

Examen d'Enquête et analyse des données

Session de janvier 2016

3<sup>ème</sup> Année Licence Techniques du Commerce International

Durée de l'épreuve : 2 heures - Documents non autorisés - Nombre de pages : 4

Une enquête a été réalisée auprès des clients d'ordinateurs portables. Un questionnaire a été administré à un échantillon représentatif de 30 personnes. Un extrait du questionnaire est présenté à l'Annexe 1.

Dans la suite, on associe à la question numéro  $i$  la variable statistique notée  $Q_i$ .

Partie I

On a effectué une ACP sur les 6 items ( $Q_{4,j}, j \in \{1, \dots, 6\}$ ) de la question 4 du questionnaire. Les résultats de cette ACP sont présentés à l'Annexe 2.

1. Quel est l'intérêt de cette ACP ?
2. Combien de composantes principales devrait-on retenir dans cette ACP ? Justifier votre réponse.

Dans la suite, on suppose que l'on ne retient que 3 composantes principales.

3. Donner une interprétation des axes retenus.
4. On a étudié le lien entre les 3 premières composantes et la variable  $Q_3$  (CSP). Les résultats sont présentés à l'Annexe 2. Commenter ces résultats.

Partie II

En utilisant les coordonnées des 30 personnes interrogées sur les 3 premiers axes de l'ACP, on a réalisé une classification hiérarchique avec le critère de Ward sur l'ensemble des 30 personnes. En vous basant sur la hiérarchie présentée dans l'Annexe 2 :

5. Déterminer, en justifiant votre réponse, la meilleure partition issue de cette analyse typologique (on indiquera la liste des individus appartenant à chaque classe de la partition choisie).
6. On considère la partition en 3 classes des 30 personnes. Soit  $P_3$  la variable indiquant la classe d'appartenance de chacune des 30 personnes dans cette partition. On a étudié le lien entre la variable  $P_3$  et la variable  $Q_2$  (Age). Les résultats sont présentés à l'Annexe 2. Interpréter ces résultats.

## Annexe 1

### Extrait du questionnaire :

1. Genre :

Homme ...

Femme ...

2. Age :

Moins de 20 ans ...

Entre 20 et 40 ans ...

Plus de 40 ans ...

3. Catégorie Socioprofessionnelle (CSP) :

Etudiant ...

Professionnel ...

Chômeur...

Retraité...

4. Pour chacun des déterminants de choix d'achat suivants, veuillez indiquer son degré d'importance pour vous ((1) si pas du tout important, (2) si peu important, (3) si neutre, (4) si important, (5) si très important) :

4.1 Prix	1	2	3	4	5
4.2 Design	1	2	3	4	5
4.3 Poids	1	2	3	4	5
4.4 Robustesse	1	2	3	4	5
4.5 Batterie	1	2	3	4	5
4.6 Service après vente (SAV)	1	2	3	4	5



## Annexe 2

### Les 4 premières valeurs

propres	
Composante	Valeurs propres
	initiales
	Total
1	1,982
2	1,367
3	1,134
4	,647

### Matrice des composantes

	Composante		
	1	2	3
Prix	,106	-,098	,921
Design	,135	,893	,017
Poids	,817	,184	,103
Robustesse	-,289	,801	-,043
Batterie	-,581	,210	,596
SAV	,729	-,296	-,059

### Descriptives

		N	Moyenne	Ecart-type
REGR factor score 1 for analysis 1	Etudiant	8	-,2294077	,85708802
	Professionnel	13	,2325946	1,21524786
	Chômeur	6	-,4413081	,78637549
	Retraité	3	,4864598	,14347081
	Total	30	0E-7	1,00000000
REGR factor score 2 for analysis 1	Etudiant	8	,6060041	,79093925
	Professionnel	13	,3078394	,74233085
	Chômeur	6	-,8936156	,79861147
	Retraité	3	-1,1627506	,94926553
	Total	30	0E-7	1,00000000
REGR factor score 3 for analysis 1	Etudiant	8	-,1438995	,86780226
	Professionnel	13	,3052840	,99744132
	Chômeur	6	-,2123510	1,27808502
	Retraité	3	-,5144634	,76826029
	Total	30	0E-7	1,00000000

### ANOVA à 1 facteur

		Somme des carrés	ddl	Moyenne des carrés	F	Signification
REGR factor score 1 for analysis 1	Inter-groupes	3,003	3	1,001	1,001	,408
	Intra-groupes	25,997	26	1,000		
	Total	29,000	29			
REGR factor score 2 for analysis 1	Inter-groupes	13,017	3	4,339	7,059	,001
	Intra-groupes	15,983	26	,615		
	Total	29,000	29			
REGR factor score 3 for analysis 1	Inter-groupes	2,442	3	,814	,797	,507
	Intra-groupes	26,558	26	1,021		
	Total	29,000	29			

Arbre hiérarchique utilisant la Distance de Ward  
Distance de combinaison des classes redimensionnées

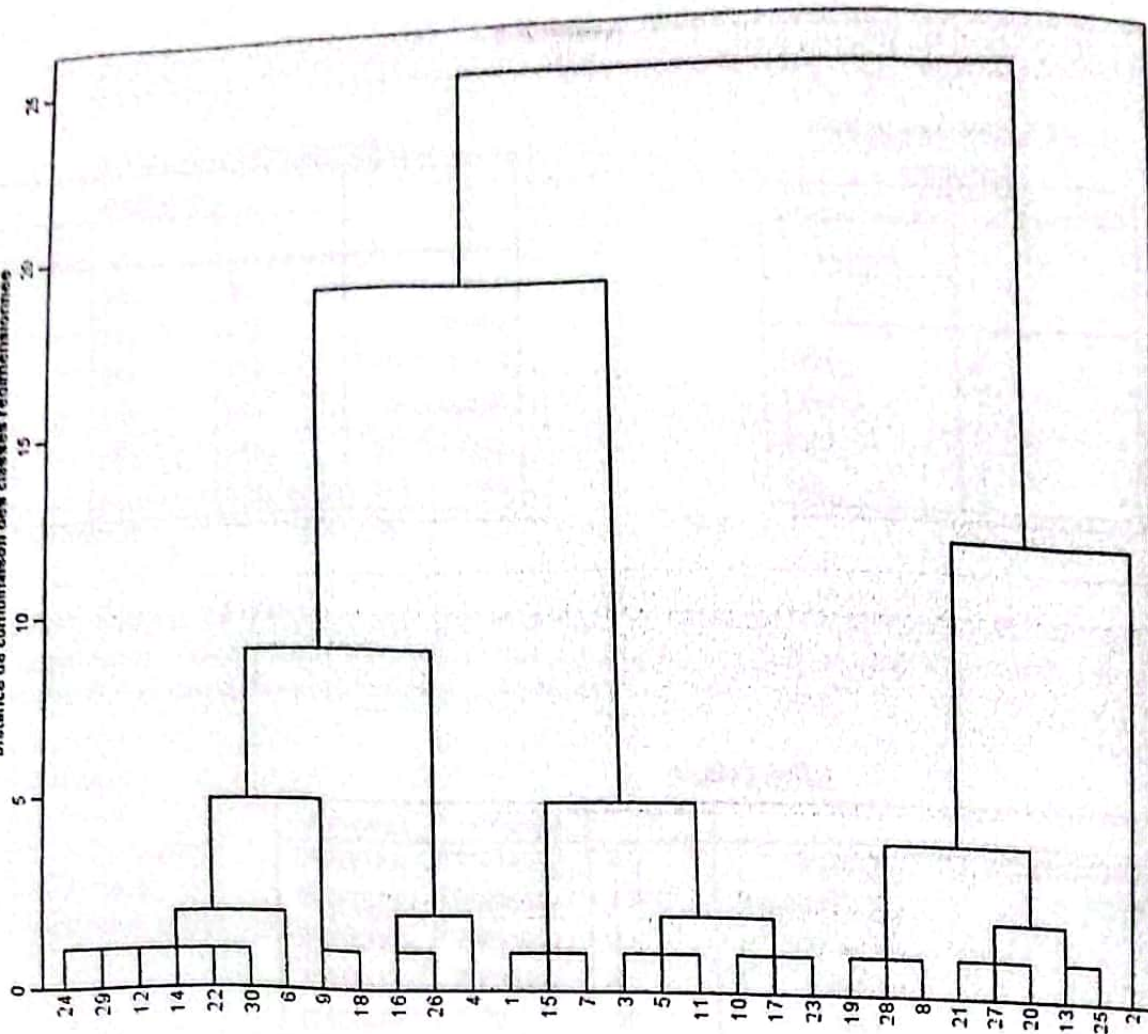


Tableau croisé P3 \* Age

% compris dans P3

		Age			Total
		moins de 20 ans	Entre 20 et 40 ans	Plus de 40 ans	
P3	1	55,6%	11,1%	33,3%	100,0%
	2	44,4%	44,4%	11,1%	100,0%
	3	8,3%	25,0%	66,7%	100,0%
Total		33,3%	26,7%	40,0%	100,0%

Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
Khi-deux de Pearson	9,924 <sup>a</sup>	4	,042
Rapport de vraisemblance	11,103	4	,025
Association linéaire par linéaire	4,961	1	,026
Nombre d'observations valides	30		



1) (2pts=1 +1) L'intérêt de cette ACP est double

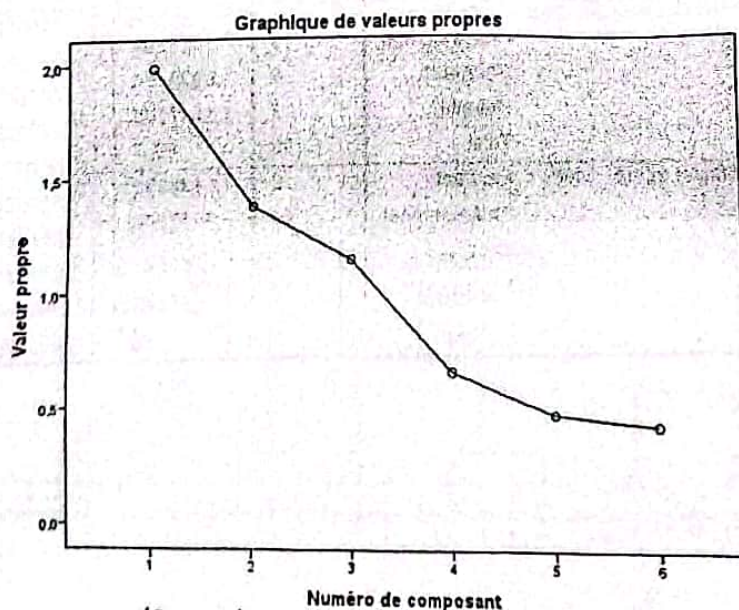
- (1pt) Réduire le nombre de variables décrivant les déterminants de choix d'achats d'un ordinateur portable en fournissant un petit nombre de nouvelles variables (les composantes principales) les décrivant.
- (1pt) Identifier les groupes de clients d'ordinateurs portables ayant les mêmes types de déterminants lors de l'achat d'un ordinateur portable et décrire ces groupes de clients.

2) (3pts=1 +1 +1) Nombre d'axes à retenir

- (1pt) Critère de kaiser : on remarque qu'il y a 3 axes dont les valeurs propres sont supérieures à 1 donc on retient 3 axes d'après ce critère.
- (1pt) Critère du taux d'inertie cumulée : On remarque que le taux d'inertie cumulé des 3 premiers axes est de 74.71% qui est un taux important : on va donc, d'après ce critère, retenir les 3 premiers axes.

Composante	Variance totale expliquée		
	Valeurs propres initiales		
	Total	% de la variance	% cumulés
1	1,982	33,033	33,033
2	1,367	22,779	55,812
3	1,134	18,900	74,712
4	,647	10,777	85,489
5	,460	7,674	93,163
6	,410	6,837	100,000

- (1pt) Critère du coude : On remarque que le coude se trouve au niveau du deuxième axe (voir figure ci-dessous), d'après ce critère, on devrait retenir les 2 premiers axes. Toutefois, on a changement de concavité au niveau du troisième axe, on pourrait donc retenir 3 axes.



Selon, au moins 2 critères on doit retenir les 3 premiers axes, on va donc retenir 3 axes.

3) (3pts=1+1+1) Interprétation des axes :

- (1pt) **Axe 1** : cet axe est fortement corrélé positivement aux variables (déterminants) poids et SAV
- (1pt) **Axe 2** : cet axe est fortement corrélé positivement aux variables design et robustesse
- (1pt) **Axe 3** : cet axe est fortement corrélé positivement aux variables prix et batterie.

4) (4pts=2 +2)

- Seul le deuxième facteur est significativement corrélé à la variable CSP (p-value =0.1% <1%). Les 2 autres p-values sont largement supérieures à 5% (40 et 50 %).
- D'autre part, à partir du tableau Descriptives, on remarque que les étudiants sont ceux qui accordent le plus d'importance au design et à la robustesse suivies des professionnels, tandis que les chômeurs et les retraités ne semblent accorder aucune importance à ces deux déterminants.

Descriptives

		N	Moyenne	Ecart-type
REGR factor score 1 for analysis 1	Etudiant	8	-.2294077	.85708802
	Professionnel	13	.2325946	1.21524786
	Chômeur	6	-.4413081	.78637549
	Retraité	3	.4864598	.14347081
	Total	30	0E-7	1.00000000
REGR factor score 2 for analysis 1	Etudiant	8	.6060041	.79093925
	Professionnel	13	.3078394	.74233085
	Chômeur	6	-.8936156	.79861147
	Retraité	3	-1.1627506	.94926553
	Total	30	0E-7	1.00000000
REGR factor score 3 for analysis 1	Etudiant	8	-.1438995	.86780226
	Professionnel	13	.3052840	.99744132
	Chômeur	6	-.2123510	1.27808502
	Retraité	3	-.5144634	.76826029
	Total	30	0E-7	1.00000000



ANOVA à 1 facteur

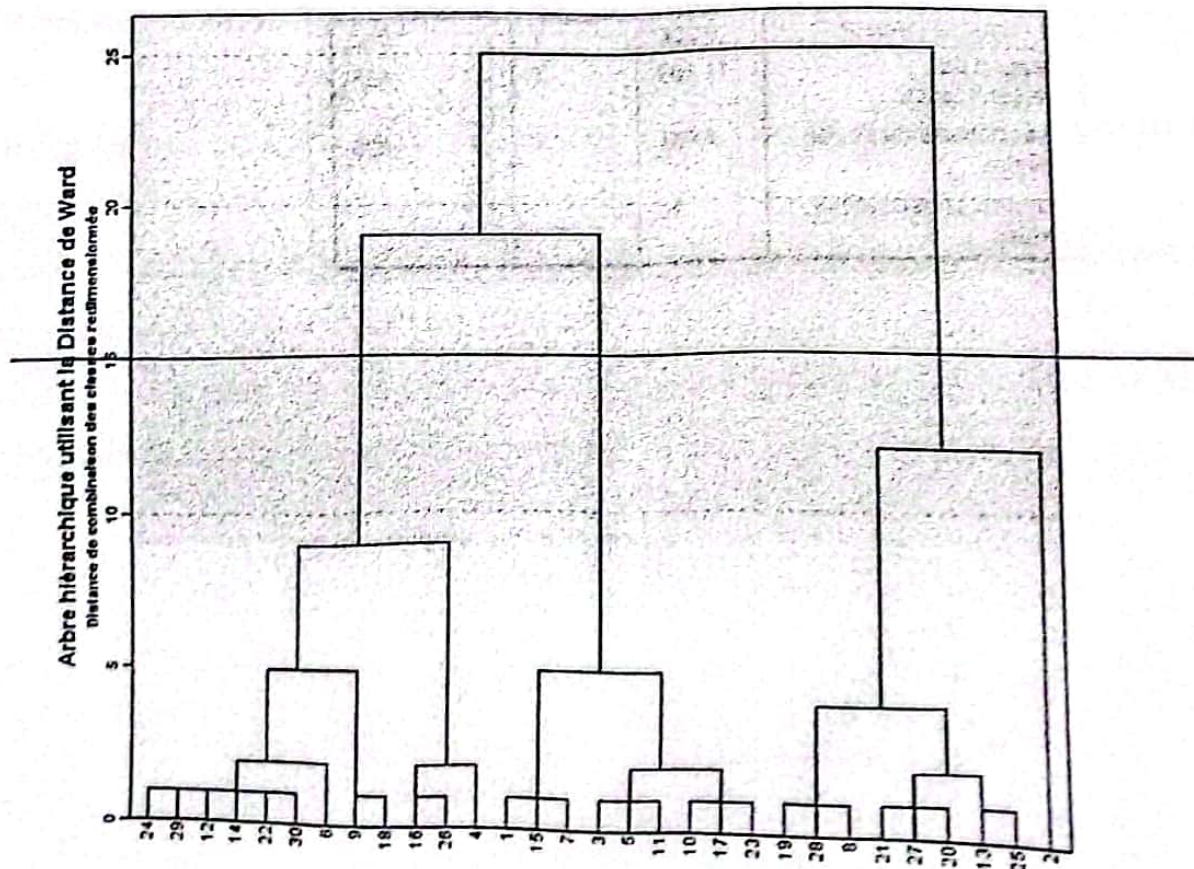
		Somme des carrés	ddl	Moyenne des carrés	F	Signification
REGR factor score 1 for analysis 1	Inter-groupes	3,003	3	1,001	1,001	,408
	Intra-groupes	25,997	26	1,000		
	Total	29,000	29			
REGR factor score 2 for analysis 1	Inter-groupes	13,017	3	4,339	7,059	,001
	Intra-groupes	15,983	26	,615		
	Total	29,000	29			
REGR factor score 3 for analysis 1	Inter-groupes	2,442	3	,814	,797	,507
	Intra-groupes	26,558	26	1,021		
	Total	29,000	29			

## 5) (4pts)

- Il est clair que pour obtenir la meilleure partition, il faut couper cette hiérarchie au niveau de l'indice 15 sur l'axe verticale, car c'est entre ces deux paliers que nous avons le plus haut saut entre deux paliers successifs.

- On obtient ainsi une partition en 3 classes qui sont les suivantes :

$$C1 = \{24, 29, 12, 14, 22, 30, 6, 9, 18, 16, 26, 4\}, C2 = \{1, 15, 7, 3, 5, 11, 10, 17, 23\}, \\ C3 = \{19, 28, 8, 21, 27, 20, 13, 25, 2\}.$$



6) (4pts=1 +3)

- La p-value du test du chi-deux (p-value=4,2%) indique qu'il est significatif au seuil de 5%. Donc il existe bien une relation entre la partition en 3 classes et la variable âge.

- Plus précisément, dans la classe 1 on a une forte présence des moins de 20 ans (55,6%) ainsi que des plus de 40 ans (33,3%). Dans la classe 2, on a une forte présence des moins de 20 ans (44,4%) ainsi que des plus de 40 ans (44,4%). Enfin dans la classe C3, on a surtout une forte présence des plus de 40 ans (66,7%).

Tableau croisé P3 ^ Age

% compris dans P3

		Age			Total
		moins de 20 ans	Entre 20 et 40 ans	Plus de 40 ans	
P3	1	55,6%	11,1%	33,3%	100,0%
	2	44,4%	44,4%	11,1%	100,0%
	3	8,3%	25,0%	66,7%	100,0%
Total		33,3%	26,7%	40,0%	100,0%

Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
Khi-deux de Pearson	9,924 <sup>a</sup>	4	,042
Rapport de vraisemblance	11,103	4	,025
Association linéaire par linéaire	4,961	1	,026
Nombre d'observations valides	30		