

# Séance 10 : Analyse factorielle des correspondances

## Sommaire

Proc CORRESP : Analyse de tableaux d'effectifs .....	2
Exemple 1 : .....	6

L'analyse en composantes principales traite des variables quantitatives.

L'analyse factorielle des correspondances (AFACO) traite des données de fréquences avec différentes variantes

- Un tableau croisé ( $X1 * X2$ )
- Un tableau croisé multiple ( $(X1, X2, X3) * (X1, X2, X3)$ )
- Un tableau disjonctif (un individu par ligne).

## Proc CORRESP : Analyse de tableaux d'effectifs

**Principe** : L'analyse factorielle des correspondances reconstitue la distance entre des « profils » sur les lignes et sur les colonnes. La différence fondamentale avec l'analyse factorielle réside dans la distance utilisée : ici la distance du  $\text{Khi}^2$  ( $\text{Chi}^2$ ) qui va pondérer l'écart au carré entre les deux profils pour chaque variable par les effectifs.

**Nature des variables et de la distance** :

- Variables discrète (Nominale ou ordinale dégradée en nominale)
- Attention à ne pas laisser des modalités trop « rares » du fait d'un poids inverse aux effectifs.
- Possibilité de prendre en compte des observations additionnelles en jouant sur le poids ( $w$  : mettre le poids à 0)

**Utilisations** :

- Analyse de la proximité de variables nominale (ou binaires)
- Profil des marques connues ou appréciées
- Profil des réponses à des variables non quantitatives (la plupart des variables socio-démographiques : sexe, csp,...)

```
options nocenter formdlim="-" ;
data in ;
input Y w X1 X2 X3 @@;
ID = put(_N_,$2.) ; /* récupérer le numéro de la ligne qui est dans _N_ */

cards ;
08 1 23 06 1 22 09 2 33 14 2 11 15 2 21 13 2 22 1 15 3 11 08 1 11 19 1 33
07 1 13 16 1 31 1 3 21 1 3 31 ;
run ;
proc format ;
    value x1_F 1="X11" 2="X12" 3="X13";
    value x2_F 1="X21" 2="X22" 3="X23";
    value x3_F 1="X31" 2="X32" 3="X33";
run ;

title1 'Etude des fréquences lignes ' ;
ods graphics on ;
proc freq data=in ;
    weight w ;
    table X1*X2 / nopercnt nocol cellchi2;
    format x1 x1_f. x2 x2_F. ;
run ;

title1 'Etude des fréquences colonnes ' ;
proc freq data=in ;
    weight w ;
    table X1*X2 / nopercnt norow cellchi2;
run ;

*****
* AFC simple X1 * X2 ;
*****

title1 'Tableau C1*C2 ' ;
proc tabulate data=in;
    var w;
    class X1 X2;
    table (X1 all),(X2 all)*w*sum*F=3.0/ RTS=10;
run;
```

```

title1 'AFC simple ligne x colonne x1*x2 (SCA)' ;
Proc corresp data=in OUTC=dat_coord1 DIMENS=2 plots=all;
  weight w;
  Tables X1 , X2 ;          /* SCA : mettre une virgule entre les variables */
  format x1 x1_f. x2 x2_F. ;
run ;
proc print data=dat_coord1 ;
run;

*****
* AFC multiple X1 * X2 *X3 ;
*****
title1 'Tableau des correspondances multiples (x1,x2,x3)*(x1,x2,x3)' ;
proc tabulate data=in;
  weight w;
  class X1 X2 X3;
  table (X1 X2 X3), (X1 X2 X3)*n*F=3.0/ RTS=10;
  format x1 x1_f. x2 x2_F. x3 x3_F. ;
Run;

title1 'AFCM Correspondances multiples (x1,x2,x3)*(x1,x2,x3) (MCA)' ;
Proc corresp data=in
  mca
  OUTC=dat_coord2 DIMENS=2 ;
  weight w;
  Tables X1 X2 X3 ;
  format x1 x1_f. x2 x2_F. x3 x3_F. ;
run ;
proc print data=dat_coord2 ;
Run;

*****
* AFC disjonctif : ID X1 X2 X3 ;
*****
title1 'Tableau disjonctif complet ' ;
proc tabulate data=in;
  var w;
  class ID X1 X2 X3;
  table (ID all),(w*sum*F=3.0 (X1 X2 X3)*N*F=3.0)/ RTS=10;
  format x1 x1_f. x2 x2_F. x3 x3_F. ;
Run;

title1 'AFC Tableau disjonctif complet ' ;
Proc corresp data=in OUTC=dat_coord3 DIMENS=3 ;
  weight w ;
  Tables id , X1 X2 X3 ;
  format x1 x1_f. x2 x2_F. x3 x3_F. ;
Run ;

proc print data=dat_coord3 ;
Run;

```

## Décisions à prendre :

### - (1) Quel tableau analyser ?

- Tableau croisé simple : (table X1, X2)
- Tableaux croisés multiples : (table X1 X2 X3) (tableau de Burt)

- **Tableau disjonctif (individus en ligne) :** (table ID, X1 X2 X3)

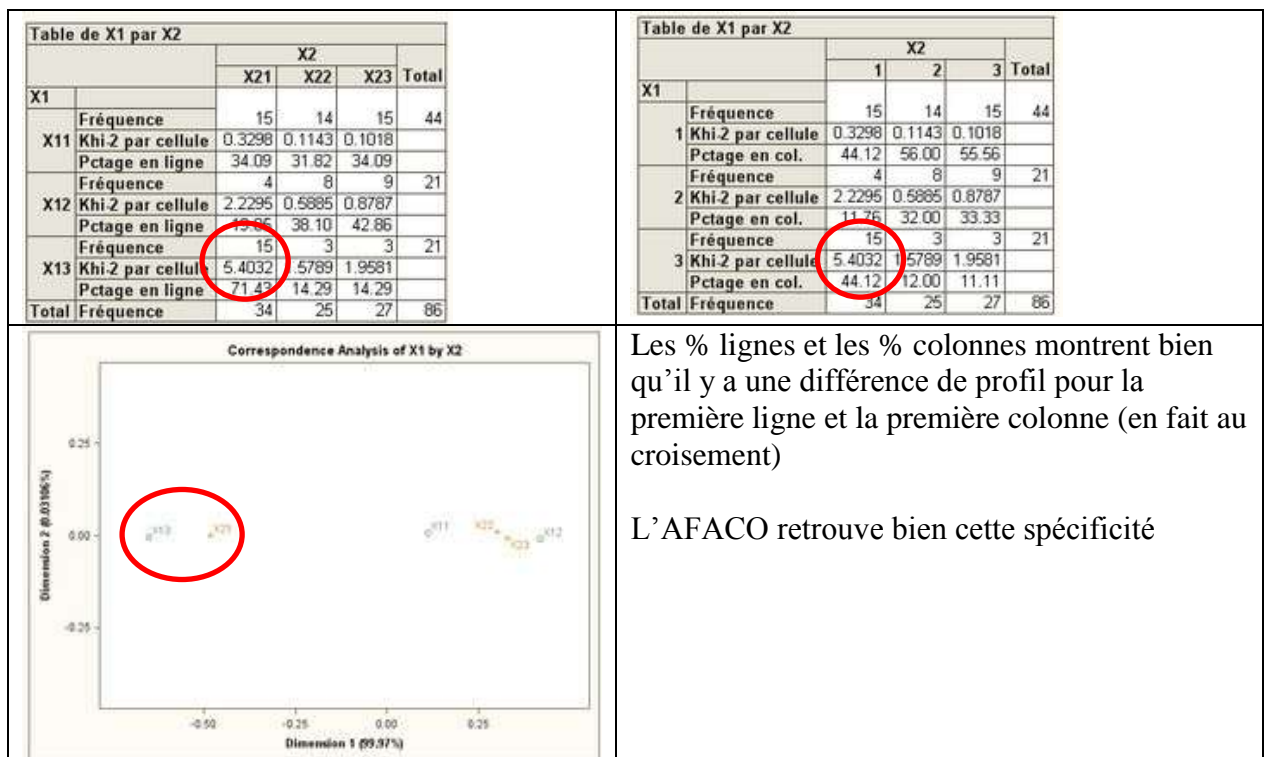
- **(2) Quelles variables éliminer ?**

- La qualité de la reconstitution de la variable est maintenant exprimée par le carré du « Cosinus » (la somme sur les axes fera 1). Un tableau donne la « qualité » de la reconstitution de la variable
- On éliminera les variables mal représentées

- **(3) Signification des axes ?**

- L'inertie est décomposée en axes qui ont une valeur singulière et représentent un % de l'inertie totale.
- Du fait du nombre des modalités, la contribution d'un axe est plus faible qu'en ACP.
- Les contributions des variables aux axes sont présentées
  - D'abord pour les colonnes (variables)
  - Puis pour les individus (lignes)

**Cas 1 : tableau croisé**



**Cas 2 : ACM analyse des correspondances multiples**

Tableau des correspondances multiples										Le tableau à analyser : croisement de toutes les variables entre elles
	X1			X2			X3			
	X11	X12	X13	X21	X22	X23	X31	X32	X33	
	N	N	N	N	N	N	N	N	N	
X1										
X11	6	.	.	2	2	2	2	1	3	
X12	.	4	.	1	2	1	2	1	1	
X13	.	.	3	1	1	1	3	.	.	
X2										
X21	2	1	1	4	.	.	3	.	1	
X22	2	2	1	.	5	.	2	2	1	
X23	2	1	1	.	.	4	2	.	2	
X3										
X31	2	2	3	3	2	2	7	.	.	
X32	1	1	.	.	2	.	.	2	.	
X33	3	1	.	1	1	2	.	.	4	

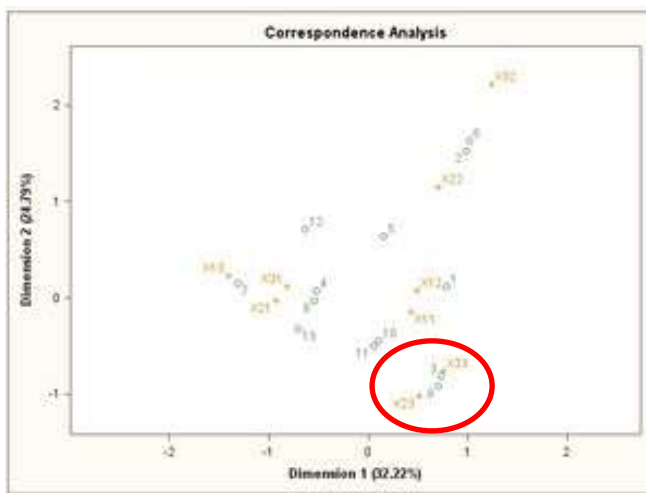
Valeur singulière	Inertie principale	Khi2	Pourcentage	Pourcent cumulé	6	12	18	24	30
0.80276	0.64442	217.809	32.22	32.22	*****	*****	*****	*****	*****
0.70415	0.49692	167.562	24.79	57.01	*****	*****	*****	*****	*****
0.55215	0.36066	118.515	17.53	74.54	*****	*****	*****	*****	*****
0.45837	0.34010	71.013	10.51	85.05	*****	*****	*****	*****	*****
0.40540	0.16436	55.548	8.22	93.27	*****	*****	*****	*****	*****
0.36896	0.13896	45.513	6.73	100.00	*****	*****	*****	*****	*****
Total	2.00000	675.980	100.00						

**Cas 3 : analyse disjonctive**

ID	w	X1			X2			X3		
		X11	X12	X13	X21	X22	X23	X31	X32	X33
		N	N	N	N	N	N	N	N	N
1	8	1	.	.	.	1	.	.	.	1
2	6	1	.	.	.	1	.	.	1	.
3	9	.	1	.	.	.	1	.	.	1
4	4	.	1	.	1	.	.	1	.	.
5	5	.	1	.	.	1	.	1	.	.
6	3	.	1	.	.	1	.	.	1	.
7	15	.	.	1	1	.	.	1	.	.
8	8	1	.	.	1	.	.	1	.	.
9	9	1	.	.	.	.	1	.	.	1
10	7	1	.	.	1	.	.	.	.	1
11	6	1	.	.	.	.	1	1	.	.
12	3	.	.	1	.	1	.	1	.	.
13	3	.	.	1	.	.	1	1	.	.
<b>Tout</b>	<b>86</b>	<b>6</b>	<b>4</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>4</b>	<b>7</b>	<b>2</b>	<b>4</b>

Le tableau à analyser est bien un tableau « classique » (individu en lignes, variables en colonnes).



On peut directement interpréter la proximité des lignes (individus) et des colonnes.

Ainsi on remarque que les profils 3 et 9 qui représentent (9+9 = 18 répondants) ont une modalité préférée pour la variable X2 (X23) et pour la variable X3 (X33). De même les individus du profil 7 (n=15) se caractérisent par la fréquence de la modalité X13

### Exemple 1 :

**Contexte :** Analyse du tableau croisé entre la CSP et le mode d'hébergement pendant les vacances. Procédez à une analyse dans les deux sens : qui va à l'hôtel ? en camping ? Pour une catégorie de CSP donnée, quels sont les modes d'hébergement privilégiés ? (données du professeur BESSE)

Table de csp par heb		heb								Total
		hôtel	locat	propri	parent	amis	tente	villag	divers	
<b>csp</b>										
Agriculteur	Fréquence	160	28	0	321	36	141	45	65	796
Salariés	Fréquence	35	34	1	178	8	0	4	0	260
Patrons	Fréquence	700	354	229	959	185	292	119	140	2978
Cadre sup	Fréquence	961	471	633	1580	305	360	162	148	4620
Cadre moy	Fréquence	572	537	279	1689	206	748	155	112	4298
Employés	Fréquence	441	404	166	1079	178	434	178	92	2972
Ouvriers	Fréquence	783	1114	387	4052	497	1464	525	387	9209
Personnels	Fréquence	65	43	21	294	79	57	18	6	583
Autres actif	Fréquence	77	60	189	839	53	124	28	53	1423
Non actifs	Fréquence	741	332	327	1789	311	236	102	102	3940
<b>Total</b>	<b>Fréquence</b>	<b>4535</b>	<b>3377</b>	<b>2232</b>	<b>12780</b>	<b>1858</b>	<b>3856</b>	<b>1336</b>	<b>1105</b>	<b>31079</b>

Tableau brut à analyser

Table de csp par heb		heb								Total
		hôtel	locat	propri	parent	amis	tente	villag	divers	
<b>csp</b>										
Agriculteur	Khi-2 par cellule	16.554	39.557	57.166	0.1222	2.8215	18.066	3.3975	47.587	
Salariés	Khi-2 par cellule	0.2276	1.1698	16.729	47.233	3.6611	32.258	4.6082	9.2442	
Patrons	Khi-2 par cellule	162.16	2.8588	1.0705	57.539	0.2725	16.249	0.635	10.994	
Cadre sup	Khi-2 par cellule	122.06	1.9147	273.44	53.83	3.0035	79.304	6.7453	1.61	
Cadre moy	Khi-2 par cellule	4.851	10.488	2.8518	3.4761	10.102	86.477	4.7933	10.9	
Employés	Khi-2 par cellule	0.1239	20.35	10.544	16.76	0.0006	11.55	19.758	1.768	
Ouvriers	Khi-2 par cellule	234.71	12.843	113.82	18.568	5.2073	90.423	42.122	10.841	
Personnels	Khi-2 par cellule	4.7352	6.536	11.402	12.28	55.91	3.2504	1.9897	10.465	
Autres actif	Khi-2 par cellule	82.196	57.904	73.731	110.12	12.091	15.643	17.987	0.1144	
Non actifs	Khi-2 par cellule	47.977	21.578	6.8548	17.593	24.171	130.77	26.797	10.354	
<b>Total</b>	<b>Fréquence</b>	<b>4535</b>	<b>3377</b>	<b>2232</b>	<b>12780</b>	<b>1858</b>	<b>3856</b>	<b>1336</b>	<b>1105</b>	<b>31079</b>

Khi<sup>2</sup> par cellule.

Décomposition de l'inertie et du Khi-2					
Valeur singulière	Inertie principale	Khi-2	Pourcentage	Pourcent. cumulé	
0.21051	0.04432	1377.28	53.01	53.01	-----+-----+-----+-----+-----
0.14230	0.02025	629.32	24.22	77.23	*****
0.09383	0.00880	273.65	10.53	87.76	*****
0.07393	0.00547	169.86	6.54	94.30	***
0.04880	0.00238	74.01	2.85	97.15	*
0.04632	0.00215	66.67	2.57	99.71	*
0.01551	0.00024	7.48	0.29	100.00	
Total	0.08360	2598.27	100.00		

Deux axes reconstituent 77% de l'inertie

