

ANÁLISIS ESTADÍSTICOS MULTIVARIADOS Y DISEÑO DE GRÁFICAS EN R

Itzue W. Caviedes-Solis

Universidad de Washington, Seattle

Área temática: Ecología

Objetivos

Los análisis multivariados incluyen técnicas descriptivas e inferenciales para analizar datos ecológicos complejos. Permiten el estudio de bases de datos compuestas de dos o más variables y son ampliamente aplicables a diferentes preguntas de investigación y a diferentes organismos. El objetivo del taller es que los participantes adquirirán los conocimientos básicos de tres métodos de análisis estadísticos: Correlación, Cluster y Análisis de Componentes Principales. Implementarán el lenguaje de programación de R, el cual es ampliamente usado para análisis ecológicos. Los análisis aprendidos son aplicables a todo tipo de datos y el diseño de gráficas ofrece una mayor claridad y un mayor atractivo visual para tesis y publicaciones.

Descripción de actividades

Cada sección incluirá un marco teórico y una práctica en R Studio. La práctica incluye el análisis de datos y el diseño de gráficas, para los cuales se incluirá un tutorial a seguir. La base de datos a utilizar corresponde al estudio de la relación entre la morfología y el rendimiento en el nado de la familia de peces Cyprinidae (Yan et al., 2013). Sin embargo, los métodos pueden ser aplicados a cualquier organismo (incluyendo anfibios y reptiles!). Las matrices de datos incluyen morfología (medidas lineares y proporciones), datos de rendimiento de nado (velocidad y aceleración) y microhábitat (tipo de corriente).

**Los participantes deben contar con un conocimiento básico de programación en R.

**Se asignarán lecturas que deberán realizarse antes de tomar el taller (máximo 3 artículos).

Materiales a Utilizar

- Computadora personal (Windows o Mac)
- Free software R Studio (<https://www.rstudio.com/products/rstudio/>)
- R packages: corrplot (Wei, 2013), Nbclust (Charrad et al., 2014), ggplot2,
- Se proveerá una base de datos y un tutorial para analizar durante el taller.

Programa por día

Día 1

Marco Teórico. Análisis estadísticos multivariados (1 hora)

Análisis de Correlación (2 horas)

Estimar la correlación entre dos o mas variables

Día 2

Análisis Cluster (2 horas)

Kmean para determinar el numero de clusters.

Análisis de Componentes Principales (2 horas)

Calcular eigenvalues y determinar si son significativos con el modelo Broken stick.

Calcular eigenvectors y su contribución a la variación a cada componente principal.

Referencias

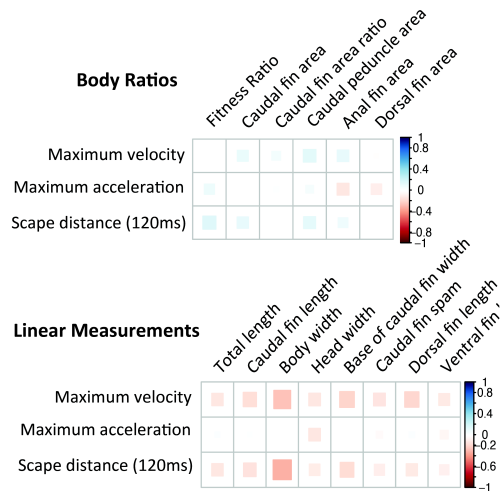
Charrad M., Ghazzali N., Boiteau V., Niknafs A. (2014). "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set.", "Journal of Statistical Software, 61(6), 1-36.", "URL <http://www.jstatsoft.org/v61/i06/>".

Wei, T., 2013. corrplot: visualization of a correlation matrix. R package version 0.60.

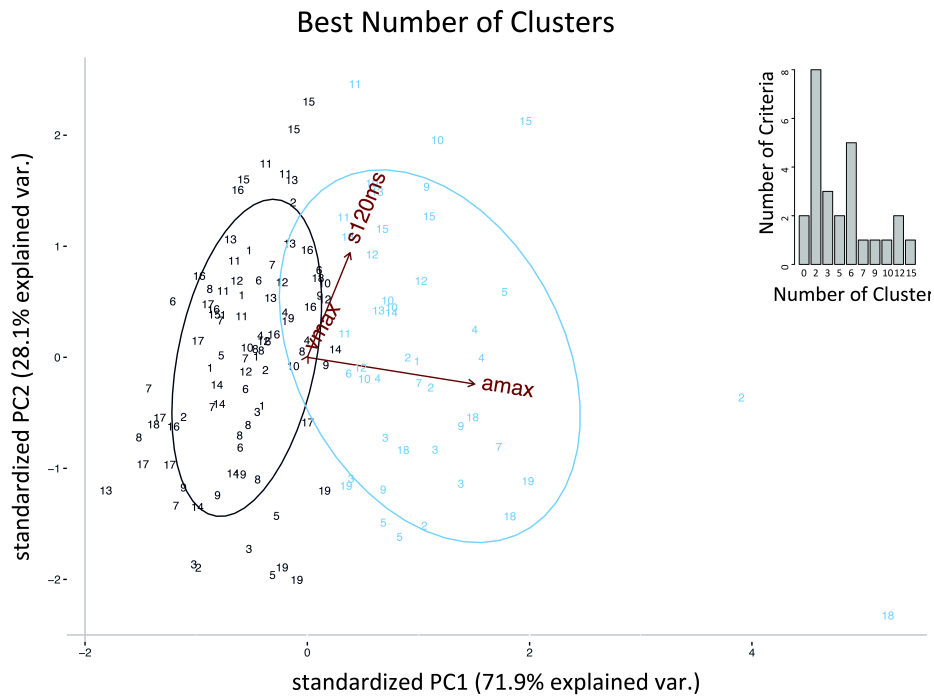
Yan, G.J., He, X.K., Cao, Z.D. and Fu, S.J., 2013. An interspecific comparison between morphology and swimming performance in cyprinids. *Journal of evolutionary biology*, 26(8), pp.1802-1815. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.vk820>

EJEMPLOS DE RESULTADOS

ANALISIS DE CORRELACION. La intensidad del color y el tamaño de los cuadros representa el valor de correlación. La escala de color de azul a rojo indica si la correlación es positiva o negativa.



ANALISIS CLUSTER. Los números corresponden a los individuos y el color al grupo al que fueron asignados. Las flechas corresponden a los parámetros que explican la variación.



ANÁLISIS DE COMPONENTES PRINCIPALES (PCA). Morfología y rendimiento del nado. El color de los puntos corresponde a las diferentes especies y el tamaño de los puntos a la longitud de del cuerpo.

