**ANÁLISIS DE CORRESPONDENCIAS NO RESTRINGIDO (CA) Y RESTRIGIDO (CCA) Y PARCIAL (pCCA)**

Método de ordenación (unimodal) no supervisada o no restringida (CA) y restringida (CCA). Son Sensibles a outliers.

**Consideraciones a la hora de su aplicación:**

* Asume una relación unimodal (de campana) con el gradiente.
* Útil para gradientes largos. Si son cortos se usarán otras técnicas lineales (PCA o RDA).
* Si la relación no es lineal ni unimodal (es decir, es monotónica) se recomienda el uso del NMDS.

**Qué objetivos se persiguen:**

* Son técnicas de ordenación y reducción de la dimensión.
* Representar la relación entre muestras e individuos mediante proximidad.
* Afinar la representación según ciertas variables explicativas maximizando la relación con dichas variables.

**Tipos:**

* CA Simple para 2 variables.
* CA Múltiple para análisis de más de 2 variables.
* CA Canónico o restringido (CCA) tenemos una matriz de restricciones.
* CCA Parcial (pCCA) cuando queremos eliminar o controlar el efecto de un grupo de variables.
* CCA sin tendencia (DCA) para eliminar efecto herradura creados por existencia de dobles ceros o ejes tienen relaciones no lineales.

**Cómo se trabaja:**

Dependiento del modelo a aplicar trabajaremos con una matriz o dos matrices. Las matrices son:

* Matriz X respuesta (VD, nominal, distribución unimodal): filas=muestras, columnas=individuos
* Matriz Y explicativas (VI, cuantitativas, y deben mostrar una rel. lineal con v.respuesta): filas=muestras, columnas=variables (para el CCA, pCA)
* Matriz Z o de control (cuantitativas, cuando deseamos controlar o eliminar el efecto de variables): filas=muestras, columnas=variables (para el pCA).

El uso de dichas matrices viene determinado por:

* Analisis de correspondencia clásico simple o múltiple (CA) usaremos solamente la matriz X
* Análisis de correspondencia restringido o canónico (CCA) usaremos las matrices Y y X
* Análisis de correspondencia parcial (pCCA) usaremos las matrices X y Z.

Con la matriz X:

* Correspondencia entre niveles de filas o columnas. Para una misma variable la correspondencia que hay entre sus distintas categorías.
* Correspondencia entre filas y columnas. Es decir, entre dos variables. Como trabajar con una tabla de contingencia.

Si utilizamos la matriz X con la matriz Y (de variables explicativas)

* Correspondencia entre individuos y variables explicativas. (CCA)

Si incluimos la matriz Z (pCA).

* Podemos controlar el efecto de un grupo de variables y analizar cualquiera de los primeros 3 objetivos.

**Metodología CA:**

1. Revisar la **Idoneidad de los datos**: Ver si el conjunto de variables a analizar (especies) tienes escales similares y no hay datos perdidos. Eliminando las observaciones que posean presencia de muchos ceros.
2. Vemos si existe relación entre filas (observaciones) y columnas (variables). Aplicación de Chi Cuadrado. Si es significativo dicho análisis se sigue con el análisis de correspondencias.
3. Representación mediante Biplot y Escalado.
4. Interpretación de la **Inercia**: Total (medida de la variabilidad total explicada).
5. Interpretación de los **Ejes**: Valores propios de cada eje que devuelve Chi cuadrado y el valor e importancia de los componentes (ejes) que nos aporta el % de explicación total.
6. Selección del número de ejes: Seleccionar todos los ejes que lleguen a explicar el 70% del total (también se puede usar el gráfico de barra quebrada o el valor propio mayor que 1).
7. Graficación en **Biplot** y **Escalado**.

**Metodología CCA:**

1. Para un conjunto determinado de variables (selección manual). Ahora sí que vamos a utilizar una evaluación directa de variables sobre observaciones.
2. Observamos a **inercia** que en este caso se divide en **restringida** y **no restringida**. Miramos la proporción de cada una de ellas.
3. Nos aporta el porcentaje de variabilidad que aportan cada variables que usamos como restricción en el modelo.
4. Interpretación y selección de los **Ejes** (viendo componentes no restringidos y restringidos).
5. Graficación con **Triplot** ya que debemos graficar también las variables explicativas.
6. Realización de pruebas de permutación: para saber si la ordenación es significativa.
7. Bondad de ajuste y selección del modelo. Y problemas de multicolinealidad.

**Metodología pCA** (eliminación del efecto de un grupo de variables)

1. Observación de la inercia dividida en condicional, restringida y no restringida.
2. Observación de los componentes restringidos y no restringidos.
3. Graficado de los componentes.
4. Aplicación del test de permutación. Cómo afectan dichos condicionantes elegidos.
5. Análisis de partición de varianzas.

**ESCALADO MULTIDIMENSIONAL MÉTRICO (MDS) Y NO MÉTRICO (NMDS)**

Son técnicas de ordenación (no restringida o indirecta) basadas en **distancias**. Lo que busca es una representación gráfica con dimensión reducida a partir de una matriz de distancias entre los objetos o individuos.

**Son de dos tipos:**

* Escalado multidimensional métrico (**MDS**).
* Escalado multidimensional no métrico (**NMDS**).

**Cómo se aplica dicha Clasificación.**

Si la matriz analizada se basa en una distancia métrica entonces estamos en un **MDS**, sino en un **NMDS**.

Escalado multidimensional métrico (Metric multidimensional scaling, MDS) o análisis de coordenadas principales (Principal coordinate analysis, PCoA) se basa en valores propios (similitud/distancias) similar al PCA pero con matriz de distancias/disimilaridad (no cov/corr), ordenación lineal pero más flexible (e.g. datos binarios)

Escalado multidimensional no-métrico (Non-metric multidimensional scaling, NMDS) se basa en distancias (rangos).

MDS es un análisis propio para una matriz de disimilaridades. Produce ordenaciones lineales, pero más flexibles. Las muestras próximas representan similaridad en el sentido de la media de asociación utilizada.

**Metodología MDS:**

1. Realizamos la pre-transformación de los datos (Hellinger).
2. Calculamos la matriz de distancias (puede ser la euclídea).
3. Se establecen el número de ejes. También se obtiene la bondad de ajuste.
4. Se grafica con **ordiplot**.

**Metodología NMDS:**

1. Realizamos la pre-transformación de los datos (Hellinger).
2. Calculamos la matriz de distancias (puede ser la euclídea).
3. Se establece en número de dimensiones y también el número de combinaciones aletorias.
4. **Stress** vs. nº dimensiones: incluir dimensiones hasta que no ganemos una reducción significativa en el valor de estrés. Si el valor de estrés es muy alto para 2 o 3 dimensiones, entonces puede que el NMDS no sea el mejor modelo.
5. Uso del **gráfico de Shepard** e interpretación de la **bondad de ajuste**. Stressplot: relación entre las distancias de puntos de la configuración final e inicial. Debemos trabajar con R cuadradados lo más cercanos a 1.
6. Si el ajuste no es malo, se grafica mediante **Biplot**. Para ver relaciones: Puntos cercanos significan similitud. Variables van representadas mediante flechas.

**MÉTODOS DE ORDENACIÓN LINEA Y REDUCCIÓN NO SUPERVISADO (PCA), Y SUPERVISADO (RDA y pRDA)**

PCA (análisis indirecto del gradiente) ordenación que resume el patrón principal de variación en la matriz respuesta. RDA (análisis directo del gradiente) ordenación que resume el patrón principal de variación en la matriz respuesta que puede explicarse por la matrix de variables explicativas.

Se parte de 2 o 3 matrices:

* Matriz X: variables respuesta.
* Matriz Y: variables explicativas.
* Matriz Z: variables a controlar.

Según la tipología:

* **PCA**: Ordenará la matriz de variables respuesta.
* **RDA**: Ordenar la matriz de variables respuesta (X) en función de la matriz de variables explicativas (Y).
* **pRDA** (RDA parcial): Controlar la ordenación a partir de una matriz (Z) de variables a controlar.

Obtendremos y analizaremos los componentes principales (ortogonales), que maximizan la varianza explicada. Identifica componentes principales, que representan la dirección de mayor variación en los datos, para representar cada muestra con un conjunto pequeño de estos PC en lugar de todo el conjunto de variables. Utiliza variables métricas (cuantitativas o numéricas) que estén correlacionadas. Preserva las distancias euclídeas entre las muestras.

**Objetivos**

* Representar de manera óptima un conjunto de variables e individuos (o casos).
* Transformar las variables originales, en general correlacionadas (o correladas), en nuevas variables no correlacionadas (o incorreladas), facilitando la interpretación de los datos (reduciendo la dimensionalidad de los datos).
* Definir qué variables contribuyen en mayor medida como fuente de variabilidad.
* Qué variables están relacionadas entre sí y cuáles no.

**Supuestos**

* Muestras independientes (diseño de estudio).
* Ausencia de outliers y pocos problemas de ceros.
* Normalidad multivariada
* Linealidad en la relación entre las variables.

**Problemas**

* Sensible a los valores perdidos (NA)
* Efecto arco o herradura por la existencia de dobles ceros.

**Metodología PCA**

1. Verificar los supuestos del análisis (normalidad, linealidad, correlación, etc..).
   1. Shapiro para normalidad. Si no es normal, o transformamos o usamos otra metodología.
   2. Linealidad mediante graficado de dispersión.
2. Pre-transformación de los datos (e.g. estandarizar, normalizar, etc.)
3. Realizar el PCA.
   1. Muestra la inercia total y no restringida (en este caso dará igual).
   2. Inercia explicada por cada componente.
4. Seleccionar el número de ejes (o PC) adecuado. Uso de varios modelos:
   1. Valores propios.
   2. Criterio de Kaiser-Guttman.
   3. Modelo broken-stick (barra quebrada).
5. Creación de gráficos Biplot (PCA).
   1. Escalado 1: Gráfico de distancias: Relación entre casos y objetos. Los casos más cercanos son los más parecidos. Los vectores son las variables y, según el tamaño determinan la contribución al eje.
   2. Escalado 2: Gráfico de Correlaciones: relación entre variables. Ángulo entre los vectores dará la correlación entre ellas.
6. Diagnóstico (spenvcor, goodness, inertcomp, VIF).

**Metodología RDA**

1. Verificar los supuestos del análisis (normalidad, linealidad, correlación, etc..).
   1. Shapiro para normalidad. Si no es normal, o transformamos o usamos otra metodología.
   2. Linealidad mediante graficado de dispersión.
2. Pre-transformación de los datos (e.g. estandarizar, normalizar, etc.)
3. Realizar el RDA.
   1. Muestra la inercia total, restringido y no restringido.
   2. Muestra los ejes restringidos y no restringidos.
4. Seleccionar el número de ejes (o PC) adecuado.
   1. Valores propios.
   2. Criterio de Kaiser-Guttman.
   3. Modelo broken-stick (barra quebrada).
5. Diagnóstico (spenvcor, goodness, inertcomp, VIF).
6. Bondad de ajuste (test de permutación). ANOVA.
7. Diagnosis del Modelo.
8. Triplot (RDA y pRDA).
   1. Uso de escalado 1: Gráfico de distancias. Relación entre casos.
   2. Uso de escalado 2: Gráfico de correlación. Relación entre variables.
9. Selección del modelo o las variables para incluir en el modelo (VIF, R2, ordistep).
10. Partición de la varianza. Que porcentaje de varianza explica cada parte del análisis.

**Metodología pRDA**

1. Verificar los supuestos del análisis (normalidad, linealidad, correlación, etc..).
   1. Shapiro para normalidad. Si no es normal, o transformamos o usamos otra metodología.
   2. Linealidad mediante graficado de dispersión.
2. Pre-transformación de los datos (e.g. estandarizar, normalizar, etc.)
3. Seleccionamos el/los subconjuntos de variables explicativas.
4. Realizar el RDA.
   1. Muestra la proporción e inercia total, condicionada, restringida y no restringida.
   2. Muestra los ejes restringidos y no restringidos.
5. Bondad de ajuste del modelo (test de permutación). ANOVA.
6. Bondad de ajuste de los ejes y efectos marginales.
7. Triplot (pRDA).
   1. Uso de escalado 1: Gráfico de distancias. Relación entre casos.
   2. Uso de escalado 2: Gráfico de correlación. Relación entre variables.
8. Partición de la varianza para todas las variables explicativas. Que porcentaje de varianza explica cada parte del análisis (varpart).
9. Partición de la varianza para las variables seleccionadas.