Analyse des correspondances multiples (TP SAS)

Exemple Pommes

Tableau croisant les modalités de 9 variables qualitatives.

Les variables sont mesurées sur 57 variétés de pommes.

La plupart de ces variétés sont génétiquement différentes, d'autres sont issues de mutations et représentent la part la plus importante des variétés commercialisées aujourd'hui car elles apportent un petit *plus* par rapport à la variété standard dont elles dérivent, notamment par leur coloration.

Les variables concernent différents types de caractères : forme de l'arbre, date de récolte, description et qualité du fruit. Certaines de ces variables ont des modalités ordonnées et d'autres non.

arb forme de l'arbre (4 modalités)

rec date de la récolte (3 modalités)

cal calibre du fruit (3 modalités)

coul couleur dominante du fruit ou intensité de la couleur rouge additionnelle (3 modalités)

pour proportion de coloration rouge additionnelle (4 modalités)

type type de coloration du fruit (3 modalités)

form forme du fruit (3 modalités)

ferm fermeté du fruit (3 modalités)

gout rapport sucre/acidité (3 modalités)

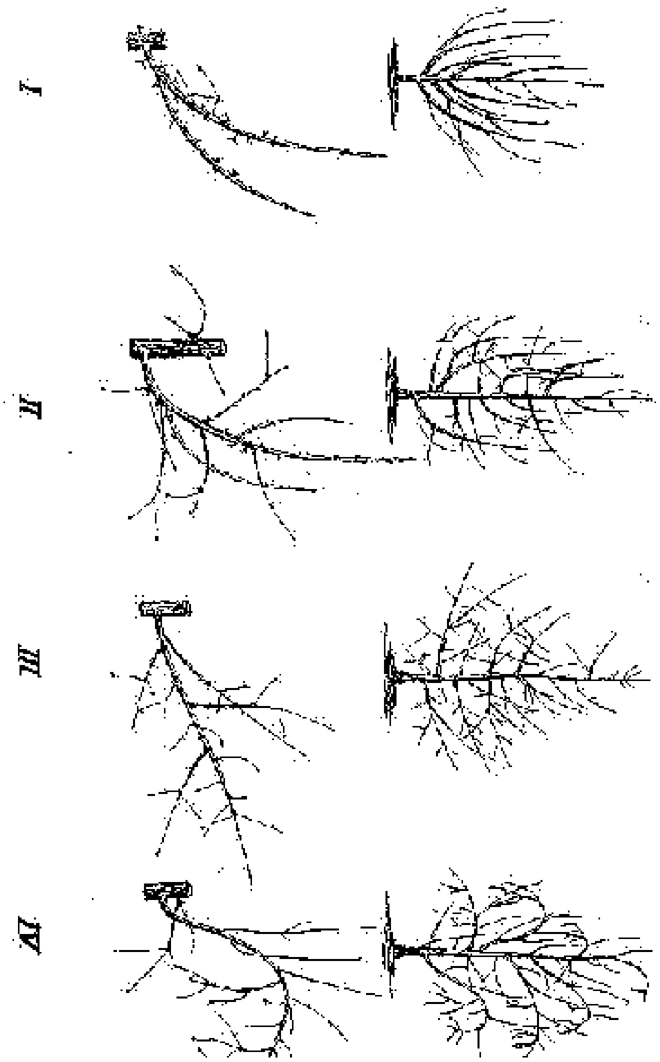
Fichiers de Données : Pommes.txt

Tableau croisant les modalités de 9 variables qualitatives mesurées sur 57 variétes de Pommes.

	arb	rec	cal	coul	pour	type	form	ferm	gout
Granny-spur	1	3	2	1	1	1	2	3	3
Granny-stand.	4	3	2	1	1	1	2	3	3
Boskoop-mut.	3	3	3	4	4	1	1	2	2
Boskoop-stand.	2	3	3	4	3	1	1	2	2
Gala-red	2	2	1	3	3	3	2	2	1
Gala	3	2	1	3	2	3	2	2	1
Elstar	3	1	1	2	2	3	2	2	3
Elista	3	1	1	3	3	1	2	2	3
Jonagold	3	2	3	3	2	2	2	2	3
Jonagored	3	2	3	4	3	3	2	2	3
New-Jonagold	3	2	3	2	3	1	2	2	3
...	...	...	...	...	...	...	...	...	...
Stayman-Winesap	4	3	3	4	4	1	2	3	1
Hyb1	3	2	1	3	4	1	2	2	1
Hyb2	3	1	2	2	3	2	1	2	2
Rubinette	3	2	1	3	3	2	2	2	2
Rubin	4	1	2	2	4	3	2	2	3
Hyb3	4	2	2	4	4	1	2	3	3
Hyb4	3	2	2	1	1	1	2	1	2

À l'intersection d'une ligne et d'une colonne, on trouve le numéro de la modalité.

Grille caractérisant les ports d'arbres chez le pommier
(d'après J.M. Lespinasse, 1977)



Création du tableau SAS

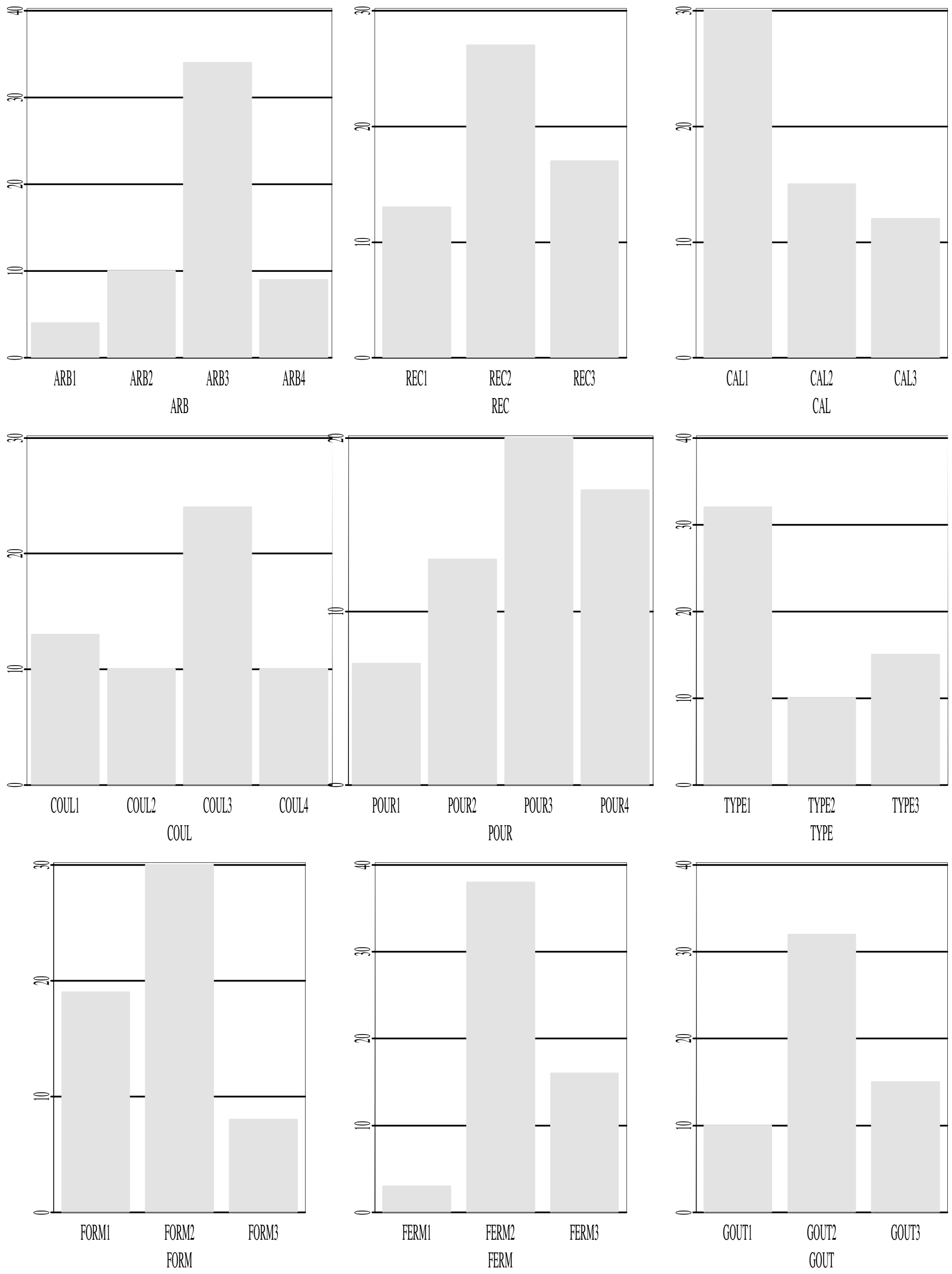
```
Data Pom ;
 infile 'pommes.txt' firstobs=2;
 input Nom :$18. arb$ rec$ cal$ coul$
       pour$ type$ form$ ferm$ gout$ ;
  id=_n_;
 length labl $8.;
 array tmp(*) arb--gout;
  do i=1 to dim(tmp);
     call vname(tmp{i},labl);
     tmp{i}=compress(labl||tmp{i});
  end;
drop i labl;
run;

Proc print data=Pom(obs=10);
run;
```

OBS	NOM	ARB	REC	CAL	COUL	POUR	TYPE	FORM	FERM	GOUT	ID
1	Granny-spur	ARB1	REC3	CAL2	COUL1	POUR1	TYPE1	FORM2	FERM3	GOUT3	1
2	Granny-stand.	ARB4	REC3	CAL2	COUL1	POUR1	TYPE1	FORM2	FERM3	GOUT3	2
3	Boskoop-mut.	ARB3	REC3	CAL3	COUL4	POUR4	TYPE1	FORM1	FERM2	GOUT2	3
4	Boskoop-stand.	ARB2	REC3	CAL3	COUL4	POUR3	TYPE1	FORM1	FERM2	GOUT2	4
5	Gala-red	ARB2	REC2	CAL1	COUL3	POUR3	TYPE3	FORM2	FERM2	GOUT1	5
6	Gala	ARB3	REC2	CAL1	COUL3	POUR2	TYPE3	FORM2	FERM2	GOUT1	6
7	Elstar	ARB3	REC1	CAL1	COUL2	POUR2	TYPE3	FORM2	FERM2	GOUT3	7
8	Elista	ARB3	REC1	CAL1	COUL3	POUR3	TYPE1	FORM2	FERM2	GOUT3	8
9	Jonagold	ARB3	REC2	CAL3	COUL3	POUR2	TYPE2	FORM2	FERM2	GOUT3	9
10	Jonagored	ARB3	REC2	CAL3	COUL4	POUR3	TYPE3	FORM2	FERM2	GOUT3	10

Exploration graphique des données

```
Goptions nodisplay;

Proc Datasets library=work;
  delete graphe/ memtype=catalog;
run;
quit;

Proc Gchart data=Pom gout=graphe;
 vbar arb rec cal coul pour type form ferm gout
      /autoref clipref frame;
run;
quit;

Goptions display;

%inc 'panels.sas';

%panels(rows=3,cols=3,plots=9,order=DOWN,equate=Y,
        gin=graphe,gout=graphe);
```

Histogrammes des modalités des différentes variables



ACM du tableau de Burt

```
* ACM du tableau de Burt: Proc Corresp avec l'option MCA;

Proc corresp data=Pom mca outc=graphem dimens=5;
 table arb rec cal coul pour type form ferm gout;
run;

The Correspondence Analysis Procedure

Inertia and Chi-Square Decomposition

Singular  Principal Chi-
Values     Inertias  Squares Percents    3    6    9   12   15
                                     ----+----+----+----+----+--
0.5641   0.31830   253.484  13.64% ***********************
0.5005   0.25056   199.538  10.74% ******************
0.4909   0.24101   191.933  10.33% *****************
0.4412   0.19473   155.075   8.35% **************
0.4171   0.17397   138.546   7.46% ************
0.3858   0.14885   118.538   6.38% ***********
0.3714   0.13797   109.877   5.91% **********
0.3627   0.13157   104.776   5.64% *********
0.3285   0.10791    85.935   4.62% ********
0.3162   0.10003    79.661   4.29% *******
0.2974   0.08848    70.459   3.79% ******
0.2835   0.08042    64.043   3.45% ******
0.2642   0.06984    55.618   2.99% *****
...
0.1937   0.03755    29.901   1.61% ***
0.1702   0.02900    23.093   1.24% **
0.1184   0.01404    11.181   0.60% *
0.1041   0.01084     8.635   0.46% *
         -------   -------
         2.33333   1858.16 (Degrees of Freedom = 841)
```

Column Coordinates

	Dim1	Dim2	Dim3	..	Dim5
ARB1	1.43593	0.86168	0.69122	..	1.53986
ARB2	-0.36819	0.00814	1.00545	..	-0.85619
ARB3	-0.37983	0.06541	-0.27529	..	0.21161
ARB4	1.20582	-0.63911	-0.38438	..	-0.53247
REC1	-0.84063	-0.23319	-0.58778	..	0.91804
REC2	-0.23762	0.34715	0.54702	..	-0.36714

Summary Statistics for the Column Points

	Quality	Mass	Inertia
ARB1	0.449865	0.007797	0.044277
ARB2	0.438455	0.019493	0.039265
ARB3	0.417789	0.066277	0.019215
ARB4	0.438550	0.017544	0.040100
REC1	0.579459	0.025341	0.036759
REC2	0.619089	0.052632	0.025063

Partial Contributions to Inertia for the Column Points

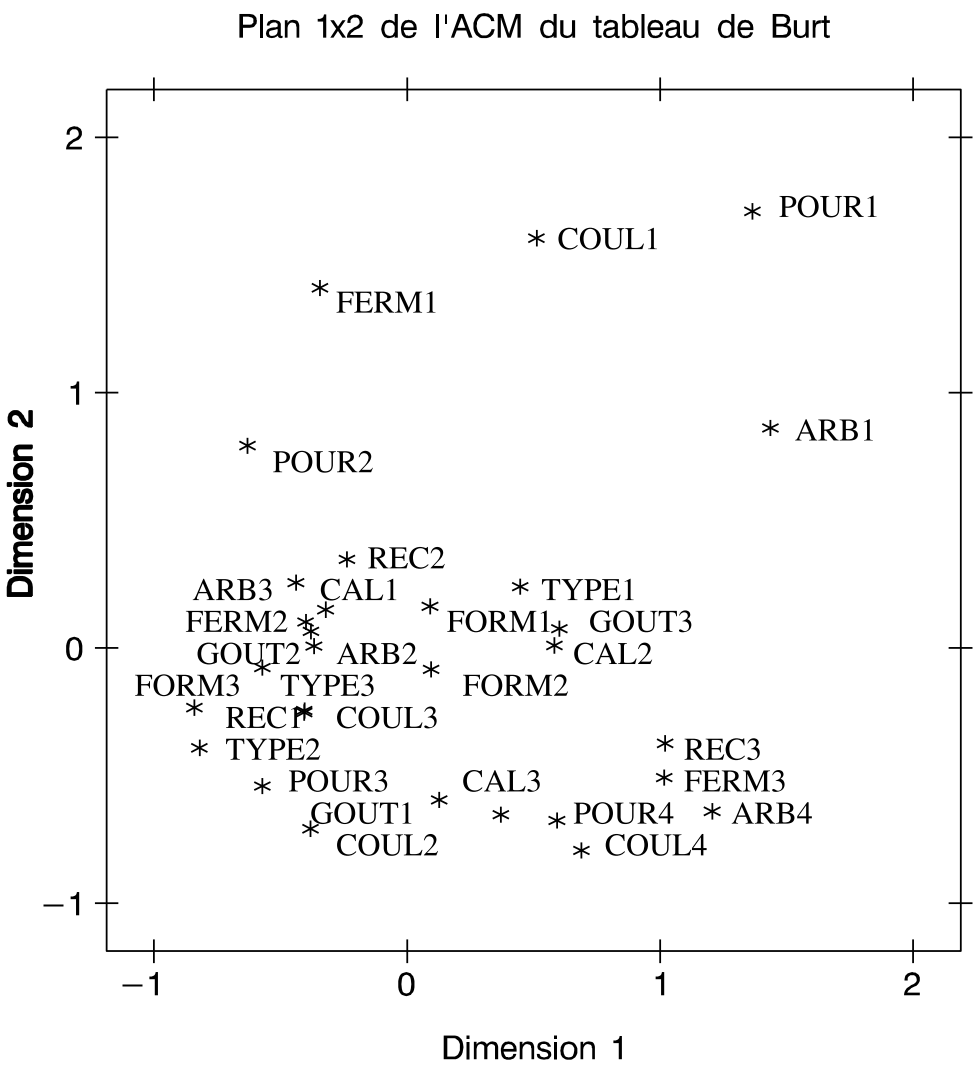
	Dim1	Dim2	Dim3	..	Dim5
ARB1	0.050509	0.023105	0.015457	..	0.106272
ARB2	0.008302	0.000005	0.081763	..	0.082137
ARB3	0.030040	0.001132	0.020841	..	0.017058
ARB4	0.080140	0.028599	0.010755	..	0.028591
REC1	0.056258	0.005500	0.036326	..	0.122760
REC2	0.009336	0.025314	0.065344	..	0.040778

Squared Cosines for the Column Points

	Dim1	Dim2	Dim3	..	Dim5
ARB1	0.155616	0.056037	0.036059	..	0.178956
ARB2	0.028844	0.000014	0.215091	..	0.155971
ARB3	0.213270	0.006324	0.112032	..	0.066194
ARB4	0.272625	0.076586	0.027702	..	0.053160
REC1	0.208783	0.016066	0.102076	..	0.249006
REC2	0.050816	0.108462	0.269304	..	0.121314

Plan 1× 2 de l'ACM du tableau de BURT

```
%inc 'plotit.sas';

title "Plan 1x2 de l'ACM du tableau de Burt";

%plotit(data=graphem,plotvars=dim2 dim1,labelvar=_name_,
       datatype=mca,method=print,post=tmp.eps,options=square,
       labfont=HWpsl005,monochro=black,
       gopprint=dev=psepsf gsfmode=replace gaccess=gsasfile);
```



Création du tableau des indicatrices : Ind

```
Proc glmmod data=Pom outdesign=ind noprint;
 class arb rec cal coul pour type form ferm gout;
 model id=arb rec cal coul pour type form ferm gout /noint;
run;

Data Ind;
 merge Pom(keep= Nom) Ind(drop=id);
run;

Proc print data=ind(obs=10);run;
```

```
                                    C C C C C C C C C
                  C C C C C C C C C O O O O O O O O O
 O N              O O O O O O O O O L L L L L L L L L
 B O              L L L L L L L L L 1 1 1 1 1 1 1 1 1
 S M              1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8

 1 Granny-spur    1 0 0 0 0 0 1 0 1 0 1 0 0 0 1 0 0 0
 2 Granny-stand.  0 0 0 1 0 0 1 0 1 0 1 0 0 0 1 0 0 0
 3 Boskoop-mut.   0 0 1 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1
 4 Boskoop-stand. 0 1 0 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0
 5 Gala-red       0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
 6 Gala           0 0 1 0 0 1 0 1 0 0 0 0 1 0 0 1 0 0
 7 Elstar         0 0 1 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0
 8 Elista         0 0 1 0 1 0 0 1 0 0 0 0 1 0 0 0 1 0
 9 Jonagold       0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 1 0 0
10 Jonagored      0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 1 0
```

```
* AFC du tableau des Indicatrices;

Proc Corresp data=Ind outc=graphe dimens=5;
 var col1-col30;
 id nom;
run;

The Correspondence Analysis Procedure

Inertia and Chi-Square Decomposition

Singular  Principal Chi-
Values     Inertias  Squares Percents    3    6    9   12   15
                                      ----+----+----+----+----+---
0.5641   0.31830   163.290  13.64% ***********************
0.5005   0.25056   128.539  10.74% ******************
0.4909   0.24101   123.640  10.33% *****************
0.4412   0.19473    99.897   8.35% **************
0.4171   0.17397    89.249   7.46% ************
0.3858   0.14885    76.360   6.38% ***********
0.3714   0.13797    70.781   5.91% **********
0.3627   0.13157    67.495   5.64% *********
0.3285   0.10791    55.358   4.62% ********
0.3162   0.10003    51.317   4.29% *******
0.2974   0.08848    45.389   3.79% ******
0.2835   0.08042    41.255   3.45% ******
0.2642   0.06984    35.828   2.99% *****
...
0.1937   0.03755    19.262   1.61% ***
0.1702   0.02900    14.876   1.24% **
0.1184   0.01404     7.203   0.60% *
0.1041   0.01084     5.562   0.46% *
         -------   -------
         2.33333      1197 (Degrees of Freedom = 1624)
```

Row Coordinates

	Dim1	Dim2	Dim3	...	Dim5
Granny-spur	1.39327	0.78567	-0.47838	...	0.52196
Granny-stand.	1.34795	0.45253	-0.72181	...	-0.03008
Boskoop-mut.	0.41507	-0.39317	0.64964	...	-0.03076
Boskoop-stand.	0.18816	-0.37585	0.68301	...	-0.43833
Gala-red	-0.51340	-0.22220	0.23002	...	-0.54687

Summary Statistics for the Row Points

	Quality	Mass	Inertia
Granny-spur	0.769807	0.017544	0.030054
Granny-stand.	0.817156	0.017544	0.023440
Boskoop-mut.	0.389304	0.017544	0.014950
Boskoop-stand.	0.380406	0.017544	0.017891
Gala-red	0.438244	0.017544	0.015736

Partial Contributions to Inertia for the Row Points

	Dim1	Dim2	Dim3	...	Dim5
Granny-spur	0.106992	0.043220	0.016658	...	0.027474
Granny-stand.	0.100145	0.014339	0.037926	...	0.000091
Boskoop-mut.	0.009496	0.010824	0.030720	...	0.000095
Boskoop-stand.	0.001951	0.009891	0.033957	...	0.019375
Gala-red	0.014528	0.003457	0.003851	...	0.030158

Squared Cosines for the Row Points

	Dim1	Dim2	Dim3	...	Dim5
Granny-spur	0.485647	0.154428	0.057252	...	0.068160
Granny-stand.	0.582828	0.065689	0.167126	...	0.000290
Boskoop-mut.	0.086648	0.077747	0.212254	...	0.000476
Boskoop-stand.	0.014879	0.059366	0.196049	...	0.080747
Gala-red	0.125942	0.023590	0.025280	...	0.142894

Column Coordinates

	Dim1	Dim2	Dim3	...	Dim5
COL1	1.43593	0.86168	0.69122	...	1.53986
COL2	-0.36819	0.00814	1.00545	...	-0.85619
COL3	-0.37983	0.06541	-0.27529	...	0.21161
COL4	1.20582	-0.63911	-0.38438	...	-0.53247
COL5	-0.84063	-0.23319	-0.58778	...	0.91804
COL6	-0.23762	0.34715	0.54702	...	-0.36714

Summary Statistics for the Column Points

	Quality	Mass	Inertia
COL1	0.449865	0.007797	0.044277
COL2	0.438455	0.019493	0.039265
COL3	0.417789	0.066277	0.019215
COL4	0.438550	0.017544	0.040100
COL5	0.579459	0.025341	0.036759
COL6	0.619089	0.052632	0.025063

Partial Contributions to Inertia for the Column Points

	Dim1	Dim2	Dim3		Dim5
COL1	0.050509	0.023105	0.015457	...	0.106272
COL2	0.008302	0.000005	0.081763	...	0.082137
COL3	0.030040	0.001132	0.020841	...	0.017058
COL4	0.080140	0.028599	0.010755	...	0.028591
COL5	0.056258	0.005500	0.036326	...	0.122760
COL6	0.009336	0.025314	0.065344	...	0.040778

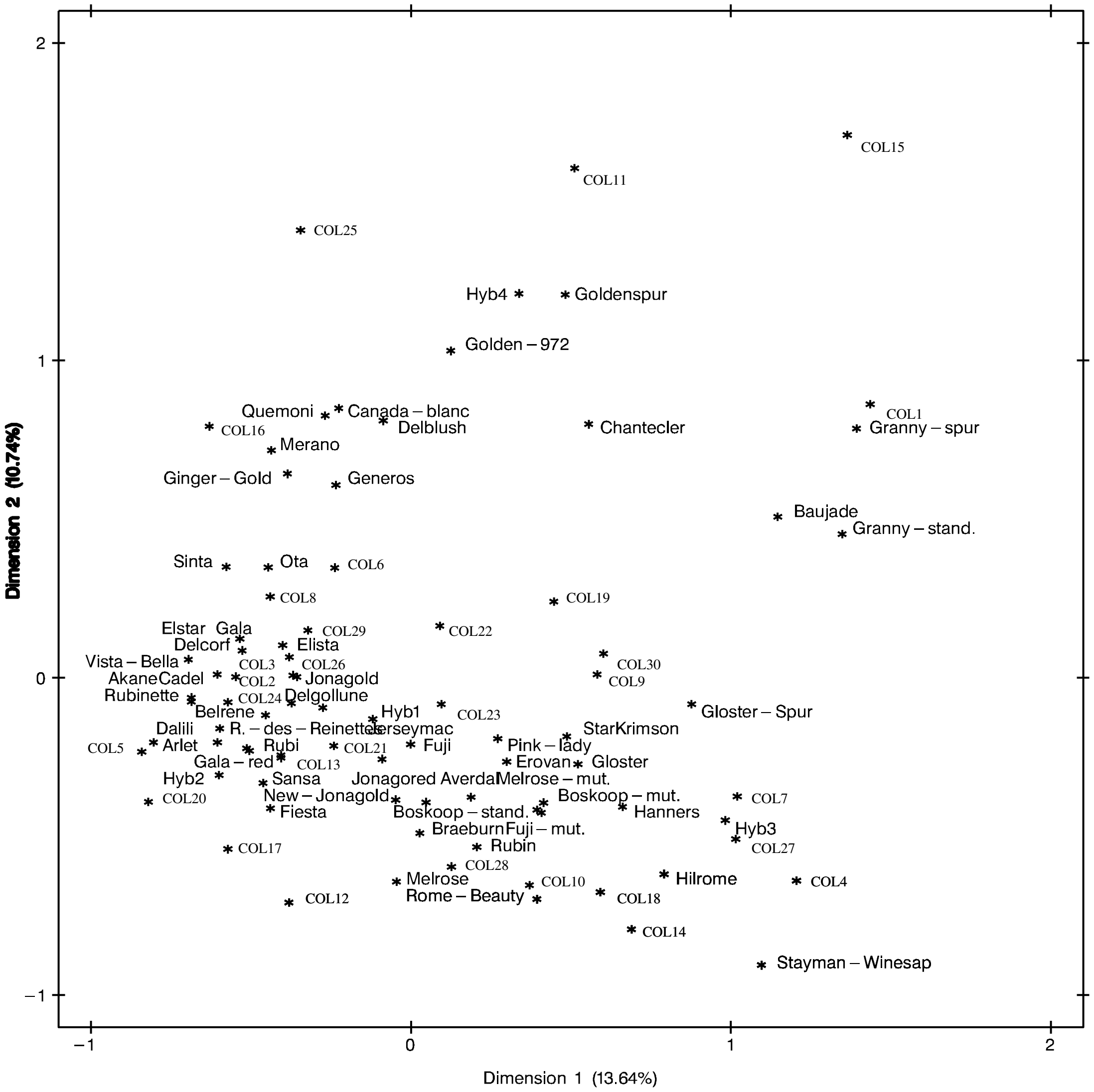
Squared Cosines for the Column Points

	Dim1	Dim2	Dim3		Dim5
COL1	0.155616	0.056037	0.036059	...	0.178956
COL2	0.028844	0.000014	0.21509	...	0.155971
COL3	0.213270	0.006324	0.112032	...	0.066194
COL4	0.272625	0.076586	0.027702	...	0.053160
COL5	0.208783	0.016066	0.102076	...	0.249006
COL6	0.050816	0.108462	0.269304	...	0.121314

Plan 1× 2 de l'AFC du tableau d'indicatrices

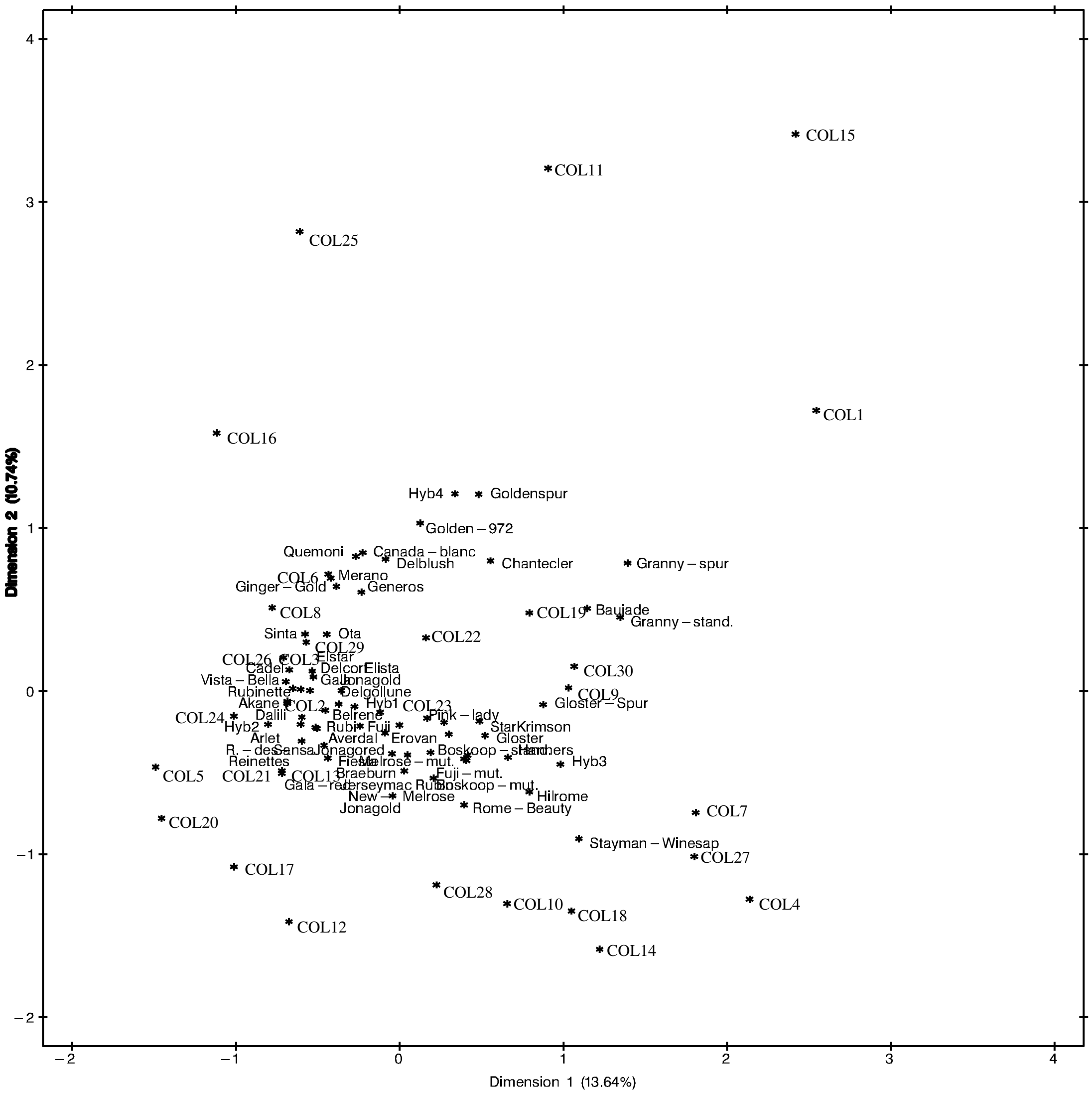
```
title "Plan 1x2: representation simultanee";

%plotit(data=graphe,plotvars=dim2 dim1,labelvar=nom,
        datatype=corresp,method=print,post=AFC12.eps,
        options=square,labfont=HWpsl005,monochro=black,
        gopprint=dev=psepsf gsfmode=replace gaccess=gsasfile);
```

Plan 1× 2 : représentation simultanée

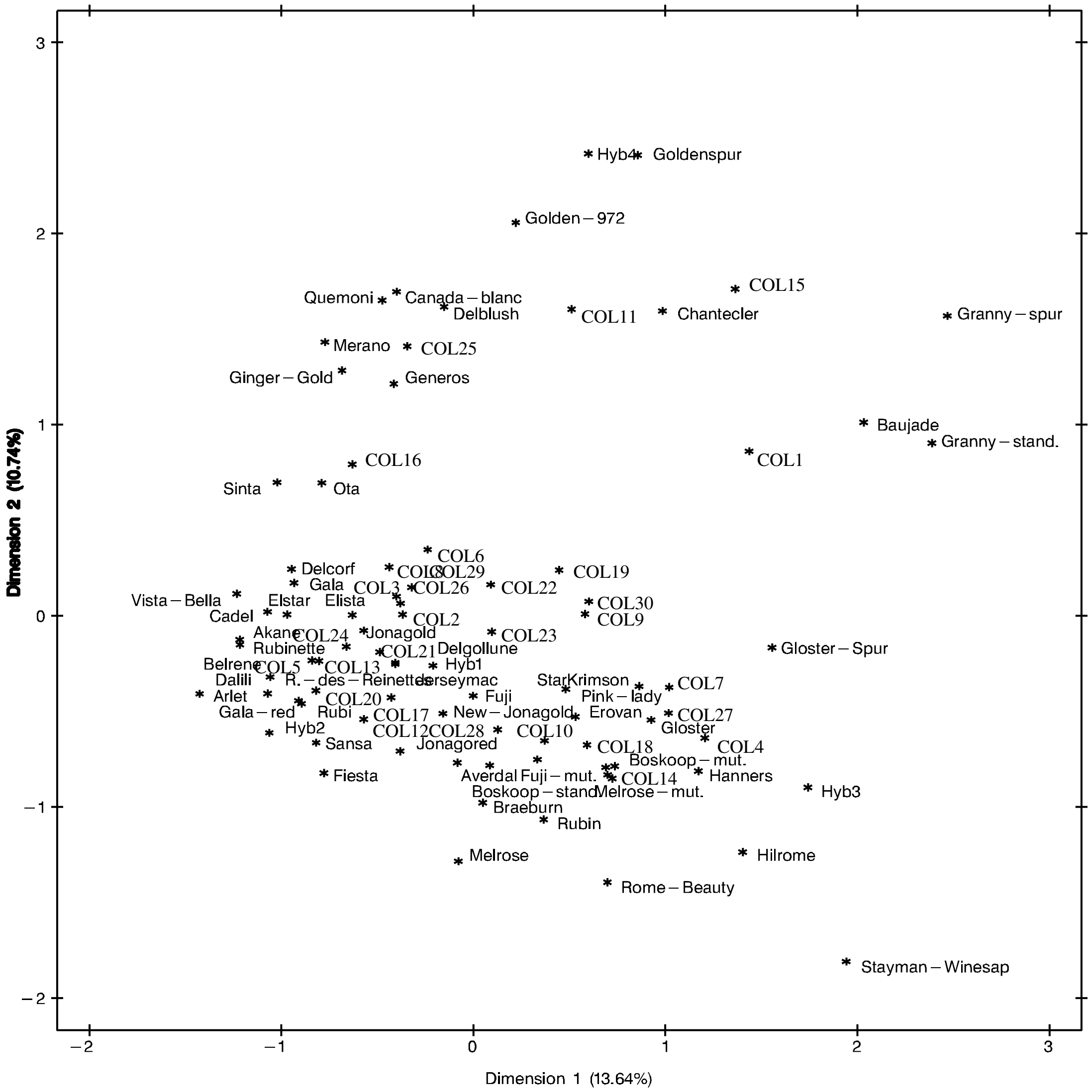


Autres représentations

```
Proc Corresp data=Ind outc=graphe noprint profile=row;
 var col1-col30;
 id nom;
run;

%plotit(data=graphe,plotvars=dim2 dim1,labelvar=nom,
        datatype=corresp,options=square,
        method=gplot,gopplot=device=xcolor);
```



```
Proc Corresp data=Ind outc=graphe noprint profile=column;
 var col1-col30;
 id nom;
run;

%plotit(data=graphe,plotvars=dim2 dim1,labelvar=nom,
        datatype=corresp,options=square,
        method=gplot,gopplot=device=xcolor);
```



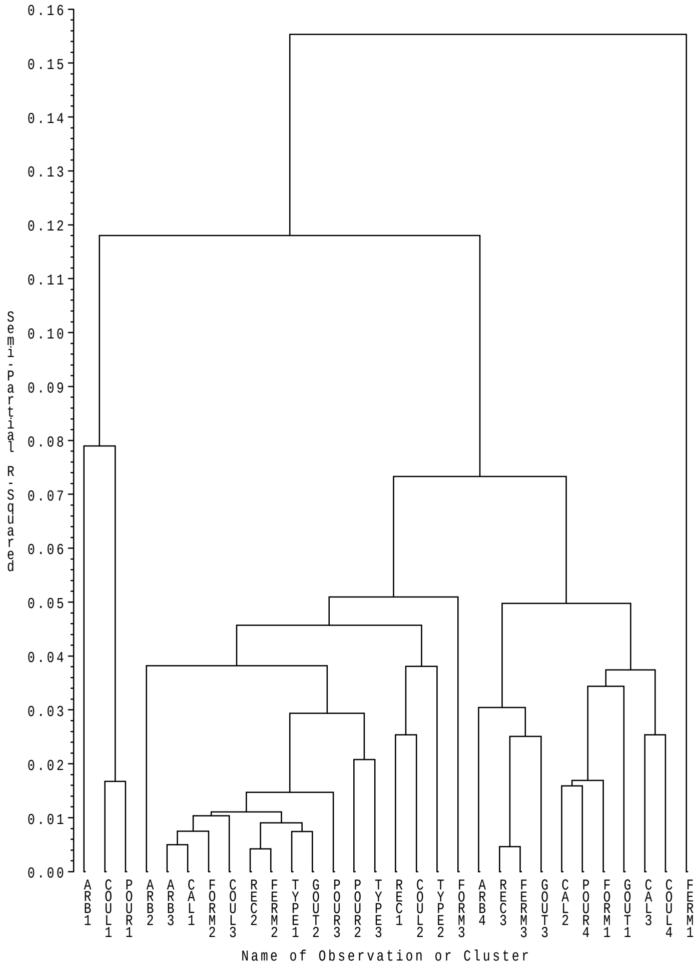
Classification hiérarchique des modalités

```
Proc corresp data=Pom dimens=21 noprint
     outc=graphem(keep=Dim1-Dim21 _name_ _type_);
 table nom, arb rec cal coul pour type form ferm gout;
run;

Proc print data=Graphem(where=(_type_='VAR'));
run;
```

```
 OBS _TYPE_  _NAME_     DIM1       DIM2       DIM3       DIM4

  59  VAR    ARB1      1.43593   0.86168   0.69122  -0.55440
  60  VAR    ARB2     -0.36819   0.00814   1.00545   0.42558
  61  VAR    ARB3     -0.37983   0.06541  -0.27529  -0.11623
  62  VAR    ARB4      1.20582  -0.63911  -0.38438   0.21261
  63  VAR    REC1     -0.84063  -0.23319  -0.58778   0.10927
  64  VAR    REC2     -0.23762   0.34715   0.54702  -0.27727
  65  VAR    REC3      1.02022  -0.37303  -0.41931   0.35681
  66  VAR    CAL1     -0.43929   0.25595  -0.27714  -0.10646
```

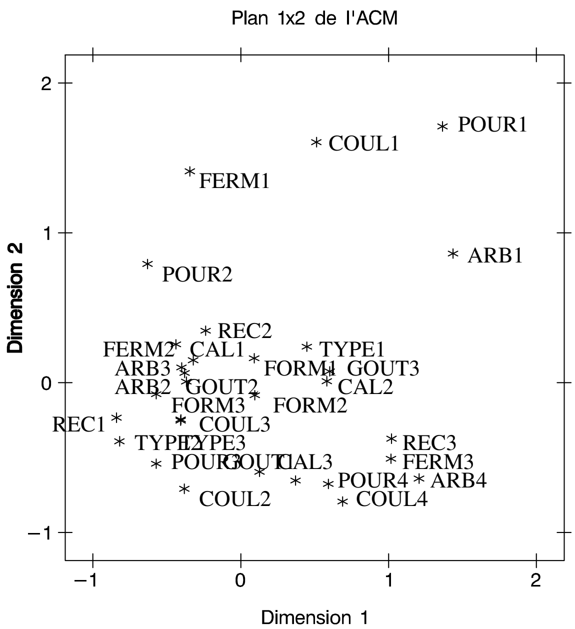
Classification hiérarchique des modalités

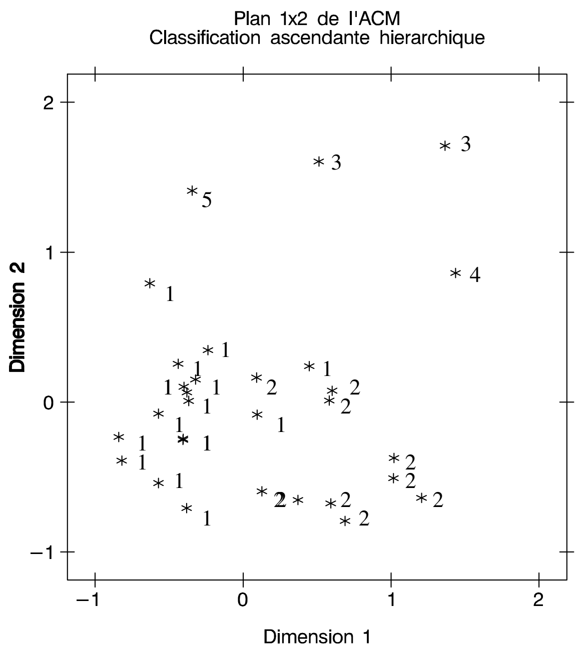
```
Proc cluster data=Graphem(where=(_type_='VAR'))
     outtree=arbre method=ward ;
 var Dim1-dim21;
 id _name_;
run;

Proc tree data=arbre n=5 out=tree graphics;
 id _name_;
 copy dim1-dim21;
run;
```



Représentation de la classification sur le plan 1×2 de l'ACM

```
title "Plan 1x2 de l'ACM du tableau de Burt";

%plotit(data=tree,plotvars=dim2 dim1,labelvar=_name_,
       datatype=mca,method=gplot,gopplot=device=xcolor,
       options=square);
%plotit(data=tree,plotvars=dim2 dim1,labelvar=cluster,
       datatype=mca,method=gplot,gopplot=device=xcolor,
       options=square);
```





Classification par les centres mobiles

```
Proc means data=Tree noprint;
 class cluster;
 var dim1-dim21;
 output out=moy(where=(_type_=1)) mean=;
run;
Proc fastclus data=Graphem(where=(_type_='VAR'))
              seed=moy mean=moy2 out=mixte maxc=6 maxiter=7;
 var dim1-dim21;
run;
```