

Epistémologie et Méthodologie
Méthodologie Quantitative

SPSS Projet Pommes Octobre 2005

Michel TENENHAUS – Professeur HEC



KHALIDI Marie-France
OUEDRAOGO Soukeyna

SOMMAIRE

<i>Introduction.....</i>	<i>1</i>
<i>Statistiques descriptives.....</i>	<i>3</i>
<i>Analyse en composante principale.....</i>	<i>8</i>
<i>Classification ascendante hiérarchique.....</i>	<i>25</i>
<i>Interprétation des classes obtenues.....</i>	<i>35</i>
<i>Annexes.....</i>	<i>38</i>

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



PROJET POMMES

L'objectif de notre étude est d'établir une cartographie, c'est à dire un regroupement, des pommes les plus couramment commercialisées en France en fonction de leurs saveurs.

Nous avons pour cela recueilli des données sur les 16 variétés de pommes suivantes : Golden, Gala, Reine des Reinettes, Braeburn, Jonagold, Pink Lady, Elstar, Tentation, Fuji, Idared, Les Rouges, Reinette grise du Canada, Belchard Chanteclerc, Belle de Boskoop, Granny Smith, Cox Orange.

Internet a été notre principale source d'informations. Nous nous sommes également inspirées d'une plaquette disponible chez la plupart des fruitiers. Les références de ces documents sont précisées à la fin de l'étude. Nous avons par la suite pallié aux imprécisions par nos connaissances personnelles dans le domaine.

Les variables utilisées pour cette étude (les différentes saveurs) sont toutes qualitatives. La saveur des pommes a donc été décrite avec des variables binaires, selon le mode suivant :

1 = présence de la caractéristique

0 = absence

Voici comment ont été codées ces caractéristiques pour le traitement des données sous le logiciel SPSS.

- Douce : *dce*
- Parfumée : *parf*
- Très Parfumée : *trparf*
- Sucrée : *sucr*
- Très sucrée : *trsucr*
- Fruitée : *fruit*
- Légèrement acidulée: *legacid*
- Acidulée: *acid*
- Très acidulée: *tracid*
- Croquante: *croq*
- Très croquante: *trcroq*
- Juteuse: *jut*
- Rafraîchissante : *rafr*
- Ferme : *ferm*

Nous avons ainsi obtenu le premier tableau suivant.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



TABLEAU DISJONCTIF COMPLET (question 1)

	Douce	Parfumée	Très parfumée	Sucrée	Très sucrée	Fruitée	Légèrement acidulée	Acidulée	Très acidulée	Croquante	Très croquante	Juteuse	Rafraîchissante	Ferme
Golden	1	1	0	1	0	1	1	0	0	1	0	1	0	1
Gala	1	0	0	1	0	1	0	0	0	1	0	1	1	1
Reine des Reinettes	1	1	0	1	0	0	1	0	0	0	0	0	0	0
Braeburn	0	0	0	1	0	0	0	1	0	0	1	1	0	1
Jonagold	0	0	1	0	1	1	1	0	0	1	0	1	0	0
Pink Lady	0	0	0	1	0	1	0	0	0	0	1	1	0	1
Elstar	0	1	0	0	1	0	1	0	0	1	0	1	1	0
Tentation	0	0	1	1	0	1	1	0	0	0	0	1	1	1
Fuji	0	0	0	0	1	0	0	0	0	1	0	1	1	0
Idared	1	0	0	1	0	0	1	0	0	1	0	1	0	1
Les Rouges	0	0	0	1	0	0	0	0	0	1	0	1	0	1
Reinette grise du Canada	0	1	0	1	0	0	0	1	0	1	0	0	0	0
Belchard Chanteclerc	0	0	1	0	1	0	0	1	0	0	0	0	0	0
Belle de Boskoop	0	1	0	0	0	0	0	1	0	0	0	0	0	0
Granny Smith	0	0	0	0	0	0	0	0	1	0	1	1	1	1
Cox Orange	1	1	0	0	0	1	0	0	0	1	0	0	0	1

$X_{ijl} = 1$ si la variété de pomme i possède la modalité l de la caractéristique j
 $= 0$ sinon

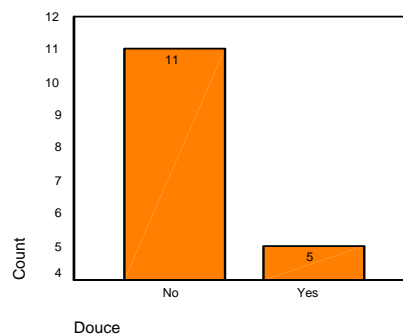
KHALIDI Marie-France et OUEDRAOGO Soukeyna
 IAE Paris/HEC



STATISTIQUES DESCRIPTIVES ET HISTOGRAMMES

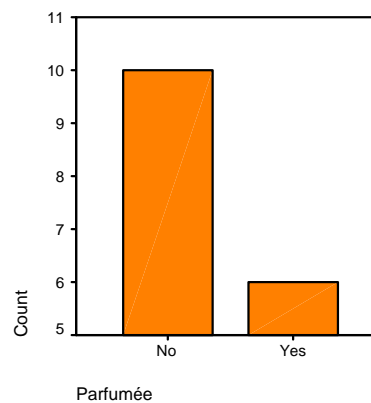
Douce

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	11	68,8	68,8	68,8
	Yes	5	31,3	31,3	100,0
	Total	16	100,0	100,0	



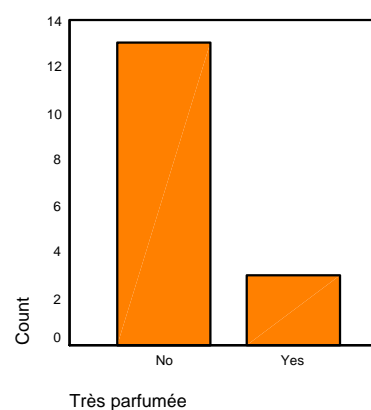
Parfumée

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	10	62,5	62,5	62,5
	Yes	6	37,5	37,5	100,0
	Total	16	100,0	100,0	



Très parfumée

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	13	81,3	81,3	81,3
	Yes	3	18,8	18,8	100,0
	Total	16	100,0	100,0	

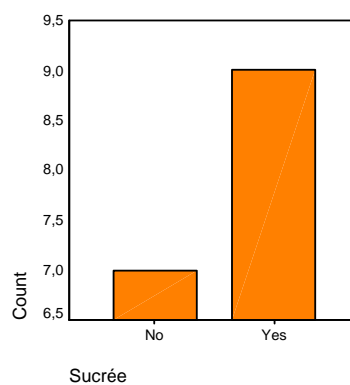


KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC

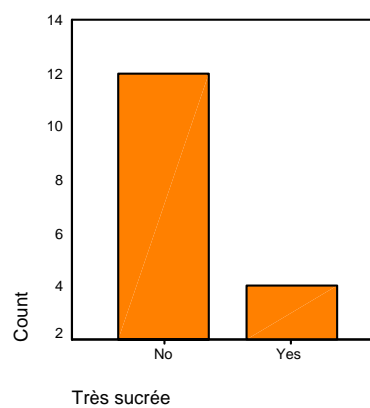


Sucrée

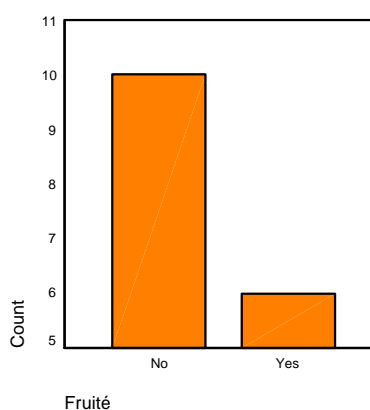
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	7	43,8	43,8	43,8
	Yes	9	56,3	56,3	100,0
	Total	16	100,0	100,0	

**Très sucrée**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	12	75,0	75,0	75,0
	Yes	4	25,0	25,0	100,0
	Total	16	100,0	100,0	

**Fruité**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	10	62,5	62,5	62,5
	Yes	6	37,5	37,5	100,0
	Total	16	100,0	100,0	

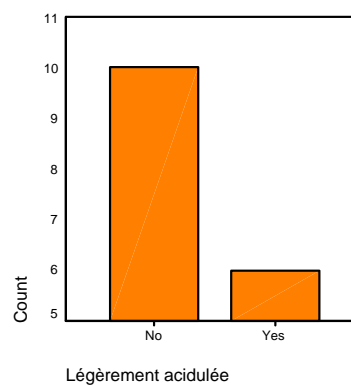


KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC

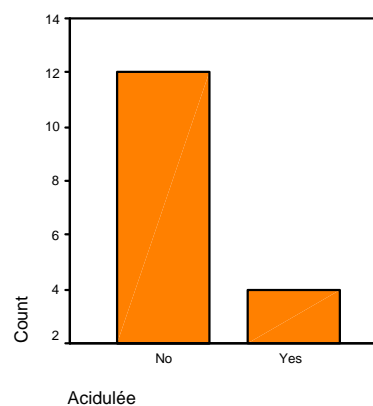


Légèrement acidulée

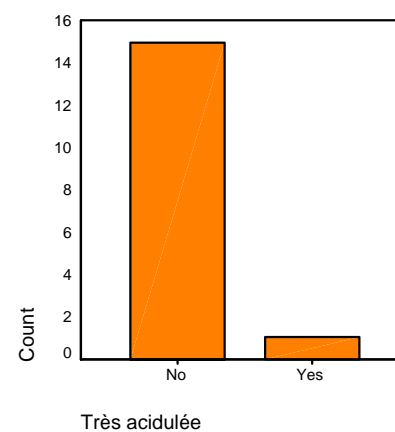
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	10	62,5	62,5	62,5
	Yes	6	37,5	37,5	100,0
	Total	16	100,0	100,0	

**Acidulée**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	12	75,0	75,0	75,0
	Yes	4	25,0	25,0	100,0
	Total	16	100,0	100,0	

**Très acidulée**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	15	93,8	93,8	93,8
	Yes	1	6,3	6,3	100,0
	Total	16	100,0	100,0	

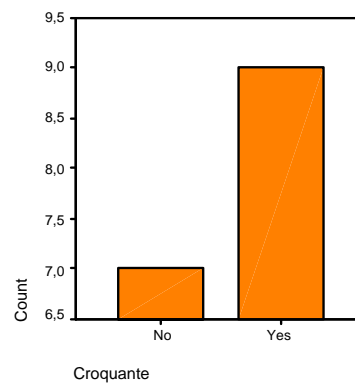


KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC

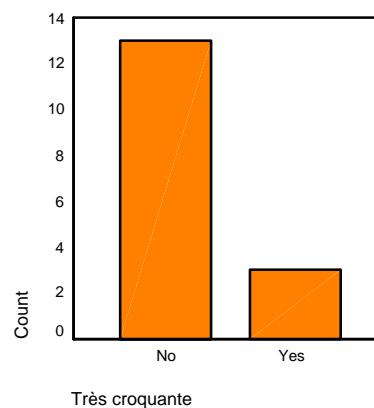


Croquante

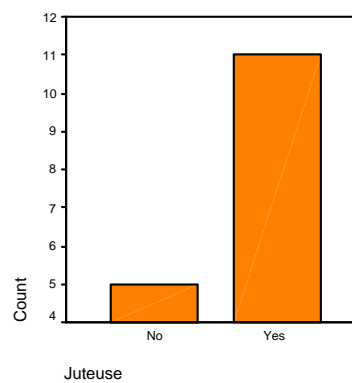
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	7	43,8	43,8	43,8
	Yes	9	56,3	56,3	100,0
	Total	16	100,0	100,0	

**Très croquante**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	13	81,3	81,3	81,3
	Yes	3	18,8	18,8	100,0
	Total	16	100,0	100,0	

**Juteuse**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	5	31,3	31,3	31,3
	Yes	11	68,8	68,8	100,0
	Total	16	100,0	100,0	

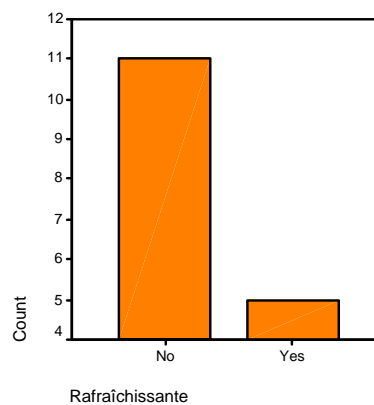


KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC

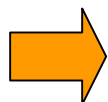
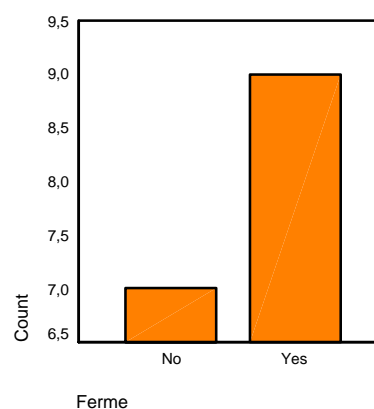


Rafraîchissante

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	11	68,8	68,8	68,8
	Yes	5	31,3	31,3	100,0
	Total	16	100,0	100,0	

**Ferme**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	7	43,8	43,8	43,8
	Yes	9	56,3	56,3	100,0
	Total	16	100,0	100,0	



Il y a 16 individus (pommes) définies dans le fichier et aucune valeur manquante. Il est intéressant de constater que toutes les variables étudiées sont relativement discriminantes.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



ANALYSE EN COMPOSANTE PRINCIPALE

Analyse factorielle des données /Construction du premier plan factoriel et interprétation des résultats (question 2)

Facteurs

L'analyse en composantes principales propose 14 facteurs qui déterminent la structure sous-jacente de la répartition des données.

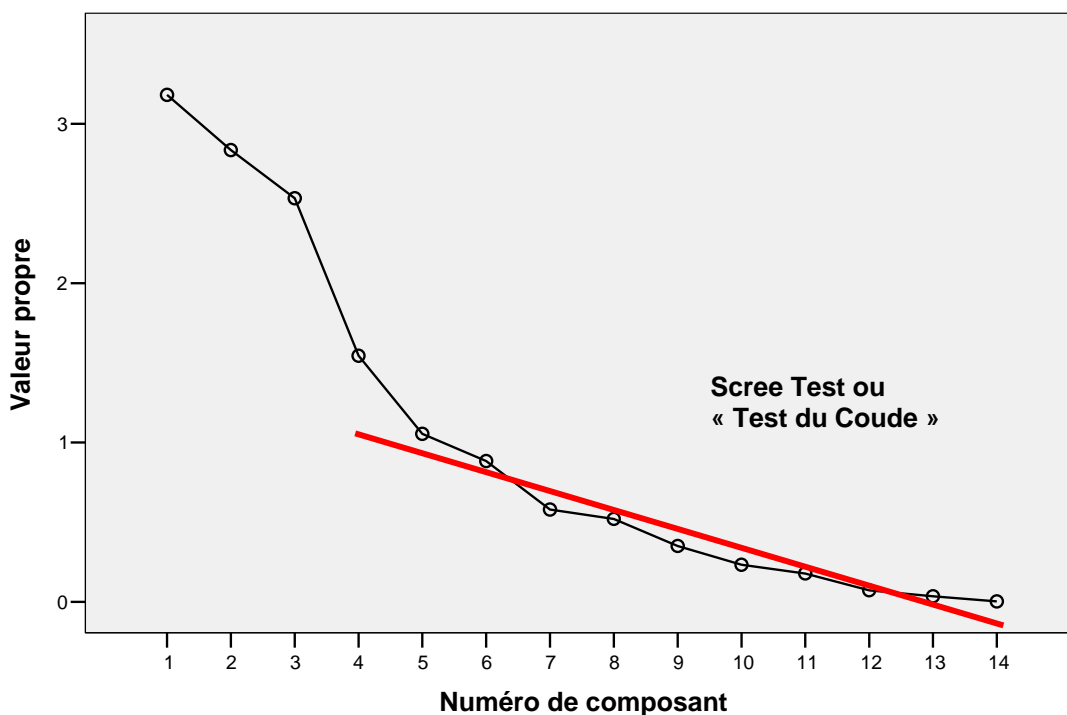
Facteurs retenus

Les facteurs retenus présentent les caractéristiques suivantes :

- Une valeur propre supérieure à 1 selon la « Règle de Kaiser » (Facteurs isolés : Facteurs 1 à 5)
Dans la mesure où l'ACP propose 14 facteurs, une valeur propre de 1 représente 1/14% de la variance, soit environ 7,14%
- Elles répondent au « Scree Test » (**Cf. Graphique 1 ci-après**). Après observation du graphique des valeurs propres, on ne retient que les valeurs qui se trouvent à gauche du point d'inflexion. Le dessin d'une droite reliant les points à droite presque alignés permet d'isoler les axes qui sont au dessus de cette ligne. Le test ainsi réalisé nous invite à retenir les facteurs 1, 2, 3 et 4 qui semblent pertinents pour notre analyse.



Graphique des valeurs propres



Graphique 2 – « Test du coude »

Les 4 premiers facteurs permettent une restitution d'information de 72% environ. (Cf. Tableau 3 ci-après : « Variance totale expliquée ») On s'intéressera cependant plus particulièrement aux facteurs 1 et 2 qui présentent la restitution marginale de l'information la plus élevée et totalisent une restitution d'information d'environ 43% (Respectivement 22.726% et 20.254% pour les composantes principales 1 et 2). On peut souligner que ce score est relativement faible et que, de ce fait, l'analyse de la dispersion des données dans l'espace qui suit est basée sur une perte importante d'informations.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Tableaux d'analyse de l'ACP (avant élimination des variables à écarter)

Qualité de représentation

	Initial	Extraction
Douce	1,000	,641
Parfumée	1,000	,430
Très parfumée	1,000	,088
Sucrée	1,000	,396
Très sucrée	1,000	,345
Fruité	1,000	,340
Légèrement acidulée	1,000	,311
Acidulée	1,000	,472
Très acidulée	1,000	,459
Croquante	1,000	,375
Très croquante	1,000	,737
Juteuse	1,000	,489
Rafraîchissante	1,000	,118
Ferme	1,000	,818

Méthode d'extraction : Analyse en composantes principales.

Tableau 1 – Qualité de représentation

Matrice des composantes ^a

	Composante	
	1	2
Douce	,380	,704
Parfumée	-,369	,542
Très parfumée	-,285	-,082
Sucrée	,575	,256
Très sucrée	-,584	-,061
Fruité	,474	,339
Légèrement acidulée	,107	,547
Acidulée	-,612	-,311
Très acidulée	,246	-,631
Croquante	,073	,608
Très croquante	,403	-,758
Juteuse	,671	-,194
Rafraîchissante	,217	-,266
Ferme	,900	-,089

Méthode d'extraction : Analyse en composantes principales.

a. 2 composantes extraites.

Tableau 2 – Matrice des composantes

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Variance totale expliquée

Composante	Valeurs propres initiales			Extraction Sommes des carrés des facteurs retenus		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	3,182	22,726	22,726	3,182	22,726	22,726
2	2,836	20,254	42,980	2,836	20,254	42,980
3	2,533	18,092	61,072			
4	1,544	11,031	72,103			
5	1,054	7,530	79,633			
6	,884	6,311	85,944			
7	,579	4,134	90,078			
8	,520	3,714	93,792			
9	,350	2,501	96,293			
10	,232	1,659	97,952			
11	,178	1,269	99,222			
12	,072	,514	99,736			
13	,035	,247	99,982			
14	,002	,018	100,000			

Méthode d'extraction : Analyse en composantes principales.

Tableau 3 – Variance totale expliquée

Matrice des coefficients des coordonnées des composantes

	Composante	
	1	2
Douce	,120	,248
Parfumée	-,116	,191
Très parfumée	-,089	-,029
Sucrée	,181	,090
Très sucrée	-,184	-,021
Fruité	,149	,120
Légèrement acidulée	,034	,193
Acidulée	-,192	-,110
Très acidulée	,077	-,223
Croquante	,023	,214
Très croquante	,127	-,267
Juteuse	,211	-,068
Rafraîchissante	,068	-,094
Ferme	,283	-,031

Méthode d'extraction : Analyse en composantes principales.

Scores composante.

Tableau 4 – Matrice des coefficients des coordonnées des composantes

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Interprétation des facteursMatrice des composantes^a

	Composante	
	1	2
Douce	,380	,704
Parfumée	,369	,542
Très parfumée	-,285	-,082
Sucrée	,575	,256
Très sucrée	-,584	-,061
Fruité	,474	,339
Légèrement acidulée	,107	,547
Acidulée	-,612	-,311
Très acidulée	,246	-,631
Croquante	,073	,608
Très croquante	,403	-,758
Juteuse	,671	-,194
Rafraîchissante	,217	-,266
Ferme	,900	-,089

Méthode d'extraction : Analyse en composantes principales.

a. 2 composantes extraites.

Il s'agit ici de construire les deux premiers axes factoriels qui s'établissent à partir des variables et des composantes principales. L'observation de la matrice des composantes permet d'amorcer un début de compréhension et d'explication de la dispersion des données de départ.

- **Premier axe principal et première composante principale**

Les variables « Ferme » et « Juteuse » sont fortement et positivement corrélées à la première composante principale et concourent ainsi le plus à la construction de l'axe 1. Les variables « Acidulée » et « Très sucrée » sont, quant à elles, négativement corrélées à la première composante, mais permettent de comprendre la nature du premier facteur pratiquement à la même hauteur. Il est à noter que la variable « Ferme » concoure de façon excessive à la construction de l'axe 1 (Corrélation avec composante principale 1 proche de 1). La variable pourrait être écartée de l'analyse, dans la mesure où sa proximité trop importante du facteur 1 risque de se traduire par une absence de gain d'informations (indispensables à l'interprétation), et pourrait constituer un biais à l'analyse.

- **Deuxième axe principal et deuxième composante principale**

Les variables « Douce » et « Croquante » sont fortement corrélées à la deuxième composante principale. Elles seront privilégiées pour l'interprétation du facteur 2. Il faut souligner que, malgré le fait que la variable « Très croquante » soit fortement et négativement

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



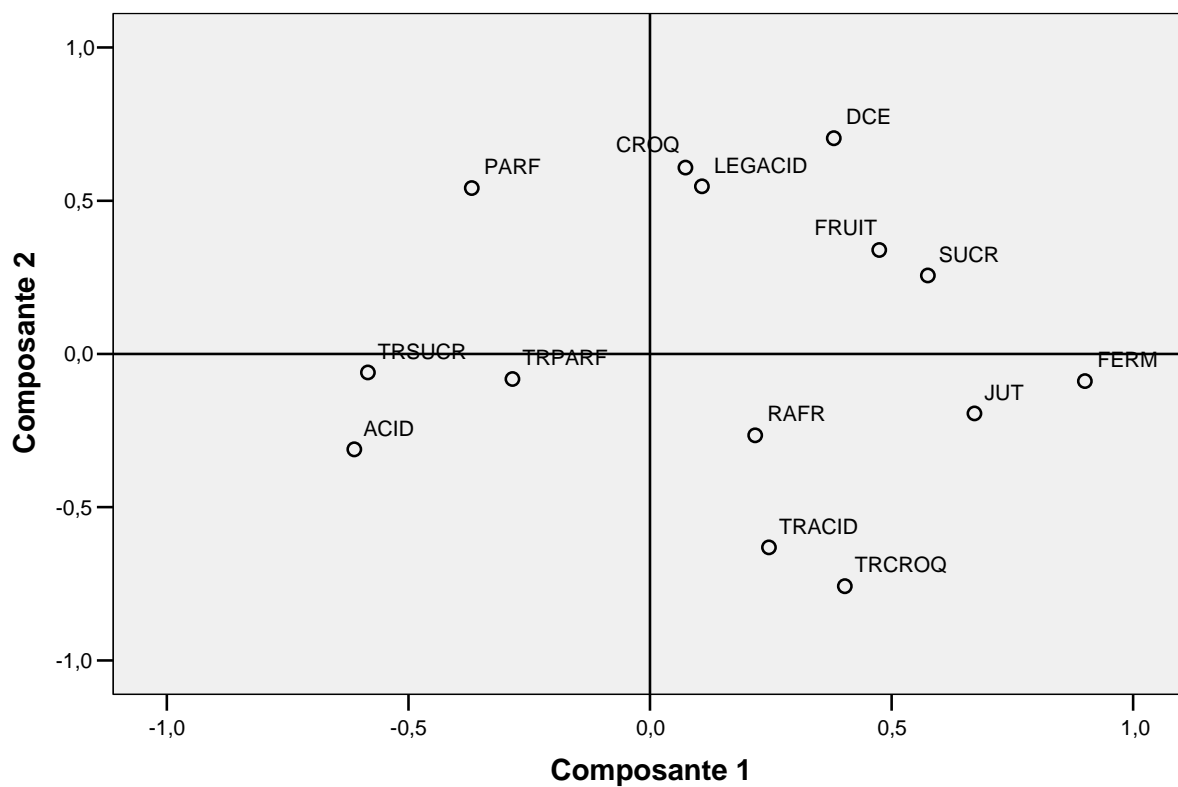
corrélée à la deuxième composante principale, elle constitue un biais évident à l'interprétation. En effet, nous avons retenu une typologie binaire lors de la constitution du tableau des données (0=Absence de caractéristique, 1=Présence). Lorsqu'il s'agit d'évaluer les corrélations sous-jacentes entre les individus et les variables sur des caractéristiques non « diamétralement opposées » (ex : Très croquante, croquante), il est logique que les variables apparaissent *in fine* traitées et représentées quasiment de manière opposées. Mais là encore, la corrélation, positive pour l'une, négative pour l'autre, qu'elles entretiennent avec une autre variable apporte très peu à l'analyse. On retiendra donc que les deux variables qui concourent de manière la plus significative à la construction de l'axe 2 sont les variables « Douce » et « Croquante ». De façon plus secondaire, les variables « Acidulée », « Très acidulée » et « Rafraîchissante » sont négativement corrélées à la seconde composante principale, alors même que, parallèlement, la variable « Légèrement acidulée » est positivement corrélée à la même composante. L'opposition est ici significative et pertinente pour l'analyse, on décide donc de privilégier plus particulièrement cette variable pour la suite de l'interprétation.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Cercle des corrélations avant élimination des variables à écarter

Diagramme de composantes



KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



- **Qualité de représentation des variables et variables à écarter de l'analyse et de son interprétation**

Qualité de représentation

	Initial	Extraction
Douce	1,000	,641
Parfumée	1,000	,430
Très parfumée	1,000	,088
Sucrée	1,000	,396
Très sucrée	1,000	,345
Fruité	1,000	,340
Légèrement acidulée	1,000	,311
Acidulée	1,000	,472
Très acidulée	1,000	,459
Croquante	1,000	,375
Très croquante	1,000	,737
Juteuse	1,000	,489
Rafraîchissante	1,000	,118
Ferme	1,000	,818

Méthode d'extraction : Analyse en composantes principales.

Variables à écarter :

L'étude du tableau de qualité de représentation permettrait d'exclure de l'étude les variables « Rafraîchissante » et « Très parfumée » dont le coefficient de saturation est très faible, et qui seront par conséquent mal représentées par les 2 dimensions extraites.

L'observation du diagramme des composantes (Cercle des corrélations 1) (Graphe « Cercle des corrélations ») permet par ailleurs de confirmer que ces deux variables se situent le plus au centre du plan.

-> Pour la suite de l'analyse, on fait le choix d'écarter ces deux variables.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Qualité de la représentation après élimination des variables à écarter :

Qualité de représentation

	Initial	Extraction
Douce	1,000	,602
Parfumée	1,000	,422
Sucrée	1,000	,393
Très sucrée	1,000	,335
Fruité	1,000	,387
Légèrement acidulée	1,000	,354
Acidulée	1,000	,487
Très acidulée	1,000	,424
Croquante	1,000	,405
Très croquante	1,000	,779
Juteuse	1,000	,464
Ferme	1,000	,836
REGR factor score 1 for analysis 1	1,000	,991
REGR factor score 2 for analysis 1	1,000	,988

Méthode d'extraction : Analyse en composantes principales.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Variance totale expliquée après élimination des variables à écarter :

Variance totale expliquée

Composante	Valeurs propres initiales			Extraction Sommes des carrés des facteurs retenus		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	4,086	29,188	29,188	4,086	29,188	29,188
2	3,783	27,020	56,208	3,783	27,020	56,208
3	2,020	14,426	70,634			
4	1,132	8,083	78,717			
5	,905	6,465	85,183			
6	,786	5,614	90,797			
7	,460	3,286	94,083			
8	,285	2,039	96,122			
9	,238	1,702	97,824			
10	,175	1,252	99,076			
11	,074	,530	99,606			
12	,052	,370	99,976			
13	,003	,024	100,000			
14	5,31E-005	,000	100,000			

Méthode d'extraction : Analyse en composantes principales.

Remarques :

Par rapport au tableau de la variance expliquée avant élimination des variables à écarter, on gagne environ 7 points de variance expliquée.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Matrice des composantes après élimination des variables à écarter :

Matrice des composantes(a)

	Composante	
	1	2
Douce	,429	,646
Parfumée	-,354	,545
Sucrée	,607	,157
Très sucrée	-,576	,058
Fruité	,540	,309
Légèrement acidulée	,179	,568
Acidulée	-,626	-,308
Très acidulée	,163	-,631
Croquante	,104	,628
Très croquante	,330	-,819
Juteuse	,650	-,203
Ferme	,894	-,190
REGR factor score 1 for analysis 1	,992	-,090
REGR factor score 2 for analysis 1	,091	,990

Méthode d'extraction : Analyse en composantes principales.
a 2 composantes extraites.

L'observation de la nouvelle matrice des composantes permet de renforcer les tendances perçues dès le départ :

Variables à retenir pour la composante principale 1 :

- Ferme (Corrélation ++)
- Juteuse (Corrélation +)
- Acidulée (Corrélation -)

Variables à retenir pour la composante principale 2 :

- Douce (Corrélation +)
- Parfumée (Corrélation +)
- Acidulée (Corrélation -) (Rappel : On a fait le choix plus haut d'écarter de l'interprétation la variable « Très croquante » du fait du biais inhérent à une typologie binaire sur des variables non diamétralement opposées)

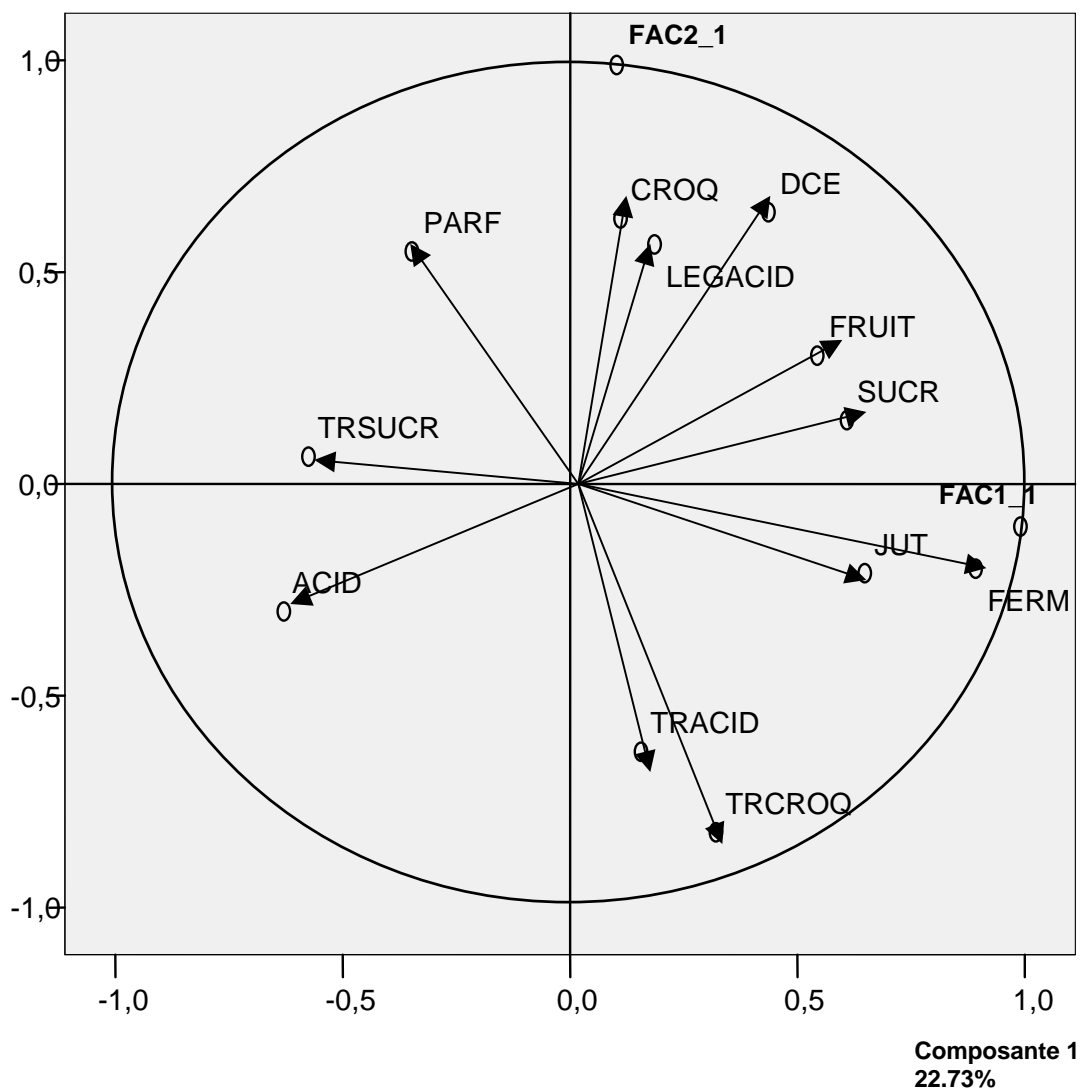
KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Cercle des corrélations et interprétation des résultats

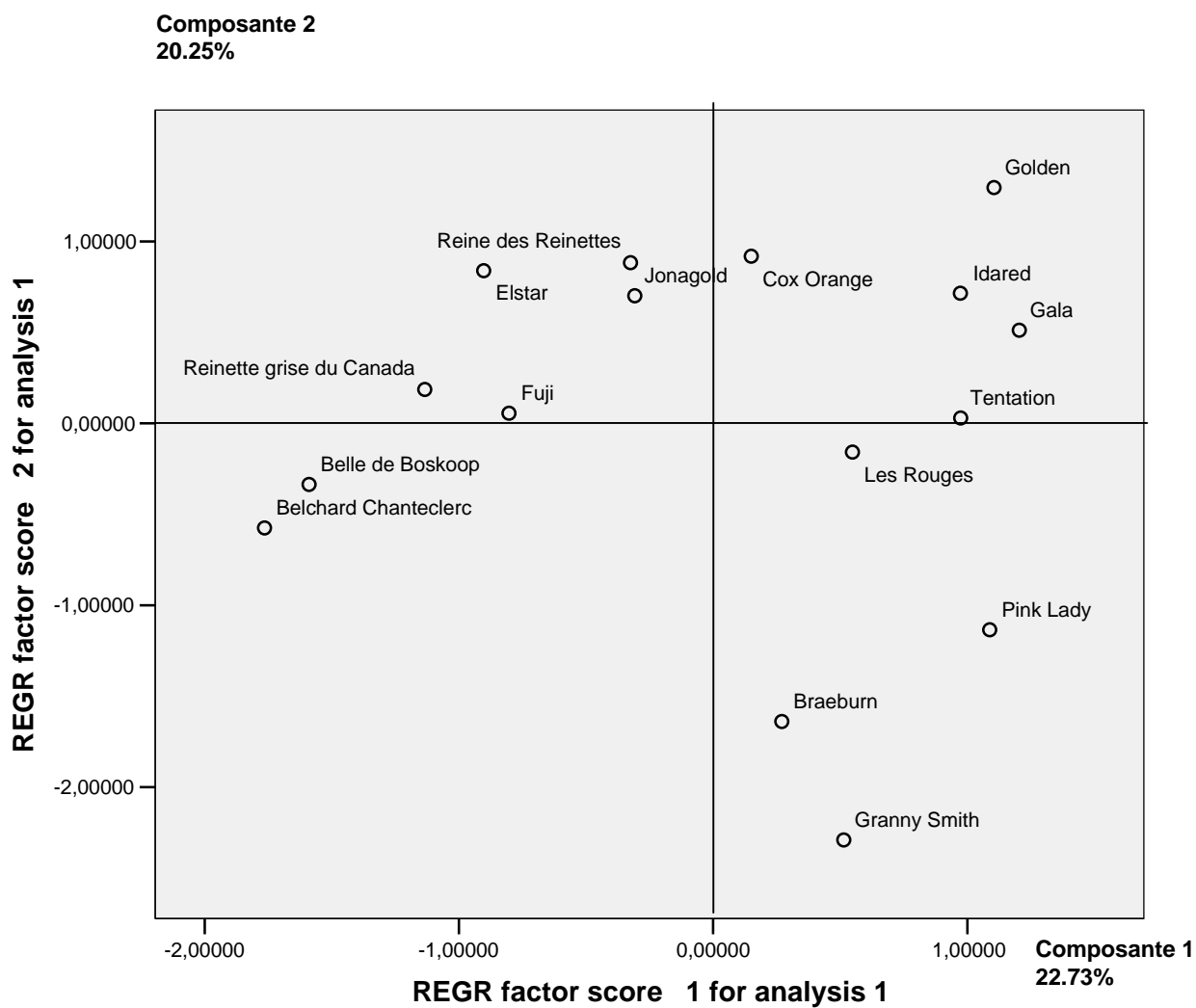
Graphique : Diagramme de composantes dans l'espace après rotation

Composante 2
20.25%



KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC





- **Construction du biplot individus/variables (Représentation simultanée des individus et des variables)**

Méthode utilisée :

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Pour des questions d'échelle, un calcul s'impose pour pouvoir positionner sur un même plan variables et individus. Nous avons repris les coordonnées des caractéristiques de pommes par rapport aux composantes principales 1 et 2 tels qu'ils ressortaient de la première analyse en composantes principales. Les valeurs des coordonnées des variables « caractéristiques » par rapport aux facteurs 1 et 2 « composantes principales » sont calculées directement dans SPSS en utilisant la fonction d'agrégation « moyenne ». La consolidation des données nous permet d'obtenir le tableau suivant, qui sert de base au lancement sous SPSS de la fonction graphique « Dispersion simple des données ». Nous obtenons le biplot inséré à la suite du tableau.

Tableau des données

Variabes	Facteur_1	Facteur_2
Golden	1,100	1,300
Gala	1,200	0,510
Reine des Reinettes	-0,320	0,880
Braeburn	0,270	-1,640
Jonagold	-0,310	0,700
Pink Lady	1,090	-1,140
Elstar	-0,900	0,840
Tentation	0,970	0,030
Fuji	-0,800	0,060
Idared	0,970	0,720
Les Rouges	0,550	-0,160
Reinette grise du Canada	-1,130	0,190
Belchard Chanteclerc	-1,760	-0,580
Belle de Boskoop	-1,590	-0,340
Granny Smith	0,510	-2,290
Cox Orange	0,150	0,920
DOUCE	0,429	0,646
PARFUMEE	-0,354	0,545
SUCREE	0,607	0,157
TRES SUCREE	-0,576	0,058
FRUITE	0,540	0,309
LEG. ACIDULEE	0,179	0,568
ACIDULEE	-0,626	-0,308
TRES ACIDULEE	0,163	-0,631
CROQUANTE	0,104	0,628
TRES CROQUANTE	0,330	-0,819
JUTEUSE	0,650	-0,203
FERME	0,894	-0,190

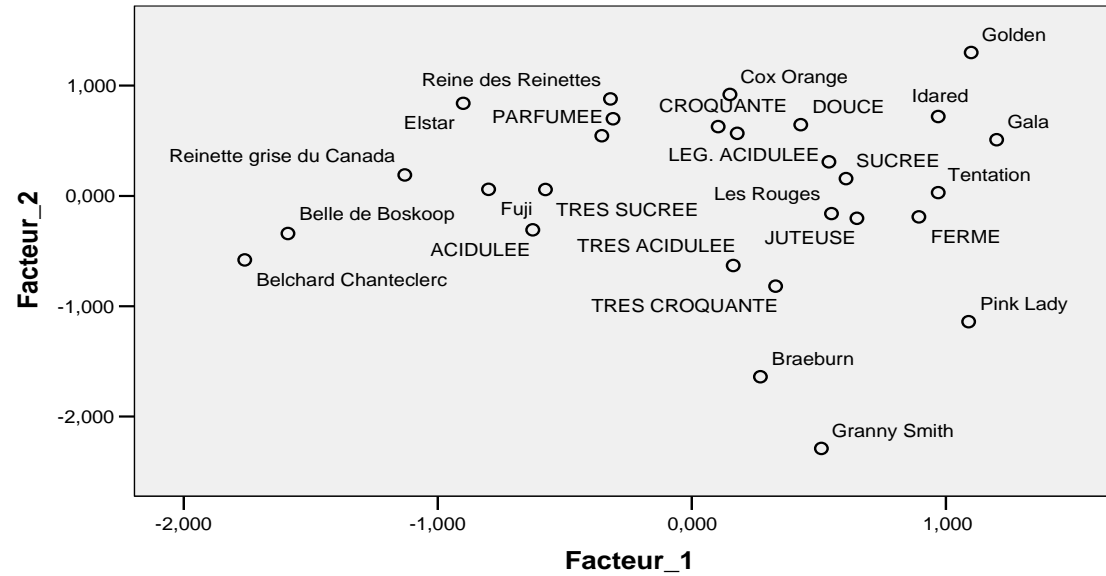
KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Graphique de représentation simultanée individus / variables (ou biplot):

Biplot Individus / Variables

Représentation simultanée des individus et des variables



KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Résultats et interprétation :

L'observation du biplot individus/variables permet d'enrichir l'interprétation. Les résultats observés seront limités aux variables fortement corrélées aux deux premières composantes principales, tels qu'elles ressortent de l'observation du cercle des corrélations.

On a cherché dans un premier temps à vérifier les résultats en observant plus particulièrement les projections des individus sur les axes des variables en s'assurant qu'ils restituent bien la répartition des données de départ. Ainsi, la projection des variétés de pommes donne le positionnement suivant : Golden, Gala, Idared, Tentation, Cox Orange. ...avec les pommes les plus douces opposées aux pommes les plus amères (Belle de Boskoop, Belchard Chanteclerc), ce qui est conforme aux données récoltées et aux résultats du premier plan factoriel.

L'analyse et l'interprétation des axes factoriels s'en trouve facilitée : La première composante principale oppose les variables « croquante », « douce », « sucrée » et « légèrement acidulée » aux variables « très sucrée » et « parfumée ». La encore, à des variables relativement équilibrées s'opposent des variables qui traduisent des caractéristiques gustatives plus soutenues. Surtout, on retrouve à l'extrémité de l'axe des pommes caractérisées par leur arôme terroir. L'opposition entre des pommes « **rustiques** » d'apparence et de goût (Reinette grise du Canada, Belle de Boskoop) , et des pommes issues de gamme plus récentes et innovantes (Gala et Golden en particulier) est très marquée. On choisit donc de retenir ses critères de différenciation pour interpréter et nommer le premier axe factoriel : A l'extrême gauche de celui-ci se trouvent des pommes « rustiques », à l'extrême droite, des pommes « sophistiquées ».

La seconde composante principale corrélée positivement aux variables « douce », « légèrement acide », « parfumée », « sucrée » qui sont des variables gustatives que l'on pourrait qualifier d' « équilibrées » ou de « tempérées », nous semble s'apparenter à un facteur de « **délicatesse** » (tout est dans la modération des qualités « gustatives »). L'observation des variétés de pommes qui se situent à proximité des centres de gravité de ces variables confirme cette interprétation ; si l'on se rapporte aux données « brutes » de départ, on retrouve les variétés de pommes dont le score dans la variable est le plus élevé à proximité directe du centre de gravité de celle-ci (En « tête de file », on retrouve la Golden par exemple)

A l'inverse, la seconde composante principale est négativement corrélée aux variables « très acidulée » et « très croquante » et dans une moindre mesure aux variables « ferme » et « juteuse ». A l'équilibre et à la modération s'opposent ici clairement des caractéristiques plus nettement formulées et plus marquées : « **très acidulée** » et « **très croquante** » (Des qualités gustatives plus soutenues). On retiendra plus particulièrement la variable « très acidulée » comme unités d'interprétation pour des raisons que nous avons déjà évoqué plus haut. Nous avons décidé d'axer l'interprétation de la manière suivante :

Composante 2 est un facteur d'échelle d'intensité gustative : Au sommet de l'axe, on retrouve les pommes « **déliçates** », appréciées pour leurs qualités gustatives et esthétiques équilibrées. A l'autre bout de l'axe, (en bas), on retrouve les pommes « **toniques** »,

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



particulièrement acidulées, croquantes, fermes et juteuses, dont la variété type est la Grany Smith.

Intuitivement, l'observation du biplot permet de dégager quatre classes de variétés de pommes : les quatre quadrants du plan représentant à peu près un type particulier de variétés. Cependant, la méthode de classification hiérarchique que nous allons à présent détailler nous a permis d'approfondir et d'infirmar cette première intuition.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



CLASSIFICATION ASCENDANTE HIERARCHIQUE (question 4)

Remarque préalable :

Les variables « Très parfumée » et « Rafrâchissante » restent écartées de l'analyse, afin de préserver la cohérence de l'analyse et de son interprétation par rapport à l'analyse factorielle des composantes principales.

Case Processing Summary^{a,b}

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
16	100,0	0	,0	16	100,0

a. Squared Euclidean Distance used

b. Ward Linkage

Nous avons réalisé une classification ascendante hiérarchique de l'ensemble des 16 pommes à l'aide du critère de Ward.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Tableau des distances entre les pommes

Proximity Matrix

Case	Squared Euclidean Distance															
	1:Golden	2:Gala	3:Reine des Reinettes	4:Braeburn	5:Jonagold	6:Pink Lady	7:Elstar	8:Tentation	9:Fuji	10:Idared	11:Les Rouges	12:Reinette grise du Canada	13:Belchard Chanteclerc	14:Belle de Boskoop	15:Granny Smith	16:Cox Orange
1:Golden	,000	3,000	4,000	7,000	6,000	5,000	6,000	5,000	8,000	2,000	4,000	6,000	11,000	8,000	9,000	3,000
2:Gala	3,000	,000	7,000	6,000	7,000	4,000	7,000	4,000	5,000	3,000	3,000	7,000	10,000	9,000	6,000	4,000
3:Reine des Reinettes	4,000	7,000	,000	7,000	8,000	7,000	6,000	7,000	8,000	4,000	6,000	4,000	7,000	4,000	9,000	5,000
4:Braeburn	7,000	6,000	7,000	,000	9,000	2,000	9,000	6,000	7,000	5,000	3,000	5,000	6,000	5,000	4,000	8,000
5:Jonagold	6,000	7,000	8,000	9,000	,000	7,000	4,000	5,000	4,000	6,000	6,000	8,000	5,000	8,000	9,000	7,000
6:Pink Lady	5,000	4,000	7,000	2,000	7,000	,000	9,000	4,000	7,000	5,000	3,000	7,000	8,000	7,000	4,000	6,000
7:Elstar	6,000	7,000	6,000	9,000	4,000	9,000	,000	7,000	2,000	6,000	6,000	6,000	7,000	6,000	7,000	7,000
8:Tentation	5,000	4,000	7,000	6,000	5,000	4,000	7,000	,000	7,000	5,000	5,000	9,000	8,000	9,000	6,000	8,000
9:Fuji	8,000	5,000	8,000	7,000	4,000	7,000	2,000	7,000	,000	6,000	4,000	6,000	5,000	6,000	5,000	7,000
10:Idared	2,000	3,000	4,000	5,000	6,000	5,000	6,000	5,000	6,000	,000	2,000	6,000	9,000	8,000	7,000	5,000
11:Les Rouges	4,000	3,000	6,000	3,000	6,000	3,000	6,000	5,000	4,000	2,000	,000	4,000	7,000	6,000	5,000	5,000
12:Reinette grise du Canada	6,000	7,000	4,000	5,000	8,000	7,000	6,000	9,000	6,000	6,000	4,000	,000	5,000	2,000	9,000	5,000
13:Belchard Chanteclerc	11,000	10,000	7,000	6,000	5,000	8,000	7,000	8,000	5,000	9,000	7,000	5,000	,000	3,000	8,000	8,000
14:Belle de Boskoop	8,000	9,000	4,000	5,000	8,000	7,000	6,000	9,000	6,000	8,000	6,000	2,000	3,000	,000	7,000	5,000
15:Granny Smith	9,000	6,000	9,000	4,000	9,000	4,000	7,000	6,000	5,000	7,000	5,000	9,000	8,000	7,000	,000	8,000
16:Cox Orange	3,000	4,000	5,000	8,000	7,000	6,000	7,000	8,000	7,000	5,000	5,000	5,000	8,000	5,000	8,000	,000

This is a dissimilarity matrix

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Chaîne des agrégations

Le tableau qui suit indique la progression de l'algorithme. Chaque classe est identifiée par son premier élément, c'est-à-dire celui qui apparaît en premier dans le fichier de données.

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	12	14	1,000	0	0	8
2	10	11	2,000	0	0	6
3	7	9	3,000	0	0	9
4	4	6	4,000	0	0	7
5	1	16	5,500	0	0	10
6	2	10	7,167	0	2	11
7	4	15	9,500	4	0	14
8	12	13	11,833	1	0	13
9	5	7	14,167	0	3	13
10	1	3	16,667	5	0	12
11	2	8	19,500	6	0	12
12	1	2	23,429	10	11	14
13	5	12	29,595	9	8	15
14	1	4	35,833	12	7	15
15	1	5	45,188	14	13	0

Ce tableau donne également le détail des regroupements de l'arbre de classification.

Plusieurs niveaux de classification se dégagent plus particulièrement de l'arbre hiérarchique (Cf. dendrogramme).

- Un premier niveau permet de distinguer 12 classes de pommes
- Un second niveau permet de dégager 10 classes de pommes
- Un troisième niveau permet de dégager 7 classes de pommes
- Un quatrième niveau permet de dégager 5 classes de pommes
- Un cinquième niveau permet de dégager 4 classes de pommes

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Choix du nombre de classes

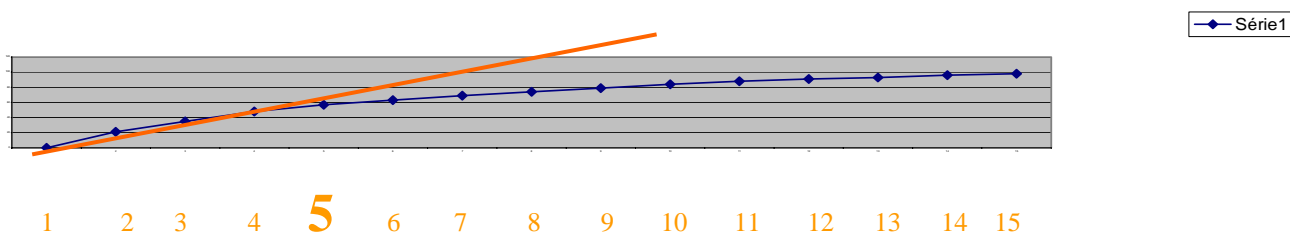
	Cluster Combined		Coefficients	Nombre de classes	Qualité de la typologie (%)
Stage	Cluster 1	Cluster 2			
1	12	14	1	15	97,79%
2	10	11	2	14	95,57%
3	7	9	3	13	93,36%
4	4	6	4	12	91,15%
5	1	16	5,5	11	87,83%
6	2	10	7,1667	10	84,14%
7	4	15	9,5	9	78,98%
8	12	13	11,8333	8	73,81%
9	5	7	14,1667	7	68,65%
10	1	3	16,6667	6	63,12%
11	2	8	19,5	5	56,85%
12	1	2	23,4286	4	48,15%
13	5	12	29,5952	3	34,51%
14	1	4	35,8333	2	20,70%
15	1	5	45,1875	1	0,00%

A l'étape initiale, chaque pomme forme une classe. Il y a donc 16 classes.



Dans le graphique ci-dessous est représenté :

- en abscisse, le nombre de classes
- en ordonnée, la qualité de la typologie (%)



KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



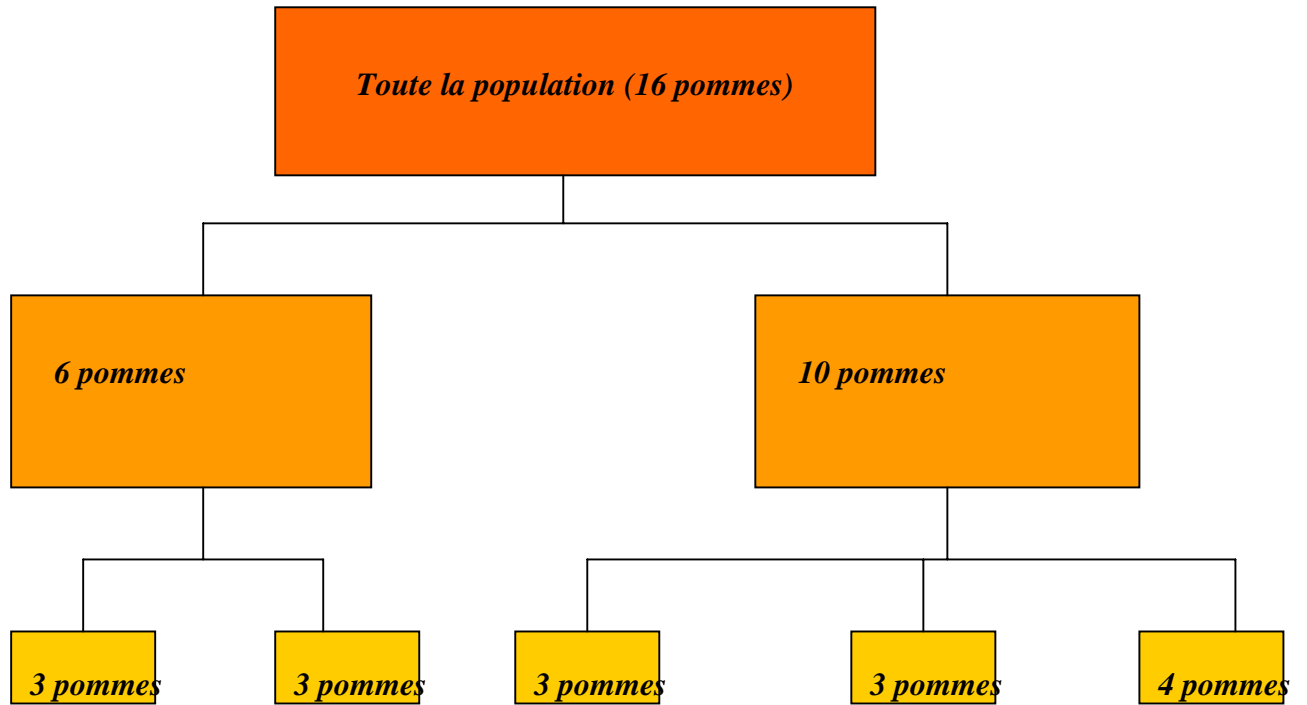
En remonte la dernière colonne du tableau pour fixer le nombre de classes. La typologie en deux classes explique 20,70% de la somme des carrés totale, la typologie en trois classes 34,51% de la somme des carrés totale, et ainsi de suite. On observe dans le graphique qui suit l'évolution du rapport entre la somme des carrés interclasses et la somme des carrés totale en pourcentage. D'où le choix de 5 classes.

Remarque : Ce regroupement infirme la pertinence d'une classification « intuitive » issue de l'observation du premier plan factoriel de l'ACP. D'où la nécessité de procéder à la classification ascendante hiérarchique, complémentaire de l'étude ACP.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Construction de la typologie



KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Qualité de la typologie en 5 classes

La qualité de la typologie est mesurée par la part de la somme des carrés totale expliquée par la typologie.

La somme des carrés expliquée par la typologie en 5 classes est égale à la somme des carrés interclasses de la typologie en 5 classes.

Somme des carrés totale = **45,188**

Somme des carrés intra classes pour la typologie en 5 classes = **19,500**

Part de somme des carrés totale expliquée par la typologie en 5 classes =

$(45,188 - 19,500)/45,188 = \mathbf{0,5684}$

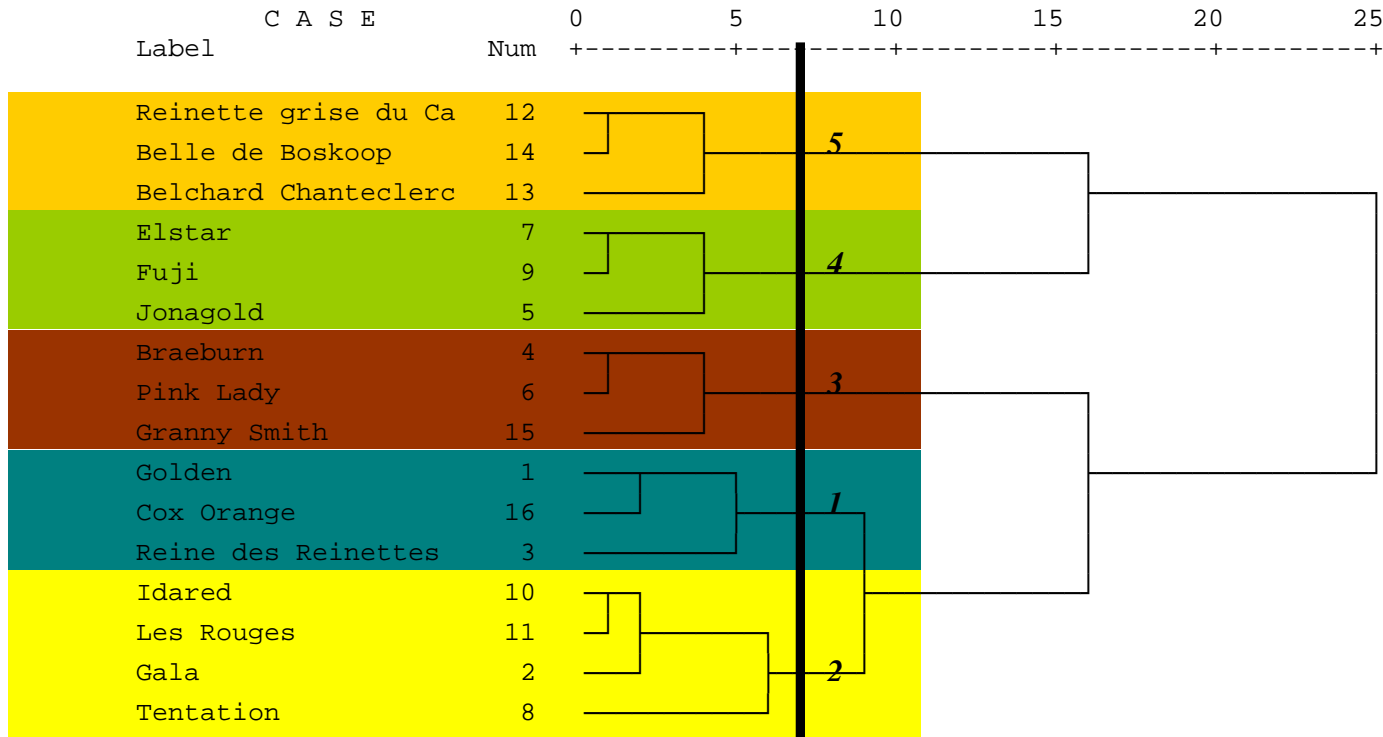


Dendrogramme ou arbre de classification : la typologie en 5 groupes explique 56,85% de la S.C. totale

* * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * *

Dendrogram using Ward Method

Rescaled Distance Cluster Combine



On voit ici nettement la construction de 5 classes.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



*Affectation des individus par classe***Cluster Membership**

Case	5 Clusters
1:Golden	1
2:Gala	2
3:Reine des Reinettes	1
4:Braeburn	3
5:Jonagold	4
6:Pink Lady	3
7:Elstar	4
8:Tentation	2
9:Fuji	4
10:Idared	2
11:Les Rouges	2
12:Reinette grise du Ca	5
13:Belchard Chanteclerc	5
14:Belle de Boskoop	5
15:Granny Smith	3
16:Cox Orange	1

Groupe n°1:

Golden
Reine des Reinettes
Cox Orange

Groupe n°2:

Idared
Tentation
Les Rouges
Gala

Groupe n°3 :

Braeburn
Pink Lady
Granny Smith

Groupe n°4 :

Elstar
Fuji
Jonagold

Groupe n°5 :

Reinette grise du Canada
Belle de Boskoop
Belchard Chanteclerc

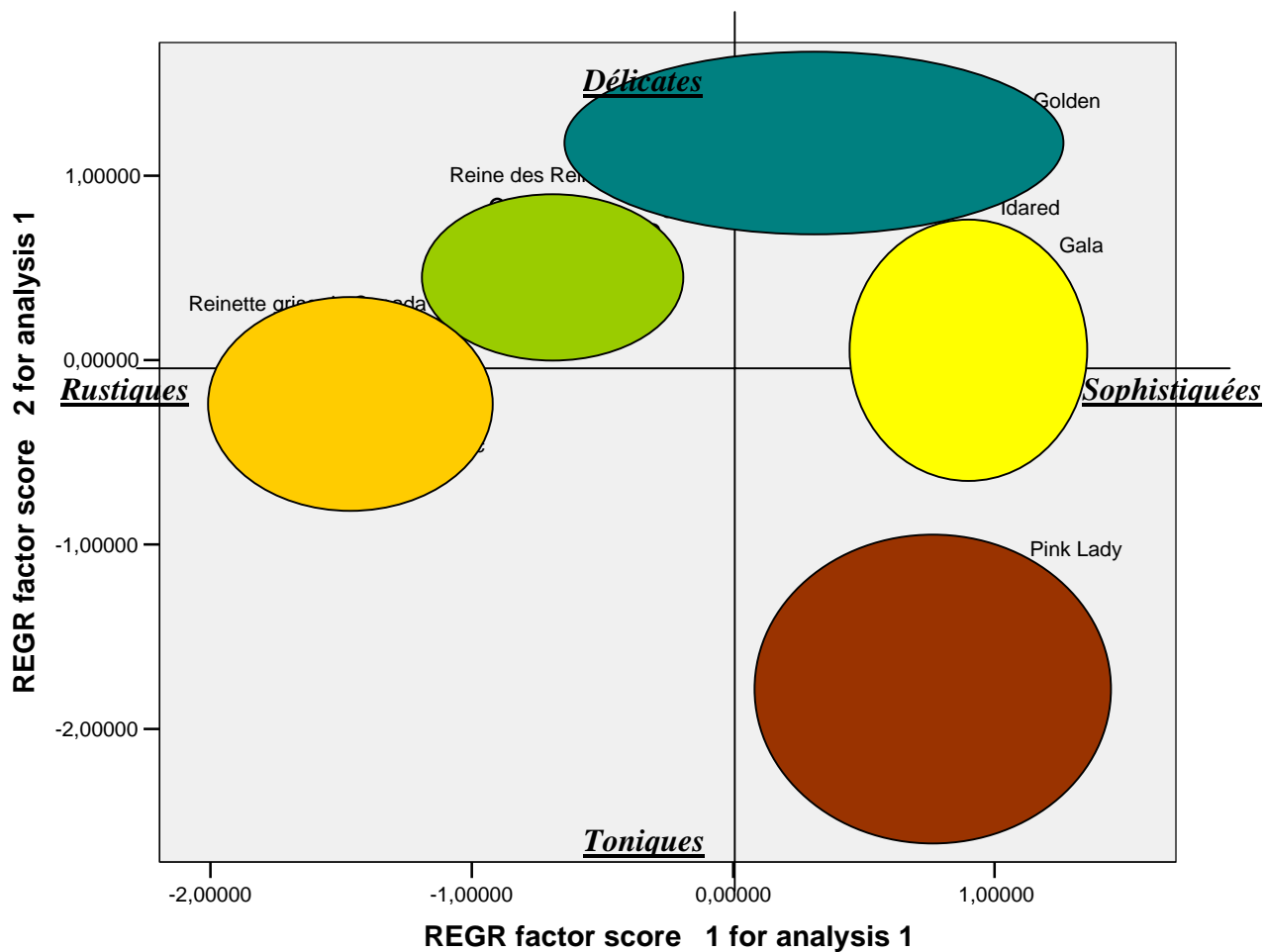
La typologie en cinq classes est représentée sur le premier plan principal de l'analyse en composantes principales réalisée sur les 16 pommes.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



D'où :

Premier plan factoriel et typologie



-  Les « équilibrées » : Fuji, Jonagold, Elstar
-  Les « délicates » : Golden, Reine des Reinettes, Cox Orange
-  Les « terroirs » : Reinettes grises du Canada, Belle de Boskoop, Belchard Chanteclerc
-  Les « toniques » : Pink Lady, Braeburn, Granny Smith
-  Les « sophistiquées » : Tentation, Les Rouges, Idared, Gala

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC

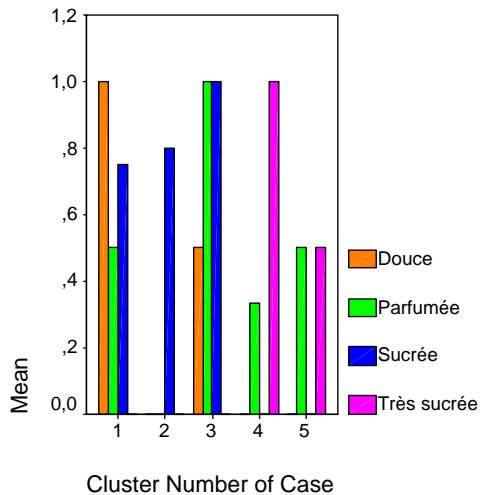


INTERPRÉTATION DES CLASSES OBTENUES (question 5)

La typologie est interprétée à l'aide des moyennes par classe des variables visualisées ci-dessous.

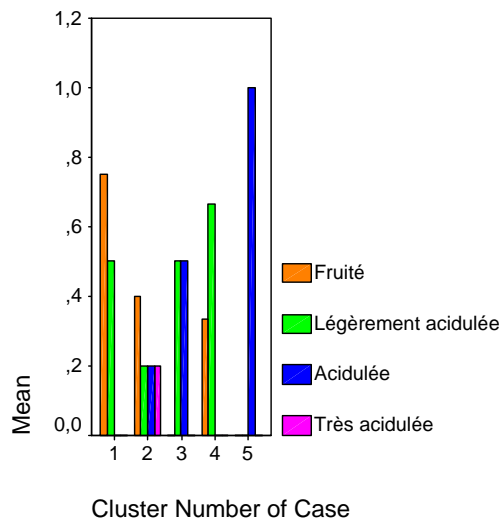
Initial Cluster Centers

	Cluster				
	1	2	3	4	5
Douce	1	1	1	0	0
Parfumée	1	0	1	1	0
Sucrée	1	1	0	0	0
Très sucré	0	0	0	0	1



Initial Cluster Centers

	Cluster				
	1	2	3	4	5
Fruité	1	1	0	0	0
Légèrement acid	1	0	1	0	0
Acidulée	0	0	0	1	0
Très acidulée	0	0	0	0	1

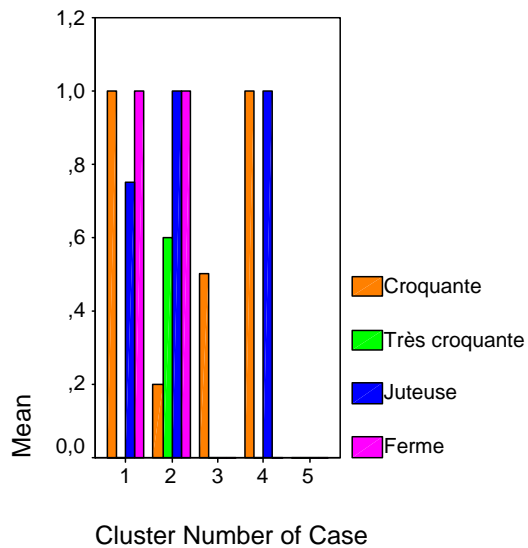


KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Initial Cluster Centers

	Cluster				
	1	2	3	4	5
Croquante	1	0	0	1	1
Très croqua	0	1	0	0	0
Juteuse	1	1	0	1	0
Ferme	1	1	0	0	1



La classe 1 est formée de pommes douces, fruitées, sucrées, croquantes, juteuses et fermes

La classe 2 est formée de pommes sucrées, très croquantes, juteuses et fermes

La classe 3 est formée de pommes parfumées et sucrées

La classe 4 est formée de pommes très sucrées, légèrement acidulées, croquantes et juteuses

La classe 5 est formée de pommes acidulées

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



Tableau de synthèse/Caractérisation des classes obtenues

	Douce	Parfumée	Sucrée	Très sucrée	Fruitée	Légèrement acidulée	Acidulée	Très acidulée	Croquante	Très croquante	Juteuse	Ferme
Classe 1	OUI	moyen	OUI	NON	OUI	moyen	NON	NON	OUI	NON	OUI	OUI
Classe 2	NON	NON	OUI	NON	moyen	moyen	moyen	moyen	moyen	OUI	OUI	OUI
Classe 3	moyen	OUI	OUI	NON	NON	moyen	moyen	NON	moyen	NON	NON	NON
Classe 4	NON	moyen	NON	OUI	moyen	OUI	NON	NON	OUI	NON	OUI	NON
Classe 5	NON	moyen	NON	moyen	NON	NON	OUI	NON	NON	NON	NON	NON

Exemples :

- « *Les Rouges* » appartiennent à la classe 2 ; elles sont en effet sucrées, très croquante juteuses et ferme
- « *Les Reinettes grises du Canada* » appartiennent à la classe 5 ; elles sont en effet acidulée et moyennement sucrées.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



ANNEXES

Présentation des différentes variétés de pommes étudiées

1. Golden



La plus connue des pommes. **Douce, parfumée, sucrée, croquante.**

Une forme régulière, une peau bien lisse, une robe jaune parfois teintée de rose, voilà qui est prometteur. Promesse tenue : on craque pour sa chair bien **sucrée**, révélant des **arômes fruités** de banane. Choisie jaune, elle est **moelleuse et douce** ; un peu verte, elle est **croquante, juteuse et un peu acidulée**. Un plaisir tout au long de l'année avec cette pomme à croquer entre août et juin ! Coloration jaune verdâtre, légère coloration rouge côté soleil.

Chair blanche-jaunâtre, **juteuse, ferme, fine, sucrée, légèrement acidulée et agréablement parfumée.**

2. Gala



Arôme **fruité, rafraîchissante, sucrée.**

Sa robe striée mélange les rouges et les jaunes pour mieux révéler un trésor de **chair ferme**. Croquée, elle libère un goût **bien sucré** et des **arômes fruités**, de banane et de poire. Politesse des reines oblige, elle est une des premières pommes de la rentrée et se distingue jusqu'en février. Sa chair est jaune, **ferme, croquante, juteuse, peu acidulée douce et sucrée.**

3. Reine des Reinettes



Légèrement acidulée, sucrée, arôme de miel.

Cette petite pomme rouge orangée, couronnée de jaune d'or vous offre une **chair sucrée, discrètement parfumée**; une texture dense et **fondante** qui libère des arômes de noisette.

Très prisée, elle nous offre ses quartiers d'août à octobre : une courte saison dont il faut profiter! Chair blanc-crème, **douce acidulée et parfumée.**

4. Braeburn



Acidulée et sucrée, très croquante, juteuse.

Que recèle cette belle pomme bicolore, aux odeurs de feuilles vertes et d'herbe coupée ?

Un goût très plaisant, à la fois **sucré et acide**, mais sans excès et une texture **ferme, croquante et juteuse**. De la séduction à revendre de novembre à avril !

Chair **ferme, juteuse et bien sucrée.**

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



5. Jonagold



Juteuse, sucrée, arôme fruité.

*Elle annonce la couleur pour mieux se laisser croquer et révéler son goût : des odeurs de pomme mûre et des arômes de poire qui imprègnent sa chair **juteuse et sucrée à souhait**.*

De quoi rassembler les amateurs du mois d'octobre jusqu'à l'été...

*Chair jaunâtre, **mi-ferme, parfois lâche, croquante, juteuse, très sucrée, légèrement acidulée, très parfumée, fond harmonieux.***

6. Pink Lady®



Sucrée, très croquante, arôme fruité.

*Ma chair de couleur blanc-crème est **croquante, ferme et juteuse.***

7. Elstar



Sucrée et acidulée, juteuse, parfumée.

Son nom évoque les paillettes et ne manque pas de personnalité.

*Des odeurs de pomme mûre et de poire, une robe bicolore toute de rouge et de jaune, une chair parfumée aux arômes de poire. Enfin un bel **équilibre entre acidulé et sucré** et une texture juteuse.*

*On comprend que sa présence entre août et mars soit si appréciée. Chair blanche-jaunâtre, **croquante, juteuse, très sucrée avec une acidité rafraîchissante, agréablement parfumée.***

8. Tentation®



Très parfumée et sucrée, juteuse et croquante, arôme fruité.

Elle porte bien son nom ! Comment résister à son parfum d'anis qui la rend unique ?

A sa forme allongée si particulière ? A cet atout majeur qu'est sa couleur - jaune moucheté avec une pointe de rose ? Comment ne pas céder au bouquet d'arômes fruités et de fleurs qu'elle libère en bouche, avec en prime croquant et juteux ? Voilà une Tentation® qui désaltère de décembre à avril.

*Le fruit est couleur jaune doré, son **parfum est très typé, floral, avec des notes de noisette et une très légère touche mentholée.** Les saveurs **sucrées et acidulées s'équilibrent.** La chair est agréable et **très rafraîchissante.** Sa texture est **ferme et juteuse.***

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



9. Fuji



Sucrée et rafraîchissante, croquante, juteuse.

Epiderme vert à lenticelles en étoile - chair blanc-verdâtre, fine, **tendre, juteuse, très sucrée** mais **peu acidulée** et **peu parfumée**. Tout en elle invite au voyage : sa forme bien ronde, sa peau vert délavé sur fond rose, son goût agréablement sucré. Mais c'est aussi et surtout son **exceptionnelle jutosité** qui fait de la Fuji une pomme **très rafraîchissante** de janvier à juin, toute en parfums de feuilles et de bois verts...

10. Idared



Légèrement acidulée, croquante.

Son aspect est diversement apprécié, mais coupons-la : des odeurs de bois vert envahissent l'espace. En bouche, l'impression se confirme : c'est un fruit **croquant et ferme**. La texture est son point fort et la plénitude de ses arômes végétaux - la pomme verte notamment - ne gâche rien ! On aime cette pomme qui se croque de janvier à juin.

Chair blanche, fine, demi-croquante, **juteuse, douce, sucrée, acidulée**. **Arôme léger**, les fruits pleinement développés ont un goût agréable.

11. Les Rouges



Délicatement sucrée, croquante.

Comme leurs noms l'indiquent, ces pommes attirent d'abord l'œil par leur couleur rouge et brillante. Puis c'est au tour de l'odorat d'être comblé, lorsque coupée en deux, elles exhalent leurs parfums de bois vert. Mais c'est en bouche qu'elles révèlent leur point fort : une **chair ferme et croquante, délicatement sucrée**. Les rouges sont de mise entre octobre et fin avril !

Chair blanche, fine, **ferme, juteuse** douçâtre.

12. Reinette grise du Canada



Acidulée, croquante, arôme terroir.

Très originale avec sa belle couleur bronze, on pourrait mal la juger par sa forme parfois irrégulière. Elle est tout simplement nature ! **Croquante, crissante et feutrée** comme la neige, les parfums de bois et de sous-bois de cette pomme authentique gagnent à être connus ! Tout comme ses arômes terroir et agrume, **un brin acide, un brin sucré...** à découvrir entre novembre et mars.

Chair blanche jaunâtre, **mi-tendre, sucrée, fine, parfumée** et dorée d'une **acidité discrète** et agréable.

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC



13. Belchard® Chanteclerc



Sucrée, acidulée, très parfumée.

Connaissez-vous la Belchard® CHANTECLER ? Voilà une pomme qui gagne à être connue ! Elle a une forme légèrement aplatie et une robe très caractéristique, jaune d'or, un peu rugueuse parsemée de lenticelles qui lui donnent un aspect rustique.

Chair très sucrée et acidulée, très bonne qualité.

14. Belle de Boskoop



Acidulée, parfumée.

Que cache-t-elle sous sa peau mate, ses formes rondes et sa robe striée où se mêlent le jaune, le vert, le rouge et le brun ? Réponse : une chair blanche où se niche des arômes plutôt verts, peu sucrés, marqués d'agrumes et d'acidité... pour se faire plaisir entre octobre et mars !

15. Granny Smith



Très acidulée et croquante, rafraîchissante, ferme et juteuse.

C'est le moment de se mettre au vert avec cette belle pomme aux parfums d'agrumes, bien ronde et régulière, à la peau très brillante.

Rien que du croquant, de la fermeté, une chair juteuse et des arômes végétaux à dévorer à pleines dents d'octobre à mai.

16. Cox Orange



Chair jaunâtre, fine, tendre à moyennement ferme, croquante et fondante à la fois, arôme très agréable,

KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC





KHALIDI Marie-France et OUEDRAOGO Soukeyna
IAE Paris/HEC

