

Séance 9 : Espace produit

Sommaire

Proc FACTOR : Synthèse des covariations (ACP).....	2
Exemple 1 : synthèse (Dépenses médias par Secteur d'activité).....	3
Proc FACTOR : Construction d'un espace (PAF /AFCS).....	7
Exemple 2 : Critères d'évaluation de la bière.....	9
Proc MDS : Analyse des similarités	14
Exemple 3 : Distances kilométriques entre villes françaises	15

Les études de positionnement concernent la place que peut avoir un produit dans l'esprit du consommateur et comment cette place peut influencer ses préférences ou son comportement.

Il y a donc 2 informations très distinctes

- Les perceptions
- Les préférences

L'espace produit peut être constitué de différentes manières :

- Par une approche directe de l'espace par les clients (placement des produits sur un espace physique par des méthodologies de type Napping®)
- A partir de données sur les proximités (similarités / dissimilarités) entre les produits, c'est l'approche de décomposition.
- Indirectement à partir des évaluations des produits sur des variables, puis de la configuration d'un espace des variables. C'est l'approche de composition.

Chaque méthode a ses avantages et inconvénients.

- L'approche de **décomposition** est respectueuse de la perception « holistique » (globale) des produits par les clients. Elle conduit souvent à un traitement individuel (un espace par client) et donc à des problèmes d'agrégation de ces résultats. Il faut ensuite, après la constitution de l'espace, pouvoir l'interpréter et comprendre les leviers permettant à l'entreprise de modifier son image.
- L'approche de **composition** est plus facile à mettre en œuvre (questions par produit). De plus cette approche se prête bien à des analyses agrégées suivies d'une segmentation. Cependant les résultats dépendent des variables qui ont été utilisées. Il faut donc bien s'assurer (par des études qualitatives) que toutes les dimensions importantes ont été prises en compte.

Proc FACTOR : Synthèse des covariations (ACP)

Principe : recherche d'un faible nombre de combinaison linéaire des variables (facteurs) reconstituant au mieux l'information fournie par une matrice de variance/covariance ou de corrélations entre ces variables. $F1 = a1. X1 + a2. X2$ et $X1 = b1.F1 + b2.F2$

Nature des variables et de la distance :

- Quantitatives
- Distance : euclidienne = $(xi-xj)^2$: valorise les grands écarts absolus

Utilisations :

- Cas 1 : Approche exploratoire : synthèse de l'information fournie par l'évaluation de plusieurs objets (marques) sur plusieurs variables
- Cas 2 : Constitution de combinaisons linéaires orthogonales des variables pour éliminer la colinéarité avant une régression
- Cas 3 : Analyse confirmatoire (vue plus tard dans le cours) d'un ensemble de groupes de variables, chaque groupe étant associé à une variable latente que l'on souhaite mesurer.

```

*-----* ;
Title4 "ACP avec rotation" ;
*-----* ;
ODS graphics on ;
PROC FACTOR DATA=in
  Plot = (SCREE                               /* demande de l'éboulis */
  initloadings loadings)                    /* graphiques des variables avant eta près rotation */
  CORR                                       /* sur corrélations ou COV covariance */
  SIMPLE                                    /* statistiques descriptives des variables */
  MSA                                       /* test d'adéquation KMO, Bartlett*/
  METHOD=PRIN          PRIORS=ONE           /* ACP */
  MINEIGEN=1          /* critère de Kaiser */
  ROTATE= Varimax Reorder                  /* rotation orthogonale, tri par importance décroissante */
                                          /* ou promax = rotation oblique */
  FLAG=.32                                  /* identification des loadings > à cette valeur */
  NFACT= 4                                  /* choix du nombre de facteurs, obligatoire si OUTPUT = */
  OUT =          out_individu ;             /* sortie d'un data avec les facteurs factor1, factor2, */
  OUTSTAT =      out_variable ;            /* sortie d'un data avec les variables */
  VAR services distribution B_to_B B_to_C; /* variables à analyser */
RUN;

```

Décisions à prendre :

- (0) Quelle approche choisir ?

- o Analyse en composantes principales (PCA) : elle cherche à reconstituer *TOUTE* la variance initiale et est adaptée à une approche exploratoire.
- o Priors = 1 : toute la variance de la variable est « à reconstituer »

- (1) Quelles variables retenir dans la constitution de la matrice à analyser ?

- o D'abord toutes celles qui sont pertinentes par rapport à l'objectif de l'analyse
- o

- (2) Faut-il standardiser les variables ? Choisir les covariances (COV) ou les corrélations (CORR).

- o Non si les différences de variances ont une signification (analyse des Variances/covariances) : COV
- o Oui si les différences de variance sont dues principalement à des questions d'unités de mesure (analyse des corrélations) CORR

- **(3) Une analyse factorielle est-elle adaptée ?** Oui si les variables ont quelque chose en commun (MSA measures of Sample Adequacy).
 - o Test 1 **KMO** (Kaiser-Meyer-Olkin)
 - o Test de **sphéricité de Bartlett** (une matrice de corrélations sphérique ne contient que des 1 sur la diagonale : les corrélations entre deux variables sont toutes nulles)

- **(4) Combien d'axes faut-il retenir ?** il faut chercher un équilibre entre le pourcentage d'inertie (information contenue dans la matrice initiale) reconstituée et le nombre d'axes retenu (plus il est important, moins la représentation est synthétique et plus elle est difficile à interpréter). Les critères sont :
 - o **Critère de qualité de la représentation** : la configuration doit être la plus simple possible tout en représentant une part importante de l'inertie de la matrice initiale (75% et au minimum 50%).
 - o **Critère de l'éboulis (scree test)** (SCREE): une représentation graphique des valeurs propres montre souvent une cassure (un « coude ») avec un fort changement de pente. On retient alors le nombre de facteurs juste avant le début de l'éboulis.
 - o **Critère de Kaiser** : Une valeur propre au moins égale à 1 pour un facteur. En effet chaque variable contribue à 1 unité d'inertie, un facteur avec une valeur propre <1 représente donc moins qu'une variable. C'est la moins bonne technique de sélection qui conduit à retenir un trop grand nombre d'axes.

- **(5) Eliminer les variables mal représentées** dans la configuration envisagée par le nombre d'axes retenus sur la base de la communauté (*communality* = somme des carrés des loadings d'une variable pour tous les facteurs retenus)
 - o processus de « raffinement » du modèle par l'élimination des variables mal représentées : moins de 50% de leur variance est représentée dans l'espace retenu

- **(6) Interpréter, donner un nom, une signification à chaque facteur.** Attention à cette phase cruciale puisque c'est ensuite sur base de ce « nom » que vous allez donner du sens à vos résultats. On interprète les corrélations (loadings) des variables avec les facteurs dans la matrice de « Structure ». A partir des variables
 - o Les plus corrélées positivement et négativement : contribution au facteur (*loading*) > 0.4 pour une taille d'échantillon de n=140, moins pour des effectifs plus importants.
 - o La matrice de « Pattern » donne la contribution spécifique d'une variable (l'effet des autres variables étant éliminé)

- **(7) Interprétation graphique.** La méthode fournit 2 graphiques
 - o *Le graphique des variables* (dont les axes sont borné à +/-1 (les corrélations). Les variables au centre du graphe sont mal représentées. Celles proches du cercle des corrélations le sont bien. La proximité avec un facteur indique que la variable contribue fortement à ce facteur.
 - o *Le graphique des individus.* Son système d'axes est différent de celui des variables, il ne faut donc pas les superposer. Des individus proches dans l'espace des facteurs ont des valeurs proches sur les variables qui forment les facteurs.

Exemple 1 : synthèse (Dépenses médias par Secteur d'activité)

Contexte : compréhension d'un tableau des dépenses de communication (promotion, TV, salons,...) selon les secteurs d'activité (services, Business to Business (B_to_B),...

Programme : deux macros sont proposées pour faciliter les sorties graphiques

En ligne : <http://marketing-science-center.com/acp-medias/>

```
*****
* Programme basique ACP avec graphique
* Pierre DESMET Décembre 2012
* Traitement des dépenses de communication par secteur d'activité*
*****

Options nocenter formdlim="-" ;
Data in ;
    length ID $ 13 ;
    input ID $ services distribution B_to_B B_to_C ;
    total = services+distribution+B_to_B+B_to_C ;
cards ;
TV 1109 217 105 2681
Cinéma 59 5 3 50
Radio 398 264 38 221
Internet 489 195 80 243
Presse 2012 696 438 1101
Annuaire 785 189 290 22
Affichage 615 272 17 551
Mark_Direct 4149 1208 1903 2299
Promotion 1285 1317 317 2187
salons_Foires 844 54 456 185
Mecenasat_Spons 590 60 242 294
Rel_Publiques 930 79 516 308
;
run ;

* MACRO Graphique Variables sur data= out_variable *****;
%Macro graphiqueV (data= &data, x=&X, y=&Y, XL=&XL, YL=&YL, ID=&ID, titre = &titre) ;
axis1 length= 24 cm; axis2 length= 24 cm; symbol1 V=none interpol=none;
data out_variable ; set &data; if _TYPE_='SCORE'; run ;
proc transpose data=&data out=t_variable ;run ; proc print data=t_variable ; run ;
data work; set t_variable ; keep X Y Xsys Ysys text size color; X=&X ; Y=&Y; text=&ID ; color='black' ;
size=1; xsys=2; ysys=2; Label Y=&YL X=&XL; run; proc gplot data=work; title4 &titre ;
plot Y*X=1 / haxis=axis1 vaxis=axis2 annotate=work Frame Href=0 Vref=0; run;
%Mend graphiqueV ;

* Macro Graphique Individus sur data= out_individu *****;
%Macro graphiquel (data= &data, x=&X, y=&Y, XL=&XL, YL=&YL, ID=&ID, titre = &titre,
red=&red, blue=&blue, green=&green ) ;
axis1 length= 24 cm; axis2 length= 24 cm; symbol1 V=none interpol=none;
data work ; set &data ; keep X Y Xsys Ysys text size color ; X=&x ; Y=&y; text=&ID ; color='black' ;
if text=&red then color='RED' ; if text=&blue then color='BLUE'; if text=&green then color='GREEN' ;
size=1; xsys=2; ysys=2; Label Y=&YL X=&XL; run; proc gplot data=work;
plot Y*X=1 / haxis=axis1 vaxis=axis2 annotate=work Frame Href=0 Vref=0; run;
%Mend graphiquel ;

*-----*
Title4 "Dépenses de communication selon les secteurs d'activité" ;
*-----*
proc tabulate data=in ;
class id ;
var services distribution B_to_B B_to_C total ;
table (id all), (services distribution B_to_B B_to_C total)*sum*F=6.0 ;
run ;

*-----*
Title4 "Corrélations" ;
*-----*
proc corr data = in ;
var services distribution B_to_B B_to_C ;
run ;

ODS graphics on ;
```

```

*-----* ;
Title4 "ACP sans rotation" ;
*-----* ;
ODS graphics on ;
PROC FACTOR DATA=in
    Plot = (SCREE                                /* demande de l'éboulis */
    loadings)                                    /* graphiques des variables */
    CORR                                          /* sur corrélations ou COV covariance */
    SIMPLE                                       /* statistiques descriptives des variables */
    MSA                                          /* test d'adéquation KMO, Bartlett*/
    METHOD=PRIN          PRIORS=ONE             /* ACP */
/* MINEIGEN=1 /* critère de Kaiser valeur propres >1 */
FLAG=.32 /* identification des loadings > à cette valeur */
NFACT= 2 /* choix du nombre de facteurs, obligatoire si OUT = */
OUT = out_individu /* sortie d'un data avec les facteurs factor1, factor2, */
OUTSTAT = out_variable ; /* sortie d'un data avec les variables */
VAR services distribution B_to_B B_to_C; /* variables à analyser */
RUN;

Title4 "Data out_individu" ;
*-----* ;
proc print data=out_individu;run ;

Title4 "Data out_variable" ;
*-----* ;
proc print data=out_variable;run ;

%graphiqueV(data= out_variable, x=factor1, y=factor2, XL="Services", YL="B_to_C",
ID=_name_ , titre= "premier plan factoriel Variables (colonnes)" ) ;
%graphiqueL( data= out_individu, x=factor1, y=factor2, XL="Services", YL="B_to_C",
ID=ID, titre="premier plan factoriel Individus (lignes)",
red= "Mark_Direct", blue=., green =.) ;

```

Résultats commentés

<p>Dépenses de communication selon les secteurs d'activité</p> <table border="1"> <thead> <tr> <th rowspan="2"></th> <th colspan="2">services</th> <th colspan="2">distribution</th> <th>B to B</th> <th>B to C</th> <th rowspan="2">total</th> </tr> <tr> <th>Sum</th> <th>Sum</th> <th>Sum</th> <th>Sum</th> <th>Sum</th> <th>Sum</th> </tr> </thead> <tbody> <tr><td>ID</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>Affichage</td><td>615</td><td>272</td><td>17</td><td>551</td><td>1455</td><td></td><td></td></tr> <tr><td>Annuaire</td><td>785</td><td>189</td><td>290</td><td>22</td><td>1286</td><td></td><td></td></tr> <tr><td>Cinéma</td><td>59</td><td>5</td><td>3</td><td>50</td><td>117</td><td></td><td></td></tr> <tr><td>Internet</td><td>489</td><td>196</td><td>80</td><td>243</td><td>1007</td><td></td><td></td></tr> <tr><td>Mark Direct</td><td>4149</td><td>1208</td><td>1903</td><td>2299</td><td>9559</td><td></td><td></td></tr> <tr><td>Mecenas Spons</td><td>590</td><td>60</td><td>242</td><td>294</td><td>1186</td><td></td><td></td></tr> <tr><td>Presse</td><td>2012</td><td>696</td><td>438</td><td>1101</td><td>4247</td><td></td><td></td></tr> <tr><td>Promotion</td><td>1265</td><td>1317</td><td>317</td><td>2187</td><td>5106</td><td></td><td></td></tr> <tr><td>Radio</td><td>398</td><td>264</td><td>38</td><td>221</td><td>921</td><td></td><td></td></tr> <tr><td>Ref Publiques</td><td>930</td><td>79</td><td>516</td><td>308</td><td>1833</td><td></td><td></td></tr> <tr><td>TV</td><td>1109</td><td>217</td><td>105</td><td>2681</td><td>4112</td><td></td><td></td></tr> <tr><td>salons Foires</td><td>844</td><td>54</td><td>456</td><td>185</td><td>1539</td><td></td><td></td></tr> <tr><td>Tout</td><td>13265</td><td>4556</td><td>4405</td><td>10142</td><td>32368</td><td></td><td></td></tr> </tbody> </table>		services		distribution		B to B	B to C	total	Sum	Sum	Sum	Sum	Sum	Sum	ID								Affichage	615	272	17	551	1455			Annuaire	785	189	290	22	1286			Cinéma	59	5	3	50	117			Internet	489	196	80	243	1007			Mark Direct	4149	1208	1903	2299	9559			Mecenas Spons	590	60	242	294	1186			Presse	2012	696	438	1101	4247			Promotion	1265	1317	317	2187	5106			Radio	398	264	38	221	921			Ref Publiques	930	79	516	308	1833			TV	1109	217	105	2681	4112			salons Foires	844	54	456	185	1539			Tout	13265	4556	4405	10142	32368			<p>Les données à analyser sont présentées sous forme d'un tableau.</p> <p>On choisit de mettre les secteurs en colonnes (les colonnes doivent toujours être la plus petite des dimensions).</p>
		services		distribution		B to B	B to C		total																																																																																																																						
	Sum	Sum	Sum	Sum	Sum	Sum																																																																																																																									
ID																																																																																																																															
Affichage	615	272	17	551	1455																																																																																																																										
Annuaire	785	189	290	22	1286																																																																																																																										
Cinéma	59	5	3	50	117																																																																																																																										
Internet	489	196	80	243	1007																																																																																																																										
Mark Direct	4149	1208	1903	2299	9559																																																																																																																										
Mecenas Spons	590	60	242	294	1186																																																																																																																										
Presse	2012	696	438	1101	4247																																																																																																																										
Promotion	1265	1317	317	2187	5106																																																																																																																										
Radio	398	264	38	221	921																																																																																																																										
Ref Publiques	930	79	516	308	1833																																																																																																																										
TV	1109	217	105	2681	4112																																																																																																																										
salons Foires	844	54	456	185	1539																																																																																																																										
Tout	13265	4556	4405	10142	32368																																																																																																																										
<p>Procédure CORR</p> <p>4 Variables : services distribution B_to_B B_to_C</p> <table border="1"> <thead> <tr> <th colspan="7">Statistiques simples</th> </tr> <tr> <th>Variable</th> <th>N</th> <th>Moyenne</th> <th>Ecart type</th> <th>Somme</th> <th>Minimum</th> <th>Maximum</th> </tr> </thead> <tbody> <tr><td>services</td><td>12</td><td>1105</td><td>1079</td><td>13265</td><td>59</td><td>0.00000</td><td>4149</td></tr> <tr><td>distribution</td><td>12</td><td>379.66667</td><td>449.81862</td><td>4556</td><td>5</td><td>0.00000</td><td>1317</td></tr> <tr><td>B to B</td><td>12</td><td>367.08333</td><td>516.34757</td><td>4405</td><td>3</td><td>0.00000</td><td>1903</td></tr> <tr><td>B to C</td><td>12</td><td>845.16667</td><td>978.24758</td><td>10142</td><td>22</td><td>0.00000</td><td>2681</td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="5">Coefficients de corrélation de Pearson, N = 12</th> </tr> <tr> <th colspan="5">Proba > r sous H0: Rho=0</th> </tr> <tr> <th></th> <th>services</th> <th>distribution</th> <th>B to B</th> <th>B to C</th> </tr> </thead> <tbody> <tr><td>services</td><td>1.00000</td><td>0.74562</td><td>0.93539</td><td>0.64679</td></tr> <tr><td>distribution</td><td>0.74562</td><td>1.00000</td><td>0.60140</td><td>0.71625</td></tr> <tr><td>B to B</td><td>< .0001</td><td>0.0386</td><td>1.00000</td><td>0.44404</td></tr> <tr><td>B to C</td><td>0.0230</td><td>0.0088</td><td>0.1481</td><td>1.00000</td></tr> </tbody> </table>	Statistiques simples							Variable	N	Moyenne	Ecart type	Somme	Minimum	Maximum	services	12	1105	1079	13265	59	0.00000	4149	distribution	12	379.66667	449.81862	4556	5	0.00000	1317	B to B	12	367.08333	516.34757	4405	3	0.00000	1903	B to C	12	845.16667	978.24758	10142	22	0.00000	2681	Coefficients de corrélation de Pearson, N = 12					Proba > r sous H0: Rho=0						services	distribution	B to B	B to C	services	1.00000	0.74562	0.93539	0.64679	distribution	0.74562	1.00000	0.60140	0.71625	B to B	< .0001	0.0386	1.00000	0.44404	B to C	0.0230	0.0088	0.1481	1.00000	<p>La matrice des corrélations montre que les corrélations sont fortes entre toutes les variables.</p> <p>Le premier facteur sera donc probablement très important.</p>																																													
Statistiques simples																																																																																																																															
Variable	N	Moyenne	Ecart type	Somme	Minimum	Maximum																																																																																																																									
services	12	1105	1079	13265	59	0.00000	4149																																																																																																																								
distribution	12	379.66667	449.81862	4556	5	0.00000	1317																																																																																																																								
B to B	12	367.08333	516.34757	4405	3	0.00000	1903																																																																																																																								
B to C	12	845.16667	978.24758	10142	22	0.00000	2681																																																																																																																								
Coefficients de corrélation de Pearson, N = 12																																																																																																																															
Proba > r sous H0: Rho=0																																																																																																																															
	services	distribution	B to B	B to C																																																																																																																											
services	1.00000	0.74562	0.93539	0.64679																																																																																																																											
distribution	0.74562	1.00000	0.60140	0.71625																																																																																																																											
B to B	< .0001	0.0386	1.00000	0.44404																																																																																																																											
B to C	0.0230	0.0088	0.1481	1.00000																																																																																																																											

Méthode de facteur initiale : Composantes principales

Corrélations partielles contrôlant toutes les autres variables				
	services	distribution	B to B	B to C
services	1.00000	0.35664	0.93712	0.54605
distribution	0.35664	1.00000	-0.18646	0.29638
B to B	0.93712	-0.18646	1.00000	-0.50474
B to C	0.54605	0.29638	-0.50474	1.00000

Mesure d'adéquation de l'échantillonnage de Kaiser : MSA globale = 0.63474216			
services	distribution	B to B	B to C
0.58653914	0.85134711	0.55113906	0.63783838

Valeurs estimées des facteurs communs a priori : ONE

Valeurs propres de la matrice de corrélation: Total = 4 Moyenne = 1				
Valeur propre	Différence	Proportion	Cumulé	
1	3.05738553	2.40415667	0.7643	0.7643
2	0.65322886	0.39444443	0.1633	0.9277
3	0.25878443	0.22818326	0.0647	0.9923
4	0.03060118		0.0077	1.0000

2 facteurs seront retenus par le critère NFACTOR.

La mesure globale de l'adéquation de la méthode factorielle est moyenne (MSA = 0.63) ce qui signifie que chaque variable garde une information originale, spécifique. Les mesures par variable montrent qu'aucune variable n'a une reconstitution de sa variance inférieure à 50%. La somme des valeurs propres fait 4 (parce qu'il y a 4 variables)

La première reconstitue 76% de l'inertie initiale. On pourrait se contenter du premier facteur car aucun autre VP n'est > à 1 et l'éboullis commence à 2.

Avec deux dimensions on perd très peu d'informations (7.3%). On la garde donc même si sa valeur propre est inférieure à 1.

Représentation du facteur		
	Factor1	Factor2
services	0.95860	-0.24112
distribution	0.87543	0.26220
B to B	0.86273	-0.48973
B to C	0.79233	0.53526

Variance expliquée par chaque facteur	
Factor1	Factor2
3.0573855	0.6532289

Valeurs estimées finales des facteurs communs : Total = 3.710614			
services	distribution	B to B	B to C
0.97704561	0.83513144	0.98414402	0.91429333

Le facteur F1 = 0.95.Ser + 0.87 Distr+ ...

Pour un espace de deux facteurs la reconstitution de la variance des variables est très bonne (de 85% à 97%)

Data out_individu

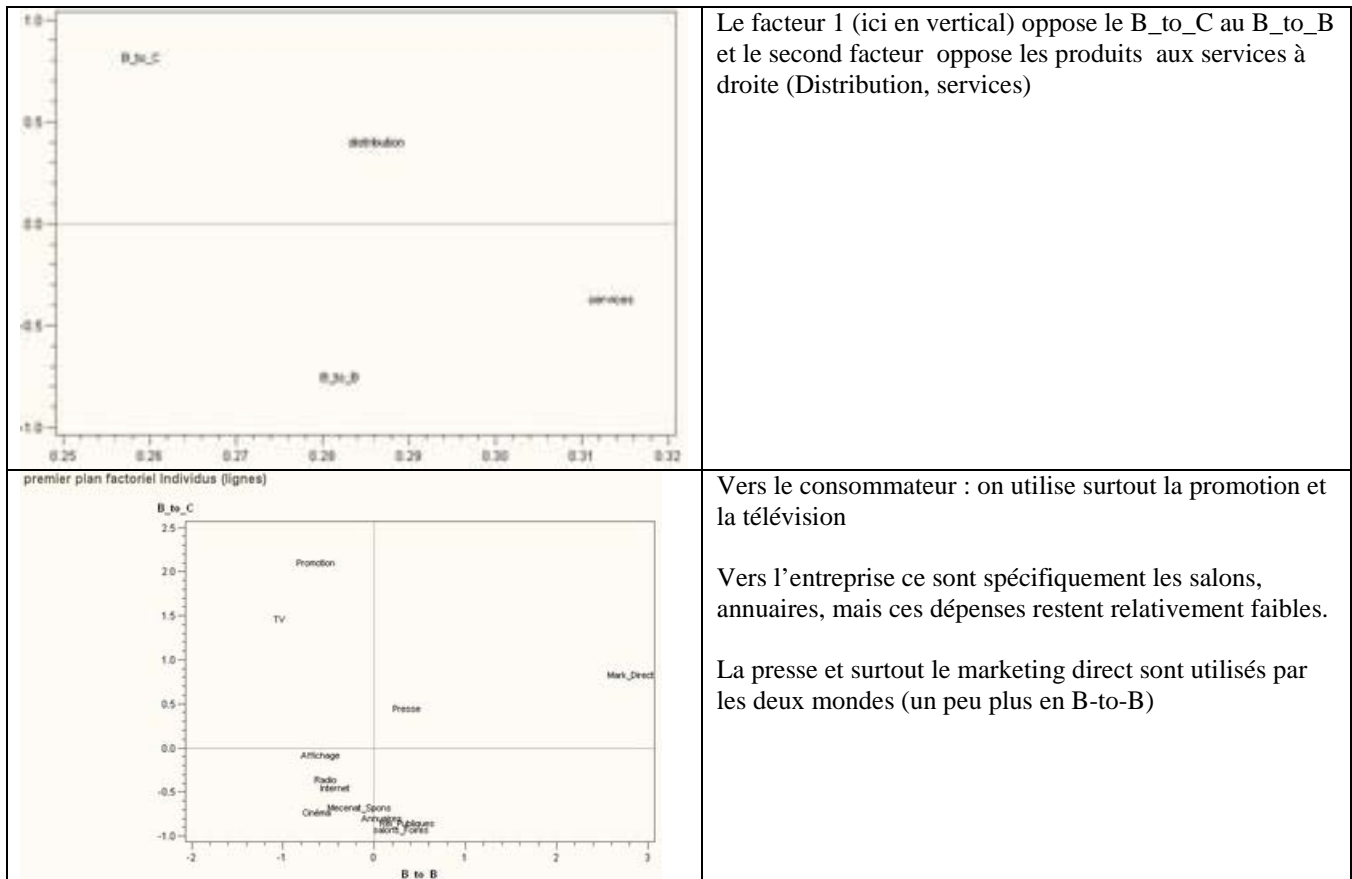
Obs_ID	services	distribution	B to B	B to C	total	Factor1	Factor2
1 TV	1109	217	105	2681	4112	-1.02858	1.46264
2 Cinéma	59	5	3	50	117	-0.62078	-0.73099
3 Radio	398	264	38	221	921	-0.52159	-0.35599
4 Internet	489	195	80	243	1007	-0.42131	-0.44736
5 Presse	2012	696	438	1101	4247	0.36217	0.44974
6 Annuaire	785	189	290	22	1286	0.08596	-0.79063
7 Affichage	615	272	17	551	1455	-0.57903	-0.08208
8 Mark_Direct	4149	1208	1903	2299	9559	2.82682	0.82930
9 Promotion	1285	1317	317	2187	5106	-0.62471	2.10988
10 salons_Foires	844	54	456	185	1539	0.30042	-0.92607
11 Mecenat_Spons	590	60	242	294	1186	-0.15163	-0.67121
12 Rei_Publiques	930	79	516	308	1833	0.37236	-0.84734

Le data out_individu contient maintenant les variables Factor1 et Factor2

Data out_variable

Obs	TYPE	NAME	services	distribution	B to B	B to C
1	MEAN		1105.42	379.667	367.083	845.167
2	STD		1078.76	449.819	516.348	978.248
3	N		12.00	12.000	12.000	12.000
4	CORR	services	1.00	0.746	0.935	0.647
5	CORR	distribution	0.75	1.000	0.601	0.716
6	CORR	B to B	0.94	0.601	1.000	0.444
7	CORR	B to C	0.65	0.716	0.444	1.000
8	COMMUNAL		0.98	0.835	0.984	0.914
9	PRIORS		1.00	1.000	1.000	1.000
10	EIGENVAL		3.06	0.653	0.259	0.031
11	UNROTATE	Factor1	0.96	0.875	0.863	0.792
12	UNROTATE	Factor2	-0.24	0.262	-0.490	0.535
13	TRANSFOR	Factor1	0.73	-0.680		
14	TRANSFOR	Factor2	0.68	0.733		
15	PATTERN	Factor1	0.87	0.463	0.966	0.217
16	PATTERN	Factor2	0.48	0.788	0.228	0.931
17	SCORE	Factor1	0.48	-0.063	0.717	-0.367
18	SCORE	Factor2	-0.06	0.489	-0.358	0.777

Le data out_variable contient les informations sur les variables.



Proc FACTOR : Construction d'un espace (PAF /AFCS)

Principe :

- En ACP simple, le tableau « existe » et toute la variance de ce tableau est importante et la méthode cherche à reconstituer la somme des variances (ou des corrélations) : la trace de la matrice Var-Covar.
- En AFCS analyse en facteurs communs et spécifiques le tableau est construit par le chercheur qui a sélectionné les variables.
- En AFCS la variance est décomposée en une partie « commune » aux variables, qui justifie l'analyse factorielle, et une partie « spécifique » à chaque variable, comparable à un résidu sans intérêt spécifique. La variance à reconstituer est la somme des variances « communes » .

Utilisations :

- Recherche de variables « latentes » formées par des groupes de variables sélectionnées.

Décisions à prendre :

- **(0) Quelle approche choisir ? Facteurs communs**
 - o **Method = Prin**
 - o **Priors = smc.** Une première estimation de la variance commune pour chaque variable est la corrélation multiple au carré (smc).
- **(1) Quelles variables retenir dans la constitution de la matrice à analyser ?**
 - o Celles qui sont corrélées à un axe et un seul

- Celles qui ont la plus faible variance spécifique (la plus forte variance commune)
 - Celles qui ont des corrélations partielles les plus faibles (corrélations spécifiques entre deux variables après avoir éliminé les variations communes)
- **(2) Faut-il standardiser les variables ?**
- Choisir les corrélations : CORR
- **(3) Une analyse factorielle est-elle adaptée ?** Oui si les variables ont quelque chose en commun (MSA measures of Sample Adequacy).
- Test 1 **KMO** (Kaiser-Meyer-Olkin) avec la statistique MSA. 0.8 = bien ; 0.5 inacceptable ; 0.6 médiocre.
 - Test de **sphéricité de Bartlett** (une matrice de corrélations sphérique ne contient que des 1 sur la diagonale : les corrélations entre deux variables sont toutes nulles)
- **(4) Combien d'axes faut-il retenir ?** il faut retenir autant d'axes qu'il y a de variables latentes attendues.
- **Critère de l'éboulis (scree test)** (SCREE): une représentation graphique des valeurs propres montre souvent une cassure (un « coude ») avec un fort changement de pente. On retient alors le nombre de facteurs juste avant le début de l'éboulis.
 - **Critère de Kaiser** : Une valeur propre au moins égale à 1 pour un facteur. En effet chaque variable contribue à 1 unité d'inertie, un facteur avec une valeur propre <1 représente donc moins qu'une variable. C'est la moins bonne technique de sélection qui conduit à retenir un trop grand nombre d'axes.
- **(5) Eliminer les variables mal représentées** dans la configuration envisagée par le nombre d'axes retenus sur la base de la communauté (*communality* = somme des carrés des loadings d'une variable pour tous les facteurs retenus)
- processus de « raffinement » du modèle par l'élimination des variables mal représentées : moins de 50% de leur variance est représentée dans l'espace retenu
- **(6) Interpréter, donner un nom, une signification à chaque facteur.** Attention à cette phase cruciale puisque c'est ensuite sur base de ce « nom » que vous allez donner du sens à vos résultats. On interprète les corrélations (loadings) des variables avec les facteurs dans la matrice de « Structure ». A partir des variables
- Les plus corrélées positivement et négativement : contribution au facteur (*loading*) > 0.4 pour une taille d'échantillon de n=140, moins pour des effectifs plus importants.
 - La matrice de « Pattern » donne la contribution spécifique d'une variable (l'effet des autres variables étant éliminé)
- **(7) Obtenir une structure « simple »** par le choix d'une rotation éventuelle.
- Une structure est « simple » si 5 critères sont validés (Thurstone, 1947) :
 - 1. Chaque variable doit avoir un loading = 0 sur au moins un facteur
 - 2. Chaque facteur doit avoir au moins autant de loadings nuls qu'il y a de facteurs
 - 3. Chaque paire de facteurs doit avoir des variables différenciées (loading important sur un facteur et loading nul sur l'autre)
 - 4. Chaque paire de facteurs doit avoir une large proportion de loading simultanément nuls (s'il y a 4 facteurs et plus)
 - 5. Chaque paire de facteurs ne doit avoir qu'un nombre très réduit de variables complexes (loading sur les deux facteurs).

- **(8) Choisir une rotation éventuelle.** Le principe de base est que les facteurs sont orthogonaux (non corrélés) et sont extraits par ordre d'importance décroissante. Une rotation permet de faire « tourner » le système d'axes pour rendre l'interprétation des facteurs plus facile.
 - o *Rotations obliques* (rotation NON orthogonale) : à utiliser si les variables représentent des facteurs non indépendants. Si deux facteurs sont corrélés à plus de 0.32, alors les facteurs partagent au moins 10% de leurs variances et cela justifie de garder la rotation oblique (Tabachnick et Fidell, 2007) sauf s'il y a de bonnes raisons de maintenir la relation orthogonale (avant régression par exemple)
 - Promax
 - o *Rotations orthogonales* : La qualité globale de la représentation n'est pas modifiée mais la répartition entre les facteurs (contribution) est modifiée.
 - Varimax : maximise le carré des loadings des variables sur un facteur
 - Quartimax : minimise le nombre de facteurs pour reconstituer la variance d'une variable. (peu utile)
 - Equamax (combinaison de Varimax et Quartimax)

- **(9) Interprétation graphique.** La méthode fournit 2 graphiques
 - o *Le graphique des variables* (dont les axes sont bornés à +/-1 (les corrélations)). Les variables au centre du graphe sont mal représentées. Celles proches du cercle des corrélations le sont bien. La proximité avec un facteur indique que la variable contribue fortement à ce facteur.
 - o *Le graphique des individus.* Son système d'axes est différent de celui des variables, il ne faut donc pas les superposer. Des individus proches dans l'espace des facteurs ont des valeurs proches sur les variables qui forment les facteurs.

Exemple 2 : Critères de choix de la bière

Contexte : Deux groupes de 10 individus évaluent l'importance (0 à 100) donnée à 7 critères pour le choix d'une bière ainsi que l'intensité de leur consommation (0 à 8).



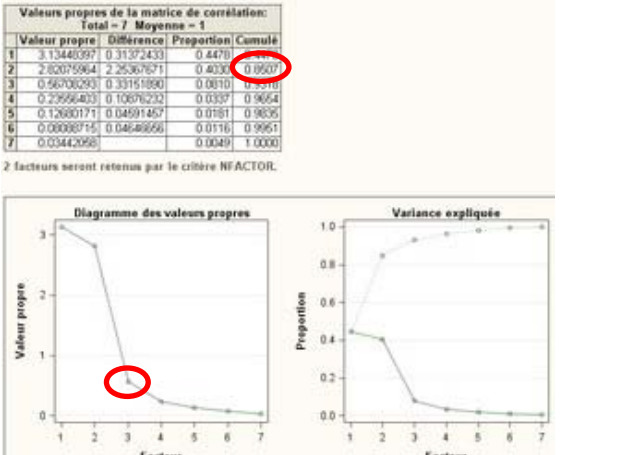
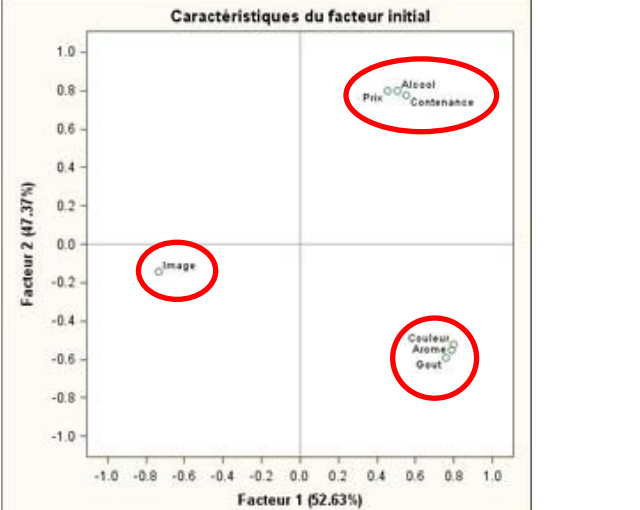
Faire un tableau récapitulatif. Puis : une ACP, un AFCS avec Varimax, une AFCS avec Promax.

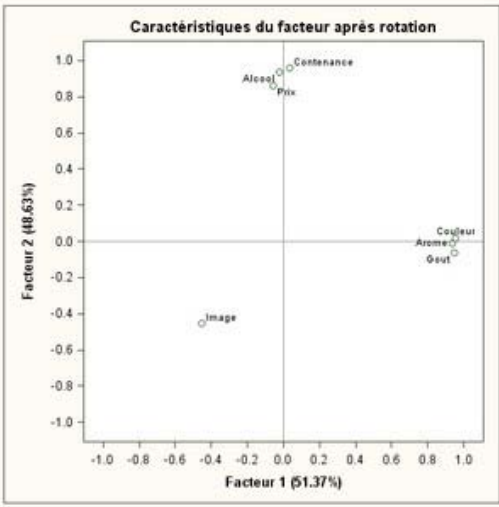
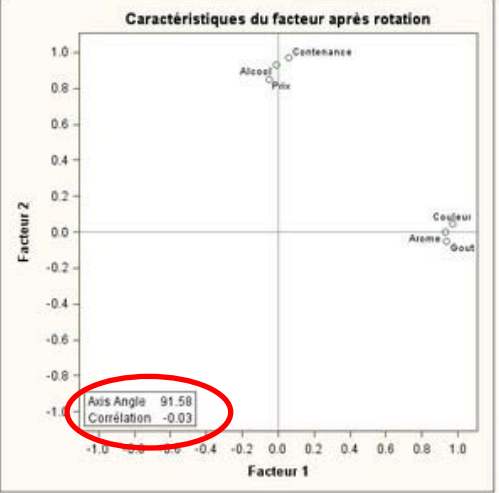
Éliminer les variables mal représentées. Donner un nom à chaque axe puis étudier les différences entre les groupes sur les axes et la capacité des facteurs à expliquer la consommation (SES).

Interprétation

	Prix	Contenance	Alcool	Image	Couleur	Arome	Gout	Ses
Num	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
1	20	10	25	35	30	35	45	4
2	50	100	100	30	90	75	100	3
3	50	15	20	50	10	5	50	2
4	70	60	70	75	10	15	25	0
5	75	95	100	50	55	40	65	1
6	75	50	95	40	0	0	40	0
7	80	90	100	50	20	20	40	0
8	80	70	50	50	40	20	50	1
9	85	80	70	40	60	50	65	2
10	90	80	70	20	50	70	60	2
11	95	95	100	0	80	70	95	2
12	0	0	20	30	80	90	100	8
13	5	15	15	75	20	10	25	2
14	5	10	15	65	50	65	85	7
15	10	25	10	100	50	40	60	5
16	10	15	20	95	40	30	50	4
17	15	10	25	30	95	80	100	8
18	20	5	10	40	60	50	95	7
19	25	35	30	40	45	30	65	3
20	65	30	35	80	80	60	90	6
21	100	70	50	30	75	60	80	3
Tout	49	46	49	48	50	44	66	3
groupe								
1	70	68	73	40	40	36	58	2
2	26	22	23	58	60	52	75	5

Matrice des données brutes (tableau)

 <p>Matrice des corrélations initiale</p>																												
 <p>Matrice des corrélations semi-partielles. Les valeurs devraient être faibles. On observe que certaines corrélations restent fortes (par « paquets » de variables).</p>																												
 <p>Extraction des valeurs propres par analyse en composantes principales. Le premier plan représente 85% de l'inertie.</p> <p>L'éboulis indique que deux facteurs sont suffisants</p>																												
 <p>Résultat avec ACP sans rotation. Le premier axe reconstitue la plus grande partie de l'information initiale mais son interprétation n'est pas claire avec trois groupes de variables.</p>																												
<table border="1" data-bbox="151 1512 630 1825"> <thead> <tr> <th colspan="3">Coefficients du score normalisés</th> </tr> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr> <td>Prix</td> <td>0.14453</td> <td>0.28322</td> </tr> <tr> <td>Contenance</td> <td>0.17637</td> <td>0.27453</td> </tr> <tr> <td>Alcool</td> <td>0.16134</td> <td>0.28316</td> </tr> <tr> <td>Image</td> <td>-0.23376</td> <td>-0.05034</td> </tr> <tr> <td>Couleur</td> <td>0.25459</td> <td>-0.18557</td> </tr> <tr> <td>Arome</td> <td>0.25070</td> <td>-0.19480</td> </tr> <tr> <td>Gout</td> <td>0.24227</td> <td>-0.20937</td> </tr> </tbody> </table> <p>Les coefficients des variables pour la construction des axes factoriels</p>	Coefficients du score normalisés				Factor1	Factor2	Prix	0.14453	0.28322	Contenance	0.17637	0.27453	Alcool	0.16134	0.28316	Image	-0.23376	-0.05034	Couleur	0.25459	-0.18557	Arome	0.25070	-0.19480	Gout	0.24227	-0.20937	
Coefficients du score normalisés																												
	Factor1	Factor2																										
Prix	0.14453	0.28322																										
Contenance	0.17637	0.27453																										
Alcool	0.16134	0.28316																										
Image	-0.23376	-0.05034																										
Couleur	0.25459	-0.18557																										
Arome	0.25070	-0.19480																										
Gout	0.24227	-0.20937																										

<p>Méthode de rotation : Varimax</p> 	<p>L'étude des résultats et du graphique montre que la réputation de la marque n'a pas vraiment sa place dans l'analyse. Il faut</p> <ul style="list-style-type: none"> - Soit l'éliminer - Soit la renforcer avec d'autres variables qui lui sont corrélées
	<p>Résultat avec Promax (rotation oblique). La faible corrélation des axes indique qu'une transformation oblique n'est pas indispensable.</p>

```
*****
* critères de choix d'une bière ;
* 7 critères de 0 à 100;
* une variable de consommation ses (0 à 8);
* une variable de groupe : Group (jeune, âgé);
* source : core.ecu.edu/psyc/wuenschk/sas/FactBeer.pdf ;
*****
```

options nocenter formdlm="-" ;

DATA in;

INPUT Prix Contenance Alcool Image Couleur Arôme Gout Ses Group ;

Num = _N_ ;

ID = PUT (Num, \$2.) ;

groupe=put(group, \$1.) ;

/* variable alpha pour ID */

/* variable alpha pour groupe */

CARDS;

```
20 10 25 35 30 35 45 4 1
50 100 100 30 90 75 100 3 1
50 15 20 50 10 5 50 2 1
70 60 70 75 10 15 25 0 1
75 95 100 50 55 40 65 1 1
75 50 95 40 0 0 40 0 1
80 90 100 50 20 20 40 0 1
80 70 50 50 40 20 50 1 1
```

```

85 80 70 40 60 50 65 2 1
90 80 70 20 50 70 60 2 1
95 95 100 0 80 70 95 2 1
0 0 20 30 80 90 100 8 2
5 15 15 75 20 10 25 2 2
5 10 15 65 50 65 85 7 2
10 25 10 100 50 40 60 5 2
10 15 20 85 40 30 50 4 2
15 10 25 30 95 80 100 8 2
20 5 10 40 60 50 95 7 2
25 35 30 40 45 30 65 3 2
65 30 35 80 80 60 90 6 2
100 70 50 30 75 60 80 3 2
;
proc format ;
    value group_format 1="Jeune" 2="Agé";
run;

proc tabulate data =in ;
    class Num groupe;
    var          Prix Contenance Alcool Image couleur arome gout ses ;
    table (Num all groupe), (Prix Contenance Alcool Image couleur arome gout ses )*mean*F=6.0 ;
run;

* MACRO Graphique Variables sur data= out_variable *****;
%Macro graphiqueV (data= &data, x=&X, y=&Y, XL=&XL, YL=&YL, ID=&ID, titre = &titre) ;
axis1 length= 24 cm; axis2 length= 24 cm; symbol1 V=none interpol=none;
data out_variable ; set &data; if _TYPE_='SCORE'; run ;
proc transpose data=&data out=t_variable ; run ; proc print data=t_variable ; run ;
data work; set t_variable ; keep X Y Xsys Ysys text size color; X=&X ; Y=&Y; text=&ID ; color='BLACK' ;
size=1; xsys='2'; ysys='2'; Label Y=&YL X=&XL; run; proc gplot data=work; title4 &titre ;
plot Y*X=1 / haxis=axis1 vaxis=axis2 annotate=work Frame Href=0 Vref=0; run;
%Mend graphiqueV ;
* Macro Graphique Individus sur data= out_individu *****;
%Macro graphiqueI (data= &data, x=&X, y=&Y, XL=&XL, YL=&YL, ID=&ID, titre = &titre,
                    red=&red, blue=&blue, green=&green ) ;
axis1 length= 24 cm; axis2 length= 24 cm; symbol1 V=none interpol=none;
data work ; set &data ; keep X Y Xsys Ysys text size color ; X=&x ; Y=&y; text=&ID ; color='BLACK' ;
if text=&red then color='RED' ; if text=&blue then color='BLUE' ; if text=&green then color='GREEN' ;
size=1; xsys='2'; ysys='2'; Label Y=&YL X=&XL; run; proc gplot data=work;
plot Y*X=1 / haxis=axis1 vaxis=axis2 annotate=work Frame Href=0 Vref=0; run;
%Mend graphiqueI ;

*-----*;
Title4 "ACP" ;
*-----*;

ODS graphics on ;
PROC FACTOR DATA=in
    Plot = (SCREE all)          /* demande de l'éboulis */
    CORR                      /* sur corrélations ou COV covariance */
    SIMPLE                    /* statistiques descriptives des variables */
    MSA                       /* test d'adéquation KMO, Bartlett*/
    METHOD=PRIN PRIORS=smc     /* ACP simple priors= 1 */
/* MINEIGEN=1 */             /* critère de Kaiser valeur propres >1 */
    FLAG=5                    /* identification des loadings > à cette valeur */
    NFACT= 2                  /* choix du nombre de facteurs, obligatoire si OUT = */
    OUT = out_individu        /* sortie d'un data avec les facteurs factor1, factor2, */
    OUTSTAT = out_variable ; /* sortie d'un data avec les variables */
    VAR Prix Contenance Alcool Image couleur arome gout;          /* variables à analyser */
RUN;

```

```

*-----*;
Title4 "AFCS avec rotation varimax" ;
*-----*;
ODS graphics on ;
PROC FACTOR DATA=in
    Plot = (SCREE                /* demande de l'éboulis */
    loadings)                   /* graphiques des variables avant et après rotation */
    CORR                        /* sur corrélations ou COV covariance */
    SIMPLE                      /* statistiques descriptives des variables */
    MSA                         /* test d'adéquation KMO, Bartlett*/
    METHOD=PRIN                 PRIORS=smc /* AFCS, smc= square multiple correlation */
/* MINEIGEN=1 */              /* critère de Kaiser valeur propres >1 */
    rotate = varimax
    FLAG=.32                   /* identification des loadings > à cette valeur */
    NFACT= 2                   /* choix du nombre de facteurs, obligatoire si OUT = */
    OUT =          out_individu /* sortie d'un data avec les facteurs factor1, factor2, */
    OUTSTAT =     out_variable ; /* sortie d'un data avec les variables */
    VAR Prix Contenance Alcool Image couleur arôme gout; /* variables à analyser */
RUN;

Title4 "Data out_individu" ;
*-----*;
proc print data=out_individu (obs=20);run ;

Title4 "Data out_variable" ;
*-----*;
proc print data=out_variable (obs=20);run ;

%graphiqueV(data= out_variable,      x=factor1, y=factor2, XL="axe 1", YL="Axe 2",
            ID=_name_ , titre= "premier plan factoriel Variables (colonnes)" ) ;
%graphiqueI( data= out_individu, x=factor1, y=factor2, XL="axe 1", YL="Axe 2",
            ID=ID, titre="premier plan factoriel Individus (lignes)",
            Red= '.', Blue='.', Green ='.') ;

PROC FACTOR DATA=in
    Plot = (SCREE                /* demande de l'éboulis */
    loadings)                   /* graphiques des variables avant et après rotation */
    CORR                        /* sur corrélations ou COV covariance */
    SIMPLE                      /* statistiques descriptives des variables */
    MSA                         /* test d'adéquation KMO, Bartlett*/
    METHOD=PRIN                 PRIORS=smc /* AFCS, smc= square multiple correlation */
    rotate = promax
    FLAG=.32                   /* identification des loadings > à cette valeur */
    NFACT= 2                   /* choix du nombre de facteurs, obligatoire si OUT = */
    OUT =          out_individu /* sortie d'un data avec les facteurs factor1, factor2, */
    OUTSTAT =     out_variable ; /* sortie d'un data avec les variables */
    VAR Prix Contenance Alcool /*Image*/ couleur arôme gout; /* variables à analyser */
RUN;

* exemple avec un filtre couleur ;
%graphiqueI(data= out_individu, x=factor1, y=factor2, XL="Pack", YL="Produit",
            ID=groupe, titre="premier plan factoriel Individus (lignes)",
            Red="1", Blue= ".", Green =".") ;

proc ttest data=out_individu ;
    class group ;
    var factor1 factor2 ses ;
    format group group_fmt. ;
run ;

```



```
proc glm data = out_individu ;
  class group ;
  model ses = factor1 factor2 group / ss3 solution ;
  format group group_fmt. ;
run ;
```

Proc MDS : Analyse des similarités

Principe :

- La méthode travaille directement avec une matrice (Produit x Produit), sans mesure sur différents attributs
 - o Evaluation des paires
 - Valeur directe de la « distance » (ou dissimilarité)
 - Classement par distance croissante
 - o Evaluation d'un sous-ensemble (plan fractionné)
 - o Evaluation simplifiée (proche/non : 0/1), éventuellement par tas
 - o Comparaisons triade (Kelly)
 - o Placement des objets dans un espace (ex : Napping® 2x3 mètres)
- Ou à partir de la comparaison de profils
 - o Quanti : distance de Minkovski : $[\sum |x_i - x_j|^m]^{1/m}$.
 - si $m=2$ dist euclidienne
 - si $m=1$ dist city-block
 - o Binaires : dist de Hamming (euclidienne) : nb de désaccords
- Il n'y a donc aucune perturbation de la constitution de l'espace du fait des variables choisies par le chercheur
- (1) la première étape est une transformation des mesures
 - o LEVEL=ABSOLUTE : pas de transformation
 - o LEVEL=RATIO : transformation par une multiplication $y' = a.y$
 - o LEVEL=INTERVAL : transformation linéaire $y' = a.y + b$
 - o LEVEL=ORDINAL : transformation respectant un ordre (monotone croissante). A partir d'une configuration initiale, le stress est minimisé (attention instabilité, première configuration doit être bien choisie).
- (2) la seconde étape est une régression entre la valeur transformée et la distance issue de la représentation spatiale

Utilisations :

- Pour la création d'une représentation spatiale à partir d'une matrice (Produit x Produit)

Nature des variables : il peut s'agir de

- distances (3 critères : $d(i,j)=d(j,i)$, $d(i,i)=0$, $d(i,j) \leq d(i,k)+d(j,k)$)
- de « similarités »
- de « dissimilarités »

Décisions à prendre :

- **(1) Quelle est la nature de la variable ?**
 - o Level =
 - **absolute** : mesures métriques
 - **ordinal** : mesures de (dis)similarités

- **Similarity (=max)**. Par défaut il s'agit de dissimilarités. Cette option indique qu'il s'agit de similarités et qu'il faut soustraire la valeur d'un maximum. Si max n'est pas indiqué c'est le maximum constaté qui est utilisé.

- **(2) Combien de dimensions retenir ?**

- **Dimension = 2** choisir le plus faible nombre de dimensions possible tout en ayant une qualité de représentation acceptable
- **La qualité de la représentation est déterminée par le Stress** : qui dépend de la somme des écarts au carré des valeurs qu'il faut ajouter à chaque distance solution pour obtenir un ajustement parfait de l'ordre des valeurs initiales, sur la somme du carré des distances. <.10 correct ; <0.05 bon ; <0.025 excellent

Exemple 3 : Distances kilométriques entre villes françaises

Contexte : Les distances kilométriques inter-villes sont-elles représentatives des distances « à vol d'oiseau » ?

Distances intervilles françaises															
Obs/ville	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12	d13	d14	d15
1 amse	0														
2 ando	1020	0													
3 arge	440	760	0												
4 bala	550	1130	770	0											
5 laba	590	830	160	140	0										
6 besa	560	970	630	160	770	0									
7 bord	730	430	340	640	400	700	0								
8 boul	120	1020	480	690	550	610	830	0							
9 bour	380	680	250	500	430	350	400	480	0						
10 bres	610	1130	380	1090	270	960	620	690	830	0					
11 brux	210	1200	800	560	760	550	690	260	550	910	0				
12 cava	240	950	220	800	350	640	580	300	350	370	450	0			
13 cala	160	1080	560	690	730	610	670	34	530	720	220	340	0		
14 cham	760	750	720	350	900	260	790	870	480	1050	760	850	910	0	
15 cher	350	1030	250	920	390	760	630	420	490	350	570	120	460	970	0

MATRIX	Nombre de données non manquantes	Pondération	Critère de dissimilarité
1	1081	1.00	0.38

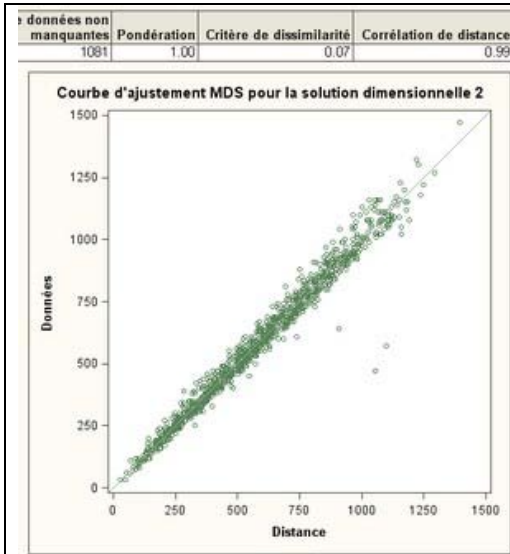
Matrice (partielle) des distances kilométriques entre les villes françaises (à comparer avec les distances « à vol d'oiseau »)

Qualité de l'ajustement avec une dimension 0.38. Très mauvaise qualité, ajustement inadmissible.

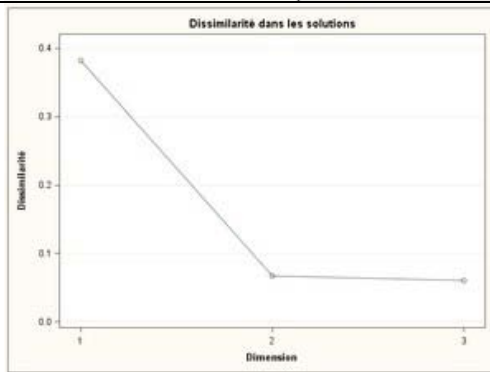
Diagramme de Sheppard avec distance initiale et distance reconstituées.

Si l'ajustement est bon, tous les points sont sur la diagonale.

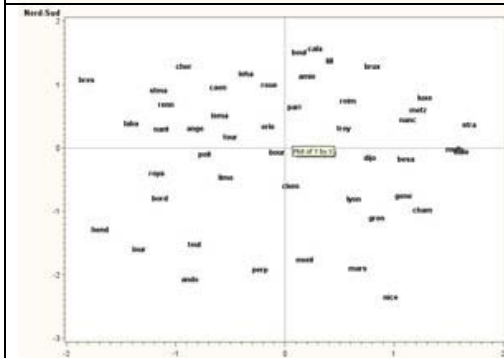
Représentation graphique en deux dimensions.(à une rotation près)



La qualité de l'ajustement est bien améliorée (la dissimilarité passe à 0.07. Elle est donc entre « correct » et « bonne ».



Graphique de l'importance du « mauvais ajustement » (badness of fit)
Le passage à 2 dimensions suffit pour bien réduire. Une troisième dimension n'apporte rien de plus.



Meilleur résultat si les données sont transformées de manière monotone (choix « ordinal »)
Rotation du graphique (Brest revient à l'ouest !)

Le programme

```
*****.
* analyse MDS : Distances intervilles françaises *;
* une seule matrice à analyser ;
* Pierre Desmet décembre 2012 ;
* http://www.psych.yorku.ca/lab/psy6140/lectures/MDS2x2.pdf ;
*****.

ods graphics on ;
Title1 "Distances intervilles françaises";
Data Villes;
    Length ville $4 ;

    Array d{47} d1-d47;
    do i=1 to 47;
        Input ville $ @@;
        do j=1 to i;
```

```

                                if j=i then input d{j} ;else input d{j} @@;
                                end;
                                output ;
                                end;
                                end;

cards;
amie 0
ando 1020 0
ange 440 760 0
bale 560 1130 770 0
laba 590 830 160 940 0
besa 560 970 620 160 770 0
bord 730 430 340 840 400 700 0
boul 120 1020 480 690 550 610 830 0
bour 380 680 260 500 430 350 400 480 0
bres 610 1130 380 1090 270 960 620 690 630 0
brux 210 1200 600 560 760 550 890 260 550 910 0
caen 240 950 220 800 350 640 580 300 360 370 450 0
cala 160 1080 580 690 730 610 870 34 530 720 220 340 0
cham 760 750 720 350 900 260 790 870 480 1050 760 850 910 0
cher 360 1030 290 920 390 760 630 420 490 390 570 120 460 970 0
clem 530 580 410 480 550 340 370 630 190 750 670 540 680 410 670 0
dijo 460 800 520 260 690 94 640 560 250 860 470 540 600 290 660 280 0
gene 680 720 640 250 800 180 690 790 380 1080 680 770 830 95 880 320 200 0
gren 710 600 640 400 840 290 660 810 390 1110 780 790 860 180 920 290 300 140 0
leha 180 1030 320 760 470 610 650 240 380 480 410 110 280 810 230 590 510 730 760 0
hend 940 440 550 1050 610 910 210 1060 620 840 1100 790 1080 1000 850 580 860 920 860 850 0
lill 120 1090 510 630 680 530 800 120 460 810 110 340 110 840 460 610 530 760 780 290 1010 0
limo 540 490 260 600 380 470 220 640 190 600 700 440 690 600 550 180 420 490 460 550 430 610 0
lour 940 270 590 1080 640 940 240 1070 580 910 1100 800 1100 900 930 520 790 830 700 870 180 1020 400 0
luxe 330 1090 650 320 860 310 960 400 480 940 230 570 410 560 660 590 320 460 620 600 1150 380 740 1060 0
lyon 610 590 540 400 700 250 550 700 280 1000 670 690 750 240 810 180 190 160 110 650 840 680 360 680 520 0
lema 350 730 94 710 250 560 410 370 230 400 510 150 410 730 270 380 460 690 710 220 650 420 290 650 590 610 0
etc;

Proc Print Data=Villes;
Run;

ODS graphics on ;
* ----- *;
Title2 "Absolute Distances : distances métriques 2 dim";
* ----- *;

Proc MDS Data=Villes Level=Absolute
    Plots(Flip)
        Dimension= 1 to 3
        PData
        PConfig Pfit
        PFinal
        out = out
        outres= res
        outfit=fitstats;
    Var d1-d47;
    Id ville;
Run;

proc print data=out (obs=20); run;
proc print data= res (obs=20); run ;
proc print data= fitstats (obs=20); run ;

* ----- *;
Title2 "Absolute Distances : distances transformées 2 dim";
* ----- *;

Proc MDS Data=Villes Level=interval
    Dimension=2 PConfig PFinal;
    Var d1-d47;
    Id ville;
Run;

```

```

* ----- *;
Title2 "Non-metric Scaling : distances ordinales 2 dim";
* ----- *;
Proc MDS Data=Villes Level=Ordinal Out=MDSout
  Dimension=2 PConfig PFinal;
  Var d1-d47;
  Id ville;
Run;
proc print data=MDSout ;
run ;

data work ; set MDSout ;
  X=-dim2 ; Y=dim1;
  text=ville ; /* variable d'identification */
  size=1; xsys='2'; ysys='2';
  Label Y='Nord-Sud' X='Ouest-Est'; /* nom des axes */
  keep X Y Xsys Ysys text size ;
run;

* ----- *;
Title4 'premier plan factoriel Profils Lignes/Individus';
* ----- *;
symbol1 V=none interpol=none; goptions reset ;
proc gplot data=work;
plot Y*X=1
  y*x=2 / overlay
  haxis=axis1 vaxis=axis2 annotate=work Frame Href=0 Vref=0; run;

/* ----- *;
Title2 "Graph avec macro plotit% ";
* ----- *;
%plotit(data=out, datatype=mds, labelvar=ville, vtoh=1.75, labfont=swissb);
run; */

```