

REVUE DE STATISTIQUE APPLIQUÉE

B. BIECHELER

Un exemple d'application : l'alcoolémie des conducteurs dans une quinzaine de départements

Revue de statistique appliquée, tome 18, n° 4 (1970), p. 59-69

http://www.numdam.org/item?id=RSA_1970__18_4_59_0

© Société française de statistique, 1970, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

UN EXEMPLE D'APPLICATION : L'ALCOOLEMIE DES CONDUCTEURS DANS UNE QUINZAINE DE DÉPARTEMENTS

par M. B. BIECHER

Ministère de l'Équipement et du Logement - Paris

LES DONNEES :

Distribution de l'alcoolémie par départements

	1 0-0,09	2 0,1-0,49	3 0,5-0,79	4 0,8-1,09	5 1,09
13	323	238	12	6	6
14	130	286	8	7	7
28	213	157	13	3	4
33	281	195	23	6	3
44	153	291	36	10	10
45	189	226	19	6	5
54	198	201	22	3	8
59	197	216	22	20	7
63	253	154	25	3	7
69	291	228	17	11	8
74	265	219	17	11	8
83	276	165	16	1	3
87	199	259	13	7	4
94	270	278	35	8	7

Les niveaux de l'alcoolémie classifiés de 1 à 5 sont mesurés en grammes par litre.

1 - ANALYSE DES CORRESPONDANCES

Les tableaux d'effectifs comme ceux de la présente enquête sont particulièrement propres à être analysés par cette méthode.

Reproduction du Listing des résultats :

DEPARTEMENTS	M	1	2	3
	VALEUR PROPRE TEST POURCENTAGE	0.30407708E-01 0.83051971E 02 72.90	0.57572685E-02 0.40741798E 02 13.80	0.46691373E-02 0.64283084E 01 11.19
Bouches du Rhône	G (701)	4 -0.14578896E 00	-0.91623172E-01	0.13966947E-01
Calvados	G (702)	15 0.37332969E 00	-0.15631091E 00	-0.18080644E-01
Eure et Loir	G (703)	4 -0.14304396E 00	-0.38211956E-01	-0.27674563E-01
Gironde	G (704)	3 -0.16234222E 00	0.2495426E-01	-0.17723333E-01
Loire Atlantique	G (705)	14 -0.32301545E 00	0.12754708E 00	-0.65637692E-01
Loiret	G (706)	11 -0.98953500E-01	-0.13105474E-01	-0.38108385E-01
Meurthe et Moselle	G (707)	9 0.22350624E-01	0.33996149E-01	-0.81734195E-01
Nord	G (708)	12 -0.10742682E 00	0.10216507E 00	0.18364217E-00
Puy de Dôme	G (709)	2 -0.21122837E 00	0.91883436E-01	-0.56768439E-01
Rhône	G (710)	6 -0.91347053E-01	-0.16263317E-01	0.65198853E-01
Haute Savoie	G (711)	7 -0.62332213E-01	-0.52443035E-02	0.68126273E-01
Var	G (712)	1 -0.25763231E 00	-0.38075059E-01	-0.56918189E-01
Haute Vienne	G (713)	13 0.13336908E 00	-0.96438080E-01	-0.12350367E-01
Val de Marne	G (714)	10 0.84939721E-01	0.73161035E-01	-0.54597281E-01
Val d'oise.	G (715)	8 0.11500259E-01	-0.76586986E-02	0.72490781E-01
ALCOOLEMIE	M	1	2	3
	VALEUR PROPRE TEST POURCENTAGE	0.30407708E-01 0.83051971E 02 72.90	0.57572685E-02 0.40741798E 02 13.80	0.46691373E-02 0.64283084E 01 11.19
Classe 1	F (1)	-0.18155163E 00	-0.74587986E-02	0.73168929E-02
" 2	F (2)	0.17020657E 00	-0.31558498E-01	-0.15369793E-01
" 3	F (3)	0.61094544E-01	0.33802384E 00	-0.12360608E 00
" 4	F (4)	0.25874292E 00	0.17958027E 00	0.50022280E 00
" 5	F (5)	0.16238617E 00	0.11989800E 00	0.55816873E-01

Les départements sont représentés dans un espace à 5 dimensions ; la distance du χ^2 calculée dans cet espace entre points-départements rend proches les départements de même profil par rapport à l'alcoolémie. Le nuage de points est alors projeté dans le meilleur sous-espace qui traduise ces proximités.

Les données ont été traitées par le programme ANAFA qui donne les résultats suivants :

ALCCOLEMIE-DEPARTEMENTS-COORDONNEES DES DEPARTEMENTS SUR LES 3 PREMIERS AXES FACTORIELS

TEST = 306,5

C'est l'inertie totale du nuage

Sur le listing des résultats :

Les valeurs successives de TEST représentant la part d'inertie non exprimée par les facteurs successifs extraits.

POURCENTAGE, le pourcentage d'inertie exprimée par les facteurs.

On trouvera ci-joint la représentation simultanée des départements et des niveaux d'alcoolémie dans le plan des deux premiers axes factoriels.

Les valeurs successives de TEST permettent d'affirmer que la part d'inertie du nuage de points non expliquée par les deux premiers facteurs est de l'ordre des aléas statistiques.

L'analyse factorielle de la correspondance alcoolémie-départements d'enquête faite montre qu'on ne peut pas conclure à une classification en deux catégories des départements : les alcoolisés et les non alcoolisés. Les oppositions entre départements sont plus nuancées. Les deux premiers facteurs extraient environ 90 % de l'inertie - dont 73 % pour le premier facteur - ce qui permet d'obtenir une bonne représentation dans le plan des deux premiers axes factoriels.

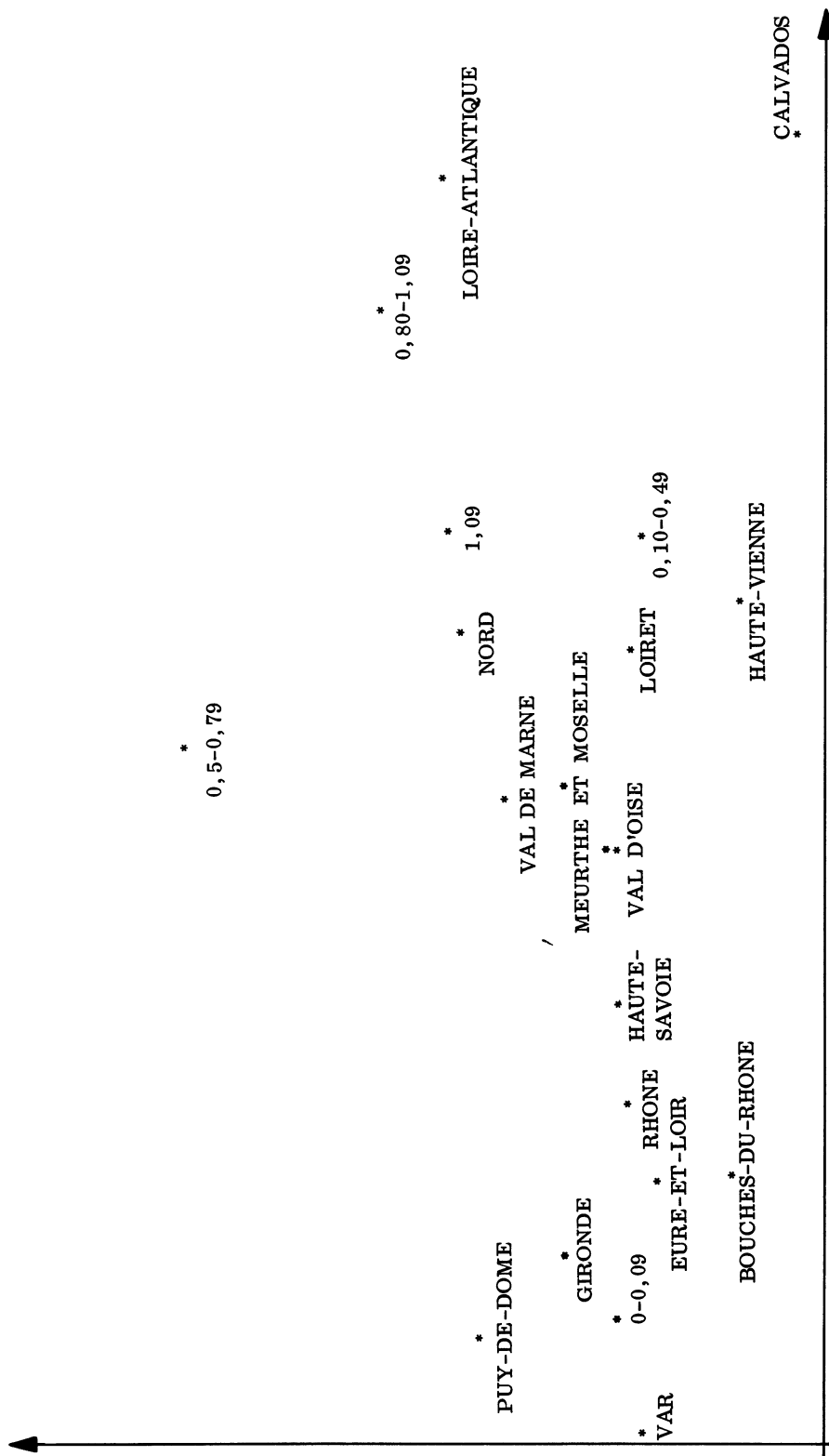
Les départements, en projection sur le premier axe, sont rangés de gauche à droite selon les pourcentages croissants de conducteurs sobres (dont l'alcoolémie est inférieure à 0,1 g/l) * . Par exemple, le Calvados et la Loire Atlantique (30 % de conducteurs sobres) s'opposent au Var et au Puy de Dôme (60 % de conducteurs sobres), alors que la Savoie, la Meurthe et Moselle ou le Loiret sont des départements dans lesquels la proportion des conducteurs sobres est moyenne. On peut donc dire que le premier facteur qui différencie les départements est le taux de sobriété.

La classe d'alcoolémie 1 est discriminée de l'ensemble des classes 2, 3, 4, 5.

En projection sur le second axe, le Nord où le nombre d'alcoolisés est relativement fort (conducteurs d'alcoolémie supérieure à 0,8 g/l) s'oppose à la Haute Vienne ou aux Bouches du Rhône qui proportionnellement comptent le moins de conducteurs alcoolisés.

Ainsi du haut en bas on distingue trois strates de départements de forte à faible alcoolisation.

* Voir liste page 63.



ANALYSE FACTORIELLE DE LA CORRESPONDANCE ALCOOLEMIE-DEPARTEMENTS

La position d'un département dans le plan permet de le caractériser selon le double critère taux de sobriété et taux d'alcoolisation ou d'ébriété.

Par exemple : si les proportions de conducteurs alcoolisés sont voisines en Loire Atlantique et Puy de Dôme, les taux de sobriété différencient nettement ces deux départements.

Par ailleurs, sur le graphique, sont représentés en même temps que les départements, les classes d'alcoolémie ; les proximités entre les divers points peuvent être interprétés :

Un département est d'autant plus proche d'une classe d'alcoolémie que cette classe intervient fortement dans le profil d'alcoolisation du département. On constate que chez les conducteurs du Nord l'alcoolémie est généralement élevée ($> 1,1$ g/l). En Gironde tout au contraire, la grande majorité des conducteurs paraît faire partie des abstinents (alcoolémie $< 0,09$ g/l).

Insistons sur l'intérêt d'une telle analyse qui permet de situer schématiquement les départements par rapport au phénomène d'alcoolémie, selon leurs positions respectives dans le plan.

- Liste ordonnée des départements selon les pourcentages de conducteurs dans la classe 1 :

Var	60,1
Puy de dôme	57,2
Gironde	55,3
Bouches du Rhône	55,2
Eure et Loire	54,6
Rhône	52,4
Haute Savoie	51
Val d'Oise	47,3
Meurthe et Moselle	45,8
Val de Marne	45,1
Nord	42,6
Loiret	42,5
Haute Vienne	41,3
Loire Atlantique	30,6
Calvados	29,7

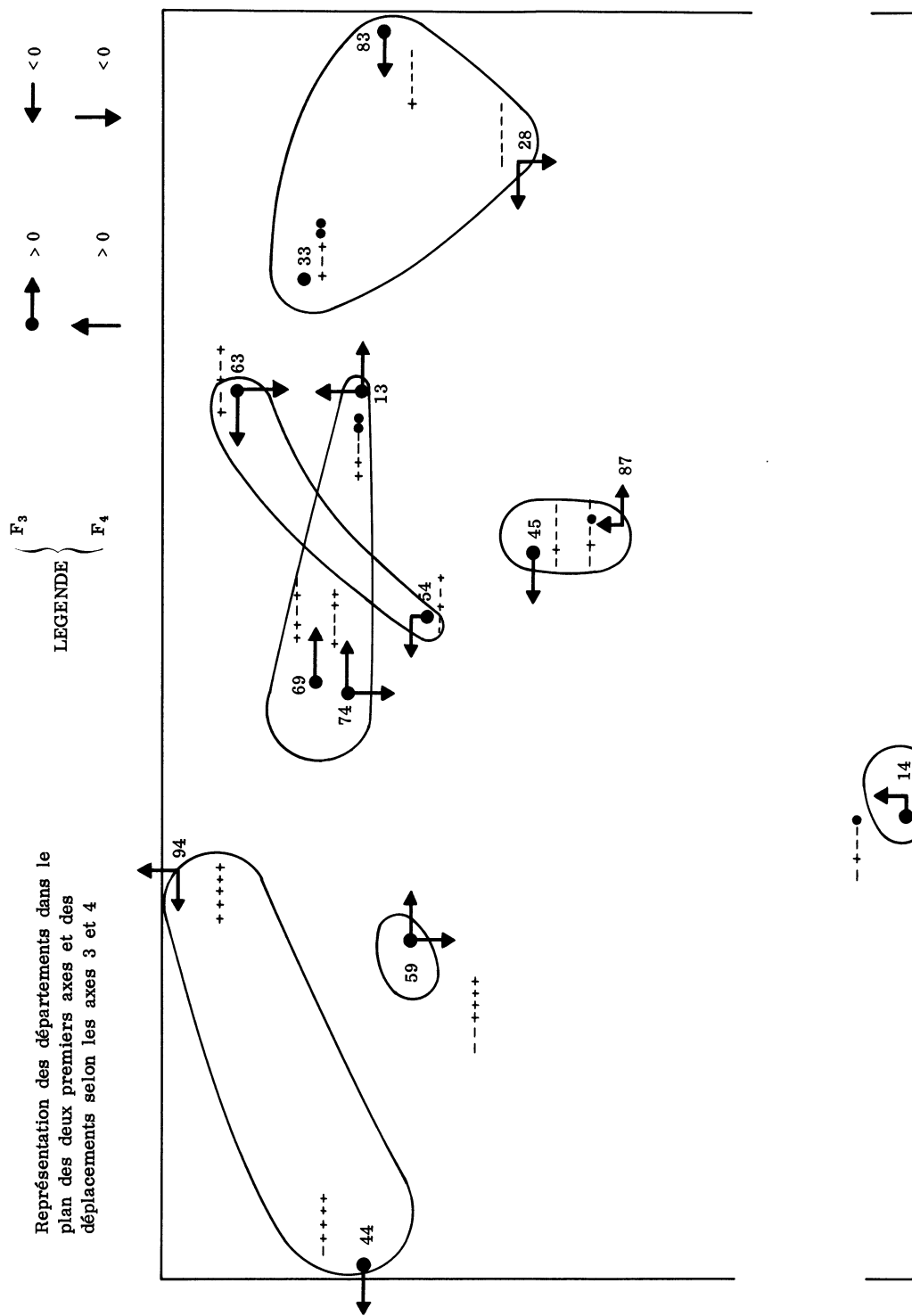
c'est la reproduction exacte de la liste ordonnée des départements selon le premier facteur. Ceci tient aux forts pourcentages d'enquêtés dans la classe 1.

Une représentation des départements dans le plan des axes 1 et 2 s'impose si on veut différencier les départements selon leur ébriété.

- Les proximités des départements ainsi traduites dans le plan peuvent être confrontées aux résultats obtenus par la classification automatique décrite au paragraphe 3.

2 - ANALYSE EN COMPOSANTES PRINCIPALES

Les données ont été soumises à un second type d'analyse factorielle, sans préjugé quant au rôle explicatif d'une analyse en composantes principales sur une distribution unidimensionnelle, afin de comparer la typologie des départements obtenus et les résultats de classification.



- Disposition géométrique dans un espace à 5 dimensions :

Les 14 points-observation (les départements d'enquête) sont disposés dans l'espace des variables, à 5 dimensions (les 5 niveaux d'alcoolémie).

Les proximités entre points-départements seront représentés dans le plan des deux premiers axes factoriels (les 2 premières composantes principales du nuage de points).

On trouvera ci-dessous une représentation avec des indications sur la déformation du nuage selon le troisième axe factoriel.

L'analyse consiste à rechercher les vecteurs propres de la matrice des corrélations des variables.

Matrice de corrélations des variables

	1	2	3	4	5
1	1	<u>-0,381</u>	-0,023	-0,152	<u>-0,267</u>
2		1	0,203	<u>0,410</u>	<u>0,469</u>
3			1	0,188	<u>0,438</u>
4				1	<u>0,449</u>
5					1

1, 2, 3, 4, 5 sont les 5 niveaux d'alcoolémie et les chiffres du tableau les coefficients de corrélation entre niveaux.

Les liaisons entre les niveaux sont étudiés graphiquement par ailleurs

Ces coefficients ne mesurent que l'intensité linéaire de la liaison ; ils sont ici assez faibles. Les coefficients soulignés sont significatifs.

Le niveau 1 est corrélé négativement avec tous les autres niveaux ce qui était prévisible.

Les niveaux 2 et 5 sont les plus fortement corrélés.

Composantes principales

● Contribution des premières composantes successives dans l'explication de la variation :

Composante	contribution cumulée
1	0,45
2	0,66
3	0,81
4	0,91

Corrélations entre les composantes principales et les variables (niveaux d'alcoolémie).

Composante niveau d'alcoolémie.	C ₁	C ₂	C ₃	C ₄	C ₅
1	0,501	0,685	0,420	0,301	-0,111
2	-0,768	-0,260	0,063	0,570	-0,100
3	-0,523	0,673	0,432	0,030	-0,290
4	-0,683	0,049	0,634	0,284	-0,216
5	-0,821	0,199	0,055	0,093	0,523

Le coefficient de corrélation entre la variable i et le facteur " k " est proportionnel à la i^{e} composante de l'axe factoriel k . Ainsi en portant les points-variables sur les deux premiers axes on peut interpréter leurs proximités en terme de corrélation :

- Le premier axe oppose la classe d'alcoolémie 1 aux autres classes 2, 3, 4, 5. L'opposition est la plus forte pour la classe 5. Les classes 2 et 5 sont également corrélés avec le premier facteur.

- Le deuxième axe oppose les classes 1 et 3 aux classes 2, 4, 5 et surtout à la classe 2.

- Le troisième facteur est corrélé aux classes 4 et 1 et 3

- Le quatrième facteur est corrélé à la classe 2 très distinctement par rapport aux autres classes

- Le cinquième facteur est essentiellement corrélé avec la classe 5

Les départements dans l'espace des deux premiers axes factoriels

Sur le premier axe s'opposent les départements à forte proportion d'enquêtés dans la classe 5 comparativement aux autres classes (et plus spécialement à la classe 1) :

{28-33-83} et {44}

Sur le second axe s'opposent les départements à densité d'enquêtés dans les classes 1 et 3

{14} et {94}, ou {14-87} et {94-63-33}

Le troisième axe oppose {44} à très forte proportion d'enquêtés dans la classe 3 et {59} à très forte proportion d'enquêtés dans la classe 4.

Dans la représentation ci-jointe, pour chaque département les écarts positifs ou négatifs des effectifs d'enquêtés dans les diverses classes d'alcoolémie, par rapport aux moyennes dans les classes, sont respectivement notées + ou -. L'examen de ces séquences de + ou - fait apparaître qu'un facteur de taille entre dans l'ordre obtenu selon l'axe 1.

Les départements regroupés dans une même classe par les méthodes de classification ci-dessous ont été entourés d'une même ligne.

3 - LA CLASSIFICATION DES DÉPARTEMENTS

Deux algorithmes de classification automatique ont été appliqués au tableau de données alcoolémie-départements :

1/ Classification ultramétrique ("Single Linkage", programme PECLAT)

2/ Classification de VAN DEN DRIESSCHE ("Average Linkage", programme PELLAS)

Ils utilisent le tableau de distances entrée départements calculé par le programme DISGE.

a - La classification ultramétrique

C'est au cours des premières étapes de la classification que s'opèrent les rapprochements les plus significatifs. La partition des départements opérés à l'étape 3 est la suivante :

{83,28}, {87,45}, 33}

{63,54}

{74,69}, 13}

94

44

59

14

b - La classification de Van Den Driessche

La partition opérée par cette classification est la suivante :

{54, 63}

{69, 74, 13}

{{28, 83, 33}, {45, 87}}

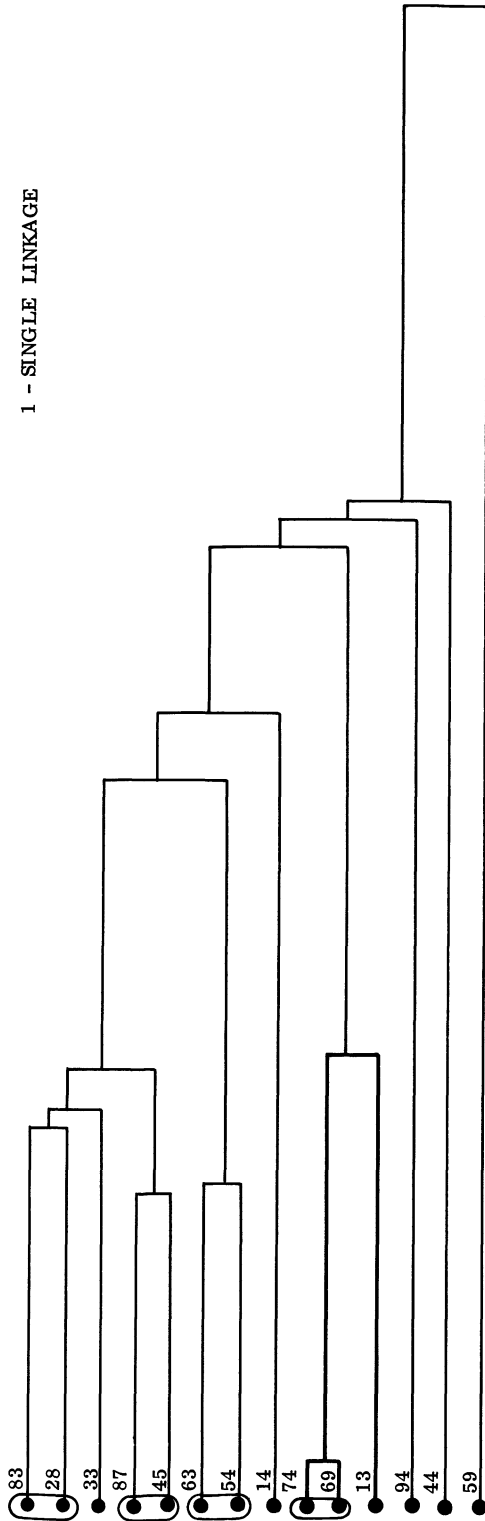
{44, 94}

{14}

{59}

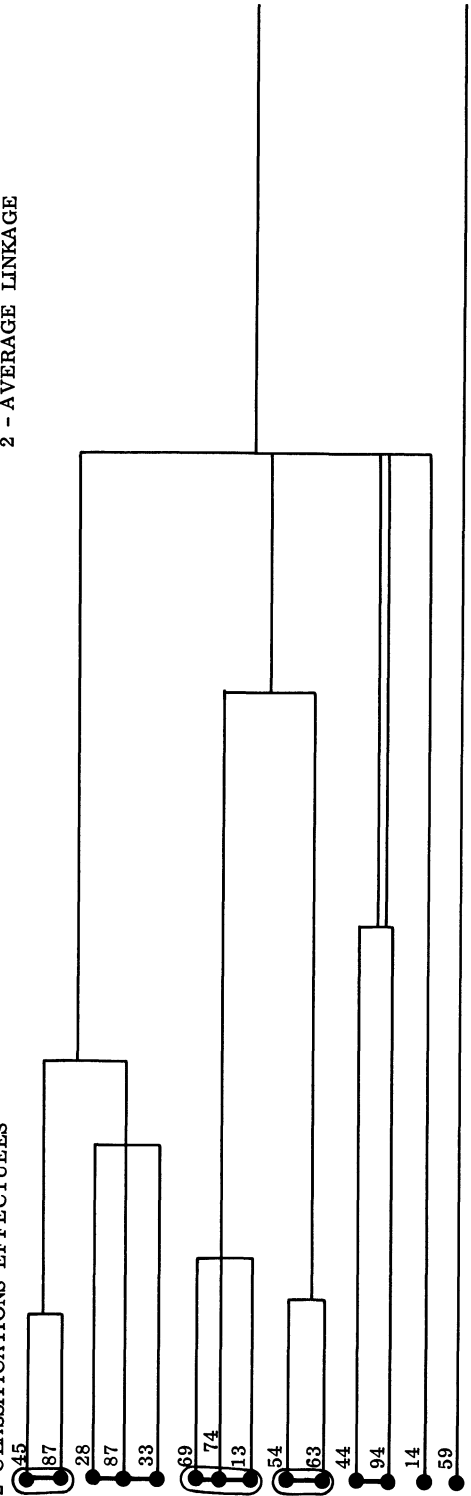
L'état des partitions à la phase 3 est quasiment le même (au regroupement 44, 94 près) que dans l'algorithme précédent.

1 - SINGLE LINKAGE



2 - AVERAGE LINKAGE

Les 2 CLASSIFICATIONS EFFECTUEES



EN CONCLUSION

On peut regretter que l'étude n'ait pas porté sur les 90 départements et se situe dans un contexte unidimensionnel ; l'usage descriptif et typologique des techniques utilisées s'en trouve un peu affaibli.

Les trois méthodes sans être équivalentes ont chacune leur apport dans l'interprétation des données étudiées et leurs résultats sont cohérents.

RESUME

Divers traitements statistiques ont été appliqués à la correspondance alcoolémie-départements (analyse factorielle des correspondances, analyse en composantes principales et classification automatique). Les trois méthodes jouent un rôle complémentaire descriptif ou explicatif dans la constitution de groupes des départements semblables.