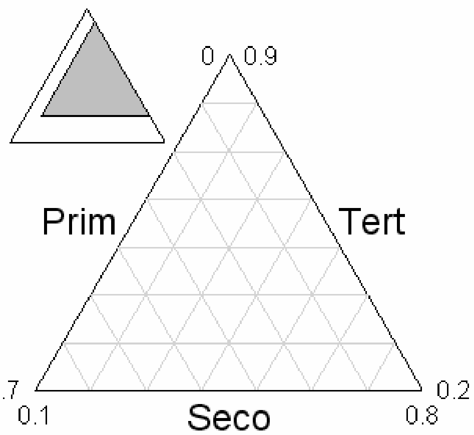
ISFA 2° ANNEE 2004-2005 - Analyse des données

Tout document autorisé – Sans ordinateur

*Les blocs de questions sont en grande partie indépendants mais de difficulté variée. **1** est une question simple de statistique pratique. **2-8** est une application directe des définitions de base. **9** et **10** sont des théorèmes à démontrer. **11** concerne l'interprétation. **12** à **17** concernent la théorie et la pratique de l'analyse en composantes principales. Merci d'utiliser l'espace imparti pour la réponse. Il vous suffira de donner une réponse précise à quelques questions que vous aurez choisies.*

1. Compléter, avec l'information numérique du tableau de gauche, le graphique de droite :



Tissu économique

Secteurs primaire, secondaire, tertiaire

	% POPULATION ACTIVE
AGRICULTURE (primaire)	67% (France 5,1%)
MINES (primaire)	2% (France 0,8%)
INDUSTRIE (secondaire)	10% (France 27%)
SERVICES (tertiaire)	21% (France 67,1%)

http://www.aquicom.org/ASIE/VIETMAN/travail.htm

2. $\mathbb{R}^2$ est muni du produit scalaire canonique $\langle \mathbf{x} | \mathbf{y} \rangle = \mathbf{x}^t \mathbf{y}$. Les colonnes de la matrice $\mathbf{A}$ définissent les coordonnées de 7 points de $\mathbb{R}^2$:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 2 & 3 & 3 & 4 \\ 1 & 2 & 4 & 3 & 2 & 4 & 5 \end{bmatrix}$$

On en fait un data frame de R :

```
w
   x y
M1 0 1
M2 1 2
M3 1 4
M4 2 3
M5 3 2
M6 3 4
M7 4 5
```

Donner (avec des indications de calcul minimales) :

- les coordonnées du centre de gravité du nuage ainsi défini (pondération uniforme) ;
- le tableau centré, qu'on notera $\mathbf{X}_0$;
- la matrice des variances-covariances (en $1/n$) des variables x et y, notée $\mathbf{C}$.

3. (suite) Calculer :

- l'inertie de ce nuage autour du centre de gravité ;

- les axes principaux du nuage centré .

4. (suite) Donner :

- les valeurs propres de la matrice $\mathbf{X}_0^t\mathbf{X}_0$;
- les valeurs propres de la matrice $\mathbf{X}_0\mathbf{X}_0^t$;
- les valeurs singulières du tableau centré $\mathbf{X}_0$;

5. (suite) Donner :

- la pente de la droite de régression de y prédite par x ;

6. (suite) Reporter graphiquement dans la fenêtre proposée l'information numérique précédente.

7. (suite) Soit $\mathbf{u} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ un vecteur unitaire de $\mathbb{R}^2$ et $\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Donner un sens géométrique aux applications linéaires dont les matrices sont respectivement :

$$\mathbf{P} = \mathbf{u}\mathbf{u}^t,\ \mathbf{Q} = 2\mathbf{P} - \mathbf{I}_2,\ \mathbf{R} = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix},\ \mathbf{X}_0\mathbf{Q} \text{ et } \mathbf{X}_0\mathbf{R}$$

8. (suite et fin) Donner un vecteur **u** pour lequel $\mathbf{X}_0\mathbf{Q}$ est le tableau $\mathbf{X}_0$ à une permutation des lignes près.

9. En toute généralité, soit $\mathbf{X}_0$ un tableau centré à n lignes et p colonnes. Soit **R** une rotation de $\mathbb{R}^p$, c'est-à-dire une application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^p$ qui vérifie $\mathbf{R}\mathbf{R}^t = \mathbf{R}^t\mathbf{R} = \mathbf{I}_p$. Montrer que, si **u** est axe principal du nuage des lignes de $\mathbf{X}_0$, **Ru** est axe principal du nuage des points images définis par la rotation.

10. En toute généralité, soit $\mathbf{X}_0$ un tableau centré à n lignes et p colonnes. Soit **u** un vecteur unitaire et **S** la symétrie droite de $\mathbb{R}^p$ par rapport à **u**, c'est-à-dire l'application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^p$ qui vérifie $\mathbf{S}^t = \mathbf{S}$, $\mathbf{S}^2 = \mathbf{I}_p$ et $\mathbf{Su} = \mathbf{u}$. Le nuage des points lignes de $\mathbf{X}_0$ est dit symétrique par rapport à **u** si $\mathbf{X}_0\mathbf{S}$ est le tableau $\mathbf{X}_0$ à une permutation des lignes près. Montrer que **u** est, dans ce cas, axe principal du nuage des points lignes de $\mathbf{X}_0$.

Problème : le lectorat de la presse d'information générale

Les données du problème sont proposées par Pierre-Alain Boscher (ISFA 2°A, Lyon, 2000) à partir d'une enquête de l'INSEE. L'introduction de P.-A. Boscher est soignée (il manque seulement la référence exacte) et est reproduite intégralement :

Les données de l'étude proviennent d'une enquête de l'INSEE d'octobre 1999. 5685 ménages ont répondu a cette enquête. Dans chaque ménage, un individu de 15 ans ou plus tiré au sort a été interrogé. Les questions posées sont :

« Combien de fois en moyenne avez-vous lu au cours des douze derniers mois: un quotidien national (code **`nat`**), *un quotidien régional (code* **`quo`**) *... un magazine ou un journal d'information générale (hors hebdomadaire télé ou magazine spécialisé, code* **`mag`**) *»*

Le niveau de lecture a été divisé en 3. On distingue les non lecteurs (Jamais, code **`0`**), *les lecteurs occasionnels (Moins de 2 fois par semaine, code* **`1`**) *et les lecteurs assidus (Au moins 2 fois par semaine, code* **`2`**).

De plus 3 critères ont été retenus afin de caractériser chaque lecteurs.

Le niveau d'étude :

`Dip0` *Aucun diplôme*
`Dip1` *CEP ou BEPC*
`Dip2` *CAP, BEP*
`Dip3` *Baccalauréat*
`Dip4` *Bac+2*
`Dip5` *Supérieur à bac+2*

La catégorie socioprofessionnelle :

`agri` *Agriculteurs*
`arti` *Artisans, commerçants*
`cadr` *Cadres*
`inte` *Professions intermédiaires*
`empl` *Employés*
`ouvr` *Ouvriers*
`etud` *Scolaires et étudiants*
`inac` *Inactifs (personnes n'ayant jamais travaillé. Les retraités, les chomeurs et les femmes au foyer sont classés selon leur ancienne profession.*

La zone géographique

`RegP` *Région parisienne (lie de France)*
`BasP` *Bassin parisien (Basse-Normandie, Bourgogne, Centre, Champagne Ardenne, Haute-Normandie, Picardie)*
`Nord` *Nord (Nord-Pas-de-Calais)*
`Est` *Est (Alsace, Franche-Comté, Lorraine)*
`Ouest` *Ouest (Bretagne, Pays de la Loire)*
`Su_Ou` *Sud-Ouest (Aquitaine, Limousin, Midi-Pyrénées)*
`Ce_Est` *Centre-Est (Auvergne, Rhône-Alpes)*
`Medi` *Méditerranée (Corse, Languedoc-Roussillon, Provence-Alpes-Côte d'Azur)*

Les données sont dans un data frame **`lecto`** :

	nato	nat1	nat2	quo0	quo1	quo2	mag0	mag1	mag2
dip0	87	8	5	38	23	39	68	12	20
dip1	79	13	8	37	20	43	56	18	26
dip2	77	14	9	29	27	44	52	21	27
dip3	63	24	13	38	28	34	41	27	32
dip4	56	33	11	39	29	32	32	33	35
dip5	38	32	30	52	25	23	26	28	46
agri	93	5	2	17	20	63	61	12	27
arti	71	19	10	32	19	49	50	18	32
cadr	42	29	29	46	23	31	25	27	48
inte	62	23	15	37	27	36	39	27	34
empl	78	14	8	37	24	39	54	19	27
ouvr	85	10	5	31	26	43	63	16	21
etud	68	24	8	47	30	23	48	26	26
inac	77	14	9	46	23	31	63	10	27
RegP	51	21	28	73	18	9 *	43	25	32
BasP	76	17	7	30	29	41	56	19	25
Nord	83	12	5	24	30	46	65	14	21
Est	79	15	6	22	20	58	51	20	29
Ouest	82	14	4	21	21	58	54	20	26
Su_Ou	74	17	9	31	28	41	45	24	31
Ce_Est	75	20	5	31	32	37	49	20	31
Medi	75	17	8	38	24	38	48	17	35

* En région parisienne (**RegP**), 73% des personnes interrogées ne lisent pas de quotidiens régionaux (**quo0**), 18% sont des lecteurs occasionnels (**quo1**) et 9% sont des lecteurs assidus (**quo2**).

En fait, cette présentation simplifie la réalité. Nous pouvons dire que les données sont une liste de 9 tableaux appelée **lili** :

```
names(lili)
[1] "csp-mag" "csp-nat" "csp-quo" "dip-mag" "dip-nat" "dip-quo" "reg-mag"
[8] "reg-nat" "reg-quo"
```

Les composantes de **lili** sont :

$"csp-mag"

	mag0	mag1	mag2
agri	61	12	27
arti	50	18	32
cadr	25	27	48
inte	39	27	34
empl	54	19	27
ouvr	63	16	21
etud	48	26	26
inac	63	10	27

$"csp-nat"

	nato	nat1	nat2
agri	93	5	2
arti	71	19	10
cadr	42	29	29
inte	62	23	15
empl	78	14	8
ouvr	85	10	5
etud	68	24	8
inac	77	14	9

$"csp-quo"

	quo0	quo1	quo2
agri	17	20	63
arti	32	19	49
cadr	46	23	31
inte	37	27	36
empl	37	24	39
ouvr	31	26	43
etud	47	30	23
inac	46	23	31

$"dip-mag"

	mag0	mag1	mag2
dip0	68	12	20
dip1	56	18	26
dip2	52	21	27
dip3	41	27	32
dip4	32	33	35
dip5	26	28	46

$"dip-nat"

	nato	nat1	nat2
dip0	87	8	5
dip1	79	13	8
dip2	77	14	9
dip3	63	24	13
dip4	56	33	11
dip5	38	32	30

$"dip-quo"

	quo0	quo1	quo2
dip0	38	23	39
dip1	37	20	43
dip2	29	27	44
dip3	38	28	34
dip4	39	29	32
dip5	52	25	23

$"reg-mag"

	mag0	mag1	mag2
RegP	43	25	32
BasP	56	19	25
Nord	65	14	21
Est	51	20	29
Ouest	54	20	26
Su_Ou	45	24	31
Ce_Est	49	20	31
Medi	48	17	35

$"reg-nat"

	nato	nat1	nat2
RegP	51	21	28
BasP	76	17	7
Nord	83	12	5
Est	79	15	6
Ouest	82	14	4
Su_Ou	74	17	9
Ce_Est	75	20	5
Medi	75	17	8

$"reg-quo"

	quo0	quo1	quo2
RegP	73	18	9
BasP	30	29	41
Nord	24	30	46
Est	22	20	58
Ouest	21	21	58
Su_Ou	31	28	41
Ce_Est	31	32	37
Medi	38	24	38

11. Rajouter des étiquettes à la figure et rédiger une légende pour la figure proposée.

On ne s'intéressera par la suite qu'à la catégorie diplôme.

```
diplo=lecto[1:6,]/100
diplo
     nato nat1 nat2 quo0 quo1 quo2 mag0 mag1 mag2
dip0 0.87 0.08 0.05 0.38 0.23 0.39 0.68 0.12 0.20
dip1 0.79 0.13 0.08 0.37 0.20 0.43 0.56 0.18 0.26
dip2 0.77 0.14 0.09 0.29 0.27 0.44 0.52 0.21 0.27
dip3 0.63 0.24 0.13 0.38 0.28 0.34 0.41 0.27 0.32
dip4 0.56 0.33 0.11 0.39 0.29 0.32 0.32 0.33 0.35
dip5 0.38 0.32 0.30 0.52 0.25 0.23 0.26 0.28 0.46
```

12. Expliquer pourquoi une fonction accepte et pas l'autre ?

```
princomp(diplo,scale=F)
Error in princomp.default(diplo, scale = F) :
        princomp can only be used with more units than variables
prcomp(diplo,scale=F)
Standard deviations:
[1] 3.053238e-01 8.641854e-02 4.132374e-02 2.522190e-02 4.327816e-03
[6] 7.648296e-17
```

13. Quel est la rang du tableau `diplo` centré ?

```
prcomp(diplo,scale=F)
...

Rotation:
            PC1        PC2         PC3         PC4         PC5         PC6
nato -0.5884209 -0.1311990  0.09372318 -0.08890651 -0.20131668 -0.37406841
nat1  0.3308054 -0.2886674  0.39242355 -0.15664127  0.47405998  0.04323299
nat2  0.2576154  0.4198664 -0.48614673  0.24554778 -0.27274330  0.02374076
quo0  0.1808952  0.5238444  0.43722730 -0.33059085 -0.20707651  0.38176763
quo1  0.0534675 -0.2607014  0.08491688  0.75248821  0.05931486  0.37621339
quo2 -0.2343627 -0.2631430 -0.52214418 -0.42189736  0.14776165  0.59644415
mag0 -0.5064457  0.3321357  0.18537662  0.18906926  0.22840939  0.09303459
mag1  0.2163156 -0.4363237  0.10520207 -0.10223032 -0.62108787 -0.05614384
mag2  0.2901301  0.1041880 -0.29057869 -0.08683894  0.39267848 -0.45033752
```

14. Observer la matrice des axes principaux. Chacun d'entre eux a une propriété très particulière. Laquelle ? Expliquer pourquoi ?

```
pr1=prcomp(diplo,scale=F)
names(pr1)
[1] "sdev"     "rotation" "center"   "scale"    "x"
dudi1=dudi.pca(diplo,scale=F,scan=F)
names(dudi1)
 [1] "tab"  "cw"   "lw"   "eig"  "rank" "nf"   "c1"   "l1"   "co"   "li"
[11] "call" "cent" "norm"
```

15. Comment retrouve-t-on la composante `sdev` de `pr1` dans `dudi1` ?

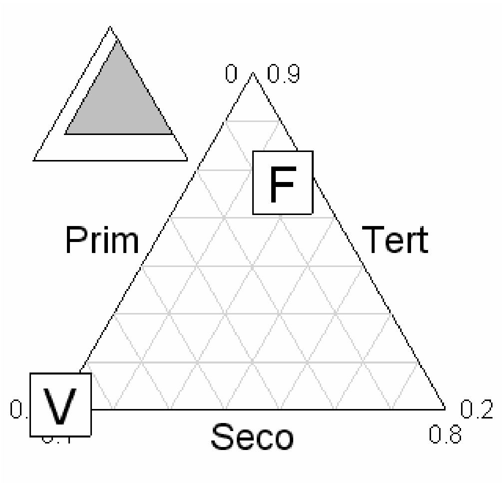
16. La composante `x` de `pr1` est-elle égale à la composante `co` de `dudi1` ?

```
provi=dudi.pca(diplo,scale=F)
Select the number of axes: 2
s.arrow(provi$c1,xlim=c(-0.8,0.6))
s.label(provi$li,add.p=T,clab=0.75)
```

17. Rédiger une légende pour la figure proposée.

ISFA 2° année 2004-2005 Analyse des données

1. Compléter, avec l'information numérique du tableau de gauche, le graphique de droite :



2. *Tous les calculs se font à la main !*

```
apply(w,2,mean)
x y
2 3

X0=t(t(w)-m0)
X0
    x  y
M1 -2 -2
M2 -1 -1
M3 -1  1
M4  0  0
M5  1 -1
M6  1  1
M7  2  2


C=t(X0)%*%X0/7
C
        x     y
  x 1.714 1.143
  y 1.143 1.714

En fait les variances valent 12/7 et la covariance vaut 8/7
```

3. (suite) Calculer :

- l'inertie de ce nuage autour de ce centre de gravité ;

```
m0=apply(w,2,mean)
apply(t(w)-m0,1,function(x) mean(x*x))
    x     y
1.714 1.714

La somme des variances est la réponse. En fait, la réponse est 24/7
```

- les axes principaux du nuage centré ;

```
$values
[1] 2.8571 0.5714

$vectors
       [,1]    [,2]
```

```
[1,] 0.7071  0.7071
[2,] 0.7071 -0.7071
```

```
En fait les valeurs propres valent 20/7 et 4/7. Les axes principaux sont les vecteurs
unitaires portés par les diagonales (y=x et y=-x).
```

4. (suite) Donner :

- les valeurs propres de la matrice $\mathbf{X}_0^t\mathbf{X}_0$: 20 et 4, car $\mathbf{X}_0^t\mathbf{X}_0 = n\mathbf{C}$
- les valeurs propres de la matrice $\mathbf{X}_0\mathbf{X}_0^t$: 20, 4 et 0 à l'ordre de multiplicité 5 (dimension du noyau)
- les valeurs singulières du tableau centré $\mathbf{X}_0$: les racines des valeurs propres

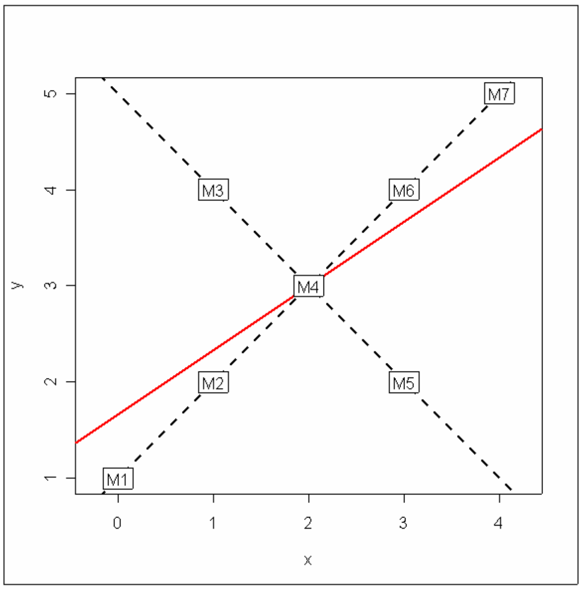
```
svd(X0)$d
[1] 4.472136 2.000000
sqrt(eigen(t(X0)%*%X0)$values)
[1] 4.472136 2.000000
```

5. (suite) Donner :

- la pente de la droite de régression de y prédite par x ;

```
La réponse est 2/3

lm(w$y~w$x)

Coefficients:
(Intercept)          w$x
     1.6667       0.6667

C[1,2]/C[1,1]
[1] 0.6666667

lm(X0[,2]~X0[,1])

Coefficients:
(Intercept)      X0[, 1]
 -2.046e-17    6.667e-01
```

6. Reporter graphiquement dans la fenêtre ci-dessous l'information numérique précédente.



7.

P est le projecteur orthogonal sur **u**. **Q** est la symétrie par rapport à la droite *D* qui porte **u**. **R** est la rotation autour de l'origine d'angle l'angle polaire de **u**. C'est aussi la matrice du changement de base de la base canonique dans une base orthonormale contenant **u**. $\mathbf{X}_0\mathbf{Q}$ est le tableau centré où chaque point ligne est remplacé par son symétrique par rapport à *D*. $\mathbf{X}_0\mathbf{R}$ est le tableau dont chaque ligne est l'image par la rotation du point ligne du tableau centré ou le tableau des points lignes du tableau centré dans la nouvelle base.

Remarque: les deux interprétations de la matrice $\mathbf{X}_0\mathbf{R}$ justifient les noms des objets dans les différentes fonctions :

```
princomp(w,scale=F)$loadings

Loadings:
  Comp.1 Comp.2
x  0.707  0.707
y  0.707 -0.707

prcomp(w,scale=F)$rotation
        PC1        PC2
x 0.7071068 -0.7071068
y 0.7071068  0.7071068
```

8. Donner un vecteur **u** pour lequel $\mathbf{X}_0\mathbf{Q}$ est le tableau $\mathbf{X}_0$ à une permutation des lignes près.

$\mathbf{u} = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}$ convient car la symétrie par rapport à la première diagonale conserve tous les points sauf M_3 et M_5 qui sont échangés.

9.

En effet, si **u** est axe principal, c'est un vecteur propre normé de la matrice de covariance ou plus simplement $\mathbf{X}_0^t\mathbf{X}_0\mathbf{u} = \lambda\mathbf{u}$. Le nuage après rotation est dans la tableau $\mathbf{Y}_0 = \left(\mathbf{R}\mathbf{X}_0^t\right)^t = \mathbf{X}_0\mathbf{R}^t = \mathbf{X}_0\mathbf{R}^{-1}$. Il suffit d'observer que **Ru** est normé (les rotations conservent les longueurs) et que :

$$\mathbf{Y}_0^t\mathbf{Y}_0\mathbf{R}\mathbf{u} = \mathbf{R}\mathbf{X}_0\mathbf{X}_0\mathbf{R}^{-1}\mathbf{R}\mathbf{u} = \mathbf{R}\mathbf{X}_0\mathbf{X}_0\mathbf{u} = \lambda\mathbf{R}\mathbf{u}$$

Les axes principaux tournent avec les nuages et sont caractéristiques de leur géométrie.

10.

En effet, puisque les lignes sont simplement permutées, on a :

$$\mathbf{X}_0^t\mathbf{X}_0 = \left(\mathbf{X}_0\mathbf{S}\right)^t\mathbf{X}_0\mathbf{S} \Rightarrow \mathbf{X}_0^t\mathbf{X}_0\mathbf{u} = \mathbf{S}^t\mathbf{X}_0^t\mathbf{X}_0\mathbf{S}\mathbf{u} = \mathbf{S}\mathbf{X}_0^t\mathbf{X}_0\mathbf{u}$$

Or, $\mathbf{S}\mathbf{z} = \mathbf{z} \Rightarrow \mathbf{S}^2\mathbf{z} = \mathbf{S}\mathbf{z} = \mathbf{z} \Rightarrow \mathbf{z} = \lambda\mathbf{u}$

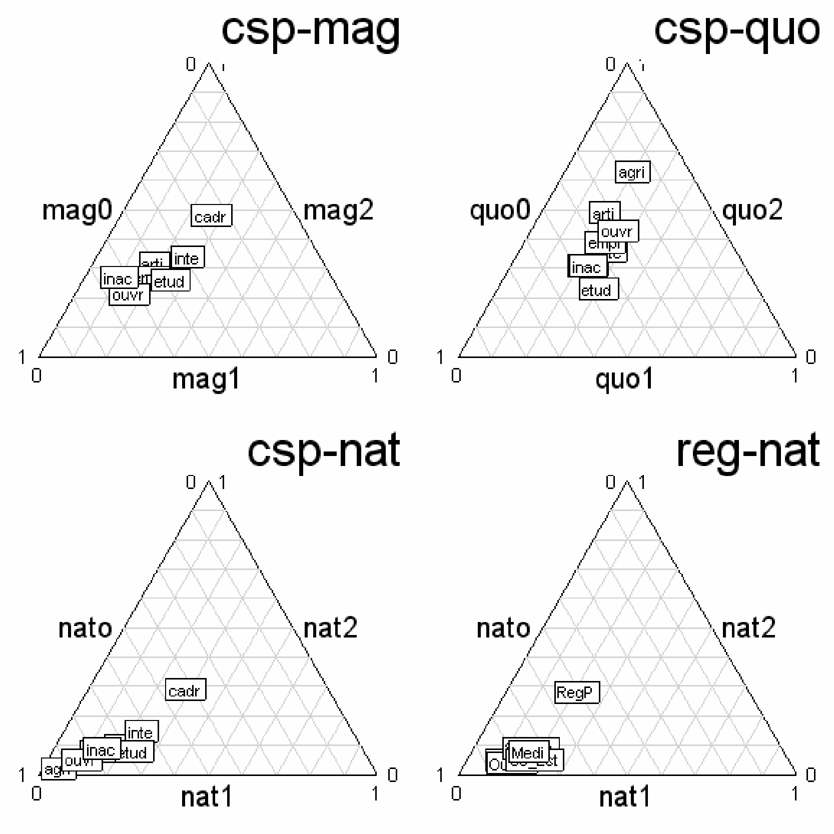
Donc $\mathbf{S}\mathbf{X}_0^t\mathbf{X}_0\mathbf{u} = \lambda\mathbf{u}$

Le vecteur est donc un vecteur propre de la matrice de covariances donc un axe principal. Les axes de symétrie sont toujours des axes principaux.

11.

Exemple de commentaire : Représentation triangulaire de 9 tableaux à trois colonnes. Pour les trois types de presse et les trois modes de répartition des lecteurs, les variations de profils sont importantes. Si on s'intéresse aux catégories socioprofessionnelles, on note que les cadres et les agriculteurs sont deux profils nettement différents des autres. Les premiers consomment nettement plus les magazines et la presse nationale, les seconds ont une lecture occasionnelle presque systématique des quotidiens régionaux. La presse nationale est de plus un phénomène parisien, dans la mesure où elle joue à Paris un rôle de presse locale. On peut choisir de mettre en évidence d'autres éléments.

Exemple d'étiquettes



12. Expliquer pourquoi une fonction accepte et pas l'autre.

princomp adopte le point de vue de la loi normale : l'acp est l'estimation des éléments principaux de la matrice de covariance qui n'a de sens que dans le cas où le nombre d'échantillons (lignes) est bien plus grand que le nombre de variables (colonnes). prcomp accepte le modèle géométrique de Pearson (recherche des lignes et plans les plus proches des données) sans contraintes de distributions de probabilité.

13. Quel est la rang du tableau **diplo** centré.

Il vaut 5 (la dernière valeur propre est nulle). Le rang du tableau est le même que celui de la matrice de covariances. Il ne peut dépasser 6 (6 lignes et 9 colonnes). Comme il est centré la somme des lignes est nulles et le rang ne peut dépasser 5. L'affichage numérique garantit qu'il vaut 5 exactement.

14. Observer la matrice des axes principaux. Chacun d'entre eux a une propriété très particulière. Laquelle ? Expliquer pourquoi.

```
provi=prcomp(diplo,scale=F)$rotation
apply(provi[1:3,],2,sum)
          PC1           PC2           PC3           PC4           PC5           PC6
-1.665335e-16 -4.163336e-16 -1.387779e-17 -1.831868e-15  2.803313e-15 -3.070947e-01
apply(provi[4:6,],2,sum)
          PC1           PC2           PC3           PC4           PC5           PC6
 0.000000e+00  5.551115e-17  2.081668e-16  8.326673e-16 -3.871903e-15  1.354425e+00
apply(provi[7:9,],2,sum)
          PC1           PC2           PC3           PC4           PC5           PC6
 2.775558e-17  2.636780e-16 -1.526557e-16 -3.469447e-16  1.332268e-15 -4.134468e-01
```

Les sommes des composantes par blocs sont nulles. Cela vient du fait que les sommes par blocs sur une même ligne du tableau valent 1 (données en fréquences), les sommes par lignes et par blocs du tableau centré valent 0, les trois vecteurs :

$$\begin{pmatrix}1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

$$\begin{pmatrix}0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0\end{pmatrix}$$

$$\begin{pmatrix}0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1\end{pmatrix}$$

vérifient donc $$\mathbf{X}_0\mathbf{u} = \mathbf{0} \Rightarrow \frac{1}{n}\mathbf{X}_0^t\mathbf{X}_0\mathbf{u} = \mathbf{C}\mathbf{u} = \mathbf{0}$$

Ils sont axes principaux (pour la valeur propre nulle), les autres leur sont orthogonaux, d'où la propriété observée.

15. Comment retrouve-t-on la composante **sdev** de **pr1** dans **dudi1** ?

sdev contient les valeurs singulières, on retrouvera leur carré dans **dudi1$eig**, avec un écart du aux variances en 1/*n* et 1/(*n*-1).

```
pr1$sdev^2
[1] 9.322264e-02 7.468164e-03 1.707651e-03 6.361441e-04 1.872999e-05
[6] 5.849644e-33
dudi1$eig
[1] 7.768554e-02 6.223470e-03 1.423043e-03 5.301201e-04 1.560833e-05
dudi1$eig*6/5
[1] 9.322264e-02 7.468164e-03 1.707651e-03 6.361441e-04 1.872999e-05
```

16. La composante **x** de **pr1** est-elle égale à la composante **co** de **dudi1** ?

Pas du tout, **pr1$x** contient les scores qui sont dans **dudi1$li**.

17. Rédiger une légende pour la figure obtenue

Biplot dans $\mathbb{R}^9$. Le plan est celui des deux premiers axes principaux. Les flèches sont les projections des vecteurs de la base canonique sur ce plan et les points sont les projections des lignes du tableau centré sur ce plan. Les magazines et la presse nationale ont une logique commune : les scores associent nat0-mag0 (à gauche), nat1-mag1 (à droite et en bas) et nat2-mag2 (à droite et en haut). Les catégories quo sont de logique différente. La grande différence des deux valeurs propres (0.3 et 0.09) se lit dans la forme du nuage des points, essentiellement un ordre de 0 à 5 régulier. L'absence de lecture (magazines et presse nationale) diminue régulièrement avec l'augmentation du niveau des diplômes. C'est un peu l'inverse, mais très atténué, pour les quotidiens régionaux.