow=c(1,2)) biplot(pca1, scaling=1, main="PCA - scaling 1") biplot(pca1, main="PCA - scaling 2")

windows(... corresponde a los parámetro definidos para construir un nuevo panel de la siguiente figura que comparará a dos biplots (variables y observaciones). *main"PCA...,* respresenta el título de la figura, Los scaling 1 y 2, son dos formas de visualizar gradientes, el primero agrupa a las observaciones y el segundo a las variables. *scaling=1,* prioriza en la relación o cercanía de las observaciones y los ángulos o magnitudes de los vectores no presentan mayor sentido. Scaling=2, es un proceso inverso al anterior, en donde los ángulos de los vectores reflejan sus correlaciones.





PCA - scaling 2



En scaling 1, es posible visualizar un gradiente de izquierda a la derecha del primer eje, representado por bahías y su relación con variables ambientales. Comenzando con un primer grupo representado por las bahías Punta Betín, Monohuaca y Taganga, relacionadas con mayores valores de oxígeno, conductividad, temperatura y capa Fótica. Luego hay un segundo grupo con el resto de bahías, asociadas principalmente al pH. El scaling 2 lo que agrupa son las variables y mejora la visualización de sus relaciones.



# Ejemplo 1. ACP CON VARIABLES AMBIENTALES (Continuación)

Se pueden usar dos criterios para definir el menor número de ejes necesarios para la ordenación de los datos, lo cual es importante a la hora de realizar la interpretación de las estructuras con pocos ejes, resumiendo así la información del análisis. Una de las opciones se define con el análisis de los valores propios o autovalores (eigenvalues). La selección de los ejes puede ser arbitraria o cualitativa, al identificar el número de ejes necesarios para representar el 75% de la varianza de los datos. Otra opción se define mediante pruebas estadísticas como se describe a continuación. (1) criterio de Kaiser-Guttman, calcula el promedio de los autovalores y escoge a los que son mayores a su media, de acuerdo a los siguientes comandos:

(av <- pca1\$CA\$eig)

| PC1   | PC2   | PC3   | PC4   | PC5   | PC6   |
|-------|-------|-------|-------|-------|-------|
| 3,406 | 1,224 | 0,984 | 0,898 | 0,463 | 0,025 |

av[av > mean(av)]

|       | \ /3  |
|-------|-------|
| PC1   | PC2   |
| 3,406 | 1,224 |

av, representa a los autovalores (\$eig) del análisis realizado, mean(av), corresponde a los promedios para los dos primeros autovalores.

(2) criterio de Vara Quebrada (vq), divide de forma aleatoria a una vara de longitud unitaria (p) entre la cantidad de piezas de los ejes del PCA (n), de la siguiente manera:

*j* representa a las piezas (ejes) y *p* a los valores calculados con el modelo de vara quebrada. Para este caso, los ejes seleccionados son aquellos en donde los valores propios son mayores a los valores del modelo,

Panel gráfico en el que incluirán dos figuras.

```
windows(title="Autovalores del PCA")
par(mfrow=c(2,1))
```

1. Figura del modelo de Kaiser

```
barplot(ev, main="Valores Propios", col="bisque", las=2)
abline(h=mean(av), col="red")
```

legend("topright", "Promedio de autovalores", lwd=1, col=2, bty="n")

2. Figura del modelo de Vara Quebrada

barplot(t(cbind(100\*av/sum(av),vq\$p[n:1])), beside=TRUE, main="% Varianza", col=c("bisque",2), las=2)

legend("topright", c("% Autovalores", "Modelo Vara Quebrada"), pch=15, col=c("bisque",2), bty="n")



**Valores Propios** 3.0 Promedio de autovalores 2.5 2.0 1.5 1.0 0.5 PC2 PC4 PC6 2 % Varianza 40 % Autovalores Modelo Vara Quebrada 30 20 10 0 PC2 PC3 PC6 PC1

Ejemplo 1. ACP CON VARIABLES AMBIENTALES (Continuación)

Para el caso de la primera figura (criterio de Kaiser-Guttman), el promedio de autovalores es superado por los dos primeros autovalores, por lo que se definen dos ejes como los de mayor importancia. Para la segunda figura (criterio de vara quebrada), muestra que 3 autovalores están por encima del modelo, por lo que se define a 3 ejes como los de mayor importancia (PC1, PC4 y PC5). Para este último caso la interpretación es más compleja.

### 1.2 Ejercicios propuestos

En la base de datos entregada por el docente (ambiental.xlsx), las filas corresponden a sitios de observación y las columnas representan a variables fisicoquímicas, de un estudio desarrollado en la Sabana de Bogotá para caracterizar localidades (ríos) en diferentes cuencas, de acuerdo con sus variables fisicoquímicas.

Con base a lo anterior, realice un ejercicio que incluya a los siguientes elementos:

- 1. Diseño de objetivos (por lo menos 2)
- 2. Elaborar una pregunta de investigación
- 3. Rédactar la hipótesis del diseño estadístico propuesto.
- 4. Análisis de algunas graficas exploratorias (univariadas y multivariadas) vistas en clase y otras que sean consultadas en documentos en físicos o en la web, Interpretar en detalle los resultados.
- 5. Realizar un análisis de PCA, enfatizando su análisis en el contexto biológico y ecológico (analizar el tipo de matriz utilizada, las matrices de autovalores, autovectores y escores).
- 6. Pendiente el punto de PCA implementando el paquete "factoextra".





- 7. De la Lectura de Borcard et al. (2018) y con la base "ambiental.xlsx", se deben realizar los procedimientos del numeral 5.3.2 (págs. 118 125) y explicar los resultados que se obtengan. Se adjunta copia del script del libro mencionado.
- 8. Investigar por el Criterio Kalser Meyer Olkin (KMO) y la prueba de esfericidad de Barlett, para el análisis PCA. Aplicar estos análisis a la base de datos del punto anterior y analizar los resultados.
- 2. Análisis de Factores (AF)

**Ejemplo 2. AF CON VARIABLES AMBIENTALES** 

- 2.1 Ejercicios propuestos
- 3. Análisis de escalamiento multidimensional (MDS y nMDS)

**Ejemplo 3. MDS CON VARIABLES BIOLÓGICAS** 

Ejemplo 4. nMDS CON VARIABLES BIOLÓGICAS

**Ejemplo 5. nMDS CON DATOS DE PRESENCIA-AUSENCIA** 

- 3.1 Ejercicios propuestos
- 4. Análisis de correspondencia (AC o CA)

**Ejemplo 6. AC CON VARIABLES HUMANAS** 

- **4.1. Ejercicios propuestos**
- 5. Análisis de correspondencia dirigido (ACD o DCA)
- 5.1 Análisis canónicos.

Ejemplo 7. ADC, RDA y ACC con arañas

Ejemplo 7. ADC, RDA y ACC con arañas (continuación)

Ejemplo 8. ADC, RDA y ACC con arañas

7.2 Ejercicios propuestos

## **Bibliografía**

<u>Nota</u>: Se requiere contar con la versión completa del libro Análisis de Datos Ambientales y Ecológicos - ANDEA

