O.de Cambry **MA 412**

ESIEE **08/09**

Traitement statistique des données

TD n^0 1

On considère le tableau de données suivant concernant 10 villes françaises.

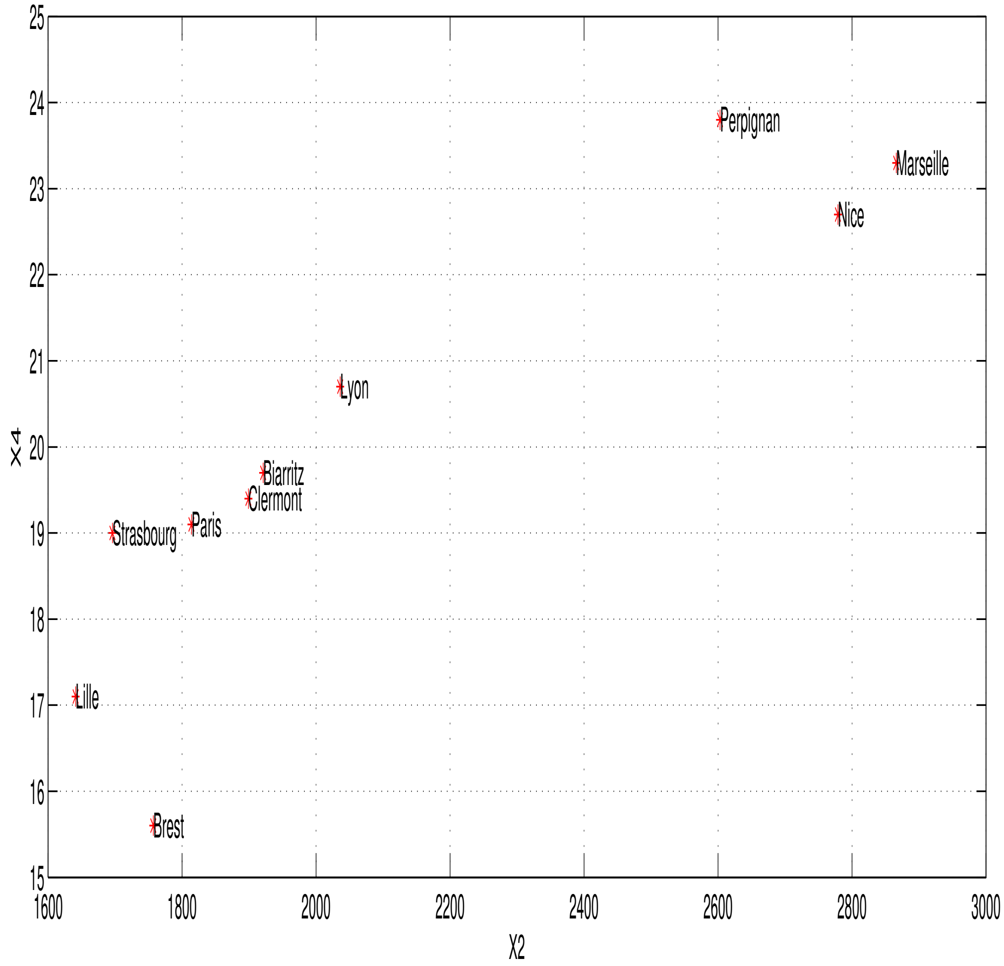
Numéro	Ville	X_1	X_2	X_3	X_4
1	Biarritz	1474	1921	7.6	19.7
2	Brest	1157	1757	6.1	15.6
3	Clermont	571	1899	2.6	19.4
4	Lille	612	1641	2.4	17.1
5	Lyon	828	2036	2.1	20.7
6	Marseille	533	2866	5.5	23.3
7	Nice	868	2779	7.5	22.7
8	Paris	624	1814	3.4	19.1
9	Perpignan	628	2603	7.5	23.8
10	Strasbourg	719	1696	0.4	19.0
Moyenne		801	2101	4.5	20.0
Ecart-type		285	442	2.51	2.51

Les variables étudiées sont :

- X_1 :Hauteur moyenne des précipitations par an
- X_2 :Durée annuelle d' ensoleillement en heures
- X_3 :Température moyenne du mois de Janvier en degré Celsius
- X_4 :Température moyenne du mois de Juillet en degré Celsius

- A - Statistiques descriptives

1. *Représenter l'histogramme en 4 classes et la boite à moustache pour X_1 et commenter ces représentations.*
2. *Déterminer l'équation de la droite de régression de X_4 par rapport à X_2.*
3. *Représenter cette droite sur la figure 1 et interpréter ce résultat.*



- B - Analyse en Composantes principales

On a effectué une Analyse en Composantes Principales sur les données normalisées, dont les résultats sont rassemblés à la suite des questions.

1. *Calculer la fidélité de la représentation des données sur le plan principal.*
2. *Déterminer les corrélations entre les caractères* X_1 *et* X_2 *et les 2 axes principaux et représenter les dans le cercle des corrélations (Figure 2).*
3. *Commenter les résultats de L'ACP.*

- C - Analyse des contributions de chaque individu

On note M_i', les coordonnées normalisées de l'individu M_i.
Pour analyser, l'information apportée par l'individu M_i, on peut utiliser les quantités suivantes, en notant $\Pi_k(M_i) =< M_i', u_k >$ la valeur de la projection de M_i sur le $k^{ième}$ axe :

- la contribution à l'inertie portée par le $k^{ième}$ axe : $CTR_k(M_i) = \dfrac{\Pi_k(M_i)^2}{n\,\lambda_k}$
- la qualité de projection sur le $k^{ième}$ axe (ou Cosinus carré) : $CO2_k(M_i) = \dfrac{\Pi_k(M_i)^2}{||M_i'||^2}$

1. *Montrer que* $\displaystyle\sum_{i=1}^{n} CTR_k(M_i) = 1$
2. *Montrer que* $CO2_{k+l}(M_i) = CO2_k(M_i) + CO2_l(M_i)$ *et* $\displaystyle\sum_{k=1}^{p} CO2_k(M_i) = 1$
3. *Déterminer ces 2 quantités pour la ville de Biarritz.*
4. Pour un axe donné, l'examen parallèle des CTR et des $CO2$ des individus qui s'y projettent peut donner lieu à quatre cas de figure :

	CTR faible	CTR forte
$CO2$ faible	Elément peu contributif quasi indépendant de l'axe	Elément très inerte contributif mais peu illustratif de l'axe
$CO2$ forte	Elément peu inerte mais bien illustratif de l'axe	Elément particulièrement caractéristique de l'axe

Analyser la position de la ville de Biarritz dans le plan principal.

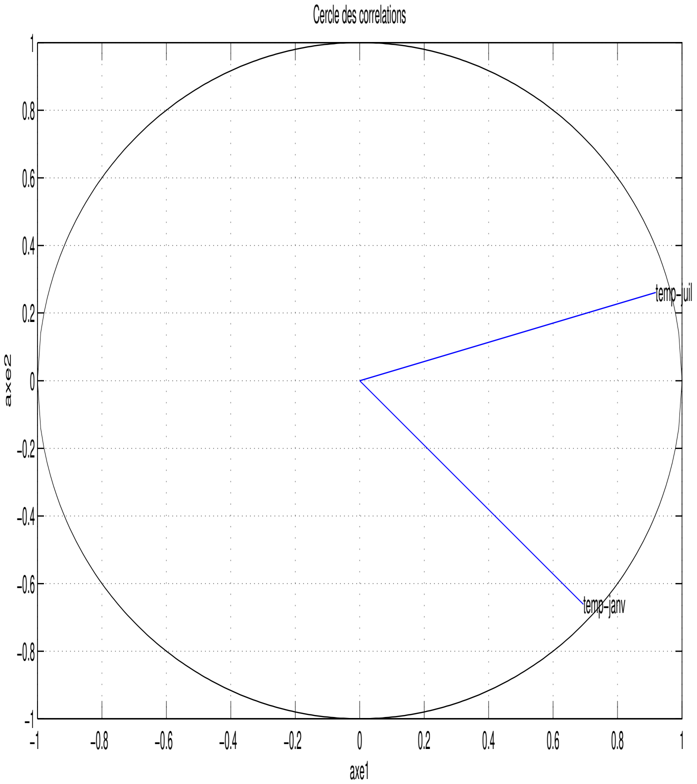
Annexe

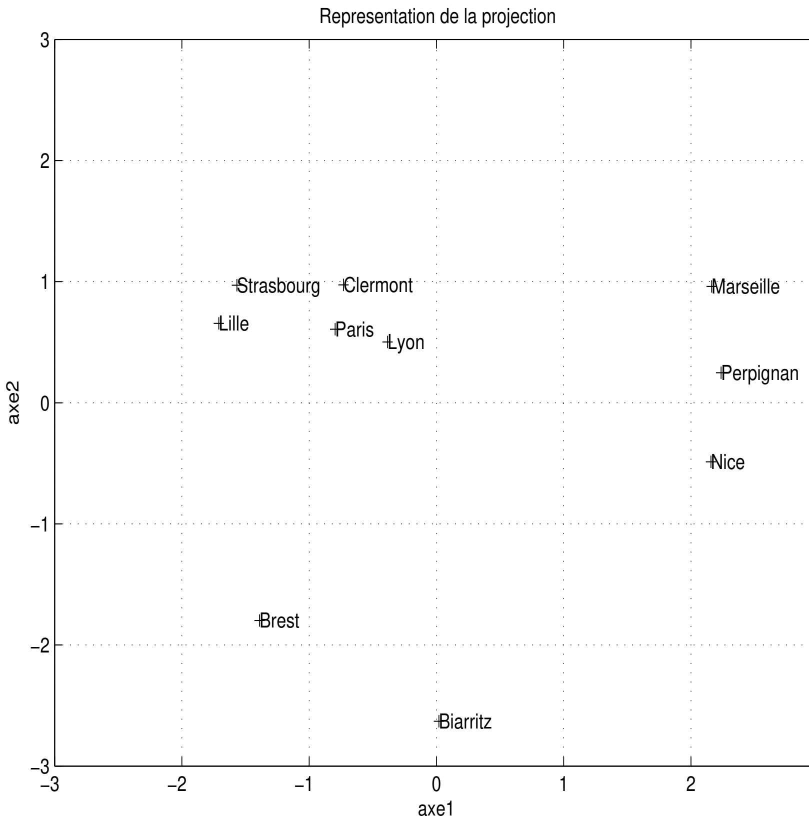
Matrice de corrélation

$$R = \begin{pmatrix} 1.0000 & -0.2178 & 0.4709 & -0.3047 \\ -0.2178 & 1.0000 & 0.6026 & 0.8925 \\ 0.4709 & 0.6026 & 1.0000 & 0.4039 \\ -0.3047 & 0.8925 & 0.4039 & 1.0000 \end{pmatrix}$$

Valeurs propres et vecteurs propres

Valeurs propres		Vecteurs propres				
$\boldsymbol{\lambda_1}$	2.30	$\boldsymbol{u_1}$	-0.08	0.64	0.46	0.61
$\boldsymbol{\lambda_2}$	1.43	$\boldsymbol{u_2}$	-0.80	0.08	-0.55	0.22
$\boldsymbol{\lambda_3}$	0.21	$\boldsymbol{u_3}$	0.56	-0.07	-0.58	0.59
$\boldsymbol{\lambda_4}$	0.06	$\boldsymbol{u_4}$	-0.21	-0.76	0.38	0.49


Cercle des correlations
temp−juil
temp−janv
axe1
axe2


Representation de la projection
Strasbourg
Clermont
Marseille
Lille
Paris
Lyon
Perpignan
Nice
Brest
Biarritz
axe1
axe2