



Análise de Agrupamentos

Exemplo 12.3 Similarities of 11 languages - p677, ou exemplo 12.2 na

Example 12.2 (Measuring the similarities of 11 languages) The meanings of words change with the course of history. However, the meaning of the numbers 1, 2, 3, ... represents one conspicuous exception. Thus, a first comparison of languages might be based on the numerals alone. Table 12.2 gives the first 10 numbers in English, Polish, Hungarian, and eight other modern European languages. (Only languages that use the Roman alphabet are considered, and accent marks, cedillas, diereses, etc., are omitted.) A cursory examination of the spelling of the numerals in the table suggests that the first five languages (English, Norwegian, Danish, Dutch, and German) are very much alike. French, Spanish, and Italian are in even closer agreement. Hungarian and Finnish seem to stand by themselves, and Polish has some of the characteristics of the languages in each of the larger subgroups.

Table 12.2 Numerals in 11 Languages

English (E)	Norwegian (N)	Danish (Da)	Dutch (Du)	German (G)	French (Fr)	Spanish (Sp)	Italian (I)	Polish (P)	Hungarian (H)	Finnish (Fi)
one	en	en	een	eins	un	uno	uno	jeden	egy	yksi
two	to	to	twee	zwei	deux	dos	due	dwa	ketto	kaksi
three	tre	tre	drie	drei	trois	tres	tre	trzy	harom	kolme
four	fire	fire	vier	vier	quatre	cuatro	quattro	cztery	negy	neljä
five	fem	fem	vijf	funf	cinq	cinco	cinque	piec	ot	viisi
six	seks	seks	zes	sechs	six	seis	sei	szesc	hat	kuusi
seven	sju	syv	zeven	sieben	sept	siete	sette	siedem	het	seitseman
eight	atte	otte	acht	acht	huit	ocho	otto	osiem	nyolc	kahdeksan
nine	ni	ni	negen	neun	neuf	nueve	nove	dziewiec	kilenc	yhdeksan
ten	ti	ti	tien	zehn	dix	diez	dieci	dziesiec	tiz	kymmenen

The words for 1 in French, Spanish, and Italian all begin with *u*. For illustrative purposes, we might compare languages by looking at the *first letters* of the numbers. We call the words for the same number in two different languages *concordant* if they have the same first letter and *discordant* if they do not.

12.3: Considere o exemplo 12.3 acima e utilize métodos hierárquicos de agrupamento para criar grupos de idiomas.

12.6. The distances between pairs of five items are as follows:

	1	2	3	4	5
1	0				
2	4	0			
3	6	9	0		
4	1	7	10	0	
5	6	3	5	8	0

Cluster the five items using the single linkage, complete linkage, and average linkage hierarchical methods. Draw the dendrograms and compare the results.

12.11. Suppose we measure two variables X_1 and X_2 for four items A , B , C , and D . The data are as follows:

Item	Observations	
	x_1	x_2
A	5	4
B	1	-2
C	-1	1
D	3	1

Use the K -means clustering technique to divide the items into $K = 2$ clusters. Start with the initial groups (AB) and (CD) .

12.12. Repeat Example 12.11, starting with the initial groups (AC) and (BD) . Compare your solution with the solution in the example. Are they the same? Graph the items in terms of their (x_1, x_2) coordinates, and comment on the solutions.

12.13. Repeat Example 12.11, but start at the bottom of the list of items, and proceed up in the order D, C, B, A . Begin with the initial groups (AB) and (CD) . [The first potential reassignment will be based on the distances $d^2(D, (AB))$ and $d^2(D, (CD))$.] Compare your solution with the solution in the example. Are they the same? Should they be the same?