

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

SULLY LEDERMANN

Structure factorielle des tables de mortalité

Journal de la société statistique de Paris, tome 100 (1959), p. 245-258

http://www.numdam.org/item?id=JSFS_1959__100__245_0

© Société de statistique de Paris, 1959, 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/>

III

STRUCTURE FACTORIELLE DES TABLES DE MORTALITÉ (1)

Les indices uniques

L'intérêt évident de pouvoir résumer en un seul chiffre la situation de la mortalité d'une population a conduit à utiliser à cet effet : le taux brut de mortalité, un taux construit sur une population-type, la mortalité infantile, l'espérance de vie à la naissance et, plus récemment, la proportion des décès dus à des maladies infectieuses ou parasitaires dans l'ensemble des décès toutes causes, la proportion des décès survenus avant l'âge de 5 ans, ou mieux encore de 50 ans, dans l'ensemble des décès tous âges (SWAROOP et UEMURA, O. M. S., 1957), etc..

Deux questions fondamentales

Le nombre des indices essayés montre qu'aucun d'eux n'a satisfait entièrement les besoins exprimés par les divers utilisateurs.

Cette situation conduit donc à se poser deux questions :

- a) Dans quelle mesure un indice unique peut-il résumer la mortalité d'une population?
- b) Quel est le nombre minimal d'indices nécessaires pour décrire la mortalité ou, plus précisément, une table de mortalité?

(1) Pour plus de détails, le lecteur pourra se reporter à l'article de la revue *Population*, 1959, n° 4, de l'Institut National d'Études Démographiques : « Les dimensions de la mortalité ».

L'analyse factorielle

Une analyse factorielle, au sens des psycho-techniciens, de 157 tables de mortalité couvrant sensiblement la période 1900-1950 apporte ici quelques éléments de réponse (1).

Par rapport au domaine psycho-technique, le rôle des *tests* est tenu par les groupes d'âges, 18 par sexe : moins d'un an, 1 à 4 ans, 5 à 9, 10 à 14, etc., jusqu'à 80 à 84 ans, auxquels a été ajoutée l'espérance de vie à la naissance e_0 , soit donc 38 variables au total; le rôle des *sujets* est tenu par les 157 tables de mortalité elles-mêmes.

La *note* ou le « score », par exemple, du *sujet* « Inde : 1891-1901 » dans le *test* « Hommes 55 à 59 ans » est le quotient 234 pour 1 000, correspondant aux 234 décès observés en 5 ans pour 1 000 hommes ayant atteint 55 ans.

Chaque table peut ainsi être représentée par un point à 38 coordonnées. Les 157 points obtenus dessinent un ellipsoïde dans cet espace à 38 dimensions, dont la forme est suggérée, par exemple, par sa projection sur le plan des $X_9 = {}_5q_{30}$ et $X_{11} = {}_5q_{40}$ du *graphique 1*.

C'est un ellipsoïde très allongé. Les inter-corrélations entre les 38 variables sont en effet élevées. Ces inter-corrélations ont été calculées d'après les logarithmes de quotients ou de $100 - e_0$, quantités qui se trouvent distribuées à peu près normalement. Les caractéristiques de ces variables logarithmiques figurent dans le *tableau 1*, et les corrélations observées dans les *tableaux 2, 3 et 4*.

Soit Z_{ji} la variable réduite du logarithme X_{ji} avec $j = 1, 2, \dots$ jusqu'à $n = 38$ et $i = 1, 2, \dots$ jusqu'à $N = 157$:

$$(1) \quad \boxed{Z_{ji} = \frac{X_{ji} - M_j}{\sigma_j}}$$

X_{ji} = logarithme du quotient de mortalité du groupe d'âges j pour la table n° i , ou bien de la quantité $100 - e_0$.

M_j = valeur moyenne de X_{ji} pour les 157 tables (*tableau 1*)

σ_j = écart-type de X_{ji} (*tableau 1*).

Les Z_{ji} constituent une matrice ($n \times N$). La matrice de corrélation \mathbf{R} ($n \times n$) est donnée par :

$$(2) \quad \boxed{\mathbf{R} = \frac{1}{N} \mathbf{Z} \mathbf{Z}'}$$

Les composantes principales

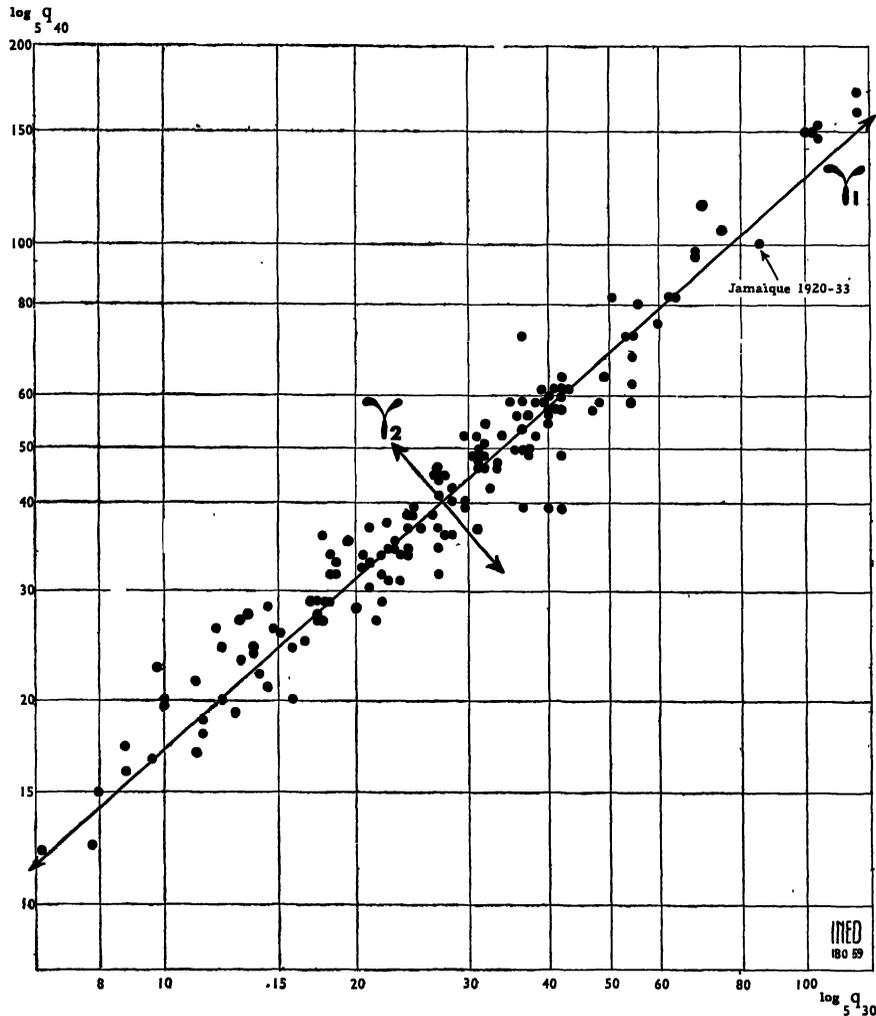
L'ellipsoïde a, en principe, 38 axes principaux, mais les corrélations entre les variables étant très élevées, on s'aperçoit qu'il peut être considéré comme plat sur, au moins, 33 dimensions.

Il n'y a pas de tests disponibles pour tester la significativité des axes négligés. Ceux qui existent sont en effet relatifs à un échantillon extrait d'une population multi-variantes normale, les observations étant tirées au hasard individuellement et indépendamment les

(1) De premiers résultats obtenus à partir de la mortalité de quelques groupes d'âges par la technique dite « centroïde » ont été indiqués dans une communication lors de la 31^e Session de l'Institut International de Statistique, Bruxelles 1958. Les différences entre ces résultats et les résultats actuels proviennent du fait que : 1^o l'ensemble des groupes d'âges a été aujourd'hui pris en considération, notamment un groupe extrême tel que 80-84 ans, et 2^o la méthode utilisée est ici celle dite des « axes principaux ». Mais la première raison est de loin ici la plus importante.

unes des autres. Or ici les 157 tables sont fournies par une cinquantaine de pays, qui chacun fournit ainsi une grappe de tables. Le mécanisme d'extraction est plutôt celui d'un tirage en grappes, et les hypothèses à la base des tests existants ne sont pas satisfaites.

L'ellipsoïde suffisant à décrire le phénomène avec un nombre de dimensions que nous prendrons égal à 3 provisoirement, on peut rapporter chaque point-table aux trois axes principaux retenus.



Graphique 1. — Corrélation entre logarithmes des ${}_5q_{30}$ et les ${}_5q_{40}$ pour les 157 tables de mortalité (sexe masculin). La table « Jamaïque 1920-33 » par exemple peut être repérée, soit par ses deux quotients, soit par les projections du point sur les deux axes $\gamma_1 \gamma_2$. Dans le cas de la figure, la projection sur l'axe γ_1 constitue la meilleure description « unique » de la situation de la « Jamaïque 1920-33 ».

Soit Y_{1i}, Y_{2i}, Y_{3i} les projections du point n° i sur les trois premiers axes principaux de l'ellipsoïde, mesurées dans une échelle normale, de moyenne nulle et de variance unité;

a_{11}, a_{12}, a_{13} les coefficients de corrélation de Z ou X avec la composante principale 1, 2 ou 3.

Chacune des 38 variables Z_{jt} peut alors se mettre sous la forme (les valeurs estimées se distinguant des valeurs observées par un accent circonflexe) :

$$(3) \quad \hat{Z}_{jt} = a_{j1} \Upsilon_{1t} + a_{j2} \Upsilon_{2t} + a_{j3} \Upsilon_{3t}$$

soit encore

$$(4) \quad \hat{\mathbf{Z}} = \mathbf{a} \Upsilon$$

avec \mathbf{a} matrice ($n \times 3$) des a_j et Υ matrice ($3 \times N$) des Υ

La valeur numérique des coefficients a figure dans le *tableau 5*. Ils fournissent l'estimation suivante des coefficients de corrélation entre deux variables j et k , avec :

$$(5) \quad \hat{r}_{jk} = a_{j1} a_{k1} + a_{j2} a_{k2} + a_{j3} a_{k3}$$

soit encore :

$$(6) \quad \hat{\mathbf{R}} = \mathbf{a} \mathbf{a}'$$

Ce que les tests que nous ne pouvons appliquer cherchent en général à tester, c'est la différence $\mathbf{R} - \hat{\mathbf{R}}$.

L'indice unique optimal

L'observation fondamentale ici est que c'est la projection d'un point sur le grand axe de cet ellipsoïde c'est-à-dire de la composante Υ_1 qui fournit la meilleure description *unique* de la table correspondante, ou encore le meilleur classement des tables si l'on veut s'en tenir à un seul indice, et ce d'autant plus, naturellement, que l'ellipsoïde est allongé. Cette projection est appelée, en général, *première composante principale*.

La première composante principale Υ_1 transformée pour commodité numérique en indice E (*tableau 6*) par :

$$(7) \quad E = 100 + 25 \Upsilon_1$$

explique 82 % de la variance totale (1) des 38 variables Z , mais bien davantage de la variance de certaines variables particulières, telle que l'espérance de vie à la naissance : $0,974^2 = 95$ %.

Il faut en conclure que l'information globale donnée par l'espérance de vie à la naissance est très proche de l'information maximale unique possible.

La mortalité infantile (variables 0-1 an, n° 2 et 21) ne se présente pas du tout dans une position aussi favorable, puisque l'indice E n'explique que $0,88^2 = 77$ % de sa variance, et réciproquement.

Il est à noter que la mortalité des femmes âgées de 35 à 44 ans fournit une information globale aussi poussée que l'espérance de vie à la naissance (*tableau 5*).

La contribution des deux autres composantes principales à la variance totale des 38 Z tombe de 81,5 % pour la première composante à, respectivement, 7,4 % et 3,7 % (*tableau 5*).

(1) La contribution de la composante E à la variance totale du Z est mesurée par $\frac{1}{38} \sum_j a_{j1}^2$.

Nous n'insisterons pas sur ces deux dernières composantes. Elles vont seulement servir à chercher une structure à trois facteurs, laquelle doit pouvoir, au total, expliciter $81,5 + 7,4 + 3,7 = 92,6 \%$ de la variance totale des 38 variables.

Changement d'axes de référence

Si, en effet, dans le sous-espace des $p = 3$ composantes principales, on effectue un changement d'axes de référence, la matrice orthogonale de la rotation étant \mathbf{T} ($p \times p$), on substitue au système \mathbf{Y} un système \mathbf{F} , avec :

$$(8) \quad \boxed{\hat{\mathbf{Z}} = \mathbf{a} \mathbf{Y} = \mathbf{A} \mathbf{F}} \quad \text{et} \quad (9) \quad \boxed{\mathbf{A} = \mathbf{a} \mathbf{T}^{-1}}$$

et, de plus :

$$(10) \quad \boxed{\hat{\mathbf{R}} = \mathbf{a} \mathbf{a}' = \mathbf{A} \mathbf{T} \mathbf{T}' \mathbf{A}' = \mathbf{A} \mathbf{A}'}$$

En d'autres termes, dans le sous-espace considéré à p dimensions, une infinité de systèmes orthogonaux de p composantes reproduisent les corrélations de la matrice \mathbf{R} .

Le système \mathbf{Y} est seulement l'un de ces systèmes, mais privilégié. Chacune des composantes principales \mathbf{a} , en effet, extrait le maximum de variance de l'univers qu'il lui revenait d'explorer :

\mathbf{Y}_1 de l'ensemble; \mathbf{Y}_2 de l'univers résiduel non explicité par \mathbf{Y}_1 ; \mathbf{Y}_3 de celui non explicité par \mathbf{Y}_1 et \mathbf{Y}_2 , etc.

C'est pour cette raison que la première composante principale \mathbf{Y}_1 ou l'indice \mathbf{E} est l'indice moyen le plus adéquat pour résumer une table de mortalité, c'est-à-dire qu'elle fournit la meilleure description moyenne d'un ensemble hétérogène.

Ceci étant, il reste maintenant à voir si cet ensemble hétérogène ne se laisserait pas analyser, par une rotation \mathbf{T} convenable, en sous-ensembles relativement homogènes auxquels pourraient être accolées des étiquettes et affectés des indices.

Un système plus économique

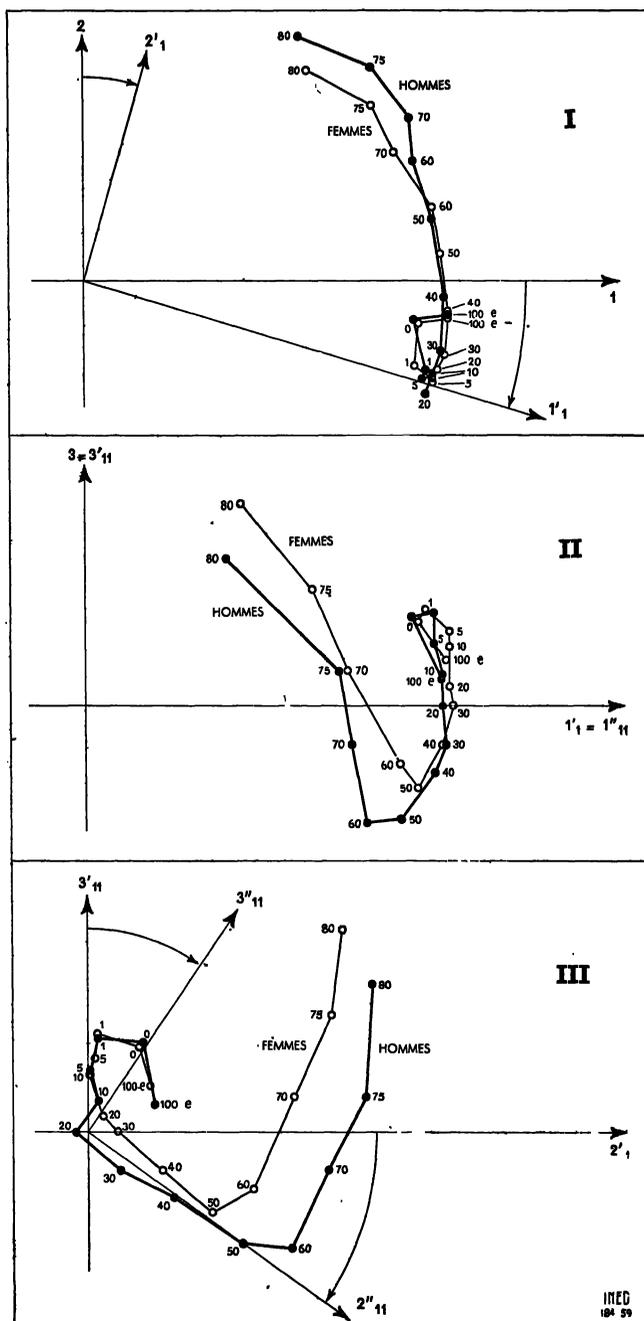
Le principe de cette recherche est celui d'une économie supplémentaire dans la description. Une première économie a été obtenue, puisqu'il est devenu possible de décrire chacun des n variables de départ en termes de composantes en nombre p inférieur à n . Si maintenant une rotation \mathbf{T} fait apparaître p nouvelles composantes qui, elles, n'interviennent pas toutes dans les n variables, une économie supplémentaire aura été réalisée, puisque certaines de variables n'exigent même plus p composantes pour être décrites.

Il est exclu qu'un tel résultat éventuel puisse être obtenu à partir d'une matrice arbitraire de rang n . Il méritera donc au moins un commentaire, car il constituera probablement un pas accompli dans la détection d'un certain ordre sous-jacent simple.

Nous nous bornons ici à la recherche d'une structure orthogonale, c'est-à-dire de composantes ou de facteurs indépendants, laissant de côté provisoirement la recherche d'une structure oblique relative à des sous-ensembles peut-être plus séduisants au point de vue descriptif, mais s'interpénétrant, c'est-à-dire non indépendants.

Sur le plan pratique, la rotation \mathbf{T} est décomposée en rotations partielles dans chaque plan de référence. Pour trois axes, 3 rotations partielles sont à envisager avec les combinaisons d'axes 1.2, 1.3 et 2.3.

Le *graphique 2* présente les rotations adoptées. Une analyse critique de ces rotations montre que la marge d'arbitraire est extrêmement faible. Elle ne dépasse pas quelques degrés dans les rotations. Trois sous-ensembles apparaissent, correspondant à des « facteurs »



Graphique 2.

Rotations conduisant à la structure factorielle F, F, F.

$F_1 F_2 F_3$, dnot les coefficients de corrélations A avec les 38 variables Z figurent dans le tableau 5, la matrice de transformation linéaire de passage de Υ à F étant la matrice T

$$T = \begin{pmatrix} 0,9658 & 0,2125 & 0,1488 \\ -0,2594 & 0,7912 & 0,5539 \\ 0 & -0,5735 & 0,8192 \end{pmatrix}$$

Dans le nouveau système, l'estimation des Z est :

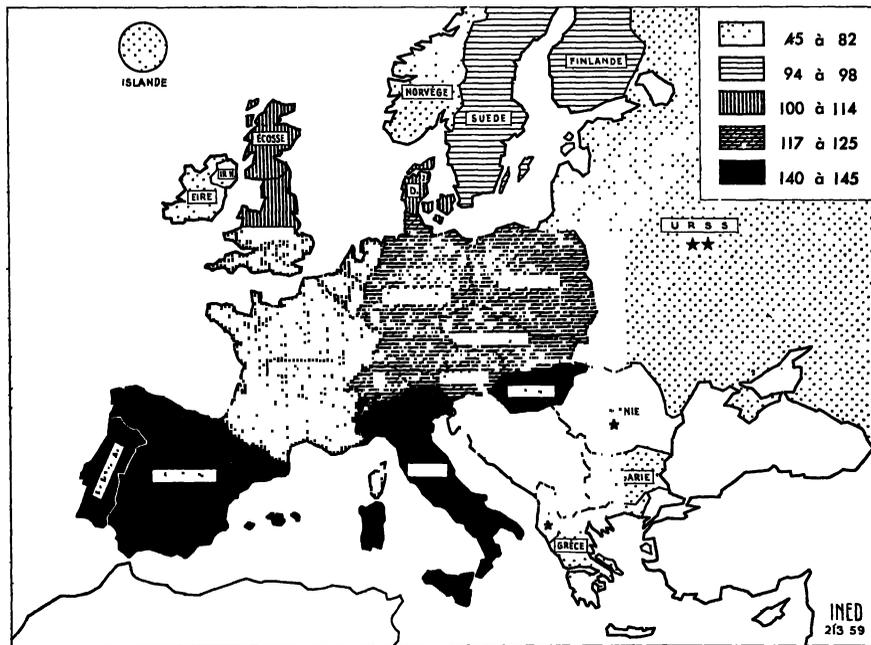
$$(11) \quad \hat{Z}_{ji} = A_{j1} F_{1i} + A_{j2} F_{2i} + A_{j3} F_{3i}$$

La toile de fond : l'indice φ_1

Le facteur F_1 , de moyenne nulle et de variance unité, a été transformé pour commodité en indice φ_1 par :

$$(12) \quad \varphi_1 = 100 + 25 F_1$$

Il explicite 77 %, c'est-à-dire les 3/4 de la variance totale des 38 variables, mais plus particulièrement la mortalité de la population âgée de 5 à 34 ans. Il diminue avec le temps (tableau 6) et semble relever de conditions de fond.



Graphique 3. — Valeur moyenne de l'indice φ_1 pour l'Europe, suivant les pays.

★ pas d'estimation.

★★ estimation d'après les taux de mortalité de la période 1953-56.

Le calcul de la valeur numérique de φ_1 pour une table donnée peut s'effectuer à partir d'une lourde régression de F_1 ou de φ_1 sur les 38 variables, ou bien on peut l'estimer en partant de quelques variables seulement, ce qui présente un intérêt pour le pays où l'on ne possède que des données fragmentaires sur la mortalité.

La situation des adultes : l'indice φ_2

Le facteur F_2 de moyenne nulle et de variance unité, transformé également en indice φ_2 pour commodité par :

$$(13) \quad \boxed{\varphi_2 = 100 + 25 F_2}$$

explique 10 % seulement de la variance totale des 38 variables Z, mais au-dessus de 40 ans, 26 % de la variance des quotients des hommes et 13 % de celle des femmes.

C'est un indice qui caractérise une mortalité additionnelle des adultes, indépendante de conditions de fond. Le fait qu'en France, ses variations coïncident avec celles de la consommation globale de vin et de la surmortalité masculine n'est peut-être pas dû au hasard : il s'agirait bien, en tout cas, d'une mortalité additionnelle-type.

Les extrémités de la vie : l'indice φ_3

Le facteur F_3 , également de moyenne nulle et de variance unité, transformé également en indice φ_3 pour commodité par :

$$(14) \quad \boxed{\varphi_3 = 100 + 25 F_3}$$

explique seulement 6,5 % de la variance totale des 38 variables Z, mais 61 % de la dispersion de la mortalité pour les groupes 80-84 ans.

Il s'agit d'une composante de mortalité additionnelle, indépendante des deux premières et qui agit surtout sur le groupe 80-84 ans (61 %), moins sur les groupes 70 à 79 (16 % pour les hommes et 26 % pour les femmes) et à l'autre extrémité, mais beaucoup moins, sur la mortalité des très jeunes enfants (6 %).

φ_3 a tendance à diminuer légèrement avec le temps. Ses faibles mouvements peuvent s'expliquer par le fait qu'elle caractérise principalement les grands âges, où la lutte contre la mort n'a enregistré que peu de progrès. De ce fait, on peut calculer sa valeur moyenne par pays (graphique 3).

En Europe, des groupes géographiques homogènes apparaissent curieusement :

- l'Europe Occidentale avec la France (106), la Belgique (110), les Pays-Bas (114), la Grande-Bretagne (107);
- l'Europe Centrale commençant à la Suisse (117), et se précisant avec l'Autriche (121), la Tchécoslovaquie (125), l'Allemagne (125);
- un ensemble de trois pays méditerranéens : l'Italie (141), l'Espagne (145) et le Portugal (142);
- l'Europe du Nord se divise en Norvège (71) et Islande (64), d'un côté, Suède (94) et Finlande (98), de l'autre. Le Danemark (109) qui n'en est pas très éloigné, entrerait peut-être plus volontiers dans un ensemble « Mer du Nord » comprenant la Grande-Bretagne (107), la Belgique (110) et les Pays-Bas (114).

L'interprétation de cette dernière composante n'est pas aisée. Elle est marginale, relative aux âges extrêmes, et peut mobiliser à la fois des conditions locales, des erreurs d'enregistrement des décès, et des variations de technique actuarielle dans l'élaboration des tables. Il n'est pas possible d'avancer une interprétation causale, mais il est exclu que les divers groupements géographiques puissent être le fruit d'un simple hasard.

La reconstitution des tables.

Une estimation des quotients à partir des valeurs $F_1 F_2 F_3$ calculées ne permet pas une reconstitution satisfaisante de la mortalité infantile observée et de la différence de mortalité entre les sexes après 25 ans. Il paraît nécessaire de prendre plus de 3 composantes.

CONCLUSION

La technique de l'analyse factorielle appliquée à l'étude des tables de mortalité permet de poser d'une façon relativement claire, le problème des indices représentatifs d'un état complexe, ici de la mortalité selon l'âge, de 1900 à 1950, dans une cinquantaine de pays.

L'indice unique optimal E — si l'on cherche à résumer en un seul chiffre les données d'une table de mortalité — ne peut expliquer plus de 82 % de la variance totale des quotients de mortalité dans les 157 tables considérées.

L'espérance de vie à la naissance e_0 et la mortalité des femmes âgées de 35 à 44 ans sont très proches, à égalité, de l'indice optimal E .

Si l'on abandonne la description par un indice unique et que l'on cherche à analyser les tables au moyen de plusieurs indices, l'analyse fait apparaître trois sous-ensembles relativement homogènes, suffisamment délimités pour qu'il puisse leur être accolé une étiquette et des indices, nommés ici $F_1 F_2 F_3$ ou $\varphi_1 \varphi_2 \varphi_3$.

L'indice φ_1 explicite les 3/4 de la dispersion des quotients. Le groupe d'âges qui le représentent le mieux, 5 à 35 ans environ, et sa diminution avec le temps donnent à penser qu'il reflète des conditions sanitaires très générales, une « toile de fond ».

φ_2 est un indice secondaire, relatif à une part additionnelle de la mortalité des adultes, variable et indépendante du contexte saisi par φ_1 . Il explicite 10 % seulement de la dispersion des quotients tous âges, mais 1/4 de la dispersion des quotients des hommes au-dessus de 40 ans (13 % chez les femmes).

Les causes de cette mortalité additionnelle, à toiles de fond égales, peuvent varier d'un pays à un autre. En France, on peut observer que φ_2 paraît avoir enregistré, dans le temps, les variations de l'alcoolisation du pays. Il s'agit bien là d'une mortalité additionnelle-type, indépendante des possibilités médico-sociales générales.

φ_3 est un indice secondaire également, indépendamment de φ_1 et de φ_2 . Il explicite 6 % de la dispersion des quotients tous âges, mais 60 % de celle du groupe 80 à 84 ans. Son interprétation est difficile, technique actuarielle, problèmes d'enregistrement des décès, etc. Constatons seulement qu'il détermine des groupements de pays impossibles à imputer à un simple hasard.

Les trois indices $\varphi_1 \varphi_2 \varphi_3$ tout en instituant les ordres de grandeur, conduisent à une estimation encore insuffisamment précise des quotients de mortalité des tables. Pour une telle reconstitution, il paraît nécessaire de faire intervenir probablement deux composantes supplémentaires, qui expliciteraient des variations locales marginales, notamment de la mortalité infantile et de la différence de mortalité entre les sexes au-dessus de 25 ans.

Sully LEDERMANN.

TABLEAU 1. — Écart-type et moyenne des variables X retenues pour l'analyse factorielle (logarithme des quotients pour 1 000 ou de 100 — e_0)

VARIABLE	n° j	HOMMES		n° j	FEMMES					
		Écart type σ_j	Moyenne M_j		Écart type σ_j	Moyenne M_j				
100 — e_0	1	0,093	76 233	1,666	369	20	0,106	59 356	1,635	274
0 à 1 an	2	0,241	5 175	1,980	917	21	0,261	3 751	1,903	121
1 à 4 ans	3	0,413	9 039	1,597	076	22	0,442	3 889	1,575	369
5 à 9 ans	4	0,320	4 443	1,152	357	23	0,374	9 564	1,101	082
10 à 14 ans	5	0,276	4 592	0,996	688	24	0,339	9 394	0,979	312
15 à 19 ans	6	0,250	8 278	1,223	363	25	0,323	7 116	1,187	529
20 à 24 ans	7	0,254	5 773	1,374	229	26	0,311	5 593	1,316	822
25 à 29 ans	8	0,258	6 694	1,391	191	27	0,297	0 094	1,362	051
30 à 34 ans	9	0,257	3 312	1,428	605	28	0,283	6 164	1,400	082
35 à 39 ans	10	0,245	8 347	1,509	896	29	0,258	4 979	1,458	204
40 à 44 ans	11	0,223	6 253	1,609	726	30	0,231	6 733	1,525	433
45 à 49 ans	12	0,192	6 032	1,736	299	31	0,202	1 655	1,619	325
50 à 54 ans	13	0,166	7 506	1,874	790	32	0,178	4 093	1,747	662
55 à 59 ans	14	0,148	5 689	2,017	121	33	0,161	5 930	1,892	732
60 à 64 ans	15	0,126	4 185	2,171	274	34	0,139	0 322	2,063	127
65 à 69 ans	16	0,103	5 824	2,329	115	35	0,132	9 841	2,219	459
70 à 74 ans	17	0,086	0 1517	2,485	841	36	0,093	45 996	2,422	306
75 à 79 ans	18	0,069	4 1226	2,632	650	37	0,081	96 895	2,586	025
80 à 84 ans	19	0,057	9 0518	2,763	204	38	0,064	54 807	2,730	045

TABLEAU 2. — Sexe masculin : coefficients de corrélation ($\times 1 000$) entre les quotients de mortalité des groupes d'âges, dans les 157 tables de mortalité *

VARIABLE	n°	100-e	0-1	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
100 — e_0	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
0-1	2	937	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1-4	3	948	905	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5-9	4	915	844	917	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
10-14	5	933	842	906	929	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
15-19	6	911	801	872	879	949	—	—	—	—	—	—	—	—	—	—	—	—	—	—
20-24	7	898	782	856	861	916	985	—	—	—	—	—	—	—	—	—	—	—	—	—
25-29	8	917	793	865	869	921	987	984	—	—	—	—	—	—	—	—	—	—	—	—
30-34	9	932	810	874	875	920	940	949	982	—	—	—	—	—	—	—	—	—	—	—
35-39	10	904	790	841	839	888	891	890	925	939	—	—	—	—	—	—	—	—	—	—
40-44	11	933	811	841	844	890	889	888	936	974	933	—	—	—	—	—	—	—	—	—
45-49	12	915	790	800	805	853	846	840	891	940	920	988	—	—	—	—	—	—	—	—
50-54	13	886	758	744	757	805	800	792	844	889	887	956	982	—	—	—	—	—	—	—
55-59	14	850	722	697	703	753	749	740	790	832	849	910	949	988	—	—	—	—	—	—
60-64	15	828	696	676	671	736	726	715	765	809	835	888	931	970	991	—	—	—	—	—
65-69	16	815	702	672	647	691	704	691	735	782	796	862	907	940	951	972	—	—	—	—
70-74	17	818	713	684	653	692	705	687	726	763	780	836	875	909	939	927	963	—	—	—
75-79	18	715	635	597	559	602	614	591	618	647	653	714	746	777	820	844	871	946	—	—
80-84	19	520	479	456	409	419	419	384	412	441	462	498	524	542	576	610	665	749	878	—

NOTE* — 1° Le tableau est symétrique. Seule la partie inférieure est notée.

— 2° Les calculs ont été effectués sur les logarithmes des variables, c'est à-dire des quotients de mortalité et de la variable supplémentaire constituée par 100 — e_0 .

TABLEAU 3. — Sexe féminin : coefficients de corrélation ($\times 1\ 000$) entre les quotients de mortalité des groupes d'âges, dans les 157 tables de mortalité *

VARIABLE	n°	100-e	0-1	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84
		(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)
100 — e _x	20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
0-1	21	945	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1-4	22	981	895	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5-9	23	954	879	928	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
10-14	24	935	856	876	961	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
15-19	25	928	840	854	926	961	—	—	—	—	—	—	—	—	—	—	—	—	—	—
20-24	26	929	833	854	923	947	991	—	—	—	—	—	—	—	—	—	—	—	—	—
25-29	27	941	841	854	932	951	983	993	—	—	—	—	—	—	—	—	—	—	—	—
30-34	28	951	848	866	941	950	972	981	993	—	—	—	—	—	—	—	—	—	—	—
35-39	29	957	858	863	935	945	959	964	979	992	—	—	—	—	—	—	—	—	—	—
40-44	30	945	884	835	914	922	935	939	959	977	991	—	—	—	—	—	—	—	—	—
45-49	31	915	801	790	864	870	887	892	917	939	963	985	—	—	—	—	—	—	—	—
50-54	32	894	779	750	826	833	856	862	888	911	939	968	990	—	—	—	—	—	—	—
55-59	33	880	775	727	797	803	825	832	858	882	914	934	974	989	—	—	—	—	—	—
60-64	34	883	785	740	797	795	815	820	843	867	897	926	953	967	986	—	—	—	—	—
65-69 (1)	35	796	708	673	705	686	699	715	738	759	783	809	833	867	892	901	—	—	—	—
70-74	36	857	789	732	768	757	765	782	787	819	833	851	877	898	940	864	—	—	—	—
75-79	37	731	683	633	672	655	646	634	650	660	678	683	696	713	738	771	717	876	—	—
80-84	38	567	529	495	518	501	494	484	499	499	507	497	508	516	540	584	556	751	872	—

1° Le tableau est symétrique. Seule la partie inférieure est notée. 2° Les calculs ont été effectués sur les logarithmes des variables, c'est-à-dire des quotients de mortalité et de la variable supplémentaire constituée par 100 — e_x.

(1) Voir note (1) tableau 4.

TABLEAU 4. — Coefficients de corrélation ($\times 1\ 000$) entre les quotients de mortalité des groupes d'âges du sexe masculin et ceux du sexe féminin *

VARIABLE Femmes	n°	100-e	0-1	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
100 — e _x	(20)	989	935	957	922	937	907	887	904	916	889	908	886	856	821	798	783	789	690	516
0-1	(21)	946	896	917	852	852	813	794	804	822	800	821	802	770	732	707	715	719	634	476
1-4	(22)	921	882	979	890	880	844	825	832	835	865	801	769	719	676	664	660	668	575	457
5-9	(23)	938	873	956	958	950	909	884	893	899	859	868	822	767	715	687	665	676	588	443
10-14	(24)	922	852	909	923	961	954	931	924	913	860	872	825	771	717	686	668	676	604	436
15-19	(25)	915	834	886	878	925	967	957	954	936	887	893	851	803	760	739	714	720	630	456
20-24	(26)	911	825	884	876	909	950	949	955	939	893	898	855	809	763	741	715	713	612	443
25-29	(27)	923	834	890	886	922	942	937	955	948	897	917	877	832	788	762	731	733	631	456
30-34	(28)	933	841	898	896	935	936	926	951	953	913	930	894	850	806	780	745	744	636	456
35-39	(29)	939	848	896	893	940	932	914	940	951	914	939	911	876	830	806	776	766	656	471
40-44	(30)	923	825	869	876	928	913	891	923	942	913	945	926	890	862	836	794	787	671	470
45-49	(31)	899	793	813	833	889	859	836	877	911	897	935	933	918	894	870	825	812	695	486
50-54	(32)	877	772	776	797	851	826	801	844	878	879	915	923	924	908	890	847	833	715	504
55-59	(33)	861	769	750	773	821	790	762	805	840	851	884	899	915	912	897	859	850	738	525
60-64	(34)	864	779	758	771	804	782	753	793	829	843	876	895	909	904	900	874	870	760	556
65-69	(35)	773	699	685	690	701	665	630	675	704	727	752	774	800	810	809	788	803	714	545
70-74	(36)	841	789	743	739	759	736	700	724	753	772	803	822	835	839	845	851	890	846	700
75-79	(37)	708	688	645	639	647	608	573	584	612	617	645	652	661	680	682	696	762	807	786
80-84	(38)	546	537	507	480	474	438	396	414	442	438	487	497	499	519	534	576	659	772	903

* Voir tableau 2, note *

(1) Une erreur s'est glissée dans le calcul des corrélations du groupe des femmes âgées de 65 à 69 ans. Elle ne modifie pas toutefois les résultats d'ensemble de l'analyse. Pour certaines utilisations, il serait préférable de remplacer les corrélations du tableau par des estimations obtenues par interpolation entre les groupes encadrants : 50—64 et 70—74 ans.

TABLEAU 5. — Coefficients de corrélation entre les quotients de mortalité *
et les 3 premières composantes principales (coefficients a_j)
et les trois composantes factorielles F (coefficients A_j)

n°	VARIABLE	a_{j_1}	a_{j_2}	a_{j_3}	A_{j_1}	A_{j_2}	A_{j_3}	$\Sigma a^2 = \Sigma A^2$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Hommes :</i>								
1	100 - e_0	97 359	— 08 868	07 335	9 633	0 946	1 559	9 612
2	0-1	88 386	— 09 919	23 198	8 798	— 0 237	2 666	8 448
3	1-4	90 610	— 22 962	24 210	9 347	— 1 280	2 059	9 324
4	5-9	89 782	— 24 808	16 447	9 315	— 0 998	1 309	8 948
5	10-14	93 319	— 23 986	07 581	9 634	— 0 530	0 681	9 340
6	15-19	92 453	— 25 953	03 049	9 602	— 0 263	0 189	9 230
7	20-24	90 725	— 23 881	— 00 338	9 511	— 0 338	— 0 279	9 065
8	25-29	93 623	— 24 254	— 07 689	9 671	— 0 511	— 0 580	9 413
9	30-34	95 090	— 18 324	— 09 753	9 658	1 130	— 0 399	9 471
10	35-39	93 258	— 10 955	— 12 480	9 291	1 881	— 0 241	8 978
11	40-44	95 941	— 03 847	— 17 212	9 366	2 721	— 0 195	9 516
12	45-49	94 752	— 05 944	— 23 087	8 997	3 806	— 0 151	9 545
13	50-54	92 610	— 16 073	— 28 942	8 527	4 899	— 0 103	9 672
14	55-59	90 593	— 31 219	— 23 628	7 939	5 750	1 141	9 739
15	60-64	88 172	— 30 980	— 29 531	7 711	6 019	0 609	9 606
16	65-69	86 286	— 36 449	— 22 295	7 387	5 997	1 477	9 271
17	70-74	86 586	— 41 994	— 10 159	7 273	5 746	2 782	9 365
18	75-79	77 169	— 54 630	— 08 853	6 036	5 454	4 899	9 018
19	80-84	58 241	— 63 435	— 37 597	3 979	4 101	7 461	8 832
<i>Femmes :</i>								
20	100 - e_0	97 416	— 10 442	11 634	9 679	0 577	1 825	9 735
21	0-1	89 195	— 10 751	21 694	8 893	— 0 200	2 509	8 542
22	1-4	88 095	— 21 550	25 146	9 067	— 1 275	2 177	8 858
23	5-9	92 809	— 24 684	19 321	9 603	— 1 089	1 597	9 595
24	10-14	92 912	— 26 047	14 671	9 649	— 0 928	1 142	9 527
25	15-19	94 180	— 22 519	06 250	9 680	— 0 139	0 666	9 417
26	20-24	94 083	— 22 939	03 630	9 681	— 0 024	0 426	9 390
27	25-29	95 436	— 20 316	01 891	9 744	0 312	0 450	9 525
28	30-34	96 455	— 18 869	— 00 310	9 804	0 575	0 365	9 658
29	35-39	97 391	— 14 758	— 03 188	9 789	1 085	0 371	9 714
30	40-44	97 237	— 09 773	— 10 068	9 645	1 872	0 081	9 654
31	45-49	95 658	— 00 154	— 17 458	9 234	3 046	0 002	9 454
32	50-54	94 674	— 06 862	— 21 347	8 965	3 770	0 040	9 465
33	55-59	92 949	— 14 600	— 19 934	8 598	4 273	0 559	9 250
34	60-64	92 939	— 18 919	— 15 180	8 485	4 343	1 187	9 227
35	65-69	82 924	— 24 681	— 10 030	7 369	4 290	1 779	(1) 7 587
36	70-74	89 170	— 32 660	— 08 725	7 765	3 979	3 851	9 096
37	75-79	76 971	— 45 075	— 29 632	6 265	3 503	6 069	8 835
38	80-84	59 810	— 53 660	— 51 949	4 384	2 538	8 118	9 156
	Σa	34,15259	0,71930	0,40199	32,7969	7,5962	5,8098	—
	Σa^2	30,98293	2,80874	1,41498	29,0871	3,6231	2,4973	35,2073
	Contribution	81,5 %	7,4 %	3,7 %	76,6 %	9,5 %	6,5 %	92,6 %

(1) Voir note 1, tableau 4.

* Les chiffres des colonnes sont décimaux : 97 359 colonne 2, par exemple, doit se lire 0,97359 et 0946 colonne 6, se lire 0,0946.

TABLEAU 6. — Valeurs numériques des indices E φ_1 φ_2 φ_3 pour quelques pays (*)

PAYS	PÉRIODE	E	φ_1	φ_2	φ_3
Egypte	1936—1938	119	122	77	120
Canada	1930—1932	85	90	77	97
	1950—1952	60	59	111	83
U.S.A.	1900—1902	113	117	98	90
	1929—1931	96	95	117	84
	1949—1951	64	60	139	62
Ceylan	1920—1922	151	144	110	143
	1951—1952	94	90	50	196
Inde	1801—1901	167	157	135	128
	1921—1931	162	154	136	114
	1941—1950	152	147	134	96
Japon	1899—1903	124	124	98	110
	1926—1930	122	120	102	115
	1949—1950	97	98	105	89
Autriche	1901—1905	127	124	98	131
	1930—1933	99	94	107	120
	1949—1951	73	67	115	113
Belgique	1891—1900	115	114	98	116
	1946—1949	75	72	113	96
Danemark	1901—1905	100	104	79	102
	1946—1950	56	55	96	104
Finlande	1901—1910	116	119	75	121
	1950—1951	78	67	144	103
France	1893—1903	118	117	98	121
	1923—1933	101	98	108	106
	1950—1951	68	61	128	98
Allemagne fédérale	1901—1910	114	109	106	129
	1932—1934	85	81	100	123
	1949—1951	66	61	104	114
Italie	1901—1911	117	115	76	149
	1930—1932	95	98	73	125
Pays-Bas	1900—1909	102	104	80	115
	1931—1940	71	68	96	116
	1947—1949	52	51	92	108
Norvège	1901—1902	98	114	49	67
	1931—1941	73	82	74	75
	1946—1950	51	58	73	82
Suède	(1841—1855)	(127)	(122)	(111)	(124)
	1901—1910	98	109	57	98
	1931—1940	77	80	85	99
	1946—1950	55	53	103	101
Suisse	1910—1911	112	104	122	126
	1933—1937	87	80	122	110
Grande-Bretagne	1901—1910	109	106	113	104
	1930—1932	90	86	106	113
	1950	60	50	135	103
Russie	1896—1897	129	138	78	77
U.R.S.S.	1926—1927	108	119	68	79
Nouvelle Zélande	1901—1905	93	95	94	92
	1934—1938	71	70	105	96
	1950—1952	53	48	127	85
Moyenne de 157 tables	—	100	100	100	100

(*) Pour le tableau complet des valeurs des composantes pour les 157 tables, voir *Population* 1959, n° 4.

DISCUSSION

M. THIONET : J'ai acquis récemment quelque expérience de l'analyse factorielle et me rends compte que les résultats présentés par M. Ledermann ont exigé des calculs considérables. Je serais intéressé de savoir quel matériel (électronique?) a été employé.

2^e question : Il existe divers procédés d'analyse factorielle, qui n'aboutissent pas forcément aux mêmes résultats (toujours d'après mon expérience). Un facteur très faible mis en évidence à l'aide d'une méthode pourrait fort bien disparaître ou se transformer si l'on usait d'une autre méthode. Nous souhaitons savoir quelle méthode a été employée par l'INED (puisqu'on n'en a pas essayé plusieurs).

3^e question : Peut-on avoir quelques détails sur les tables qui ont « mal » supporté l'analyse factorielle? et peut-on connaître les interprétations que ce fait a pu susciter?

M. ROSENFELD. — La méthode présentée semble permettre de représenter l'ensemble d'une table de mortalité par la valeur numérique de trois facteurs. Peut-on considérer qu'un seul de ces facteurs, à savoir φ_1 que l'on a présenté comme une sorte de toile de fond, traduise d'une façon suffisante le niveau moyen de la mortalité générale d'une population? D'après les chiffres qui sont donnés et leurs variations dans le temps pour un pays donné, ou dans l'espace de pays à pays, ce facteur φ_1 semble bien suivre ce que l'on imagine être le niveau de la mortalité générale.

M. LEDERMANN. — Je répondrai à M. Thionet que les calculs ont été effectués sur 704 IBM. Une programmation est actuellement au point qui permet d'extraire par exemple les cinq ou six premières valeurs et premiers vecteurs propres d'une matrice de corrélation, par ordre de grandeur décroissante. Je signalerai qu'une programmation existe aussi pour la méthode centroïde.

Dans le cas présent, les deux méthodes « axes principaux » et « centroïdes » ont été essayées. Elles conduisent sensiblement aux mêmes résultats. Ceci est probablement en rapport avec la qualité des données des 157 tables de mortalité utilisées dans l'analyse et sélectionnées par les Nations Unies.

Malgré cette sélection, certaines tables se séparent un peu de l'ensemble des autres au point de vue homogénéité: la Guyane britannique, par exemple, l'Inde, en ce qui concerne la mortalité infantile particulièrement, il y en a quelques autres encore. Je n'ai toutefois pas étudié spécialement ces tables sous l'angle de leur dissidence, trois facteurs ne conduisant pas à une reconstitution numérique générale encore assez poussée.

A M. Rosenfeld, je répondrai que l'indice φ_1 n'a pas exactement le sens d'une moyenne, mais seulement d'une « toile de fond » résumant les conditions de vie d'une population les plus générales et dominantes, celles qui rendent compte de la grande différence de mortalité entre le Danemark et l'Égypte par exemple.

A ces conditions de fond, s'ajoutent des particularités locales se traduisant par des décès supplémentaires pour les adultes (φ_2) par exemple, mortalité additionnelle spécifique pouvant exister dans tel pays mais non dans tel autre, à « toiles de fond » égales.

Le niveau de la mortalité générale qui, lui, résume tout, conditions de fond et les diverses mortalités additionnelles spécifiques, serait plutôt représenté par l'indice E porté par le premier axe principal de l'ellipsoïde et dont est très voisin l'espérance de vie à la naissance par exemple.

Il n'en reste pas moins que, sur le plan numérique, φ_1 et E sont très voisins.