

TP: Analyse des correspondances élémentaire

Résumé

Exemple d'analyse l'Analyse factorielle des correspondances élémentaire et comparaison avec l'ACP sur un exemple plus complexe (résultats d'élection).

1 Introduction

Contrairement aux autres techniques factorielles qui ont été développées dans SAS il y a plusieurs années, la procédure `corresp` est plus récente. En conséquence, son usage pour mettre en œuvre une analyse des correspondances simple ou multiple répond bien aux besoins des présentations récentes de ces techniques. Il n'est donc pas indispensable de faire appel à des macros pour en faciliter l'usage, seuls les graphiques ont été "carrossés". Du côté de R, la librairie `FactoMineR` est la plus complète. Le premier exemple traité est une illustration élémentaire de l'AFC tandis que le 2ème a pour objectif d'illustrer le rôle de la métrique du χ^2 en comparant ACP et AFC.

2 CSP et lieux de vacances

Les données sont sous une forme particulière avec trois colonnes. La première contient un effectif conjoint associé aux deux modalités contenues dans les colonnes suivantes. La deuxième colonne contient un code de catégorie socio-professionnelle tandis que la 3ème contient un code correspondant à un type d'hébergement pour les vacances.

2.1 Lecture et description des variables

Charger le fichier `vaccsp.dat`.

La lecture des données se fait par le programme :

```
data sasuser.vaccsp;
```

```
infile "vaccsp.dat" ;
input eff csp $ heber $;
run;
```

identifier les différentes modalités et représenter leurs distributions avec `sas/insight`. Tracer également les profils lignes ou colonnes. Commentaire sur ces profils. Attention, ne pas oublier de déclarer la variable `eff` comme une fréquence.

Calculer le test du χ^2 d'indépendance entre les variables. Que dire de la liaison entre celles-ci ?

```
proc freq data=sasuser.vaccsp;
tables heber*csp /chisq;
weight eff;
run;
```

2.2 AFC avec SAS

L'AFC est obtenue avec la gestion ODS des sorties graphiques par :

```
ods html;
ods graphics on;
proc corresp data=sasuser.vaccsp observed
out=resul dim=3;
tables csp,heber; /* croisement des deux variables */
weight eff; /* contient les effectifs de la table */
run;
ods graphics off;
ods html close;
```

Les graphiques fournis par le système ODS ne sont pas toujours très explicites. Une autre façon de faire : ouvrir la table `work.resul` dans `SAS/INSIGHT`, déclarer la variable `NUM` en label, et construire les graphiques : distributions des composantes principales `dim1` et `dim2`, nuage des modalités dans ces axes.

Les graphiques sont également construits par la macro `sas` :

```
%gafcx (ident=num);
```

produit également la représentation graphique.

2.3 AFC avec R

Il est possible d'exporter les données de SAS puis de les relire avec R mais aussi de lire directement le fichier `vaccsp.dat` dans R. La fonction `table` construit bien une table de contingence mais sans tenir compte d'une variable d'effectif. C'est la fonction `xtabs` qui permet de construire la bonne table de contingence utilisée par la librairie `FactoMineR`.

```
vaccsp=read.table("vaccsp.dat",
  col.names=c("effectif", "CSP", "heber"))
summary(vaccsp)
library(FactoMineR)
contin=xtabs(effectif~CSP+heber,vaccsp)
contin
res.afc=CA(contin)
summary(res.afc)
```

Interprétation des graphiques obtenus : positions relatives des modalités hébergement puis des modalités CSP et enfin des modalités des deux variables.

3 les élections présidentielles de 95

3.1 Les données

Pour le premier tour des élections présidentielles de 1995, on connaît, pour chacun des 95 départements métropolitains et la Corse¹ les informations suivantes :

- le nombre d'inscrits,
- le nombre de votants,
- le nombre de suffrages exprimés,
- et, dans l'ordre, le nombre de voix des candidats Villiers, Le Pen, Chirac, Laguillier, Chem., Jospin, Voynet, Balladur, Hue.

Charger le fichier `elec95.dat` puis lire les données et calculer le nombre des abstentions, des votes blancs ou nuls.

1. Les résultats des deux départements de Corse sont agrégés.

```
data sasuser.elec95;
infile "elec95.dat" dlm='09'x;
input num $ inscrits votants exprimes villiers
      le_pen chirac laguill cheminad jospin
      voynet balladur hue poids;
abstent=inscrits-votants;
blancs=votants-exprimes; run;
```

3.2 AFC simple

Exécuter le programme suivant :

```
proc corresp data=sasuser.elec95 observed
  out=resul dim=3;
var abstent blancs villiers le_pen chirac laguill
  cheminad jospin voynet balladur hue;
id num;
run;
%gafcx(num);
```

Commentaire à propos du vote "Villiers" en Vendée. Ce département est mis en "supplémentaire" dans l'analyse suivante en modifiant les pondérations :

```
proc corresp data=sasuser.elec95 observed
  out=resul dim=3;
var abstent blancs villiers le_pen chirac laguill
  cheminad jospin voynet balladur hue;
id num;
weight poids; run;
%gafcx(num);
```

Choisir le nombre d'axes, tenter de les interpréter. Que dire des problèmes posés par cette représentation graphique, que fait ressortir en priorité l'AFC ?

3.3 Comparaison avec l'ACP

Faut-il considérer des nombres de suffrages exprimés ou des taux de suffrage exprimés dans l'ACP ? Pourquoi ? Commenter le rôle de la métrique du χ^2 dans le cadre de l'AFC en comparant les résultats avec ceux de l'ACP des taux de

suffrage exprimés pour chacun des candidats. Interpréter les résultats.

```
data sasuser.telec95;
set sasuser.elec95;
villiers=villiers/inscrits;
le_pen=le_pen/inscrits;
chirac=chirac/inscrits;
laguill=laguill/inscrits;
cheminad=cheminad/inscrits;
jospin=jospin/inscrits;
voynet=voynet/inscrits;
balladur=balladur/inscrits;
hue=hue/inscrits; run;
%acp(telec95,num,villiers--hue);
%gacpbx(p=9); %gacpsx(p=9); %gacpix; %gacpvx;
```

En conclusion de cette comparaison ACP/AFC, que dire du rôle de la métrique du chi-deux ?