

HÉLÈNE MASSÉ

**Méthode de classement des départements à partir de la mortalité
par cause : application de l'analyse discriminatoire**

Journal de la société statistique de Paris, tome 118, n° 4 (1977), p. 311-320

http://www.numdam.org/item?id=JSFS_1977__118_4_311_0

© Société de statistique de Paris, 1977, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

MÉTHODE DE CLASSEMENT DES DÉPARTEMENTS A PARTIR DE LA MORTALITÉ PAR CAUSE : APPLICATION DE L'ANALYSE DISCRIMINATOIRE

Docteur Hélène MASSÉ

Unité de recherches sur les mécanismes physiopathologiques des nuisances de l'environnement (U. 123), I.N.S.E.R.M.

On a appliqué la méthode de l'analyse discriminatoire pour classer les départements à partir des causes de décès. On a utilisé les taux de mortalité, par sexe, du groupe d'âge 45-64 ans, au cours de la période 1968-1970, pour neuf causes principales de décès. Les résultats obtenus montrent le rôle important joué dans la détermination de la fonction discriminante par les lésions vasculaires cérébrales, les accidents et les maladies du cœur.

The discriminant method has been applied to file the French « Départements » from causes of death. We used the mortality rate by sex, for the 45-64 age group, during the 1968-1970 period, for nine main causes of death. The results obtained show the important role played in the determination of the discriminant fonction by cerebral vascular diseases, accidents and heart diseases.

Man verwendet die Methode der diskriminatorischen Analyse um die französischen Départements zu klassifizieren, indem man von den Todesursachen ausgeht. Man verwendet dafür den Prozentsatz der Sterblichkeit, nach dem Geschlecht, für die Altersgruppen von 45-64 Jahren für den Zeitabschnitt 1968-1970. Man verwendet neun Todesursachen. Die Resultate zeigen als Hauptursachen der Todesfälle die Erkrankungen der Hirngefäße, die Unfälle und die Herzkrankheiten.

INTRODUCTION

Les tables de mortalité établies par l'Institut national de la statistique et des études économiques donnent, pour chaque département, des indications sur l'évolution de la mortalité par âge suivant le sexe. Si on veut classer les départements suivant leur niveau de mortalité, on utilise le plus souvent l'espérance de vie à la naissance, indice calculé à partir de la table de mortalité et qui représente la durée moyenne de vie pour un individu soumis depuis la naissance à la mortalité indiquée par la table considérée [3].

On peut également chercher un indicateur plus élaboré. Parmi les travaux entrepris dans ce domaine, on peut citer une étude réalisée en 1968 par Damiani [1] qui a appliqué la méthode de l'analyse discriminatoire aux quotients de mortalité tirés des tables de mortalité départementales.

On a essayé, dans cette étude, d'appliquer cette méthode de l'analyse discriminatoire aux taux de mortalité départementaux des principales causes de décès. La fonction discriminante ainsi trouvée permet de classer les départements et montre quelle est l'influence des différentes causes de décès dans ce classement.

DONNÉES DE BASE

Les données de base sont les statistiques départementales des causes de décès, par âge et sexe, pour la période 1968-1970 [5].

Pour éliminer l'influence de la structure par âge, on n'a considéré qu'un seul groupe d'âge. On a choisi le groupe 45-64 ans car les décès sont suffisamment nombreux et représentatifs de la mortalité.

Les causes de décès retenues sont indiquées ci-après avec mention des numéros correspondants de la nomenclature internationale, 8^e Révision, 1968 (liste abrégée et certains numéros de la liste détaillée) :

1 — Tuberculose toutes formes	B 5, B 6
2 — Cancers	B 19
3 — Diabète	B 21
4 — Maladies du cœur	B 25 - B 29
5 — Lésions vasculaires cérébrales	B 30
6 — Alcoolisme et psychose alcoolique (liste détaillée)	303, 291
7 — Cirrhose du foie	B 37
8 — Accidents	BE 47, BE 48
9 — Suicides	BE 49

Correction des données

Dans la statistique des causes de décès, il existe une proportion non négligeable de décès pour lesquels la cause n'est pas spécifiée : cause non déclarée, cause mal définie (mentions de manifestations purement symptomatiques le plus souvent) et décès attribués à la sénilité (rubrique B 45 de la liste abrégée). Cette proportion atteint 8,2 % des décès en 1969 et varie beaucoup d'un département à l'autre.

Pour corriger les statistiques de causes de décès des causes non spécifiées, on a utilisé une méthode inaugurée par l'I.N.S.E.E. et exposée par Ledermann [4]. On suppose que la proportion des décès de cause k dissimulés dans les décès de cause non spécifiée est constante quel que soit le département, pour le groupe d'âge et le sexe considérés. Cette proportion est alors égale au coefficient de régression changé de signe de la proportion des décès observés de cause k en fonction de la proportion des décès de cause non spécifiée, calculé à partir des données départementales. On rectifie les décès observés de cause k en y ajoutant la proportion ainsi calculée des décès de cause non spécifiée attribuables à la cause k .

On calcule des taux de mortalité corrigés par cause en divisant les nombres corrigés de décès par la population correspondante. On dispose donc, pour chaque sexe et par département, des taux corrigés de mortalité du groupe d'âge 45-64 ans, pour chacune des 9 causes définies précédemment.

PRINCIPE DE LA MÉTHODE

On définit tout d'abord deux ensembles de départements :

- Ensemble 1 : constitué par les départements où plus de 5 taux de mortalité sont supérieurs à la moyenne. Cet ensemble est dit « à mortalité forte ». Il comprend 33 départements pour le sexe masculin et 22 pour le sexe féminin.
- Ensemble 2 : constitué par les départements où moins de 3 taux sont supérieurs à la moyenne. Cet ensemble est dit « à mortalité faible ». Il comprend 27 départements pour le sexe masculin et 22 pour le sexe féminin.
- Ensemble 3 : constitué par les autres départements.

L'analyse discriminatoire permet de remplacer les 9 taux de mortalité par cause par une variable unique y , fonction linéaire de ces taux, telle que cette nouvelle variable permette de classer les départements de l'ensemble 3 dans chacun des ensembles 1 ou 2 avec une probabilité d'erreur minimum. Cette variable y est la *fonction discriminante* [2].

Expression de la fonction discriminante

On note, pour chacun des ensembles 1 et 2 respectivement :

\bar{x}_{i1} et \bar{x}_{i2} les moyennes des taux de mortalité de la cause i ($i = 1, 2, \dots, 9$),

\bar{x}_1 et \bar{x}_2 , les matrices (9×1) composées des moyennes des taux par cause,

S_1 et S_2 , les matrices (9×9) de dispersion des variables.

On pose : $d = \bar{x}_1 - \bar{x}_2$, matrice (9×1),

$S = n_1 S_1 + n_2 S_2$, matrice (9×9),

où n_1 et n_2 sont les nombres de départements des ensembles 1 et 2.

La fonction discriminante y s'écrit :

$$y = l_1 x_1 + l_2 x_2 + \dots + l_9 x_9$$

Si on appelle l la matrice (1×9) des coefficients l_i , on démontre que l'on a :

$$l = k d' S^{-1}$$

où k est un facteur de proportionnalité arbitraire.

Vérification des hypothèses

On calcule la valeur de y pour chaque département. On appelle \bar{y}_1 , \bar{y}_2 les moyennes et $s^2(y_1)$, $s^2(y_2)$ les variances des distributions de y dans les ensembles 1 et 2.

On vérifie que les distributions de y dans chacun des ensembles sont normales et de même variance. On prend comme estimation de la variance commune :

$$s^2 = \frac{n_1 s^2(y_1) + n_2 s^2(y_2)}{n_1 + n_2 - 2}$$

Valeur de séparation

On choisit comme valeur de séparation :

$$y_0 = \frac{1}{2} (\bar{y}_1 + \bar{y}_2)$$



On classe alors chacun des départements de l'ensemble 3 dans l'ensemble 1 ou dans l'ensemble 2, suivant la valeur prise par la fonction discriminante y pour ce département par rapport à la valeur de séparation y_0 .

Probabilité d'erreur

La valeur centrée réduite correspondant à y_0 est :

$$u_0 = \frac{y_0 - \bar{y}_2}{s}$$

La probabilité d'erreur de classement est donnée par : $1 - \Phi(u_0)$ où $\Phi(u_0)$ représente la valeur de la fonction de répartition de la loi normale réduite correspondant à u_0 .

RÉSULTATS

1. Fonction discriminante

On a déterminé le facteur de proportionnalité k de façon que la somme des carrés des coefficients l_i soit égale à 1 :

$$\sum l_i^2 = 1$$

Afin de pouvoir comparer ces coefficients entre eux, on utilise la forme centrée réduite z_i des variables d'origine x_i :

$$z_i = \frac{x_i - \bar{x}_i}{s_i},$$

où \bar{x}_i et s_i représentent la moyenne et l'écart type de x_i .

La fonction discriminante s'écrit alors :

$$y = h_1 z_1 + h_2 z_2 + \dots + h_9 z_9 + a$$

avec :

$$h_i = l_i s_i \quad \text{et} \quad a = \sum l_i \frac{\bar{x}_i}{s_i}$$

Les valeurs des coefficients l_i et h_i par sexe sont données dans le tableau I ci-après.

TABLEAU I

Valeur des coefficients de la fonction discriminante

Cause de décès	Coefficients l_i		Coefficients h_i	
	Sexe masculin	Sexe féminin	Sexe masculin	Sexe féminin
1. Tuberculose	0,209	0,561	1,991	1,771
2. Cancers	0,095	0,054	5,096	1,337
3. Diabète	0,539	0,145	3,714	0,759
4. Maladies du cœur	0,125	0,194	5,449	3,357
5. Lésions vasculaires cérébrales	0,399	0,352	10,130	4,057
6. Alcoolisme	0,092	-0,089	1,643	-0,425
7. Cirrhose du foie	-0,079	0,105	-2,869	2,104
8. Accidents	0,233	0,635	6,466	5,931
9. Suicides	-0,013	0,278	-0,212	1,845

Si on classe les causes de décès suivant la valeur absolue des coefficients correspondants, dans l'ordre décroissant, on obtient les résultats suivants. Pour le sexe masculin, arrivent en tête les lésions vasculaires cérébrales, puis vient le groupe accidents, maladies du cœur, cancers, ensuite le diabète, la cirrhose du foie, la tuberculose, l'alcoolisme; le coefficient des suicides est négligeable. Pour le sexe féminin, on trouve d'abord les accidents, les lésions vasculaires cérébrales, les maladies du cœur, puis la cirrhose du foie, les suicides, la tuberculose, les cancers; on peut négliger les coefficients du diabète et de l'alcoolisme.

2. Vérification des hypothèses, valeur de séparation, probabilité d'erreur

On teste tout d'abord que les distributions de y dans les ensembles 1 et 2 peuvent être considérés comme normales.

On vérifie ensuite l'hypothèse de l'égalité des variances de ces deux distributions en calculant le rapport :

$$F = \frac{n_1}{n_1 - 1} s^2(y_1) / \frac{n_2}{n_2 - 1} s^2(y_2)$$

Les valeurs de F ainsi trouvées, pour chaque sexe, sont inférieures à la valeur de F de la table de Fisher-Snedecor pour $(n_1 - 1)$ et $(n_2 - 1)$ degrés de liberté et le niveau de confiance de 95 %.

On calcule la variance commune s^2 , la valeur de séparation y_0 et la probabilité d'erreur.

Les moyennes et variances des distributions de y dans les ensembles 1 et 2 sont données dans le tableau II, ainsi que la variance commune, la valeur de séparation et la probabilité d'erreur.

TABLEAU II

Caractéristique de la distribution de y

	Sexe masculin	Sexe féminin
<i>Moyenne</i>		
\bar{y}_1	178,697	104,156
\bar{y}_2	140,548	78,059
<i>Variance</i>		
$s^2(y_1)$	280,845	48,904
$s^2(y_2)$	195,533	36,090
<i>Variance commune</i>		
s^2	250,815	44,726
<i>Valeur de séparation</i>		
y_0	159,622	91,107
<i>Probabilité d'erreur</i>		
$1 - \Phi(u_0)$	11,5 %	2,6 %

3. Classement des départements

Les départements de l'ensemble 3, dont la valeur de y est supérieur ou égale à la valeur de séparation y_0 , sont classés dans un sous-ensemble 3.1, appelé « à mortalité forte probable ». Les autres départements de l'ensemble 3 sont classés dans un sous-ensemble 3.2., appelé « à mortalité faible probable ».

On trouvera, dans le tableau III, la liste des départements classés suivant les ensembles 1, 2, 3.1 et 3.2 avec la valeur de la fonction discriminante y correspondante.

Deux cartes illustrent ces résultats.

4. Comparaison avec d'autres données

On constate une très forte analogie entre le classement des départements obtenu dans cette étude et le classement trouvé dans l'étude citée précédemment [1], qui appliquait la méthode de l'analyse discriminatoire aux quotients de mortalité par âge.

Comme les résultats de cette dernière étude étaient en forte corrélation avec, en particulier, l'espérance de vie à la naissance, il en est donc de même pour la fonction discriminante définie ici ($r = -0,77$ pour le sexe masculin, $r = -0,82$ pour le sexe féminin).

TABLEAU III

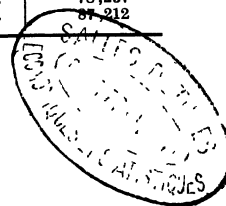
Fonction discriminante par département

Sexe masculin		Sexe féminin	
Départements	Fonction discriminante	Départements	Fonction discriminante
<i>Ensemble 1. Mortalité forte</i>		<i>Ensemble 1. Mortalité forte</i>	
Aisne	164,160	Aisne	103,732
Ardennes	172,610	Ardennes	109,118
Calvados	182,915	Calvados	94,860
Côtes du-Nord	215,676	Côtes du Nord	116,714
Creuse	197,363	Eure	99,581
Eure	163,234	Ille et Vilaine	101,550
Finistère	202,999	Manche	97,676
Ille et Vilaine	208,158	Mayenne	102,477
Isère	160,078	Meurthe et Moselle	96,093
Landes	178,696	Meuse	113,094
Loire	155,512	Morbihan	113,926
Haute Loire	161,450	Moselle	102,388
Loire Atlantique	188,793	Nord	119,385
Lozère	157,230	Orne	107,979
Manche	185,585	Pas de Calais	106,515
Mayenne	173,388	Bas Rhin	98,306
Meuse	155,996	Haut Rhin	108,600
Morbihan	215,595	Haute Saône	99,140
Moselle	192,933	Seine Maritime	95,843
Nord	184,439	Seine et Marne	97,797
Orne	171,495	Somme	108,271
Pas de Calais	179,796	Tarn et Garonne	98,392
Puy de Dôme	173,340		
Bas Rhin	191,659	<i>Ensemble 3. 1. Mortalité forte probable</i>	
Haut Rhin	196,107	Doubs	91,817
Haute Saône	171,499	Eure et Loir	91,224
Savoie	169,430	Loire Atlantique	114,766
Haute Savoie	173,881	Oise	91,179
Seine Maritime	161,668	Pyrénées Atlantiques	92,868
Somme	169,108	Savoie	92,120
Vosges	171,586	Haute Savoie	97,232
Yonne	178,398	Vienne	93,169
Territoire de Belfort	162,220	Vosges	92,663
		Yonne	95,617
<i>Ensemble 3.1. Mortalité forte probable</i>			
Allier	165,400		
Ardennes	166,196		
Doubs	164,417		
Var	185,933		
Vendée	161,441		

TABLEAU III (suite et fin)

Fonction discriminante par département

Sexe masculin		Sexe féminin	
Départements	Fonction discriminante	Départements	Fonction discriminante
<i>Ensemble 1. Mortalité faible</i>		<i>Ensemble 1. Mortalité faible</i>	
Alpes de Haute Provence	145,166	Alpes Maritimes	83,682
Ariège	133,067	Aude	63,134
Aude	126,794	Aveyron	81,768
Aveyron	130,856	Cantal	79,315
Charente	149,633	Charente	82,065
Charente Maritime	133,060	Cher	80,633
Corrèze	160,822	Corse	66,829
Corse	139,354	Côte d'Or	82,608
Côte d'Or	139,524	Drôme	83,041
Drôme	144,402	Gers	81,688
Haute Garonne	148,726	Hérault	72,953
Gironde	158,517	Loiret	87,474
Hérault	122,644	Lot et Garonne	84,676
Indre et Loire	137,488	Lozère	70,462
Jura	149,369	Hautes Pyrénées	81,762
Lot et Garonne	133,698	Pyrénées Orientales	77,824
Pyrénées Atlantiques	127,853	Deux Sèvres	74,167
Sarthe	146,723	Var	75,853
Yvelines	125,732	Vaucluse	79,502
Deux Sèvres	118,616	Haute Vienne	76,973
Tarn	141,994	Hauts de Seine	72,821
Tarn et Garonne	143,414		
Vienne	136,105	<i>Ensemble 3.2. Mortalité faible probable</i>	
Haute Vienne	148,958	Ain	90,795
Essonne	127,936	Allier	90,089
Hauts de Seine	140,957	Alpes de Haute Provence	89,194
Val de Marne	132,264	Hautes Alpes	88,572
		Ardèche	78,420
<i>Ensemble 3.2. Mortalité faible probable</i>		Ariège	85,905
Ain	152,220	Aube	72,471
Hautes Alpes	129,220	Bouches du Rhône	85,440
Alpes Maritimes	145,177	Charente Maritime	80,281
Aube	143,537	Corrèze	72,778
Bouches du Rhône	154,756	Creuse	83,750
Cantal	153,127	Dordogne	88,242
Cher	159,439	Finistère	87,143
Dordogne	151,328	Gard	87,678
Eure et Loir	149,796	Haute Garonne	88,300
Gard	152,313	Gironde	84,484
Gers	144,099	Indre	88,589
Indre	145,329	Indre et Loire	89,897
Loir et Cher	146,646	Isère	84,633
Loiret	145,041	Jura	85,650
Lot	153,072	Landes	81,452
Maine et Loire	131,885	Loir et Cher	86,906
Marne	152,271	Loire	87,574
Haute Marne	144,570	Haute Loire	82,514
Meurthe et Moselle	158,517	Lot	74,175
Nièvre	142,142	Maine et Loire	83,432
Oise	150,606	Marne	89,961
Hautes Pyrénées	155,732	Haute Marne	87,166
Pyrénées-Orientales	149,557	Nièvre	88,629
Rhône	150,706	Puy de Dôme	90,343
Saône et Loire	156,777	Rhône	87,568
Paris	105,123	Saône et Loire	86,166
Seine et Marne	135,640	Sarthe	88,413
Vaucluse	151,010	Paris	84,275
Seine Saint Denis	145,898	Yvelines	80,515
Val d'Oise	131,714	Tarn	81,069
		Vendée	85,958
		Territoire de Belfort	82,590
		Essonne	78,098
		Seine Saint Denis	88,919
		Val de Marne	78,237
		Val d'Oise	81,212



CONCLUSION

La méthode de l'analyse discriminatoire est un moyen de classement d'éléments pour lesquels on dispose de données sur plusieurs variables. Cette étude en est une nouvelle application. Les résultats obtenus sont en concordance avec les données déjà existantes. Ils apportent cependant une information supplémentaire en montrant quelles sont les causes de décès dont le rôle est le plus important dans la détermination de la fonction discriminante : lésions vasculaires cérébrales, accidents, maladies du cœur.

RÉFÉRENCES

- [1] DAMIANI P. — « Recherche d'une méthode de classement des départements suivant la mortalité ». *Études et conjoncture*, avril 1968, I. N. S. E. E.
- [2] DAMIANI P. — « Méthodes d'analyse statistique ». 1976, I. N. S. E. E.
- [3] LABAT J.-C., VISEUR J. — « Données de démographie régionale, 1968 ». Collection I. N. S. E. E. D. 23, septembre 1973.
- [4] LEDERMANN S. — « La répartition des décès de cause indéterminée ». *Revue de l'Institut international de statistique*. 1956, I-III, 47-5.
- [5] « Statistique des causes médicales de décès ». Volumes annuels, INSERM.

Nous tenons à remercier M. Paul Damiani, administrateur de l'I. N. S. E. E., dont les conseils nous ont permis de mener à bien cette étude.