

Rapport de travaux Pratique MA412

TP1 : Analyse en composantes principales

Table des matières

Partie A : Statistiques Descriptives.....	3
Question 1.....	3
Question 2.....	3
Question 3.....	4
Partie B : Analyse en composantes principales.....	4
Question 1.....	4
Question 2.....	5
Question 3.....	6
Question 4.....	6
Question 5.....	7
Annexes.....	7
Script Matlab.....	7

Partie A : Statistiques Descriptives

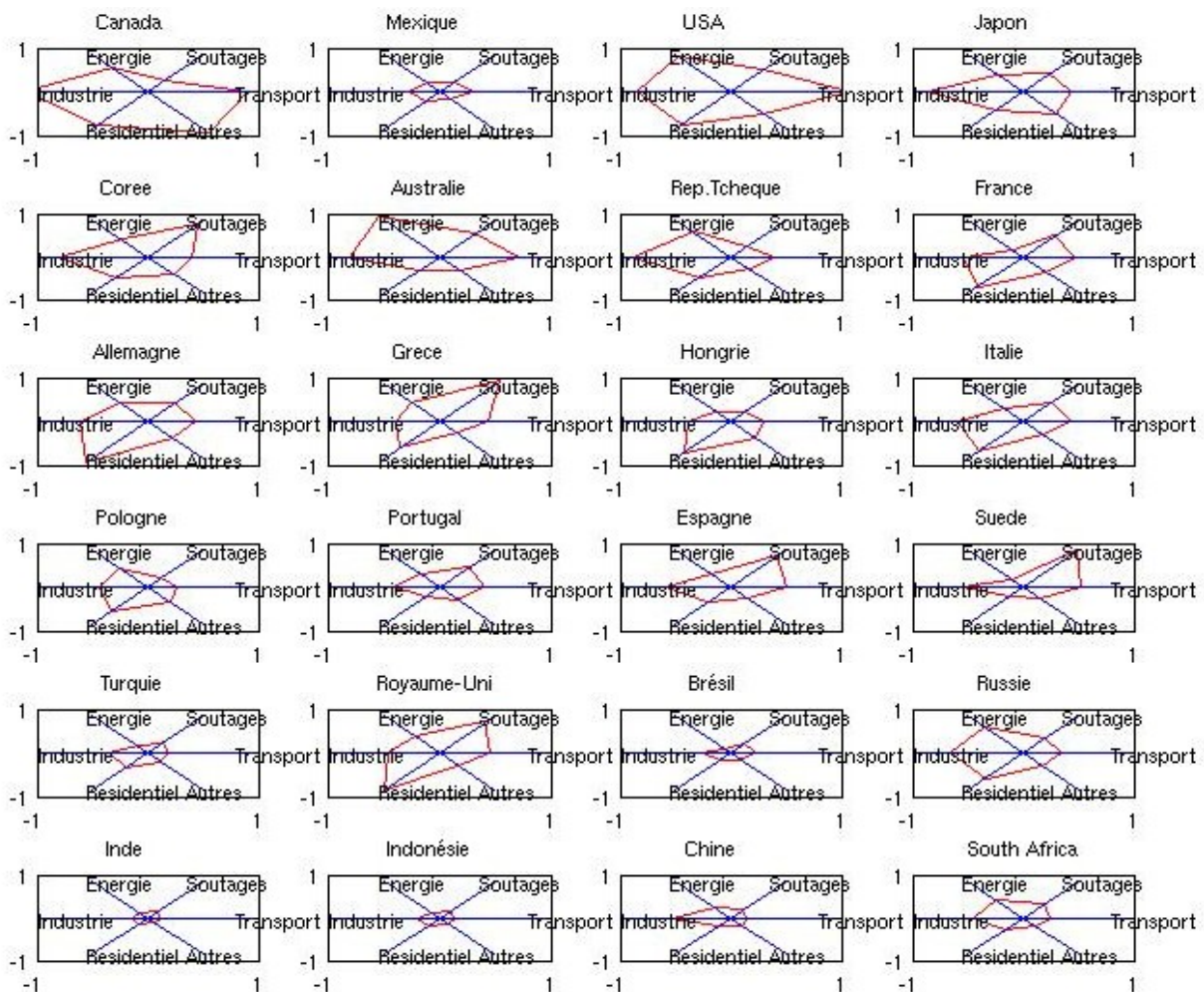
Question 1.

Fig 1.1 - Tableau des statistiques descriptives principales :

	Transport	Soutages	Energie	Industrie	Résidentiel	Autres
Moyenne	1.8353	0.3332	3.5902	1.2655	0.6280	0.4812
Écart type	1.3674	0.2998	2.6556	0.6295	0.4333	0.4017
Médiane	1.8104	0.2697	3.2620	1.2234	0.5990	0.4219
Minimum	0.0833	0.0090	0.3060	0.2079	0.0602	0.0353
Maximum	5.9104	1.0177	10.7854	2.6973	1.4756	1.9703
Distance interquartile	1.2377	0.3832	3.0889	0.9045	0.8295	0.4201

Question 2.

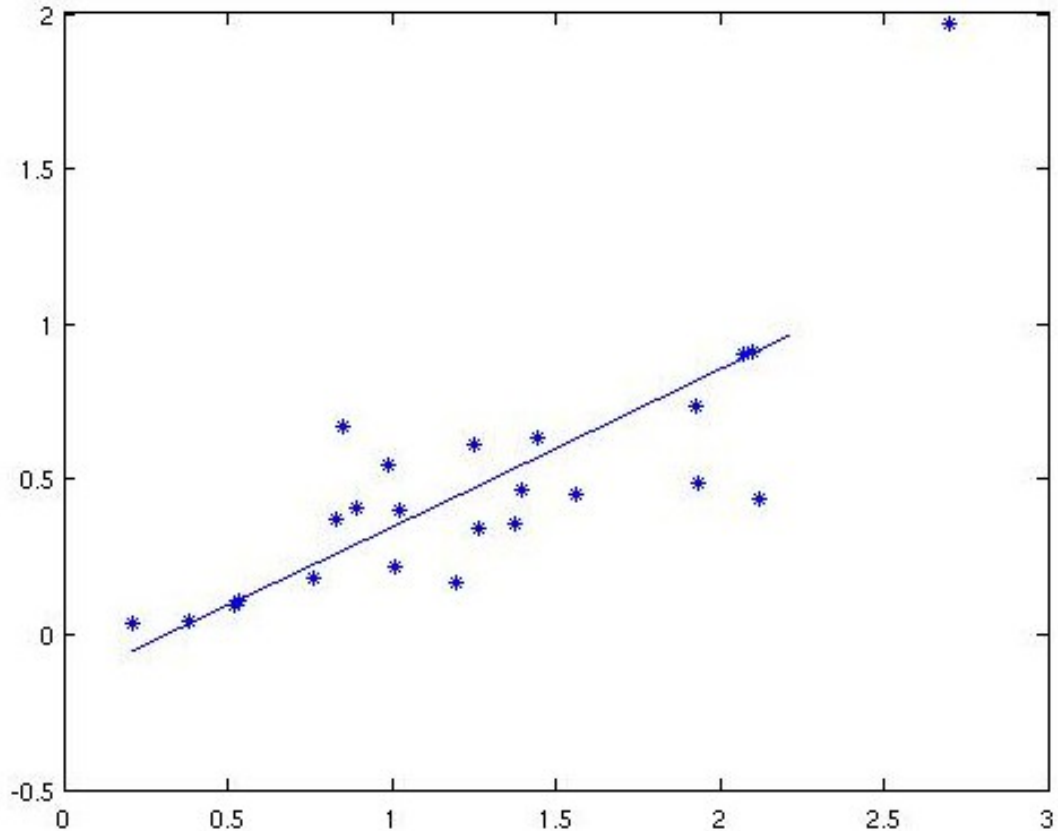
Fig 1.2 - Profil en étoile des individus :



Le profil en étoile obtenu permet d'avoir une vue d'ensemble des sources d'émissions de CO2 pour chacun des différents pays à la fois au niveau de la quantité d'émission que sa source.

Question 3.

Fig 1.3 -Régression linéaire de X6 par rapport à X4 :



En calculant le coefficient de corrélation, on obtient $R = 0.7927 > 0,75$, on peut donc voir que les autres sources d'émission peuvent se déduire des sources d'émission industrielles.

Partie B : Analyse en composantes principales

Question 1.

On normalise les données de manière à l'aide de la formule :

$$x'_{ij} = \frac{x_{ij} - \bar{x}}{\sigma_i}$$

Ceci nous permet d'obtenir une interprétation plus facile des données recueillies.

Question 2.

Fig 2.1 - Boxplot non normalisée :

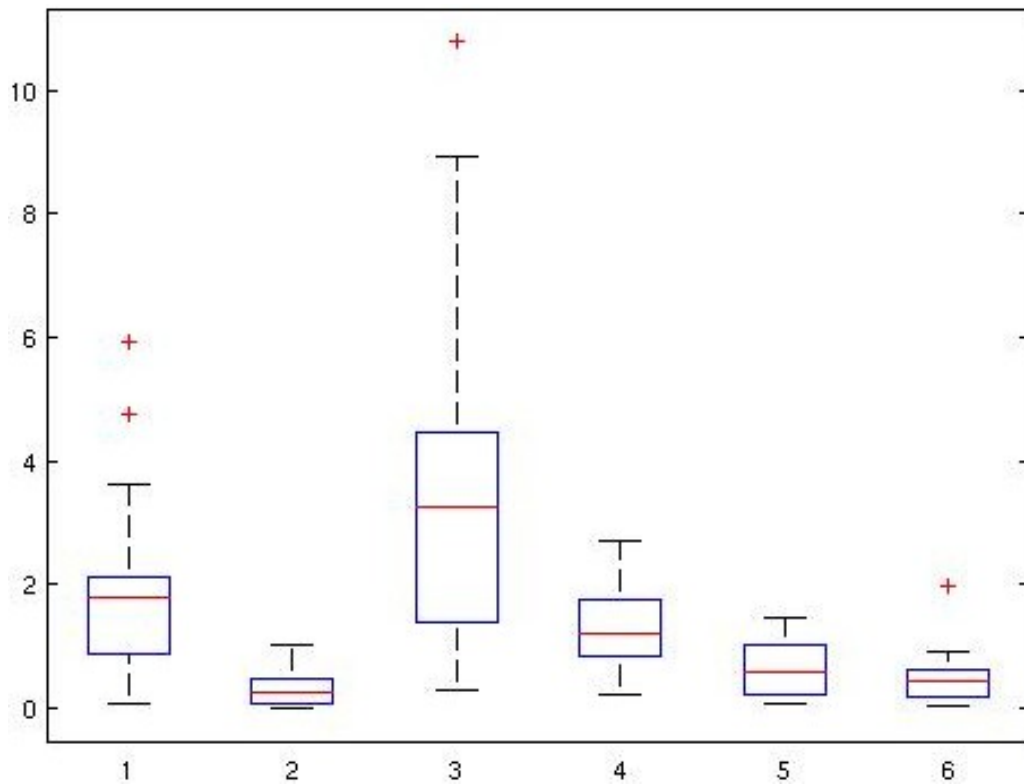
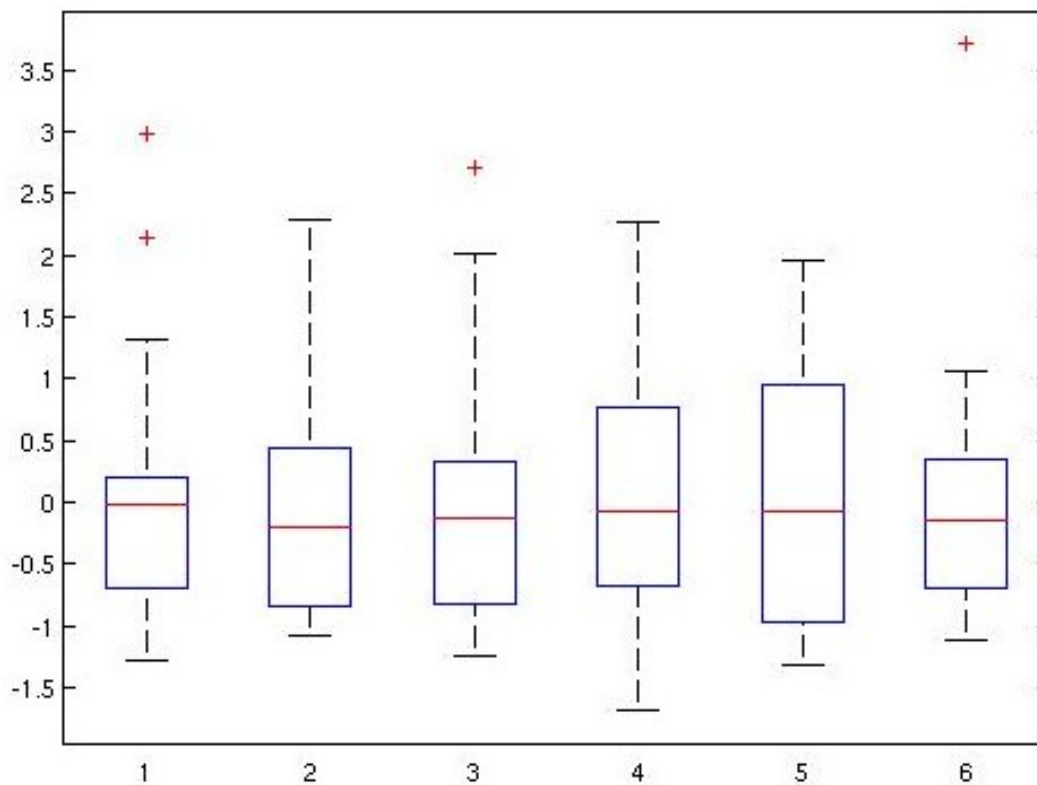


Fig 2.2 - Boxplot normalisée :



Question 3.

Les données étant normalisée, il n'est pas nécessaire de diviser par la norme du vecteur.

Les axes principaux sont les vecteurs propres suivants :

u6	u5
0.4765	-0.1154
0.1958	-0.9281
0.4133	0.0120
0.4757	0.1453
0.3679	0.0532
0.4496	0.3181

La fidélité de la représentation est obtenue à l'aide des valeurs propres de u6 et u5 :

$$\text{fidélité} = (3.5582 + 0.9693) / 6 = 0.7546$$

La fidélité étant supérieur à 0.75, un troisième axe n'est pas nécessaire.

Question 4.

Fig 2.3 – Projection des points sur le plan principal :

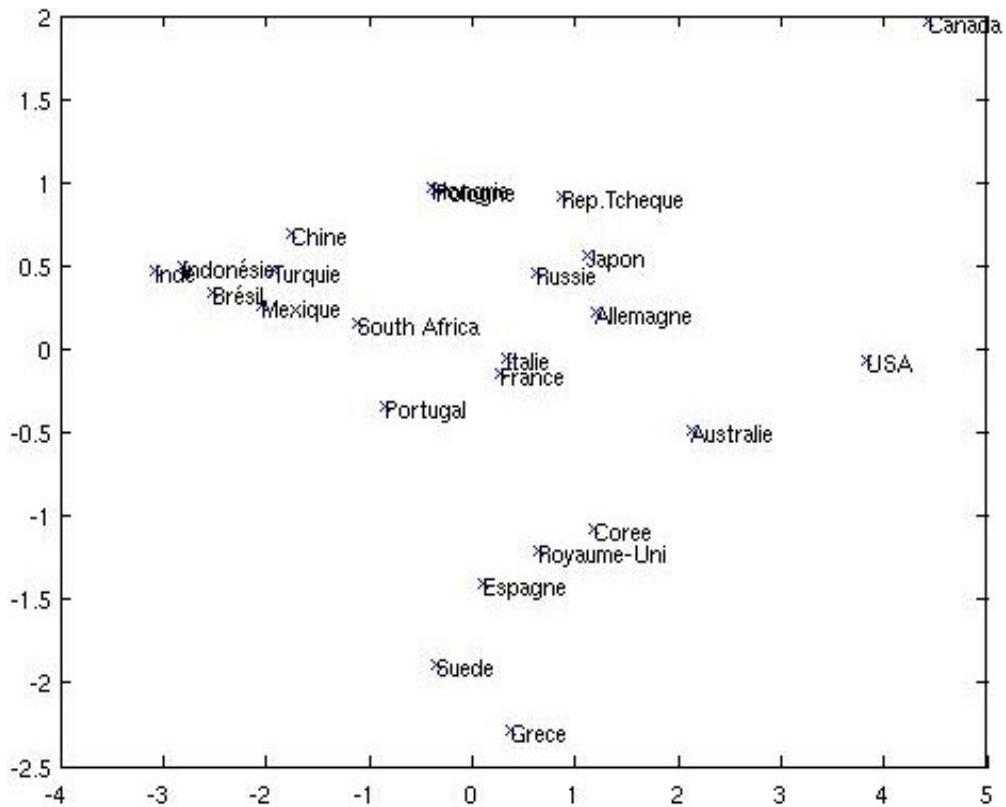
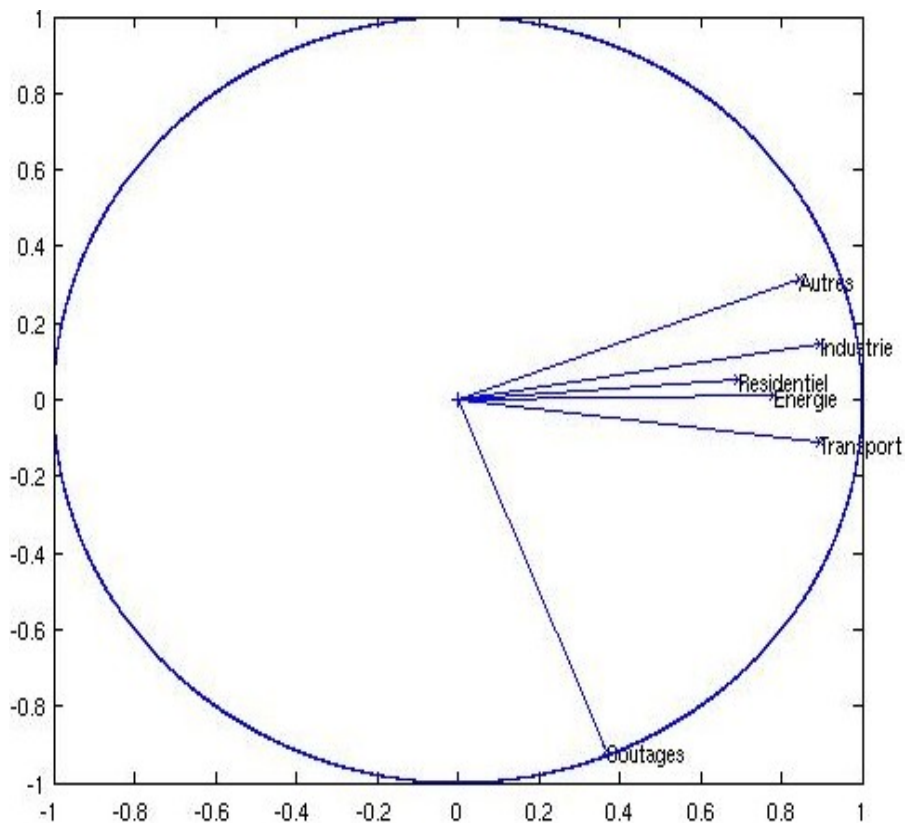


Figure 2.4 : cercle de corrélation



Question 5.

On observe que les données des sources sont fortement corrélées avec les composantes de ce plan. Plus particulièrement avec l'axe u1, sauf pour le soutage.

On observe que la pollution des pays côtier est principalement due au soutage, notamment pour la Grèce, la Suède et la Corée tandis que des pays comme les USA fortement industrialisé génèrent plus de pollution par les activités de leur société (transports, industries, résidentielle, énergie...).

Annexes

Script Matlab.

```
close all
```

```
% Statistiques descriptives
```

```
moyenne = mean(X)
```

```
ecartType = std(X)
```

```
mediane = median(X)
```

```
minimum = min(X)
```

```
maximum = max(X)
```

```
Q1 = prctile(X,25);
```

```
Q3 = prctile(X,75);
```

```
distanceInterquartile = Q3-Q1
```

```
% Profil etoile
```

```
etoile(X, Pays, Source, [6 4])
```

```
% Regression lineaire
```

```
P = polyfit(X(:,4), X(:,6), 1)
```

```
x = min(X(:,4)):max(X(:,4));
```

```
y = P(1)*x+P(2);
```

```
figure
```

```
plot(x,y)
```

```
hold on
```

```
plot(X(:,4),X(:,6), 'k')
```

```
Sxy = mean(X(:,4).*X(:,6))-moyenne(6)*moyenne(4)
```

```
Sxx = mean(X(:,4).^2)-moyenne(4)^2
```

```
Syy = mean(X(:,6).^2)-moyenne(6)^2
```

```
R = Sxy/sqrt(Sxx*Syy)
```

```
% Normalisation des données
```

```
donneesNormalisees = zeros(24,6);
```

```
for N=1:6
```

```
    donneesNormalisees(:,N) = (X(:,N)-moyenne(N))/ecartType(N);
```

```
end
```

```
% Boite a moustaches
```

```
figure
```

```
boxplot(X)
```

```
figure
```

```
boxplot(donneesNormalisees)
```

```
% Axes principaux
```

```
matriceCorrelation = corrcoef(donneesNormalisees)
```

```
[vecteursPropres,valeursPropres] = eig(matriceCorrelation)
```

```
vp = eig(matriceCorrelation)
```

```
valeursPropresSort = sort(vp, 'descend')
```

```
fidelite = (valeursPropresSort(1) + valeursPropresSort(2))/6
```

```
%Représentation graphique
```

```
figure
```

```
%Plan de projection
```

```
Px = zeros(24,1);
```

```
Py = zeros(24,1);
```

```
for i = 1 : 24
```

```
    Px(i,1) = donneesNormalisees(i,:)*vecteursPropres(:,6);
```

```
    Py(i,1) = donneesNormalisees(i,:)*vecteursPropres(:,5);
```

```
end
```

```
plot(Px,Py,'x');
```

```
hold on
```

```
text(Px,Py,Pays);
```

```
%Cercle correlation
```

```
figure
```

```
CercleCorrX = zeros(6,1);
```

```
CercleCorrY = zeros(6,1);
```

```
for j = 1: 6;
```

```
    CercleCorrX(j,1) = sqrt(valeursPropres(6,6))*vecteursPropres(j,6);
```

```
    CercleCorrY(j,1) = sqrt(valeursPropres(5,5))*vecteursPropres(j,5);
```

```
end
```

```
plot(CercleCorrX, CercleCorrY, 'x')
```

```
text(CercleCorrX, CercleCorrY, Source)
```

```
hold on
```

```
t = 0:0.01:360;
```

```
x = cos(t);
```

```
y = sin(t);
```

```
plot(x,y);
```

```
plot(0,0,'+');
```

```
for j = 1:6;
```

```
    plot([0,CercleCorrX(j,1)],[0,CercleCorrY(j,1)])
```

```
end
```