

ANALYSE DES DONNÉES

Fiche de TD n°7

La comparaison de trente caméscopes a été réalisée par le magazine Camera Vidéo pour son guide achat 2008. Sur chacun de ces caméscopes a été attribuée une note sur 10 pour les huit critères suivants : image, son, focales, ergonomie, automatismes, réglages, photos, et rapport qualité/prix (cf. tableau de données en annexe). On se propose, dans le cas présent, de réaliser une classification ascendante hiérarchique de ces trente caméscopes sur les notes obtenues dans les huit critères. Pour mener à bien ce travail, on décide de faire usage de la procédure CLUSTER de SAS avec la méthode de Ward. Au préalable, on va effectuer un travail exploratoire afin de s'appropriier les données, notamment celles qui sont associées aux variables actives.

Phase A : analyse exploratoire

1. A partir des statistiques simples (cf. Tab. 1), identifier : le critère qui est globalement le plus sévèrement noté, ainsi que le critère admettant les variations les plus importantes.

Critère	Moyenne	Ecart-type	Q1	Médiane	Q3
Image	6.55	0.888	6	6.5	7.3
Son	6.60	0.723	6	7	7
Focales	7.48	0.831	7	7.25	8
Ergonomie	8.05	0.472	8	8	8
Automatismes	7.75	0.668	7.63	8	8
Réglages	7.48	0.953	7	8	8
Photos	6.33	1.650	5	5.75	7.5
Qualité/Prix	8.17	0.860	8	8	9

Tab. 1 - Statistiques usuelles

2. A partir de la matrice des corrélations (cf. Tab. 2), préciser l'existence ou l'absence de corrélation entre les notes des différents critères. Interpréter alors selon le contexte.

	Image	Son	Focales	Ergonomie	Automatismes	Réglages	Photos	Qualité.Prix
Image	100.00	69.48	-14.55	6.92	22.77	64.51	50.55	23.29
Son	69.48	100.00	12.75	-8.79	-3.45	54.64	25.83	37.52
Focales	-14.55	12.75	100.00	-21.03	-11.26	-20.02	-54.87	12.05
Ergonomie	6.92	-8.79	-21.03	100.00	-22.49	31.70	32.12	-2.06
Automatismes	22.77	-3.45	-11.26	-22.49	100.00	-12.44	-7.56	13.07
Réglages	64.51	54.64	-20.02	31.70	-12.44	100.00	39.58	8.48
Photos	50.55	25.83	-54.87	32.12	-7.56	39.58	100.00	-1.57
Qualité.Prix	23.29	37.52	12.05	-2.06	13.07	8.48	-1.57	100.00

Tab. 2 - Matrice des corrélations

3. On envisage maintenant de réaliser une ACP normée. Préciser dans quel espace seront représentés les caméscopes. Combien doit-on trouver théoriquement d'axes factoriels ? Justifier votre réponse.
4. Dans le cadre d'une ACP normée, indiquer à quoi est égale la somme des valeurs propres. Compléter alors le tableau Tab. 3, puis préciser le nombre d'axes factoriels que l'on serait à conserver. Justifier votre réponse.

Valeur propre	Pourc. Inertie	Inertie cum.
2.7515	34.39	34.39
1.6962	21.20	55.60
.	15.34	70.94
0.8464	.	.
0.6667	8.33	89.85
0.4287	5.36	95.21
0.2249	2.81	98.02
0.1581	1.98	100.00

Tab. 3 - Tableau des inerties

5. A partir de la représentation graphique du nuage des observations (cf. Fig. 1) identifier des groupes de caméscopes. Visualise-t-on la présence de caméscopes atypiques ? Si oui, lesquels ?

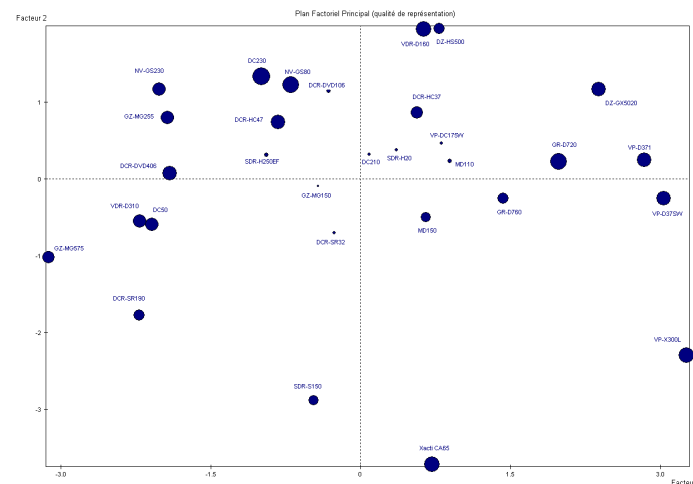


Fig. 1 - Nuage des observations

6. A partir de la représentation graphique du nuage des variables (cf. Fig. 2), proposer une interprétation de chacun des groupes de caméscopes préalablement identifiés. Proposer également une interprétation de la position des éventuels caméscopes atypiques. Que peut-on dire de la variable **Automatismes** ?

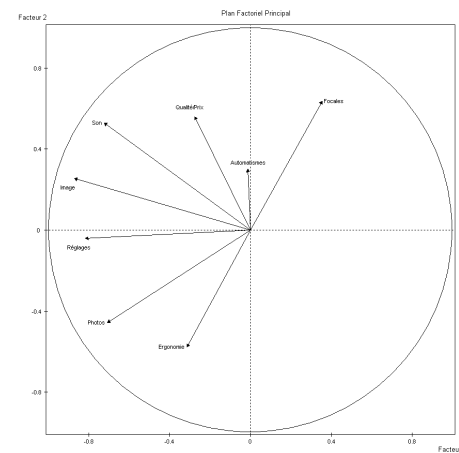


Fig. 2 - Nuage des variables

Phase B : classification ascendante hiérarchique

1. Via la procédure CLUSTER, on a obtenu l'historique suivant :

NCL	-Classifications jointes-	FREQ	SPRSQ	RSQ	BSS	Tie
29	VDR-D310 DC50	2	0.0001	1.00	0.0245	
28	NV-GS80 DC230	2	0.0007	.999	0.1532	
27	VP-D371 VP-D37SW	2	0.0028	.996	0.6541	
26	CL28 DCR-HC47	3	0.0030	.993	0.6983	
25	GR-D760 MD150	2	0.0031	.990	0.7098	
24	DCR-HC37 DC210	2	0.0039	.986	0.9047	
23	CL29 DCR-DVD4	3	0.0051	.981	1.1825	
22	GR-D720 CL25	3	0.0059	.975	1.373	
21	SDR-H20 GZ-MG150	2	0.0067	.969	1.5495	
20	NV-GS230 GZ-MG255	2	0.0079	.961	1.8276	
19	DCR-SR19 GZ-MG575	2	0.0081	.953	1.8906	
18	CL24 MD110	3	0.0109	.942	2.5209	
17	CL26 VDR-D160	4	0.0123	.930	2.8639	
16	CL21 DCR-SR32	3	0.0124	.917	2.8787	
15	DZ-GX502 DZ-HS500	2	0.0144	.903	3.3473	
14	CL23 SDR-H250	4	0.0163	.886	3.7839	
13	CL22 CL18	6	0.0173	.869	4.0064	
12	CL27 VP-DC175	3	0.0220	.847	5.1056	
11	CL17 CL20	6	0.0222	.825	5.1559	
10	CL15 DCR-DVD1	3	0.0272	.798	6.3061	
9	VP-X300L Xacti CA	2	0.0277	.770	6.4294	
8	CL13 CL16	9	0.0403	.730	9.34	
7	CL11 CL14	10	0.0548	.675	12.716	
6	CL8 CL10	12	0.0654	.609	15.177	
5	CL9 SDR-S150	3	0.0794	.530	18.41	
4	CL12 CL6	15	0.0815	.449	18.915	
3	CL7 CL19	12	0.1076	.341	24.969	
2	CL4 CL5	18	0.1256	.215	29.144	
1	CL2 CL3	30	0.2154	.000	49.963	

Tab. 4 - Historique de classification

Interpréter chacune des colonnes du tableau de l'historique de classification, puis rappeler les formules permettant de calculer les indicateurs SPRSQ, RSQ, et BSS.

2. Via la procédure TREE de SAS, on obtient à partir de l'historique de classification le dendrogramme suivant :

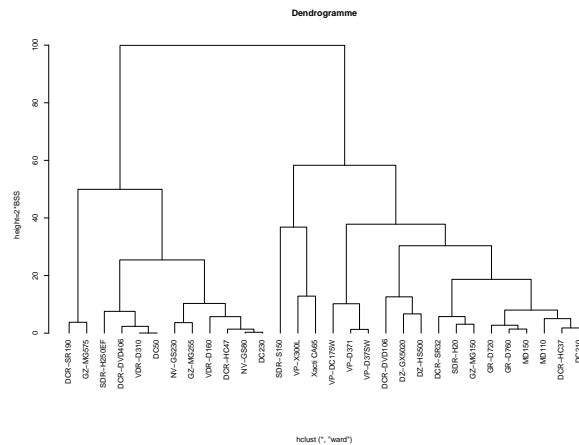


Fig. 3 - Dendrogramme

En se basant sur cette représentation graphique, proposer un nombre de classes à conserver. Justifier votre choix.

3. On se propose pour confirmer ce choix de représenter les variations de perte d'inertie inter-classe en fonction des agrégations (cf. Fig. 4).

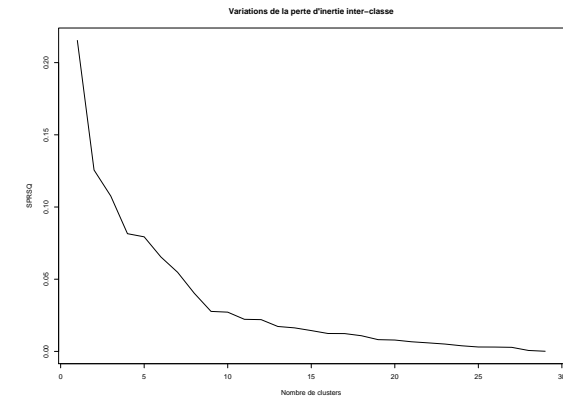


Fig. 4 - Perte d'inertie inter-classe

A partir de ce graphique, indiquer un choix judicieux pour le nombre de classes. Ce choix est-il en accord avec celui obtenu via le dendrogramme ?

4. Selon un découpage approprié du dendrogramme, compléter le tableau 5.

5. Une étude des variables actives à l'intérieur de chacune des classes a permis d'obtenir les boxplots suivants (cf. Fig. 5 et Fig. 6) :

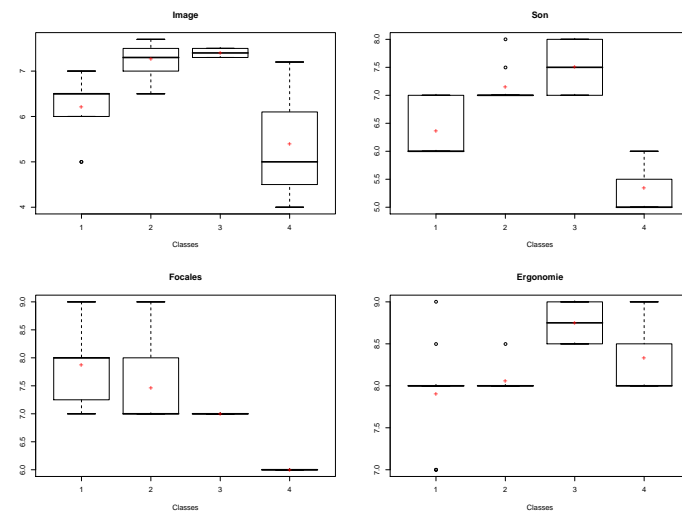


Fig. 5 - Boxplots par classe

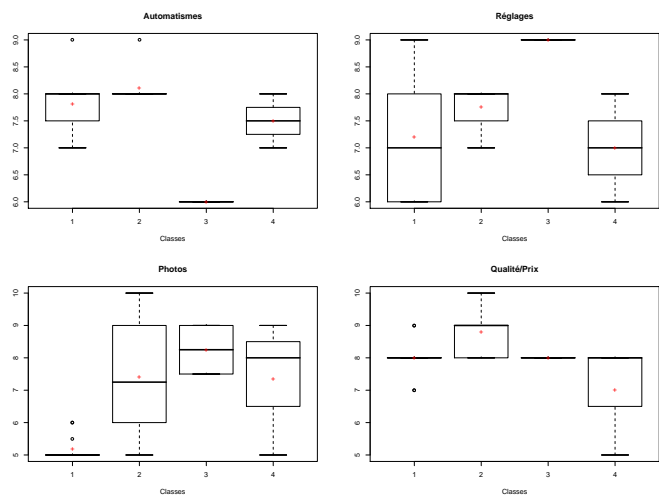


Fig. 6 - Boxplots par classes

Pourquoi n'est-il pas pertinent de commenter les boxplots associés aux classes 3 et 4? Pour les classes 1 et 2, commenter les boxplots pour chacune des variables actives. Proposer alors une caractérisation de chacune des classes.

6. Le graphique suivant (cf. Fig. 7) permet de comparer l'écart entre les moyennes par classe, et les moyennes globales. Ce graphique confirme-t-il votre première analyse (i.e. la caractérisation des classes) ?

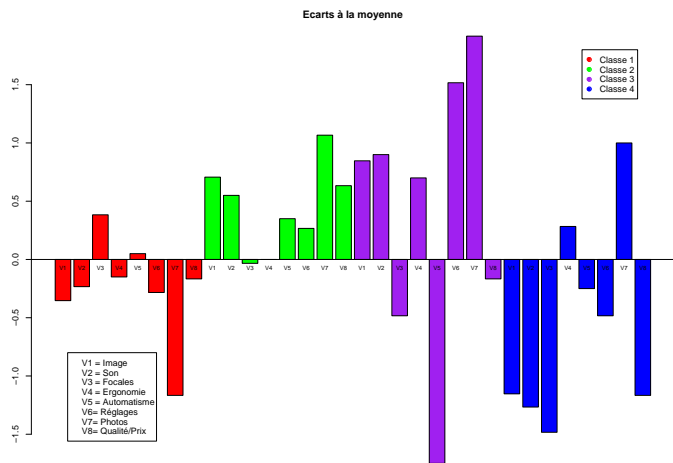


Fig. 7 - Ecart à la moyenne

7. On s'intéresse maintenant aux variables illustratives comme la marque, le type ou le prix du caméscope. En vous aidant des graphiques suivants essayer d'affiner la caractérisation des classes.

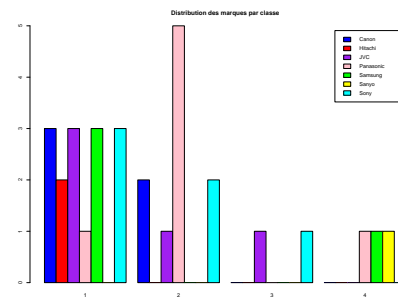


Fig. 8 - Répartition des marques par classe

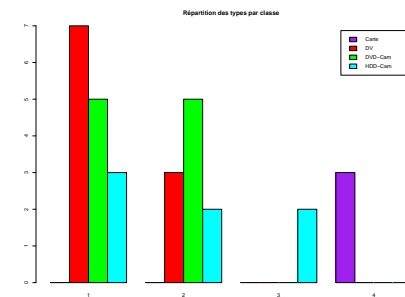


Fig. 9 - Répartition des types par classe

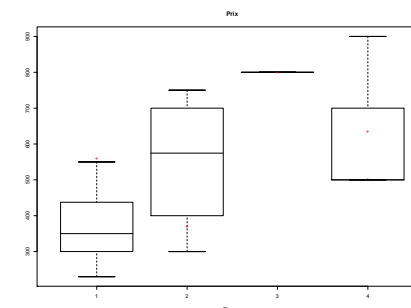


Fig. 10 - Boxplots du prix par classe

8. Si vous deviez acheter un caméscope, préciser dans quelle classe vous vous tourneriez selon les informations dont vous disposez maintenant. Justifier votre réponse.

9. A l'intérieur de cette classe, identifier le ou les meilleurs caméscopes.

10. **Question subsidiaire** : prouver à l'aide de la formule de décomposition de la variance (cf. Théorème de Huygens) que l'on a les relations suivantes :

$$RSQ(k) = \frac{I_{inter}(n-k)}{I} \quad \text{et} \quad \frac{I_{intra}(n-k)}{I} = \sum_{i=1}^k SPRSQ(i)$$

pour tout $0 \leq k \leq n-1$.

CLASSE=1											
Modèle	Marque	Type	Prix	Image	Son	Focales	Ergonomie	Automatismes	Réglages	Photo	Qualité/Prix
VP-D371	Samsung	DV	230	5.0	6.0	9.0	8.0	7	6	5.0	8
VP-D37SW	Samsung	DV	300	5.0	6.0	9.0	8.0	7	6	5.0	7
GR-D760	JVC	DV	300	6.5	6.0	7.0	8.0	8	6	5.0	8
MD150	Canon	DV	369	6.5	6.0	7.0	8.0	8	7	6.0	8
DCR-HC37	Sony	DV	280	6.5	7.0	8.0	8.0	8	7	5.0	8
.	Canon	DVD-Cam	400	6.0	7.0	7.5	8.0	8	8	5.5	8
GR-D720	JVC	DV	249	6.0	6.0	8.0	8.0	8	6	5.0	8
SDR-H20	Panasonic	HDD-Cam	500	7.0	6.0	8.0	8.0	8	8	5.0	8
GZ-MG150	JVC	HDD-Cam	550	7.0	6.0	7.5	8.5	8	8	6.0	9
.	Canon	DV	300	6.0	6.5	7.0	8.0	9	7	5.0	8
DCR-SR32	Sony	HDD-Cam	500	6.5	6.0	8.0	9.0	8	9	5.0	8
DZ-GX5020	Hitachi	DVD-Cam	329	6.0	6.0	8.0	7.0	8	6	5.0	8
DZ-HS500	Hitachi	DVD-Cam	475	6.5	7.0	8.0	7.0	7	7	5.0	9
VP-DC175W	Samsung	DVD-Cam	400	6.0	7.0	9.0	8.0	7	8	5.0	7
DCR-DVD106	Sony	DVD-Cam	350	6.5	7.0	7.0	7.0	8	9	5.0	8
Moyenne			557.800	6.200	6.367	7.867	7.900	7.800	7.200	5.167	8.000
Ecart-type			163.617	0.592	0.481	0.719	0.541	0.561	1.082	0.362	0.535
CLASSE=2											
Modèle	Marque	Type	Prix	Image	Son	Focales	Ergonomie	Automatismes	Réglages	Photo	Qualité/Prix
VDR-D310	Panasonic	DVD-Cam	649	7.7	7.0	7.0	8.0	8	8.0	10.0	8
.	Canon	DVD-Cam	750	7.5	7.0	7.0	8.0	8	8.0	10.0	8
NV-GS80	Panasonic	DV	300	7.0	7.0	8.0	8.0	8	8.0	6.0	9
DC230	Canon	DVD-Cam	500	7.5	7.0	8.0	8.0	8	8.0	6.0	9
DCR-HC47	Sony	DV	400	7.0	7.0	7.5	8.0	8	7.5	7.0	9
DCR-DVD406	Sony	DVD-Cam	700	7.3	7.0	7.0	8.0	8	8.0	9.0	9
NV-GS230	Panasonic	DV	500	7.3	7.5	7.0	8.0	8	8.0	7.5	10
GZ-MG255	JVC	HDD-Cam	650	7.3	8.0	7.0	8.5	8	8.0	6.0	9
VDR-D160	Panasonic	DVD-Cam	379	6.5	7.0	9.0	8.0	8	7.0	5.0	9
SDR-H250EF	Panasonic	HDD-Cam	750	7.5	7.0	7.0	8.0	9	7.0	7.5	8
Moyenne			368.800	7.260	7.150	7.450	8.050	8.100	7.750	7.400	8.800
Ecart-type			99.329	0.347	0.337	0.685	0.158	0.316	0.425	1.761	0.632
CLASSE=3											
Modèle	Marque	Type	Prix	Image	Son	Focales	Ergonomie	Automatismes	Réglages	Photo	Qualité/Prix
DCR-SR190	Sony	HDD-Cam	800	7.3	7	7	9.0	6	9	7.5	8
.	JVC	HDD-Cam	800	7.5	8	7	8.5	6	9	9.0	8
Moyenne			800	7.400	7.500	7.000	8.750	6.000	9.000	8.250	8.000
Ecart-type			0	0.141	0.707	0.000	0.354	0.000	0.000	1.061	0.000
CLASSE=4											
Modèle	Marque	Type	Prix	Image	Son	Focales	Ergonomie	Automatismes	Réglages	Photo	Qualité/Prix
Xacti CA65	Samsung	Carte	500	4.0	5	6	8	7.0	6	5	8
SDR-S150	Panasonic	Carte	900	7.2	6	6	8	8.0	8	8	5
Moyenne			633.333	5.400	5.333	6.000	8.333	7.500	7.000	7.333	7.000
Ecart-type			230.940	1.637	0.577	0.000	0.577	0.500	1.000	2.082	1.732

Tab. 5 - Descriptif des classes

Camescope	Marque	Type	Prix	Image	Son	Focales	Ergonomie	Automatismes	Réglages	Photos	Qualité/Prix
VP-D371	Samsung	DV	230	5	6	9	8	7	6	5	8
GR-D720	JVC	DV	249	6	6	8	8	8	6	5	8
DCR-HC37	Sony	DV	280	6.5	7	8	8	8	7	5	8
VP-D37SW	Samsung	DV	300	5	6	9	8	7	6	5	7
MD110	Canon	DV	300	6	6.5	7	8	9	7	5	8
GR-D760	JVC	DV	300	6.5	6	7	8	8	6	5	8
NV-GS80	Panasonic	DV	300	7	7	8	8	8	8	6	9
MD150	Canon	DV	369	6.5	6	7	8	8	7	6	8
DCR-HC47	Sony	DV	400	7	7	7.5	8	8	7.5	7	9
NV-GS230	Panasonic	DV	500	7.3	7.5	7	8	8	8	7.5	10
DZ-GX5020	Hitachi	DVD-Cam	329	6	6	8	7	8	6	5	8
DCR-DVD106	Sony	DVD-Cam	350	6.5	7	7	7	8	9	5	8
VDR-D160	Panasonic	DVD-Cam	379	6.5	7	9	8	8	7	5	9
VP-DC175W	Samsung	DVD-Cam	400	6	7	9	8	7	8	5	7
DC210	Canon	DVD-Cam	400	6	7	7.5	8	8	8	5.5	8
DC230	Canon	DVD-Cam	500	7.5	7	8	8	8	8	6	9
DZ-HS500	Hitachi	DVD-Cam	475	6.5	7	8	7	7	7	5	9
VDR-D310	Panasonic	DVD-Cam	649	7.7	7	7	8	8	8	10	8
DC50	Canon	DVD-Cam	750	7.5	7	7	8	8	8	10	8
DCR-DVD406	Sony	DVD-Cam	700	7.3	7	7	8	8	8	9	9
SDR-H20	Panasonic	HDD-Cam	500	7	6	8	8	8	8	5	8
GZ-MG150	JVC	HDD-Cam	550	7	6	7.5	8.5	8	8	6	9
DCR-SR32	Sony	HDD-Cam	500	6.5	6	8	9	8	9	5	8
GZ-MG255	JVC	HDD-Cam	650	7.3	8	7	8.5	8	8	6	9
SDR-H250EF	Panasonic	HDD-Cam	750	7.5	7	7	8	9	7	7.5	8
DCR-SR190	Sony	HDD-Cam	800	7.3	7	7	9	6	9	7.5	8
GZ-MG575	JVC	HDD-Cam	800	7.5	8	7	8.5	6	9	9	8
VP-X300L	Samsung	Carte	500	4	5	6	8	7	6	5	8
Xacti CA65	Sanyo	Carte	500	5	5	6	9	7.5	7	9	8
SDR-S150	Panasonic	Carte	900	7.2	6	6	8	8	8	8	5

Tab. 6 - Données brutes