

# ANALYSE EXPLORATOIRE D'UNE SUITE DE TABLEAUX DE DONNEES INDICES PAR LE TEMPS

**Catherine Pardoux**

*Université PARIS IX-DAUPHINE  
LISE-CEREMADE  
Place du M<sup>re</sup> de Lattre de Tassigny  
75775 – Paris cedex 16  
Email : Catherine.Pardoux@dauphine.fr*

Dans le cadre de l'analyse exploratoire des tableaux multiples, nous nous intéressons dans cet article à une succession de tableaux « individus  $\times$  variables » indicés par le temps.

Le théorème d'Eckart et Young ([LEB99]) sur lequel reposent les méthodes factorielles n'admet pas de généralisation en ce sens qu'il n'existe pas de décomposition optimale unique d'un tableau à trois entrées en tableaux de rangs 1.

Plusieurs méthodologies de traitement vont être présentées et comparées à l'aide d'un exemple, notamment pour le suivi (construction des trajectoires) des individus, des modalités et des variables.

## 1. Introduction

Considérons  $T$  tableaux de données, indicés par le temps.

Lorsque tous les tableaux sont définis avec les mêmes variables (au nombre de  $p$ ), on peut représenter, dans l'espace  $\mathbb{R}^p$ , la succession des  $T$  tableaux par  $T$  nuages d'individus, associer aux individus communs leur trajectoire, et représenter les trajectoires de centres de gravité.

Lorsque tous les tableaux sont définis avec les mêmes individus (au nombre de  $n$ ), on peut représenter, dans l'espace  $\mathbb{R}^n$ , la succession des  $T$  tableaux par  $T$  nuages de variables, associer aux variables communes leur trajectoire, et calculer les  $T$  matrices de covariances.

Un ensemble de  $T$  tableaux doublement appariés (mêmes individus, mêmes variables) est un ensemble de données à trois indices (un cube) sur lequel on peut opérer plusieurs analyses de tableaux à deux indices :  $T$  tableaux individus  $\times$  variables,  $p$  tableaux individus  $\times$  temps,  $n$  tableaux variables  $\times$  temps. Chacun de ces ensembles de tableaux peut être juxtaposé horizontalement ou verticalement.

L'objectif étant ici l'étude de l'évolution, on choisit l'analyse des  $T$  tableaux individus  $\times$  variables.

Le traitement va être envisagé selon deux optiques :

1. ne pas tenir compte de la structure induite par le temps pour le tableau global (formé des  $T$  tableaux juxtaposés), et interpréter les résultats a posteriori à l'aide cette structure,
2. tenir compte a priori de la structure en faisant
  - une étude de l'« interstructure », c'est-à-dire une analyse des différences entre instants,
  - une étude de l'« intrastructure », c'est-à-dire une analyse des ressemblances – ou des différences - au cours de l'évolution.

Notons que la prise en compte de la structure n'implique pas la prise en compte de l'ordre chronologique.

Quatre méthodologies vont être présentées et comparées.

La première approche consiste à faire l'analyse factorielle du tableau moyen qui peut être considéré comme un tableau « compromis », et à utiliser la technique de projections de points supplémentaires ([CAZ90]). Elle permet le suivi dans le temps des individus et des variables, et la construction des trajectoires de modalités illustratives à condition d'introduire les centres de gravité des modalités en éléments supplémentaires.

La seconde approche consiste à faire l'analyse factorielle du tableau de données obtenu par juxtaposition verticale des  $T$  tableaux. Elle ne nécessite pas l'observation des mêmes individus à chaque date. Elle permet une analyse « globale », une analyse « interclasse » et une analyse « intraclasse » ([CAZ90]). Sur les plans factoriels, il y a pour chaque unité

statistique (observée à tous les instants) autant de points que d'unités de temps. On peut donc suivre l'évolution de chaque unité (active ou supplémentaire) observée à tous les instants et construire sa trajectoire. On peut aussi construire les trajectoires d'une modalité illustrative (centre de gravité des unités prenant cette modalité). Cette seconde approche ne permet pas le suivi des variables.

L'analyse du tableau de données obtenu par juxtaposition horizontale des  $T$  tableaux ne nécessite pas l'observation des mêmes variables à chaque date. On lui préférera l'Analyse Factorielle Multiple [ESC98] qui est une analyse factorielle de l'ensemble des  $T$  tableaux juxtaposés horizontalement, mais avec une pondération des variables afin d'équilibrer l'influence des groupes. L'AFM constitue la troisième approche. Elle réalise directement les études de l'« interstructure » et de l'« intrastructure ». Elle permet le suivi des variables et le suivi « intraclasse » des modalités et des individus.

La quatrième approche repose sur la méthode « STATIS » dont l'objectif essentiel est la recherche d'une structure commune aux tableaux de données ([LAV88]). Cette méthode ne nécessite pas l'observation des mêmes variables à chaque instant. La méthode « STATIS duale » ne nécessite pas l'observation des mêmes individus à chaque instant.

Toutes les analyses factorielles ont été réalisées avec logiciel SPAD.

## 2. Présentation des données

### 2.1. Les trois tableaux de données indicés par le temps

Les tableaux de données traités (extraits de [JAM89]) concernent l'étude des composantes de la qualité de service du réseau téléphonique par région en France métropolitaine.

Chaque tableau (cf. annexe) comporte vingt-deux lignes correspondant aux vingt-deux régions « Télécom » identifiées par le nom de la ville représentative de la région, et sept colonnes correspondant à sept indicateurs mesurant la qualité de service offerte aux usagers : six indicateurs élémentaires et un indice global de qualité de service (noté IGQS), combinaison d'indicateurs élémentaires (cf. annexe). Ces indicateurs, représentant des pourcentages, des taux de signalisation, des vitesses de relevé..., définissent un ensemble de variables hétérogènes.

On dispose de trois tableaux de ce type pour trois années successives : 1984, 1985 et 1986, et d'un autre tableau (pour 1984) avec huit indicateurs technico-économiques. Les variables de ce dernier tableau peuvent être utilisées comme éléments illustratifs.

Ces trois tableaux de données sont particuliers en ce sens que les variables et les individus sont les mêmes pour les trois années. Lorsqu'on dispose de données évolutives, ces conditions ne sont bien souvent pas satisfaites.

## 2.2. Individus supplémentaires

La variable IGQS étant une variable synthétique, une première analyse exploratoire des données a consisté à représenter les boîtes-à-pattes en parallèle de cette variable pour les trois années (figure 1). La valeur d'IGQS pour Ajaccio apparaît comme « valeur extrême » pour chacune des trois années. C'est la raison pour laquelle cette région ne sera pas prise en élément actif dans les analyses factorielles et sera introduite en élément supplémentaire.

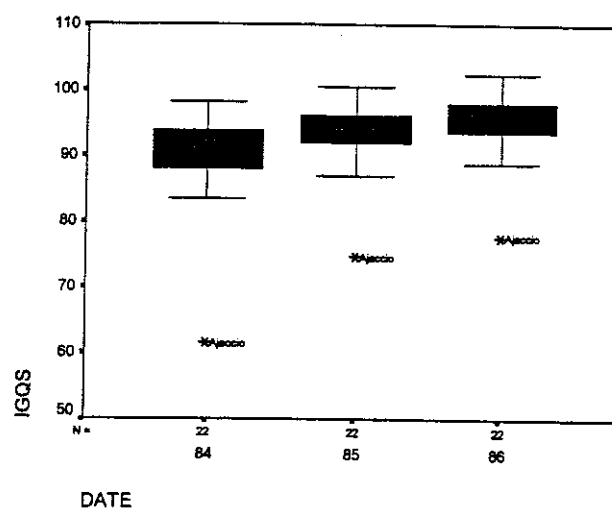


Figure 1 : Boîtes-à-pattes en parallèle (logiciel SPSS)

Pour chaque année, trois unités statistiques fictives ont été créées pour être introduites en éléments supplémentaires : le centre de gravité (ou point moyen), ainsi que le maximum et le minimum, calculés sur les vingt et une régions actives (figure 2). Ces deux dernières unités statistiques servent de cadre pour les positions des régions.

Pour chaque année, on dispose de quatre « individus supplémentaires » dont on peut construire les trajectoires : « Ajaccio », le centre de gravité « G », le maximum « Max » et le minimum « Min ».

### 2.3. Variable nominale illustrative

Une variable nominale à cinq modalités est définie par l'indicatif téléphonique régional : 1 pour l'Ile-de-France, 2 pour l'ouest, 3 pour le nord-est, 4 pour le sud-est et 5 pour le sud-ouest. Cette variable sera illustrative, et on construira les trajectoires de ses modalités, c'est-à-dire des centres de gravité de chaque modalité, quand la méthodologie le permettra.

### 3. Notations

A chaque instant  $t$ , on dispose du tableau  $X_t$  des observations de l'ensemble  $J_t$  des  $p_t$  variables sur l'ensemble  $I_t$  des  $n_t$  individus (figure 2). A chaque tableau, on ajoute trois éléments supplémentaires : le centre de gravité (ou point moyen), le maximum et le minimum calculés sur les  $n_t$  unités statistiques actives de chaque variable  $j$  (§2.2).

		$J_t$		
		1	$j$	$p_t$
$I_t$	1	$x_{ijt}$		
	$i$			
	$n_t$			
	$G_t$			
	$Max_t$			
	$Min_t$			

Figure 2 : Le tableau  $X_t$

### 4. Analyse factorielle du tableau moyen

Les trois tableaux de données de l'exemple étant doublement appariés (mêmes individus, mêmes variables), on peut calculer le tableau moyen  $X_m$ , et juxtaposer à ce tableau de données les trois tableaux  $X_t$  en lignes et en colonnes. Dans ce cas,  $n_t$  et  $p_t$  sont constants et respectivement égaux à  $n$  et  $p$  (figure 3). Les huit variables du tableau des indicateurs économiques de 1984 peuvent être introduites en éléments illustratifs.

Cette première approche consiste à faire une analyse en composantes principales normée, à cause de l'hétérogénéité des variables, du tableau moyen  $X_m$  et à utiliser la technique de projections de points supplémentaires ([CAZ90]).

Amiens	$X_m$	$X_1$	$X_2$	$X_3$	Données Économiques 1984
Paris Amiens 84					
Paris 84 Amiens 85	$X_1$	$X_2$	$X_3$		
Paris 85 Amiens 86					
Paris 86 Ajaccio Ajaccio 84	$X_1$	$X_2$	$X_3$		
Min 86					

Figure 3 : Trois tableaux doublement appariés avec tableaux supplémentaires

Les résultats de l'ACP du tableau moyen  $X_m$  montrent qu'il est pertinent de retenir trois axes factoriels. Avec ces trois axes, on prend en compte plus de 90% de l'inertie totale

NUMERO	VALEUR PROPRE	POURCENT	POURCENT CUMULE
1	4.0981	58.54	58.54
2	1.2867	18.38	76.93
3	0.9503	13.58	90.50
4	0.3277	4.68	95.18
5	0.2394	3.42	98.60
6	0.0830	1.19	99.79
7	0.0148	0.21	100.00

La 3<sup>ème</sup> valeur propre n'étant pas significativement différente de la 2<sup>ème</sup> valeur propre au vu des intervalles de confiance de niveau 95%, les axes correspondants sont définis à une rotation près ([LEB99]).

INTERVALLES LAPLACIENS D'ANDERSON, INTERVALLES AU SEUIL 0.95

NUMERO	BORNE INFERIEURE	VALEUR PROPRE	BORNE SUPERIEURE
1	1.5581	4.0981	6.6381
2	0.4892	1.2867	2.0843
3	0.3613	0.9503	1.5393

L'examen de la matrice des corrélations montre que les quatre variables IZAA, EZAA, TSI et VR2 sont fortement corrélées à l'indice global IGQS, et que tous les indicateurs varient dans le même sens que l'indice global IGQS, à l'exception de TSI et TCOM.

MATRICE DES CORRELATIONS

	IGQS	IZAA	EZAA	TSI	VR2	TCR	TCOM
IGQS	1.00						
IZAA	0.56	1.00					
EZAA	0.80	0.80	1.00				
TSI	-0.90	-0.52	-0.61	1.00			
VR2	0.86	0.27	0.57	-0.80	1.00		
TCR	0.38	0.71	0.69	-0.36	0.31	1.00	
TCOM	0.01	-0.01	-0.14	-0.04	-0.04	-0.23	1.00

Le 1<sup>er</sup> facteur résume à lui seul plus de 58% de l'information. L'ensemble des variables du sous-groupe défini par les six premières variables étant toutes corrélées fortement au premier

facteur, celui-ci apparaît comme un facteur de qualité globale générale (figure 4). Une ACP réalisée sans la variable synthétique IGQS ayant donné des résultats très voisins (52% de l'information résumée par le 1<sup>er</sup> facteur lié à IGQS avec un coefficient de corrélation égal à 0,89), la structure engendrée par le sous-groupe des variables IZAA, EZAA, TSI, VR2 et TCR apparaît très forte.

Les facteurs 2 et 3, définis à une rotation près, introduisent une dimension supplémentaire pour la mesure de la qualité de service du réseau téléphonique (figure 5). Ils sont principalement liés à la variable TCOM qui a dans le plan factoriel 2-3 une qualité de représentation égale à 0,98 (les autres variables ayant toutes des qualités de représentation inférieures à 0,4). C'est cette variable « temps d'établissement des communications », non corrélée avec le 1<sup>er</sup> facteur, qui doit être prise en compte en sus du 1<sup>er</sup> facteur pour expliquer la qualité de service du réseau téléphonique dans les 21 régions (Ajaccio n'étant pas élément actif dans l'analyse).

COORDONNEES DES VARIABLES SUR LES AXES 1 A 5 : VARIABLES ACTIVES

VARIABLES	COORDONNEES				
	1	2	3	4	5
IGQS - IGQS	-0.93	-0.31	-0.10	-0.12	0.09
IZAA - IZAA	-0.77	0.38	0.40	-0.23	-0.13
EZAA - EZAA	-0.91	0.22	0.09	-0.13	0.30
TSI - TSI	0.86	0.36	0.09	0.05	0.33
VR2 - VR2	-0.79	-0.43	-0.32	0.23	0.09
TCR - ICR	-0.68	0.58	0.15	0.41	-0.06
TCOM - TCOM	0.09	-0.58	0.80	0.13	0.05



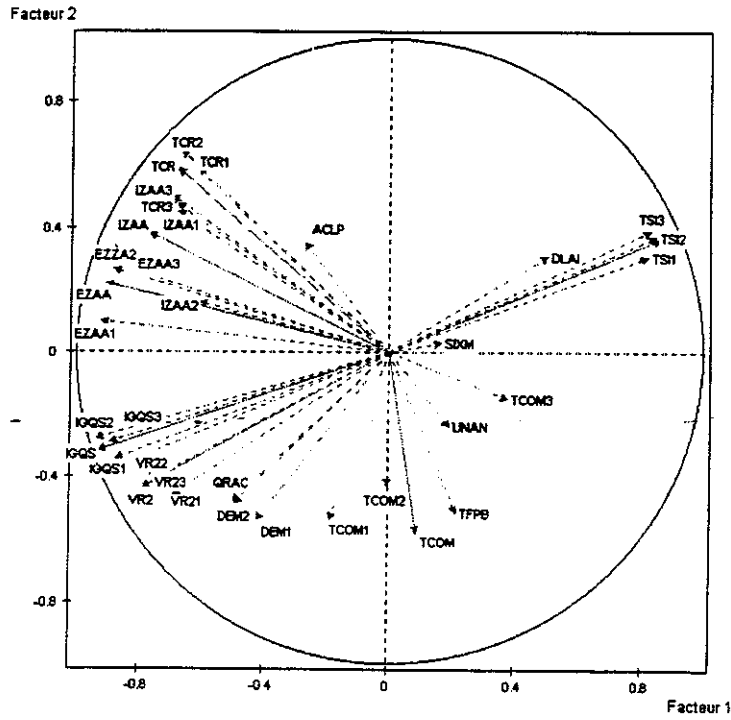


Figure 4 : Cercle des corrélations du 1<sup>er</sup> plan factoriel avec 7 variables actives et 29 variables illustratives

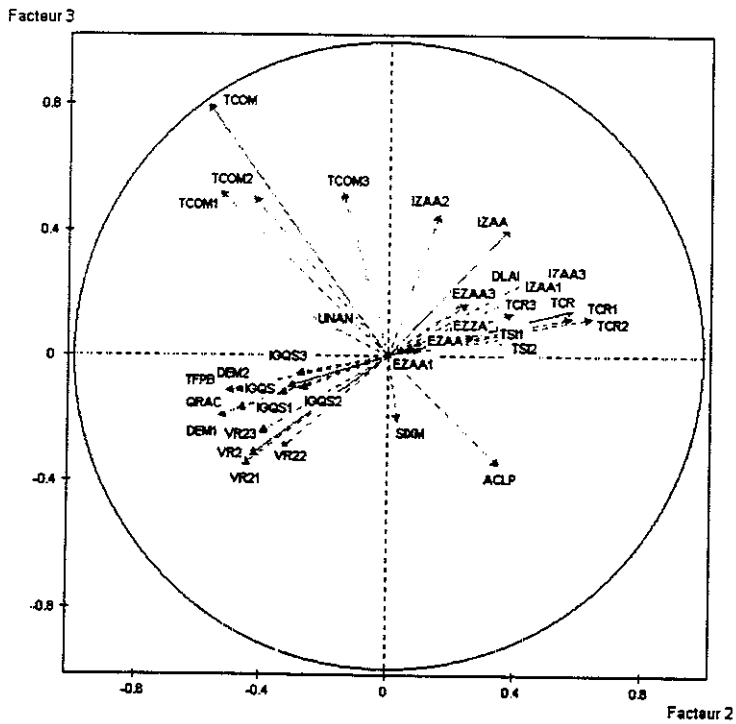


Figure 5 : Cercle des corrélations du plan factoriel 2-3 avec 7 variables actives et 29 variables illustratives

Pour la visualisation du suivi des régions Télécom au cours des années 1984-1985-1986, les trajectoires ont été construites avec des unités statistiques supplémentaires, les trois années successives ayant été désignées par 1, 2, 3.

Les trajectoires représentées sont celle du centre de gravité et celles de six régions significatives quant à leur contribution aux trois premiers axes factoriels.

Le premier facteur variant en sens inverse de la qualité (figure 4), les trajectoires vont dans le sens d'une amélioration de la qualité le long du 1<sup>er</sup> axe (figure 6).

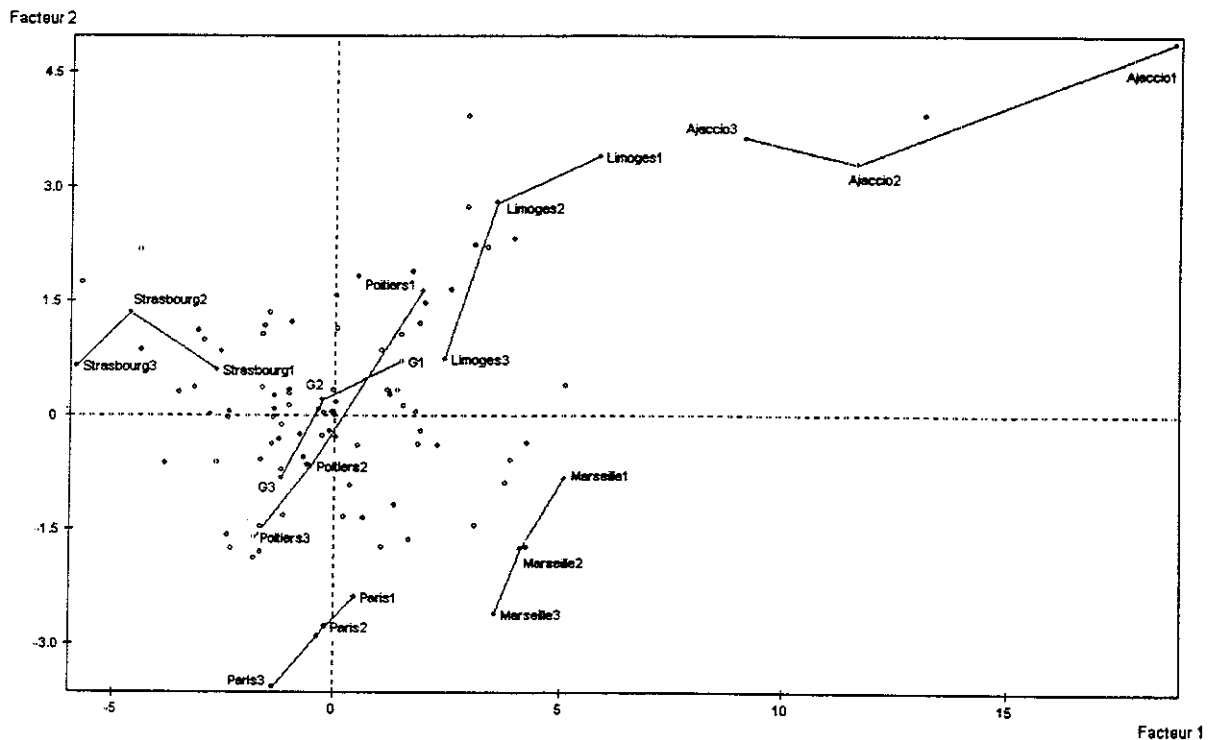


Figure 6 : *Premier plan factoriel avec suivi de six régions et du centre de gravité*

Dans le plan factoriel 2-3, on a représenté les projections des axes unitaires des variables. Les sept trajectoires varient dans le même sens que la variable TCOM : le temps d'établissement des communications diminue entre 1984 et 1986 (figure 7).

Ces deux séries de trajectoires montrent une amélioration de la qualité de service du réseau téléphonique de 1984 à 1986.

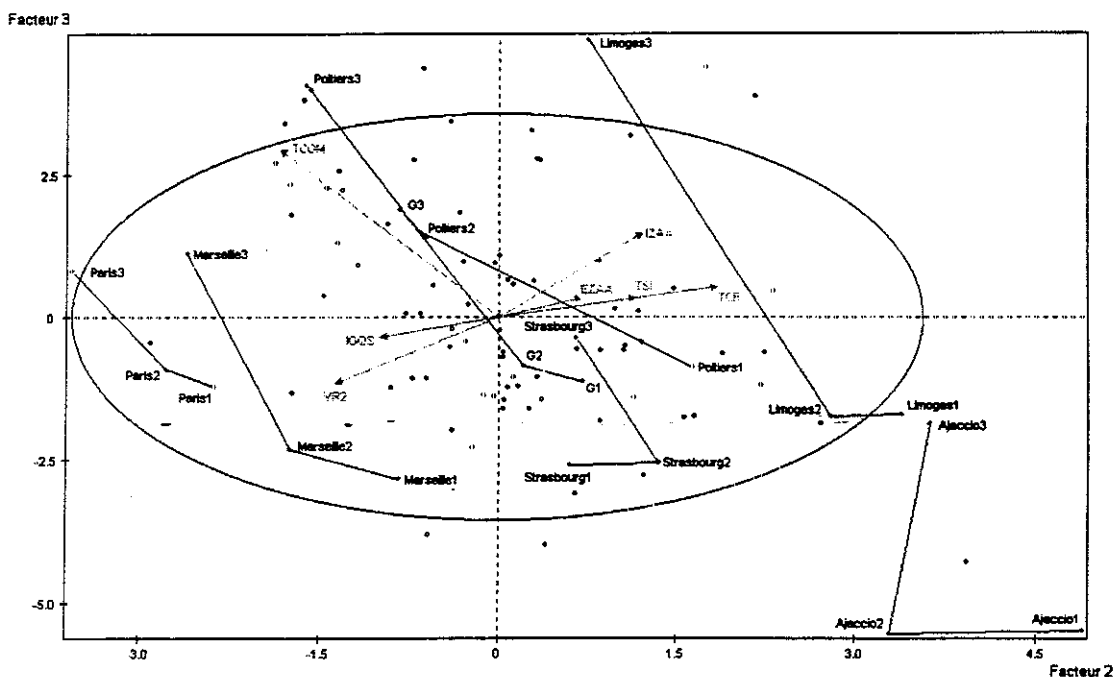


Figure 7 : *Plan factoriel 2-3 avec suivi de six régions et du centre de gravité, et représentation des anciens axes unitaires*

Cette méthode ne permet le suivi de modalités illustratives qu'à condition d'introduire auparavant les centres de gravité de ces modalités aux différentes dates en éléments supplémentaires (figures 8 et 9).

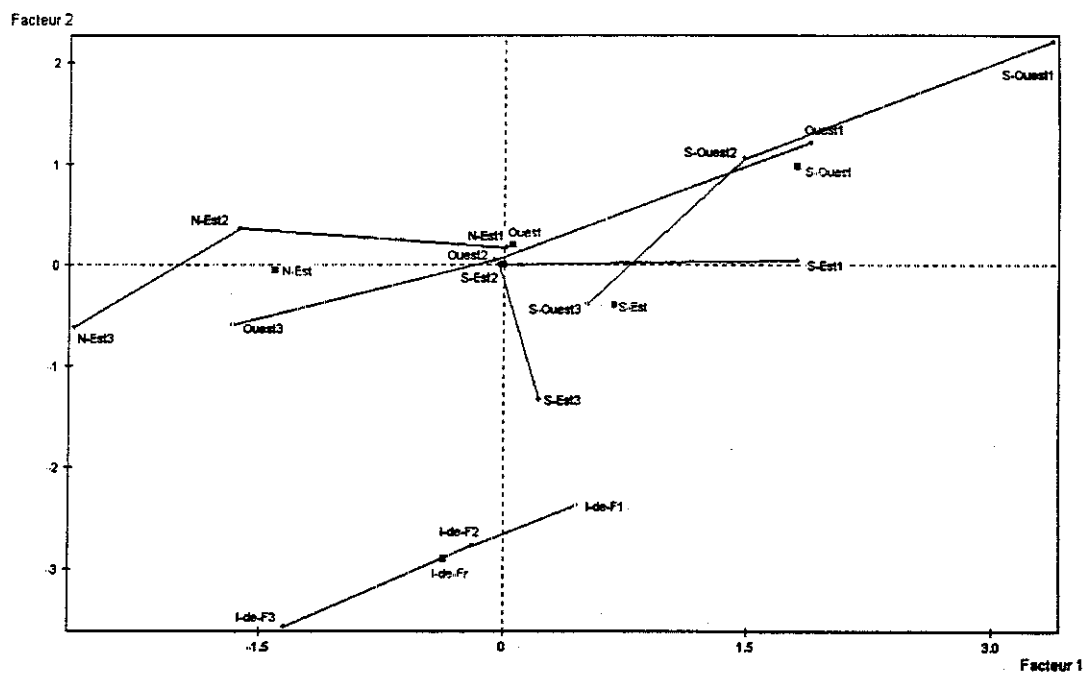


Figure 8 : *Premier plan factoriel avec suivi de l'indicatif régional*

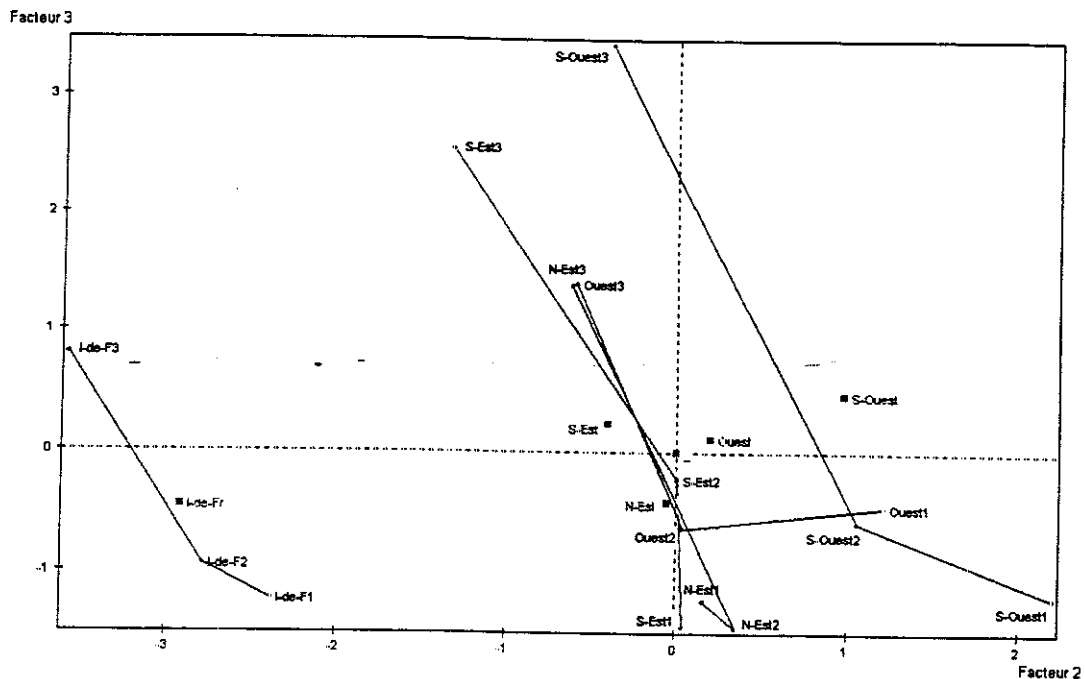


Figure 9 : Plan factoriel 2-3 avec suivi de l'indicateur régional

Une variante de cette méthode consiste à analyser le tableau des données de la 1<sup>ère</sup> date, prise comme date de référence, et à projeter les variables et les individus observés aux dates suivantes en éléments supplémentaires. Cette méthode permet de traiter les cas où :

- les variables ne sont pas les mêmes à chaque date, mais alors on n'a pas de suivi des individus puisque la juxtaposition verticale n'est pas possible,
- les individus ne sont pas les mêmes à chaque date, mais alors on n'a pas de suivi des variables puisque la juxtaposition horizontale n'est pas possible.

## 5. Juxtaposition verticale des tableaux

Cette approche nécessite l'observation des mêmes variables à toutes les dates.

Soit  $G$  le tableau à  $p$  lignes et  $T$  colonnes des centres de gravité de terme général  $g_{jt}$ .

Soit  $Y_t$  le tableau à  $n_t$  lignes et  $p$  colonnes de terme général :

$$y_{ijt} = x_{ijt} - g_{jt}$$

Dans ce tableau de données, chaque observation a été centrée non sur le centre de gravité global des  $T$  tableaux, mais sur le centre de gravité du tableau  $X_t$ . On peut alors réaliser les trois analyses préconisées par [CAZ90]:

- l'analyse globale avec l'ACP normée du tableau  $X$  formé des  $T$  tableaux  $X_t$  juxtaposés verticalement,
- l'analyse intraclasse avec, soit l'ACP normée du tableau  $Y$  formé des  $T$  tableaux  $Y_t$  juxtaposés verticalement (ce qui revient à pondérer chaque variable  $j$  par l'inverse de son écart-type global  $s_j$ ), soit l'ACP sur matrice de covariances après avoir pondéré chaque variable  $j$  par l'inverse de son écart-type  $s_{jt}$  à la date  $t$ ,
- l'analyse interclasse avec l'ACP normée du tableau transposé  $G'$  du tableau  $G$ .

L'ACP du tableau  $Y$  est une ACP conditionnelle à la structure sous-jacente induite par la chronologie. Elle permet de s'affranchir de l'effet de taille dû à l'évolution. Pour les données de l'exemple, on va opérer l'analyse globale pour le suivi des régions au cours des trois années et l'analyse intraclasse pour étudier les différences concernant l'évolution. On pourrait aussi faire l'analyse du tableau  $X$  en mettant les tableaux  $G$  et  $Y$  en supplémentaire.

### 5.1. Analyse globale

Une ACP est réalisée sur le tableau  $X$ , à 63 lignes actives ( $21 \times 3$ ) et 7 colonnes actives, formé des trois tableaux  $X_t$  juxtaposés verticalement. Douze lignes supplémentaires ( $4 \times 3$ ) sont introduites:  $Ajaccio_t$ ,  $G_t$ ,  $Min_t$  et  $Max_t$  pour chaque date  $t$ . Une variable nominale supplémentaire à 15 modalités est définie par l'indicatif téléphonique régional qui prend cinq modalités à chaque date.

Les valeurs propres ne sont pas très différentes de celles de l'ACP du tableau moyen. Il faut toujours retenir trois axes factoriels avec lesquels on prend en compte presque 88% de l'inertie totale :

NUMERO	VALEUR PROPRE	POURCENT	POURCENT CUMULE
1	3.9562	56.52	56.52
2	1.1584	16.55	73.07
3	1.0248	14.64	87.71
4	0.4234	6.05	93.75
5	0.2419	3.46	97.21
6	0.1755	2.51	99.72
7	0.0198	0.28	100.00

Les résultats et représentations graphiques des cercles de corrélation conduisent à interpréter les trois premiers facteurs de la même manière que dans l'analyse du tableau moyen, la différence étant que le 1<sup>er</sup> facteur varie ici dans le même sens que l'indice IGQS (figures 10 et 11).

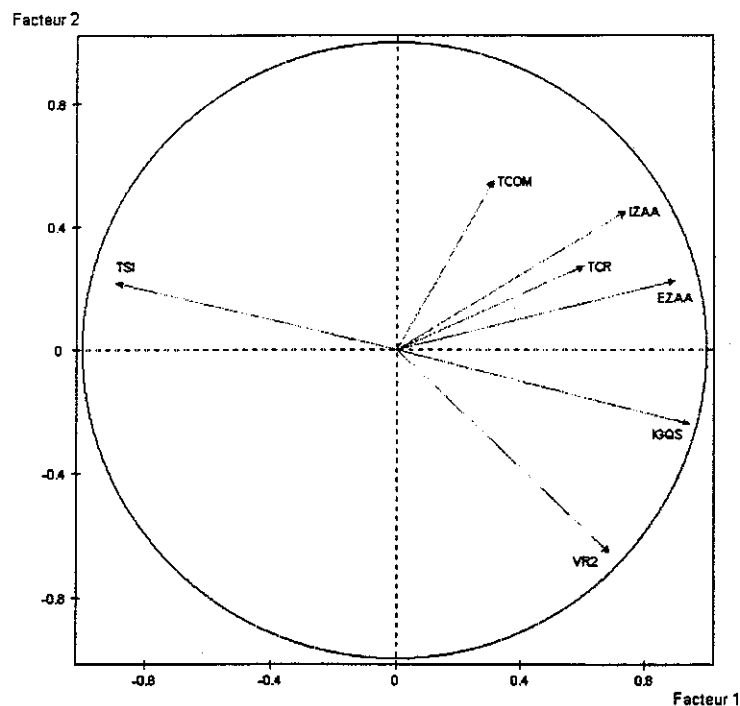


Figure 10 : Cercle des corrélations du 1<sup>er</sup> plan factoriel

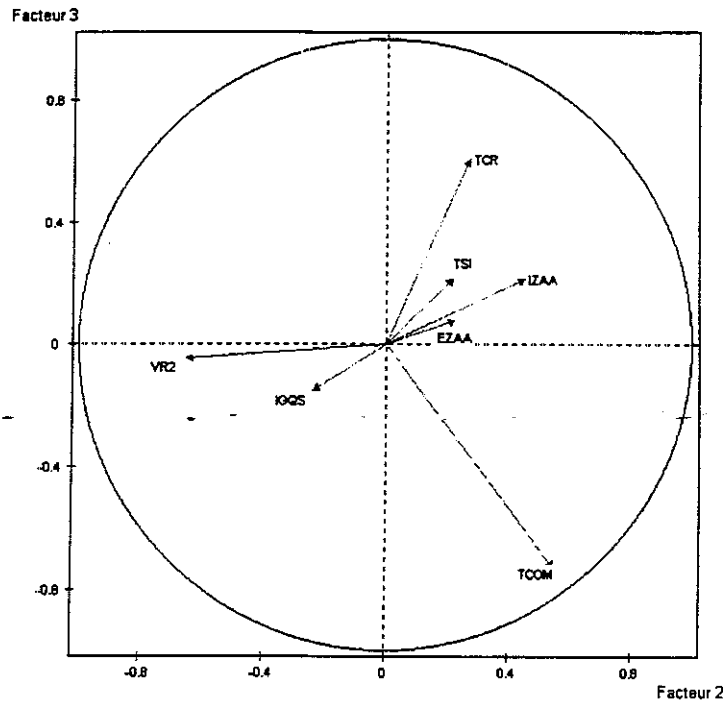


Figure 11 : Cercle des corrélations du plan factoriel 2-3

On peut compléter cette analyse globale par le suivi des modalités de la variable nominale « Indicatif téléphonique régional ». Les modalités dont les valeurs absolues des valeurs-tests (tableau 1) sont supérieures à 2 ont une coordonnée significativement différente de 0 sur l'axe ([LEB99]).

VALEURS-TEST DES MODALITES  
AXES 1 A 3

MODALITES				VALEURS-TEST		
IDEN	LIBELLE	EFF.	P. ABS	1	2	3
1	R*DATE		[AA]			
AA_1	I-de-Fr1	1	1.00	-0.1	-1.5	-1.8
AA_2	Ouest1	5	5.00	-1.9	0.9	1.3
AA_3	N-Est1	7	7.00	-0.2	-1.9	0.9
AA_4	S-Est1	4	4.00	-1.7	-1.2	-0.3
AA_5	S-Ouest1	4	4.00	-3.1	1.1	2.2
AA_6	I-de-Fr2	1	1.00	0.1	-1.5	-2.0
AA_7	Ouest2	5	5.00	0.0	-1.1	0.5
AA_8	N-Est2	7	7.00	1.7	-2.3	1.8
AA_9	S-Est2	4	4.00	-0.1	-0.5	-0.1
AA10	S-Ouest2	4	4.00	-1.4	0.5	1.2
AA11	I-de-Fr3	1	1.00	0.7	-1.0	-2.6
AA12	Ouest3	5	5.00	1.7	0.9	-0.8
AA13	N-Est3	7	7.00	3.2	0.9	-0.6
AA14	S-Est3	4	4.00	0.1	1.8	-2.0
AA15	S-Ouest3	4	4.00	-0.2	3.8	-1.4

Tableau 1

Pour tous les groupes de régions définies par l'indicatif régional, on constate une amélioration de la qualité globale générale au cours des trois années (figure 12), et une amélioration du complément de qualité de service défini par la variable TCOM (figure 13).

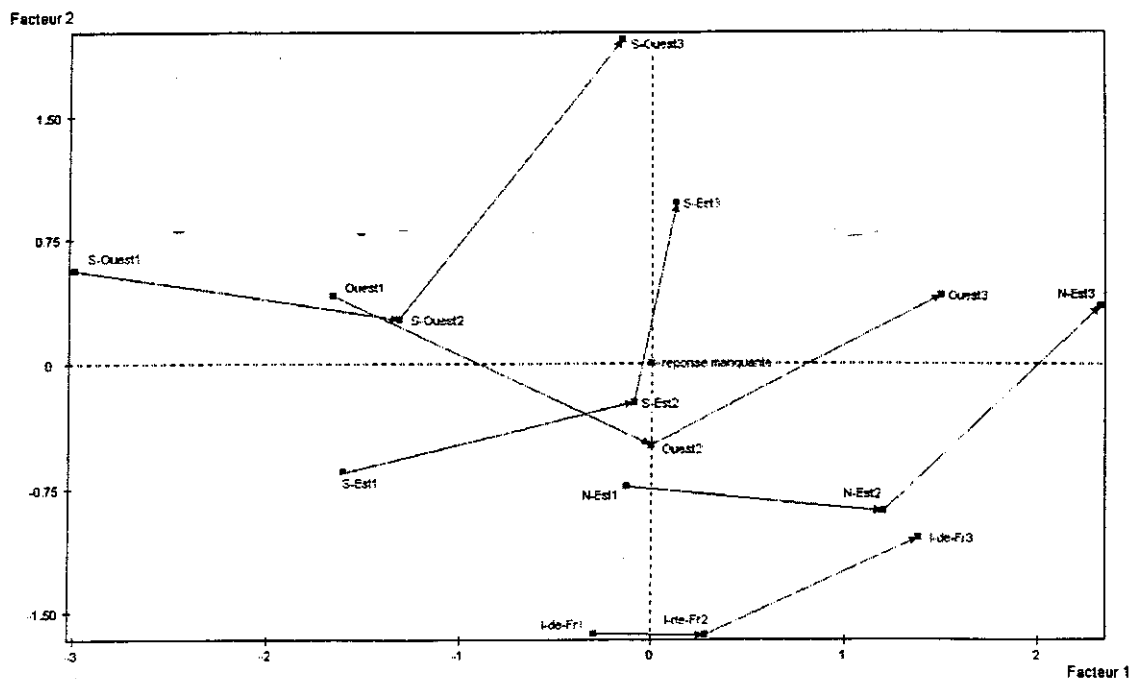


Figure 12 : Premier plan factoriel avec suivi de l'indicatif régional

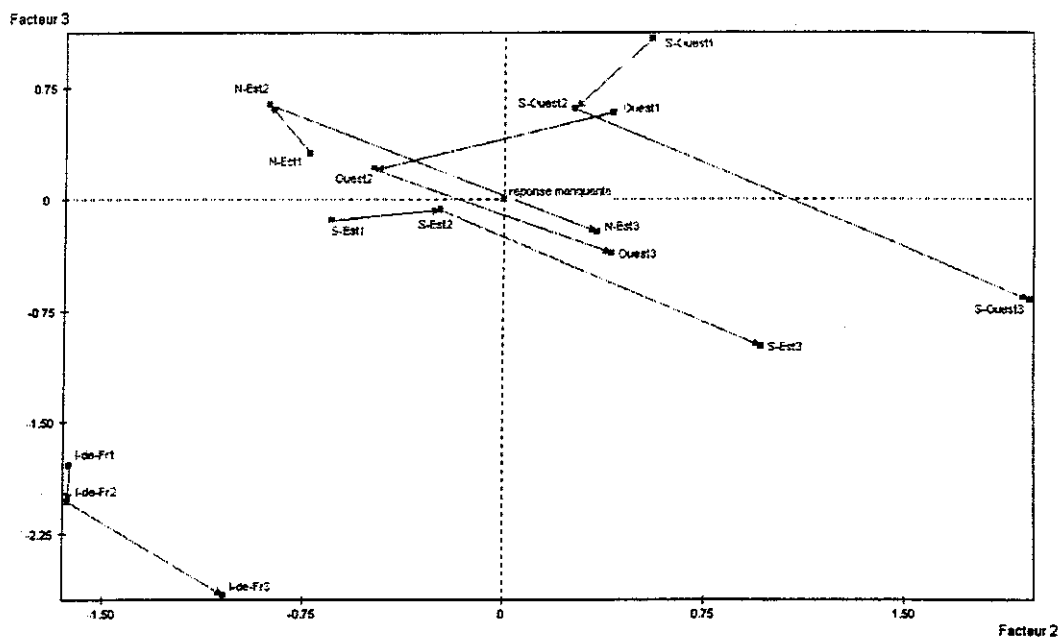


Figure 13 : Plan factoriel 2-3 avec suivi de l'indicatif régional



## 5.2. Analyse intraclasse

Pour cette analyse, les trois nuages sont chacun centrés sur leur centre de gravité pour s'affranchir de l'effet dû à l'évolution des moyennes de chaque variable. On peut faire, soit une ACP normée (pondération de chaque variable  $j$  par l'inverse de son écart-type  $s_j$ ), soit une ACP sur matrice de covariances après avoir pondéré chaque variable  $j$  par l'inverse de son écart-type  $s_j$ . Ces deux ACP donnent des résultats extrêmement voisins. Ce sont les résultats de l'ACP normée qui vont être exposés. Les noms des variables sont suivis de la lettre « C » pour indiquer l'opération de centrage. Les valeurs propres sont peu différentes de celles de l'analyse globale, et il faut toujours prendre en compte les trois premiers axes factoriels.

NUMERO	VALEUR PROPRE	POURCENT	POURCENT CUMULE
1	3.7745	53.92	53.92
2	1.2601	18.00	71.92
3	0.9850	14.07	85.99
4	0.4836	6.91	92.90
5	0.2820	4.03	96.93
6	0.1926	2.75	99.68
7	0.0223	0.32	100.00

Au signe près, les corrélations des variables avec les facteurs sont quasiment les mêmes que celles de l'analyse globale (figures 14 et 15). On peut compléter cette analyse intraclasse en projetant les modalités de la variable nominale « Indicatif téléphonique régional » sur les axes factoriels et en examinant les valeurs-tests (tableau 2). Le 1<sup>er</sup> facteur, corrélé négativement à la qualité globale, oppose le Nord-Est avec une qualité globale significativement positive au Sud-Ouest avec une qualité globale significativement négative. Le second facteur oppose l'Ile-de-France au Sud-Ouest en 1984.

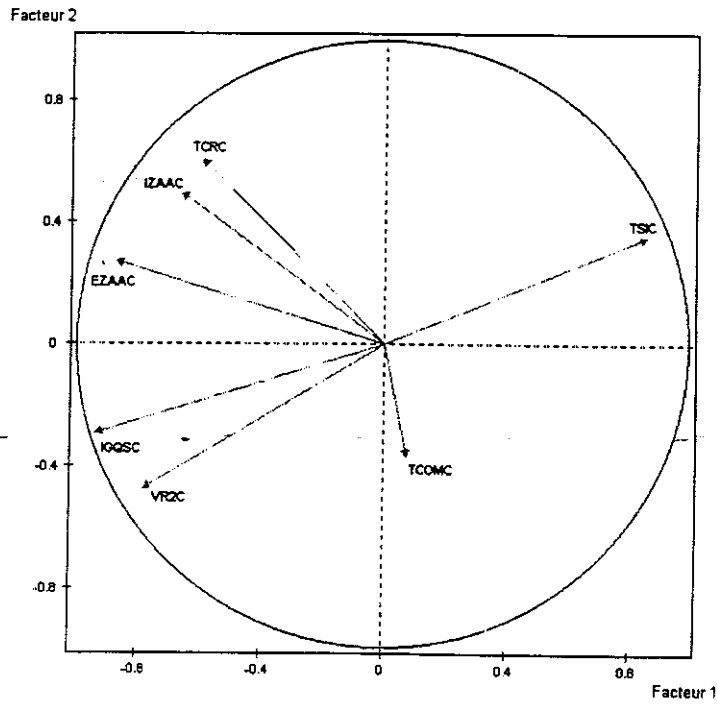


Figure 14 : Cercle des corrélations du 1<sup>er</sup> plan factoriel

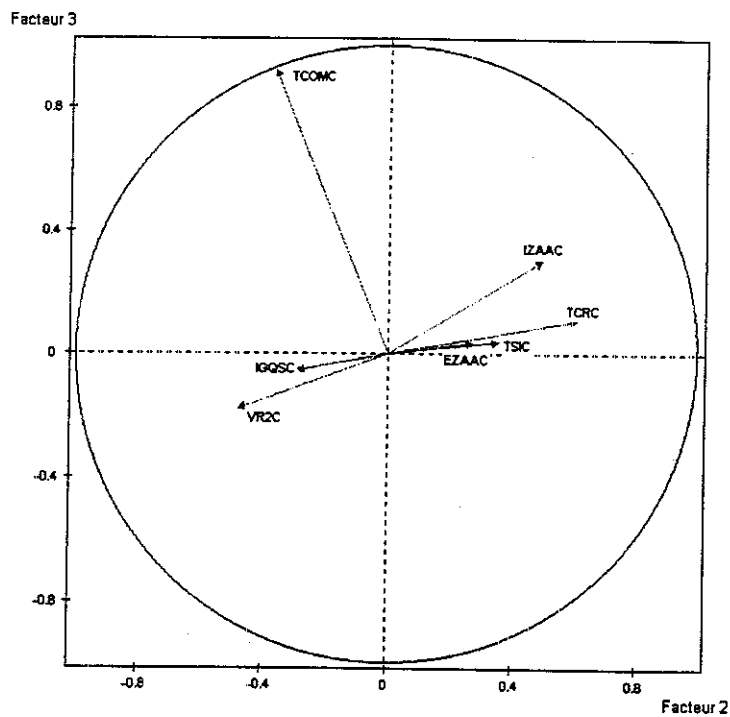


Figure 15 : Cercle des corrélations du plan factoriel 2-3

VALEURS-TESI DES MODALITES				AXES 1 A 3		
MODALITES				VALEURS-TEST		
IDEN	LIBELLE	EFF.	P. ABS	1	2	3
1	R*DATE		[AA]			
AA_1	I-de-Fr1	1	1.00	-0.6	-2.3	0.1
AA_2	Ouest1	5	5.00	0.5	1.3	1.0
AA_3	N-Est1	7	7.00	-2.0	-1.0	-0.2
AA_4	S-Est1	4	4.00	0.3	-1.2	-0.4
AA_5	S-Ouest1	4	4.00	2.0	2.2	-0.5
AA_6	I-de-Fr2	1	1.00	-0.1	-2.3	0.1
AA_7	Ouest2	5	5.00	0.2	-0.1	0.2
AA_8	N-Est2	7	7.00	-1.8	0.0	-1.1
AA_9	S-Est2	4	4.00	0.3	-0.1	0.7
AA10	S-Ouest2	4	4.00	1.8	1.5	0.1
AA11	I-de-Fr3	1	1.00	-0.2	-2.4	-0.7
AA12	Ouest3	5	5.00	-0.5	0.1	-0.8
AA13	N-Est3	7	7.00	-1.9	0.2	-0.9
AA14	S-Est3	4	4.00	1.4	-0.5	0.8
AA15	S-Ouest3	4	4.00	1.8	1.3	1.8

Tableau 2

Les trois nuages partiels étant centrés sur leur centre de gravité, cette analyse rend compte des évolutions par rapport à l'évolution moyenne (tableau 3). Les valeurs de la variable IGQSC étant quasi-constantes et négatives pour le Sud-Ouest, et quasi-constantes et positives pour le Nord-Est, ces deux régions évoluent le long du 1<sup>er</sup> axe comme la moyenne de IGQS et avec toujours le même décalage par rapport à la moyenne (figure 16). Pour l'ouest, l'amélioration est plus rapide que pour la moyenne. Quant à l'Ile-de-France et au sud-est, l'amélioration est moins rapide que pour la moyenne. C'est la raison pour laquelle la trajectoire de l'ouest est opposée à celles de l'Ile-de-France et du sud-est (figure 16). On peut de même interpréter les trajectoires du plan factoriel 2-3 en examinant les évolutions de TCOMC (figure 17).

IND_TEL	IGQSC	IZAAC	EZAAC	TSIC	VR2C	TCRC	TCOMC
I-de-F1	6,59	-0,9	1,11	-3,57	3,47	-15,11	0,67
I-de-F2	4,92	-1,46	0,65	-1,88	1,39	-18,32	0,7
I-de-F3	5,1	-1,04	0,05	-2,68	3,06	-19,92	0,06
Ouest1	-1,33	0,28	0,13	2,37	-1,45	2,01	0,21
Ouest2	-0,34	-0,2	-0,03	1,3	0,63	-0,1	0,16
Ouest3	0,6	0,36	0,33	0,28	1,7	0,14	-0,32
N-Est1	2,3	0,27	0,81	-3,14	2,94	2,7	0,04
N-Est2	2,03	0,11	0,82	-2,69	1,96	4,95	-0,27
N-Est3	2,04	0,45	1,02	-2,39	2,26	4,21	-0,3
S-Est1	-0,13	-0,37	-0,71	-1,52	0,12	-4,58	-0,03
S-Est2	-0,98	1,29	-0,9	-0,48	-0,61	-5,02	0,15
S-Est3	-2,25	-0,86	-1,6	0,8	-1,96	-1,49	0,41
S-Ouest1	-3,88	-0,2	-1,16	4,93	-4,3	1,14	-0,45
S-Ouest2	-3,38	-0,88	-0,67	4,02	-3,93	1,06	-0,02
S-Ouest3	-3,37	-0,09	-0,6	3,7	-4,89	-1,07	0,51

Tableau 3

Valeurs des variables centrées sur la moyenne de la variable à la date  $t$ :  $x_{ijt} - \bar{x}_{jt}$

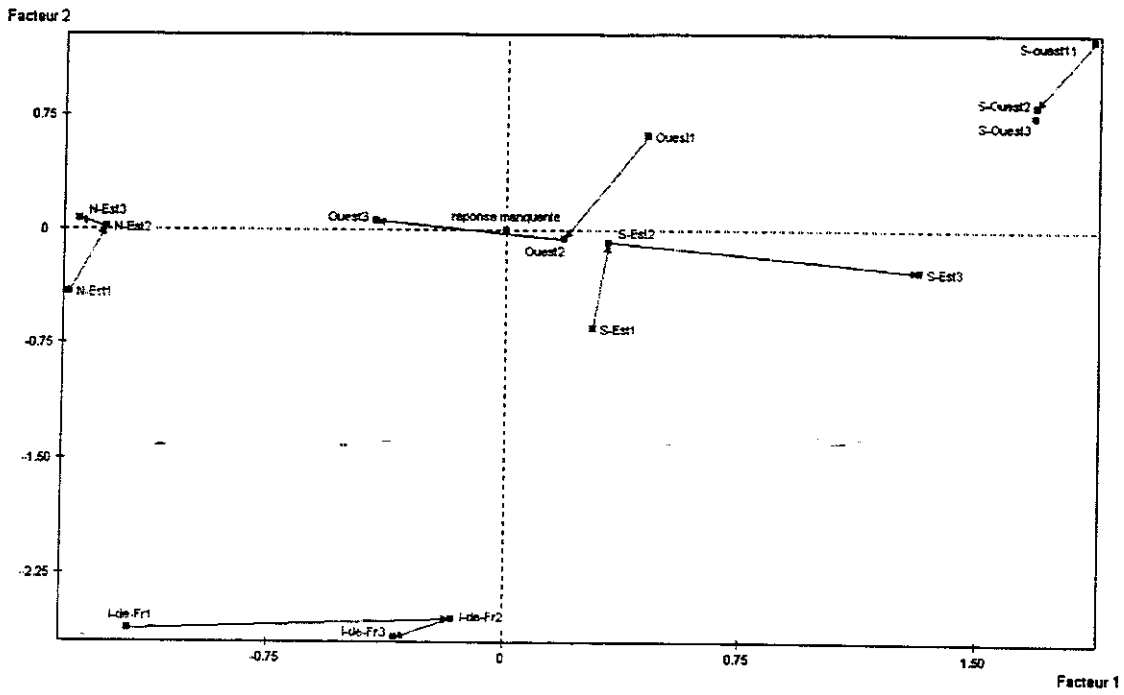


Figure 16 : Premier plan factoriel de l'analyse intraclasse

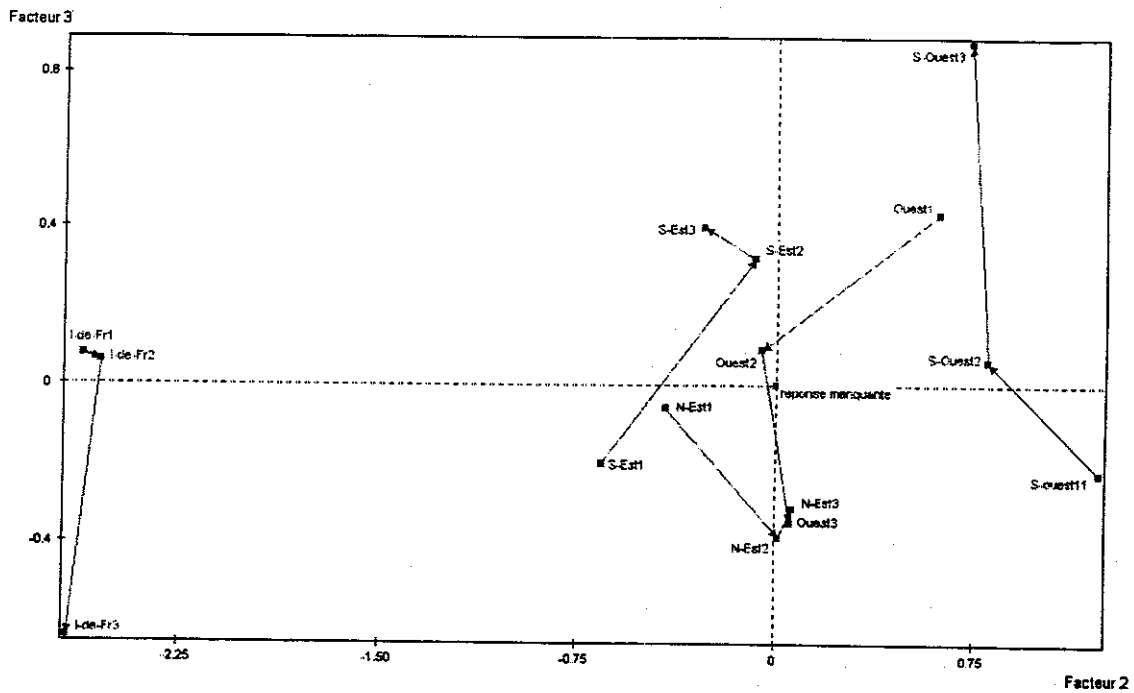


Figure 17 : Plan factoriel 2-3 de l'analyse intraclasse

## 6. Analyse Factorielle Multiple

Le principe de l'AFM ([ESC98]) repose sur une analyse factorielle de l'ensemble des groupes juxtaposés horizontalement, les variables de chaque groupe étant pondérées de façon à équilibrer leur influence. L'inertie axiale maximum de chaque groupe est ramenée à 1 avec la pondération de chaque variable du groupe par l'inverse de la racine carrée de la première valeur propre de son analyse factorielle séparée.

Comme en Analyse Canonique Généralisée [CAR68], on cherche en AFM, une suite de *variables générales*, orthogonales entre elles, les plus liées à l'ensemble des groupes de variables et, pour chaque variable générale et dans chaque groupe, des *variables canoniques* les plus liées aux variables générales. L'AFM prend en compte la structure en groupes de variables, permet la comparaison des groupes, la mise en évidence de facteurs communs aux différents groupes et une visualisation des ressemblances entre groupes.

### 6.1. Mise en œuvre

Pour les trois tableaux  $X_1$ ,  $X_2$  et  $X_3$ , les valeurs propres maxima des ACP normées ne sont pas très différentes et respectivement égales à 3,821, 3,846 et 4,129. Après pondération des variables de chaque groupe, on réalise l'analyse des trois tableaux juxtaposés horizontalement. Trois groupes sont introduits en éléments supplémentaires :

- le groupe des indicateurs économiques en 1984,
- le groupe formé par la variable nominale « Indicatif téléphonique régional » à cinq modalités, et une nouvelle variable nominale à deux modalités « Nord » et « Sud » introduite pour avoir un groupe comportant au moins deux variables de même nature (une variable ne pouvant former à elle seule un groupe),
- le groupe défini par le tableau moyen  $X_m$ .

En AFM, la liaison  $Lg$  entre une variable quantitative  $z$  et un groupe de variables est égale à la valeur de l'inertie projetée des variables du groupe sur  $z$  : la 1<sup>ère</sup> variable générale  $z_1$  maximise la somme de ses liaisons avec chacun des groupes, la 2<sup>nd</sup>e variable générale  $z_2$ , orthogonale à  $z_1$ , maximise la somme de ses liaisons avec chacun des groupes... Dans le cas de variables

continues, la liaison  $Lg$  entre la variable  $z$  et le groupe  $t$  s'écrit : 
$$Lg(z, t) = \sum_{j=1}^{p_t} (r(z, j))^2 / \lambda_{1,t}$$
,

$r(z, j)$  étant le coefficient de corrélation linéaire entre les variables  $z$  et  $j$ , et  $\lambda_{1,t}$  la plus grande valeur propre de l'ACP normée du groupe  $t$ . On montre que les variables générales

sont les composantes principales normées de l'ACP pondérée du tableau complet. En notant

$\lambda_\alpha$  la valeur propre de rang  $\alpha$  et  $\psi_\alpha$  le facteur associé, on a :  $\Psi_\alpha = \sqrt{\lambda_\alpha} z_\alpha$ .

LES CINQ PREMIERES VALEURS PROPRES			
NUMERO	VALEUR PROPRE	POURCENTI.	POURCENT. CUMULE
1	2.8544	53.38	53.38
2	0.8816	16.49	69.87
3	0.3618	6.77	76.63
4	0.3467	6.48	83.11
5	0.2491	4.66	87.77

La 1<sup>ère</sup> valeur propre, égale à 2,85, est proche de la valeur maximale possible, à savoir 3. Ceci signifie que les 1<sup>ers</sup> axes de chaque groupe sont très voisins du 1<sup>er</sup> axe de l'analyse du tableau complet. Ce résultat sera confirmé au §6.2. Pour l'analyse du tableau complet, on retient les deux premiers axes factoriels puisqu'on observe un « coude » après la 2<sup>nde</sup> valeur propre (critère de Catell). Les facteurs sont interprétés à l'aide de leurs corrélations avec les 21 variables actives (tableau 4). Le 1<sup>er</sup> facteur résume à lui seul plus de 53% de l'inertie. Les variables étant toutes fortement corrélées à ce facteur (sauf ICOM1, TCOM2 et TCOM3), celui-ci apparaît toujours comme un facteur de qualité globale générale. Les variables TCR1, ICR2, ICR3, IZAA1 et IZAA3 permettent d'interpréter le 2<sup>ème</sup> facteur qui explique 16,5% de l'inertie. Ce facteur nuance le facteur global de qualité de service (figure 18).

COORDONNEES ET AIDES A L'INTERPRETATION DES VARIABLES ACTIVES												
	COORDONNEES				CONTRIBUTIONS				COSINUS CARRES			
	1	2	3	4	1	2	3	4	1	2	3	4
<b>GROUPE 1</b>												
IGQS1	-0.9	0.3	0.0	-0.2	7.0	3.3	0.1	3.5	0.77	0.11	0.00	0.05
IZAA1	-0.7	-0.5	-0.1	-0.3	4.0	8.6	0.7	5.4	0.43	0.29	0.01	0.07
EZAA1	-0.9	-0.1	-0.1	0.0	7.8	0.3	0.3	0.1	0.85	0.01	0.00	0.00
TSI1	0.8	-0.3	0.0	0.4	6.1	2.8	0.0	12.9	0.66	0.09	0.00	0.17
VR21	-0.7	0.5	0.3	0.0	4.6	7.3	4.9	0.1	0.50	0.25	0.07	0.00
TCR1	-0.6	-0.6	0.0	0.2	3.3	11.7	0.0	4.6	0.36	0.40	0.00	0.06
TCOM1	-0.2	0.4	-0.6	0.3	0.4	4.1	27.2	8.4	0.05	0.14	0.38	0.11
ENSEMBLE					33.2	38.2	33.3	34.9	0.52	0.18	0.07	0.07
<b>GROUPE 2</b>												
IGQS2	-0.9	0.2	0.0	-0.1	8.1	1.8	0.1	0.2	0.89	0.06	0.00	0.00
IZAA2	-0.6	-0.2	-0.5	-0.3	2.9	1.1	17.5	7.4	0.32	0.04	0.24	0.10
EZAA2	-0.9	-0.3	0.0	0.0	7.0	2.7	0.0	0.2	0.72	0.12	0.00	0.00
TSI2	0.8	-0.3	0.0	0.3	6.6	3.5	0.0	6.5	0.72	0.12	0.00	0.09
VR22	-0.8	0.4	0.2	0.2	5.9	4.0	2.7	3.9	0.65	0.14	0.04	0.05
TCR2	-0.6	-0.7	0.1	0.1	3.8	14.0	0.3	0.3	0.42	0.47	0.00	0.00
TCOM2	0.0	0.3	-0.7	0.3	0.0	2.9	35.9	6.7	0.00	0.10	0.50	0.09
ENSEMBLE					34.4	30.0	56.5	25.2	0.54	0.15	0.11	0.05
<b>GROUPE 3</b>												
IGQS3	-0.9	0.2	0.0	0.2	7.0	1.5	0.1	1.6	0.83	0.06	0.00	0.02
IZAA3	-0.7	-0.6	-0.1	0.1	4.0	9.9	0.3	1.2	0.47	0.36	0.00	0.02
EZAA3	-0.8	-0.3	-0.1	0.3	5.3	3.3	0.4	6.2	0.63	0.12	0.01	0.09
TSI3	0.8	-0.4	0.0	0.1	6.0	4.1	0.0	0.8	0.71	0.15	0.00	0.01
VR23	-0.8	0.4	0.1	0.2	5.0	4.8	0.6	4.4	0.59	0.17	0.01	0.06
TCR3	-0.7	-0.5	0.0	-0.1	3.7	8.2	0.0	0.6	0.44	0.30	0.00	0.01
TCOM3	0.4	-0.1	-0.4	-0.6	1.4	0.1	8.8	25.1	0.16	0.00	0.13	0.36
ENSEMBLE					32.4	31.9	10.2	39.8	0.55	0.17	0.02	0.08

Tableau 4

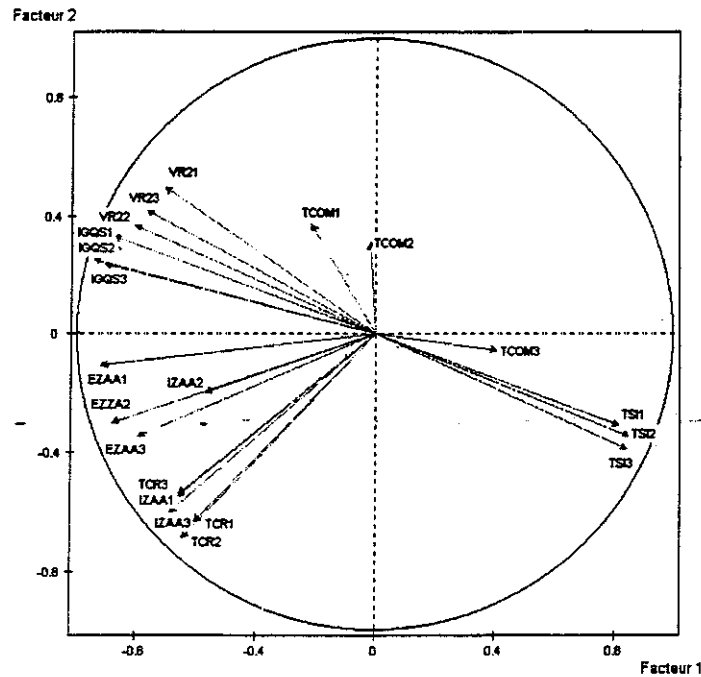


Figure 18 : Cercle des corrélations du 1<sup>er</sup> plan factoriel de l'AFM

## 6.2. Représentation des groupes

Le coefficient de corrélation vectoriel  $RV$ , toujours compris entre 0 et 1, est une mesure classique de la ressemblance entre deux groupes de variables ([ESC98], [ROB76]). En AFM, on complète cette mesure par la mesure de liaison  $Lg$ , généralisation de la mesure de liaison entre une variable et un groupe de variables (§6.1) :  $Lg(t, t') = \sum_{j \in \text{Groupe } t} Lg(j, t')$ . Cet indicateur

est d'autant plus élevé que chacune des variables d'un groupe est plus liée à chacune des variables de l'autre groupe ([ESC98]).

	Lg			RV		
	1	2	3	1	2	3
1 : 1984	1.22			1.00		
2 : 1985	1.03	1.19		0.86	1.00	
3 : 1986	0.87	0.97	1.16	0.73	0.82	1.00

Les valeurs des mesures de liaison  $Lg$  et  $RV$  entre les trois groupes, nous montrent que :

- les trois groupes présentent des dimensionnalités de valeurs voisines,
- la structure commune aux groupes 1 et 2 est la plus riche avec  $Lg(1,2) = 1,03$  et  $RV(1,2) = 0,86$ ,

- deux groupes correspondant à deux années successives ont des structures plus proches que les deux groupes 1 et 3.

Les valeurs des coefficients de corrélation entre les variables canoniques d'un groupe et les facteurs de l'analyse factorielle du tableau complet montrent dans quelle mesure les facteurs de cette analyse peuvent être considérés comme des structures du groupe. Ici, les deux premiers facteurs sont communs aux trois groupes.

CORRELATIONS ENTRE LES VARIABLES CANONIQUES  
ET LES VARIABLES GENERALES

- CORRELATIONS -					
FAC.	1	2	3	4	5
GR 1	0.97	0.92	0.77	0.63	0.55
GR 2	0.99	0.95	0.86	0.70	0.67
GR 3	0.96	0.92	0.40	0.57	0.75

Le rapport de corrélation, défini comme le quotient de l'inertie inter par l'inertie totale, est calculé pour chaque axe de la façon suivante : à chaque individu actif, est associé une classe regroupant l'ensemble des points représentant cet individu, ici chaque classe contient trois points puisqu'on a trois groupes actifs. Ce rapport montre dans quelle mesure chaque axe met en évidence une structure commune aux groupes.

Le tableau suivant montre que ce rapport de corrélation est élevé pour les deux premiers axes qui sont communs aux trois groupes.

RAPPORT : INERTIE INTER/INERTIE TOTALE

FAC.	1	2	3	4	5
	0.95	0.85	0.49	0.36	0.43

A chaque groupe, on associe sa matrice des produits scalaires entre individus qui est un élément de l'espace  $R^{n^2}$ . L'Analyse Factorielle Multiple fait aussi intervenir les matrices  $z_\alpha z'_\alpha$  associées à chaque variable générale  $z_\alpha$  et qui forment une base orthonormée d'un sous-espace de  $R^{n^2}$ . Les points-groupes sont projetés sur cette base pour visualiser leurs proximités. L'intérêt de cet espace de projection est que les axes de projection possèdent la même interprétation que ceux de l'analyse du tableau complet.

On montre que la coordonnée du groupe  $t$  sur l'axe de rang  $\alpha$  est égale à la quantité  $Lg(z_\alpha, t)$  qui est toujours comprise entre 0 et 1 du fait de la pondération des variables à l'intérieur des



groupes. La somme des coordonnées des groupes sur un axe donné est ainsi égale à l'inertie de l'axe de même rang de l'analyse du tableau complet.

L'Analyse Factorielle Multiple des groupes de variables permet la représentation des trois groupes actifs et des groupes supplémentaires. La coordonnée d'un groupe sur un axe égale à l'inertie projetée de l'ensemble du groupe sur l'axe peut, de ce fait, être considérée comme une mesure de la liaison entre le groupe et le facteur correspondant.

GRP.	P. REL DISTO		COORDONNEES				COSINUS CARRES			
	1	2	1	2	3	4	1	2	3	4
GR 1	0.33	1.22	0.95	0.34	0.12	0.12	0.73	0.09	0.01	0.01
GR 2	0.33	1.19	0.98	0.26	0.20	0.09	0.81	0.06	0.04	0.01
GR 3	0.33	1.16	0.93	0.28	0.04	0.14	0.74	0.07	0.00	0.02

Les valeurs des cosinus carrés montrent que les trois groupes actifs sont surtout bien représentés sur le 1<sup>er</sup> axe factoriel.

### 6.3. Représentation des individus et des modalités

A chaque groupe  $t$  de variables, on associe le nuage partiel des individus qui contient les individus vus par ce groupe de variables, et on projette les nuages partiels sur les axes principaux de l'analyse factorielle du tableau complet. La projection du nuage associé au groupe  $t$  sur l'axe factoriel de rang  $\alpha$  de cette analyse définit la variable canonique  $\Psi_{\alpha}^t$  du groupe  $t$  qui se trouve être la combinaison linéaire des variables du groupe la plus liée à la variable générale  $z_{\alpha}$ .

A un coefficient près, la représentation d'un individu  $\omega_i$  par sa coordonnée  $\Psi_{\alpha i}$  est au barycentre de ses représentations par les  $\Psi_{\alpha i}^t$  : on dit que l'image globale d'un individu est au centre de gravité de ses images partielles. On retrouve cette propriété pour les modalités des variables qualitatives puisqu'on représente une modalité par le centre de gravité des individus qui la possèdent.

En AFM, on représente simultanément le nuage de l'analyse du tableau complet et en éléments supplémentaires, les nuages associés à chaque groupe. Ces nuages étant centrés sur leur centre de gravité, l'analyse des nuages partiels correspond à une analyse intraclasse.

Les modalités sont traitées comme des individus, car elles sont représentées par des centres de gravité d'individus. L'examen des valeurs-tests (tableau 5) des modalités de l'indicatif régional, ainsi que la représentation graphique des trajectoires (figure 19), montrent que ces résultats s'interprètent (au signe du 2<sup>nd</sup> facteur près) comme ceux de l'analyse intraclasse du § 5.2. Ceci s'explique par le fait que pour chacune de ces deux analyses, les variables sont centrées sur la moyenne de leur groupe. L'analyse factorielle des tableaux juxtaposés verticalement, chaque tableau étant centré sur le centre de gravité de ses variables, se révèle donner des résultats presque équivalents à ceux de l'AFM qui opère sur les tableaux juxtaposés horizontalement avec, pour l'exemple, des pondérations peu différentes pour les variables de chaque tableau. On distingue toujours les régions qui évoluent le long du 1<sup>er</sup> axe avec toujours le même décalage par rapport à la moyenne (Sud-Ouest et Nord-Est), plus rapidement que la moyenne (Ouest), plus lentement que la moyenne (Ile de France et Sud-Est).

CENTRES DE GRAVITE		VALEURS-TESTI			
MODAL. / GROUPE	P. REL	1	2	3	4
Ile-de-France	4.76	0.9	1.9	-0.9	0.5
POUR GROUPE 1		-0.3	2.5	-0.3	-0.1
2		-0.5	2.2	-0.2	-1.1
3		-0.2	2.5	-0.4	0.5
Ouest	23.81	0.0	-0.5	-0.4	2.0
POUR GROUPE 1		0.4	-1.2	-1.2	1.2
2		0.2	0.0	-0.2	1.3
3		-0.6	-0.1	0.9	1.1
Nord-Est	33.33	-2.2	0.3	1.6	-0.1
POUR GROUPE 1		-2.1	0.9	0.9	-1.5
2		-2.2	-0.2	1.4	-1.2
3		-2.2	0.0	1.0	1.7
Sud-Est	19.05	0.8	1.0	-0.8	-2.0
POUR GROUPE 1		0.4	1.1	0.3	-0.6
2		0.5	0.9	-1.2	-1.2
3		1.5	0.6	-0.5	-1.8
Sud-Ouest	19.05	2.0	-2.1	-0.5	0.1
POUR GROUPE 1		2.0	-2.0	0.1	1.7
2		2.0	-2.1	-0.1	0.9
3		1.9	-1.7	-1.7	-1.7

Tableau 5

En AFM, la représentation des groupes constitue l'étude de l'« interstructure », et la représentation simultanée des nuages partiels constitue l'étude de l'« intrastructure ».

Avec l'AFM, on ne dispose pas de l'analyse de l'évolution globale comme au §5.1.

Pour l'exemple traité, les groupes étant quasiment équilibrés, l'ACP de l'ensemble des tableaux juxtaposés horizontalement ne donne pas des résultats très différents de ceux de l'analyse du tableau complet de l'AFM. Mais l'AFM est beaucoup plus intéressante de par la richesse de ses résultats. C'est la raison pour laquelle l'ACP de l'ensemble des tableaux juxtaposés horizontalement n'est pas traitée ici.

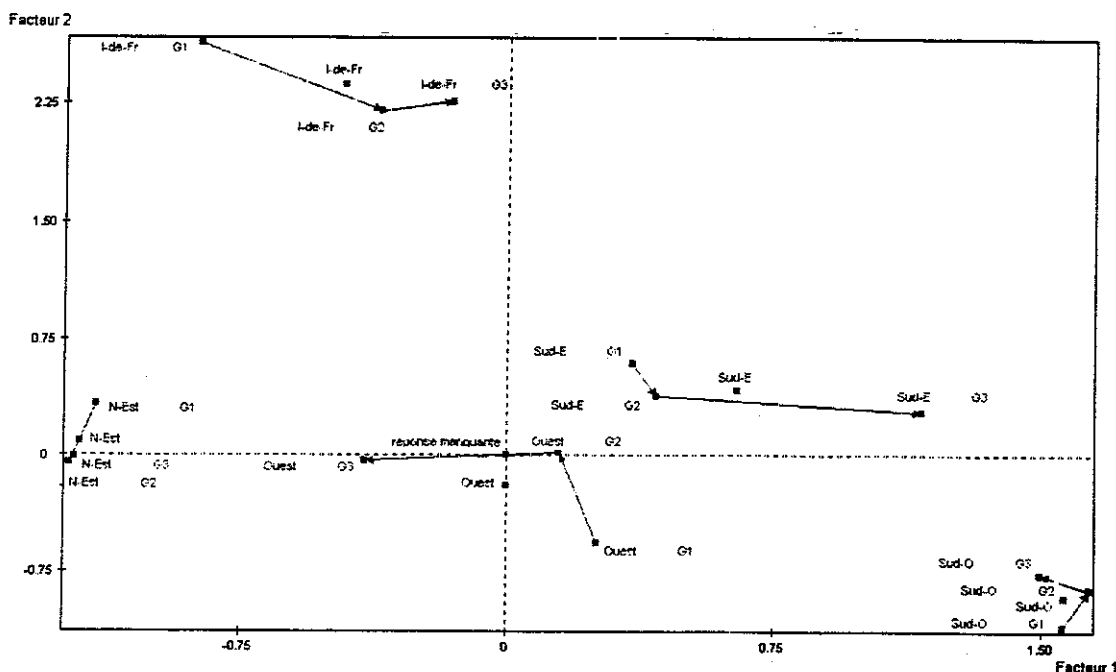


Figure 19 : Premier plan factoriel de l'AFM

## 7. Méthode STATIS

Cette méthode développée par [LAV88] s'applique à des tableaux de données quantitatives. L'idée essentielle est la recherche d'une structure commune aux différents tableaux.

La méthode STATIS s'applique à des tableaux dans lesquels un ensemble d'individus est décrit par plusieurs groupes de variables.

L'étude de l'« interstructure » est réalisée avec l'Analyse en Composantes Principales de la matrice des coefficients de corrélation vectoriel  $RV$  (définis au §6.2).

Comme en AFM, on tient compte de la structure en pondérant spécifiquement les variables de chaque groupe. Avec la méthode STATIS, le poids affecté aux variables d'un groupe est d'autant plus important que le groupe est lié à l'ensemble des groupes ([PAG96]). Finalement, on réalise une ACP d'un tableau « compromis » qui fournit un référentiel commun pour représenter le nuage des individus et les nuages partiels (les nuages partiels ne peuvent en fait pas être représentés avec la version disponible actuellement dans SPAD).

Le programme SPAD décompose l'analyse en quatre étapes :

- analyse et comparaison de la forme des différents nuages de lignes par l'analyse des  $T$  matrices carrées de dimension  $n$  des produits scalaires entre individus (tableau 6),
- analyse de la situation des nuages les uns par rapport aux autres (analyse de l'« interstructure »), interprétation des distances entre nuages (tableaux 7 et 8),
- étude du nuage des groupes (figure 20) et construction du nuage compromis,
- étude du nuage compromis (analyse de l'« intrastructure ») et représentation des individus (figure 21).

Pour l'exemple, les trois nuages de lignes ont à peu près la même allure, car les pourcentages d'inertie des axes des trois nuages ne sont pas très différents (tableau 6). Par contre, les distances entre nuages étant non négligeables, les nuages ne sont pas superposables et diffèrent par leur structure interne (tableaux 7 et 8).

NUMERO	VALEUR PROPRE	POURCENT	VALEUR PROPRE	POURCENT	VALEUR PROPRE	POURCENT
1	0.8540	57.70	0.9521	69.83	0.9308	67.83
2	0.5152	34.81	0.2984	21.89	0.3618	26.36
3	0.0671	4.53	0.0577	4.23	0.0448	3.27
4	0.0247	1.67	0.0266	1.95	0.0241	1.75
5	0.0111	0.75	0.0218	1.60	0.0073	0.53
Nuage 1			Nuage 2		Nuage 3	

**Tableau 6**

Ces résultats peuvent être vus comme une première analyse exploratoire des différences entre les tableaux de données. Ils permettent de mettre en évidence les unités statistiques qui contribuent le plus à leur différence et qui correspondent ici aux régions qui ont déjà été repérées comme évoluant le plus : Paris, Marseille, Limoges (§4).

Le nuage compromis est très bien représenté dans le 1<sup>er</sup> plan factoriel avec plus de 86% de l'inertie (figure 21).

REPRESENTAIIION GRAPHIQUE DU NUAGE COMPROMIS  
VALEURS PROPRES TRACE DE LA MATRICE : 1.4912

NUMERO	VALEUR PROPRE	POURCENT. CUMULE	POURCENT.
1	0.9293	62.32	62.32
2	0.3579	24.00	86.33
3	0.0705	4.73	91.05
4	0.0408	2.74	93.79
5	0.0294	1.97	95.76

La procédure STATIS de SPAD ne propose pas l'utilisation d'éléments supplémentaires. Le dessin des trajectoires, obtenu en projetant les nuages partiels, n'est pas possible, car « le plan principal du compromis n'est pas un plan optimal pour ces trajectoires » [SPAD • TM, 2000]. La décomposition des distances entre nuages en contribution des individus permet de repérer les régions qui ont une trajectoire différente des autres.

INTERPRETAIIION DES DISTANCES ENTRE NUAGES  
DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LE NUAGE 1 ET LES AUTRES  
CONTRIBUTIONS DES LIGNES EN POURCENTAGE

LIGNE	MOYENNE	NUAGE	
		2	3
Amiens	2.2	2.2	2.2
Besançon	5.8	6.1	5.5
Bordeaux	4.1	4.1	4.2
Caen	5.8	5.1	6.4
Chalons	3.5	3.7	3.3
Clermont	2.5	2.2	2.7
Dijon	1.5	2.3	0.8
Lille	8.1	11.0	5.1
Limoges	7.3	7.0	7.6
Lyon	1.5	1.3	1.6
Marseille	9.1	12.7	5.5
Montpellier	3.3	2.1	4.4
Nancy	1.7	2.1	1.4
Nantes	1.7	1.5	2.0
Orléans	0.6	0.9	0.3
Poitiers	4.1	3.2	5.0
Rennes	3.1	0.5	5.7
Rouen	7.4	7.1	7.8
Strasbourg	6.0	7.0	5.0
Toulouse	2.0	1.3	2.8
Paris	18.7	16.6	20.8

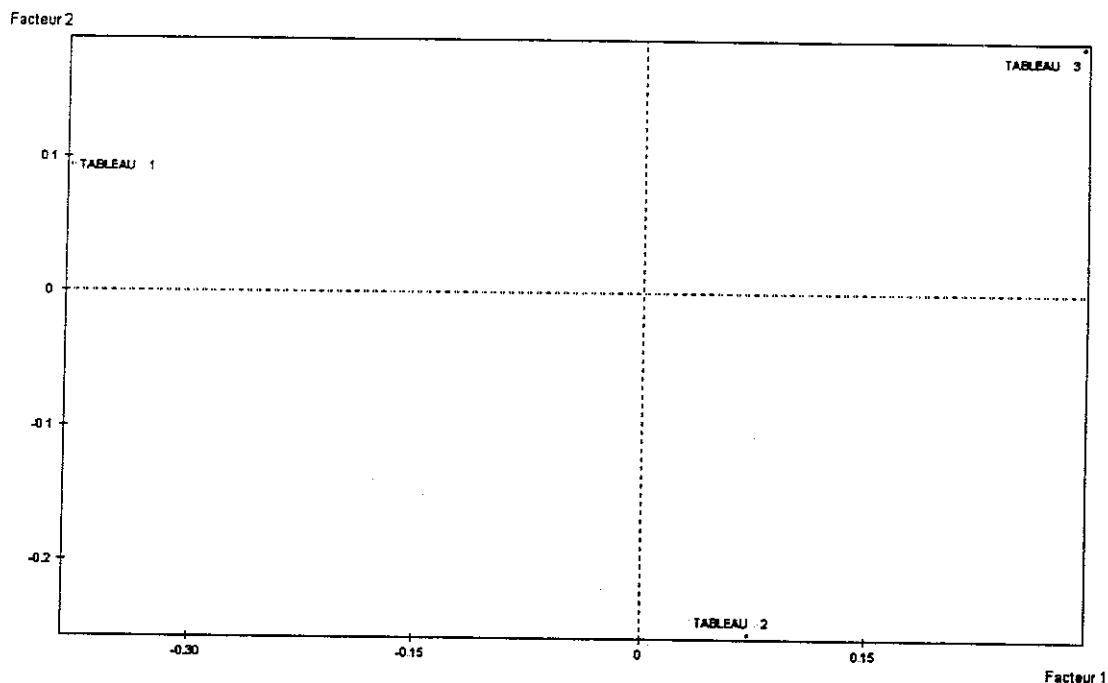
Tableau 7

LA SOMME DES DISTANCES AU CARRE SE DECOMPOSE SELON LES LIGNES

Amiens	=	2.22 %
Besancon	=	4.66 %
Bordeaux	=	3.40 %
Caen	=	4.85 %
Chalons	=	3.10 %
Clermont	=	2.57 %
Dijon	=	1.68 %
Lille	=	7.03 %
Limoges	=	8.18 %
Lyon	=	1.53 %
Marseille	=	9.73 %
Montpellier	=	5.18 %
Nancy	=	1.87 %
Nantes	=	1.57 %
Orleans	=	0.79 %
Poitiers	=	3.83 %
Rennes	=	4.88 %
Rouen	=	6.32 %
Strasbourg	=	5.96 %
Toulouse	=	2.21 %
Paris	=	18.44 %

**Tableau 8**

Le tableau moyen pouvant être envisagé comme un tableau compromis, la méthode STATIS et l'analyse factorielle du tableau moyen (§4) donnent des résultats presque équivalents, mais la contrainte pour l'analyse du tableau moyen, à savoir, l'observation des mêmes variables à chaque instant, n'existe pas pour STATIS.

**Figure 20 : Représentation des groupes avec la méthode STATIS**

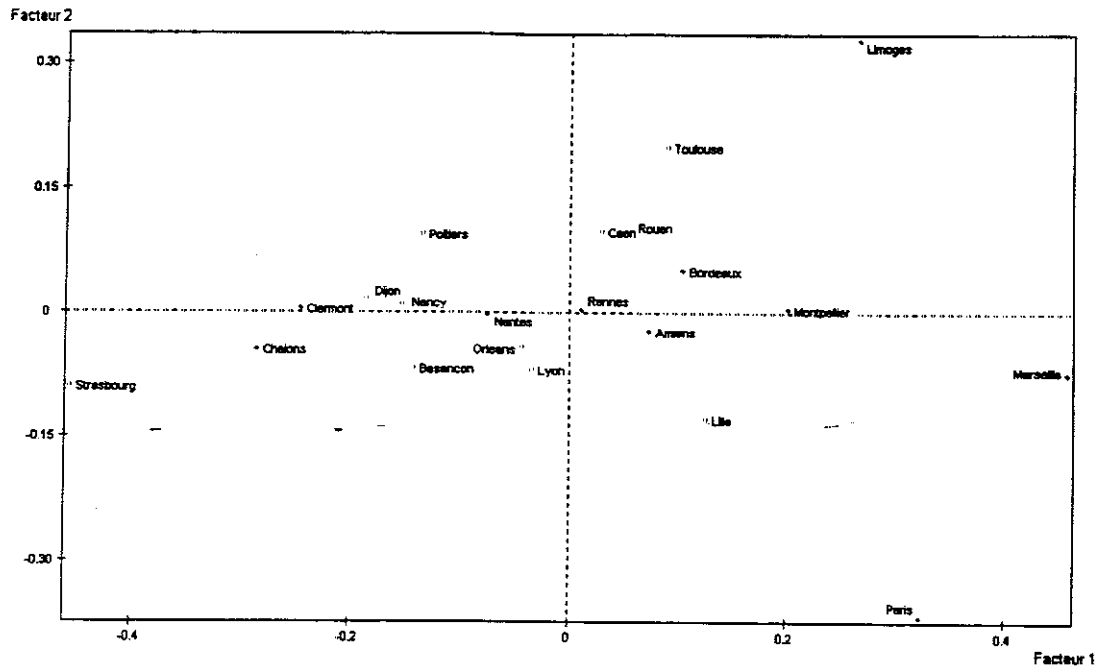


Figure 21 : Représentation des régions (méthode STATIS)

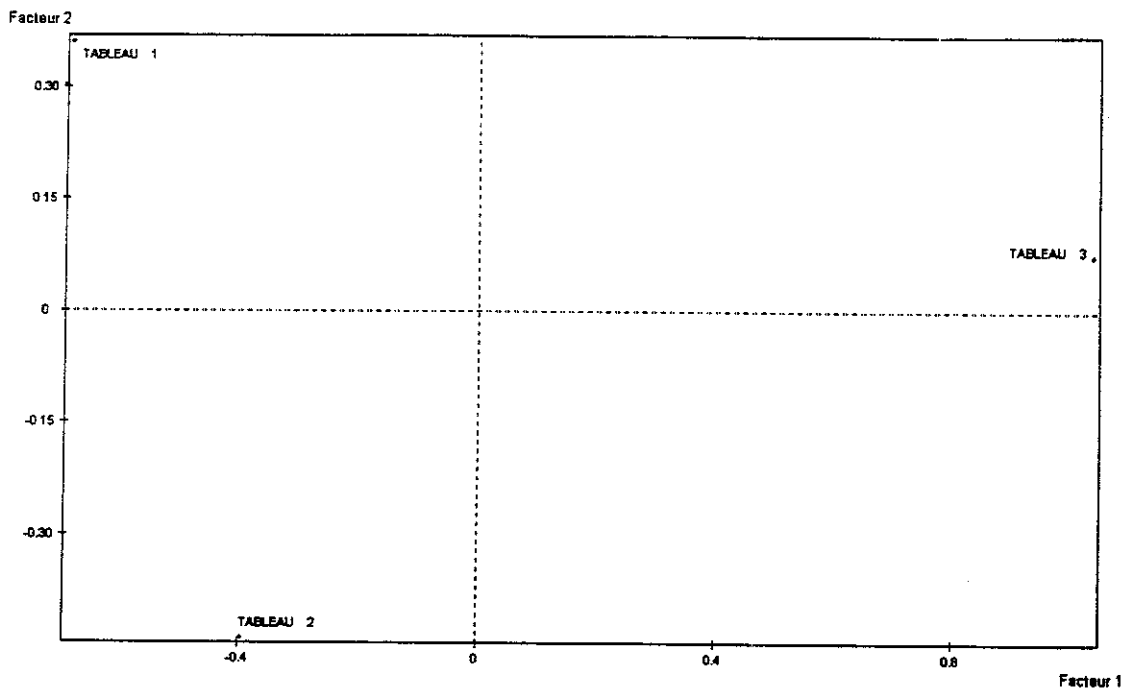


Figure 22 : Représentation des groupes avec la méthode STATIS duale

La méthode STATIS duale s'applique à des tableaux de données définis avec les mêmes variables. Elle permet de comparer les matrices de covariances. La méthode STATIS duale et l'analyse intraclasse envisagée comme une ACP conditionnelle (§5.2) donnent des résultats presque équivalents ([DAZ96]) On le constate en comparant les figures 14 et 23. Mais en

opérant directement l'analyse intraclasse du §5.2, on obtient une représentation des différences d'évolution des modalités illustratives (figures 16 et 17), et on aurait pu aussi représenter les différences d'évolution des unités actives.

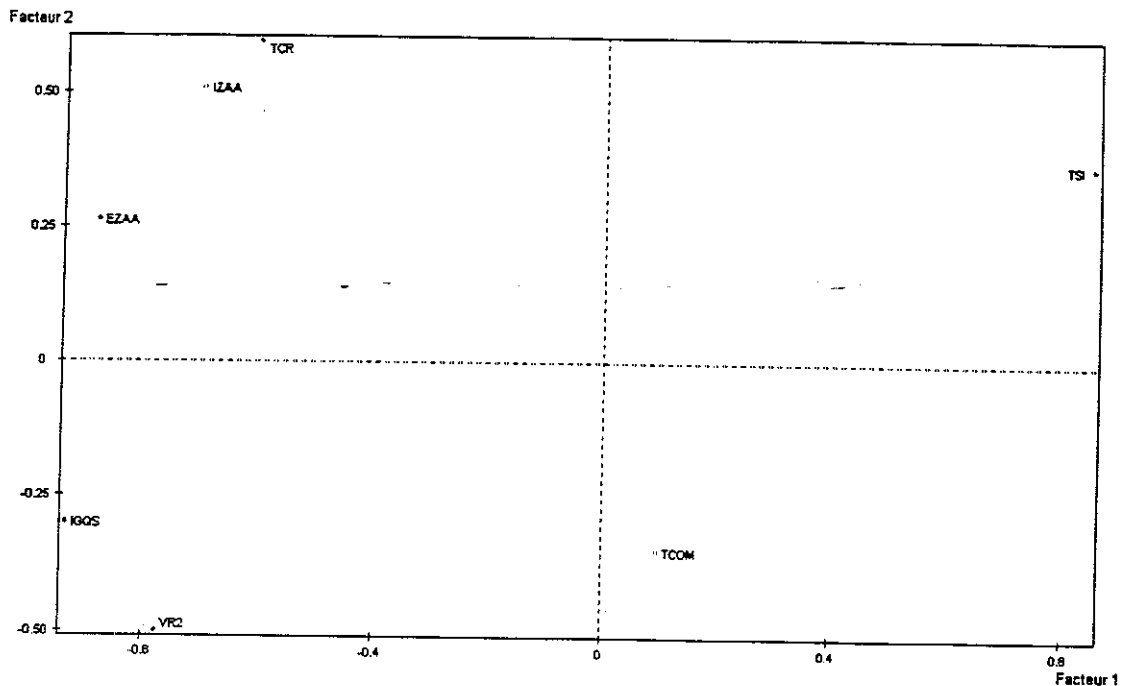


Figure 23 : Représentation des variables (méthode STATIS duale)

## 8. Conclusion

Il est important de souligner qu'aucune des méthodes ne prend en compte l'aspect temporel des données. Si on permute les  $T$  tableaux de données, les analyses donnent les mêmes résultats. L'ordre chronologique est récupéré au niveau des interprétations.

La méthodologie est choisie en fonction des objectifs selon qu'on s'intéresse à l'évolution des variables, des individus, à l'évolution globale, à l'interstructure, à l'intrastructure ....

La méthodologie est aussi choisie en fonction du type de données : mêmes individus, mêmes variables, variables quantitatives, variables qualitatives. Seule la méthode STATIS ne permet pas de traiter des tableaux de données avec variables qualitatives.

Si les données possèdent une structure un tant soit peu forte, les résultats donnés par les différentes méthodologies sont cohérents. Il faut choisir sa stratégie en fonction du type des données et des objectifs.



**Annexe : Les tableaux de données (extraits de [JAM89])****Indicateurs de qualité de service du réseau téléphonique : Variables actives**

**IGQS** : Indice global de qualité de service (combinaison de 6 autres indicateurs dont IZAA, EZAA, TSI et VR2),

**IZAA** : Pourcentage d'appels efficaces par rapport au nombre des appels correctement effectués par les usagers en ne tenant compte que des appels intérieurs à la zone d'autonomie d'acheminement ou à la zone urbaine,

**EZAA** : Définition identique à l'indicateur précédent en remplaçant « intérieurs » par « extérieurs »,

**TSI** : Taux de signalisation des dérangements (pourcentage du nombre de signalisation de dérangements reçus et traités durant une période donnée par rapport au nombre de lignes principales en service),

**VR2** : Vitesse de relevé des dérangements dans les 2 jours (pourcentage de dérangements relevés le jour même et le lendemain de la signalisation par rapport au nombre de dérangements qui ont donné lieu à déplacement d'agents),

**TCR** : Taux d'efficacité à l'arrivée des centres de renseignements (pourcentage d'appels efficaces par rapport au nombre des appels au service des renseignements),

**TCOM** : Temps d'établissement des communications (délai exprimé en secondes de mise en relation de deux abonnés).

**Indicateurs technico-économiques : Variables illustratives**

**ACLP** : Accroissement brut du parc de lignes principales,

**DLAI** : Délai moyen de raccordement,

**QRAC** : Indice de qualité de raccordement,

**SIXM** : Demande en instance inférieure à 6 mois,

**UNAN** : Demande en instance inférieure à 1 an,

**DEMI** : Délai de réalisation inférieur à 15 jours,

**DEM2** : Délai de réalisation inférieur à 30 jours,

**TFPB** : Trafic financier des postes publiés.

Régions	Indicatif tél.	IGQS1	IZAA1	EZAA1	TSI1	VR21	TCR1	TCOM1
Ajaccio	4	61,4	58,6	55,6	43,3	53	31,6	13,4
Amiens	3	89,1	66,3	66,2	28	83,9	68	18,7
Besançon	3	95,7	69,5	68	19	87,5	68,3	15,2
Bordeaux	5	89,3	67,2	66,1	29,2	84,6	70,6	15,9
Caen	2	88	67,1	66,6	31,5	82,9	74,1	15,6
Châlons	3	95,8	67,9	68,1	22,5	93,4	78,9	16,6
Clermont	4	96	70,8	69,3	22,5	87,5	75,6	16,6
Dijon	3	93,4	68,1	68,3	23,2	86,7	74,6	16,1
Lille	3	91,8	67,1	67,1	28,5	90,2	58,5	16,3
Limoges	5	83,4	66,7	64,7	38,6	78,9	66,4	15,5
Lyon	4	93,5	67,4	68,4	21,2	86,1	67,8	17,3
Marseille	4	87,1	64	63,9	29,7	86,3	54,4	16,2
Montpellier	4	89,3	66,3	64,7	27,6	85,9	60,7	15,5
Nancy	3	93,8	67,6	68,9	25,4	89,7	73,1	16,6
Nantes	2	91,8	67,8	67,8	28,9	88,3	74,3	16,7
Orléans	2	92,7	67,6	67	25,7	89	70,2	18,5
Poitiers	5	90,2	67,7	66,8	28,5	84,3	74,2	16,3
Rennes	2	91,5	68,3	67,9	28,1	85,5	66,2	17,3
Rouen	2	87,4	68,1	67,8	31,5	78,7	71,3	15,1
Strasbourg	3	97,8	67,9	70,1	18,8	93,5	82	15,8
Toulouse	5	88	67,6	66,9	30,5	80,3	70,2	16,2
Paris	1	98,2	66,6	68,4	23,2	89,8	54,1	17,1
Moyenne		91,6	67,5	67,3	26,8	86,3	69,2	16,4
Ecart-type		3,8	1,3	1,5	4,6	4,0	7,2	0,9

**Tableau 1 : Régions Télécom – indicateurs de qualité pour l'année 1984**

Régions	Indicatif tél.	IGQS2	IZAA2	EZZA2	TSI2	VR22	TCR2	TCOM2
Ajaccio	4	74,6	61,7	60,2	35,3	67,4	42,1	13,7
Amiens	3	93,6	66,8	67,6	21	87,6	68,1	16,8
Besançon	3	96,1	68,9	68,3	18,5	88,4	76,9	15,4
Bordeaux	5	92,2	66,9	67,2	23,9	86,6	68,7	16,3
Caen	2	92	66,8	67	25,7	88,3	72,3	18
Châlons	3	97,6	68,5	69,3	19,9	93,9	80,7	16,5
Clermont	4	97,9	70,8	70,8	19,6	89,1	80	16,7
Dijon	3	96,1	69	69,3	20,2	88,9	80	16,6
Lille	3	95	67,8	68,3	22	89	69,8	16,3
Limoges	5	87	66,3	67,4	33,6	81,6	72,2	15,7
Lyon	4	95	76,4	69	20,1	88,4	71,8	17,6
Marseille	4	89,1	64,5	64,1	25,9	87,3	53,4	16,6
Montpellier	4	91,6	66,5	65,9	23,6	87,2	63,6	16,1
Nancy	3	96	68	69,4	22,2	91,8	78,7	17,2
Nantes	2	95,6	68,3	69,4	23,8	91,5	73,8	16,6
Orléans	2	94,6	68,2	68,1	23,3	90,4	72,2	16,5
Poitiers	5	94,5	68,7	68	22,6	88,7	78,4	18
Rennes	2	94,6	68,4	68,4	22,2	88,5	69,7	16,9
Rouen	2	93,4	68,6	68,7	25,4	87,5	72,6	15,8
Strasbourg	3	100,5	69,6	72	16,8	94,4	86	15,5
Toulouse	5	90,3	67,6	68,1	27,1	81,8	73,8	16,3
Paris	1	99,3	66,8	69	20,9	90	53,9	17,3
Moyenne		94,4	68,3	68,3	22,8	88,6	72,2	16,6
Ecart-type		3,2	2,2	1,6	3,5	3,0	7,9	0,7

**Tableau 2 : Régions Télécom – indicateurs de qualité pour l'année 1985**

Régions	Indicatif tél.	IGQS3	IZAA3	EZAA3	TSI3	VR23	TCR3	TCOM3
Ajaccio	4	77,6	64	61,1	34,3	68,5	66,5	15,4
Amiens	3	97,3	69,4	70,9	18,7	87,6	69,6	18,7
Besançon	3	96,6	69,3	69,2	17,9	87,4	79,3	19,1
Bordeaux	5	93,6	68,5	68,6	22	84,8	67,8	18
Caen	2	95,1	68,4	69	23,3	90,4	70,9	17,3
Châlons	3	98,9	70,2	72,1	19,2	90,6	81	17,2
Clermont	4	98,1	71,1	72,3	20,1	86,7	82,2	18,1
Dijon	3	97,4	69,9	71,5	19,6	87,9	78,5	19,3
Lille	3	97,2	69	70,2	18,8	90,1	65,5	17,5
Limoges	5	88,8	69,6	70,4	31,2	75,6	66,3	19
Lyon	4	95,8	68,8	69,9	19,3	86,5	73,1	18,4
Marseille	4	89,5	65,6	64,8	24,3	84,6	58,6	18,4
Montpellier -	4	91,2	68	67,2	23,8	82,5	71	19,3
Nancy	3	95,8	69	70,5	21	87,2	78,2	16,5
Nantes	2	98,1	69,3	71,3	20,7	90,9	73,3	18,6
Orléans	2	98,2	70,4	72,2	20,1	88,6	73,7	17,4
Poitiers	5	96,4	69,7	70,1	20,9	88,3	80,4	19,4
Rennes	2	95,5	69,6	69,8	22	86,9	77,2	18,5
Rouen	2	95,6	70,3	70,1	20,7	86,9	69,2	17,3
Strasbourg	3	102,4	71	73,8	15,6	94,3	86,4	16,6
Toulouse	5	91,3	68,8	69,1	25	79,9	72,1	18,2
Paris	1	101	68,2	70,2	18,4	90,1	52,8	18,2
Moyenne		95,9	69,2	70,1	21,1	87,0	72,7	18,1
Ecart-type		3,4	1,2	1,9	3,2	4,0	7,8	0,8

Tableau 3 : Régions Télécom – indicateurs de qualité pour l'année 1986

Régions	ACLP	DLAI	QRAC	SIXM	UNAN	DEM1	DEM2	TFPB
Ajaccio	79	2	72	5722	17	40	60	58
Amiens	58	0	90	22	0	75	87	7
Besançon	58	0	85	0	0	69	84	13
Bordeaux	60	1	82	2813	0	49	71	17
Caen	57	1	82	18	0	43	70	15
Châlons	50	0	93	1	1	79	90	12
Clermont	48	1	87	16	0	60	82	15
Dijon	58	1	82	0	0	49	72	13
Lille	60	1	84	1000	0	60	80	11
Limoges	54	1	80	61	0	50	74	13
Lyon	52	1	88	7	6	68	87	20
Marseille	51	1	87	35	5	63	79	23
Montpellier	66	1	83	11	0	56	76	22
Nancy	54	0	90	0	0	70	84	13
Nantes	55	0	94	0	0	72	91	15
Orléans	53	0	91	0	0	69	87	13
Poitiers	53	1	81	25	0	51	73	16
Rennes	61	1	83	1	0	44	70	16
Rouen	54	1	82	0	0	53	73	11
Strasbourg	96	0	93	0	0	72	88	15
Toulouse	58	1	86	54	1	54	75	17
Paris	34	0	96	0	0	92	97	31
Moyenne	56,7	0,6	86,6	193,5	0,6	61,8	80,5	15,6
Ecart-type	10,7	0,5	4,7	622,4	1,6	12,4	7,8	5,0

Tableau 4 : Régions Télécom – données économiques pour l'année 1984

**REFERENCES BIBLIOGRAPHIQUES**

- [CAR68] J.D. Carroll (1968) A generalization of canonical correlation analysis to three or more sets of variables, Proc. 76<sup>th</sup> Conv. Amer. Psych. Assoc., 227-228.
- [CAZ90] P. Cazes (1990) Quelques Considérations sur l'Utilisation des Eléments Supplémentaires en Analyse Factorielle. Revue MODULAD n°5, pp. 29-44.
- [DAZ96] F. Dazy, J.-F. Le Barzic (1996) L'analyse des données évolutives, méthodes et applications. Technip.
- [ESC98] B. Escofier, J. Pagès (1998) Analyses Factorielles Simples et Multiples. 3<sup>e</sup> édition, Dunod.
- [JAM89] M. Jambu (1989) Exploration informatique et statistique des données. Dunod.
- [LAV88] Ch. Lavit (1988) Analyse conjointe de tableaux quantitatifs. Masson.
- [LEB99] L. Lebart, A. Morineau, M. Piron (1999) Statistique exploratoire multidimensionnelle. 2<sup>e</sup> édition, Dunod.
- [PAG96] J. Pagès (1996) Eléments de comparaison entre l'analyse factorielle multiple et la méthode STATIS. RSA, 44, 4, pp.81-95.
- [ROB76] P. Robert, Y. Escoufier (1976) A unifying tool for linear multivariate methods : the RV coefficient, Applied Statistics, 25, 3, 257-265.
- SPAD version 4.5 (2000) Système Portable d'Analyse des Données, CISIA.CERESTA.
- SPAD • TM (2000) Analyse des Tableaux Multiples, Manuel de référence, CISIA.CERESTA.