

Ejercicios sobre Análisis Estadístico en R

Introducción

Este manual proporciona una explicación detallada del código utilizado para realizar diversos análisis estadísticos en ecología de comunidades y poblaciones. Se incluyen análisis de correlación, Análisis de Componentes Principales (PCA), selección de modelos de regresión y selección de indicadores biológicos. Cada sección del manual describe la función de los métodos empleados y proporciona explicaciones detalladas de cada línea de código.

1. Instalación de Paquetes y Carga de Datos

Para ejecutar los análisis, es necesario instalar y cargar los paquetes requeridos:

Instalar paquetes necesarios

```
install.packages("lme4")
```

```
install.packages("ggplot2")
```

```
install.packages("reshape2")
```

```
install.packages("factoextra") # Para visualización de PCA
```

```
install.packages("psych") # Para análisis de cargas
```

Cargar librerías

```
library(lme4)
```

```
library(ggplot2)
```

```
library(reshape2)
```

```
library(factoextra)
```

```
library(psych)
```

Estos paquetes permiten la manipulación de datos, la visualización de los resultados y la realización de análisis estadísticos.

1.1 Configuración del Directorio de Trabajo

Se establece el directorio donde se encuentran los archivos de trabajo:

```
setwd("C:/Users/TU_DIRECTORIO/")
```

1.2 Carga del Workspace y de la Matriz de Datos

Se carga un archivo con funciones predefinidas y la matriz de datos:

```
load("funciones_iv.RData")
```

```
# Asignar la matriz a un objeto de trabajo
```

```
clima <- ejercicio_clase
```

2. Análisis de Correlación

El análisis de correlación evalúa la relación entre variables ambientales.

```
# Selección de variables ambientales
```

```
clima_variables <- clima[2:13]
```

```
# Escalar las variables (transformarlas a valores Z)
```

```
clima_variables <- scale(clima_variables)
```

```
# Cálculo de la matriz de correlación de Spearman
```

```
cor_clima <- cor(clima_variables, method = "spearman")
```

```
cor_clima_2dec <- round(cor_clima, 2)
```

La matriz de correlación se visualiza con un mapa de calor:

```
# Transformación para visualización
```

```
melted_cor_clima <- melt(cor_clima_2dec)
```

```
# Gráfico de correlación
```

```
ggheatmap <- ggplot(data = melted_cor_clima, aes(Var2, Var1, fill = value)) +
```

```
  geom_tile(color = "white") +
```

```
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0, limit = c(-1,1),  
    space = "Lab", name = "Spearman\nCorrelation") +
```

```
  theme_minimal() +
```

```
  theme(axis.text.x = element_text(angle = 90, vjust = 1, size = 10, hjust = 1)) +
```

```
coord_fixed()  
print(ggheatmap)
```

3. Análisis de Componentes Principales (PCA)

El PCA reduce la dimensionalidad de los datos y permite visualizar patrones:

```
# Calcular PCA
```

```
pca_result <- prcomp(clima_variables, center = TRUE, scale. = TRUE)
```

```
# Scree plot para visualizar la varianza explicada
```

```
fviz_eig(pca_result)
```

```
# Biplot del PCA
```

```
fviz_pca_biplot(pca_result, repel = TRUE, col.var = "blue", col.ind = "red")
```

```
# Cargas de las variables en los componentes principales
```

```
pca_loadings <- data.frame(pca_result$rotation)
```

```
print(pca_loadings)
```

```
# Selección de variables importantes según el PCA
```

```
pca_importance <- apply(abs(pca_result$rotation), 2, function(x) which(x > 0.3))
```

```
print(pca_importance)
```

4. Selección del Mejor Modelo de Regresión

Para identificar las variables ambientales más importantes para la riqueza de especies, se ajustan varios modelos:

```
# Modelo nulo
```

```
lm.vacio <- lm(clima$riqueza10 ~ 1)
```

```
summary(lm.vacio)
```

Bases ecológicas y genómicas de la interacción organismo-ambiente

Modelo con todas las variables seleccionadas

```
lm.completo <- lm(clima$riqueza10 ~ clima$Precip_media_anual + clima$Tmed_anual +  
clima$riesgo + clima$efectoborde)
```

Selección de modelo por Forward Selection

```
lm.forward <- step(lm.vacio, scope = list(lower = lm.vacio, upper = lm.completo), direction =  
"forward")
```

```
BIC(lm.forward)
```

```
summary(lm.forward)
```

Selección por Backward Elimination

```
lm.backward <- step(lm.completo, scope = list(lower = lm.vacio, upper = lm.completo),  
direction = "backward")
```

```
BIC(lm.backward)
```

```
summary(lm.backward)
```

Selección por Stepwise Regression

```
lm.stepwise <- step(lm.vacio, scope = list(lower = lm.vacio, upper = lm.completo), direction =  
"both")
```

```
BIC(lm.stepwise)
```

```
summary(lm.stepwise)
```

5. Selección de Indicadores Biológicos

Se identifican los taxones indicativos de ciertas condiciones ambientales mediante `test.indval()`:

Instalar y cargar paquete

```
devtools::install_github("EnriqueSanchezBio/envpoll")
```

```
library(envpoll)
```

Seleccionar datos de especies y sitios

```
taxa <- clima[,c(1,19:415)]
```

```
sitios <- as.factor(clima[,1])
```

```
# Cálculo del índice de valor indicador
```

```
iv <- test.indval(taxa, factor(sitios), n=1000)
```

6. Aplicaciones en Ecología de Comunidades y Poblaciones

Los análisis empleados en este manual tienen diversas aplicaciones:

- **Análisis de correlación:** Evalúa la interdependencia entre factores ambientales.
- **PCA:** Permite reducir la dimensionalidad de datos ecológicos y explorar patrones en la distribución de especies.
- **Modelos de regresión:** Identifican factores determinantes de la riqueza de especies.
- **Valor indicador de especies:** Determina qué especies están asociadas a condiciones ambientales específicas.