Se puede argumentar, igual que como se hizo para la ecuación (a) en el single Linkage, que para el caso del algoritmo complete Linkage

$$d(R,\{P,Q\}) = \max\{d(R,P), d(R,Q)\},$$

razón por la cual también se le conoce como el "algoritmo del vecino más lejano". En principio complete Linkage producirá grupos de forma que todos los individuos en un grupo "son parecidos"

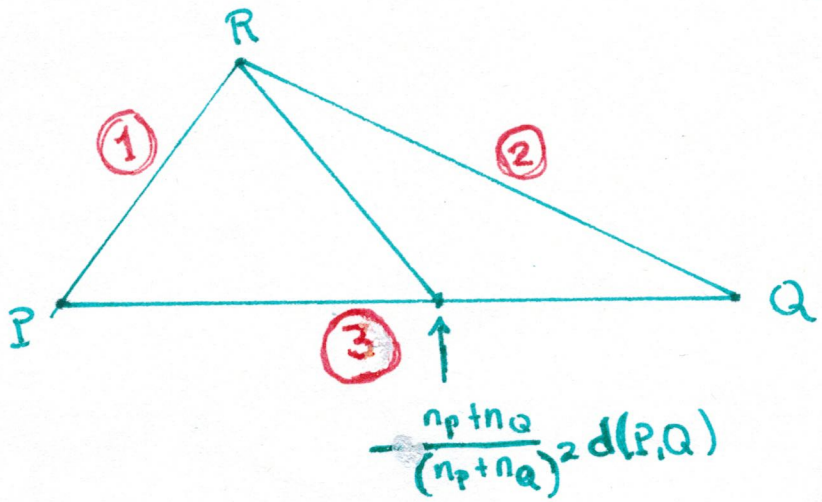
El algoritmo average Linkage (weighted o unweighted) propone un punto intermedio entre el single Linkage y el complete Linkage, ya que para este algoritmo se calcula la distancia promedio

$$d(R,\{P,Q\}) = \frac{n_P}{n_P+n_Q} d(R,P) + \frac{n_Q}{n_P+n_Q} d(R,Q)$$

El algoritmo del centroide es muy similar al de average Linkage pero tiene una corrección que corresponde a una proporción de la distancia entre P y Q

$$d(R,\{P,Q\}) = \frac{n_P}{n_P+n_Q} d(R,P) + \frac{n_Q}{n_P+n_Q} d(R,Q) - \frac{n_P n_Q}{(n_P+n_Q)^2} d(P,Q)$$

(1) (2) (3)



El algoritmo de agrupamientos de Ward, cuyos pesos se enuncian en la tabla 2, tiene una forma más sofisticada[1], ya que decide unir dos grupos sólo si al calcular una "medida de heterogeneidad" asignada al nuevo grupo (la unión de los dos grupos), esta medida no es "muy grande". En otras palabras el algoritmo une a dos grupos solo si el grupo resultante es tan homogéneo como "sea posible".

La heterogeneidad de un grupo R se mide usando la "inercia"

$$I_R = \frac{1}{n_R} \sum_{i=1}^{n_R} d^2(x_{Gi}, \bar{x}_{G_R}),$$

(1) Yo diría inteligente

donde $\overline{xG}_R = \frac{1}{n} \sum_{i=1}^{n_R} xG_i \qquad R = \{xG_1, \ldots, xG_{n_R}\}$

I_R es una medida de dispersión del grupo al rededor de su centro de gravedad ($\overline{xG}_R$). Si usamos d = distancia euclideana, entonces I_R representa la suma de las varianzas de las p componentes $x_{i1}, \ldots, x_{ip}$ de xG_i ; $i = 1, 2, \ldots, n_R$.

Cuando dos grupos P y Q son agrupados en $\{P, Q\}$, la inercia del nuevo grupo $\{P, Q\}$ se incrementa. Se puede probar que el correspondiente incremento está dado por

$$\Delta(P,Q) = \frac{n_P n_Q}{n_P + n_Q} d^2(P,Q).$$

El algoritmo de Ward une los grupos P y Q, sólo si $\Delta(P,Q)$ es mínimo.

Ejemplo Datos de comida en Francia

Se tienen registrados gastos (promedio) de comida para diferentes tipos de familias en Francia

Trabajadores Manuales = MA
Empleados = EM

Gerentes = CA

Además los datos se han registrado dependiendo del numero de hijos (2, 3, 4 ó 5).

	bread	veg.	fruits	meat	poultry	milk	wine
MA2	332	428	354	1437	526	247	427
EM2	293	559	388	1527	567	239	258
CA2	372	767	562	1948	927	235	433
MA3	406	563	341	1507	544	324	407
EM3	386	608	396	1501	558	319	363
CA3	438	843	689	2345	1148	243	341
MA4	534	660	367	1620	638	414	407
EM4	460	699	484	1856	762	400	416
CA4	385	789	621	2366	1149	304	282
MA5	655	776	423	1848	759	495	486
EM5	584	995	548	2056	893	518	319
CA5	515	1097	887	2630	1167	561	284

Haciendo un análisis de componentes principales normalizados obtenemos los siguientes porcentajes de varianza muestral explicada

Valor Propio	Proporcion de Var	Proporcion de Var acumulada
4.33	0.6190	61.9
1.83	0.2620	88.1
0.631	0.09	97.1
0.128	0.0180	98.9
0.058	0.0080	99.7
0.019	0.0030	99.9
0.001	0.0001	100

Procedemos a trabajar con las primeras dos componentes las cuales explican el 88.1 % de la varianza muestral

	r_{xiz_1}	r_{xiz_2}	$r^2_{xiz_1} + r^2_{xiz_2}$
X_1 = Pan	-0.499	0.842	0.957
X_2 = Vegetales	-0.970	0.133	0.958
X_3 = Frutas	-0.929	-0.278	0.941
X_4 = Carne	-0.962	-0.191	0.962
X_5 = Aves	-0.911	-0.266	0.901
X_6 = Leche	-0.584	0.707	0.841
X_7 = Vinos	0.428	0.648	0.604

$$Z_1 = -0.24X_1 - 0.466X_2 - 0.446X_3 - 0.462X_4 - 0.438X_5 - 0.281X_6 + 0.206X_7$$

$$Z_2 = 0.622X_1 + 0.09X_2 - 0.205X_3 - 0.141X_4 - 0.197X_5 + 0.523X_6 + 0.479X_7$$

La primera componente depende fuertemente de las cantidades gastadas en vegetales, frutas, carne y aves (mientras más grandes estos gastos, más pequeña es Z_1).