

## Análisis de Factores (AF)

- El análisis de factores intenta explicar la varianza en un número de variables originales utilizando un número menor de índices o factores.
- **Cada variable original puede expresarse como una combinación lineal de estos factores, más un término residual**, que refleja el grado en que la variable es independiente de las otras variables.
- Por ejemplo, un modelo de 2 factores para los datos de los gorriones viene dado por:

$$X_1 = a_{11}F_1 + a_{12}F_2 + e_1$$

$$X_2 = a_{21}F_1 + a_{22}F_2 + e_2$$

$$X_3 = a_{31}F_1 + a_{32}F_2 + e_3$$

$$X_4 = a_{41}F_1 + a_{42}F_2 + e_4$$

$$X_5 = a_{51}F_1 + a_{52}F_2 + e_5$$

$a_{ij} = \text{ctes.}$

$F_1$  y  $F_2 = \text{factores}$

$e_i = \text{la varianza en } X_i \text{ que es independiente de la varianza en las otras variables } X.$

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Análisis de Factores (AF)

- $F_1$  puede ser el factor de tamaño. En este caso, los coeficientes  $a_{i1}$  serán todos positivos, reflejando que algunos pájaros tienden a ser mayores y otros menores en todas las medidas de su cuerpo.
- $F_2$  puede medir un aspecto de la forma de los pájaros, con algunos coeficientes positivos y otros negativos.
- Un modo de realizar el análisis de factores consiste en tomar unos pocos **componentes principales como factores iniciales**, a los que luego se les aplica una **rotación** para facilitar la interpretación. Existen también otros métodos para determinar los factores iniciales.

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Análisis de la Función Discriminante (AFD)

---

- **Propósito:** Separar las **observaciones** en diferentes grupos, en base a las medidas disponibles. Ej.: ¿Cuán bien pueden separarse los gorriones sobrevivientes de los no sobrevivientes, utilizando las medidas de su cuerpo (**ejemplo 1**)? ¿Cuán bien pueden separarse los cráneos de diferentes épocas, utilizando también sus medidas (**ejemplo 2**)?
- Del mismo modo que el ACP, el análisis de la función discriminante se basa en la idea de encontrar una **combinación lineal adecuada de las variables originales**, para alcanzar el objetivo previsto.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Análisis de Conglomerados (Clusters) (AC)

---

- **Propósito:** Identificación de grupos compuestos por objetos similares.
- **Ej. 1 y 2:** Este análisis no tiene mucho sentido puesto que los grupos (sobrevivientes y no sobrevivientes, así como épocas) son ya conocidos.
- **Ej. 3:** Puede tener interés el agrupamiento de colonias en base a las variables ambientales o frecuencias Pgi.
- **Ej. 4:** El principal interés radica en la similitud entre perros prehistóricos thailandeses y otros animales.
- **Ej. 5:** Los países europeos posiblemente pueden agruparse en términos de su similitud en los patrones de empleo.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Análisis de Correlaciones Canónicas (ACC)

---

La correlación canónica separa **las variables** (no los objetos o individuos) en **dos grupos** y su interés se centra en las relaciones entre éstos.

Así, en el ej. 3, las primeras cuatro variables están relacionadas con el medio ambiente, mientras que las seis variables restantes reflejan la distribución genética en las diferentes colonias de *Euphydryas editha*. Encontrar qué relaciones existen, si las hay, entre estos dos grupos de variables tiene un interés biológico considerable.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Escalado MultiDimensional (EMD)

---

Este método mide distancias entre los individuos y produce, a partir de estas distancias, un "mapa" que muestra cómo se relacionan entre si los objetos/individuos.

Esta es una facilidad útil, ya que a menudo es posible medir cuán lejos están separados los pares de objetos sin tener ninguna idea de cómo los objetos están relacionados en un sentido geométrico.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Escalado MultiDimensional (EMD)

---

### Ej. 4

Hay maneras de medir las distancias entre los perros modernos y los chacales dorados, los perros modernos y los lobos chinos, etc.

Teniendo en cuenta que cada par de grupos de animales da 21 distancias en total, y a partir de estas distancias se puede usar el escalamiento multidimensional para producir un mapa de las relaciones entre los grupos.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Escalado MultiDimensional (EMD)

---

**Mapa 1-D:** los grupos se ubican a lo largo de una línea recta.

**Mapa 2-D:** los grupos están representados por puntos en un plano.

**Mapa 3-D:** los grupos están representados por puntos en el interior de un cubo.

**Mapas >3-D:** también son posibles, no se pueden visualizar de una manera sencilla.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Ordenación

---

El análisis de componentes principales y el escalamiento multidimensional se denominan a veces métodos de **ordenación**. Es decir, son métodos para producir ejes contra los que se puede representar un conjunto de objetos de interés.

Otros métodos de ordenación también están disponibles.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Análisis de Coordenadas Principales (ACoP)

---

El análisis de coordenadas principales puede entenderse como un tipo de análisis de componentes principales, que comienza con información sobre la **medida en que los pares de objetos son diferentes** en un conjunto de objetos, en lugar de los valores para las mediciones en los objetos.

Como tal, se pretende hacer lo mismo que en el escalamiento multidimensional. Sin embargo, las suposiciones hechas y los métodos numéricos utilizados no son los mismos.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Análisis de Correspondencias (Acorr)

---

El análisis de correspondencias comienza con **datos** sobre la abundancia de cada una de varias **características** para cada uno de un conjunto de **objetos**.

Esto es útil en ecología, por ejemplo, donde los objetos de interés son a menudo sitios diferentes, las características son especies diferentes, y los datos consisten en abundancias de las especies en muestras tomadas de los sitios. El propósito del análisis de correspondencias sería entonces aclarar **las relaciones entre los sitios** tal como se expresan por las distribuciones de especies, y **las relaciones entre las especies** expresadas por las distribuciones del sitio.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## MULTINORMALIDAD

---

La correcta aplicación de cualquiera de las técnicas del análisis multivariado está sujeta a dos hipótesis básicas:

- ❑ La matriz de datos está exenta de errores.
- ❑ Las variables están distribuidas normalmente. Más precisamente, los datos bajo análisis deben cumplir que ajustan a una **distribución normal multivariante**.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## MULTINORMALIDAD

---

En particular, si todas las variables individuales estudiadas parecen estar normalmente distribuidas, entonces se supone que la distribución conjunta es **normal multivariante**.

Este es, de hecho, un requisito mínimo porque la definición de normalidad multivariada requiere más que esto.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## MULTINORMALIDAD

---

En general, las variables que se miden en la **naturaleza** (datos hidrológicos, hidrogeológicos, hidroquímicos) **se distribuyen de forma no normal**. En estas condiciones, no es estrictamente válido utilizar estos datos en modelos estadísticos de análisis de varianza. No obstante, la aplicación de procedimientos estadísticos paramétricos que involucran regresiones lineales de datos no distribuidos normalmente no necesariamente conduce a resultados erróneos, lo cual no significa que no deba efectuarse algún intento de cumplir con las hipótesis inherentes a esos métodos.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## ALGORÍTMOS PARA LA NORMALIZACIÓN DE DISTRIBUCIONES DE FRECUENCIA

---

**Transformación de Box-Cox:** 
$$Y_i = \frac{(X_i^\lambda - 1)}{\lambda} \quad [1]$$

Algoritmo de transformación en el que  $\lambda$  es la constante de transformación o potencia normalizante.

$$\lambda \rightarrow 0, \quad Y_i \rightarrow \text{Ln}(X_i)$$

$$\lambda = 1, \quad Y_i = X_i - 1$$

El primer caso representa la **transformación logarítmica** de los datos.

En el segundo, apenas se afecta el primer momento de los datos, equivale a **no transformación**.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## SOFTWARE

---

- BMDP
- SAS [www.sas.com](http://www.sas.com)
- SPSS [www.spss.com](http://www.spss.com)
- XLSTAT [www.xlstat.com](http://www.xlstat.com)
- NCSS [www.ncss.com](http://www.ncss.com)
- MATLAB
- R (software libre) [www.r-project.org/](http://www.r-project.org/)

La disponibilidad de software condujo a un aumento en la variedad de métodos gráficos disponibles para datos multivariados, que se utilizan con mayor frecuencia como parte de su análisis con métodos multivariados.

---

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.



## SOFTWARE

**Table A1** Some Statistical Packages with Information about the Multivariate Analyses that They Provide, an Assessment of How Easy the Packages Are to Use for Someone New to Multivariate Analysis, and an Indication of the Graphical Facilities in the Package

Package	Ease of Use <sup>a</sup>	Source of Information	Graphs <sup>b</sup>	Tests and Analyses Available on Menus													
				Tests of Significance <sup>c</sup>	Distance <sup>d</sup>	Mantel Test	PCA <sup>e</sup>	FA <sup>f</sup>	DFA <sup>g</sup>	CA <sup>h</sup>	Log. Reg <sup>i</sup>	CCA <sup>j</sup>	MDS <sup>k</sup>	PCCA <sup>l</sup>	CORA <sup>m</sup>		
GenStat	*	www.vsn-intl.com	**	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MINITAB 14	**	www.minitab.com	***	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes
MVSP 3.1	*	www.kovcomp.com	*	No	Yes	No	Yes	No	No	Yes	No	No	No	No	Yes	Yes	Yes
NCSS 2004	**	www.ncss.com	***	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
SPSS 12	**	www.spss.com	***	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Stata 8.0	*	www.stata.com	***	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
Statistica 6.1	**	www.statsoft.com	***	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

<sup>a</sup> Ease of use for beginners: \* = reasonably easy after some practice; \*\* = the easiest.

<sup>b</sup> Graphics capabilities: \* = limited; \*\* = reasonably good; \*\*\* = best.

<sup>c</sup> Tests of significance as discussed in Chapter 4, although none of the packages have all of these tests as menu items.

<sup>d</sup> Calculation of distances as discussed in Chapters 5 and 9. Generally only some of these distances are available.

<sup>e</sup> Principal components analysis.

<sup>f</sup> Factor analysis.

<sup>g</sup> Discriminant function analysis.

<sup>h</sup> Cluster analysis.

<sup>i</sup> Logistic regression.

<sup>j</sup> Canonical correlation analysis.

<sup>k</sup> Multidimensional scaling.

<sup>l</sup> Principal coordinates analysis.

<sup>m</sup> Correspondence analysis.

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.

## Datos y métodos multivariados - RESUMEN

- ❑ Introducción de 5 conjuntos de datos multivariados, que se utilizarán en el curso.
- ❑ Breve descripción de los métodos multivariados que se desarrollarán en el curso.
- ❑ Condición que deben cumplir los datos multivariados: distribución normal multivariante.
- ❑ Introducción de software y la importancia de los métodos gráficos.

AMARN 2018 - IMFIA.FI.UDELAR -  
Ing. Luis Silveira, Ph.D.