

# Cours 8

## CLASSIFICATION

Master 2 – 2006/2007

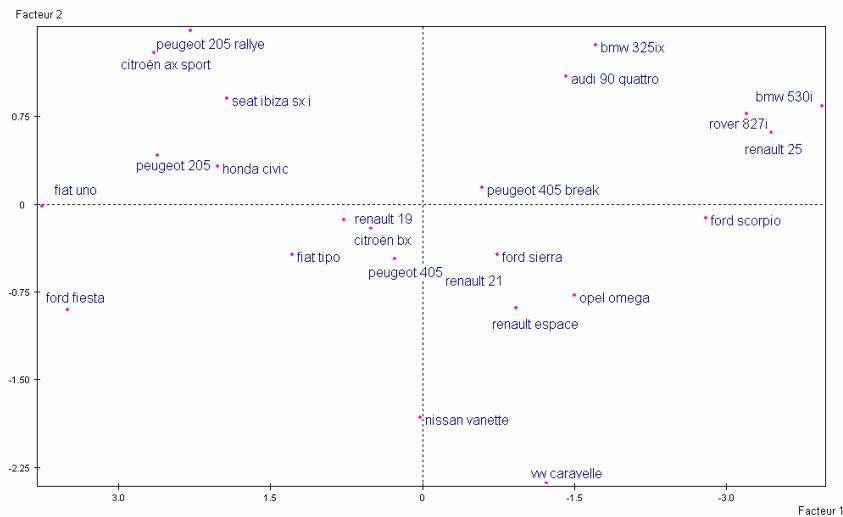
1

### Les données

Modèle	Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur
Honda Civic	1396	90	174	850	369	166
Renault 19	1721	92	180	965	415	169
Fiat Tipo	1580	83	170	970	395	170
Peugeot 405	1769	90	180	1080	440	169
Renault 21	2068	88	180	1135	446	170
Citroën BX	1769	90	182	1060	424	168
BMW 530i	2986	188	226	1510	472	175
Rover 827i	2675	177	222	1365	469	175
Renault 25	2548	182	226	1350	471	180
Opel Omega	1998	122	190	1255	473	177
Peugeot 405 Break	1905	125	194	1120	439	171
Ford Sierra	1993	115	185	1190	451	172
BMW 325ix	2494	171	208	1300	432	164
Audi 90 Quattro	1994	160	214	1220	439	169
Ford Scorpio	2933	150	200	1345	466	176
Renault Espace	1995	120	177	1265	436	177
Nissan Vanette	1952	87	144	1430	436	169
VW Caravelle	2109	112	149	1320	457	184
Ford Fiesta	1117	50	135	810	371	162
Fiat Uno	1116	58	145	780	364	155
Peugeot 205	1580	80	159	880	370	156
Peugeot 205 Rallye	1294	103	189	805	370	157
Seat Ibiza SXI	1461	100	181	925	363	161
Citroën AX Sport	1294	95	184	730	350	160

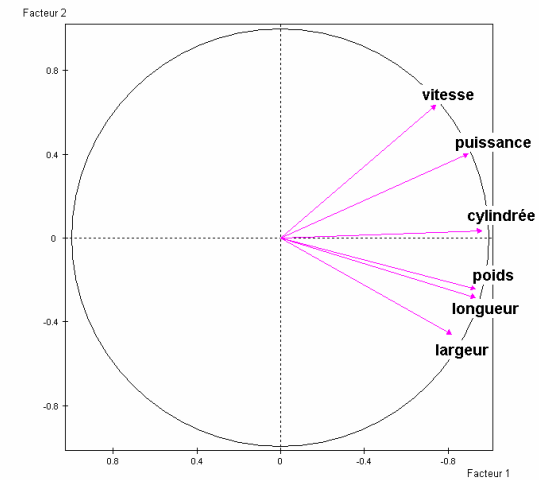
2

### Exemple Auto88 Le premier plan principal



3

### Exemple Auto88 Le cercle des corrélations



4

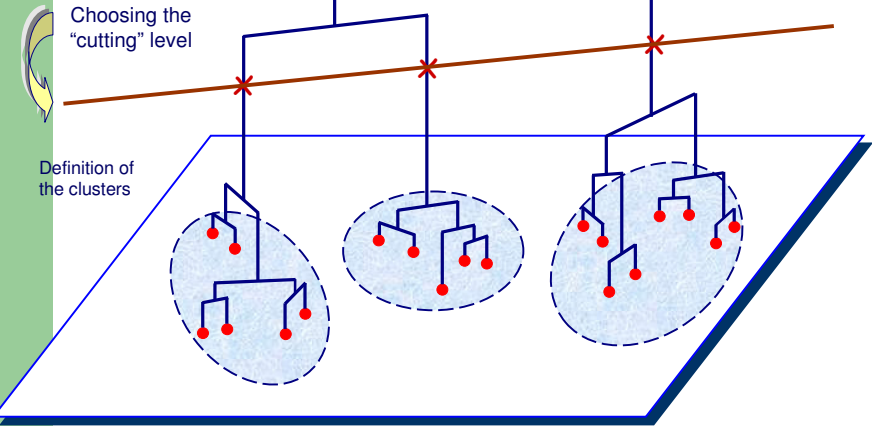
# La classification :

« Qui se ressemble s'assemble »

5

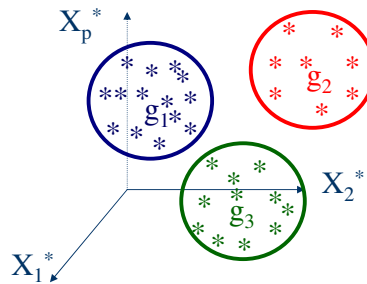
# Dendrogramme

18 groups



6

# Classification ascendante hiérarchique (Méthode de Ward)



$$\text{Distance de Ward : } D(G_i, G_j) = \frac{n_i n_j}{(n_i + n_j)} d^2(g_i, g_j)$$

7

Données centrées-réduites

	Modèle	Cylindrée*	Puissance*	Vitesse*	Poids*	Longueur*	Largeur*
1	Honda Civic	-.966	-.610	-.360	-1.133	-1.272	-.370
2	Renault 19	-.351	-.559	-.122	-.633	-.159	.022
3	Fiat Tipo	-.618	-.791	-.519	-.612	-.643	.152
4	Peugeot 405	-.260	-.610	-.122	-.134	.445	.022
5	Renault 21	.307	-.662	-.122	-.105	.591	.152
6	Citröen BX	-.260	-.610	-.043	-.221	.058	-.109
7	BMW 530i	2.046	1.917	1.702	1.733	1.220	.806
8	Rover 827i	1.456	1.633	1.543	1.104	1.147	.806
9	Renault 25	1.216	1.762	1.702	1.039	1.195	1.459
10	Opel Omega	.174	.215	.274	.626	1.244	1.067
11	Peugeot 405 Break	-.002	.292	.433	.040	.421	.283
12	Ford Sierra	.165	.034	.076	.344	.712	.414
13	BMW 325ix	1.114	1.478	.988	.821	.252	-.631
14	Audi 90 quattro	.166	1.195	1.226	.474	.421	.022
15	Ford Scorpio	1.945	.937	.671	1.017	1.074	.936
16	Renault Espace	.168	.163	-.241	.669	.349	1.067
17	Nissan Vanette	.087	-.688	-1.550	1.386	.349	.022
18	VW Caravelle	.384	-.043	-1.352	.908	.857	1.982
19	Ford Fiesta	-1.495	-1.642	-1.907	-1.306	-1.224	-.893
20	Fiat Uno	-1.497	-1.435	-1.510	-1.437	-1.393	-1.807
21	Peugeot 205	-.618	-.868	-.955	-1.002	-1.248	-1.677
22	Peugeot 205 rallye	-1.160	-.275	-.235	-1.328	-1.248	-1.546
23	Seat Ibiza sxi	-.843	-.352	-.083	-.807	-1.417	-1.023
24	Citröen AX sport	-1.160	-.481	.036	-1.654	-1.732	-1.154
Mean		.000	.000	.000	.000	.000	.000
Std. Dev.		1.000	1.000	1.000	1.000	1.000	1.000

Calculer la distance de Ward entre la BX et la 405 ?

8

## Tableau des distances entre les voitures

Proximity Matrix

Case	Squared Euclidean Distance								
	1: Honda Civic	2: Renault 19	3: Fiat Tipo	4: Peugeot 405	5: Renault 21	6: Citroën BX	...	23: Seat Ibiza sxi	24: Citroën AX sport
1: Honda Civic									
2: Renault 19	2.079								
3: Fiat Tipo	1.119	.534							
4: Peugeot 405	4.657	.626	1.748						
5: Renault 21	6.954	1.567	3.064	.419					
6: Citroën BX	3.270	.252	1.100	.181	.787				
...									
23: Seat Ibiza sxi	.713	2.992	2.453	5.423	7.665	3.766			.964
24: Citroën AX sport	1.309	5.582	4.676	9.283	12.400	7.183			

This is a dissimilarity matrix

$$D(\text{Citroën BX}, \text{Peugeot 405}) = \frac{1 \times 1}{(1+1)} \times .181 = .0905$$

9

## Classification Ascendante Hiérarchique

### Etape initiale

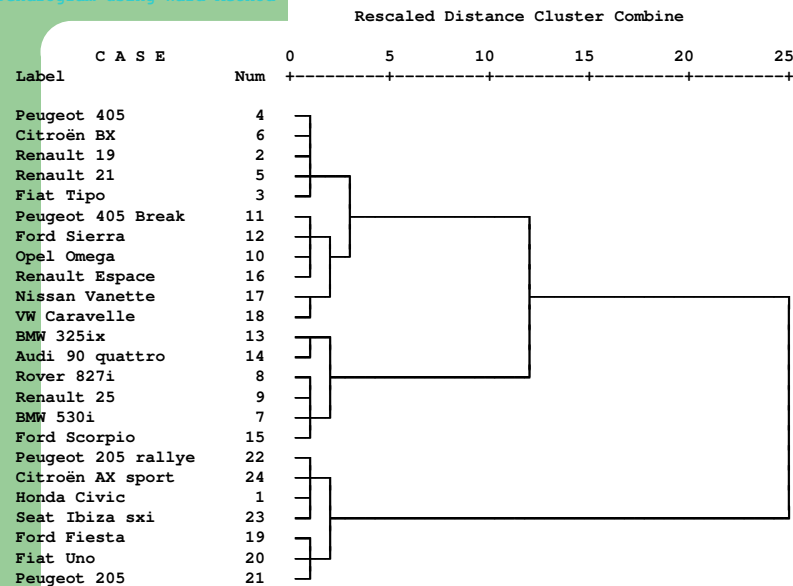
Chaque individu forme une classe. On regroupe les deux individus les plus proches.

### Etape courante

A chaque étape, on regroupe les deux classes  $G_i$  et  $G_j$  minimisant le critère de Ward  $D(G_i, G_j)$ .

10

## \*\*\*\*\* HIERARCHICAL CLUSTER ANALYSIS \*\*\*\*\* Dendrogram using Ward Method



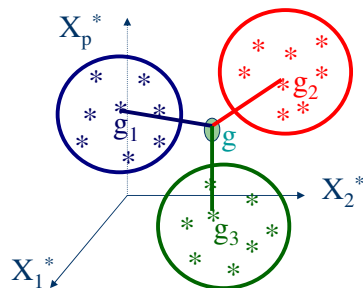
11

## Qualité de la typologie en K classes

- La somme des carrés expliquée par la typologie en K classes est égale à la somme des carrés inter-classes de la typologie en K classes.
- La qualité de la typologie est mesurée par la part de la somme des carrés totale expliquée par la typologie.

12

## Décomposition de l'inertie totale



$$I(N, g) = \sum_{i=1}^K \frac{n_i}{n} d^2(g_i, g) + \sum_{i=1}^K \frac{n_i}{n} I(G_i, g_i)$$

Inertie totale = p = Inertie inter-classes + Inertie intra-classes

13

Coefficient : Somme des carrés intra-classes de la typologie en K classes

Résultats SPSS : Somme des carrés intra-classes

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	6	.09035	0	0	3
2	11	12	.29804	0	0	12
3	2	4	.56054	0	1	9
4	8	9	.82705	0	0	10
5	22	24	1.11486	0	0	13
6	1	23	1.47116	0	0	13
7	10	16	2.00686	0	0	12
8	19	20	2.54783	0	0	15
9	2	6	3.15285	3	0	16
10	7	8	3.89960	0	4	14
11	13	14	4.69472	0	0	20
12	10	11	5.53718	7	2	18
13	1	22	6.45948	6	5	19
14	2	3	7.55715	9	0	21
15	19	21	8.91356	8	0	19
16	7	15	10.32859	10	0	20
17	17	18	12.76377	0	0	18
18	10	17	17.35471	12	17	21
19	1	19	22.53230	13	15	23
20	7	13	27.74035	16	11	22
21	2	10	34.47233	14	18	22
22	2	7	67.34381	21	20	23
23	1	2	138.00000	19	22	0

Part de somme des carrés totale expliquée par la typologie en K classes :  $(138 - \text{Coeff}[n-K])/138$

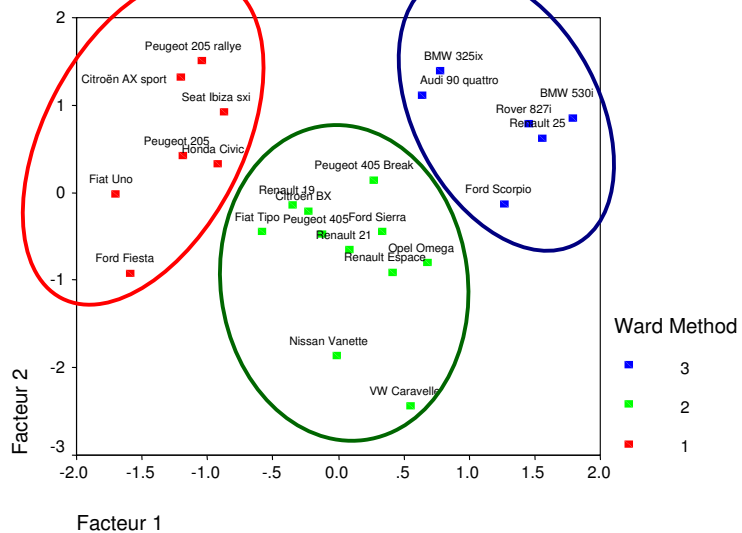
Part de somme des carrés totale expliquée par la typologie en 2 classes :  $(138 - 67.344)/138 = 0.512$

Somme des carrés intra-classes pour la typologie en K=2 classes

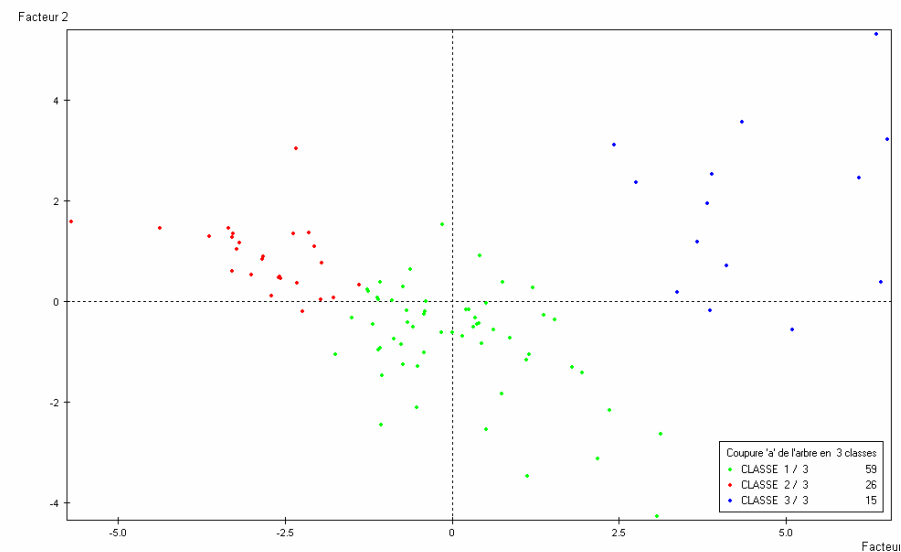
Somme des carrés totale =  $p*(n-1)$

14 Distance de Ward(4,6)

## Premier plan principal et typologie



15



16

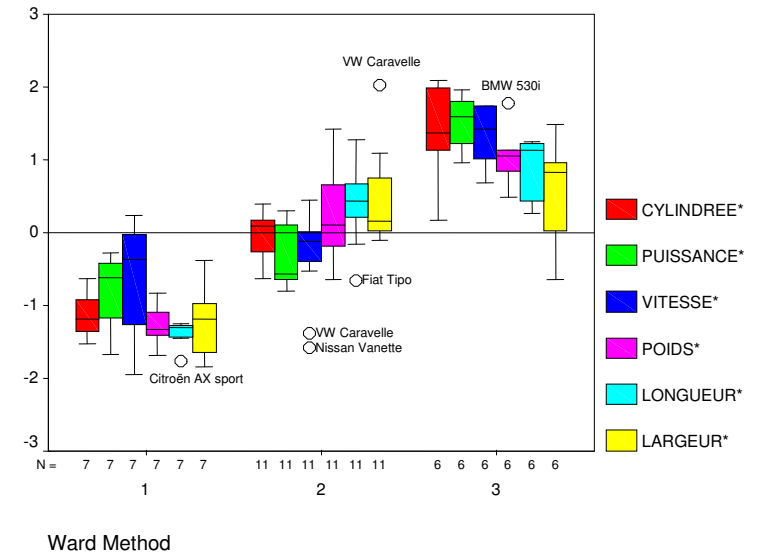
## Interprétation des classes

### Report

Mean	Ward Method			
	1	2	3	Total
Cylindrée	1322.57	1896.27	2605.00	1906.12
Puissance	82.29	102.18	171.33	113.67
Vitesse	166.71	175.55	216.00	183.08
Poids	825.71	1162.73	1348.33	1110.83
Longueur	365.29	437.45	458.17	421.58
Largeur	159.57	172.36	173.17	168.83

17

## Interprétation des classes



18