

# « LE MONDE PEUT-IL SUFFIRE? » POUR UNE ACP

## Familiarisation à la réalisation et à l'interprétation des résultats d'une Analyse en Composantes Principales normées (ACP normée)

**Myriam BARON, Université Paris 7 – Denis Diderot**  
**Christine ZANIN, Université Paris 7 – Denis Diderot**

**Type :** TD, enseignements intégrés

**Niveau :** 3<sup>e</sup> année de Licence et au-delà

**Durée :** au moins 1 séance de 3 heures, 2 séances de 2 heures

**Thèmes :**

Géographie de la population

Utilisation de l'analyse de données pour y répondre

Etudes de cas

Cartographie thématique

**Objectifs :**

Cette ou ces séances de travaux dirigés constitue(nt) un exemple de réalisation et de commentaire et d'interprétation de résultats d'analyse de données utilisées en géographie à partir d'un exemple d'Analyse en Composantes Principales normée (ACP normée). Elle(s) s'adresse(nt) à des étudiants qui ont pu déjà acquérir des notions de statistiques uni- et bivariées au cours des deux premières années du cursus de géographie. On insiste, lors des commentaires à proprement parler, sur le fait qu'il existe deux types de résultats dans une ACP normée (comme dans la plupart des méthodes d'analyse de données) : d'une part ceux qui rendent compte de l'efficacité de l'analyse effectuée ; d'autre part ceux qui sont détaillés et se rapportent aux variables étudiées mais aussi aux unités spatiales sur lesquelles ces phénomènes sont observés. La démarche proposée vise à insister sur la relation qui doit exister entre traitements statistiques et questionnements, interrogations préalables : pour faire bref, l'outil ou les outils ne doivent pas être déconnectés d'une problématique, d'une thématique. Enfin, on propose de travailler sur les différentes formes graphiques et cartographiques des résultats d'une ACP normée.

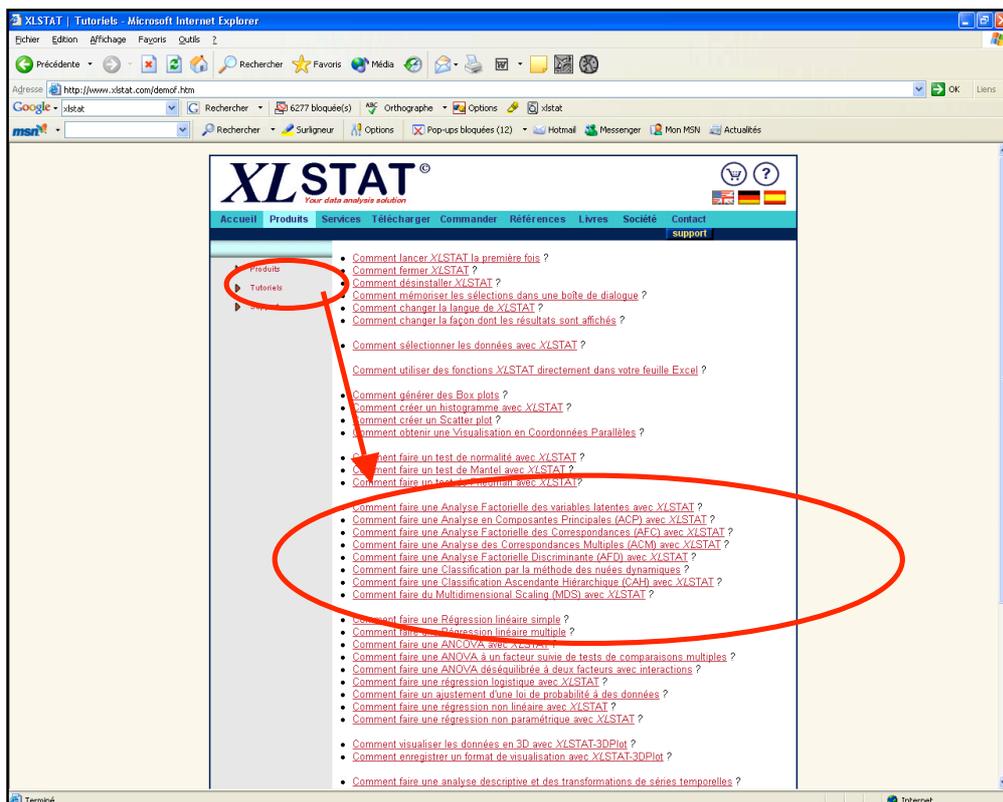
Cette séance peut être une première entrée dans l'univers des méthodes d'analyse des données. Selon les publics concernés, elle pourra être envisagée uniquement sur supports papier ou combinée avec la prise en main d'un logiciel qui permet de réaliser des analyses de données. Ici, il s'agit de résultats d'ACP obtenus sous le logiciel Xlstat, logiciel qui est téléchargeable gratuitement sur Internet pour un nombre limité d'utilisations (30 sessions de travail) à l'adresse suivante :

**Myriam BARON, Christine ZANIN, Familiarisation à la réalisation et à l'interprétation des résultats d'une Analyse en Composantes Principales normées (ACP normée), Feuilles de Géographie, III-2006, Feuilles n° 58, 28 pages**

<http://www.xlstat.com/indexfr.html>



Sur le site de téléchargement, vous trouverez un tutorial (autrement dit « Xlstat pour les Nuls » !!!) dans lequel sont expliquées les principales phases des analyses de données que l'on peut réaliser sous Xlstat, avec en prime un exemple traité de A à Z :



**Myriam BARON, Christine ZANIN**, Familiarisation à la réalisation et à l'interprétation des résultats d'une Analyse en Composantes Principales normées (ACP normée), Feuilles de Géographie, III-2006, Feuilles n° 58, 28 pages

Enfin, toujours sur ce site, et comme on s'apprête à réaliser de « grandes choses » en analyses factorielles, on pourra également télécharger le module de visualisation graphique en 3 dimensions qui permet de mieux se rendre compte des volumes des nuages de points. Il s'agit du module qui s'appelle *Xlstat-3Dplot*. Lors des commentaires de résultats et des prolongements envisagés, nous proposerons la visualisation de quelques graphiques en 3 dimensions.

Les principaux résultats, qui concernent les pays, sont cartographiés sous Philcarto, logiciel de cartographie (et bien plus que cela) téléchargeable gratuitement sur Internet à l'adresse suivante : <http://philgeo.club.fr/Index.html>

Page sans titre - Microsoft Internet Explorer

Adresse <http://philgeo.club.fr/Index.html>

Accueil / Welcome

Philippe Waniez  
Curriculum vitae  
Mémorial de titres et travaux Publications

Téléchargements / Downloads

- Windows®  
[Licence Philcarto / xPhil](#)  
[Philcarto, Phildigit, xPhil](#)  
[Fonds de cartes / Maps](#)
- Macintosh®  
[Philcarto, Philexplo, Fonds de cartes / Maps](#)

Cartographie / Cartography

Types de cartes

Géographie du Brésil  
Geography of Brazil

Habilitation à Diriger des Recherches  
Les données et le territoire au Brésil

1. Les données
2. La cartomatiques
3. La chorématique et les atlas
4. La pauvreté
5. L'habitat et les migrations
6. La mortalité
7. L'espace agropastoral
8. La modernisation agricole
9. La Réforme Agraire
10. Les religions
11. Les élections
12. Vers la mondialisation

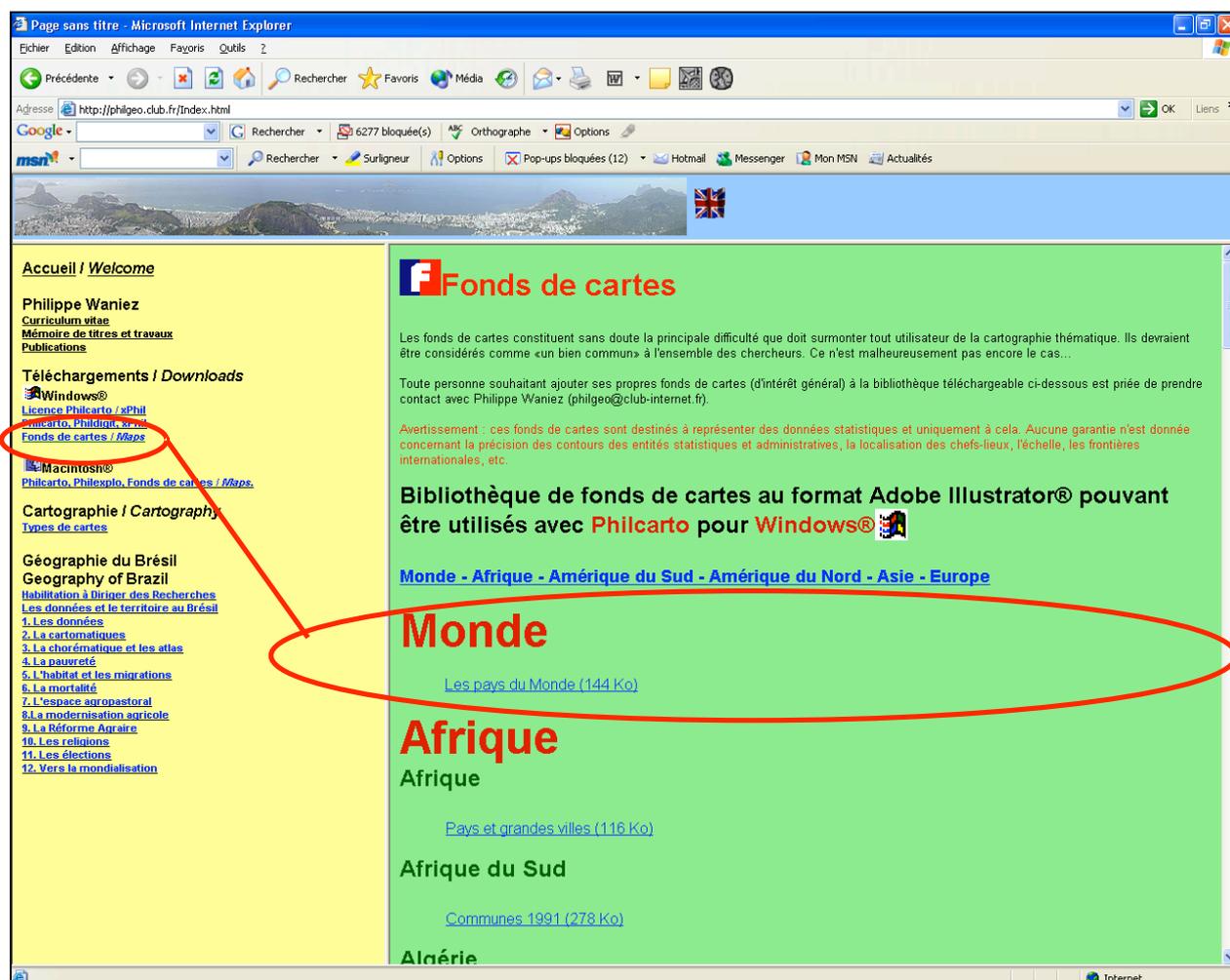
Bienvenue sur le site de Philippe WANIEZ

Le cadre ci-contre vous donne accès aux informations suivantes :

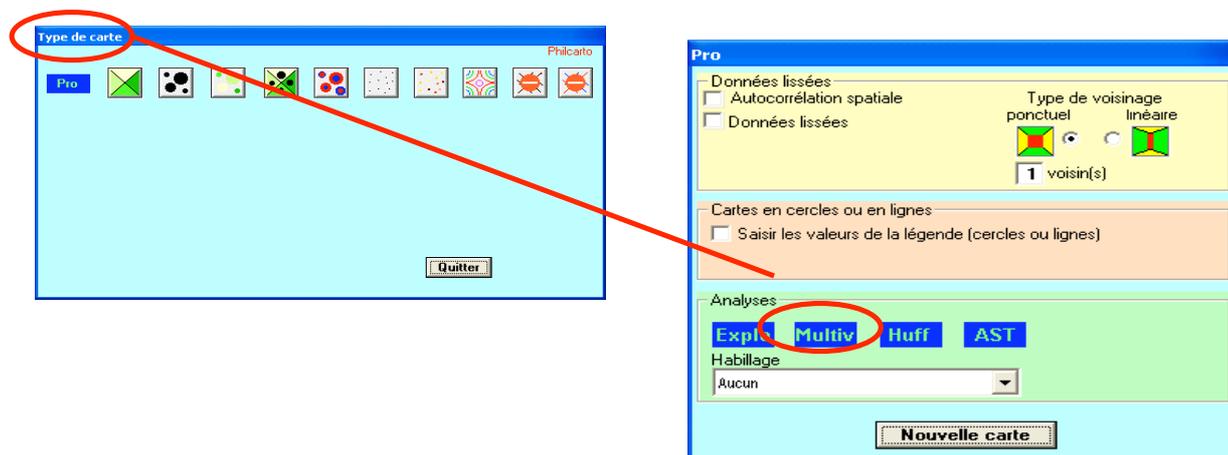
- Mon curriculum-vite
- La liste de mes publications
- Une présentation de mes recherches passées, présentes et futures (!)
- Une partie des résultats de mes recherches sur la géographie du Brésil
- Téléchargement de mes logiciels de cartographie  
[Philcarto](#) et [xPhil](#) pour PC/Windows®  
[Philcarto](#) et [Philexplo](#) pour pour Macintosh®

[Ecrivez moi!](#) [Support aux utilisateurs de Philcarto / xPhil](#)

Quant au **fond de carte** des pays du Monde, il est également **téléchargeable gratuitement** sur le site de Philippe Waniez dans sa **bibliothèque de fonds de cartes** :



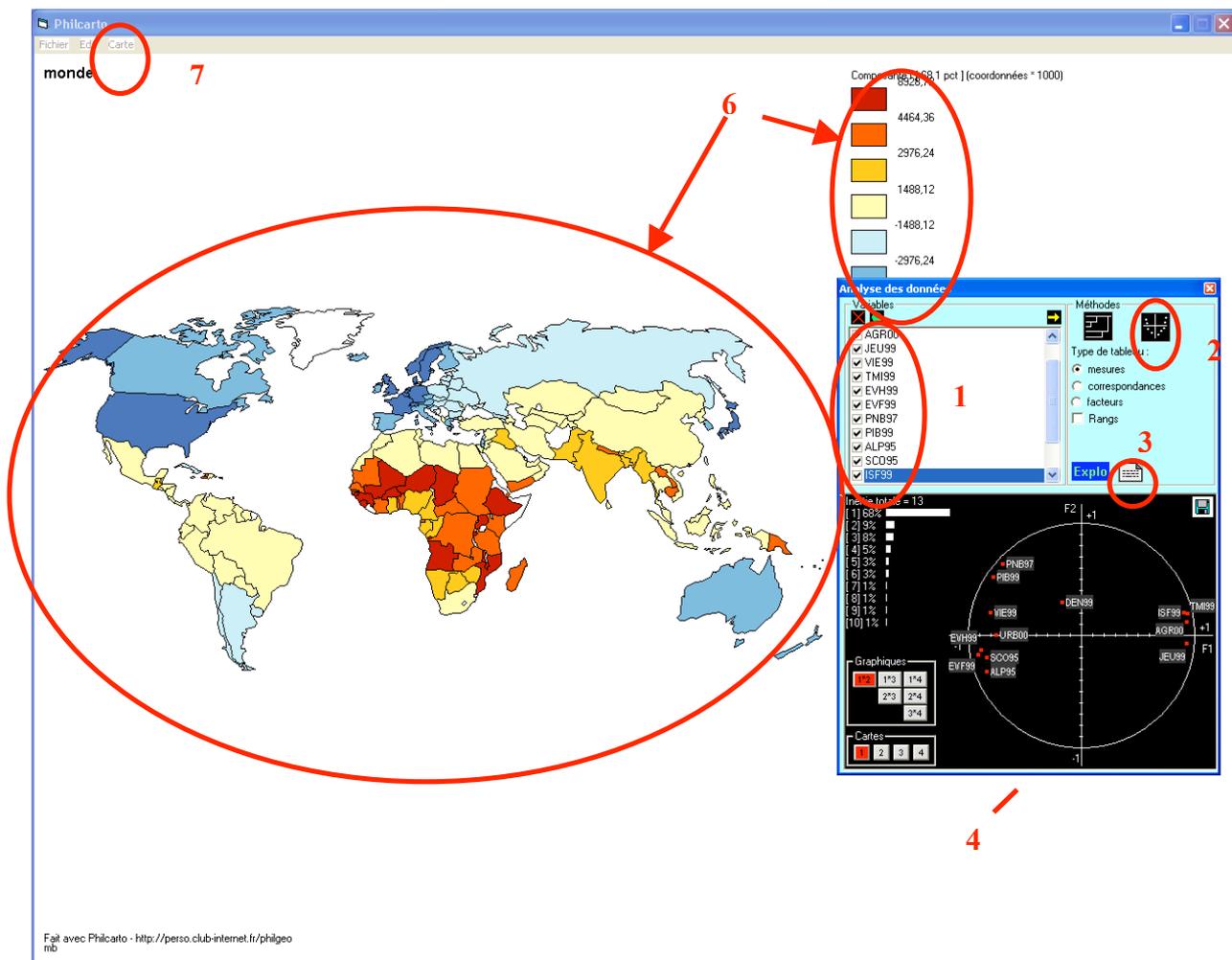
On signalera enfin que pour les enseignants ... et les étudiants, qui désirent rester dans le même environnement et donc éviter de changer de logiciel ou qui souhaitent dans une première approche exploratoire une visualisation de l'ensemble des résultats, **Phicarto** propose un **module Pro** dans la réalisation des cartes dans lequel figure **un onglet MULTIV** :



Après avoir pris la précaution de vérifier que les codes des pays sont bien les mêmes dans le fichier de données et ceux qui correspondent aux fonds de cartes, on peut :

- réaliser une ACP normée sous Philcarto (1 et 2),
- visualiser et enregistrer l'ensemble des résultats (3),
- visualiser et enregistrer les graphiques des principaux plans factoriels (4)
- visualiser et enregistrer les cartes des coordonnées des pays sur les 4 premiers facteurs (5, 6 et 7) !!!

Le Pérou non ?!

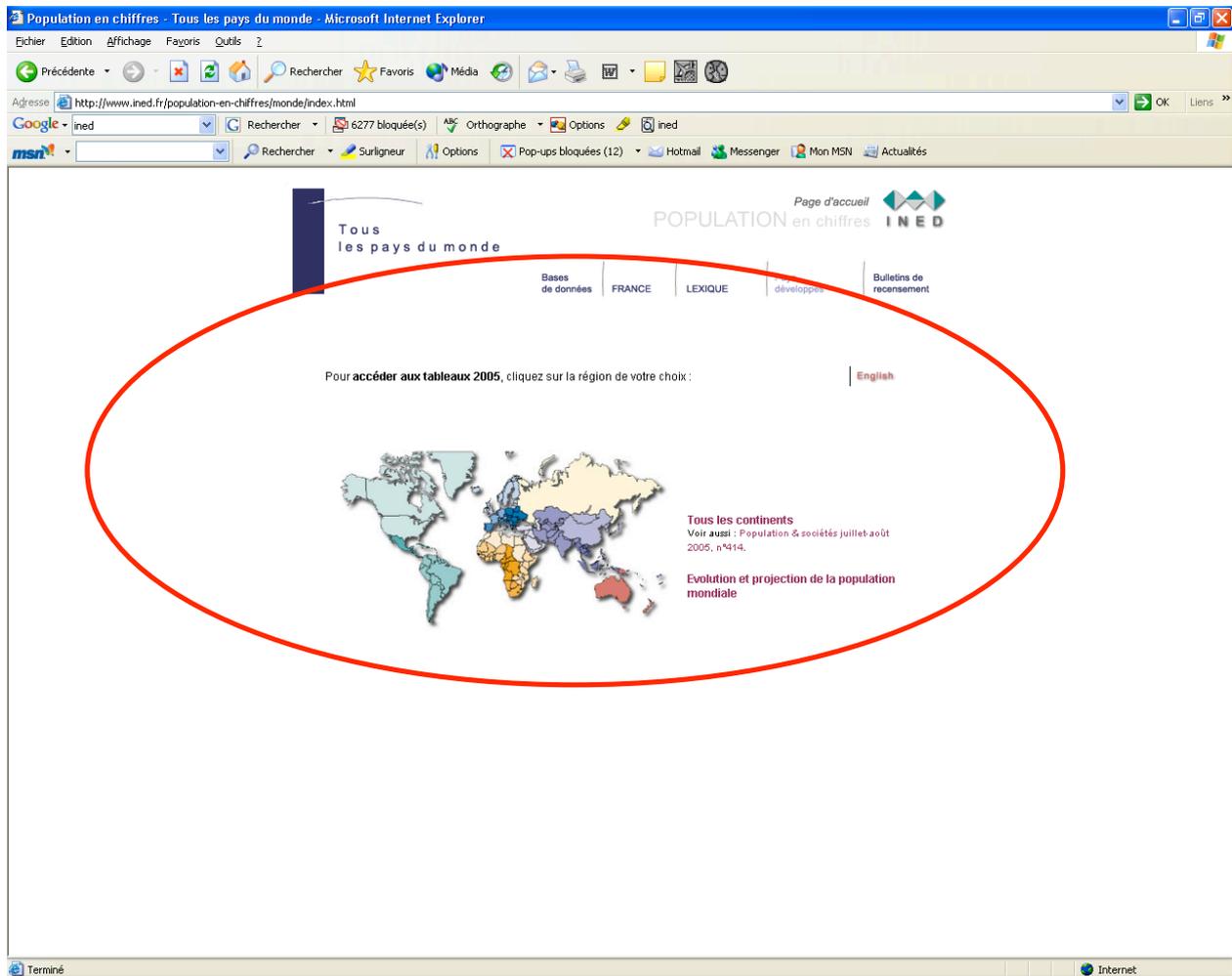


Pour plus de précisions, nous vous recommandons le manuel d'utilisation de Philcarto qui est également disponible sur le site, ainsi que la rubrique FAQ.

### Déroulement :

La séance de travail est basée sur la familiarisation à l'interprétation des résultats d'ACP normée par les étudiants, voire à la réalisation d'une ACP normée. L'ensemble des documents (ou et des fichiers de données) est distribué au début de la séance. On travaille sur les résultats d'ACP normées menées sur un tableau élémentaire de données qui concerne la situation démographique, économique et sociale de 146 pays du monde appréhendée à partir de données mises à disposition gratuitement par l'INED et collectées au cours de la dernière décennie (entre 1990 et 2000) à l'adresse suivante :

<http://www.ined.fr/population-en-chiffres/monde/index.html>



Cette séance comporte trois temps :

- **le premier temps** est celui de **la familiarisation et de la maîtrise du vocabulaire propre à l'ACP normée**, de son appropriation à partir des résultats généraux qui concernent la qualité, l'efficacité de l'analyse factorielle employée.
- **le deuxième temps** est consacré au **commentaire des résultats détaillés de l'ACP normée sur les pays du Monde**. Ces résultats concernent d'une part l'organisation de l'information en fonction des variables, d'autre part l'organisation de cette même information en fonction des unités spatiales, en l'occurrence ici les 146 pays du Monde pour lesquelles l'information est disponible. Il vise à bien vérifier que les étudiants sont capables de relier les résultats statistiques à des contenus thématiques, mais aussi de formuler clairement des conclusions à partir de ces premiers résultats : il s'agit en particulier de sensibiliser les étudiants à la mise en regard du classement des pays du Monde par l'IDH et du même classement par des indicateurs démographiques, sociaux et économiques.
- Enfin, **le troisième temps** est consacré aux différents types de valorisation graphique et cartographique des résultats détaillés de l'ACP normée, des valeurs de l'IDH, et enfin des écarts à la relation qui peut exister entre coordonnées des pays sur le premier facteur et valeur de l'IDH.

## « LE MONDE PEUT-IL SUFFIRE ? » POUR UNE ACP

### 1. Recherche Indice de Développement Humain ... désespérément

**Documents** : Tableau concernant la description du tableau élémentaire caractérisant les Etats du Monde

Tableaux 2.1 et 2.2 résumant le tableau élémentaire de données

**Sources** : INED, 2000-2002

Il s'agit de faire acquérir une méthode de dépouillement et d'interprétation des résultats statistiques. On sensibilise les étudiants non seulement aux interprétations des résultats mais aussi à leurs valorisations graphiques et cartographiques.

L'ensemble des 14 caractères retenus pour ce travail sont des variables quantitatives de taux continues : ce qui signifie que la plupart des traitements statistiques uni- ou bivariés sont possibles. Il est indispensable dans tout résumé de données chiffrées d'avoir une idée claire des ordres de grandeur, de l'étalement (ou dispersion) des valeurs, enfin de pouvoir comparer voire classer les variables les unes par rapport aux autres... Dans ce dernier cas on peut chercher à évaluer les différenciations introduites par les variables entre les 146 pays du Monde : on utilise alors un paramètre de dispersion. Les caractéristiques de dispersion ont pour but d'indiquer globalement, par une seule valeur, l'hétérogénéité d'une distribution. Comme on doit comparer la dispersion de plusieurs variables qui ont certes la même unité de mesure mais pas toujours les mêmes ordres de grandeur, on doit choisir un paramètre de dispersion relative qui exprime l'hétérogénéité indépendamment de l'unité de mesure et sans effet de taille. Le paramètre de dispersion relative dont nous disposons ici est le **coefficient de variation**. Ce coefficient permet de comparer les variables ayant des unités de mesure différentes et de classer les variables en fonction du degré de différenciation qu'elles introduisent entre les pays du Monde.

On rappelle alors que les variables qui introduisent le plus de différenciations entre les Etats du monde sont celles qui sont caractérisées par le coefficient de variation le plus élevé. Si l'on se limite aux coefficients de variation supérieurs à 1 (en ayant soin de rappeler que cette valeur n'est nullement remarquable), on constate que de manière relativement attendue le produit intérieur brut en 1997 (dont le coefficient de variation est égal à 1,06) et le produit national brut en 1997 (dont le coefficient de variation est égal à 1,61) font partie des phénomènes sélectionnés qui introduisent le plus de différenciations entre les 146 pays du Monde. On pense en effet immédiatement aux écarts qui existent entre pays très riches comme les Etats-Unis pour lesquels le produit national brut atteint 29 080\$ par an et par habitant, voire comme la Suisse où ce dernier dépasse 43 000\$ et certains pays africains, comme l'Ethiopie où ce même PNB dépasse à peine 100\$ par an et par habitant, ou asiatiques comme le Cambodge où le PNB est égal à 300\$ par an et par habitant. En revanche, car c'est une dimension que l'on oublie trop souvent, il est plus surprenant de trouver dans ce groupe de variables la densité de population (hab/km<sup>2</sup>) en 1999, dont le coefficient de variation est égal 1,26. On peut tout de même mentionner les écarts qui peuvent exister entre l'occupation humaine moyenne d'un pays comme l'Islande, la Namibie ou encore l'Australie dont la densité de population n'est que de 3 habitants/km<sup>2</sup> alors qu'elle est 3 fois supérieure dans la Fédération Russe elle et des pays comme le Japon, l'Inde ou encore Israël dans lesquels cette même densité de population atteint 300 habitants/km<sup>2</sup>, soit 100 fois plus. A l'opposé, les phénomènes, qui introduisent le moins de différenciations entre les pays du Monde et qui sont donc ceux qui ont le coefficient de variation le plus faible, sont l'espérance de vie des hommes et l'espérance de vie des femmes en 1999 (coefficients de variation respectivement égaux à 0,18 et 0,19).

Après avoir rappelé ce qu'est un coefficient de corrélation (à ne pas confondre bien évidemment avec le coefficient de variation vu auparavant), son intervalle de variation (compris entre -1 et +1)

et ses différentes significations on insiste sur ce qu'il nous apprend des différentes combinaisons de variables dans les pays du Monde. On insiste ici sur le fait que, comme on travaille sur un ensemble important de données, le seuil à partir duquel le coefficient de corrélation est significatif est plutôt faible : dès  $|0,168|$ . D'où la nécessité de ne construire les graphes de corrélation qu'à partir des valeurs absolues les plus fortes en l'occurrence supérieures à  $|0,6|$ . Sur 83 corrélations significatives, 84 % sont en effet supérieures à  $|0,6|$ .

### Matrice des corrélations

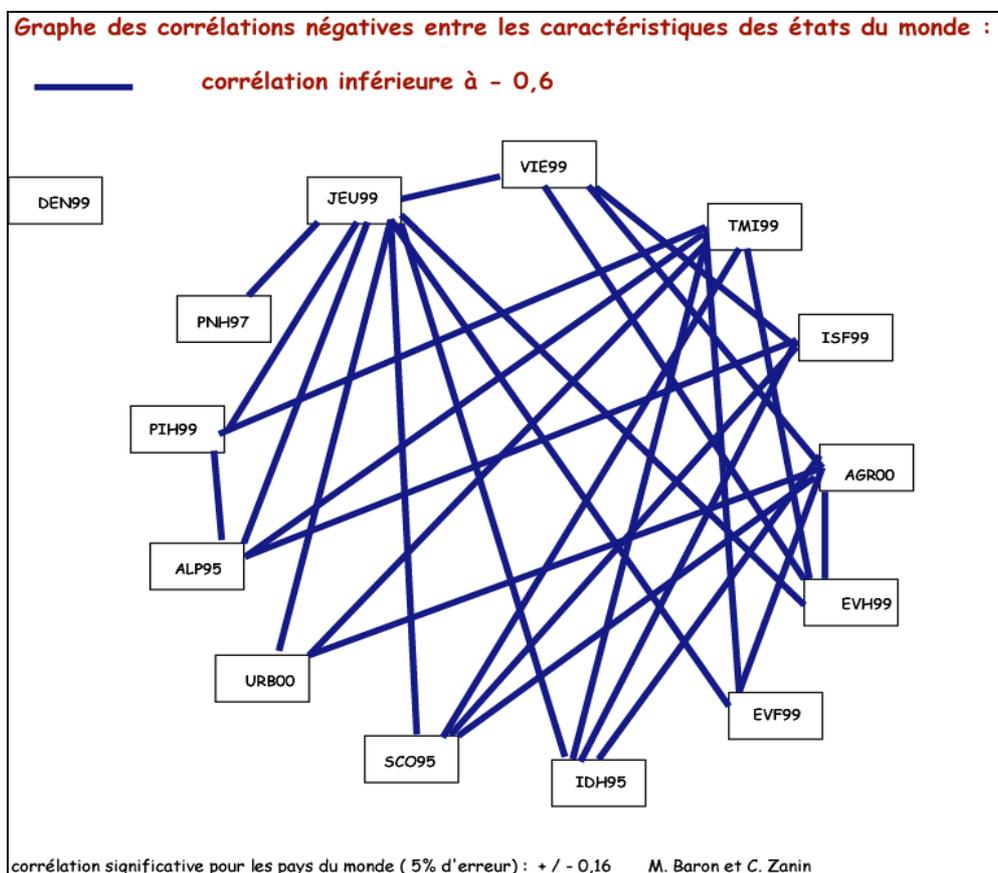
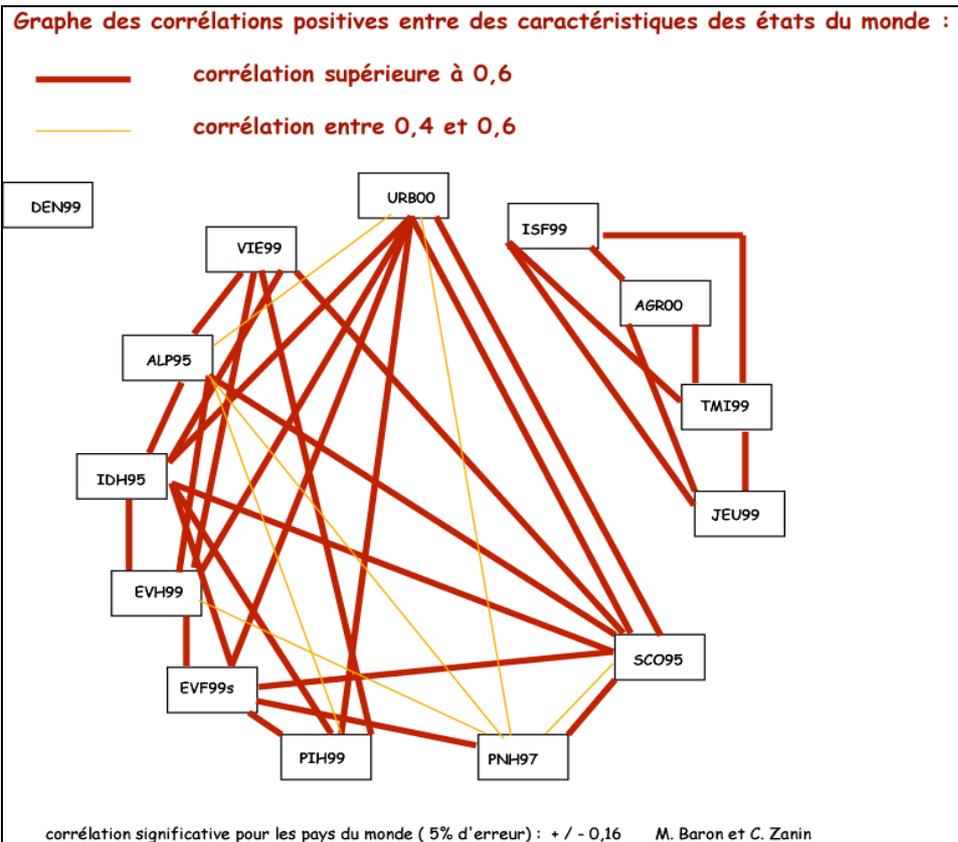
Codes des variables	DEN99	URB00	AGR00	JEU99	VIE99	TMI99	EVH99
DEN99	1	0,002	-0,092	<b>-0,200</b>	0,146	-0,155	<b>0,218</b>
URB00	0,002	1	<b>-0,824</b>	<b>-0,621</b>	<b>0,519</b>	<b>-0,655</b>	<b>0,656</b>
AGR00	-0,092	<b>-0,824</b>	1	<b>0,767</b>	<b>-0,648</b>	<b>0,856</b>	<b>-0,840</b>
JEU99	<b>-0,200</b>	<b>-0,621</b>	<b>0,767</b>	1	<b>-0,908</b>	<b>0,807</b>	<b>-0,751</b>
VIE99	0,146	<b>0,519</b>	<b>-0,648</b>	<b>-0,908</b>	1	<b>-0,669</b>	<b>0,600</b>
TMI99	-0,155	<b>-0,655</b>	<b>0,856</b>	<b>0,807</b>	<b>-0,669</b>	1	<b>-0,905</b>
EVH99	<b>0,218</b>	<b>0,656</b>	<b>-0,840</b>	<b>-0,751</b>	<b>0,600</b>	<b>-0,905</b>	1
EVF99	<b>0,181</b>	<b>0,678</b>	<b>-0,866</b>	<b>-0,806</b>	<b>0,665</b>	<b>-0,922</b>	<b>0,983</b>
PNB97	<b>0,168</b>	<b>0,552</b>	<b>-0,561</b>	<b>-0,634</b>	<b>0,664</b>	<b>-0,522</b>	<b>0,566</b>
PIB99	0,122	<b>0,640</b>	<b>-0,656</b>	<b>-0,704</b>	<b>0,694</b>	<b>-0,624</b>	<b>0,654</b>
ALP95	0,052	<b>0,585</b>	<b>-0,768</b>	<b>-0,773</b>	<b>0,655</b>	<b>-0,827</b>	<b>0,720</b>
SCO95	0,061	<b>0,630</b>	<b>-0,804</b>	<b>-0,695</b>	<b>0,615</b>	<b>-0,789</b>	<b>0,727</b>
IDH95	0,091	<b>0,757</b>	<b>-0,899</b>	<b>-0,813</b>	<b>0,676</b>	<b>-0,909</b>	<b>0,894</b>
ISF99	<b>-0,239</b>	<b>-0,551</b>	<b>0,767</b>	<b>0,907</b>	<b>-0,774</b>	<b>0,868</b>	<b>-0,788</b>

En gras, valeurs significatives (hors diagonale) au seuil  $\alpha=0,050$  (test bilatéral)

Codes des variables	EVF99	PNB97	PIB99	ALP95	SCO95	IDH95	ISF99
DEN99	<b>0,181</b>	<b>0,168</b>	0,122	0,052	0,061	0,091	<b>-0,239</b>
URB00	<b>0,678</b>	<b>0,552</b>	<b>0,640</b>	<b>0,585</b>	<b>0,630</b>	<b>0,757</b>	<b>-0,551</b>
AGR00	<b>-0,866</b>	<b>-0,561</b>	<b>-0,656</b>	<b>-0,768</b>	<b>-0,804</b>	<b>-0,899</b>	<b>0,767</b>
JEU99	<b>-0,806</b>	<b>-0,634</b>	<b>-0,704</b>	<b>-0,773</b>	<b>-0,695</b>	<b>-0,813</b>	<b>0,907</b>
VIE99	<b>0,665</b>	<b>0,664</b>	<b>0,694</b>	<b>0,655</b>	<b>0,615</b>	<b>0,676</b>	<b>-0,774</b>
TMI99	<b>-0,922</b>	<b>-0,522</b>	<b>-0,624</b>	<b>-0,827</b>	<b>-0,789</b>	<b>-0,909</b>	<b>0,868</b>
EVH99	<b>0,983</b>	<b>0,566</b>	<b>0,654</b>	<b>0,720</b>	<b>0,727</b>	<b>0,894</b>	<b>-0,788</b>
EVF99	1	<b>0,547</b>	<b>0,639</b>	<b>0,773</b>	<b>0,742</b>	<b>0,900</b>	<b>-0,840</b>
PNB97	<b>0,547</b>	1	<b>0,950</b>	<b>0,435</b>	<b>0,521</b>	<b>0,630</b>	<b>-0,451</b>
PIB99	<b>0,639</b>	<b>0,950</b>	1	<b>0,527</b>	<b>0,616</b>	<b>0,751</b>	<b>-0,535</b>
ALP95	<b>0,773</b>	<b>0,435</b>	<b>0,527</b>	1	<b>0,823</b>	<b>0,843</b>	<b>-0,847</b>
SCO95	<b>0,742</b>	<b>0,521</b>	<b>0,616</b>	<b>0,823</b>	1	<b>0,860</b>	<b>-0,759</b>
IDH95	<b>0,900</b>	<b>0,630</b>	<b>0,751</b>	<b>0,843</b>	<b>0,860</b>	1	<b>-0,818</b>
ISF99	<b>-0,840</b>	<b>-0,451</b>	<b>-0,535</b>	<b>-0,847</b>	<b>-0,759</b>	<b>-0,818</b>	1

En gras, valeurs significatives (hors diagonale) au seuil  $\alpha=0,050$  (test bilatéral)

On peut alors mettre en regard le graphe qui traduit les associations entre variables (autrement dit celui qui est construit à partir des corrélations positives et significatives) et celui qui traduit les oppositions (autrement dit celui qui est construit à partir des corrélations négatives et significatives).



En ne conservant que les corrélations les plus fortes, on remarque que :

- La variable densité de population, qui est parmi celle qui introduit le plus de différenciations entre les pays du Monde, n'a pas de relation avec les 13 autres variables
- Parmi les corrélations positives, on peut relever deux groupes de variables bien distinctes. Le premier regroupe le taux de mortalité infantile (TMI99), la part des jeunes dans la population totale (JEUN99), la part des agriculteurs dans la population active (AGR00) et enfin l'indice synthétique de fécondité (ISF99). Le second comporte le taux de population urbaine (URB00), la part des populations âgées (VIE99), l'espérance de vie des hommes (EVH99), celle des femmes (EVF99), le produit national brut (PNB97), le produit intérieur brut (PIB99), le taux d'alphabétisation (ALP95), le taux de scolarisation des populations (SCO95) et enfin l'indice de développement humain (IDH95).
- Les corrélations négatives traduisent certaines oppositions évidentes et n'amènent aucune information supplémentaire : la part de la population urbaine dans la population totale s'oppose à la part de la population agricole dans la population active, la part des populations jeunes s'oppose aux espérances de vie des hommes et des femmes. Les autres oppositions complètent les associations vues précédemment : le taux de mortalité infantile (TMI99) s'oppose au taux d'alphabétisation (ALP95), au taux d'urbanisation (URB00) et à l'indice de développement humain (IDH95). Tandis que le taux de scolarisation (SCO95) s'oppose à l'indice synthétique de fécondité (ISF99), à la part des agriculteurs dans la population active (AGR00) et au taux de mortalité infantile (TMI99) etc.

Ces associations et ces oppositions permettent d'émettre quelques hypothèses concernant la structuration de l'information dans le cadre de l'ACP normée qui va être réalisée. Si l'on suppose que le taux de mortalité infantile participe à la définition de la principale composante démographique, économique et sociale qui différencie et classe les pays du Monde, on trouvera à proximité l'indice synthétique de fécondité mais aussi la part des jeunes dans la population totale. On peut penser qu'on trouvera à l'autre extrémité, en opposition, les espérances de vie des hommes et des femmes, les produits intérieurs et nationaux bruts.

Au final, on s'interroge sur la place de l'Indice de Développement Humain (IDH95) dans cette analyse exploratoire. En effet, on peut rappeler sa définition ou renvoyer les étudiants vers quelques adresses Internet qui présentent de manière très complète cet indicateur.

*« Le Programme des Nations Unies pour le développement (PNUD) publie depuis 1990 un indice synthétique du développement humain (IDH). Cet indice correspond à la volonté de compléter les indicateurs traditionnels (PNB, PIB, Revenu national) par un **chiffre plus synthétique construit à partir de critères fondamentaux liés au développement.***

*On retrouve ainsi les idées du développement humain, à savoir **préserver les facultés et droits de l'homme à tous les niveaux et instaurer la capacité de gestion du développement. L'IDH privilégie la longévité, le savoir, le niveau de vie.** Cet indice est actuellement calculé à partir de quatre variables de base : **revenu, espérance de vie, alphabétisation des adultes, nombre moyen d'années d'étude, en différenciant le primaire, le secondaire et le supérieur.** Le développement est lié à trois composantes : **mener une vie longue et saine, accéder à la connaissance et à l'information, enfin bénéficier de ressources assurant un niveau de vie décent.***

*L'IDH comporte une valeur maximale et une valeur minimale **pour chaque critère** ; ce qui permet d'exprimer **la position de chaque pays entre 0 et 1.** Ainsi, l'espérance de vie à la naissance s'étale entre 25 et*

**Myriam BARON, Christine ZANIN**, Familiarisation à la réalisation et à l'interprétation des résultats d'une Analyse en Composantes Principales normées (ACP normée), Feuilles de Géographie, III-2006, Feuilles n° 58, 28 pages

85 ans, l'alphabétisation des adultes et le taux de scolarisation entre 0 % et 100 %, et le PIB réel par habitant entre 100 \$ et 40 000 \$.

Le PNUD distingue trois groupes de pays : les pays à développement humain élevé (IDH > 0,804), les pays à développement moyen (0,803 > IDH > 0,507) et les pays à faible développement humain (IDH < 0,506). »

extrait de <http://www.socialinfo.ch/cgi-bin/dicoposso/show.cfm?id=418>

« 21 -L'Indice synthétique du développement humain (IDH).

- Mode de calcul: l'IDH est un indice normé.

L'IDH, utilisé depuis 1990 par le PNUD, privilégie la longévité, le savoir, le niveau de vie. Conçu au départ comme la moyenne arithmétique des indicateurs de durée de vie, de niveau d'éducation, et de PIB réel corrigé par la PPA, il est actuellement calculé à partir de quatre variables de base: revenu, espérance de vie, alphabétisation des adultes, nombre moyen d'années d'étude, en différenciant le primaire, le secondaire et le supérieur. Le développement a trait à la "possibilité" fondamentale (et non plus à la détention de biens matériels) d'intégration d'un ou plusieurs individus dans la société. Cette possibilité a trois composantes : mener une vie longue et saine, accéder à la connaissance et à l'information, enfin bénéficier de ressources assurant un niveau de vie décent.

L'IDH comporte une valeur maximale et une valeur minimale pour chaque critère; ce qui permet d'exprimer la position de chaque pays entre 0 et 1. Ceci veut dire que l'IDH, normé, au départ sur le niveau du Japon, puis de la Suisse, l'est désormais par rapport aux valeurs maximales des variables retenues:

- Espérance de vie à la naissance: 25- 85 ans. - Alphabétisation des adultes; 0% - 100 % - Taux de scolarisation: 0%- 100 % - PIB réel par habitant (PPA): 100 \$ - 40 000 \$. Le calcul de l'IDH soulève de nombreux problèmes de traitement, le principal étant le traitement du revenu au dessus du seuil de pauvreté, dont le rendement est décroissant par hypothèse. D'où l'utilisation du logarithme et d'un coefficient de pondération nul au revenu au dessus du seuil de pauvreté.

Classement. Le PNUD distingue trois groupes de pays:

- les pays à développement humain élevé dont l' IDH est supérieur à 0,804

- les pays à développement moyen dont l'IDH > 0,507

- et en deçà les pays à faible développement humain, le Niger (0,207) et la Sierra Leone (0,185) détenant l'IDH le plus faible. L'intérêt revient à accentuer la différence entre l'IDH et le PIB, à diminuer la corrélation de rang entre les deux classements. Ainsi dans le classement du rapport 1995, des écarts positifs importants caractérisent les pays nordiques (sauf le Danemark) et certains pays très pauvres tels Madagascar ou Sao Tomé et Príncipe. Le Canada devient la nation économique la plus développée dans le cadre d'un bouleversement de hiérarchie des pays industrialisés.»

Extrait de <http://www.globenet.org/horizon-local/webdev/9905indic.html#21>

Il résulte de ces extraits de définition que **l'IDH fait en partie double emploi avec des variables comme l'espérance de vie des femmes et des hommes, le taux de scolarisation et les PIB et PNB par habitant**. On peut alors soit le supprimer, soit le **garder dans la future analyse en prenant soin qu'il ne joue aucun rôle dans la structuration des résultats : autrement dit en le déclarant en variable supplémentaire**. Une précaution indispensable sinon il y a fort à parier que le rôle joué par l'espérance de vie des femmes et des hommes, le taux de scolarisation et les PIB et PNB par habitant verront leur poids sensiblement accru dans l'analyse.

## **2. Les pays riches, les pays pauvres... et l'effet Guttman**

**Documents** : Tableau concernant la description du tableau élémentaire caractérisant les Etats du Monde

Tableaux 3 rendant compte des principaux résultats de l'ACP normée

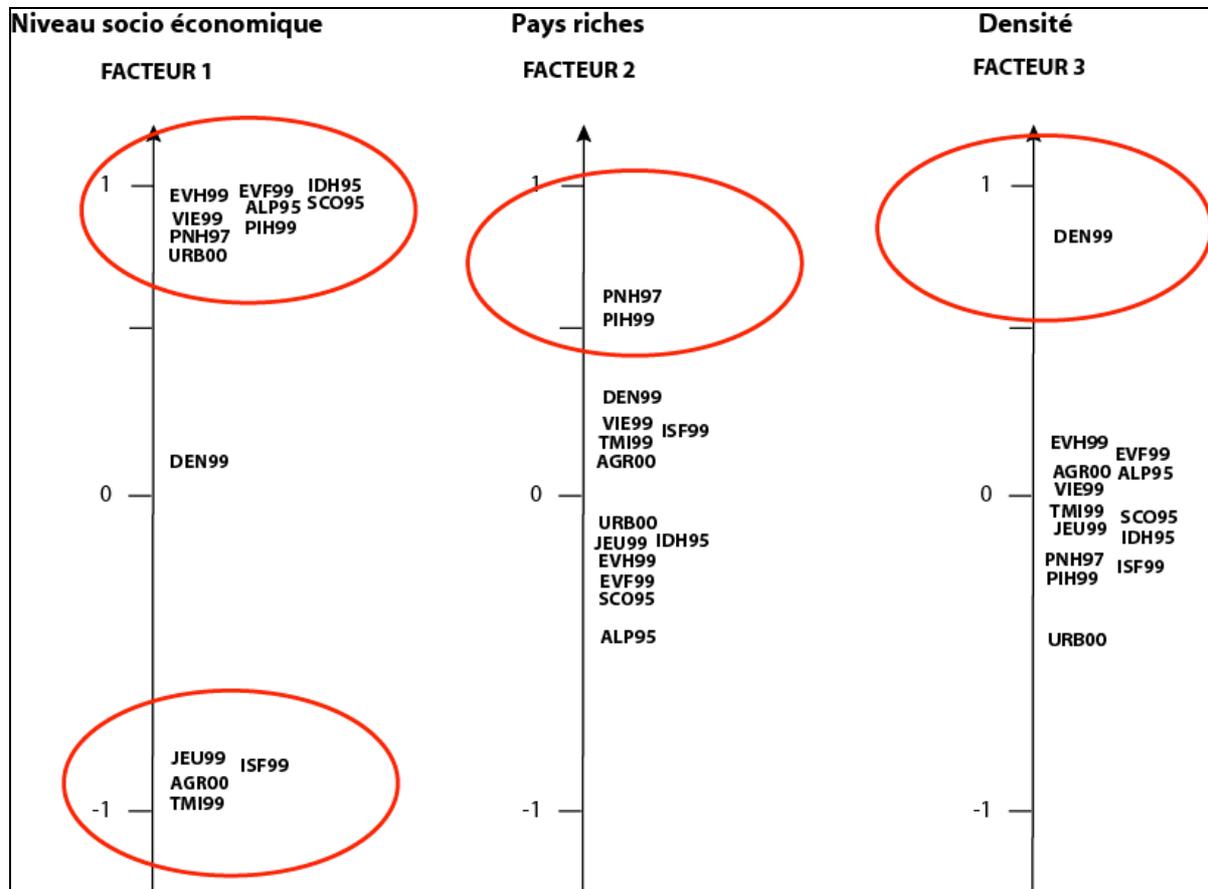
Graphiques concernant les positions des variables et des pays dans le premier plan factoriel

**Sources** : INED, 2000-2002

• *Pour connaître le nombre de facteurs à retenir pour analyser les résultats en détail de l'ACP normée, on étudie le **tableau des valeurs propres***. Ici nous avons 13 variables, un facteur devient plus intéressant qu'une variable lorsque sa valeur propre est supérieure à 1 ou lorsque la part de variance totale qu'il prend en compte est supérieure à 7,69% ( $100/13 = 7,69\%$  de la variance totale). Le premier facteur explique 68,14 % de la variance : c'est une part d'information très importante et on peut considérer que ce facteur présente une forte structure (peut-être est-elle triviale et assez évidente...il faut analyser la signification du facteur pour répondre à la question). Le 2<sup>e</sup> facteur porte 8,70 % et le 3<sup>e</sup> 8,15 % de la variance. Ces parts d'information sont nettement plus faibles mais méritent que nous nous attachions à connaître leur signification. Les trois premiers facteurs présentent tous une valeur propre supérieure à 1. On peut donc considérer que l'ensemble de ces 3 facteurs est porteur d'une information intéressante.

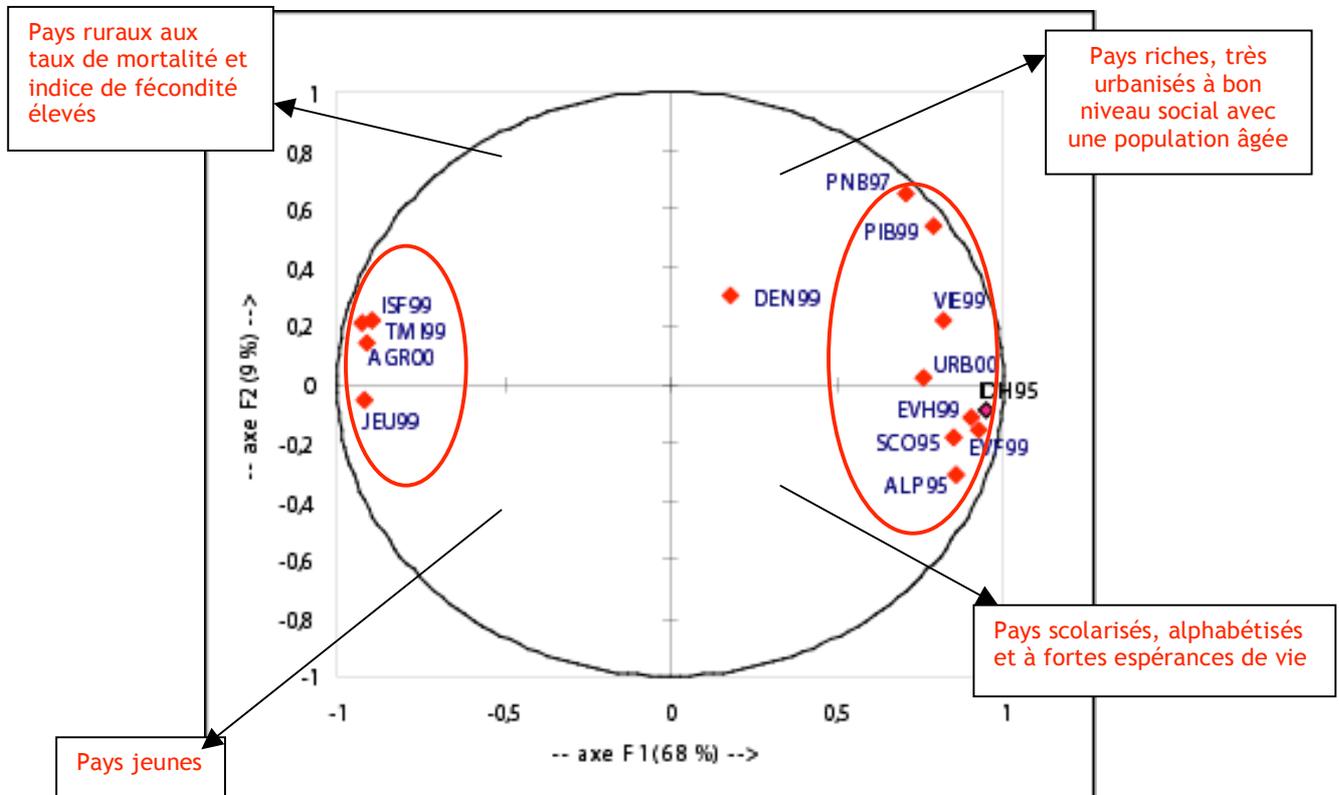
• Dans un second temps, on caractérise les axes factoriels en fonction des associations et des oppositions de variables, qui les définissent. Chaque coordonnée des variables sur chacun des 3 premiers facteurs correspond au coefficient de corrélation des variables sur chaque facteur. Par exemple, l'espérance de vie des femmes a une coordonnée de +0,928 sur le premier facteur alors que la part des agriculteurs dans la population active a, quant à elle, une coordonnée égale à -0,913. On le voit : les hypothèses de structuration de l'information grâce à l'étude de la matrice de corrélation étaient donc correctes. On remarque aussi que toutes les variables ont des coordonnées dont les valeurs absolues peuvent également être proches de 0 : autrement dit ces variables ont des positions qui sont proches du centre de gravité du nuage de points et ne jouent aucun rôle dans la structuration d'un facteur donné. C'est le cas de la densité de population sur le premier facteur (+0,181), de la part de population urbaine dans la population totale sur le deuxième facteur (+0,020) ou encore de l'espérance de vie des femmes sur le troisième facteur (+0,009). Enfin, on remarque la place de l'IDH95 en fin de tableau des coordonnées, séparé des autres variables par un trait : il s'agit de la place réservée aux variables supplémentaires. Ces variables ont la particularité d'être intégrées dans l'ACP normée mais de ne jouer aucun rôle actif, autrement dit d'avoir une contribution nulle dans les résultats de l'analyse de données. La valeur de sa coordonnée sur le premier facteur confirme les réserves que nous émettions sur les risques de redondance de l'information : +0,950. Cela signifie que cette variable se projette du même côté que les espérances de vie des femmes et des hommes, des taux de scolarisation et d'alphabétisation mais aussi des PIB et PNB.

*Graphiques concernant les trois premiers axes factoriels*

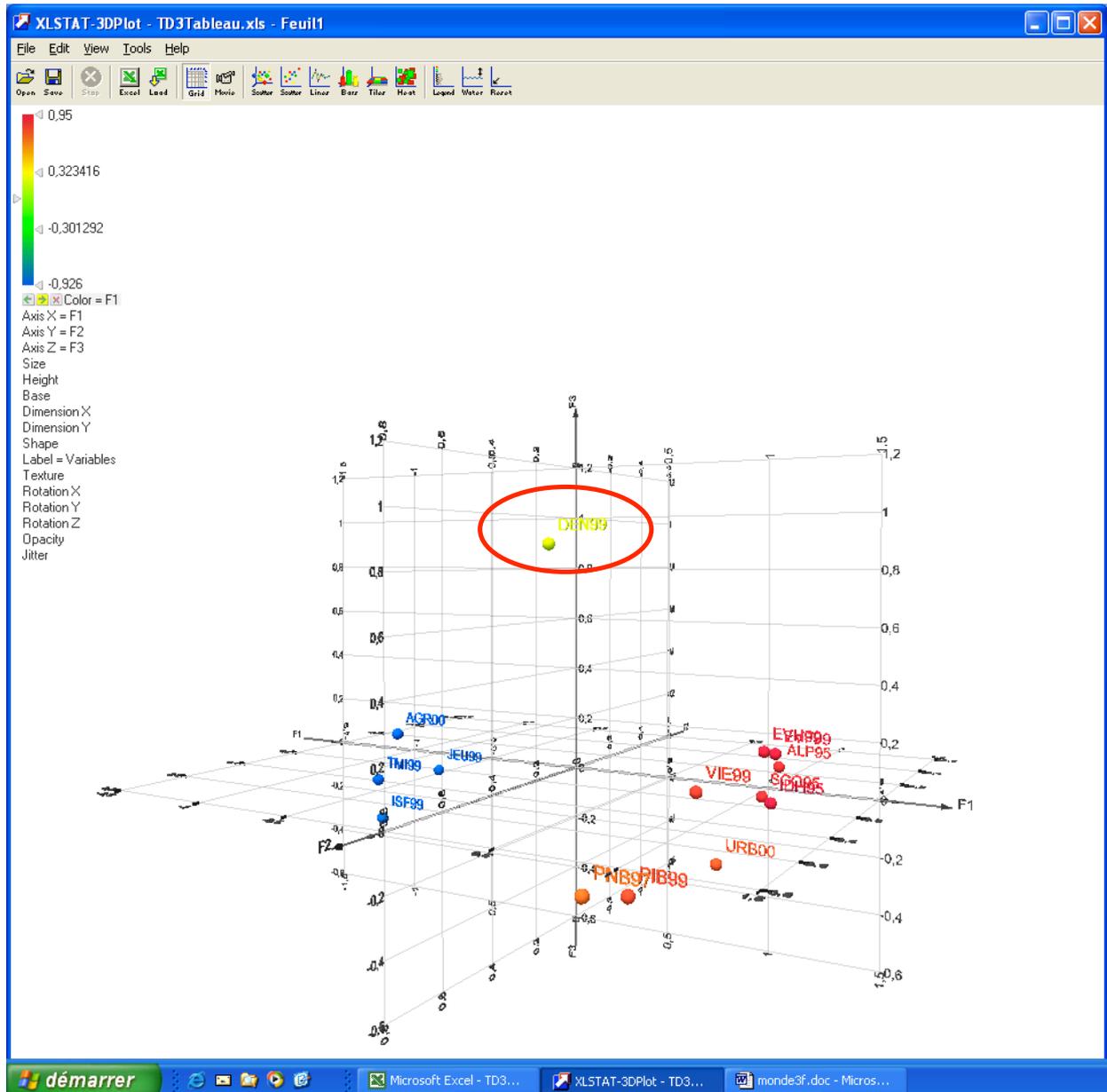


- Le premier facteur (ou axe factoriel ou composante principale) est caractérisé par l'opposition entre les Etats plutôt urbains, aux niveaux élevés socialement et économiquement et du bon niveau de santé, et Etats plus agricoles avec une population plutôt jeune, un indice synthétique de fécondité et un taux de mortalité infantile élevés. La première dimension mise en valeur par l'ACP est liée au niveau de richesse et rend compte également du stade d'avancement des pays dans la Transition démographique.
- Tandis que le deuxième facteur ne met en valeur qu'une seule association : PNB et PIB. Les pays dans lesquels ces 2 indicateurs de niveau de richesse sont élevés s'opposent à tous les autres pays. Les autres caractéristiques ne contribuent pas à la formation de ce facteur.
- Enfin, le troisième facteur présente une similitude avec le deuxième dans la mesure où il est défini par un phénomène qui se différencie de tous les autres. Il s'agit ici de la densité de population, qui oppose les Etats du monde à forte densité de population aux Etats peu densément peuplés.
- Quant au premier plan factoriel, il permet d'analyser 78% de l'information totale contenue dans le tableau initial de données. Ce plan factoriel a une allure particulière : les variables se situent plutôt aux extrémités du premier axe. Mises à part les variables PIB et PNB, l'ensemble des variables ne sont pas significatives sur le deuxième facteur. Ce cas particulier rend assez difficile la caractérisation du plan factoriel :

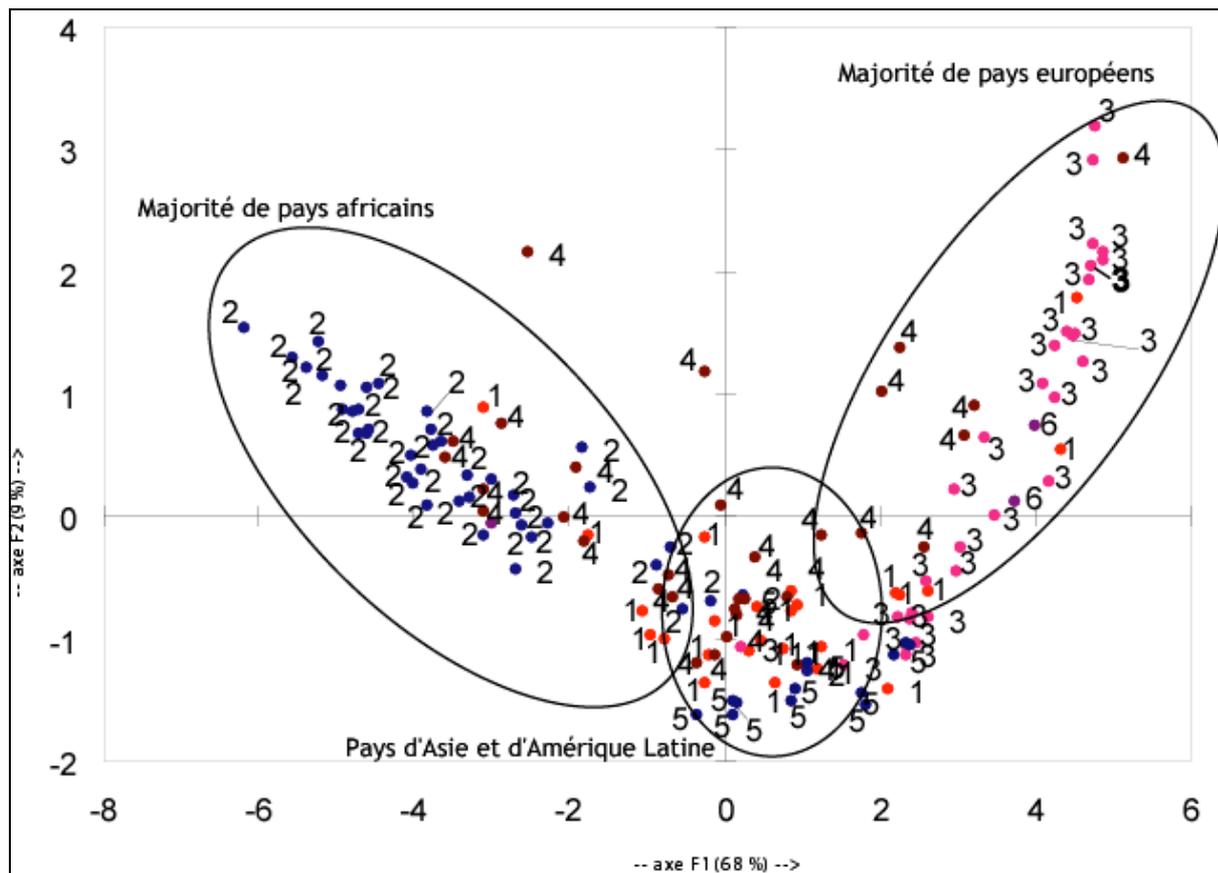
*Graphiques concernant le premier plan factoriel*



Pour rendre compte du caractère multidimensionnel de l'ACP normée, on peut montrer quelques exemples de visualisation des 3 premiers facteurs, grâce au module Xlstat-3Dplot. Dans le cas qui est proposé ci-dessous on comprend mieux le positionnement des variables dans un espace dont la dimension est supérieure à 3. Cette représentation se justifie d'autant plus dans le cas présent que les parts d'information prises en compte par les 2<sup>e</sup> et 3<sup>e</sup> facteurs sont très voisines. On se rend alors bien compte que la densité de population, qui ne joue aucun rôle dans la structuration du premier facteur, occupe une place majeure sur le 3<sup>e</sup> facteur.



Quant au graphique concernant le positionnement des pays dans le premier plan factoriel, il permet d'individualiser 3 « groupes » ou ensembles de pays. Un premier groupe correspondant à une majorité de pays africains correspondant aux pays plutôt ruraux et/ou jeunes. Le deuxième groupe plus compact décrit la situation d'une partie des pays d'Asie et d'Amérique latine qui sont caractérisés par une forte espérance de vie et de forts taux de scolarisation et d'alphabétisation. Enfin le troisième groupe concerne plutôt les pays européens, d'Amérique du Nord et tous les pays dits très développés. Au final, ces trois ensembles se positionnent bien les uns à la suite des autres, dessinant une parabole. Cette **courbe est caractéristique de l'effet Guttman** et traduit des **inter relations fortes entre les 146 pays**. Autrement dit ce qui caractérise un pays est fortement influencé par ce qui caractérise ses « voisins » ou les pays appartenant au même sous-ensemble continental.



### 3. Le grand jeu ... cartographique

**Documents** : Tableaux concernant les principaux résultats de la corrélation et de la régression entre les coordonnées des pays sur le premier facteur de l'ACP normée et leur valeur de l'IDH en 1995

Graphique concernant les résidus de la régression linéaire

Planches de cartes à faire ou à commenter mettant en regard les principaux résultats

**Sources** : INED, 1995-1999

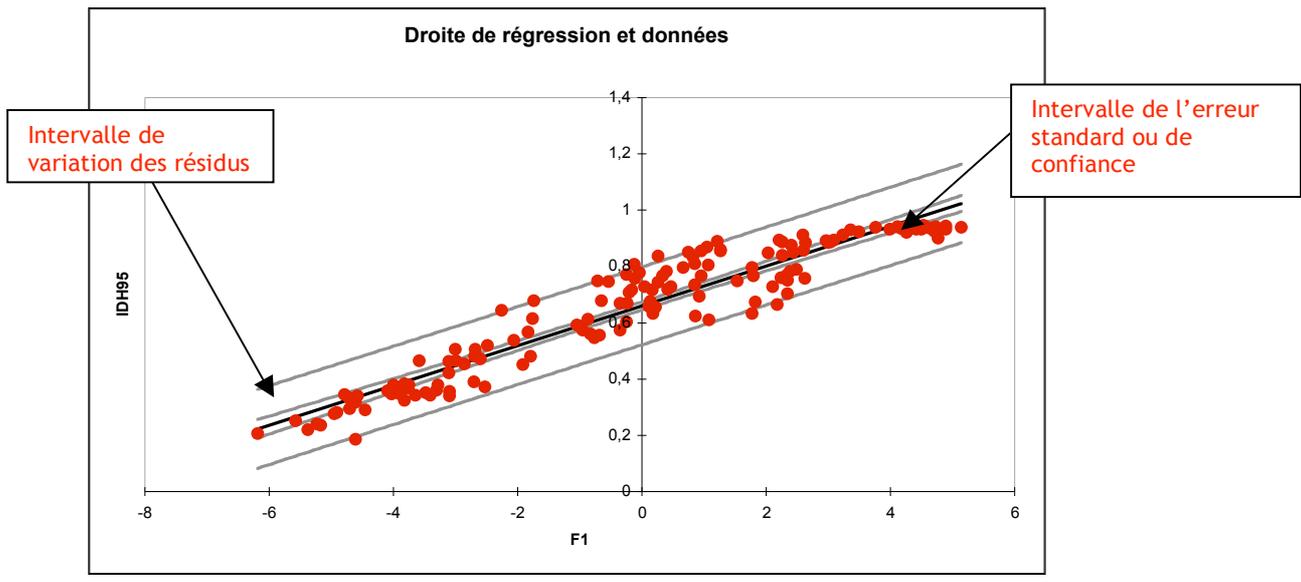
Au final, on peut être tenté de vérifier une bonne fois pour toute si les hypothèses que nous faisons initialement sur la redondance de l'information en conservant l'IDH dans le tableau élémentaire de données. On propose alors aux étudiants soit de commenter, soit de calculer la corrélation entre l'IDH95 pour chacun des 146 pays du Monde et leurs coordonnées respectives sur le 1<sup>er</sup> facteur de l'ACP normée. On obtient les résultats suivants :

<b>Coefficient de corrélation R</b>	<b>0,950</b>
<b>Coefficient de détermination R<sup>2</sup></b>	<b>0,902</b>
<b>Coefficient de détermination ajusté</b>	0,902

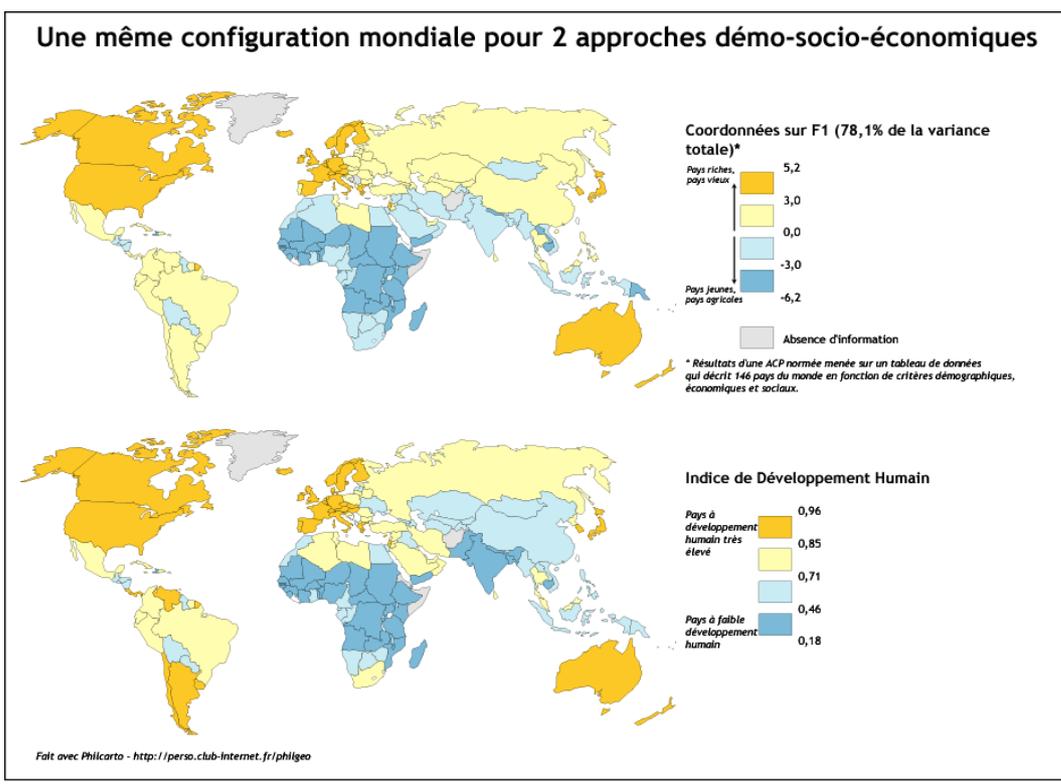
On est ainsi en mesure de conclure qu'il existe une relation positive et très significative entre les coordonnées des pays sur le premier facteur et les valeurs de l'IDH95 pour ces mêmes pays : 90% des variations de l'IDH95 dans les 146 pays retenus pour cette étude sont expliquées par la simple prise en compte de la coordonnée de chacun des pays sur le premier facteur de l'ACP normée. Autrement dit : il n'y a guère qu'un peu plus de 10% des variations de l'IDH95 dans ce même groupe de pays qui ne sont pas expliqués par leurs coordonnées sur le premier facteur. On peut alors en déduire l'équation de la droite de régression linéaire suivante :

Paramètre	Valeur	Ecart-type	t de Student	Pr > t
Constante	0,660	0,006	114,411	0,000
F1	0,071	0,002	36,467	0,000

**L'équation du modèle est :  $Y = 0,07 \cdot F1 + 0,66$**

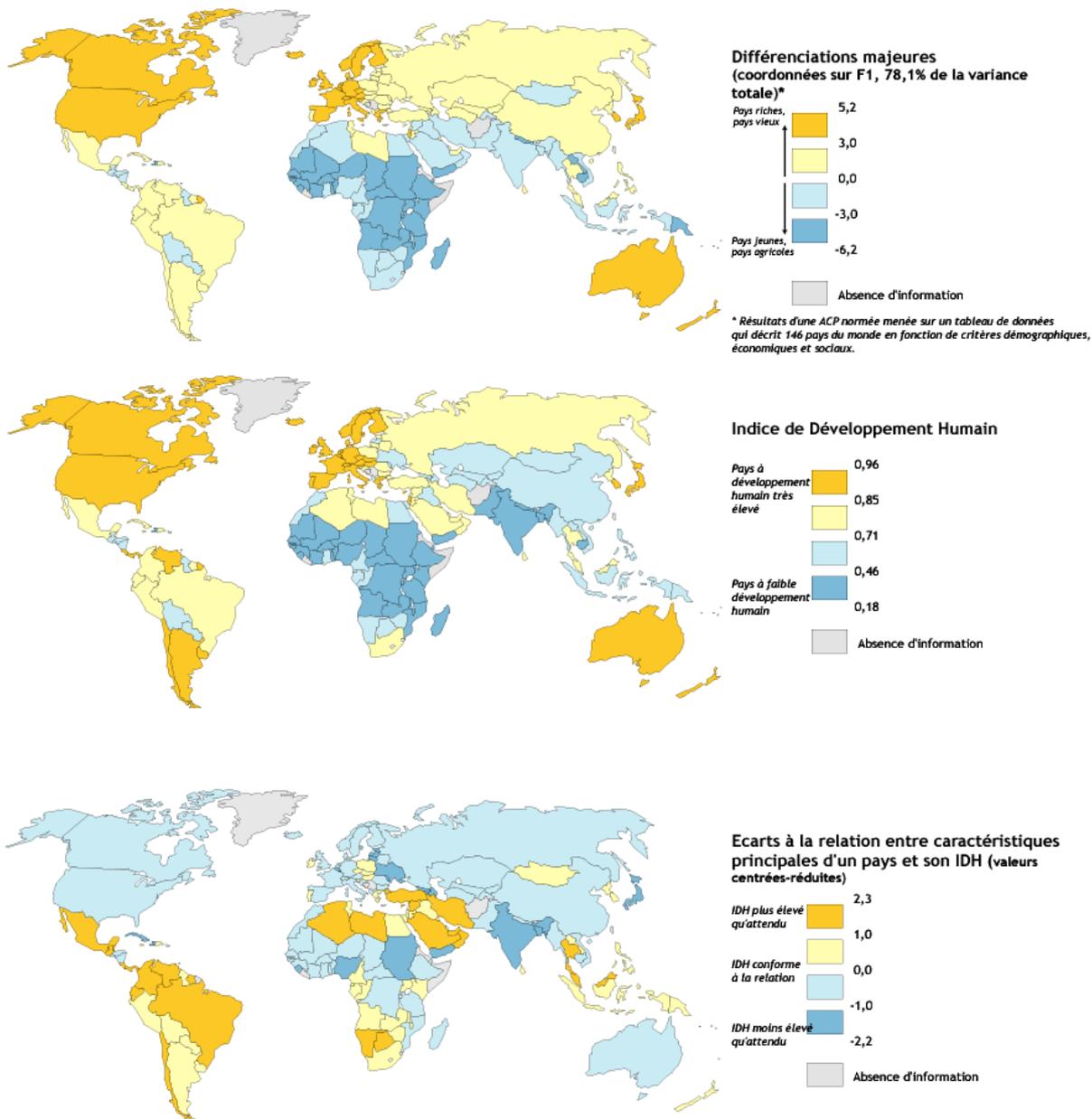


On peut enfin produire 3 cartes soit sur 2 planches différentes, soit sur une même planche. Sur une première planche, on décide de ne rendre compte que des similitudes de répartition entre coordonnées des pays sur le premier facteur et valeurs de l'IDH pour ces mêmes pays :



Sur une seconde planche, on décide de visualiser non seulement les similitudes de répartition entre coordonnées des pays sur le premier facteur et valeurs de l'IDH pour ces mêmes pays, mais aussi les résidus centrés-réduits de la relation :

## Une même configuration mondiale pour 2 approches démo-socio-économiques ?



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>

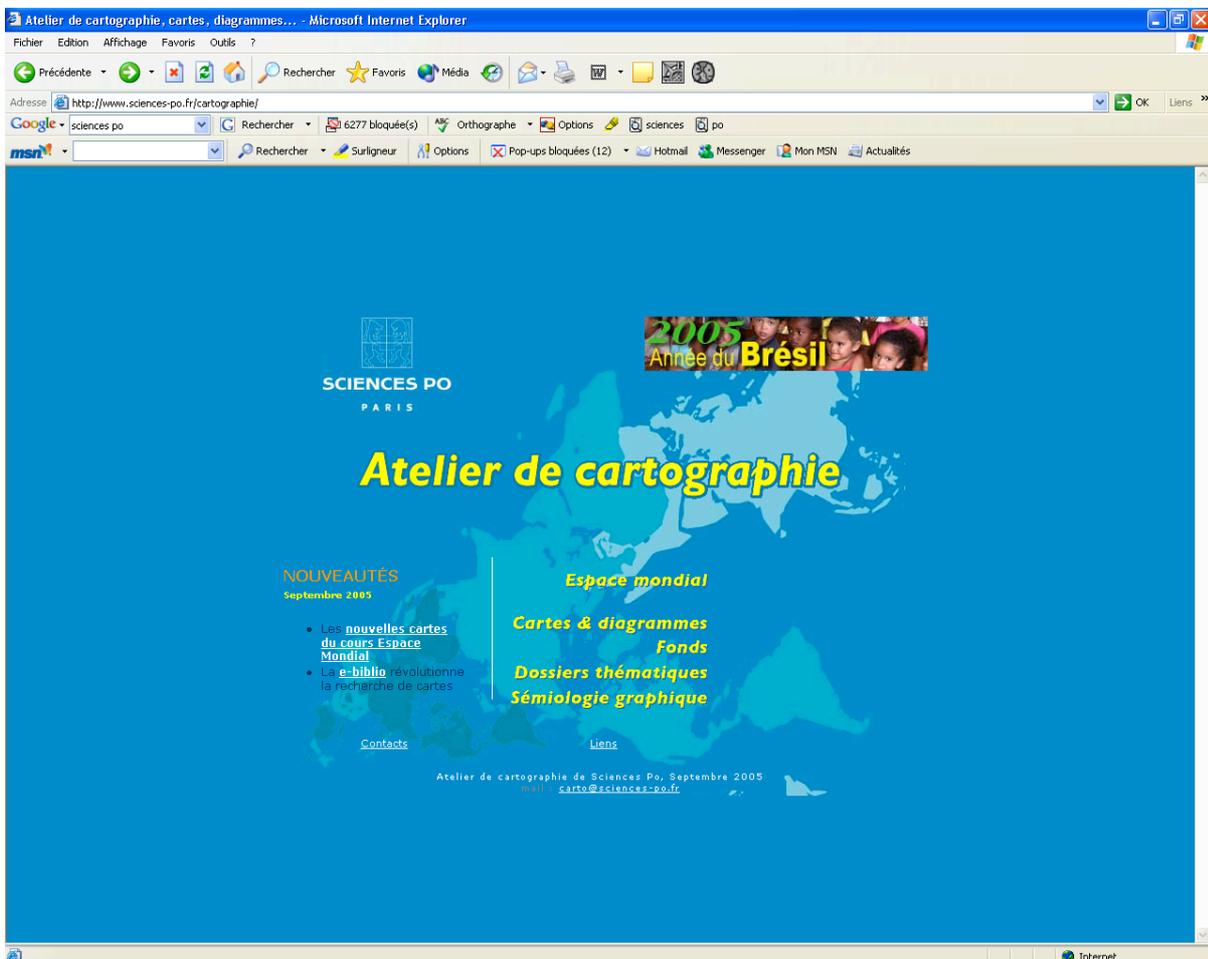
Dans ce dernier cas, on met en évidence qu'une partie des pays de l'Amérique latine mais aussi du Maghreb et du Proche-Orient sont caractérisés par un IDH bien supérieur à ce que laisserait attendre la simple prise en compte de leur situation décrite à partir des résultats de l'ACP normée (résidus centrés-réduits significativement positifs compris entre +1 et +2,3). A l'opposé, quelques pays d'Europe de l'Est comme les pays Baltes mais aussi d'Afrique comme le Nigéria, sans oublier l'Asie méridionale comme l'Inde mais aussi le Bangladesh, sont caractérisés par des valeurs d'IDH bien inférieures à ce que laisserait attendre la simple prise en compte de leur situation décrite à partir des résultats de l'ACP normée (résidus centrés-réduits significativement négatifs compris

entre -1 et -2,2). Ces écarts pourraient éventuellement être expliqués par le fait que les taux de scolarisation traités ici sont généraux et non détaillés par niveau de sortie du système éducatif (primaire, secondaire, supérieur). On peut alors émettre l'hypothèse que, dans le cas des pays d'Amérique latine par exemple, les taux de scolarisation désagrégés rendraient compte de situations beaucoup plus favorables que le taux global de scolarisation.

## Liens utiles pour étoffer et approfondir cet exemple

<http://www.sciences-po.fr/cartographie/>

Il s'agit du célèbre site de l'atelier de cartographie de Sciences Po : une véritable mine d'or. On ira consulter avec beaucoup d'intérêt les cartes qui concernent l'Indice de développement humain en 2002 et son évolution entre 1995 et 2002 ; mais aussi les supports de cours sur l'espace mondial sans oublier les liens vers d'autres sites.



**Myriam BARON, Christine ZANIN, Familiarisation à la réalisation et à l'interprétation des résultats d'une Analyse en Composantes Principales normées (ACP normée), Feuilles de Géographie, III-2006, Feuilles n° 58, 28 pages**

<http://www.grasland.cicrp.jussieu.fr/>

Il s'agit du site de Claude Grasland, en pleine évolution et restructuration. Dans les 2 parties qu'il distingue, on ira consulter avec beaucoup de profit les archives qui correspondent à ce qu'il a construit entre 1999 et 2004 et qu'il n'actualise pas et en particulier l'item « Géographie du Monde contemporain », dans lequel figurent non seulement les supports de cours mais aussi des idées de travaux à demander aux étudiants et enfin un répertoire des liens utiles pour compléter cet enseignement que nous ne reproduisons pas ici car cela ferait double emploi avec son site.



[https://dspace.msh-alpes.prd.fr/bitstream/1801/449/1/bk\\_multiv.pdf](https://dspace.msh-alpes.prd.fr/bitstream/1801/449/1/bk_multiv.pdf)

Sur ce site, vous trouverez au format pdf, ni plus ni moins qu'un manuel intitulé *Analyses multivariées de données géographiques* conçu par Pierre Dumolard. Ce manuel qui vient d'être mis en lignes à la Maison des Sciences de l'Homme des Alpes présente l'avantage d'être accessible plus facilement que l'ouvrage de Lena Sanders. Il présente enfin l'avantage, comme le livre de Lena Sanders, de n'être illustré que par des exemples pris en géographie.

## **Familiarisation à la réalisation et à l'interprétation des résultats d'une Analyse en Composantes Principales normées (ACP normée)**

### ***Les différenciations démographiques et économiques des Etats du Monde***

---

#### **Documents joints :**

- Description du tableau élémentaire caractérisant les Etats du Monde
- Tableaux des principaux résultats d'une Analyse en Composantes Principales sur l'ensemble des Etats du Monde
- Graphiques

#### **Documents à rendre :**

- ◆ Au moins une carte commentée de résultats d'ACP normée sur les pays du Monde

---

On analyse la situation sociale et économique des différents pays du monde en 1995-1999, les différenciations et les associations caractéristiques. Pour ce faire, on dispose de données collectées par l'INED et concernant les caractéristiques démographiques et économiques des pays du monde entre 1995 et 1999.

#### **I- ANALYSES STATISTIQUES UNIVARIÉE ET BIVARIEE**

A partir des tableaux 1 et 2 résumant le tableau élémentaire de données caractérisant les Etats du Monde

1. Vous indiquerez les variables qui introduisent le plus de différenciations entre les Etats, et celles qui en introduisent le moins. Vous justifierez votre réponse.
2. Vous construirez le ou les graphes de corrélation, en mettant en évidence les principales associations et oppositions entre les variables. Vous conclurez en émettant des hypothèses sur l'organisation de l'information lors de la réalisation d'une Analyse en Composantes Principales normée (ACP).
3. Vous indiquerez enfin si certains caractères sont redondants ou non avec d'autres. Si tel est le cas, vous préciserez le ou les caractères et justifierez votre réponse.
4. Vous émettrez des hypothèses sur les conséquences que l'existence d'éventuelles variables redondantes peuvent avoir dans la réalisation d'ACP normées.

#### **II- ANALYSE EN COMPOSANTES PRINCIPALES NORMEES**

A partir des tableaux 3 et des graphiques 1 regroupant les principaux résultats d'une ACP sur l'ensemble des pays du Monde, vous tenterez de répondre à la question suivante : **que vous apprend l'ensemble de ces résultats sur la situation des pays du monde en 1995-1999 ?**

1. Vous justifierez du nombre de facteurs retenus, en ayant soin de justifier complètement votre réponse.

2. Vous définirez ce que l'on entend par coordonnées, qualité de représentation et contribution des variables pour chacun des facteurs. Vous concluez sur la pertinence de l'ACP menée en fonction de ces 3 résultats.
3. Vous tenterez de comprendre pourquoi l'indice de développement humain de 1995 (IDH95) possède une coordonnée sur chacun des facteurs, a une qualité de représentation mais n'a pas de contribution.
4. Vous construirez et caractériserez les axes factoriels retenus, en leur donnant un titre.
5. Vous préciserez enfin ce que vous apprend l'analyse du premier plan factoriel quant au positionnement des pays de cette étude. Vous indiquerez en particulier où se projette l'IDH95 par rapport aux autres variables.

### **III- CARTOGRAPHIE DES RESULTATS D'UNE ANALYSE EN COMPOSANTES PRINCIPALES NORMEES**

A partir des résultats de l'ACP menée sur les Etats du monde, vous réaliserez au moins 1 représentation cartographique de ces résultats, en accordant un soin particulier aux variables visuelles, à la présentation de la légende et plus généralement à l'habillage de votre document.

Cette carte peut être réalisée à l'aide du logiciel Philcarto. Attention à la correspondance entre les codes des individus dans le fichier fond de carte et le fichier Excel de données.

1. Vous pourrez mettre en page votre carte à l'aide du logiciel Adobe Illustrator en accordant un soin particulier à l'habillage.
2. Pour comparaison, vous réaliserez une carte de l'Indice de Développement Humain (IDH95) dans les pays du Monde.

Vous ferez un commentaire géographique précis et concis dans lequel vous apporterez des réponses aux questions suivantes :

1. Quelle(s) est(sont) la(les) configuration(s) territoriale(s) mise(s) en évidence par les résultats de l'ACP ?
2. Quelle(s) est(sont) la(les) ressemblance(s) et la(les) différence(s) entre la représentation cartographique des résultats de l'ACP et celle de l'IDH ?
3. Quelle(s) est(sont) la(les) hypothèse(s) explicative(s) que vous avanceriez ?

**Tableau 1 : Description du contenu du tableau concernant les Etats du Monde**

Code de la variable	Description de la variable
id	Identificateur du pays à 3 chiffres, le premier chiffre à gauche correspondant au continent ou sous-continent : 1 pour l'Amérique, 2 pour l'Afrique, 3 pour l'Europe, 4 pour l'Asie, 5 pour les anciennes républiques soviétiques et 6 pour l'Océanie
nom	nom du pays
DEN99	densité de population (habitants/km <sup>2</sup> ) en 1999
URB00	part de la population urbaine dans la population totale (%) en 2000
AGR00	part des agriculteurs dans la population active (%) en 2000
JEU99	part des moins de 15 ans dans la population totale (%) en 1999
VIE99	part des plus de 60 ans dans la population totale (%) en 1999
TMI99	taux de mortalité infantile (pour 1000 naissances) en 1999
EVH99	espérance de vie des hommes (ans) en 1999
EVF99	espérance de vie des femmes (ans) en 1999
PNB97	produit national brut en 1997
PIB99	produit intérieur brut en 1997
ALP95	taux d'alphabétisation en 1995
SCO95	taux de scolarisation en 1995
ISF99	indice synthétique de fécondité (nombre d'enfants par femme en âge de procréer) en 1999
IDH95	indice de développement humain en 1995

Source : INED 1999, 2000

**Tableau 2.1. Moyennes et écart-types des phénomènes caractérisant les Etats du Monde**

Variabes	Codes	Moyennes	Ecart-types	Coefficients de variation
espérance de vie des hommes (ans) en 1999	EVH99	62,23	11,01	0,18
espérance de vie des femmes (ans) en 1999	EVF99	66,98	12,53	0,19
taux d'alphabétisation en 1995	ALP95	77,83	22,69	0,29
taux de scolarisation en 1995	SCO95	62,77	19,28	0,31
part des moins de 15 ans dans la population totale (%) en 1999	JEU99	33,41	10,50	0,31
indice de développement humain en 1995	IDH95	0,66	0,22	0,34
part de la population urbaine dans la population totale (%) en 2000	URB00	56,37	22,25	0,39
indice synthétique de fécondité (nombre d'enfants par femme en âge de procréer) en 1999	ISF99	3,48	1,83	0,53
part des plus de 60 ans dans la population totale (%) en 1999	VIE99	6,85	4,67	0,68
part des agriculteurs dans la population active (%) en 2000	AGR00	33,70	27,27	0,81
taux de mortalité infantile (pour 1000 naissances) en 1999	TMI99	45,62	39,91	0,87
produit intérieur brut en 1997	PIH99	6162,89	6545,85	1,06
densité de population (hab/km <sup>2</sup> ) en 1999	DEN99	88,61	111,51	1,26
produit national brut en 1997	PNH97	5964,32	9603,62	1,61

**Tableau 2.2** Matrice des corrélations

Codes des variables	DEN99	URB00	AGR00	JEU99	VIE99	TMI99	EVH99
DEN99	1						
URB00	0,002	1					
AGR00	-0,092	-0,824	1				
JEU99	-0,200	-0,621	0,767	1			
VIE99	0,146	0,519	-0,648	-0,908	1		
TMI99	-0,155	-0,655	0,856	0,807	-0,669	1	
EVH99	0,218	0,656	-0,840	-0,751	0,600	-0,905	1
EVF99	0,181	0,678	-0,866	-0,806	0,665	-0,922	0,983
PNB97	0,168	0,552	-0,561	-0,634	0,664	-0,522	0,566
PIB99	0,122	0,640	-0,656	-0,704	0,694	-0,624	0,654
ALP95	0,052	0,585	-0,768	-0,773	0,655	-0,827	0,720
SCO95	0,061	0,630	-0,804	-0,695	0,615	-0,789	0,727
IDH95	0,091	0,757	-0,899	-0,813	0,676	-0,909	0,894
ISF99	-0,239	-0,551	0,767	0,907	-0,774	0,868	-0,788

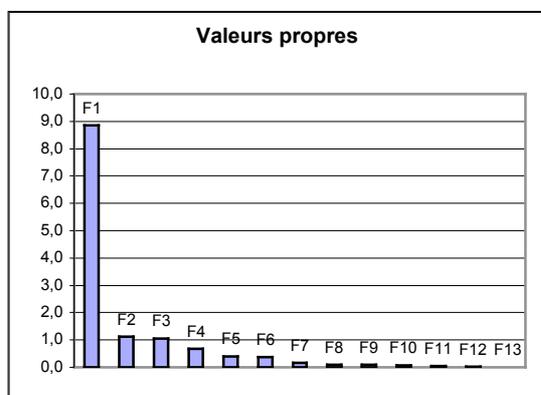
*En gras, valeurs significatives (hors diagonale) au seuil alpha=0,050 (test bilatéral)*

Codes des variables	EVF99	PNB97	PIB99	ALP95	SCO95	IDH95	ISF99
DEN99							
URB00							
AGR00							
JEU99							
VIE99							
TMI99							
EVH99							
EVF99	1						
PNB97	0,547	1					
PIB99	0,639	0,950	1				
ALP95	0,773	0,435	0,527	1			
SCO95	0,742	0,521	0,616	0,823	1		
IDH95	0,900	0,630	0,751	0,843	0,860	1	
ISF99	-0,840	-0,451	-0,535	-0,847	-0,759	-0,818	1

*En gras, valeurs significatives (hors diagonale) au seuil alpha=0,050 (test bilatéral)*

**Tableaux 3 : Principaux résultats de l'Analyse en Composantes Principales concernant les caractéristiques démographiques des Etats du Monde**

**Tableau 3.1 : Les valeurs propres**



	F1	F2	F3
Valeur propre	8,858	1,131	1,059
% variance	68,140	8,697	8,146
% cumulé	68,140	76,837	84,983

**Tableau 3.2 : Coordonnées des variables sur les différents facteurs**

Variables	F1	F2	F3	F4
DEN99	0,181	0,305	0,892	-0,215
URB00	0,763	0,020	-0,299	-0,355
AGR00	-0,913	0,137	0,105	0,213
JEU99	-0,915	-0,058	-0,095	-0,299
VIE99	0,816	0,213	0,028	0,453
TMI99	-0,926	0,212	-0,068	0,068
EVH99	0,902	-0,115	0,100	-0,247
EVF99	0,928	-0,161	0,083	-0,156
PNB97	0,707	0,652	-0,173	-0,021
PIB99	0,790	0,536	-0,210	-0,046
ALP95	0,854	-0,306	-0,009	0,166
SCO95	0,849	-0,180	-0,095	-0,007
ISF99	-0,894	0,215	-0,222	-0,233
IDH95	0,950	-0,093	-0,087	-0,114

**Tableau 3.3 : Qualité de représentation des variables sur les différents facteurs**

Variables	F1	F2	F3	F4
DEN99	0,033	0,093	0,795	0,046
URB00	0,582	0,000	0,089	0,126
AGR00	0,833	0,019	0,011	0,046
JEU99	0,837	0,003	0,009	0,089
VIE99	0,666	0,045	0,001	0,205
TMI99	0,858	0,045	0,005	0,005
EVH99	0,814	0,013	0,010	0,061
EVF99	0,862	0,026	0,007	0,024
PNB97	0,500	0,425	0,030	0,000
PIB99	0,624	0,288	0,044	0,002
ALP95	0,729	0,094	0,000	0,027
SCO95	0,722	0,032	0,009	0,000
ISF99	0,800	0,046	0,049	0,054
IDH95	0,902	0,009	0,008	0,013

**Tableau 3.4 : Contributions des variables sur les différents facteurs (%)**

Variables	F1	F2	F3	F4
DEN99	0,369	8,247	75,068	6,726
URB00	6,570	0,035	8,414	18,404
AGR00	9,400	1,655	1,048	6,633
JEU99	9,446	0,302	0,855	12,998
VIE99	7,522	4,013	0,074	29,851
TMI99	9,685	3,965	0,439	0,666
EVH99	9,186	1,176	0,947	8,902
EVF99	9,728	2,304	0,649	3,544
PNB97	5,642	37,592	2,811	0,066
PIB99	7,041	25,435	4,158	0,312
ALP95	8,234	8,307	0,008	4,000
SCO95	8,146	2,867	0,855	0,007
ISF99	9,031	4,103	4,673	7,892
IDH95	0,000	0,000	0,000	0,000

**Graphiques 1. Le premier plan factoriel**

