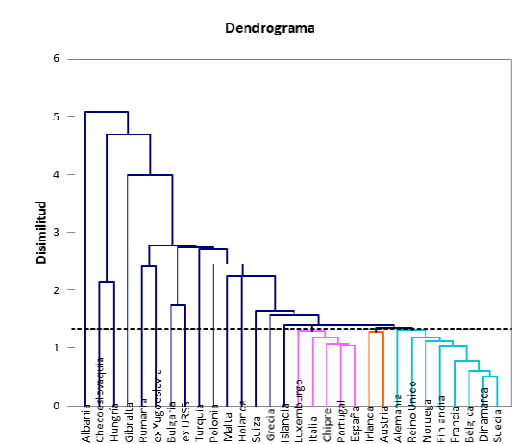


APLICACIONES DEL AC

3) Dendrograma

- A medida que la distancia aumenta, más y más países se combinan, y la fusión finaliza con Albania uniéndose a todos los demás países en un grupo.

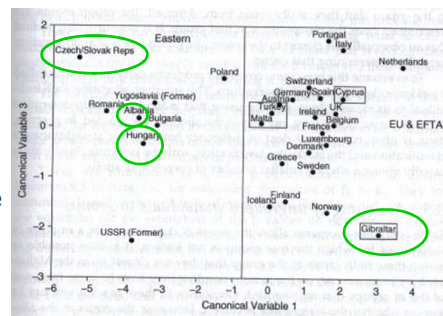


AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL AC

3) Dendrograma

Una interpretación del dendrograma es que solo hay **cuatro conglomerados definidos por el vecino más cercano**. Estos son entonces **(1) Albania, (2) Hungría y Checoslovaquia, (3) Gibraltar y (4) todos los demás países**. Esto separa tres países del este y Gibraltar de todos los demás, lo que sugiere que la clasificación en la UE, la EFTA, el este y otros países no es un buen indicador de los patrones de empleo. **Esto contradice la separación razonablemente exitosa de los países del este y de la UE / EFTA realizada con el AFD.**

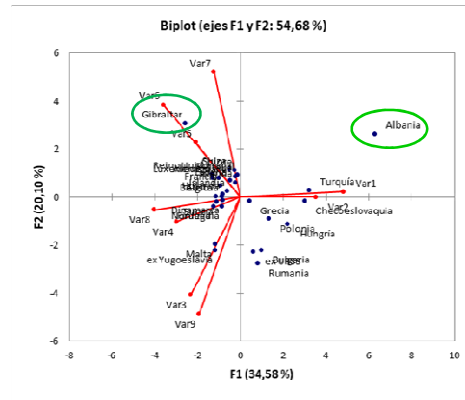


AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL AC

3) Dendrograma

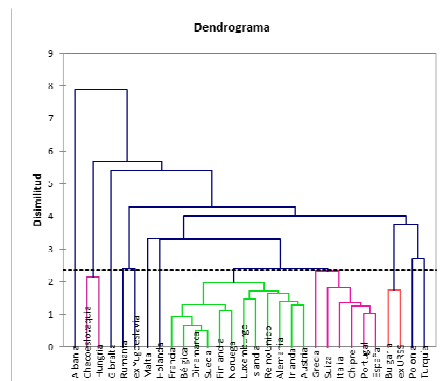
Sin embargo, hay un acuerdo limitado con la separación de países en función de los dos primeros **componentes principales**, donde Albania y Gibraltar muestran valores de datos muy extremos.



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

APLICACIONES DEL AC

3) Dendrograma Utilizando otro método disponible en XLSTAT: Promedio no ponderado par-grupo



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

APLICACIONES DEL AC

Ej. 8.2 Relaciones entre especies caninas

Datos: Tabla 1.4 (Medidas de 7 grupos caninos)

Objetivo del A.C.: Analizar la relación entre el perro prehistórico y las restantes especies.

Recordar que fue tratado en el Ejemplo 4.1

- Medidas estandarizadas (Tabla 4.1)
- Distancias Euclideanas (Tabla 4.2)

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

APLICACIONES DEL AC

Ej. 8.2 Relaciones entre especies caninas

Clústeres encontrados a diferentes niveles de distancia para un análisis jerárquico de vecino más cercano:

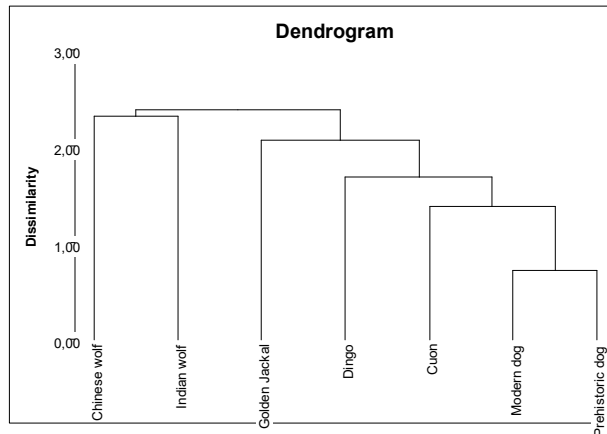
Distancia	Cluster	Nro. de Clusters
0,00	MD, PD, GJ CW, IW, CU, DI	7
0,72	(MD, PD), GJ CW, IW, CU, DI	6
1,38	(MD, PD, CU), GJ CW, IW, DI	5
1,63	(MD, PD, CU), GJ CW, IW, DI	5
1,68	(MD, PD, CU, DI), GJ CW, IW	4
1,80	(MD, PD, CU, DI), GJ CW, IW	4
1,84	(MD, PD, CU, DI), GJ CW, IW	4
2,07	(MD, PD, CU, DI, GJ), CW, IW	3
2,31	(MD, PD, CU, DI, GJ), (CW, IW)	2
2,37	(MD, PD, GJ CW, IW, CU, DI)	1

MD=modern dog, GJ=Golden jackal, CW=Chinese Wolf, IW=Indian wolf,
CU=Cuon, DI=Dingo, PD=prehistoric dog

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

APLICACIONES DEL AC

El perro prehistórico parece estar relacionado con el perro moderno



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

APLICACIONES DEL AC

- **Resultado:** El **perro prehistórico** se relaciona estrechamente con el **perro tailandés moderno**, con éstos dos algo relacionados con las especies **cuon** y **dingo** y menos relacionado con el **chacal dorado**. Los **lobos indios y chinos** están próximos entre si, pero la diferencia entre ellos es relativamente grande.
- **Conclusión:** El análisis de conglomerados ha producido una descripción sensata de la relación entre los grupos diferentes.

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

SOFTWARE-BIBLIOGRAFÍA

SOFTWARE

Prácticamente todos los paquetes computacionales incluyen AC.

BIBLIOGRAFÍA

Hartigan, J. (1975) *Clustering Algorithms*, Wiley, New York.

Romesburg, H.C. (1984) *Cluster Analysis for Researchers*, Lifetime Learning Publications, Belmont, CA.

Everitt, B., Landau, S., and Leese, M. (2001) *Cluster Analysis*, 4th ed., Edward Arnold, London.

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

AC - RESUMEN

- ❖ Se analizaron los motivos para realizar un A.C. Estos incluyen la definición de grupos subyacentes verdaderos y la identificación de una pequeña cantidad de objetos (uno por grupo) que representen el conjunto completo de objetos.
- ❖ Se describieron dos tipos de análisis de conglomerados. Uno resulta en un **dendrograma**, que muestra una **relación jerárquica entre los objetos** según sus similitudes. El otro implica un procedimiento iterativo de **particionamiento**, para hallar el mejor conjunto de **n clusters** para un conjunto de datos, comenzando con clusters arbitrarios, para luego mejorarlos al mover individuos entre ellos.

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

AC - RESUMEN

- ❖ Existe una variedad de algoritmos de **agrupación jerárquica**. Se describieron los basados en las distancias del **vecino más cercano**, las distancias del **vecino más distante** y la **media del grupo**. Estos comienzan con todos los objetos en grupos por sí solos y gradualmente los fusionan en un único grupo.
- ❖ Los **métodos jerárquicos divisivos** se describieron brevemente, puesto que no se utilizan tan a menudo como los métodos aglomerativos. Los métodos divisivos comienzan con todos los objetos en un único grupo y gradualmente los separan hasta que cada objeto está en un grupo por sí mismo.

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

AC - RESUMEN

- ❖ Se analizaron los problemas en detectar clusters con formas inusuales.
- ❖ Se analizaron las medidas de distancias entre objetos, y la distancia euclidiana en particular. Se discutió la necesidad de estandarización de variables.
- ❖ El A.C. que utiliza A.C.P. se usa a veces para reducir los tiempos de cálculo. Se sugiere evitarlo.
- ❖ Se manejaron un par de ejemplos utilizando los datos presentados en la sección 1.
- ❖ Se discutió brevemente el software disponible para llevar a cabo el A.C.

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

AC - RESUMEN

- ❖ Se manejaron fuentes de información adicional sobre el A.C. y se identificaron algunos enfoques alternativos para agrupar.

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

AC con XLSTAT

¿Cómo realizar una Clasificación Ascendente Jerárquica (CAJ) con XLSTAT?

Datos: Son los utilizados en el tutorial del A.C.P.

US Census Bureau (http://eire.census.gov/popest/states_dataset.csv)

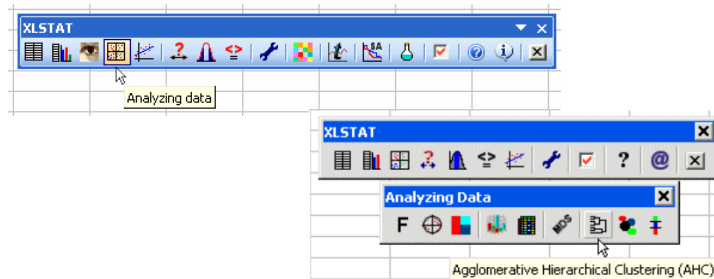
Medición de parámetros demográficos en 51 Estados de los Estados Unidos. La aplicación utiliza solo los datos del año 2001. Además, con el fin de suprimir los efectos de escala, las variables iniciales fueron convertidas en índices por 1000 habitantes.

Objetivo: Crear grupos homogéneos de estados.

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

AC con XLSTAT

- ❑ Inicie XLSTAT-Pro y seleccione en el menú XLSTAT/ Análisis de datos/ Clasificación Ascendente Jerárquica (CAH), o haga clic en el botón correspondiente de la barra "Análisis de datos":

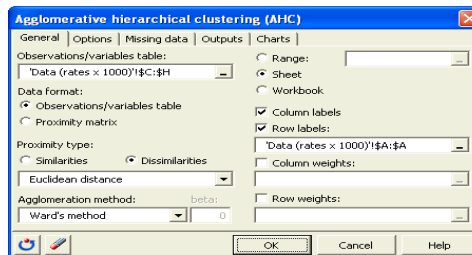


AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

AC con XLSTAT

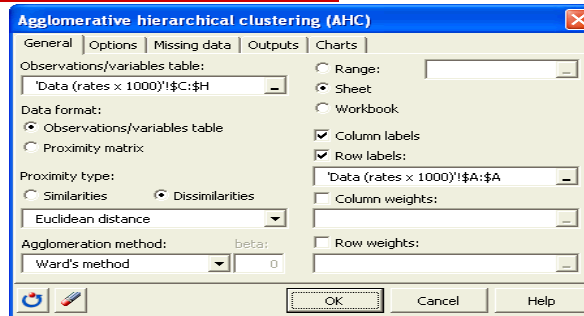
Cuadro de diálogo:

Existen varias maneras de seleccionar los datos en el cuadro de diálogo de XLSTAT (consultar el tutorial). En el ejemplo los datos comienzan desde la primera línea, por lo que es más rápido elegir el modo de selección por columnas. Por esta razón, en el cuadro de diálogo la selección aparece en forma de columnas.



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

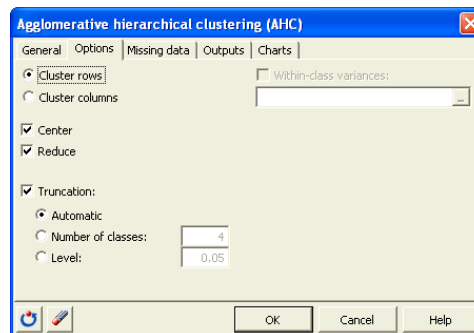
AC con XLSTAT



- La variable "Población total" no fue seleccionada ya que solo interesan los aspectos dinámicos de la población. La última columna tampoco fue seleccionada puesto que está correlacionada con la anterior (ver ACP).

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

AC con XLSTAT



- Se activó la opción "Centrar" "Reducir" para evitar la creación de grupos por efectos de escala.

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