Analyse en composantes principales (TP SAS)

©1998 - Formation statistique - Module 7/8 (ACP TP SAS)-1

Lecture des données

```
data tp;
infile 'forsy' firstobs=2 expandtabs;
input Nech$ lp lcal lsep ltc pgl ltp lte lf del;
format Nech 4. lp 5.2 lcal 4.2 lsep 4.2 ltc 5.2 pgl 5.2 ltp 4.2
lte 4.2 lf 4.2 del 5.2;
run;

proc print;
title 'Tableau de Données';
run;
```

Édition des statistiques élémentaires

```
proc means data=tp noprint;
var lp--del;
output out=stat1;
run;
data stat2 (drop=_type_ _freq_);
set stat1;
if _stat_='N' or _stat_='MEAN' or _stat_='STD';
run;
proc transpose data=stat2 out=statel;
id _stat_;
run;
proc print noobs;
title 'Statistiques élémentaires';
run;
```


Création du fichier "tp" à partir du fichier "forsy". Ce fichier provient d'un tableur et les données sont séparées par des tabulations : l'option expandtabs en permet la lecture. La lecture est effectuée à partir de la deuxième ligne du fichier (la première contient le nom des variables).

La variable "Nech" (alpha-numérique) est l'identificateur de l'individu. Le fichier "tp" comporte 54 individus et 9 variables.

Calcul et édition des statistiques élémentaires (N = nombre d'individus, mean = moyenne, std = écart-type) pour les variables "lp" à "del".

Analyse univariée

```
proc univariate data=tp plot;
var lp;
run;
```

```
Variable=LP

                        Moments

        N                 54    Sum Wgts          54
        Mean         8.00463    Sum           432.25
        Std Dev     2.823691    Variance    7.973233
        Skewness    0.232894    Kurtosis    -0.71589
        USS         3882.583    CSS         422.5813
        CV          35.27573    Std Mean    0.384256
        T:Mean=0    20.83152    Pr>|T|        0.0001
        Num ^= 0          54    Num > 0           54
        M(Sign)           27    Pr>=|M|       0.0001
        Sgn Rank       742.5    Pr>=|S|       0.0001
```


Analyse univariée de la variable "lp" du fichier "tp" ;
(l'option plot permet d'obtenir le tracé du boxplot et du graphique "quantile-quantile").
Une analyse univariée simultanée de toutes les variables ("lp" à "del") peut être obtenue par l'instruction :
proc univariate data=tp plot;
var lp–del;
run;
Cette analyse sera demandée aux stagiaires

N : nombre d'observations (chaque observation a le même poids).

Skewness : paramètre de symétrie et **Kurtosis** : paramètre d'aplatissement (Kurtosis = 0 pour une distribution normale).
USS et **CSS** : somme des carrés non corrigée et corrigée,
CV = coefficient de variation,
Std Mean = écart-type de la moyenne.
L'hypothèse de nullité de la moyenne des observations est testée de 3 façons différentes (Student, test du signe et test des rangs). Ces tests ne présentent ici aucun intérêt.

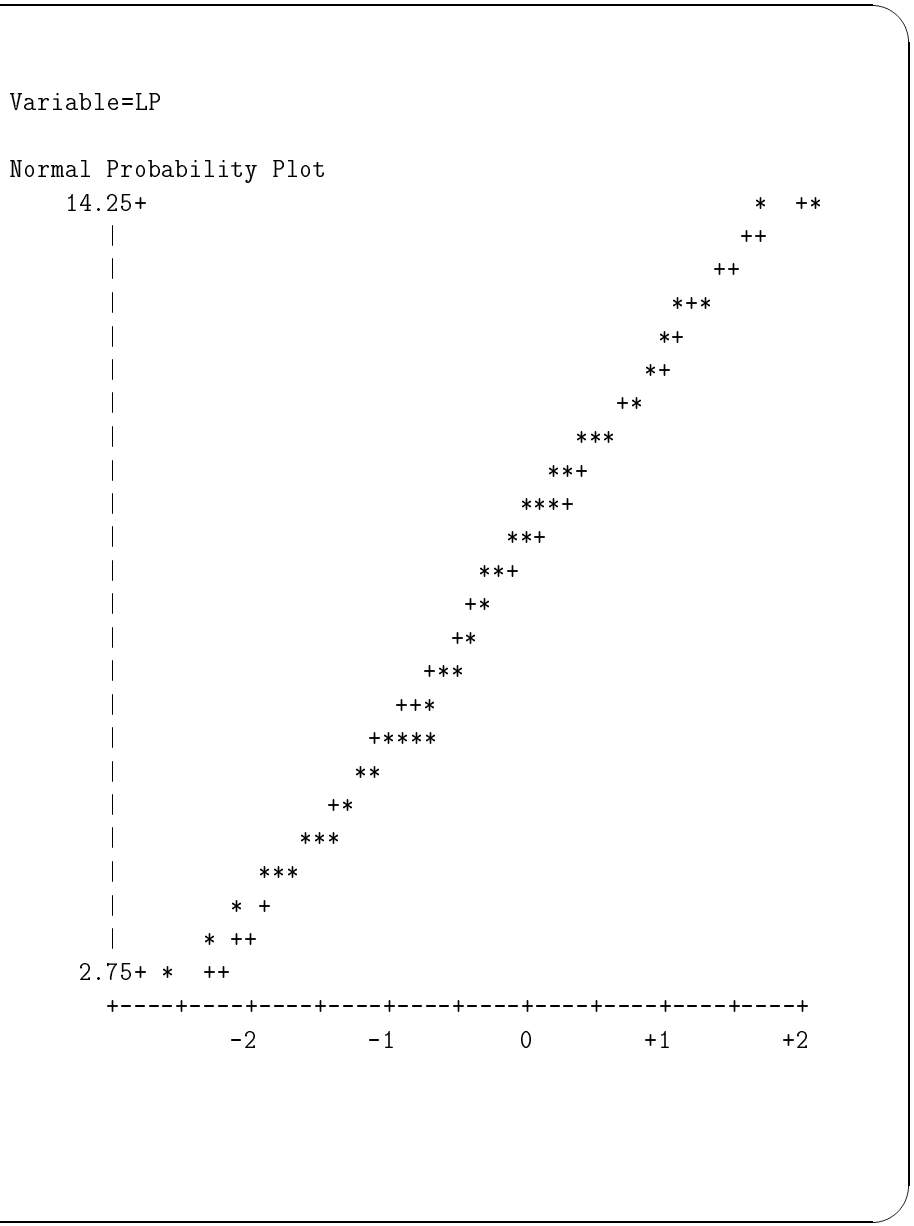
Sur les sorties (non reproduites ici en intégralité) on trouvera de plus : les quartiles et certains centiles (1 %,5 %, ...), l'amplitude, l'amplitude inter-quartile, le mode, les valeurs extrêmes. L'identification des valeurs extrêmes peut être obtenue par l'instruction : id nech;

```
Variable=LP

          Stem Leaf            #          Boxplot
            14 03              2              |
            13                                |
            13                                |
            12 58              2              |
            12 1               1              |
            11 5               1              |
            11 3               1              |
            10 57799           5              |
            10 011             3          +--------+
             9 5578            4          |        |
             9 0012            4          |        |
             8 68              2          |        |
             8 2               1          |   +    |
             7 68              2          *--------*
             7 02244           5          |        |
             6                            |        |
             6 0002334         7          |        |
             5 699             3          +--------+
             5 4               1              |
             4 6699            4              |
             4 0014            4              |
             3                                |
             3 04              2              |
             2
```

Stem and Leaf : diagramme “branche et feuille”

(branche = partie entière ; feuille = 1ère décimale - les valeurs de “LP” sont arrondies à la 1ère décimale la plus proche pour le graphique).
: effectif dans chaque classe.

```
Boxplot :  +           (moyenne)
           *--------*  (médiane)
```



Graphique “quantile-quantile” (qqplot) :

+ : répartition des observations pour une loi normale (tracé linéaire).

* : répartition des observations du jeu de données (x = quantile “théorique” calculé en fonction du rang de l'observation ; y = valeur observée. Un écart important à la droite théorique (++) signifie que la distribution de la variable ne peut être considérée comme normale.

Les différents paramètres et graphiques édités montrent que la distribution des valeurs observées de “LP” est homogène (pas de valeurs “aberrantes”) et n'est pas trop “éloignée” d'une loi normale.

Édition de la matrice de corrélation

```
proc corr data=tp outp=cor noprint;
var lp--del;
run;
data mcor (drop= _type_);
set cor;
if _type_ = 'CORR';
format lp--del 6.3;
run;
proc print data=mcor noobs;
title 'Matrice de corrélation';
run;
```

Séparation des variétés brévi et longistylées

```
data brevi (drop=del);
set tp;
if del<0 then output;
run;

data longi (drop=del);
set tp;
if del>0 then output;
run;
```


	LP	LCAL	LSEP	LTC	PGL	LTP	LTE	LF	DEL
LP	1.000	0.816	0.341	0.650	0.672	-0.085	0.579	0.446	-0.337
LCAL	0.816	1.000	0.455	0.560	0.657	-0.098	0.569	0.416	-0.344
LSEP	0.341	0.455	1.000	0.495	0.686	0.041	0.280	0.136	-0.115
LTC	0.650	0.560	0.495	1.000	0.794	-0.167	0.539	0.414	-0.358
PGL	0.672	0.657	0.686	0.794	1.000	-0.221	0.649	0.517	-0.445
LTP	-0.085	-0.098	0.041	-0.167	-0.221	1.000	-0.687	-0.781	0.936
LTE	0.579	0.569	0.280	0.539	0.649	-0.687	1.000	0.960	-0.881
LF	0.446	0.416	0.136	0.414	0.517	-0.781	0.960	1.000	-0.915
DEL	-0.337	-0.344	-0.115	-0.358	-0.445	0.936	-0.881	-0.915	1.000

Des corrélations importantes entre variables peuvent être mises en évidence (en particulier sur le groupe de variables "LTP", "LTE", "LF", "DEL" dont la bi-modalité aura été visualisée).

Création du fichier "brevi" (valeur de la variable del négative) à partir du fichier tp ; la variable del est éliminée.

Le fichier "brevi" comporte 34 individus et 8 variables.

Création du fichier "longi" (valeur de la variable del positive) à partir du fichier tp ; la variable del est éliminée.

Le fichier "longi" comporte 20 individus et 8 variables.

ACP sur le groupe des brévistylées
Déroulement de l'analyse, présentation des macro-fonctions SAS

SAS editor :
%acp(brevi,nech,lp–lf,red=);

SAS output :

Valeurs propres, variances expliquées

K	LAMBDA	PCTVAR	CUMPCT
1	4.71	0.59	0.59
2	1.54	0.19	0.78
3	0.63	0.08	0.86
4	0.47	0.06	0.92
5	0.40	0.05	0.97
6	0.11	0.01	0.98
7	0.09	0.01	0.99
8	0.04	0.01	1.00

Macro-fonction : %acp

Réalisation de l'acp sur le fichier de données initial ; création de l'ensemble des fichiers utilisés par les autres macros pour l'interprétation des résultats (voir annexe).

Sortie papier : inertie de chaque axe k (LAMBDA), pourcentage d'inertie expliqué par chaque axe (PCTVAR), pourcentage d'inertie cumulée (CUMPCT).

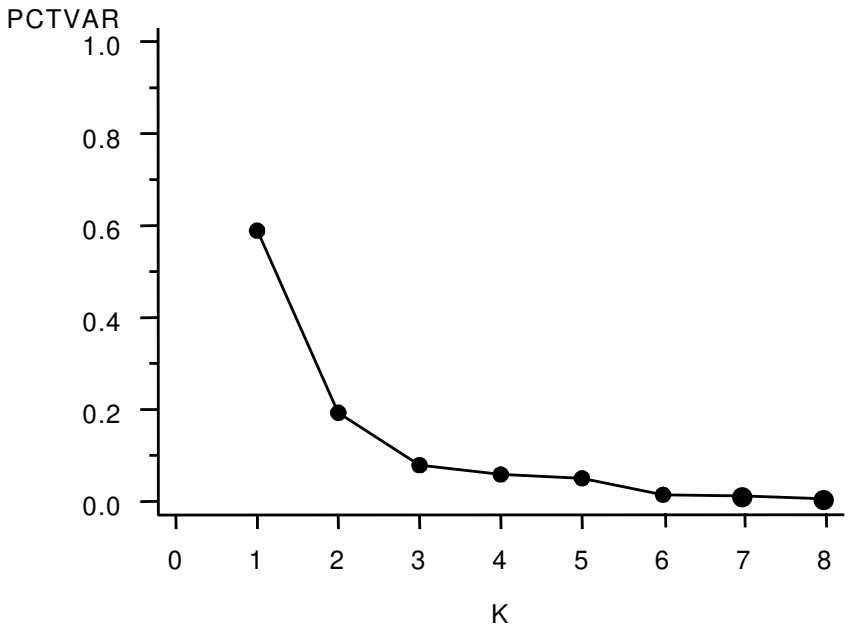
Syntaxe :

%acp(nom du fichier, variable d'identification des individus, liste des variables actives, red=)

Syntaxe de l'argument red :

...,red=) $\Longrightarrow$ ACP sur matrice de corrélation
...,red=cov) $\Longrightarrow$ ACP sur matrice de variance-covariance

SAS editor :
% eboulis(p=8);



Macro-fonction : %eboulis

Tracé de l'éboulis des valeurs propres
Syntaxe : %éboulis(p=nombre d'axes retenus pour le tracé)

Sortie fenêtre graphique.

SAS editor :
%varacp;

SAS output :

```
Coordonnées des variables

_NAME_V1    V2    V3    V4    V5    V6    V7    V8

LP    0.86  0.15 -0.30 -0.28  0.03 -0.23  0.05  0.00
LCAL  0.78  0.17 -0.56  0.11  0.09  0.18 -0.01  0.03
LSEP  0.70 -0.53  0.02  0.43 -0.15 -0.06  0.15  0.01
LTC   0.74 -0.37  0.28 -0.17  0.45  0.07  0.07  0.00
PGL   0.90 -0.32  0.09  0.14  0.02 -0.06 -0.25  0.01
LTP   0.79 -0.25  0.14 -0.35 -0.41  0.12  0.01  0.02
LTE   0.78  0.58  0.15  0.09 -0.03  0.04  0.01 -0.16
LF    0.55  0.76  0.31  0.12 -0.01 -0.02  0.02  0.13

Corrélations  variables x facteurs

_NAME_V1    V2    V3    V4    V5    V6    V7    V8

LP    0.86  0.15 -0.30 -0.28  0.03 -0.23  0.05  0.00
LCAL  0.78  0.17 -0.56  0.11  0.09  0.18 -0.01  0.03
LSEP  0.70 -0.53  0.02  0.43 -0.15 -0.06  0.15  0.01
LTC   0.74 -0.37  0.28 -0.17  0.45  0.07  0.07  0.00
PGL   0.90 -0.32  0.09  0.14  0.02 -0.06 -0.25  0.01
LTP   0.79 -0.25  0.14 -0.35 -0.41  0.12  0.01  0.02
LTE   0.78  0.58  0.15  0.09 -0.03  0.04  0.01 -0.16
LF    0.55  0.76  0.31  0.12 -0.01 -0.02  0.02  0.13
```


Macro-fonction : %varacp

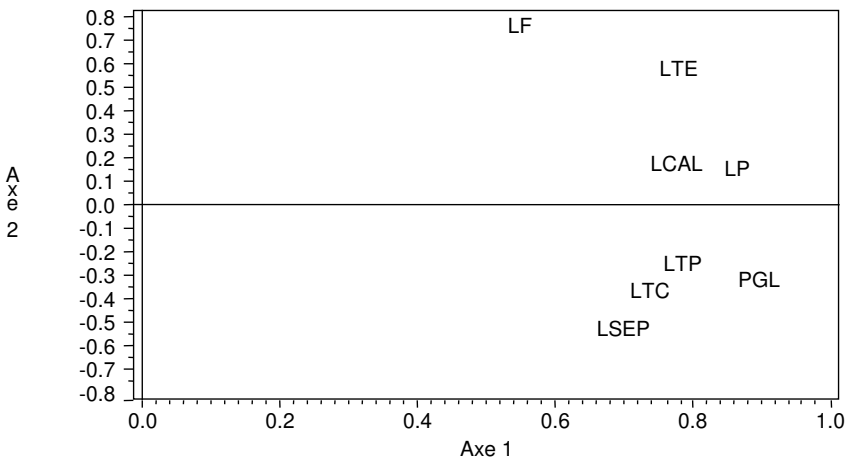
Calcul des coordonnées des variables et des corrélations entre variables et axes principaux.

Création du fichier coordvar.

Sortie papier : coordonnées des variables (V1, ...,VK) et corrélations aux axes principaux (V1, ...,VK). Les coordonnées utilisées pour la représentation graphique sont égales aux $\lambda_s v_j^s$ ($\sim$ tracé du bi-plot isométrique colonne sous S+).

Syntaxe : %varacp;

SAS editor :
%grafvar(x=1,y=2,nc=4)



Macro-fonction : %grafvar

Représentation des variables.

Syntaxe : %grafvar(x = numéro $1er$ axe choisi pour la représentation, y = numéro numéro $2ème$ axe ; nc = nombre maximum de caractères pour l'identification des variables).

Sortie fenêtre graphique (échelles modifiées à l'issue de la première sortie).

Remarquons qu'ici l'échelle est déformée. Avec la macro %cerclec(x=1,y=2,nc=4) on a une représentation non déformée avec le cercle des corrélations.

SAS editor :
%indacp(q=3);

SAS output :

nech	prin1	prin2	prin3	contg	cont1	cont2	cont3	cosca1	cosca2	cosca3
C18	-2.42	0.98	1.02	0.03	3.64	1.85	4.84	0.64	0.11	0.11
C17	-0.97	1.46	-0.35	0.02	0.58	4.07	0.56	0.22	0.51	0.03
C16	-1.02	2.22	-0.27	0.03	0.64	9.43	0.34	0.15	0.72	0.01
C27	0.95	-0.46	0.73	0.01	0.57	0.40	2.50	0.44	0.10	0.26
C36	-1.89	1.06	1.39	0.03	2.23	2.16	9.06	0.43	0.14	0.23
C29	-2.33	0.44	0.37	0.02	3.39	0.38	0.63	0.81	0.03	0.02
C32	-2.47	-0.42	0.49	0.02	3.80	0.34	1.12	0.91	0.03	0.04
C8	-2.02	0.52	0.96	0.02	2.55	0.51	4.35	0.69	0.05	0.16
C37	0.54	0.31	0.43	0.01	0.18	0.19	0.86	0.13	0.04	0.08
C20	-0.70	0.06	0.05	0.01	0.31	0.01	0.01	0.22	0.00	0.00
C28	1.35	0.78	0.03	0.02	1.14	1.17	0.00	0.39	0.13	0.00
C2	5.53	-0.58	1.12	0.12	19.11	0.65	5.84	0.94	0.01	0.04
C3	2.73	0.43	0.53	0.03	4.66	0.35	1.31	0.85	0.02	0.03
C35	1.01	0.56	-0.02	0.01	0.64	0.61	0.00	0.64	0.20	0.00
C24	-0.62	0.22	0.47	0.01	0.24	0.09	1.05	0.28	0.03	0.16
C25	2.00	-1.16	1.01	0.02	2.51	2.56	4.79	0.61	0.20	0.15
C10	0.69	1.96	-1.76	0.04	0.30	7.29	14.57	0.05	0.39	0.31
C11	1.72	0.73	0.67	0.02	1.85	1.02	2.07	0.61	0.11	0.09
C13	2.03	-0.02	-0.90	0.02	2.58	0.00	3.79	0.75	0.00	0.15
C47	1.06	-0.18	-1.36	0.01	0.70	0.06	8.65	0.31	0.01	0.51
C14	1.16	0.02	-0.22	0.01	0.85	0.00	0.22	0.69	0.00	0.02
C4	0.91	-0.04	-0.37	0.01	0.52	0.00	0.63	0.60	0.00	0.10
C5	1.89	0.32	0.34	0.02	2.23	0.20	0.54	0.59	0.02	0.02
C6	-0.19	0.17	-0.84	0.02	0.02	0.06	3.26	0.01	0.01	0.15
C39	0.12	-0.24	-0.12	0.01	0.01	0.11	0.07	0.00	0.02	0.00
C40	0.02	-0.19	-0.87	0.01	0.00	0.07	3.51	0.00	0.01	0.23
C41	-5.61	-2.03	-0.90	0.14	19.68	7.87	3.82	0.86	0.11	0.02
C38	-3.63	-2.71	0.01	0.08	8.25	14.01	0.00	0.64	0.35	0.00
C44	0.44	0.04	-0.07	0.00	0.12	0.00	0.02	0.21	0.00	0.00
C45	-2.16	0.59	0.37	0.02	2.92	0.67	0.65	0.77	0.06	0.02
C55	2.01	-1.81	0.70	0.03	2.52	6.22	2.29	0.49	0.39	0.06
C50	-3.36	-0.60	0.10	0.05	7.06	0.69	0.05	0.91	0.03	0.00
C52	0.73	1.62	-1.68	0.03	0.33	5.02	13.18	0.07	0.36	0.38
C56	2.49	-4.09	-1.07	0.09	3.86	31.92	5.41	0.25	0.67	0.05

Macro-fonction : %indacp

Calcul des coordonnées des individus, de leur contribution générale à l'analyse, de leur contribution et de leur cosinus carré pour chaque axe.

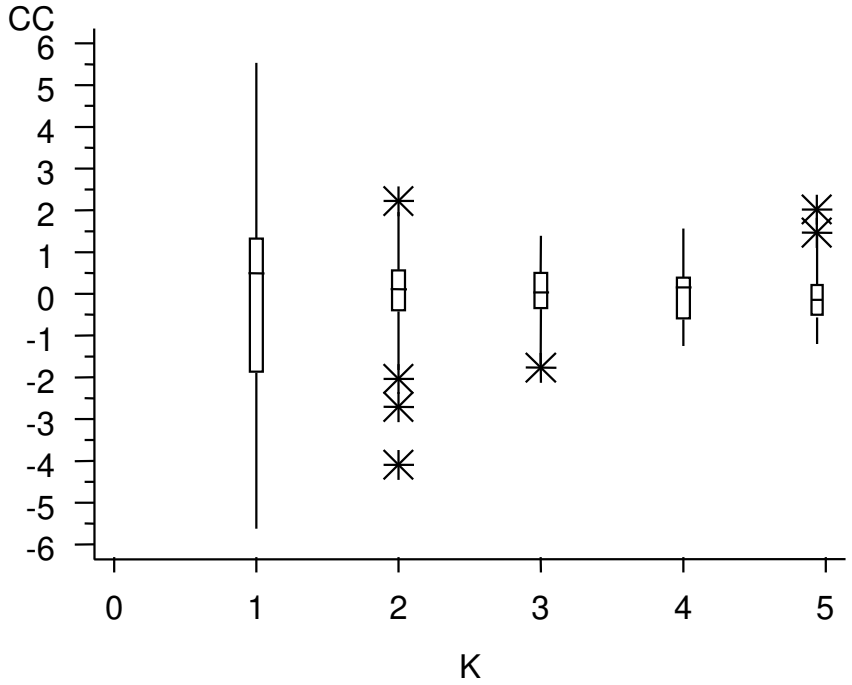
Création du fichier coorindq.

Sortie papier :
coordonnées des individus (prin1, ...),
contribution générale à l'analyse (contg),
contribution à chaque axe (cont1, ...),
cosinus carrés pour chaque axe (cosca1, ...).
Les coordonnées prin1, ..., utilisées pour la représentation graphique sont égales aux $\lambda_s u_s^i$ ($\sim$ tracé du biplot isométrique ligne sous Splus).

Syntaxe : %indacp(q= nombre d'axes retenus pour le calcul);

Ici, tout est en centième sauf la variable contg.

SAS editor :
% moustind(p=5);



Macro-fonction : %moustind

Boxplot des coordonnées des individus sur chaque axe.

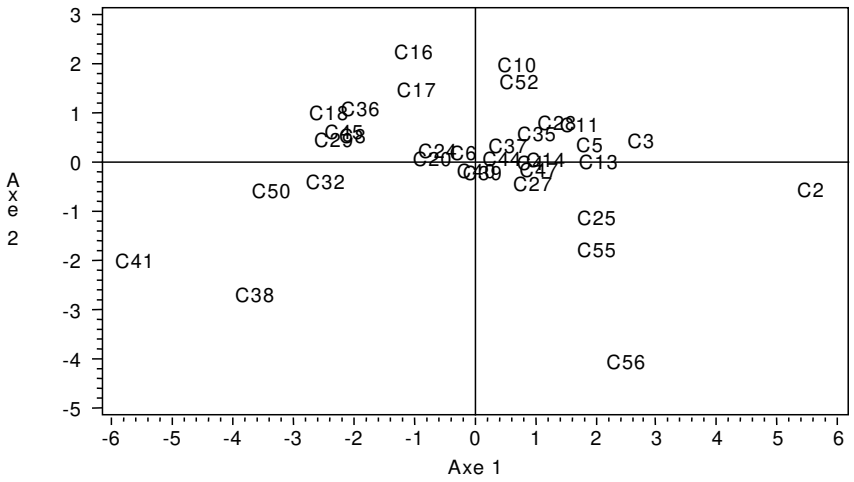
Syntaxe : %moustind(p= nombre d'axes retenus pour le tracé);

Sortie fenêtre graphique.

Remarque : * représente une observation dont la distance à la médiane est supérieure à 1.5 fois l'intervalle interquartile.

INRA ©1998 - Formation statistique - Module 7/8 (ACP TP SAS)-12

SAS editor :
%grafind(x=1,y=2,nc=4)



Macro-fonction : %grafind

Représentation des individus.

Syntaxe : %grafind (x = numéro $1er$ axe choisi pour la représentation, y = numéro numéro $2ème$ axe ; nc = nombre maximum de caractères pour l'identification des variables).

Sortie fenêtre graphique.

Ici aussi le graphique est déformé (pas d'isométrie).

Programme

```
*****************************************************************
* Programme avec envoi des graphiques sur des fichiers .eps (donc *
* la mise en instructions actives des commentaires permettant cet *
* envoi), et changement de nom du fichier .eps d'accueil de la    *
* macro grafvar (qui avait le meme nom que le fichier .eps        *
* d'accueil de la macro grafind).                                 *
*****************************************************************;

%include lecture acpmaze ebouliw varacp grafvarw indacp moustinw grafindw;

%acp(brevi,nech,lp--lf,red=);
%eboulis(p=8);
%varacp;
%grafvar(x=1,y=2,nc=4);
%indacp(q=8); * afin de faire la classif sur les 8 axes;
              * sinon il faut modifier la macro;
%moustind(p=5);
%grafind(x=1,y=2,nc=4);
```


Programme d'ACP préliminaire à la classification

Le programme d'ACP utilisé en entrée de la classification est très légèrement différent de celui utilisé pour l'ACP proprement dite. La principale différence consiste à garder les 8 axes (correspondant aux 8 variables de départ) dans la macro indacp. Si l'on avait gardé %indacp(q=3), on aurait fait la classif sur les 3 premiers axes uniquement.

Une autre différence concerne l'envoi direct des fichier graphiques dans un fichier poscript encapsulé, et non dans la fenêtre graphique de SAS. Ce qui est utile dans 2 cas :

- on n'a pas l'environnement permettant d'accéder à cette fenêtre graphique,
- on veut rapatrier le graphique dans un éditeur (word, latex, ...), une fois qu'on est satisfait du graphique obtenu.

Tous les programmes SAS relatifs à l'ACP sont appelés après le *%include*.

```
%include lecture acpmaze ebouliw varacp grafvarw indacp moustinw grafindw;
```

Il serait revenu au même de faire, au sein du programme SAS :

```
%inc lecture;
%inc acpmaze;
...
```

ou, dans la zone program editor :

```
inc lecture (retour chariot)
inc acpmaze (retour chariot)
...
```

Les macro-programmes sont exécutés tout de suite après :

```
%acp(brevi,nech,lp--lf,red=);
%eboulis(p=8);
%varacp;
%grafvar(x=1,y=2,nc=4);
%indacp(q=8); * afin de faire la classif sur les 8 axes;
              * sinon il faut modifier la macro;
%moustind(p=5);
%grafind(x=1,y=2,nc=4);
```

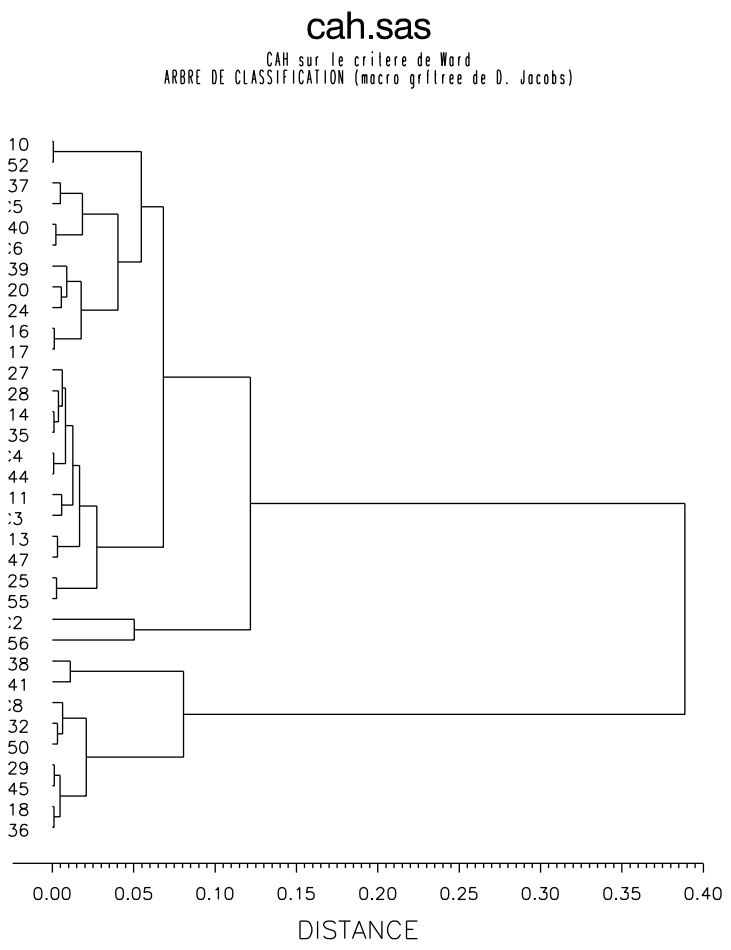
```
title1 "cah.sas";
*****************************************************************
* Ce programme fait  une CAH (critère de Ward).                 *
* Cette classification est faite sur l'ensemble des axes de l'ACP *
* faite sur les brévistylés. Avant de passer ce programme, il faut*
* passer l'ACP (programme acptout.sas).                         *
*****************************************************************;
%include "jaccard+";
/* inclusion du programme permettant de faire des beaux arbres   */
footnote;
proc cluster data=coorindq outtree=treeward method=ward;
  var prin1--prin8;
  id nech;
run;
title2 "CAH sur le critere de Ward";
title3 "ARBRE DE CLASSIFICATION (macro grftree de D. Jacobs)";
filename figure "acpcah.ps";
%grftree(clusdsn=treeward,items=34,axis=d);
run;
quit;

proc tree data=treeward n=5 noprint out=trees;
  id nech;
  copy prin1--prin8;
run;

/* cette étape permet de projeter les individus sur les axes */
data tree2;
  set trees;
clus=put(cluster,2.);   * text ne peut prendre que des chaînes de;
text = substr(clus,2);  * caractères. On traduit donc cluster en clus;
*text=nech  ;        * Cette ligne aurait fait imprimer le numéro des ;
                     * individus au lieu de celui des classes.;
y = prin1;
x = prin2;
xsys = '2';
ysys = '2';
size = 1;
label y = "Dimension 1";
label x = "Dimension 2";
keep x y text xsys ysys size;
run;

filename figure "projcah.ps";
title2 "CAH sur le critère de Ward";
title3 "Projection des individus issus de la CAH sur les 2 premiers axes";
proc gplot data=tree2;
  symbol1 i=none;
  plot y*x=1 / annotate=tree2 frame;
run;
quit;
```


Le programme de classification

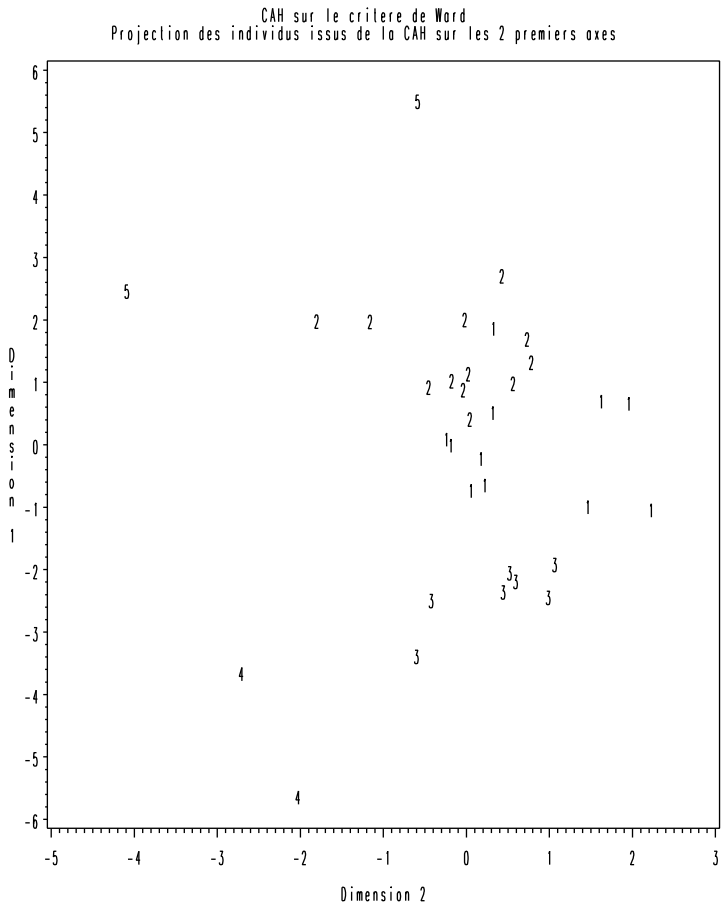
La classification proprement dite est une CAH, critère de Ward.

La macro grftree est utilisée. Sur la version 6.11 de SAS, il est possible de spécifier l'option GRAPHICS dans l'instruction TREE, qui fait quasiment le même graphique.

Contrairement à la sortie S+, SAS indique le nom des individus classifiés, et non le numéro d'observation.


cah.sas
CAH sur le critere de Ward
ARBRE DE CLASSIFICATION (macro grltree de D. Jacobs)
10
52
37
:5
40
:6
39
20
24
16
17
27
28
14
35
:4
44
11
:3
13
47
25
55
:2
56
38
41
:8
32
50
29
45
18
36
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
DISTANCE

cah.sas


CAH sur le critere de Ward
Projection des individus issus de la CAH sur les 2 premiers axes
Dimension 1
Dimension 2

Annexe technique : réalisation de l'ACP (document formateur)

%macro acp(ldataset, lident, llistev, red=);

```
%global dataset ident listev;
%let dataset=&ldataset;
%let ident=&lident;
%let listev=&llistev;
%* Acp de dataset ;
%*          ident : variable contenant les identificateurs;
%*                  des individus;
%*        llistev : liste des variables (numériques);
%*     par defaut : réduites sinon red=cov;
%* options  édition;
options linesize=80 pagesize=62 nonumber nodate;
title "A.c.p. des donnees de &dataset";
footnote;

data donnees;
set &dataset (keep=&ident &listev);
run;
```

proc princomp data=donnees noprint

```
            outstat=eltpr out=compr
          vardef=N &red;
    var &listev;

run;
```


Cette Annexe technique ne fait pas partie du document stagiaire (transparents 1 à 13).

Réalisation de l'acp du fichier données

Création des fichiers :
- *eltpr* contenant moyennes des variables, nombre d'observations, matrice de variance-covariance ou matrice de corrélation, inerties et vecteurs propres.
- *compr* contenant les variables initiales et les scores des individus ($\lambda_s u_i^s$).

```
%* nettoyage des résultats;

data tlambda (drop=_type_)
     tvectp (drop=_type_)
     sigma (drop=_type_)
     stat;
set eltpr;
select (_type_);
when ('EIGENVAL') do;
         _name_ = 'lambda';
         output tlambda;
         end;
when ('CORR','COV')  output sigma;
when ('SCORE')       output tvectp;
otherwise delete;
end;
run;
```

%* vecteur contenant les écarts types;

```
data ectype (keep=std);
set sigma;
array covcor{*} _numeric_;
std=sqrt(covcor{_n_});
run;
```

À partir de *eltpr* création des fichiers :

tlambda : inertie de chaque axe.

sigma : matrice de variance-covariance ou de corrélation.
tvectp : vecteurs propres.

Création du fichier *ectype*

calcul des écart-types pour chaque variable : éléments diagonaux de la matrice de variance-covariance ou de corrélation.

***décomposition de l'inertie**

```
data lambda (keep=k lambda pctvar cumpct);
set tlambda (drop= _name_);
array l{*} _numeric_;
tr=sum(of l{*});
cumpct=0;
do k=1 to dim(l);
lambda=l{k};
pctvar=l{k}/tr;
cumpct=pctvar + cumpct;
output;
end;
run;

proc print noobs round;
title2 'Valeurs propres, variances expliquées';
var k lambda pctvar cumpct;
run;
%mend acp;
```

Création du fichier *lambda*.

Calcul de la trace (inertie totale).

Inertie associée à l'axe i.
Pourcentage d'inertie expliqué par l'axe i.
Pourcentage d'inertie cumulé.

```
%macro eboulis(p=5);
%* Éboulis des valeurs propres;
%* p: nombre de variables ;
*filename gsasfile  "&dataset.qs.eps";
*goptions device=psepsf gend='0a'x  gaccess=gsasfile
vsize=10cm hsize=12cm htext=2.2pct;
proc gplot data=lambda;
title;
footnote;
symbol1 i=join v=dot;
 axis1 order=( 0 to &p by 1) length=7cm; /* attention taille */
 axis2 order=(0 to 1 by 0.2) length=5cm;
plot pctvar*k=1 / haxis=axis1 hminor=0 vaxis=axis2 vminor=1;
run;
quit;
goptions reset=all;
```

Étude des variables

```
%macro varacp;
%* calcul des coordonnées des variables;
data coordvar (drop=i lambda);
set tvectp;
set lambda (keep=lambda);
array coord{*} &listev;
do i = 1 to dim(coord);
     coord{i}=coord{i}*sqrt(lambda);
end;
run;

proc transpose out=coordvar prefix=v;
var _numeric_;
run;

proc print noobs round;
title2 'Coordonnées des variables';
run;
```

Création du fichier *coordvar*.

À partir du fichier des vecteurs propres *tvectp* et du fichier des inerties *lambda* on calcule les $\lambda_s u_i^s$ qui seront utilisées pour les représentations graphiques ($\lambda_s = \sqrt{(inertie)}$).

%* calcul des correlations variables x facteurs;

```
data covarfac (drop=i std);
set  coordvar;
set ectype;
array coord{*} _numeric_;
do i = 1 to dim(coord);
     coord{i}=coord{i}/std;
end;
run;

proc print noobs round;
title2 'Corrélations  variables x facteurs';
var _name_ _numeric_;
run;
```

%mend varacp;

Création du fichier *covarfac*.

Corrélation de la variable i à l'axe j calculée d'après:
cor ij = coord var i sur axe j / écart-type de i.

```
%macro grafvar(x=1,y=2,nc=4);
%* Graphique des variables;
%*    x : numéro axe horizontal;
%*    y : numéro axe vertical;
%*   nc : nombre max de caractères;
*filename gsasfile  "&dataset.v.eps";
*goptions device=psepsf gend='0a'x  gaccess=gsasfile
         vsize=12cm htext=1.8pct;
 data anno;
 retain  xsys ysys '2';
 set coordvar nobs=nind;
 style='swiss';
 y= v&y;
 x= v&x;
 text=substr(_name_,1,&nc);
 size=1;
 label y = "Axe &y"
       x = "Axe &x";
run;

proc gplot data=anno;
 title;
 footnote ;
 axis1  length=14cm; /* attention taille */
 axis2  length=7.8cm;
 symbol1 v='none';
 plot y*x=1 / annotate=anno frame  href=0 vref=0
  haxis=axis1 vaxis=axis2;
run;
goptions reset=all;
quit;
%mend grafvar;
```

Représentation graphique des variables.

Étude des individus

```
%macro indacp(q=3);
%*   q : nombre de composantes retenues;
%* Calculs des contributions et cos carres;
data coorindq;
if _n_ = 1 then set tlambda;
set compr (drop= &listev) nobs=nind;
array c{*} _numeric_;
array cosca{&q};
array cont{&q};
array l{*} &listev;
disto=uss(of c{*})-uss(of l{*});
do i = 1 to &q;
       j=i+dim(l);
       cosca{i}=c{j}*c{j}/disto;
       cont{i}=100*c{j}*c{j}/l{i}/nind;
end;
contg=disto/(sum(of l{*})*nind);
keep &ident prin1-prin&q contg cont1-cont&q cosca1-cosca&q ;
run;

proc print noobs round;
title2 'Coordonnées des individus contributions et cosinus
        carrés';
var &ident prin1-prin&q contg cont1-cont&q cosca1-cosca&q ;
run;
%mend indacp;
```

Création du fichier *coorindq*.

disto = norme du vecteur individu i.

cosca(i) = proportion de la norme de l'individu i projetée sur l'axe j.
cont(i) = contribution de l'individu i à l'axe j.

contg = contribution globale de l'individu i à l'analyse.

```
%macro moustind(p=5);
%* boîtes à moustaches parallèles des composantes principales;
%* p: nombre de variables ;
*filename gsasfile  "&dataset.qb.eps";
*goptions device=psepsf gend='0a'x  gaccess=gsasfile
 vsize=10cm hsize=12cm htext=2pct;

data conccomp (keep=k cc);
set compr (drop= &listev);
array c{*} _numeric_;
do k = 1 to &p;
cc=c{k};
output;end;
run;

proc gplot data=conccomp;
title;
footnote;
 axis1 order=( 0 to &p by 1) length=7cm; /* attention taille */
 axis2 length=5cm;
symbol1 i=box v=star;
plot cc*k=1 / haxis=axis1 hminor=0 vaxis=axis2;
run;
quit;
goptions reset=all;
%mend moustind;
```

```
%macro grafind(x=1,y=2,nc=4);
%*    Graphique des individus;
%*    x : numéro axe horizontal;
%*    y : numéro axe vertical;
%*   nc : nombre max de caractères;
*filename gsasfile  "&dataset.i.eps";
*goptions device=psepsf gend='0a'x  gaccess=gsasfile
         vsize=12cm htext=1.8pct;
 data anno;
 retain  xsys ysys '2';
 set coorindq nobs=nind;
 style='swiss';
 y= prin&y;
 x= prin&x;
 text=substr(&ident,1,&nc);
 size=1;*2-max(0,nind/100)+(cosca&x+cosca&y);
 label y = "Axe &y"
       x = "Axe &x";
run;

proc gplot data=anno;
 title;
 footnote ;
 axis1  length=14cm; /* attention taille */
 axis2  length=7.8cm;
 symbol1 v='none';
 plot y*x=1 / annotate=anno frame  href=0 vref=0
  haxis=axis1 vaxis=axis2;
run;
goptions reset=all;
quit;
%mend grafind;
```

Représentation graphique des individus

La représentation simultanée des variables (bi-plot isométrique ligne) peut être obtenue à partir de fichier *tvectp* (vecteurs propres) initial par :
proc transpose out=tvectp out=vectp prefix=v;run;
et en effectuant %grafvar sur *vectp* (set vectp ...)