

Séance 11 : Typologies

Sommaire

Proc CLUSTER : Typologie hiérarchique	3
Proc FASTCLUS : Typologie nodale	8
Proc MODECLUS : Typologie non paramétrique.....	11

- Les phénomènes observés (attitudes, comportements, préférences) ne sont souvent que la résultante de l'agrégation de comportements ou attitudes parfois très différents.
- La moyenne est donc trompeuse et il faut prendre en compte l'hétérogénéité sous-jacente, c'est la problématique de la segmentation. Poussée à l'extrême cette segmentation conduit au « marketing personnalisé » (one-to-one).
- Ces variables vont jouer le rôle de variables modératrices et influencer (1) le niveau de la moyenne, (2) la sensibilité du comportement à une autre variable.
- Pour qu'une segmentation soit intéressante sur le plan managérial les segments de clients doivent être
 - o **Pertinents** sur le phénomène observé : (1) *homogènes* au sein du groupe, (2) *différents* entre les groupes
 - o **Opérationnels** : (3) *identifiables* pour pouvoir faire l'objet d'un marketing ciblé et (4) d'une *taille suffisante* pour que cette personnalisation soit rentable
- Les variables de segmentation sont
 - o **Objectives, liés à l'individu**, ce sont les déterminants économiques (revenu, richesse, propriété du logement), démographiques (sexe, âge-génération, composition du foyer, présence enfants) ou socio-culturels (pays, région, urbanité) classiques qui dirigent les comportements. Pour le B_to_B la taille de l'entreprise, son secteur d'activité, ...
 - o **Subjectives**, il s'agit alors de variables latentes individuelles proposées en marketing pour expliquer les comportements :
 - **liés à l'individu** : leadership d'opinion, innovativité, culture,...
 - **liés au phénomène étudié** : implication (dans un achat, une catégorie de produit), expertise de l'individu, satisfaction d'une transaction ou d'une relation, attachement/fidélité à la marque,...
 - o **Comportementales**, avec la fréquence des transactions (Frequency), la durée de la relation (première et dernière transaction) (Recency), le montant du panier (Monetary), le type de marchandises achetées (Type) ou le profil d'interaction avec l'entreprise (canal utilisé, services, ...). Ce qui donne les segmentation RFM et FRAT.
- La méthode de segmentation détermine comment la sélection des variables constituant les groupes va être faite
 - o **Segmentation « a priori »** : les variables déterminantes sont connues et présélectionnées (variables objectives). On recherche un lien direct entre ces variables (X) et une variable de comportement (Y) dans une démarche *explicative* de sa variance (à base de χ^2 ou d'anova) : Typologie monothétique ou segmentation par arbre.

- **Segmentation « a posteriori »** : la variable qui va servir de variable de segmentation est construite à partir des informations collectées. Les groupes sont constitués pour restituer la diversité (variance) *sur cette variable*. On relie ensuite l'appartenance à un groupe avec un comportement d'achat. C'est une typologie polythétique ou simplement typologie (nodale, Hiérarchique)
 - **Segmentation « concomitante »** : c'est une approche hybride dans laquelle on cherche simultanément à obtenir une bonne qualité explicative du comportement (Y) et une bonne cohérence (variance de X) des groupes ainsi constitués.
- Principes d'une typologie
- Bien choisir les variables, leur distribution et les pré-traitements (valeurs aberrantes, standardisation), leurs relations (corrélation, ACP), leurs transformations (variable nominale en variable quantitative par AFACO).
 - Choisir une fonction de distance : Comment définir « proches » ? (distances, similarité, dissimilarité)
 - Choisir la nature de la typologie
 - Appartenance exclusive simple (0001)
 - Appartenance multiple (0101)
 - Appartenance probabiliste (floue) (0.5 0.1 0.2 0.2)
 - Choisir la méthode :
 - Décomposition hiérarchique de l'échantillon de 1 groupe à n groupes (typologie hiérarchique)
 - Solution « spatiale » unique d'un nombre déterminé de groupes (typologie nodale)
 - Approche séquentielle : hiérarchique sur un sous-échantillon puis nodale.
 - Approche itérative ou par simulation :
 - quels points se retrouvent toujours ensembles (nuées dynamiques et formes fortes) ?
 - quels points sont à neutraliser ?
 - Choisir l'algorithme :
 - comment décider de regrouper deux segments ? (hiérarchique)
 - quelle solution de départ (seed) ? (nodale)
 - Combien de groupes faut-il retenir ?
 - Les critères internes (sur la variance des variables X) : Maximiser la variance inter-groupes / minimiser la variance intra-groupe, Pseudo R², Pseudo F (pseudo car la même information est utilisée pour constituer les groupes et mesurer la qualité de la typologie)
 - Les critères externes : capacité à expliquer une variable comportementale,
 - Souvent entre 3 (intérêt de la segmentation) et 8 (capacité à gérer les groupes).
 - Difficultés : faiblesse du cadre théorique
 - Est-ce qu'une typologie existe vraiment ?
 - Pas de critères explicites pour guider ces différents choix
 - Il existe toujours plusieurs solutions (! même avec le même algorithme, par exemple en fonction des valeurs de départ)



Proc CLUSTER : Typologie hiérarchique

Caractéristiques :

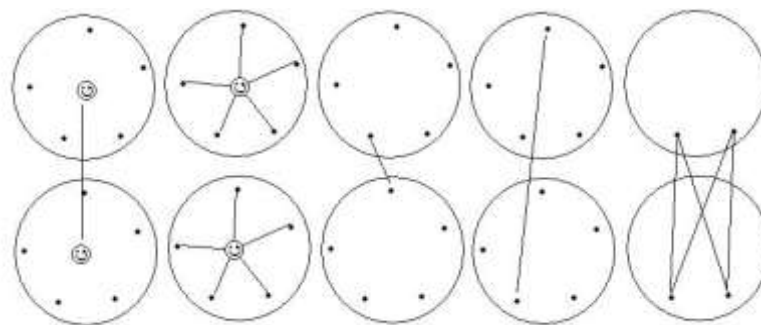
- Peu adapté à de gros volumes de données
- Choix délicat entre les nombreuses méthodes donnant des résultats différents
- Choisir la méthode de Ward si les données sont quantitatives (elle donne des groupes sphériques)
- Choix important dans le choix des variables (... « je suis beau »), des pré-traitements des données (standard, factorielle,...) et des individus (aberrants)
- Statistiques sur les variances inter et intra cluster, et pseudo Anova (pseudo F) car la variable est un découpage construit a posteriori qui maximise la variance inter-groupes

Principe :

- Approche séquentielle offrant un arbre typologique (dendrogramme, tree) de 1 seul groupe à n groupes (autant qu'il y a d'individus)
 - o Ascendante : partir de n et regrouper progressivement
 - o Descendante : partir de 1 groupe et décomposer progressivement.
- C'est ensuite à l'analyste de décider « où couper » l'arbre (choisir le nombre de groupes). Le dendrogramme reporte les distances inter-groupes pour chaque fusion/scission des clusters (R^2 semi-partiel)
- Il permet d'identifier les regroupements les plus pertinents (faible section) et ceux qui « coûtent » le plus en termes de perte d'information (longs segments horizontaux)
- (ne pas le demander si les individus sont trop nombreux !)

Choix du regroupement

- Source : <http://v8doc.sas.com/sashtml/stat/chap23/sect4.htm>
- (1CENTROID) Méthode des centroïdes : Regroupement selon la proximité des centres de gravité (centroïdes)
- (2 WARD) Méthode de la variance (Ward) : Minimisation de la variance intra-groupe
- (3 à 5 SINGLE, AVERAGE, COMPLETE) Méthode de chaînage par le plus proche voisin : distances entre les points de chaque groupe : simple (le plus court, le plus long), moyen, complet,



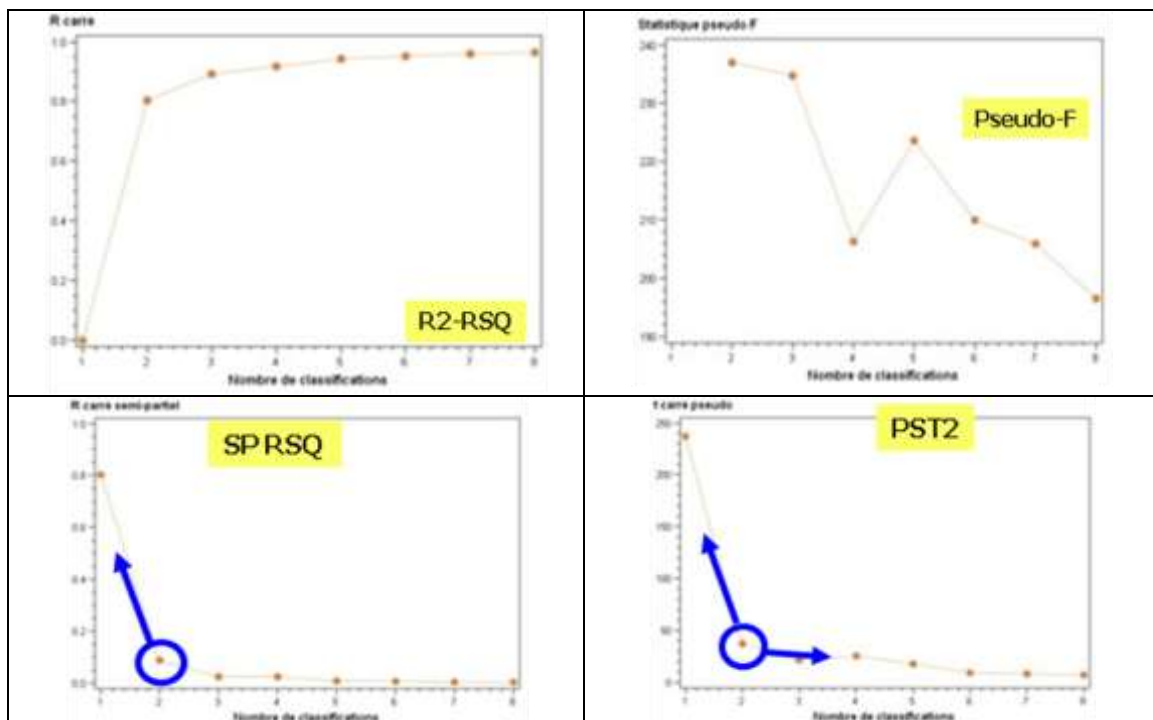
Options usuelles

- | | |
|--|----------|
| - create output data set | OUTTREE= |
| - Copy variables in data set to the | COPY |
| - specify clustering method | METHOD= |
| - suppress normalizing of distances | NONORM |
| - suppress squaring of distances | NOSQUARE |
| - standardize variables | STANDARD |
| - omit points with low probability densities | TRIM= |
| - cubic clustering criterion | CCC |

- | | |
|---|---------|
| - pseudo F and t2 statistics | PSEUDO |
| - statistiques simples (my, écart-type,...) | SIMPLE |
| - root-mean-square standard deviation | RMSSTD |
| - R2 and semipartial R2 | RSQUARE |

Choisir le nombre de clusters

- Général : Les critères ne sont valides que pour des clusters « compacts » (non étirés). Il faut rechercher un consensus entre ces statistiques : une forte valeur pour CCC et pseudo F combinée avec un accroissement important du pseudo t² pour un découpage supplémentaire.
- Les critères globaux reconstitution de la variance totale
 - o RMSSTD (root-mean-square standard deviations) écart-type
 - o RSQ (R²) : % de variance représenté par les clusters (intra) = $1 - \frac{\sum w_k}{VT}$. A maximiser mais sans trop de groupes
 - o PSF (Pseudo F) écart entre tous les groupes = $\frac{(T - \sum w_k) / (G - 1)}{\{(\sum w_k) / (n - G)\}}$. A maximiser
- Les critères marginaux (contribution d'un découpage supplémentaire) S'arrêter à celui qui commence le « plat » juste après une chute
 - o SPRSQ (Semipartial R-Squared) : % de variance expliquée gagnée par le regroupement de 2 clusters. Choisir juste avant une forte hausse
 - o PST2 (Pseudo t²) écart entre les deux derniers clusters regroupés (marginal). Le meilleur découpage a une forte différence de pente avant (petite) et après (forte)



- CCC (Cubic Clustering Criterion) http://support.sas.com/documentation/onlinedoc/v82/techreport_a108.pdf
 - o Rapport du R² empirique sur le R² théorique d'une distribution uniforme (sans cluster)
 - o Hypothèses : les données sont échantillonnées à partir d'une distribution
 - H0 : distribution uniforme
 - H1 : d'un mélange de distributions multivariées normales sphériques aux variances égales
 - o efficacité des différents niveaux de regroupement
 - > à 2 ou 3 = OK; 0-2 à valider ;

- largement <0 (-30) : problème avec des individus très atypiques à retirer
- Interprétation en fonction du nombre de clusters
 - Si clusters existent : augmentation rapide suivie d'une évolution régulière
 - Si $CCC < 0$ et en baisse pour >2 groupes, distribution unimodale ou asymétrique
 - Si CCC augmente toujours, les données sont granulaires ou manquent de finesse (arrondies)

Root-Mean-Square Total-Sample Standard Deviation = 10.89784

Root-Mean-Square Distance Between Observations = 30.82375

Historique des classifications										
NCL	Classifications jointes		FREQ	SPRSQ	RSQ	ERSQ	CCC	PSF	PST2	T
8	CL16	CL24	8	0.0037	.964	.944	3.50	197	7.2	
7	CL15	CL18	13	0.0047	.959	.936	3.65	206	8.4	
6	CL9	CL8	18	0.0078	.951	.926	3.54	210	9.3	
5	CL13	CL11	20	0.0090	.942	.911	3.71	224	17.8	
4	CL7	CL12	16	0.0250	.917	.888	1.87	206	25.4	
3	CL6	CL10	24	0.0253	.892	.842	2.55	235	21.4	
2	CL4	CL3	40	0.0884	.803	.709	3.23	237	37.3	
1	CL5	CL2	60	0.8034	.000	.000	0.00	.	237	

Master Marketing Paris-Dauphine
Source : Analyse des données appliquée au marketing

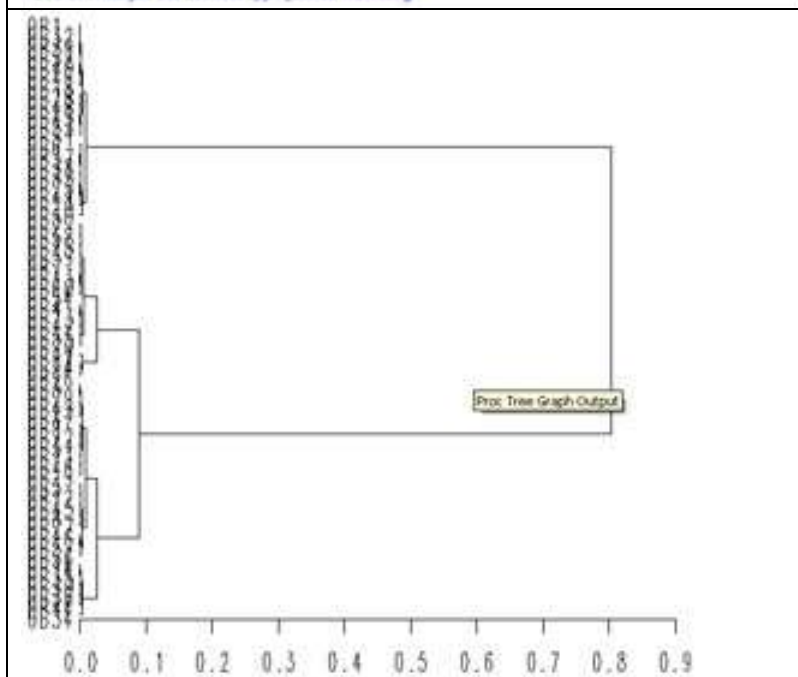
Choisir 3 groupes

car passer à 4

- ne contribue que peu à l'amélioration du RSQ (SPRSQ faible)
- n'améliore pas beaucoup RSQ
- fait beaucoup baisser le CCC et le PSF

Passer à 2

- donne un meilleur CCC
- mais au prix d'un plus mauvais ajustement global (détériore le RSQ)



Exemple 1 : distances intercités USA (entrée directe des distances)

```
Title 'Cluster Analysis of Flying Mileages Between 10 American Cities';
Data USA(type=distance); /* entrée directe de la demi-matrice des distances */
  input (atlanta chicago denver houston losangeles
        miami newyork sanfran seattle washdC) (5.)
        @55 city $15.;
Cards;
  0 ATLANTA
  587 0 CHICAGO
  1212 920 0 DENVER
  701 940 879 0 HOUSTON
  1936 1745 831 1374 0 LOS ANGELES
  604 1188 1726 968 2339 0 MIAMI
  748 713 1631 1420 2451 1092 0 NEW YORK
  2139 1858 949 1645 347 2594 2571 0 SAN FRANCISCO
  2182 1737 1021 1891 959 2734 2408 678 0 SEATTLE
  543 597 1494 1220 2300 923 205 2442 2329 0 WASHINGTON D.C.
;
/* Ward's minimum variance with pseudo-F and t2 statistics */
ODS graphics on ;
Proc Cluster data=mileages plots = all
  outtree = tree_USA
  method=ward /* méthode */
  RSQ pseudo ; /* statistiques : pas de CCC pour des distances*/
  id city; /* identifiant */
run;
Proc Tree data= tree_USA
  horizontal spaces=2;
  id city;
run;
```

Résultats

Historique des classifications		FREQ	SPRSQ	RSQ	PSF	PST2	T
NCL	Classifications jointes						i
9	NEWYORK WASHINGTON D.C.	2	0.0019	.998	66.7	.	
8	LOS ANGELES SAN FRANCISCO	2	0.0054	.993	39.2	.	
7	ATLANTA CHICAGO	2	0.0153	.977	21.7	.	
6	CL7 CL9	4	0.0296	.948	14.5	3.4	
5	DENVER HOUSTON	2	0.0344	.913	13.2	.	
4	CL8 SEATTLE	3	0.0391	.874	13.9	7.3	
3	CL6 MIAMI	5	0.0586	.816	15.5	3.8	
2	CL3 CL5	7	0.1488	.667	16.0	5.3	
1	CL2 CL4	10	0.6669	.000	.	16.0	

- Premier regroupement NY et W DC : il apporte 0.0019 en explication de la variance totale. Il reste donc (RSQ = 1- SPRSQ = 0.0998). Le pseudo F indique que c'est regroupement très intéressant.
- A la fin le pseudo t² indique que le dernier regroupement est peu pertinent (brûque augmentation du t² partiel).



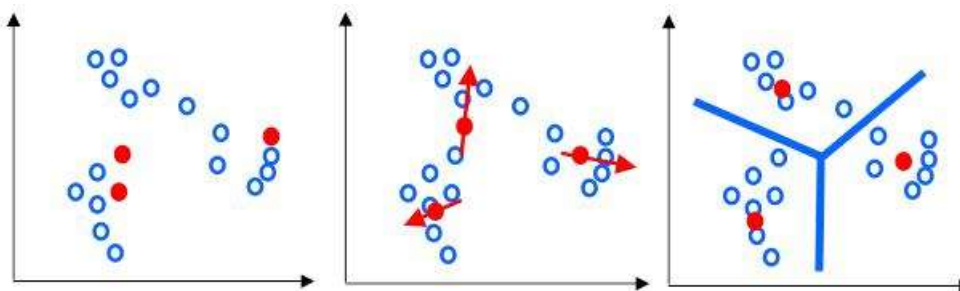
Proc FASTCLUS : Typologie nodale

Caractéristiques :

- Méthode des « K-means »
- Adaptée à de très gros volumes de données (-ne stocke pas la matrice des distances)
- Apporte une solution unique en fonction des choix initiaux donc
 - o tester différents nombre de clusters pour choisir le meilleur
 - o tester différentes valeurs de départ
 - o ou donner des valeurs initiales issues de Proc cluster

Principe :

- Choix initial des paramètres et des « points de départ »
 - o un nombre de groupes et
 - o un seuil de regroupement (S)
 - o les points de départ des groupes (seed)
 - o le calcul des centres de gravité en lot ou immédiat (Drift)
- L'algorithme calcule ensuite la quantité de déplacement : $DAB = 2 \cdot dAB \cdot (NA \cdot NB / (NA + NB))$
 - o si $DAB \leq$ Seuil regroupement de A et B
 - o sinon nouveau cluster ou individu « non classé »
- Le calcul du centre de gravité peut être effectué
 - o après chaque affectation
 - o après le traitement du lot (batch) par une passe sur l'ensemble des données



Options usuelles:

- SEED=SAS-data-set specifies an input data set from which initial cluster seeds are to be selected.
- REPLACE=RANDOM : selects a simple pseudo-random sample of complete observations as initial cluster seeds.
- OUT=SAS-data-set creates an output data set to contain all the original data, plus the new variables CLUSTER and DISTANCE
- OUTSEED=SAS-data-set is another name for the MEAN= data set, provided because the data set may contain location estimates other than means.
- OUTSTAT=SAS-data-set creates an output data set to contain various statistics
- INSTAT=SAS-data-set : reads a SAS data set previously created by the FASTCLUS procedure using the OUTSTAT= option. Only cluster assignment and imputation are performed as an OUT= data set is created.
- DISTANCE computes distances between the cluster means.
- DRIFT : les centres de classes sont ajustés après chaque observation
- DELETE=n supprime les centres avec $\leq n$ individus rattachés (pour éviter d'avoir pour centre un « outlier »)

- MAXITER=n specifies the maximum number of iterations for recomputing cluster seeds.

Exemple 2

```

options nocenter formdlim="-" dflang=Swiss_French; ;
title1 h=2 j=1 "TYPOLOGIES Hiérarchique et Nodale";
title2 h=1 j=1 "Données MM";
title3 ; * une ligne blanche pour séparer les titres ;

title4 "1- Lecture ";
footnotel h=0.5 j=1 'Master Marketing Paris-Dauphine';
footnote2 h=0.5 j=1 'Source : Analyse des données appliquée au marketing ' j=r
'2012 (c) Pierre Desmet';
data in;
  input q1-q4 type @@ ; * @@ indique que les variables sont en continu ;
  ID=_N_ ; * création d'un numéro automatique dans i (car _N_
existe dans le système) ;

cards ;
50 33 14 02 1 64 28 56 22 3 65 28 46 15 2
67 31 56 24 3 63 28 51 15 3 46 34 14 03 1
69 31 51 23 3 62 22 45 15 2 59 32 48 18 2
46 36 10 02 1 61 30 46 14 2 60 27 51 16 2
65 30 52 20 3 56 25 39 11 2 63 30 55 18 3
58 27 51 19 3 68 32 59 23 3 51 33 17 05 1
57 28 45 13 2 62 34 54 23 3 77 38 67 22 3
63 33 47 16 2 67 33 57 25 3 76 30 66 21 3
49 31 15 02 1 77 26 69 23 3 60 22 50 15 3
54 39 17 04 1 66 29 46 13 2 52 27 39 14 2
60 34 45 16 2 50 34 15 02 1 44 29 14 02 1
50 20 35 10 2 55 24 37 10 2 58 27 39 12 2
47 32 13 02 1 46 31 15 02 1 51 34 15 02 1
50 35 13 03 1 49 31 15 01 1 67 31 47 15 2
54 37 15 02 1 56 30 41 13 2 63 25 49 15 2
61 28 47 12 2 64 29 43 13 2 54 39 13 04 1
51 35 14 03 1 45 23 13 03 1 51 37 15 04 1
52 35 15 02 1 53 37 15 02 1 67 30 50 17 2
63 33 60 25 3 58 28 51 24 3 57 25 50 20 3
68 30 55 21 3 64 27 53 19 3 65 32 51 20 3
;
ODS graphics on ;
*****;
* proc CLUSTER : typologie Hiérarchique *;
* ici choix de la méthode WARD *;
* sur les données brutes *;
*****;
Proc Cluster data=in plots = all
  outtree = tree_MM /* pour faire le dendrogramme */
  method=ward /* méthode */
  RSQ pseudo ; /* statistiques */
  id ID; /* identifiant */
run;
*****;
* proc Tree : dendrogramme de la typologie Hiérarchique *;
* horizontal (sinon vertical) *;
* sortir un fichier avec und écoupage en 3 clusters *;
*****;
Proc Tree data= tree_MM
  horizontal spaces=2
  out= clust_group nclusters = 3; /* nombre de clusters retenu */
  id ID; run;
  
```

```

proc print data=clust_group (obs=20); run ;

*****;
* Affectation du cluster aux données de base (Merge après Tri)      *;
*****;
proc sort data=in ;          by ID ; run ;
proc sort data=clust_group ; by ID ; run ;
data in_2 ;
  merge in clust_group (keep =ID cluster) ;
  by ID ;
run ;

*****;
* interprétation des clusters                                     *;
*****;
title4 "Graph CLUSTER ";
proc gplot data= in_2;
  plot q3*q2=cluster;
  run ;

proc tabulate data=in_2 ;
  class cluster ;
  var q2-q3 ;
  table cluster, (q2 q3)*mean ;
  run ;

*****;
* Calcul des moyennes pour servir de "seed" dans une proc FastClus *;
*****;
proc sort data= in_2 ; by cluster ;run ;
proc means data=in_2 ; by cluster ;
  var q2 q3 ;
  output out=clust_seed mean= q2 q3 ;
  run ;
proc print data=clust_seed;
  run ;

*****;
* proc FASTCLUS : typologie nodale                               *;
*****;
title4 "Typo nodale ";
proc fastclus data= in
  maxc=3          /* maxc= 3 : on veut 3 groupes */
  maxiter=10     /* maximum 10 itérations (vérifier la convergence*/
  out=clus1      /* l'appartenance au groupe est dans CLUSTER */
  outstat = data_stat; /* résultats statistiques*/
  var q2-q3;
run ;

*****;
* contenu du data : datastat pour la qualité de la typo        *;
*****;
proc print data=data_stat ;run ;

proc gplot data= clus1;
  plot q3*q2=cluster;
  run ;

*****;
* proc FASTCLUS : typologie nodale sur les seed du Culster     *;
*****;
title4 "Typo nodale sur seed Cluster ";

```

```

proc fastclus data= in
    seed = clust_seed
           /* drift : recalcul des centres à chaque affectation */
    maxc=3 /* maxc= 3 : on veut 3 clusters */
    maxiter=10 /* maximum 10 itérations (vérifier la convergence*/
    out=clus2 /* l'appartenance au groupe est dans CLUSTER */
    outstat = data_stat; /* résultats statistiques*/
    var q2-q3;
run ;

proc tabulate data=clus2;
    var q2 q3;
    class cluster ;
    table q2 q3 , cluster* (N mean std);
run;

proc gplot data= clus2;plot q3*q2=cluster;run ;

proc gplot data= clus2;plot q1*q4=cluster;run ;
proc gplot data= clus2;plot (q1 q4)*cluster;run ;

proc glm data= clus2 ;
    class cluster ;
    model q1 q4 = cluster ;
run ;

```

Proc MODECLUS : Typologie non paramétrique

Caractéristiques :

- Bien adapté pour détecter des clusters très différents en terme de taille, de forme et de dispersion.
- Peut aussi s'adapter aux autres cas mais alors moins bien adapté que les méthodes clusters pour la recherche de clusters de tailles et de dispersions identiques (Nécessite un nombre assez important de données pour les retrouver).
- Estime les fonctions de densité des groupes de manière non paramétrique (recherche de maxima sur la densité). Recherche des voisins dans un espace sphérique.

Principe :

- Estimation des densités
- Choix d'une méthode (1 à 6), SAS recommande = 1
- Tests de plusieurs radius (rayons) : plus le radius est élevé moins il y a de groupes

```

*****;
* Proc MODECLUS                                     *;
* typologie non paramétrique                       *;
* choix d'une méthode de 1 à 6 (1 recommandées par SAS *;
* choix du radius ave possibilité d'en tester plusieurs *;
*****;
*
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug\_modeclus\_sect012.htm;
data example;
    input x y @@;
        ID = _N_ ;
cards;
18 18 20 22 21 20 12 23 17 12 23 25 25 20 16 27
20 13 28 22 80 20 75 19 77 23 81 26 55 21 64 24
72 26 70 35 75 30 78 42 18 52 27 57 41 61 48 64
59 72 69 72 80 80 31 53 51 69 72 81

```

```

;
proc sgplot;
    scatter y=y x=x;
run;

Title2 "Typologie non paramétrique ";
*-----*;
proc modeclus data=example
    method=1
    r=10 15 35
    out=out_mode_clus;
    var x y ;
run;

proc sgplot data=out_mode_clus noautolegend;
    by _r_;
    scatter y=y x=x
        / group=cluster markerchar=cluster;
run;

```

Proc VARCLUS : Typologie des variables

Caractéristiques :

- Partition automatique de nouvelles variables (composante) obtenues par le regroupement, sans recouvrement, des variables quantitatives (ou binaires) corrélées
- Options
 - o ACP avec classification hiérarchique « single linkage »
 - o ACP oblique quartimax et classification hiérarchique (Proc Varclus)
- Utilisé pour construire des échelles de mesure
 - o Sans option : analyse des corrélations et facteur
 - o Centroid cov : analyse de la variance et moyenne arithmétique

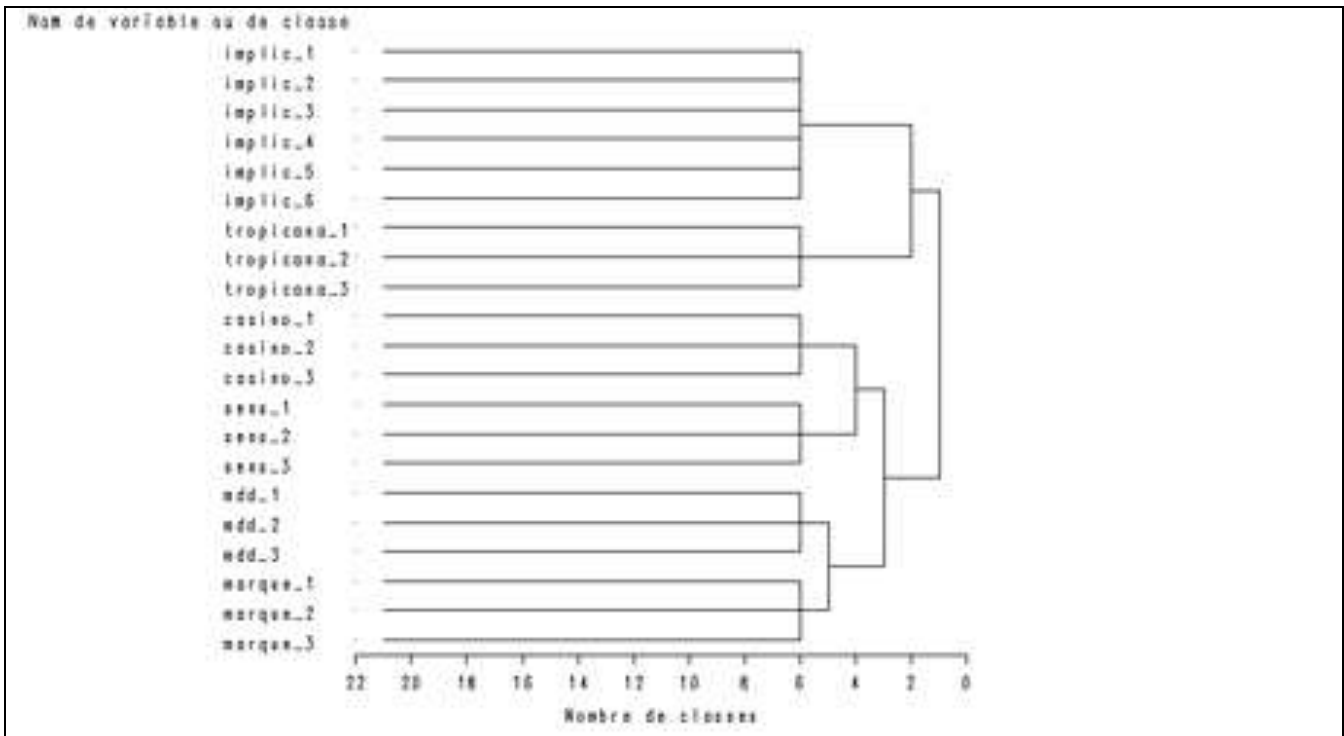
```

proc corr data=analysis alpha;
    var x1-x4;
run;
proc varclus data=analysis;
    var x1-x4;
run;
proc varclus data=analysis centroid cov;
    var x1-x4;
run;

```

Exemple sur Orange

Nombre de classes	Variation totale expliquée par classe	Proportion de variation expliquée par classe	Proportion min. expliquée par une classe	Seconde valeur propre maximum dans une classe	R carré minimum pour une variable	Ratio 1-R ² maximum pour une variable
1	4.421405	0.2105	0.2105	3.963702	0.0023	
2	7.947717	0.3785	0.3247	2.245303	0.1577	0.8824
3	10.140951	0.4829	0.3247	2.062240	0.1577	0.8908
4	11.978383	0.5704	0.4746	1.438379	0.2367	0.5224
5	13.387183	0.6375	0.4811	1.129198	0.3294	0.7092
6	14.512005	0.6910	0.6285	0.834567	0.4724	0.5418



Total variation expliquée = 14.512 Proportion = 0.6910

6 classes		R carré avec			Libellé de variable
Classe	Variable	Propre classe	Le plus proche	Ratio 1-R ²	
Cluster 1	implic_1	0.6527	0.0649	0.3715	Le jus d'orange frais est un produit qui compte vraiment beaucoup pour moi
	implic_2	0.6697	0.0562	0.3500	En matière de jus d'orange, il y a beaucoup à perdre si l'on choisit la mauvaise marque
	implic_3	0.4724	0.0034	0.5294	Il est possible de faire un mauvais choix lors de l'achat d'un jus d'orange frais
	implic_4	0.7293	0.0272	0.2783	Le choix de son jus d'orange frais a une forte valeur symbolique
	implic_5	0.7695	0.0503	0.2427	Je prends du plaisir à boire du jus d'orange frais et j'apprécie particulièrement en boire
	implic_6	0.4777	0.0060	0.5255	On peut dire que le choix de mon jus d'orange frais m'intéresse
Cluster 2	casino_1	0.8141	0.1599	0.2213	Si j'achète ce jus d'orange casino, je vais probablement l'aimer
	casino_2	0.7649	0.1049	0.2626	Je pense que la majorité des personnes qui achètent ce jus d'orange Casino est satisfaite
	casino_3	0.7910	0.1050	0.2335	D'une manière globale, je décrirais ce produit comme attrayant

Les arbres de segmentation

Caractéristiques :

- Découpage d'une population en fonction d'un critère à maximiser
- Voir Entreprise miner.