

PAUL DAMIANI

HÉLÈNE MASSÉ

Définition d'un indicateur du niveau de santé lié aux causes de décès

Journal de la société statistique de Paris, tome 119, n° 4 (1978), p. 357-366

http://www.numdam.org/item?id=JSFS_1978__119_4_357_0

© Société de statistique de Paris, 1978, 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/legal.php>). 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/>

DÉFINITION D'UN INDICATEUR DU NIVEAU DE SANTÉ LIÉ AUX CAUSES DE DÉCÈS

Paul DAMIANI (I. N. S. E. E.)

et Dr Hélène MASSÉ (I. N. S. E. R. M.)

On a dans cette étude, appliqué la méthode des composantes principales aux taux de mortalité des principales causes de décès constatés en France par sexe, pour le groupe d'âge 45-64 ans et pour la période 1968-1970. La première composante principale peut être prise comme indicateur du niveau de santé; elle est liée surtout à la cirrhose du foie, à l'alcoolisme et aux accidents. La deuxième composante principale est expliquée, en grande partie, par les maladies du cœur.

In this paper, the method of the principal components has been applied to the death rates of the leading causes of death in France, by sex, for 45-64 years age group and for the 1968-1970 period. The first principal component can be seen as an indicator of the health level. It is mostly in connection with cirrhosis of liver, alcoholism and accidents. The second principal component is explained, for a great part, by the heart diseases.

INTRODUCTION

De nombreuses tentatives ont été faites pour mesurer le niveau de santé d'une population. Les échelles de mesure que l'on peut établir diffèrent suivant les aspects par lesquels on aborde le concept de santé et suivant les méthodes utilisées [1].

Un indicateur simple et précis est *l'espérance de vie à la naissance* ou *durée moyenne de vie*, qui fait intervenir la mortalité aux différents âges de la vie.

Des indices plus complexes ont été élaborés faisant appel à des méthodes d'analyse statistique. Ils permettent de mesurer l'action réelle sur la mortalité des différents facteurs qui entrent dans leur composition. On peut citer, en France, les travaux de P. Damiani [2] et H. Massé [3]. On trouvera une bibliographie d'indicateurs de santé calculés dans le monde dans les publications de « Clearinghouse on Health Indexes » [4].

Dans cette étude, on a essayé de déterminer un indicateur du niveau de santé en appliquant la méthode des composantes principales aux taux de mortalité des principales causes de décès du groupe d'âge 45-64 ans, suivant le sexe.

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 — Suicide	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 (mention 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 [6]. 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 rectifiés par cause en divisant les nombres corrigés de décès par la population correspondante. On dispose donc, pour chaque sexe et par département, des taux rectifiés de mortalité du groupe d'âge 45-64 ans, pour chacune des 9 causes définies précédemment. Ces taux, calculés pour 100 000 habitants sont notés : x_k ($k = 1, 2, \dots, 9$).

MÉTHODE

Principe

La méthode utilisée est celle des composantes principales [7].

Cette méthode permet de remplacer les p variables x_k ($p = 9$) par p nouvelles variables y_j , fonctions linéaires des variables x_k et possédant les propriétés suivantes :

- les nouvelles variables y_j sont sans corrélation entre elles;
- la première des nouvelles variables calculées y_1 est la fonction linéaire des variables x_k possédant la variance maximum; d'une manière générale, la j^e variable y_j est

parmi les fonctions linéaires des variables x_k , sans corrélation avec les $(j - 1)$ nouvelles variables précédentes, y_1, y_2, \dots, y_{j-1} , celle qui possède la variance maximum.

Les nouvelles variables y_j sont appelées *composantes principales*.

La première composante principale y_1 peut être considérée comme la fonction linéaire des variables x_k représentant le mieux l'ensemble de ces variables.

Pratiquement, on calcule seulement les deux premières composantes principales.

Formules

1) On note X_k les valeurs des p variables x_k exprimées sous forme centrée réduite. On suppose que la matrice C des corrélations échantillon de ces variables possède p valeurs propres λ_k distinctes, notées par ordre décroissant :

$$\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$$

La j^e composante principale s'écrit :

$$y_j = l_{j1} X_1 + l_{j2} X_2 + \dots + l_{jp} X_p \quad (1)$$

avec :

$$\sum_k l_{jk}^2 = 1$$

On démontre les résultats suivants :

— la matrice l_j ($1 \times p$), composée des coefficients l_{jk} , représente le vecteur propre de la matrice C correspondant à la j^e valeur propre λ_j ,

— la variance de y_j est donnée par :

$$s^2(y_j) = \lambda_j$$

La somme des variances des composantes principales est égale à p :

$$\sum_j s^2(y_j) = \sum_j \lambda_j = p$$

et la part de la variance totale exprimée par y_j est égale à : λ_j/p .

2) Si on note Y_j la valeur de la composante principale y_j sous forme centrée réduite le modèle (1) s'écrit :

$$Y_j = c_{j1} X_1 + c_{j2} X_2 + \dots + c_{jp} X_p \quad (2)$$

où : $c_{jk} = r_{jk}$, coefficient de corrélation entre Y_j et X_k .

Les coefficients sont appelés *saturations* .

La part de la variance de X_k expliquée par Y_j est égale à c_{kj}^2 . En particulier, la part de la variance de X_k expliquée par les deux premières composantes principales est donnée par :

$$h_k^2 = c_{k1}^2 + c_{k2}^2$$

On l'appelle *communauté* relative aux deux premières composantes principales.

RÉSULTATS

Moyennes, écart-types, coefficients de corrélation des taux de mortalité

On a tout d'abord indiqué, dans le tableau I, la moyenne et l'écart-type des taux de mortalité x_k par cause, suivant le sexe, et dans le tableau II les coefficients de corrélation entre ces taux pris deux à deux.

TABLEAU I

Moyenne et écart-type des taux de mortalité par cause, suivant le sexe
Taux rectifiés pour 100 000, groupe d'âge 45-64 ans, moyenne 1968-1970

Cause de décès	Sexe masculin		Sexe féminin	
	Moyenne	Écart type	Moyenne	Écart type
x_1 - Tuberculose toutes formes	24,245	9,322	7,687	3,156
x_2 - Cancers	387,596	53,510	221,277	24,922
x_3 - Diabète	15,517	4,428	13,982	5,239
x_4 - Maladies du cœur	246,300	43,666	88,972	17,343
x_5 - Lésions vasculaires cérébrales	117,981	25,441	66,389	11,526
x_6 - Alcoolisme	38,860	17,837	7,573	4,792
x_7 - Cirrhose du foie	124,952	35,479	44,521	20,089
x_8 - Accidents	131,408	27,807	34,141	9,333
x_9 - Suicide	52,374	16,933	15,872	6,615

TABLEAU II

Coefficients de corrélation entre les taux de mortalité par cause, suivant le sexe

Cause de décès	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
<i>Sexe masculin</i>									
x_1 - Tuberculose toutes formes	1								
x_2 - Cancers	0,414	1							
x_3 - Diabète	0,115	-0,040	1						
x_4 - Maladies du cœur	0,495	0,452	0,382	1					
x_5 - Lésions vasculaires cérébrales	0,462	0,309	0,126	0,298	1				
x_6 - Alcoolisme	0,526	0,369	-0,076	0,254	0,547	1			
x_7 - Cirrhose du foie	0,540	0,685	0,101	0,463	0,501	0,656	1		
x_8 - Accidents	0,340	0,420	-0,058	0,185	0,466	0,651	0,549	1	
x_9 - Suicide	0,331	0,287	-0,089	-0,020	0,494	0,658	0,430	0,569	1
<i>Sexe féminin</i>									
x_1 - Tuberculose toutes formes	1								
x_2 - Cancers	0,029	1							
x_3 - Diabète	0,007	0,019	1						
x_4 - Maladies du cœur	-0,032	0,063	0,515	1					
x_5 - Lésions vasculaires cérébrales	0,106	-0,021	0,139	0,521	1				
x_6 - Alcoolisme	0,292	0,134	-0,016	0,188	0,284	1			
x_7 - Cirrhose du foie	0,325	0,253	0,120	0,153	0,189	0,697	1		
x_8 - Accidents	0,310	0,095	-0,091	-0,081	0,011	0,387	0,426	1	
x_9 - Suicide	0,038	0,061	-0,054	0,015	0,047	0,341	0,367	0,366	1

Valeurs propres, variance expliquée

Les valeurs propres correspondant aux cinq premières composantes principales et la part de la variance totale expliquée par chacune de ces composantes sont indiquées dans le tableau III, par sexe.

TABLEAU III

Valeur propre et proportion expliquée de la variance pour les cinq premières composantes principales

Composante principale	Sexe masculin			Sexe féminin		
	Valeur propre	Variance expliquée		Valeur propre	Variance expliquée	
		Proportion	Proportion cumulée		Proportion	Proportion cumulée
v_1	4,155	0,462	0,462	2,642	0,294	0,294
v_2	1,515	0,168	0,630	1,801	0,200	0,494
v_3	0,912	0,101	0,731	1,026	0,114	0,608
v_4	0,615	0,069	0,800	0,959	0,106	0,714
v_5	0,513	0,057	0,857	0,845	0,094	0,808

On remarque que la 1^{re} composante principale explique 46,2 pour cent de la variance totale, pour le sexe masculin, et seulement 29,4 pour cent pour le sexe féminin.

Les deux premières composantes expliquent, à elles deux 63,0 pour cent de la variance, pour le sexe masculin, et 49,4 pour cent pour le sexe féminin.

Saturation et communauté des deux premières composantes principales

Le tableau IV fournit, par sexe, pour chacune des deux premières composantes principales les coefficients c_{jk} du modèle (2) permettant de calculer Y_1 et Y_2 . Pour chacune des variables X_k , figure la valeur de la communauté relative aux deux premières composantes principales.

TABLEAU IV

Saturation et communauté des deux premières composantes principales

Cause de décès	Sexe masculin			Sexe féminin		
	Saturation		Communauté h_k^2	Saturation		Communauté h_k^2
	ck_1	ck_2		ck_1	ck_2	
x_1 - Tuberculose toutes formes	0,711	0,259	0,573	0,469	-0,177	0,251
x_2 - Cancers	0,677	0,174	0,488	0,270	-0,043	0,075
x_3 - Diabète	0,075	0,684	0,473	0,142	0,692	0,499
x_4 - Maladies du cœur	0,514	0,707	0,765	0,321	0,837	0,803
x_5 - Lésions vasculaires cérébrales	0,728	-0,018	0,581	0,400	0,583	0,500
x_6 - Alcoolisme	0,830	-0,279	0,768	0,826	-0,059	0,686
x_7 - Cirrhose du foie	0,843	0,119	0,725	0,850	-0,075	0,728
x_8 - Accidents	0,739	-0,317	0,647	0,616	-0,416	0,563
x_9 - Suicide	0,666	-0,507	0,701	0,530	-0,258	0,347

On constate que, pour la première composante principale, les causes de décès qui interviennent le plus dans sa variance sont, pour les deux sexes, la cirrhose du foie, l'alcoolisme, puis les accidents.

Pour la deuxième composante principale, les maladies du cœur et le diabète sont les causes prépondérantes, pour les deux sexes.

Enfin, pour les deux sexes, ce sont l'alcoolisme, la cirrhose du foie et les maladies du cœur dont la variance est la mieux expliquée par les deux premières composantes principales.

Valeurs des deux premières composantes principales

Le tableau V donne, par département, les valeurs des deux premières composantes principales, par sexe.

Des diagrammes représentent, pour chaque sexe, la dispersion des départements suivant les deux premières composantes principales. Ces graphiques font apparaître les départements ayant des valeurs extrêmes. Pour le sexe masculin, par exemple, les départements ayant une première composante principale élevée sont : les Côtes-du-Nord (26), le Morbihan (56), le Finistère (29), la Loire-Atlantique (44), l'Ille-et-Vilaine (35); les départements ayant une 2^e composante principale élevée sont : le Bas-Rhin (67), le Haut-Rhin (68), la Moselle (57), le Nord (59), le Pas-de-Calais (62).

Définition d'un indicateur du niveau de santé

On choisit, comme indicateur du niveau de santé lié aux causes de décès, la première composante principale. On observe que le niveau de santé varie en sens inverse des valeurs de cet indice.

Cet indicateur est en forte corrélation avec l'espérance de vie à la naissance ($r = -0,763$ pour le sexe masculin et $r = -0,673$ pour le sexe féminin).

Les cartes 1 et 2 illustrent, pour chaque sexe, les valeurs de cet indicateur par département.

TABLEAU V

Valeur des deux premières composantes principales, par département, suivant le sexe

Département	Sexe masculin		Sexe féminin	
	Y_1	Y_2	Y_1	Y_2
1 - Ain	-0,639	-0,660	-0,023	-1,343
2 - Aisne	1,194	-0,363	3,188	0,880
3 - Allier	0,306	1,765	-0,927	-1,188
4 - Alpes-de-Haute-Provence	-1,820	-0,056	-0,956	0,577
5 - Alpes (Hautes-)	-1,652	0,140	-0,924	-1,993
6 - Alpes-Maritimes	-0,609	-0,774	0,095	-1,540
7 - Ardèche	-0,264	2,226	-1,229	1,439
8 - Ardennes	1,910	-0,421	2,852	1,929
9 - Ariège	-2,871	0,206	-0,886	1,003
10 - Aube	0,324	-0,637	-0,451	-1,656
11 - Aude	-3,501	1,204	-3,173	0,053
12 - Aveyron	-2,845	-0,030	-2,075	2,879
13 - Bouches-du-Rhône	-1,429	1,081	-1,081	0,237
14 - Calvados	2,438	-0,959	1,052	-1,343
15 - Cantal	0,557	-0,781	-2,310	2,809
16 - Charente	-1,047	-0,389	-1,102	-0,310
17 - Charente-Maritime	-2,235	-1,140	-0,664	-1,601
18 - Cher	-0,281	-0,420	-0,633	-1,085
19 - Corrèze	-1,104	-0,175	-2,661	2,015
20 - Corse	-1,119	-0,766	-2,437	-0,011
21 - Côte-d'Or	-1,863	-0,760	-0,463	-1,908
22 - Côte-du-Nord	6,827	-0,657	4,124	-0,337
23 - Creuse	2,027	1,513	-1,361	0,778
24 - Dordogne	-0,881	-0,318	-0,060	0,579
25 - Doubs	-0,092	0,269	-0,254	1,539
26 - Drôme	-1,462	0,164	-1,167	-0,729
27 - Eure	1,184	-0,904	1,471	-1,270
28 - Eure-et-Loir	-0,287	-1,278	1,053	-1,435
29 - Finistère	5,709	-1,595	1,139	-0,700
30 - Gard	-1,006	1,327	-1,048	1,095

TABLEAU V (suite)

Valeur des deux premières composantes principales,
par département, suivant le sexe

Département	Sexe masculin		Sexe féminin	
	Y ₁	Y ₂	Y ₁	Y ₂
31 - Garonne (Haute-)	-2,481	0,781	-0,780	-0,224
32 - Gers	-1,629	-0,687	-1,014	-1,254
33 - Gironde	-0,798	-0,200	-0,349	-1,317
34 - Hérault	-3,589	0,168	-2,043	-0,136
35 - Ille-et-Vilaine	3,795	-0,427	2,225	-0,384
36 - Indre	-0,103	-1,286	-0,807	-0,221
37 - Indre-et-Loire	-1,432	-2,298	0,077	-1,647
38 - Isère	0,047	0,303	-0,515	-0,061
39 - Jura	-0,827	-0,758	-0,248	1,030
40 - Landes	0,833	1,010	-0,910	0,608
41 - Loir-et-Cher	-0,657	-1,106	-0,256	-2,168
42 - Loire	0,971	1,354	-0,503	1,065
43 - Loire (Haute-)	0,522	0,585	-1,007	0,475
44 - Loire-Atlantique	3,877	-2,708	3,016	-1,147
45 - Loiret	-0,860	-0,785	-0,064	-0,686
46 - Lot	-1,403	1,256	-1,528	1,409
47 - Lot-et-Garonne	-2,361	-0,428	-1,237	-0,096
48 - Lozère	1,097	0,891	-2,423	1,082
49 - Maine-et-Loire	-0,577	-1,994	0,124	-1,057
50 - Manche	2,247	-0,296	1,110	-0,525
51 - Marne	0,677	-0,675	0,859	-0,209
52 - Marne (Haute-)	0,198	-0,114	-0,022	1,369
53 - Mayenne	2,237	-0,638	2,454	0,766
54 - Meurthe-et-Moselle	0,409	0,261	0,912	0,513
55 - Meuse	1,409	0,189	2,490	0,991
56 - Morbihan	6,634	-1,645	6,355	0,918
57 - Moselle	1,730	3,161	0,622	3,662
58 - Nièvre	-0,371	-0,561	-0,350	0,119
59 - Nord	2,151	2,947	3,615	2,741
60 - Oise	1,005	-2,228	0,311	-0,955
61 - Orne	2,341	-2,582	3,654	-0,925
62 - Pas-de-Calais	3,121	2,551	1,433	2,992
63 - Puy-de-Dôme	1,074	0,664	-0,478	1,316
64 - Pyrénées-Atlantiques	-2,376	-0,081	-1,050	0,145
65 - Pyrénées (Hautes-)	-1,061	1,349	-0,941	-0,994
66 - Pyrénées-Orientales	-1,934	0,087	-1,999	0,903
67 - Rhin (Bas-)	2,772	3,710	0,922	2,314
68 - Rhin (Haut-)	3,254	3,697	2,412	2,253
69 - Rhône	-0,712	0,136	-0,444	-0,740
70 - Saône (Haute-)	1,443	-0,337	0,553	0,058
71 - Saône-et-Loire	-0,246	-0,525	-0,718	-0,114
72 - Sarthe	-0,162	-0,998	0,242	-2,037
73 - Savoie	1,017	0,130	0,018	0,004
74 - Savoie (Haute-)	1,432	1,038	0,415	0,949
75 - Ville de Paris	-0,995	1,327	-0,057	-1,683
76 - Seine-Maritime	1,668	-0,961	1,115	-0,276
77 - Seine-et-Marne	0,330	-0,912	1,607	-1,226
78 - Yvelines	-1,883	-1,373	-0,460	-2,186
79 - Sèvres (Deux-)	-2,323	-0,330	-1,968	-0,240
80 - Somme	1,371	0,204	2,622	0,233
81 - Tarn	-2,728	0,332	-1,599	1,272
82 - Tarn-et-Garonne	-2,376	0,618	-0,153	1,279
83 - Var	-0,726	0,277	-1,312	-0,195
84 - Vaucluse	-1,257	0,987	-1,357	1,389
85 - Vendée	-0,357	-0,705	-0,358	-0,302
86 - Vienne	-1,653	-0,719	0,273	-1,481
87 - Vienne (Haute-)	-1,384	-0,090	-1,442	0,540
88 - Vosges	1,926	-0,053	0,136	1,318
89 - Yonne	0,770	-0,243	0,953	-1,704
90 - Belfort (Territoire de)	-0,096	1,524	-0,424	0,906
91 - Essonne	-1,999	-1,316	0,135	-2,132
92 - Hauts-de-Seine	-1,542	0,396	-1,404	-1,531
93 - Seine-Saint-Denis	-0,503	0,503	1,023	-0,338
94 - Val-de-Marne	-2,207	0,443	-0,347	-1,617
95 - Val-d'Oise	-1,329	0,015	0,742	-1,620

CONCLUSION

La méthode utilisée a permis de définir un indicateur du niveau de santé à partir des taux de mortalité par cause. Cet indicateur est en forte corrélation avec l'espérance de vie à la naissance. Son calcul fait apparaître de plus, la part réelle des différentes causes dans

la mortalité. Pour la première composante principale, choisie comme indicateur, il s'agit de la cirrhose du foie, de l'alcoolisme et des accidents.

Les valeurs de cet indice sont complétées par celles de la deuxième composante principale qui est liée surtout aux maladies du cœur.

RÉFÉRENCES

- [1] DAMIANI P., La mesure du niveau de santé. *Journal de la Société de statistique de Paris*, t. 114, n° 2, 1973, pp. 129-144.
- [2] DAMIANI P., Recherche d'indicateur du niveau de santé : définition d'un indicateur sanitaire de mortalité. *Bulletin Statistiques* du ministère de la Santé publique et de la population, n° 2, 1964, pp. 89-95.
DAMIANI P., Recherche d'une méthode de classement des départements suivant la mortalité. *Études et conjoncture*, n° 4, 1968, pp. 131-139.
DAMIANI P., Contribution à la mesure du niveau de santé : définition d'un indicateur lié aux facteurs sanitaires. *International statistical review*, vol. 42, n° 1, 1974, pp. 29-37.
- [3] MASSÉ H., Méthode de classement des départements à partir de la mortalité par cause : application de l'analyse discriminatoire. *Journal de la Société de statistique de Paris*, t. 118, n° 4, 1977, pp. 311-320.
- [4] Clearinghouse on Health Indexes. Division of Analysis. National Center for Health Statistics. 3 700 East West Highway. Hyattsville, Maryland 20782 U. S. A.
- [5] Statistiques des causes médicales de décès. Volumes annuels. I. N. S. E. R. M.
- [6] LEDERMANN S., La répartition des décès de cause indéterminée. *Revue de l'Institut international de statistique*, 1956, I-III, pp. 47-5.
- [7] DAMIANI P., Méthodes d'analyse statistique, 1976. I. N. S. E. E.

Diagramme de dispersion des départements suivant les 2 premières composantes principales

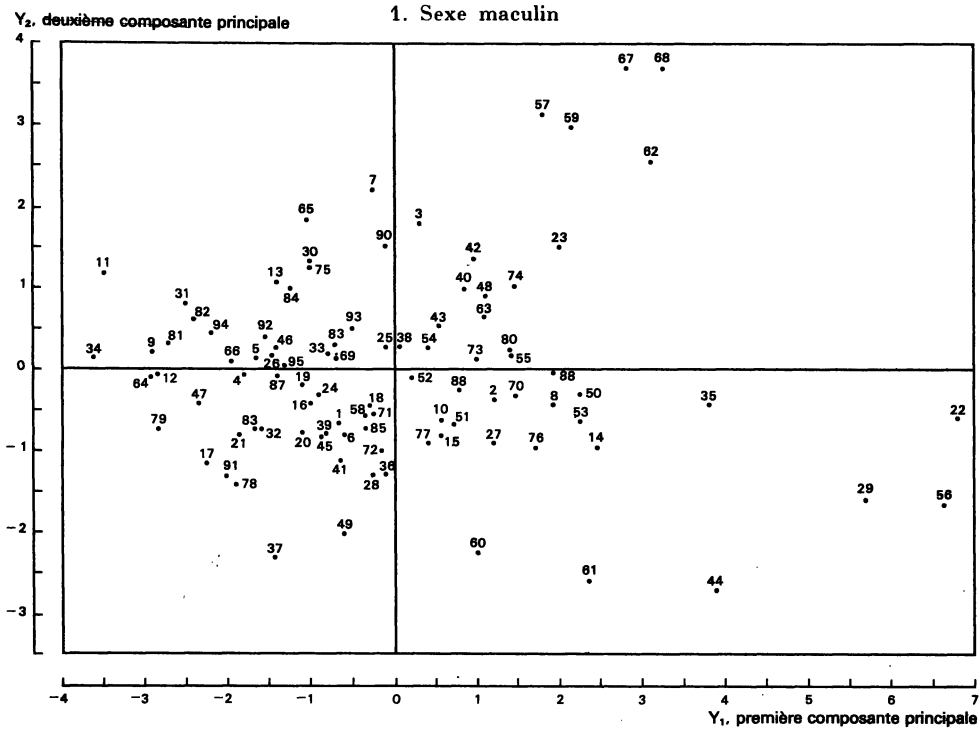
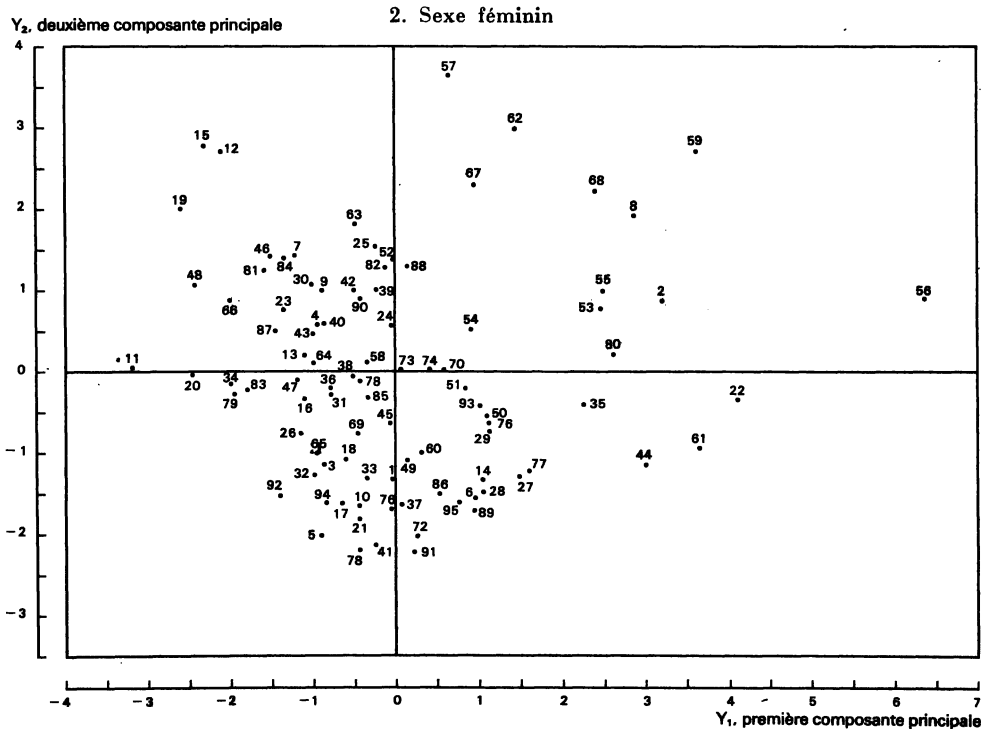
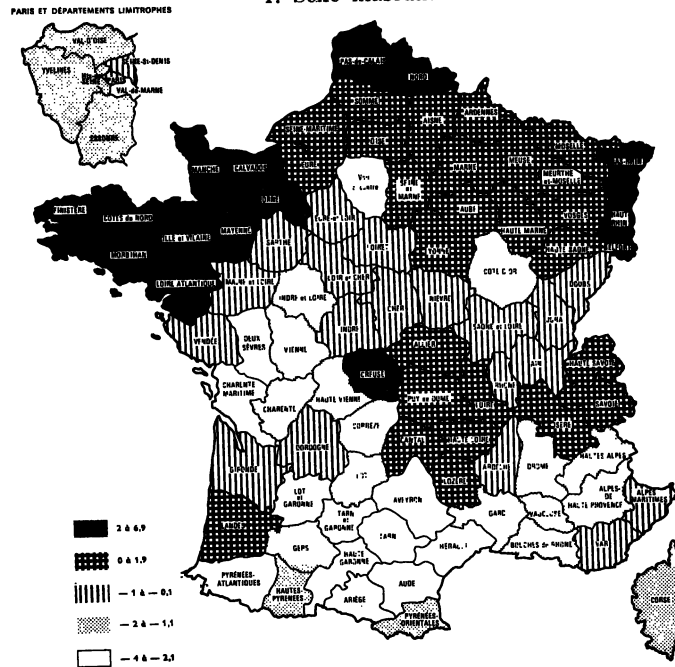


Diagramme de dispersion des départements suivant les 2 premières composantes principales



*Indicateur de santé lié aux causes de décès
par département*

1. Sexe masculin



*Indicateur de santé lié aux causes de décès
par département*

2. Sexe féminin

