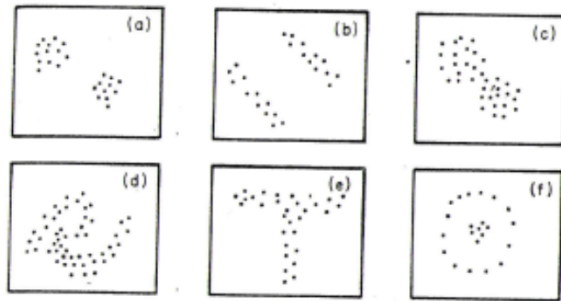


PROBLEMAS DEL AC

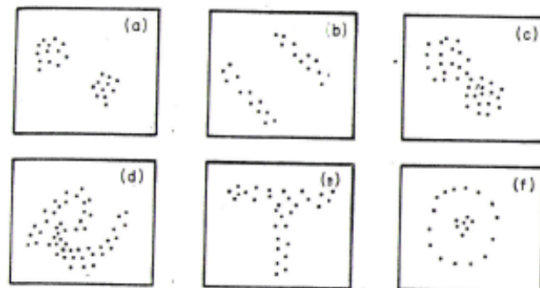
- ✓ **(c):** Algunos algoritmos son incapaces de detectar dos conglomerados debido a los puntos intermedios.



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

PROBLEMAS DEL AC

- ✓ **(d), (e) y (f):** La mayoría de los algoritmos tienen problemas al manejar casos como estos.



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

MEDIDAS DE DISTANCIA

- Los datos consisten, generalmente, en valores de ***p* variables** X_1, X_2, \dots, X_p y ***n* observaciones**. En los algoritmos jerárquicos, estos valores producen una matriz de distancia entre cada una de las observaciones.
- La distancia euclídea $d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$ es la más frecuentemente utilizada para variables cuantitativas, donde k representa la variable y i, j las observaciones
- Menú de distancias de XLSTAT: Euclídea, Euclídea cuadrada, Mahalanobis, Bhattacharya, Chi-cuadrado, Manhattan, Chebychev, Canberra, Chord, Bray y Curtis, etc.

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

MEDIDAS DE DISTANCIA

- **Estandarización:** Usualmente las variables se estandarizan antes de calcular las distancias, de modo que las ***p* variables** tengan la misma importancia en su determinación.
- **Modos de estandarización:**
 - 1) $m = 0, s^2 = 1$
 - 2) Codificación: mínimo = 0 y máximo = 1

Obs! La estandarización puede minimizar la diferencia entre los grupos.

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

ANÁLISIS DE COMPONENTES PRINCIPALES CON ANÁLISIS DE CONGLOMERADOS

- Algunos algoritmos para el AC comienzan efectuando un ACP para reducir las variables originales a un número menor de componentes principales.
- Es necesario tener presente que los resultados de un AC pueden ser bastante diferentes, con y sin, el ACP inicial. En consecuencia, probablemente deba evitarse un ACP inicial, puesto que el tiempo de computo rara vez es un problema en la actualidad.

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

ANÁLISIS DE COMPONENTES PRINCIPALES CON ANÁLISIS DE CONGLOMERADOS

- No obstante, cuando los primeros dos C.P. explican un alto porcentaje de la varianza, una representación gráfica de las observaciones en función de los C.P. es ciertamente un modo útil para identificar los conglomerados (o racimos).

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

APLICACIONES DEL AC

Ej. 8.1 Conglomerados de países europeos

- ✓ **Datos:** Tabla 1.5 (Porcentaje de personas empleadas en 9 grupos industriales, en los países europeos)
- ✓ **Objetivo del AC:** Mostrar que países tienen patrones similares de empleo y que países son diferentes.

Existe un agrupamiento político: (1) UE, (2) EFTA Zona de Libre Comercio Europea, (3) países del Este) y (4) Chipre, Gibraltar, Malta y Turquía.

INTERÉS: Comprobar si el A.C. puede reconstruir este agrupamiento.

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

APLICACIONES DEL AC

1) Estandarización de los datos

GRUPO	AGR	MIN	MAN	PS	CON	SER	FIN	SPS	TC
EU Bélgica	-0.78	-0.37	0.05	0.00	-0.45	0.24	0.51	1.13	0.28
EU Dinamarca	-0.54	-0.38	0.01	-0.16	-0.41	-0.22	0.61	1.07	0.44
EU Francia	-0.58	-0.35	-0.01	0.16	-0.16	0.21	0.89	0.70	-0.04
EU Alemania	-0.73	-0.31	0.48	0.32	0.68	0.30	0.74	0.16	-0.69
EU Grecia	0.81	-0.33	-0.11	0.32	-0.27	0.50	-0.34	-0.82	0.36
EU Irlanda	0.13	-0.32	-0.05	0.64	-0.16	0.42	0.44	-0.17	-0.53
EU Italia	-0.31	-0.26	0.17	-1.29	0.57	1.16	-0.51	0.12	-0.94
EU Luxemburgo	-0.72	-0.38	-0.07	-0.16	0.87	1.08	0.51	0.30	0.28
EU Holanda	-0.65	-0.38	-0.11	-0.16	-2.54	0.55	1.22	1.29	0.28
EU Portugal	-0.06	-0.33	0.35	-0.16	0.25	0.81	-0.09	-0.27	-1.34
EU España	-0.19	-0.33	0.09	-0.32	0.72	0.86	-0.19	-0.03	-0.53
EU Reino Unido	-0.81	-0.31	0.11	0.64	-0.19	0.88	1.44	0.16	0.04
EFTA Austria	-0.39	-0.35	0.70	0.64	0.35	0.67	0.01	-0.42	-0.04
EFTA Finlandia	-0.30	-0.37	-0.10	0.64	-0.27	-0.20	0.49	0.71	0.85
EFTA Islandia	-0.14	-0.39	-0.17	0.16	0.90	-0.22	0.34	0.42	0.20
EFTA Noruega	-0.52	-0.26	-0.60	0.48	-0.38	0.38	0.24	1.20	1.34
EFTA Suecia	-0.73	-0.35	-0.14	0.00	-0.41	-0.28	0.69	1.43	0.61
EFTA Suiza	-0.54	-0.39	0.47	-1.29	0.61	0.94	1.02	-0.45	-0.21
Este Albania	3.52	1.80	-2.15	-1.29	-1.51	-2.39	2.17	-3.09	-2.80
Este Bulgaria	0.55	-0.39	1.56	-1.29	-0.30	-1.21	-1.29	-0.70	0.85
Este Checoslovaquia	0.05	3.82	-2.15	-1.29	0.32	-1.05	-1.27	-0.47	0.36
Este Hungría	0.25	2.87	-2.15	-1.29	-0.41	-0.45	-1.67	0.04	1.90
Este Polonia	0.93	0.05	0.40	0.16	-0.45	-1.03	-1.34	-0.29	-1.02
Este Rumania	0.80	-0.10	1.86	1.93	-0.63	-1.69	-1.52	-1.34	0.28
Este ex URSS	0.51	-0.39	0.90	-1.29	0.98	-1.50	-1.52	-0.16	1.58
Este ex Yugoslavia	-0.58	-0.14	1.95	2.25	0.21	-0.36	-0.89	-0.90	1.09
Otro Chipre	0.11	-0.35	-0.14	-0.48	0.57	1.56	0.01	-0.66	-0.37
Otro Gibraltar	-0.99	-0.39	-1.43	1.93	3.43	1.72	1.04	0.80	-1.18
Otro Malta	-0.78	-0.32	0.81	1.13	-1.07	-1.05	-0.69	1.67	0.61
Otro Turquía	2.65	-0.29	-0.53	-0.97	-0.85	-0.63	-1.07	-1.43	-1.66

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

APLICACIONES DEL AC

2) Cálculo de distancias Euclídeas

Matriz de similitud (Distancia euclídea):

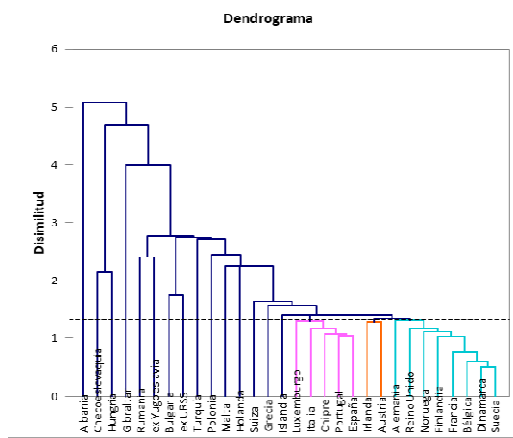
	Albani	Armenia	Francia	Países Bajos	Polonia	Rusia	Ucrania	Reino Unido	Eslovenia	Eslovaquia	Letonia	Lituania	Malta	Países Bajos	Polonia	Rusia	Ucrania	Reino Unido	Eslovenia	Eslovaquia	Letonia	Lituania	Malta
Albani	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Armenia	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Francia	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Países Bajos	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Polonia	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rusia	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ucrania	0.000	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Reino Unido	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Eslovenia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Eslovaquia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Letonia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lituania	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Malta	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

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

APLICACIONES DEL AC

3) Dendrograma

- Los países más cercanos son Suecia y Dinamarca ($d_e = 0,5$)
- A una distancia ligeramente mayor, Bélgica se une a estos dos países para crear un grupo de tres.



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