

Análisis Multivariado I

Ejercicio 1

Señale cuál de las siguientes afirmaciones es verdadera y cual es falsa.
Fundamente su respuesta.

En la clasificación jerárquica agregativa:

- a. Se agrupa a los individuos de modo que si dos individuos se unen en la etapa *j* éstos permanecerán en el mismo cluster hasta el final del proceso.
- b. La forma de definir distancias entre dos clusters no incide en el resultado de la clasificación.
- c. Las observaciones que se unen en la primera etapa no dependen del algoritmo utilizado.
- d. En el vecino más lejano las observaciones que se unen en la primera etapa son aquellas que tienen mayor distancia.
- e. La distancia de Mahalanobis se recomienda para el caso de variables cuantitativas correlacionadas.
- f. El procedimiento comienza con todas las observaciones en un grupo y termina con *I* grupos compuestos por una observación cada uno.

En la clasificación no jerárquica

- a. En la clasificación no jerárquica, al ser una clasificación no supervisada, no se sabe a priori la cantidad de grupos.
- b. Los métodos de *kmeans* necesitan definir a priori la cantidad de grupos y el vector inicial de *centroides*.
- c. El método de clasificación *fuzzy* requiere de la cantidad de grupos y la definición de individuos representativos.
- d. En el método de clasificación *fuzzy* se obtiene, como resultado del procedimiento, los coeficientes de pertenencia de las observaciones a cada grupo.

Ejercicio 2

Para **cada uno** de los siguientes ítems indicar **verdadero o falso**. Fundamente su respuesta.

Dada la siguiente matriz de distancia:

	A	B	C	D
A	0	6	3	2
B		0	5	8
C			0	4
D				0

Al aplicar clasificación jerárquica y el algoritmo del vecino más cercano:

- Los primeros individuos que se unen son A y C.
- En la penúltima etapa los cluster formados son {A, C, D} y {B}.
- En la etapa final, el nivel al que se unen los dos últimos cluster es 8.
- La distancia del cluster {A C, D} al individuo {B} es 6.

Ejercicio 3

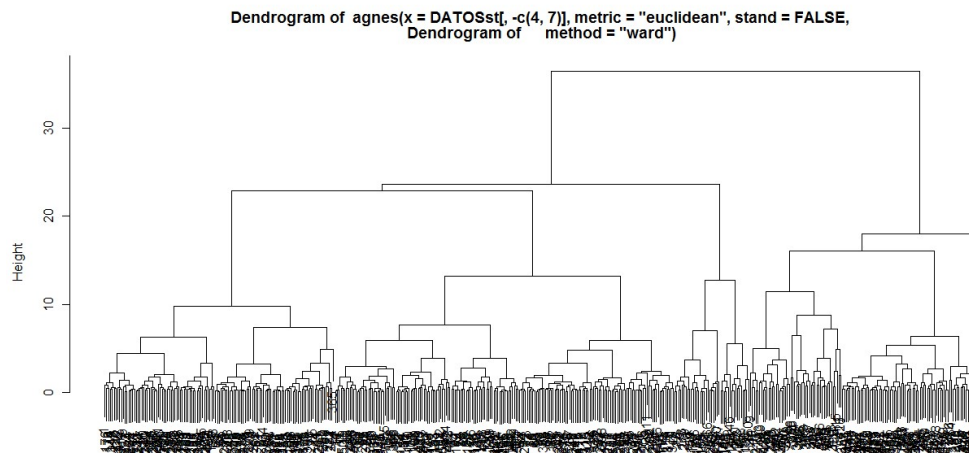
Se busca hacer una tipología de individuos por lo cual se desea aplicar la técnica de análisis de cluster. Describa que estrategia llevaría adelante ante la presencia de:

- Observaciones atípicas. Existen dos observaciones atípicas, con valores extremos en la gran mayoría de las variables.
- Observaciones atípicas. Existen dos observaciones atípicas con valores extremos en una variable.
- La totalidad de variables que se tienen son variables binarias.
- J-1 variables cuantitativas y una variable cualitativa.

Ejercicio 4

Se busca construir una tipología de estudiantes avanzados en función de variables asociadas al rendimiento. **Para ello se aplico el método de ward.**

En función de la información que se proporciona **¿Que estructura de grupos recomendaría? Fundamente su respuesta. Indique que elementos toma en cuenta y como se utilizan los mismos.**



DATOSst[, -c(4, 7)]
Agglomerative Coefficient = 0.99

.	history		Freq	Rcuad	psF	psT
454	448	431	58	0.8572806	136.3534	21.16414
455	415	435	16	0.8507639	136.5190	19.58387
456	450	427	84	0.8435296	136.5717	32.84033
457	434	445	63	0.8362189	137.2536	31.54792
458	449	439	60	0.8280397	137.8378	35.10405
459	454	433	72	0.8196133	139.0355	22.53015
460	395	428	11	0.8107149	140.7283	18.86347
461	443	361	25	0.8003392	142.1474	37.30615
462	446	451	17	0.7893744	144.2888	12.07381
463	437	452	65	0.7779886	147.4981	44.26657
464	457	447	100	0.7656963	151.6336	40.21023
465	460	462	28	0.7495759	154.6500	11.10848
466	458	463	125	0.7294293	157.0357	50.28126
467	453	465	50	0.7020099	157.1666	19.21117
468	461	455	41	0.6680655	156.9861	47.36788
469	464	456	184	0.6312101	160.5453	99.30165
470	467	459	122	0.5765631	159.9911	50.23662
471	470	442	125	0.5081371	162.1946	44.83705
472	466	469	309	0.3978791	155.9479	194.72134
473	472	468	350	0.2802043	184.1309	118.36041
474	473	471	475	0.0000000	NaN	184.13094