

IMM con R. Lab3: Medidas de Asociación: similitudes

Dra. Maite Mascaro y Dr. Edlin Guerra Castro

10/03/2021

Caso hipotético: Análisis de cambio de macrofauna en un gradiente de contaminación.

Parte 1: Análisis univariados a datos de estructura multivariada

Se requiere identificar si existen cambios en la composición de especies bentónicas en una línea de costa en cuyo extremo oriental se están vertiendo aguas residuales de una planta de tratamiento de un gran hotel. Estas aguas, aun siendo tratadas, siguen siendo ricas en materia orgánica, y se dispersan según el patrón de corrientes en sentido Este-Oeste. En principio, la línea de costa es bastante homogénea en características ambientales, siendo la contaminación orgánica la única variable que genera gradiente ambiental. Se presume que el aporte orgánico puede estar modificando la estructura trófica del ecosistema marino-costero según el gradiente de dispersión. Para identificar si esto ocurre, se decidió evaluar la composición de especies y las abundancias de los poliquetos de la macrofauna, ya que es bien conocido que son excelentes bioindicadores de contaminación orgánica. Se tomaron 10 muestras de sedimento en seis localidades a lo largo de la bahía, desde el punto de vertido hasta 30 km alejados de la fuente de contaminación. Las abundancias por especies fueron agrupadas por localidad y se encuentran en el archivo “gradiente.csv”. La actividad consiste en identificar el gradiente biológico. NOTA: Sí existe gradiente biológico, fallar en detectarlo implica librar al hotel de responsabilidades ambientales y administrativas ante el gobierno.

1. Importe la matriz de datos “gradiente.csv” i nómbre *gradiente*. Examine el objeto y responda:
 - a) ¿Qué tipo de objeto se creó? Si usaste la función `read_csv` del paquete `readr`, asegure de transformar el objeto resultante a *data frame* con la función `as.data.frame`.
 - b) ¿Cuáles son las dimensiones del objeto?
 - c) ¿qué representan las filas y qué representan las columnas?
 - d) ¿Con base en las dimensiones, se puede inferir cuántas especies hay en la matriz y en cada localidad?
 - e) ¿Son similares las magnitudes de abundancia de las especies en cada localidad?
 - f) Considerando que esta es una matriz pequeña ¿puede apreciar algún patrón biológico viendo solo la tabla?
2. Genere una tabla de nombre “uni” que describa las propiedades UNIVARIADAS bióticas de las localidades, incluya: la riqueza de especies, la abundancia de individuos y el índice de diversidad de Simpson. Le recomiendo buscar ayuda sobre las funciones `specnumber`, `diversity`. Acá un script que le ayudará a avanzar rápido. Nota: ninguna de estas funciones filtran texto

```
uni <-  
data.frame(  
  "S" = specnumber(gradiente),  
  "N" = apply(gradiente, 1, sum),  
  "Simpson" = diversity(gradiente, "simpson")  
)
```

- a) ¿Obtuvo un error? ¿qué cree que pudo ocurrir?
- b) Subseleccione sólo las columnas con información de las abundancias de las especies y llámelas “dat”.

- c) Repita el script para generar la tabla de descriptores univariados.
- d) ¿volvió a obtener un problema? ¿qué cree que pudo ocurrir ahora?
- e) ¿qué significan los NA? ¿cómo convertimos los NA en números? ¿Qué números le asignamos?
- f) Calculados los descriptores univariados ¿puede apreciar algún patrón biológico viendo solo la tabla? Explique el patrón. Si considera necesario realizar algún gráfico, hágalo. El siguiente script puede ser útil, pero no es la única forma.

```
dat.s <- data.frame(stack(dat), loc = row.names(dat))
qplot(
  data = dat.s,
  x = loc,
  y = values,
  group = ind,
  col = ind,
  shape = ind,
  geom = c("point", "line"),
  xlab = "Localidad",
  ylab = "Abundancia"
)
```

- g) Aparentemente no hay un patrón biológico según el gradiente, sin embargo, la ausencia de patrón puede ser producto del reduccionismo de los descriptores univariados. Por ello, usaremos métodos multivariados para apreciar mejor la información de la matriz.

Parte 2: Análisis en modo Q

1. Hemos identificado que nuestra matriz tiene 6 muestras (localidades) y 10 especies. Además, no todas las especies están en todas las muestras, hay especies con abundancias muy elevadas y otras muy escasas. ¿qué representa esto para iniciar un análisis multivariado? Exploremos esto usando varios métodos. Sugiero leer el material de ayuda `vegdist` del paquete `vegan`, tipeando `??vegdist` en la consola. Para esta práctica compararemos el desempeño del índice Bray-Curtis y la distancia Euclídeana.
- a) Iniciemos con las distancias Euclídeanas. Use la función `vegdist`, y llame al resultado `euc`.
- b) ¿cuántos valores se estimaron teniendo en cuenta que hay 6 localidades?
- c) ¿Qué tipo de objeto es `euc`?
- d) ¿cómo se interpretan las unidades medidas?
- e) ¿cuáles son las localidades más parecidas? ¿qué puede decir de la relación entre la localidad 1 y 6?
- f) ¿Aprecia el gradiente? Si no, genere un dendrograma que nos ayude a visualizar el patrón de cambio

```
plot(hclust(euc, "average"))
```

- g) Interprete el gráfico. ¿Hay patrón de cambio según el gradiente ambiental?
- h) ¿Por qué no se aprecia el patrón? ¿qué puede estar ocurriendo?
- i) Usemo ahora Bray-Curtis. Use la función `vegdist`, y llame al resultado `bray1`.
- j) ¿cuántos valores se estimaron teniendo en cuenta que hay 6 localidades?
- k) ¿Qué tipo de objeto es `bray1`?
- l) ¿qué fue lo que se estimaron, similitudes o disimilitudes? ¿cómo se interpretan estas unidades? ¿cuál es su escala? Para obtener las similitudes e interpretarlas con mayor facilidad aplique el siguiente comando

1-bray1

- n) ¿cuáles son las localidades más parecidas? ¿qué puede decir de la similitud entre la localidad 1 y 6? ñ) ¿Aprecia el gradiente? Si no, genere un dendograma que nos ayude a visualizar el patrón de cambio

```
plot(hclust(bray1, "average"))
```

- o) Este gráfico le permitirá identificar el par de localidades, o subgrupos de localidades, más semejantes en forma jerárquica y pareada ¿Aprecia el patrón?
- p) El patrón aun no se aprecia, probablemente se deba al peso de la especie 5. Aplique transformación raíz cuadrada a las abundancias y llame a esa matriz nueva “dat2”. Calcule nuevamente el índice Bray-Curtis y llámelo bray2.
- q) ¿cuáles con las localidades más parecidas? ¿qué puede decir de la similitud entre la localidad 1 y 6? ¿Aprecia el gradiente? Genere un dendograma que nos ayude a visualizar el patrón de cambio

```
plot(hclust(bray2, "average"))
```

- r) ¿Ahora sí aprecia el patrón de cambio con el gradiente ambiental?
- s) Compare los tres dendogramas
- t) Genere un nuevo gráfico multivariado, es un nMDS (escalamiento multidimensional no métrico - tema de la Unidad 5). Por ahora solo importa indicar que la distancia de los puntos proyectados en el nMDS refleja las similitudes expresadas en la matriz. En este sentido ¿aprecia mejor el gradiente de cambio en la composición de especies?

```
mds <- metaMDS(dat2, "bray")
plot(mds, type = "n")
text(mds,
      display = "sites",
      cex = 1,
      col = "blue")
```

2. El enunciado hace referencia a la composición de especies, explícitamente indica que solo importa el cambio en la identidad de las especies, no en sus abundancias. Esto implica que a pesar de la transformación raíz cuadrada, las diferencias en abundancias pueden estar influyendo en el patrón observado. Para satisfacer plenamente la pregunta enunciada, se aplicará un índice que sólo considere la presencia de la especie. Hay más de 25 índices para evaluar matrices de incidencia, uno de los más populares es Jaccard. Aplique este índice con la función `vegdist` y llame al objeto “jac”.

- a) ¿Qué puede inferir ahora del patrón y el gradiente? ¿cuál es la similitud en composición de especies entre las localidades 1 y 6?
- b) ¿Cuál es la similitud en la composición de especies entre las localidades 6 y 5?
- c) Genere un dendograma para proyectar la nueva matriz. Interprete el agrupamiento del dendograma
- d) Veamos ahora globalmente los cuatro dendogramas generados. Para ello use el siguiente código

```
par(mfrow = c(2, 2), mar = c(2, 4, 2, 2)) # este script indica a R que se
#proyectarán cuatro gráficos en uno solo
plot(hclust(euc, "average"), main = "Dist. Euclidean")
plot(hclust(bray1, "average"), main = "Bray-curtis sin transformación")
plot(hclust(bray2, "average"), main = "Bray-curtis con transformación")
plot(hclust(jac, "average"), main = "Jaccard")
```

- e) Existen métodos cuantitativos para comparar matrices. Uno de ellos es el método de Mantel, y se base en correlacionar los rangos de las jerarquías de las similitudes en cada matriz. Por ahora solo interesa conocer la correlación, y no la probabilidad asociada. Como reto para usted, busque en el paquete

vegan la función **mantel**, y correlacione la matriz de similitud en composición de especies respecto a las matrices Bray-Curtis y la Matriz de distancias Euclidianas.