JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

ALFREDO NICEFORO

Contribution à l'étude des corrélations entre le bien-être économique et quelques faits de la vie démographique

Journal de la société statistique de Paris, tome 52 (1911), p. 322-341
<http://www.numdam.org/item?id=JSFS_1911__52__322_0>

III

CONTRIBUTION A L'ÉTUDE DES CORRÉLATIONS ENTRE LE BIEN-ÊTRE ÉCONOMIQUE ET QUELQUES FAITS DE LA VIE DÉMOGRAPHIQUE

1. L'étude des rapports entre les phénomènes démographiques ou autres, et le degré d'aisance, à l'aide de la méthode des quartiers aisés ou pauvres de la même ville.
2. Méthodes pour l'étude des corrélations.
3. Classe professionnelle, prix du loyer et *mortalité*.
4. Classe professionnelle, prix du loyer et *causes de décès*.
5. Conditions hygiéniques des logements et *mortalité*.
6. Prix du loyer et *conditions du logement*.
7. Conditions sociales, conditions économiques et *mortalité*.
8. Classe sociale, prix du loyer, conditions economiques et sociales et *natalité*.
9. Corrélations entre la classe sociale, les conditions économiques et la *taille* des conscrits.
10. Corrélations entre la *natalité* et la *mortalité*.
11. Différences de mortalité, de causes de décès, de *nombre d'ouvriers* et de *patrons*, et de *nombre d'enfants* dans les quartiers riches et pauvres de la ville de Paris.

1. — L'étude des rapports entre les phénomènes démographiques ou autres et le degré d'aisance, à l'aide de la méthode des quartiers aisés ou pauvres de la même ville

Les rapports existant entre les différents degrés d'aisance et les différents caractères démographiques (mortalité, natalité, morbidité, etc.), ou autres (par exemple la taille) des hommes, aisés ou pauvres, ont été plusieurs fois étudiés à l'aide de la méthode si simple et si élégante des quartiers riches et pauvres de la même ville Après avoir classé les quartiers ou les arrondissements d'une ville d'après leur degré d'aisance, indiqué par des indices différents, tels que le prix moyen du loyer, le revenu probable, le nombre d'ouvriers, de domestiques, de patrons, de contrats de mariage, d'illettrés, d'indigents secourus par les bureaux de bienfaisance, de transports funèbres gratuits, de logements surpeuplés (plus de deux habitants par pièce), etc., on recherche quel est le taux moyen de la natalité, de la mortalité, et ainsi de suite pour chaque groupement de quartiers ou d'arrondissements d'aisance différente, et on arrive ainsi à établir l'allure générale que prennent les différents phénomènes étudiés (mortalité, natalité, morbidité, etc.), en passant des groupements aisés aux groupements moins aisés. Je rappelle, à ce propos, pour la France, la très ancienne étude de J.-B. Baumes et Vincens, sur la mortalité dans les différentes paroisses, riches et pauvres, de la ville de Nîmes de 1770 à 1780 (1); les recherches de Villermé sur la mortalité des différents arrondissements de la ville de Paris en 1822-1826 et sur la distribution de la grandeur de la taille à Paris en l'année XIII et à Amiens en 1835 (2); les études de Loua sur les données démo-

(1) J.-B. Baumes et M. Vincens, *Topographie de la ville de Nîmes*, Nîmes, 1802, ouvrage présenté à la Société de Médecine de Paris en 1782 et récompensé par l'Académie.

(2) L.-R. Villermé, *Annales d'Hygiène publique*, 1829 et 1830; et *Tableau de l'état physique et moral des ouvriers des manufactures*, Paris, 1840, 2 volumes. Le lecteur trouvera l'historique des recherches faites par les économistes, les démographes et les anthropologues sur les caractères

graphiques de la ville de Paris, partagée en arrondissements; les études de Jacques Bertillon pour les quartiers de Paris, de Berlin, de Vienne (1); les recherches de L. March (2), les travaux, enfin, de Shirley Murphy, de Ch. Booth, de Oloriz, de Gini, de Wilton, pour les autres capitales d'Europe (3).

Toutes ces recherches — y compris les nôtres — se sont bornées à comparer les moyennes des phénomènes démographiques ou autres pour chaque quartier ou pour chaque groupe de quartiers, avec les différents indices du degré d'aisance, en se limitant à noter si, avec l'augmentation de l'aisance, les phénomènes démographiques ou autres marquaient une augmentation ou une diminution. En constatant, par exemple, que la natalité, aussi bien générale que spécifique, diminue en passant des quartiers très pauvres aux quartiers pauvres, puis aux quartiers aisés, aux quartiers très aisés, aux quartiers riches — ou bien en constatant qu'il arrive exactement le contraire pour la taille moyenne des conscrits — nous avons déduit l'existence d'une corrélation directe entre la taille et l'aisance et une corrélation inverse entre la natalité et l'aisance.

Nous voudrions aujourd'hui établir sur les chiffres tels que nous les offrent les statistiques des villes, des *indices* constatant non seulement l'*existence,* mais aussi le *degré* de corrélation entre les différentes classes d'aisance et les différents phénomènes démographiques ou autres, présentés par chaque groupement de population.

Mais, avant de présenter les résultats de nos recherches, qu'il nous soit permis d'indiquer en quelques mots les méthodes que nous avons suivies pour l'étude des corrélations.

2. — Méthodes pour l'étude des corrélations

L'existence et le degré de corrélation entre deux séries de phénomènes peuvent être fixés en partant du principe du rapport entre les écarts de chaque classe (de la série) et la moyenne, et l'indice de variabilité de la série elle-même. En appelant x l'écart entre chaque classe et la moyenne, prise pour origine des écarts, et en désignant avec σ l'indice de déviation de la sériation, le rapport $\frac{x}{\sigma}$ donne un indice d' « anormalité » qui constitue le point de départ d'une étude sur le degré de cor-

---

démographiques et autres des hommes appartenant aux différentes classes sociales, aux différentes professions, et vivant en différents degrés d'aisance, à la première partie (p. 1-57) de notre travail : *Antropologia delle classi povere,* 1 volume du *Trattato di Medicina sociale;* Fr. Vallardi, éditeur, Milan, 1910.

(1) *Bulletin de l'Institut international de Statistique,* 1897.

(2) L. March, *Familles parisiennes,* etc, *Journ. Soc. statist.*, Paris, 1904.

(3) Voir aussi les études de Manouvrier, sur la distribution de la taille dans les différents arrondissements de la ville de Paris, *Bulletin de la Société d'anthropologie,* Paris, 1888, et nos recherches analogues sur la distribution de la taille à Paris et en France, et sur la distribution de la mortalité et de la natalité des différents quartiers de la ville de Lausanne dans nos volumes : *Forza e Ricchezza,* 1 volume de la *Biblioteca di scienze moderne*, Bocca, édit., Turin, 1906 et deux volumes de la *Bibliot. Int. de sociologia,* Barcelona, Henrich y C^{ie}, éditeurs 1907 ; — *Anthropologie der Nichtbesitzenden Klassen*, Liepzig und Amsterdam, Maas et Suchtelen, éditeurs, 1910, et *Antropologia,* etc., déjà citée, Milan, 1910, p. 217-251. Les comparaisons anthropologiques entre les hommes appartenant à différentes classes sociales se trouvent aussi dans notre ouvrage : *Les classes Pauvres, recherches anthropologiques et sociales,* 1 volume de la *Biblioth. Int. de sociologia,* dirigée par R. Worms, Giard et Brière, éditeurs, Paris, 1905.

rélation entre les séries à examiner (1), car il y aura corrélation parfaite entre les séries lorsque le rapport entre $\frac{x}{\sigma}$ pour chaque classe d'un des phénomènes à comparer et $\frac{x}{\sigma}$ pour chaque classe d'un autre phénomène est égale à 1, c'est-à-dire lorsque $\frac{\text{Indice d' « anormalité » de chaque classe du caractère A}}{\text{Indice d' « anormalité » de chaque classe du caractère B}} = 1$. Il y aura, au contraire, manque absolu de corrélation lorsque ce même rapport sera égal à zéro, le *coefficient de corrélation* oscillant ainsi entre 1 (corrélation parfaite) et zéro (manque absolu de corrélation). K. Pearson, en partant de ce principe, a appliqué la formule $r = \frac{\Sigma\,(dév.\ x \times dév.\ y \times f)}{n\sigma_1\,\sigma_2}$ où *dév. x* représente la déviation de chaque classe dans l'un des phénomènes, *dév. y* la déviation de chaque classe associée à la première, pour l'autre phénomène et *f* la fréquence de cas pour chaque association; n le nombre total des observations et $\sigma_1\,\sigma_2$ les deux indices de déviation pour les deux sériations; r mesure, entre 0 et $\pm$ 1 le degré de corrélation (2). L'erreur moyenne de cet indice de corrélation est donnée par

$$E_r = \frac{1 - r^2}{\sqrt{n}} \quad (3)$$

Nous donnerons plus loin un exemple d'application de cette méthode pour l'étude de la corrélation entre la natalité et la mortalité dans deux séries formées par les données de 713 villes françaises; mais pour nos études sur la corrélation entre les différents phénomènes économiques, démographiques et autres, dans les quartiers et les arrondissements des grandes villes, où nous nous trouvions en présence de séries formées par un nombre plutôt restreint d'éléments (80, 25, 20), nous avons préféré nous servir de la méthode Yule, plus rapide et donnant, quoique moins précise que la méthode précédente, des résultats très satisfaisants. G.-U. Yule (4) — que nous remercions ici pour les renseignements qu'il a bien voulu nous fournir épistolairement à propos de quelques applications de sa méthode — partage tous les éléments composant les deux séries dont on veut étudier la corrélation, en quatre cadrans qu'on appelle *a* (le cadran à gauche et en haut de celui qui écrit), *b* (le cadran à droite), en haut, *c* (le cadran à gauche en bas) et *d* (le cadran à droite en bas). Dans le cadran *a* viennent se placer tous les cas qui sont, en même temps, dans les séries A et B supérieures à la moyenne (ou à la médiane) respective; dans le cadran *d* tous les cas qui sont, au contraire, inférieurs;

(1) Voir C. B. Davenport, *Statistical Methods with special reference to biological variation.* Second, revised edition, New-York-London, 1901, p. 23, 42, 105.

(2) K. Pearson, *Mathematical contributions to the theory of evolution,* III, *Phil. Trans, Roy. Soc.* London, CLXXXVII, A, 253-318. Voir aussi le mémoire de L. March, *Les représentations graphiques et la statistique comparative,* communication faite à la Société de Statistique de Paris, 1904-1905, où l'on trouvera la théorie des « indices de dépendance », des « coefficients de dépendance » et de leur précision, avec des exemples d'applications pratiques.

(3) Pearson and L.-N.-G. Filon, *Mathematical contributions,* etc, IV, *Phil. Trans.* etc., etc, CXCI, A, 229-311.

(4) E.-U. Yule, *On the Association of attributes in statistics,* etc., in *Phil. Trans.*, A, CXCIV, 257-319.

dans le cadran b les cas qui sont inférieurs dans la série A, mais supérieurs dans la série B; et dans le cadran c les cas qui sont supérieurs dans la série A, mais inférieurs dans la série B.

Puisque dans les cas où

$$ad = bc$$

les deux séries sont indépendantes l'une de l'autre, l'indice de corrélation sera donné par

$$R = \frac{ad - bc}{ad + bc}$$

et R oscillera entre 0 et $\pm$ 1. Le zéro indiquera le manque absolu de corrélation, et 1 indiquera la corrélation parfaite, directe ou inverse. R aussi aura besoin d'être comparé à l'erreur moyenne qui sera donnée par

$$E_R = \frac{1 - R^2}{2\sqrt{N}} \sqrt{\frac{1}{\varphi_1} + \frac{1}{\varphi_2} + \frac{1}{\varphi_3} + \frac{1}{\varphi_4}}$$

où φ_1, φ_2, φ_3, φ_4, sont respectivement égaux à $\frac{a}{N}$, $\frac{b}{N}$, $\frac{c}{N}$, $\frac{d}{N}$; N étant le nombre total des cas compris dans les cadrans (1).

3. — Classe professionnelle — Prix du loyer et *mortalité*

Nous nous sommes servis pour nos recherches des données statistiques relatives aux 80 quartiers de la ville de Paris, aux 20 arrondissements de la même ville et aux 25 quartiers de la ville de Lausanne (2).

En prenant d'un côté le nombre d'ouvriers ou celui des patrons dans chacun des 80 quartiers de Paris et en mettant en corrélation chacune de ces données avec le taux de la mortalité générale (nombre de décès pour 1.000 habitants), ou de la mortalité de 0-1 an (décès de 0-1 an pour 1.000 naissances), dans chaque quartier; — ou bien en mettant en corrélation, pour les vingt arrondissements de Paris, le prix du loyer avec la mortalité générale et avec la mortalité par différents groupes d'âges, — et enfin en mettant en corrélation pour les 25 quartiers de la ville de

(1) K. Pearson a proposé de modifier la formule de Yule comme suit :

$$R_2 = \sin. \frac{\pi}{2} \frac{1}{\sqrt{1 + K_2}}$$

$$\text{avec } K_2 = \frac{4\ abcd \cdot N^2}{(ad - bc)^2\ (a + d)\ (b + c)}$$

K. Pearson, *Mathematical contributions*, etc., VII, *Phil.*, *Trans.* A. CXCV, 1-47, Aug. 16.

(2) Voir les *Résultats statistiques du dénombrement pour la ville de Paris* (dén. du 1891 et du 1901); l'*Annuaire statistique de la ville de Paris; La fréquence des principales causes de mort à Paris*, par Jacques Bertillon, 1906; *Le Livre foncier de Paris*, par A. Fontaine, Paris, 1908, II et *Ville de Lausanne*, *enquête sur les conditions du logement*, Lausanne, 1896.

Lausanne, le prix du loyer avec la mortalité générale ou la mortalité de 0-1 an, nous avons obtenu le tableau suivant :

Séries entre lesquelles on recherche la corrélation	Indice de corrélation $R = \frac{ad - bc}{ad + bc}$	Erreur moyenne de R $E_R = \frac{1 - R^2}{2\sqrt{N}} \sqrt{\frac{1}{\varphi_1} + \frac{1}{\varphi_2} + \frac{1}{\varphi_3} + \frac{1}{\varphi_4}}$
—	—	—
Nombre d'ouvriers et mortalité générale (Paris) .	0,583	± 0,165050
— de patrons et mortalité générale (Paris) .	— 0,849	± 0,076930
— d'ouvriers et mortalité de 0-1 an (Paris).	0,598	± 0,150964
— de patrons et mortalité de 0-1 an (Paris).	— 0,595	± 0,148580
Prix du loyer et mortalité générale (Paris). . . .	— 0,786	± 0,143325
— — de 0-1 an (Paris) [*]. .	— 0,866	± 0,178350
— — de 1-4 ans (Paris) [*] .	— 0,920	± 0,057600
— — de 5-19 ans (Paris) [*].	— 0,951	± 0,040630
— — de 20-39 ans (Paris) [*]	— 0,951	± 0,040630
— — de 40-59 ans (Paris) [*]	— 0,951	± 0,040630
— — de 60 ans et plus (Paris) [*]	— 0,814	± 0,204127
— — générale (Lausanne) . .	— 0,780	± 0,1874
— — de 0-1 an (Lausanne) .	— 0,714	± 0,233145

[*] Sur 1000 habitants de chaque groupe d'âges, combien y a-t-il de décès en un an ?

Voici l'exemple détaillé d'un de nos calculs, et précisément celui de la corrélation, méthode Yule, entre le nombre des patrons et la mortalité générale pour les 80 arrondissements de la ville de Paris.

Cadran *a*. — Nombre des cas où la mortalité et le nombre des patrons sont supérieurs à la moyenne, 11.
Cadran *b*. — Nombre des cas où le nombre des patrons est inférieur à la moyenne et la mortalité supérieure, 35.
Cadran *c*. — Nombre des cas où le nombre des patrons est supérieur, et la mortalité inférieure, 27.
Cadran *d*. — Nombre des cas où le nombre des patrons et la mortalité sont inférieurs, 7.

$$R = \frac{11 \times 7 - 27 \times 35}{11 \times 7 + 27 \times 35} = \frac{77 - 945}{77 + 945} = \frac{-868}{1022} = -0{,}849$$

$$E_R = \frac{1 - 0{,}720801}{\cdot\, 2} \sqrt{\frac{1}{11} + \frac{1}{35} + \frac{1}{27} + \frac{1}{7}}$$

$$E_R = 0{,}139599 \sqrt{0{,}30374700}$$

$$\text{Log. } E_R = \text{log. } 0{,}139599 + \text{log. } 0{,}5511$$
$$\text{d'où } E_R = 0{,}07693$$

Ainsi, en considérant la colonne des indices de corrélations on voit que *la corrélation entre la mortalité, soit générale, soit par groupes d'âges, et le prix du loyer est inverse : le prix du loyer augmente et la mortalité diminue : l'indice de corrélation est donc donné par un chiffre négatif. La corrélation est également négative pour la mortalité et le nombre de patrons* (pour 1.000 individus exerçant eux-mêmes une profession combien y a-t-il de patrons ?) *Il y a au contraire corrélation directe entre le nombre d'ouvriers, dans chaque quartier, et la mortalité. Avec l'augmentation du nombre des ouvriers ou avec la diminution du nombre des patrons dans chaque quartier, la mortalité augmente.*

Mais il est nécessaire d'étudier l'intensité de la corrélation pour les différents phénomènes et de comparer chaque indice de corrélation à son erreur moyenne, car si R était égale, ou supérieure d'une fois ou de deux fois seulement à son erreur moyenne, on peut conclure que l'existence d'une corrélation entre les deux séries n'est pas démontrée (quoiqu'elle puisse exister), — mais si R est plus grande de quatre fois, et davantage, que son erreur moyenne on peut dire que R est réellement l'indice d'une corrélation bien accentuée. On voit alors que *les corrélations les plus fortes sont celles existant entre le prix du loyer et la mortalité par groupes d'âge,*

— et celle, inverse, *entre le nombre des patrons et la mortalité générale.* Et parmi les différents groupes d'âges la corrélation est plus forte pour les âges d'adultes que pour le premier âge.

4. — Classe professionnelle — Prix du loyer et *causes de décès*

Nous avons calculé les indices de corrélations, et leurs erreurs moyennes, contenues dans le tableau suivant, se référant aux causes de décès sur les chiffres de la ville de Paris pour 80 quartiers en ce qui concerne le nombre d'ouvriers, de patrons, la mortalité par phtisie pulmonaire (pour 100.000 habitants, combien de décès par phtisie?) et par diarrhée infantile (pour 1.000 enfants de 0 à 1 an, combien de décès y a-t-il pour diarrhée infantile?) Les autres *indices* du même tableau sont calculés d'après la distribution des causes de décès par 20 arrondissements de la même ville.

Il existe une corrélation très forte entre la phtisie pulmonaire (causes de décès) *et la classe professionnelle : la corrélation est directe entre la phtisie pulmonaire et le nombre d'ouvriers ; elle est inverse si on prend en considération le nombre de patrons. Cette corrélation est tellement forte* (elle est la plus forte parmi toutes les corrélations indiquées au tableau), *que son indice* (0,971 et — 0,985) *se rapproche très sensiblement de ± 1, indice de la corrélation parfaite ; en outre l'erreur moyenne est minime.*

Séries entre lesquelles on recherche la corrélation	Indices de corrélation	Erreurs moyennes de l'indice de corrélation
—	—	—
Nombre d'ouvriers et mortalité par phtisie pulmonaire . .	0,971	± 0,021164
— de patrons et mortalité par phtisie pulmonaire .	— 0,985	± 0,012367
— d'ouvriers et mortalité par diarrhée infantile . .	0,786	± 0,119214
— de patrons et mortalité par diarrhée infantile.. .	— 0,846	± 0,075313
Prix du loyer et tuberculose en général.	— 0,977	± 0,024970
— — des poumons	— 0,977	± 0,024970
— — abdominale.	— 0,920	± 0,057600
— et maladies organiques du cœur..	— 0,920	± 0,057600
— et pneumonie et broncho-pneumonie. . . .	— 0,920	± 0,057600
— et diphtérie	— 0,920	± 0,057600
— et scarlatine	— 0,909	± 0,108500
— et cirrhose et autres maladies du foie . . .	— 0,882	± 0,077700
— et alcoolisme.	— 0,866	± 0,153750
— et méningite simple	— 0,836	± 0,102408
— et diarrhée infantile et débilité congénitale .	— 0,836	± 0,102068
— et coqueluche	— 0,836	± 0,102408
— et rougeole	— 0, 836	± 0,102408
— et maladies de l'appareil respiratoire . . .	— 0,777	± 0,133062
— et bronchite aiguë	— 0,702	± 0,172448
— et variole	— 0,702	± 0,172448
— et congestion, hémorragie cérébrale et ramollissement du cerveau	— 0.702	± 0,172448
— et cancer	— 0,615	± 0,312143
— et diabète.	— 0,500	± 0,36630
— et bronchite chronique.	— 0,363	± 0,435836
— et ataxie locomotrice progressive	— 0,363	± 0,435836
— et fièvre typhoïde.	— 0,250	± 0,435891

Il existe aussi une corrélation très forte entre la *classe professionnelle* et la *mortalité par diarrhée infantile et par débilité congénitale.*

En passant à l'examen des corrélations entre le prix du loyer et les différentes causes de décès on trouve toujours des indices de corrélation inverse, ce qui indique que pour toute cause de décès lorsque le prix du loyer — indice du bien-être économique — augmente, la mortalité spécifique pour cause déterminée de décès diminue. Le diabète et la fièvre typhoïde font exception, présentant des indices positifs. *L'indice de corrélation pour le diabète est positif,* étant de 0,500, ce qui indiquerait que cette cause de décès est toujours plus fréquente au fur et à mesure qu'on remonte vers les groupes sociaux les plus aisés. Cela serait tout à fait en rapport avec ce que nous savons des diathèses arthritiques occasionnées par l'abus de l'alimentation carnée et de la suralimentation, si fréquentes chez les classes aisées. L'erreur moyenne de l'indice, cependant, est trop haute (± 0, 36630) pour qu'on puisse affirmer que l'indice obtenu constitue la démonstration de l'existence d'une corrélation directe entre les deux séries examinées.

La corrélation entre la *fièvre typhoïde et le prix du loyer* indiquée par l'autre chiffre positif du tableau : 0,250, comparée à l'erreur moyenne, assez haute, de ± 0,15521, indiquerait le manque de corrélation entre les deux séries. D'autre part, une simple inspection des moyennes tant à Paris, qu'à Berlin, qu'à Vienne, indique qu'il n'y a pas de relation apparente entre le degré d'aisance des différents arrondissements et la fréquence de la fièvre typhoïde (1).

Par contre, pour presque toutes les autres causes de décès les corrélations sont assez étroites.

Les corrélations les plus fortes sont celles entre le prix du loyer et les différentes formes de tuberculose (ce que les corrélations pour les séries des 80 quartiers nous avaient déjà appris) et, en outre, celles entre le prix du loyer et les *maladies organiques du cœur.*

Suivent en seconde ligne les corrélations, toujours très fortes, du loyer et les formes de maladies infectieuses, telles que la *diphtérie* et la *scarlatine*, la *pneumonie,* la *coqueluche*, la *rougeole,* et même la *variole.*

Suivent finalement, en dernier lieu, les corrélations entre le prix du loyer et les décès pour *cirrhose du foie* et pour *alcoolisme,* et ceux par *maladies de l'appareil respiratoire.*

Les autres indices sont faibles ou nuls. Ainsi, l'indice de corrélation inverse pour le *cancer* (— 0,615) a une erreur moyenne trop grande (± 0,312) pour qu'il puisse avoir quelque valeur, et l'indice de corrélation pour *l'ataxie locomotrice* (— 0 363) est dépassé par son erreur moyenne (± 0,435).

5. — Conditions hygiéniques des logements et *mortalité*

L'enquête pour le logement dans la ville de Lausanne ayant donné une certaine quantité d'indications numériques sur les conditions hygiéniques des logements, nous pouvons dresser, à l'aide des statistiques de cette ville, des séries indiquant

(1) Voir le tableau donné par Jacques Bertillon, p. 143 de son ouvrage sur *La Fréquence des principales causes de décès à Paris,* déjà cité.

les conditions hygiéniques des différents quartiers et les mettre en rapport, tour à tour, avec la mortalité des différents quartiers et les différents prix du loyer.

Ce calcul a été déjà fait par nous depuis longtemps, et les résultats ont été exposés à l'aide de la méthode des moyennes (1). Ici, nous reprenons ce matériel pour en calculer les indices de corrélation. Une première partie de ces indices, concernant la mortalité générale, est groupée au tableau suivant.

On y voit que *toutes les conditions indiquant le degré d'hygiène* (cube d'air par pièce et par local, surface des pièces, nombre de fenêtres, etc.) *dans les logements, présentent des indices de corrélation avec la mortalité, précédés du signe — ; il s'agit donc de corrélations inverses, indiquant que là où ces conditions hygiéniques sont plus accentuées, la mortalité est moindre et vice versa. La corrélation la plus forte est celle existant entre la quantité d'air respirable, dans les pièces où l'on couche, et la mortalité. Suivent de près, comme importance dans leur corrélation avec la mortalité, les cubes d'air pour l'ensemble des pièces, la surface en mètres carrés des logements, l'ampleur des fenêtres et le nombre d'habitants par local.*

Séries entre lesquelles on recherche la corrélation	Indices de corrélation	Erreurs moyennes
—	—	—
Mortalité et mètres cubes d'air par habitant dans les pièces où l'on couche	— 0,946	± 0,061575
— et cube d'air par habitant sur l'ensemble des pièces	— 0,923	± 0,08880
— et surface en mètres carrés par habitant et par pièce sur l'ensemble des pièces	— 0,894	± 0,118472
— et surface moyenne des fenêtres par pièce . . .	— 0,891	± 0,118472
— et nombre d'habitants par local.	— 0,860	± 0,0152334
— — de locaux par logement	— 0,780	± 0,187968
— — moyen de fenêtres par logement . . .	— 0,714	± 0,212845
— — des logements (sur 1000) ayant 1 water-closet pour ménage (les autres logements ayant 1 water-closet pour 2, 3, 4, 5, 6, 7, 8 ménages)	— 0,647	± 0,255816
Mortalité et surface en mètres carrés par local dans les pièces où l'on couche	— 0,636	± 0,282910
— et nombre des logements sur 100 ayant des water-closets dans l'appartement	— 0,454	± 0,345303
— et nombre des pièces sur 100 au soleil.	0,400	± 0,344400
— — de logements humides	0,176	± 0,416670

Les autres indices de l'hygiène du logement sont en corrélation moins accentuée avec la mortalité ; quelques-uns parmi eux, même, — comme le *nombre de pièces sans soleil* et le *nombre de logements humides*, — offrent des indices de corrélation à qui on ne pourrait pas accorder une valeur sérieuse étant donnée la grande ampleur de leur erreur probable.

Il ressort de ceci la grande importance de l'air respirable : la corrélation entre l'air respirable et la mortalité est bien plus forte que la corrélation entre le prix du loyer et la mortalité. Nous avions d'ailleurs déjà indiqué ce fait dans nos premières recherches à la simple inspection des moyennes et des groupes des moyennes (Voir ouvrages cités).

(1) Voir nos ouvrages déjà cités.

6. — Prix des loyers et *conditions du logement*

Il est banal d'affirmer que le prix du loyer est en rapport étroit avec les conditions hygiéniques ou autres du loyer, mais il peut être de quelque intérêt de rechercher quelles sont les conditions du logement qui sont plus étroitement liées au prix du loyer. Nous pouvons faire cette recherche pour quelques-unes des conditions du logement, à l'aide des données déjà indiquées pour la ville de Lausanne. Voici le résultat de nos calculs à ce propos :

Séries entre lesquelles on recherche la corrélation	Indices de corrélation R	Erreurs moyennes E_R
—	—	—
Prix du loyer et nombre d'habitants par local	— 0,968	± 0,025830
— et cube d'air par habitant sur l'ensemble des pièces	0,943	± 0,068142
— et surface en mètres carrés par habitant sur l'ensemble des pièces	0,766	± 0,266514
— et logements humides	— 0,542	± 0,335445

Le prix du loyer est en corrélation inverse, très accentuée, avec le nombre d'habitants par local ; par conséquence, d'autant plus les conditions économiques sont inférieures, d'autant plus le surpeuplement augmente. Parmi les trois conditions hygiéniques fondamentales : cube d'air, surface des pièces, humidité, *c'est le cube d'air qui est le plus étroitement en rapport avec le prix du loyer* et c'était aussi cette condition que nous avions déjà vue être en rapport étroit avec la mortalité générale. Il semble donc que le prix du loyer est surtout en rapport avec le cube de l'appartement et que celui-ci influe très sensiblement sur la mortalité.

La corrélation entre *la surface en mètres carrés et le prix du loyer* est bien moindre que celle entre le prix et le cube de l'appartement ; — la corrélation entre *le prix du loyer et l'humidité* du logement n'est pas démontrée, d'après nos chiffres, étant donnée la grandeur de l'erreur moyenne. L'indice, cependant, est négatif et supérieur à 0,50, ce qui pourrait être considéré comme un commencement de preuve qu'il existe, entre les deux séries examinées, une corrélation inverse.

7. — Conditions sociales — Conditions économiques et *mortalité*

Nous avons jusqu'à présent considéré la classe professionnelle et le prix du loyer en rapport à la mortalité. Les données pour les vingt arrondissements de la ville de Paris nous donnent la possibilité de prendre en considération quelques autres indices des conditions sociales et économiques. Les résultats de nos calculs à ce sujet sont groupés dans le tableau suivant.

Série entre lesquelles on recherche la corrélation	Indices de corrélation R	Erreurs moyennes E_R
—	—	—
Mortalité et nombre d'illettrés	0,975	± 0,036773
— — d'indigents secourus par les bureaux de bienfaisance.	0,975	± 0,0365
— — de convois funèbres gratuits.	0,975	± 0,0365
— — d'individus mal logés (population surpeuplée, plus de 2 habitants par pièce).	0,944	± 0,070

Tous ces indices de corrélation sont très forts et positifs ; en les comparant à ceux du premier tableau, on voit que *les corrélations entre la mortalité et le nombre d'illettrés ou d'indigents, ou d'individus se laissant conduire à l'extrême demeure par le service gratuit des pompes funèbres, ou bien encore, des mal logés, sont plus fortes que les corrélations entre la classe professionnelle (patrons ou ouvriers) et la mortalité, qui cependant étaient déjà assez étroites.*

8. — Classe sociale — Prix du loyer — Conditions économiques et sociales et *natalité*

Après avoir ainsi étudié les corrélations existant entre la mortalité ou les causes de décès et les différentes conditions sociales économiques et autres pour chaque quartier ou chaque arrondissement, nous pouvons répéter, d'après les données offertes par les sources déjà citées, les mêmes calculs pour étudier les corrélations entre la natalité et les phénomènes les plus typiques indiquant la classe sociale et le degré d'aisance. Les résultats de ces nouveaux calculs sont groupés au tableau qui va suivre. A remarquer que la natalité dont ici il s'agit (pour la ville de Paris) n'est pas la natalité générale, mais la natalité spécifique, plus exacte. (Pour 1.000 femmes de 15 à 49 ans, combien y a-t-il de naissances ?)

Série entre lesquelles on recherche la corrélation —	Indices de corrélation R —	Erreurs moyennes E_R —
Nombre d'ouvriers et nombre de familles ayant 0 enfant vivant (80 quartiers de Paris)	— 0,161	± 0,224020
Nombre d'ouvriers et nombre de familles ayant 1 enfant vivant (80 quartiers de Paris)	— 0,801	± 0,107848
Nombre d'ouvriers et nombre de familles ayant 7 enfants vivants	0,638	± 0,151215
Nombre des indigents secourus par les bureaux de bienfaisance et natalité (20 arrondissements de Paris)	0,977	± 0,024970
— d'illettrés et natalité (20 arrondissements de Paris)	0,977	± 0,035842
— de mal logés (population surpeuplée) et natalité (20 arrondissements de Paris)	0,875	± 0,1315
Prix du loyer et natalité (20 arrondissements de Paris)	— 0,836	± 0,101408
— — (25 quartiers, Lausanne)	— 0,780	± 0,187968
Nombre d'habitants par local et natalité (25 quartiers, Lausanne)	0,707	± 0,212585

Il existe donc une corrélation directe très forte entre la natalité et le nombre d'indigents, le nombre d'illettrés et le nombre de mal logés ; la corrélation est aussi très forte entre *le prix du loyer et la natalité ;* cependant cette dernière corrélation est moins accentuée que celle reliant le nombre d'indigents et d'illettrés à la natalité.

Une corrélation assez prononcée existe entre le *nombre d'ouvriers,* pour chaque arrondissement, et le *nombre de familles ayant 7 fils vivants* (indice d'une très haute natalité) ; et une corrélation plus forte encore, mais inverse, existe entre le *nombre d'ouvriers* pour chaque arrondissement et le *nombre de familles ayant seulement 1 enfant vivant. Ainsi lorsque le nombre d'ouvriers croît, le nombre des familles avec 7 enfants augmente et celui des familles à 1 enfant diminue.*

En comparant ces données à celles obtenues dans nos tableaux précédents on constate que *les corrélations entre le nombre d'illettrés (ou d'indigents) et la mortalité ou la natalité sont à peu près de la même intensité; il en est de même* pour la corrélation entre le *prix du loyer* et la *natalité* et pour celle entre le *prix du loyer* et la *mortalité*. Par contre, le *nombre d'habitants* par local est en corrélation plus forte avec la *mortalité qu'avec la natalité;* et également le *nombre d'habitants mal logés* est plus étroitement en corrélation avec la *mortalité* qu'avec la *natalité*.

9. — Corrélation entre la classe sociale, les conditions économiques et la *taille* des conscrits

Dans nos travaux déjà cités, nous avions plusieurs fois montré, à l'aide des simples comparaisons entre les moyennes, que la grandeur de la taille est en raison inverse du bien-être économique et social, et nous avions aussi, dans ces mêmes travaux destinés à créer une anthropologie des classes pauvres, étudié une grande quantité de caractères anthropologiques et autres, en rapport à la condition sociale, en tâchant de mettre en lumière les causes de tout genre qui déterminent une différenciation si profonde entre les caractères, anthropologiques et autres, des classes supérieures et ceux des classes inférieures.

En ce qui concerne la taille, nos comparaisons avaient porté, non seulement sur plusieurs milliers d'enfants mesurés par nous, mais aussi sur des séries de la taille des conscrits et des indices du bien-être économique pour les vingt arrondissements de la ville de Paris, et sur les séries de la taille des conscrits français dans les départements et dans les arrondissements (fractions des départements) à terrain primitif et de transition — ou dans les terrains sédimentaires. On sait que les terrains de la première qualité sont plutôt des terrains pauvres, soutenant des sociétés pauvres, tandis que les terrains de la seconde catégorie sont des terrains riches donnant la vie, généralement, à des sociétés aisées et riches. Nous sommes encore obligé de renvoyer le lecteur à nos travaux démonstratifs de ces faits. Ici, nous désirons simplement mettre en corrélation avec la méthode déjà indiquée, la taille des conscrits avec les indices du bien-être pour les vingt arrondissements de la ville de Paris, et avec la nature du sol pour les départements et les arrondissements de la France tout entière, ainsi qu'on le voit au tableau qui suit :

Séries entre lesquelles on recherche la corrélation	Indices de corrélation R	Erreurs moyennes E^n
—	—	—
Taille et prix du loyer	1	
— et mortalité	— 0,945	± 0,045458
— et revenu probable moyen (*)	0,941	± 0,049807
— et nombre des illettrés	— 0.898	± 0,16744
— — des mal logés	— 0,866	± 0,088764
— — de convois gratuits	— 0,684	± 0,202175
— — des indigents secourus, etc.	— 0,684	± 0,202175
— et arrondissements à terrain sédimentaire	0,629	± 0,09337
— et départements à terrain sédimentaire	0,525	± 0,105038

(*) Calculé d'après le prix du loyer.

Ainsi les *corrélations entre la taille des conscrits, le bien-être économique et la condition sociale sont très étroites; elles se rapprochent de beaucoup à ± 1* (indice de corrélation parfaite) dans beaucoup de cas examinés, en ayant aussi des erreurs moyennes très petites : *lorsque la mortalité, le nombre des illettrés, des mal logés, des indigents croît, la taille diminue* (les indices donnent des chiffres négatifs); *lorsque le prix du loyer et le revenu probable augmentent, la taille diminue.* Également, *lorsque le pourcentage des départements ou de leurs arrondissements à terrain sédimentaire augmente, la taille moyenne augmente, et vice versa.*

10. — Corrélations entre la *natalité* et la *mortalité*

Les données qui précèdent ne nous permettent pas seulement de constater l'existence de corrélations entre les faits économiques ou sociaux et les faits démographiques, mais aussi de rechercher s'il existe — et de quel genre et avec quelle intensité — une corrélation entre les différents faits démographiques ainsi que, par exemple, entre la mortalité et la natalité.

Nous avons donc accompli cette recherche tant pour les quartiers de la ville de Paris que pour ceux de la ville de Lausanne. Mais nous avons aussi étudié des nouvelles séries formées par les 87 départements français, par les 713 villes françaises dont les statistiques du département de l'intérieur donnent, chaque année, les chiffres de la mortalité et de la natalité (1), par les 52 villes de la Bretagne et les 70 villes de Seine et Seine-et-Oise, dont les statistiques susdites donnent les taux de mortalité et de natalité, — et, finalement, par les 69 provinces italiennes d'aujourd'hui et les 68 provinces italiennes de 1863-1867.

En consultant le tableau où nous avons groupé le résultat de nos calculs, le lecteur s'apercevra de suite que la corrélation entre la mortalité et la natalité est une corrélation pour ainsi dire *locale* et *temporaire*, car, à côté de zones où cette corrélation est très forte, il se trouve des zones où la corrélation est très faible et où elle n'existe même pas. Le degré de corrélation, en outre, entre la mortalité et la natalité peut changer sensiblement à travers le temps ainsi qu'il est arrivé en Italie. La cause de ces faits est très probablement à rechercher dans les profondes différences que présente la vie démographique dans les différentes zones du même pays et à travers le temps. Ici nous nous bornerons à exposer les résultats de nos calculs pour l'examen des corrélations entre la mortalité et la natalité dans les zones indiquées.

L'étude de la corrélation entre la mortalité et la natalité pour la France entière, sur les données des 713 villes françaises dont le ministère de l'intérieur publie régulièrement les taux de mortalité et de natalité, a été faite par nous, puisque nous disposions de séries composées d'un grand nombre de données, à l'aide de la méthode, quelque peu longue et laborieuse, dite de Bravais et largement appliquée par K. Pearson, méthode dont nous avons déjà parlé au paragraphe 2 (2).

---

(1) Voyez notre étude ; *Quelques observations sur la dispersion et la comparaison des courbes de mortalité et de natalité en France*, dans le *Journal des économistes*, Paris, 1911.

(2) Nous avons suivi, pour simplifier les calculs, la méthode dite des « origines arbitraires » des écarts; par conséquent nous nous sommes servis de la formule qui prend en considération la différence entre l'origine arbitraire et la moyenne : $r = \left(\frac{\Sigma \text{ dév. } x \times \text{ dév. } y \times f}{n} - d_x d_y \right) \frac{1}{\sigma_x \sigma_y}$ où $d_x d_y$ exprime

Le total de la colonne Σ donne $-2321 + 3932 = 1611$. Nous savons que $n = 713$; et nous avions déjà calculé $\sigma_x = 6,24$; $\sigma_y = 6,05$; les différences d_x et d_y sont de 1,91 et 0,93; donc :

$$r = \left(\frac{1611}{713} - 1,7763\right) \frac{1}{6,24 \times 6,05} = 0,0125530$$

Ce chiffre est la mesure de la corrélation existant entre la série des 713 mortalités et celle des 713 natalités pour les villes françaises. Il indique qu'il n'existe pas de corrélation entre les deux séries examinées, car la mesure de la corrélation oscille entre 0 et $\pm$ 1, le zéro indiquant absence de corrélation et $\pm$ 1, corrélation parfaite, directe ou indirecte. En outre, l'erreur moyenne E_r, à laquelle il faut comparer l'indice de corrélation, est donnée par

$$\frac{1 - 0,000156}{\sqrt{713}} = 0,038$$

ce qui confirme le manque de corrélation entre les deux séries examinées, mortalité et natalité, pour le total des 713 zones territoriales françaises.

La représentation graphique de cette corrélation — méthode Galton (1) — nous a donné aussi, pour ces deux séries, l'indication du manque de corrélation. En inscrivant d'un côté, en ordre croissant, les unités de la mortalité et en mettant en regard la natalité moyenne de toutes les zones présentant la mortalité en question, on obtient pour les 713 villes les éléments nécessaires pour construire la représentation graphique de la corrélation, méthode Galton. Cette représentation ne donne comme résultat, ni une diagonale de droite à gauche, ni une diagonale de gauche à droite, indices d'une corrélation directe ou inverse, mais une ligne presque verticale indiquant l'absence de corrélation, qui accuse cependant des traces de corrélation seulement pour les valeurs extrêmes (natalité très haute, mortalité très haute; natalité très basse, mortalité très basse).

D'ailleurs, en recourant à la méthode extrêmement grossière des moyennes de cinq éléments, nous pouvons construire le tableau suivant de corrélation :

Aux mortalités de (1) correspond une natalité moyenne de (2) :

Mortalité (1)	Natalité (2)
9/1.000	18,1/1.000
16 —	21,0 —
21 —	20,9 —
26 —	20,6 —
31 —	22,4 —
35 —	27,0 —

le produit des différences entre l'origine arbitraire et la moyenne, pour la mortalité et la natalité. Voir DAVENPORT, ouvrage cité, p. 45 et E.-U. YULE, *On the theory of correlation* (*Journ. roy. stat. Society*, LX, Déc., p. 1-44).

(1) Voir F. GALTON, *Correlations and their Measurement, etc.*, in *Proc. Roy. Soc.*, London, p. 136 et suivantes, 1888; *Natural Inheritance*, London, Macmillan, 1889.

Où l'on voit qu'à l'élévation régulière du taux de la mortalité ne correspond pas une élévation ou une diminution régulière de la mortalité. Aux mortalités extrêmes, seulement, il correspond des natalités extrêmes, mais à l'augmentation moyenne de la mortalité, il correspond, au contraire, une diminution de la natalité.

Voici, maintenant, le tableau résumant nos calculs pour la corrélation (méthode Yule : $R = \frac{ad - bc}{ad + bc}$; $E_R = \frac{1 - R^2}{2}\sqrt{\frac{1}{\varphi_1} + \frac{1}{\varphi_2} + \frac{1}{\varphi_3} + \frac{1}{\varphi_4}}$ où $\varphi = \frac{f}{n}\Big)$ entre la natalité et la mortalité dans les séries formées par les 87 départements français, les 70 villes de Seine et Seine-et-Oise, les 69 provinces italiennes d'aujourd'hui, les 68 provinces italiennes de 1863-1867, les 25 provinces du sud d'Italie en 1863-1867 (1), les 20 arrondissements de la ville de Paris et les 25 quartiers de la ville de Lausanne.

Corrélations entre la mortalité et la natalité	Indices de corrélation R	Erreurs moyennes E_R
—	—	—
Dans les 87 départements français	0,162	± 0,1544
— les 70 villes de Seine et Seine-et-Oise	0,297	± 0,223391
— 69 provinces italiennes (1901-1905)	0,772	± 0,11514
— 68 provinces italiennes (1863-1867)	0,595	± 0,1625
— les 25 provinces du Sud-Italie (1863-1867)	0,818	± 0,153822
— les 25 quartiers de la ville de Lausanne	0,730	± 0,210150
— les 20 arrondissements de la ville de Paris	0,977	± 0,032450

Pour la *France tout entière* (série des 87 départements) on ne trouve pas de corrélation entre la natalité et la mortalité; pour les villes de deux départements (*Seine* et *Seine-et-Oise*) l'indice de corrélation est très faible; il est à peine supérieur à son erreur moyenne; la corrélation, au contraire, de signe positif et par conséquent directe, est très forte à *Paris* où les arrondissements à haute natalité sont aussi ceux qui présentent une très haute mortalité. La corrélation est aussi assez forte à *Lausanne*.

Tandis qu'il n'apparaît pas de corrélation entre la natalité et la mortalité pour la France dans son ensemble, l'*Italie* — également dans son ensemble — présente, au contraire, cette corrélation positive, d'une façon assez marquée : mortalité et natalité sont entre elles en rapport direct, l'indice étant de 0,772 et l'erreur moyenne de ± 0,11514. Il n'en était pas ainsi il y a trente ans, où la corrélation tout en existant, et toujours positive, était plus faible pour l'ensemble des provinces italiennes. Dans la zone du midi, toutefois, elle était très accentuée.

Ceci nous donne une idée des changements qui ont lieu dans la vie et dans la composition démographique des zones et des groupes humains à travers l'espace et le temps.

(1) Nous avons calculé les indices de corrélation pour les données italiennes de 1863-1867 sur les chiffres de la mortalité et de la natalité recueillis dans l'excellent ouvrage du professeur Adolphe Musco, *Il movimento naturale della popolazione italiana per provincie e regioni,* nel periodo 1862-1907, vol. I, Napoli, 1910.

Différences de mortalité, des causes de décès, du nombre *d'ouvriers* et de *patrons* et du nombre *d'enfants par famille*, dans les quartiers riches et pauvres de la ville de Paris

Après avoir ainsi encore une fois démontré l'existence de corrélations assez étroites entre les indices de la classe sociale ou du bien-être économique et hygiénique et les différents caractères démographiques — natalité, mortalité, causes de décès; — et après avoir montré les différents degrés d'intensité de ces corrélations, nous ne voulons pas abandonner ce sujet sans exposer les résultats d'une dernière recherche sur les matériaux offerts par les séries démograhiques des 80 quartiers de la ville de Paris.

Sur ces séries formées d'un nombre relativement élevé de données, il nous était possible de calculer le *standard deviation* de chaque série, ou *indice* de variabilité; sigma (le rapport de sigma à la moyenne donne le *coefficient* de variabilité $C = \frac{\sigma}{M} \times 100\,\%$ [1]) et, à l'aide des valeurs de sigma, on a pu soumettre les différences trouvées entre les moyennes de chaque phénomène dans les quartiers riches et dans les quartiers pauvres de la ville de Paris, à un examen précis et délicat permettant d'établir la valeur qu'on peut accorder à ces différences.

Étant donné, par exemple, la différence x entre la moyenne de la mortalité dans les quartiers pauvres et la moyenne de la mortalité dans les quartiers riches (la moyenne de la mortalité dans les quartiers pauvres étant plus haute que celle des quartiers riches), *quelle est la valeur qu'il faut attribuer à cette différence?* S'agit-il d'une différence qu'on peut réputer comme étant due à des causes accidentelles, de manière à ce que les deux moyennes empiriques (moyenne de la mortalité pour les pauvres et moyenne de la mortalité pour les riches) puissent être considérées comme deux simples expressions de la même moyenne abstraite? Dans ce cas, la différence trouvée n'aurait aucune signification comme indice d'une véritable différence entre les deux phénomènes étudiés. Ou bien cette différence est-elle tellement importante pour que les deux moyennes empiriques dont elle résulte puissent être considérées comme étant deux expressions bien distinctes de deux différentes moyennes abstraites? Dans ce cas, la différence constatée serait vraiment l'indice d'une profonde et sensible différence entre les deux séries de faits pris en examen.

Cette recherche, d'autre part, constitue, pour ainsi dire, une sorte de complément à celle que nous venons de faire à l'aide des corrélations, car, pour continuer l'exemple de la mortalité, après avoir démontré qu'il existe une corrélation inverse entre le bien-être économique et la mortalité des différents quartiers d'une ville, l'étude que nous allons rapidement aborder nous dira si la différence trouvée entre la moyenne de la mortalité pour les quartiers riches et la moyenne de la mortalité pour les quartiers pauvres est une différence tellement importante qu'on puisse l'interpréter comme étant l'indice de l'existence de deux moyennes abstraites, l'une différente de l'autre, de deux groupes de mortalités, dont l'un est spécial aux quartiers riches, et l'autre aux quartiers pauvres.

---

(1) K. Pearson, *Mathematical contribution*, etc, III *Phil. Trans. Roy. Soc.*, London CLXXXVII, A. Voir aussi E.-T. Brewster, in *Man. Proc. Boston Soc. Nat. Hist.*, XXIX, Yuly, 1899, p. 45 et suivantes.

Nous avons déjà dit que nous nous servirons, dans cette recherche, seulement des séries de 80 données, fournies par les 80 quartiers de la ville de Paris. Nous étudierons ainsi : la mortalité générale, celle par groupes d'âge, quelques-unes parmi les plus importantes causes de décès, le nombre de familles ayant 0, 1, 2, 3, 4, 5, 6 et 7 enfants, et aussi la distribution des patrons et des ouvriers dans les différents quartiers de la ville, — en rapport à la classification, déjà faite, de la ville elle-même, en quartiers riches et pauvres, — la classification des 80 quartiers de Paris, en riches et pauvres, étant faite d'après les indications données par Jacques Bertillon dans sa communication à l'Institut international de Statistique (1897).

Si nous appelons M'_1 et M'_2 deux moyennes empiriques et M_1 et M_2 les deux moyennes abstraites qui leur correspondent, nous savons que M'_1 étant plus grand que M'_2 et d leur différence, il existe une probabilité P que la moyenne abstraite M_1 est aussi plus grande que M_2. La probabilité P est donnée par

$$P = \frac{1}{2}\left[1 + \Theta(\gamma)\right]$$

et d'autant plus P se rapprochera de 1, d'autant plus nous pourrons dire que la différence trouvée entre les deux moyennes empiriques constitue l'indice d'une différence entre les deux moyennes abstraites et par conséquent entre les deux phénomènes étudiés.

Pour obtenir la valeur de $\Theta(\gamma)$, nous commencerons par établir l'indice de variabilité σ, pour chaque sériation à étudier, σ étant donné par

$$\sigma = \sqrt{\frac{\Sigma\,[f(x - M)^2]}{n}}$$

où f représente la fréquence de chaque classe de la série; $(x - M)^2$ le carré de chaque écart des classes de la moyenne ; et n le nombre des observations. Nous nous sommes servi, pour établir le sigma et la moyenne de chacune des seize sériations que nous avons dressées, de la méthode dite des ν_1 et ν_2 :

$$\nu_1 = \frac{\Sigma(yf)}{n} \quad \text{et} \quad \nu_2 = \frac{\Sigma(y^2 f)}{n}$$

où y représente l'écart de chaque classe d'une origine arbitraire égale à zéro. La moyenne alors est donnée par $M = y_m + \nu_1 \lambda$ et par $\sigma = \lambda \sqrt{\nu_2 - \nu_1^2}$ où y_m indique la valeur de la classe qui a été choisie comme origine arbitraire et λ la grandeur qu'on a choisie comme échelle de mesure de la série (1).

Ayant ainsi calculé pour chaque série la valeur de *sigma*, on trouve à l'aide de sigma la valeur de γ qui est donnée en résolvant

$$d = \gamma \sqrt{\frac{2\sigma_1^2}{n_1} + \frac{2\sigma_2^2}{n_2}}$$

d étant la différence entre les deux moyennes empiriques, σ_1 et σ_2 les deux *sigma* des deux sériations à comparer, et n_1 n_2 le nombre d'observations; γ étant connu, la valeur de $\Theta(\gamma)$ est donnée par

$$\Theta(\gamma) = \frac{2}{\sqrt{\pi}} \int_0^{\gamma} e^{-\gamma^2}\, d\gamma.$$

(1) Voir Davenport, ouvrage cité, p. 15, 20.

qu'on trouve dans les tableaux de tout traité de calcul des probabilités. Ainsi la valeur de $\Theta(\gamma)$ se rapprochant sensiblement de 1 indiquera la grande valeur que nous devons accorder aux différences constatées entre les moyennes.

Également, ainsi que l'a fait Lexis, en donnant à γ la valeur de 3, nous pouvons, sans qu'il soit besoin de considérer la valeur $\Theta(\gamma)$ résoudre

$$D = 3\sqrt{\frac{2\sigma_1^2}{n_1} + \frac{2\sigma_2^2}{n_2}}$$

où D représente la « différence maximum » et constater si d, ou différence trouvée entre les moyennes empiriques, est plus grande que D. Dans ce cas, nous pouvons croire que les deux moyennes empiriques correspondent réellement à deux moyennes abstraites.

Nous pouvons aussi tout simplement calculer l'erreur moyenne

$$\rho = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

et, en comparant ρ à la différence trouvée entre les moyennes, constater si cette différence est de beaucoup plus grande que la valeur de ρ. Dans ce cas (comme dans les cas où $\Theta(\gamma)$ se rapproche de 1, et où D est plus petit que d_1), nous pourrons conclure à la réelle importance des différences constatées entre les moyennes.

Nous nous sommes servi pour toutes les séries examinées de la méthode de $\Theta(\gamma)$ en recherchant la valeur de γ; nous avons également recherché la valeur de D et pour quelques séries nous avons aussi donné la valeur de ρ.

Les résultats de nos calculs sont groupés au tableau suivant. Mais nous donnons avant tout la reproduction abrégée d'une page de nos calculs, qui servira d'exemple pour trouver la valeur de σ, de γ, de D et de ρ, pour la mortalité de 15 à 34 ans dans les 80 quartiers de la ville de Paris.

Mortalité de 15 à 34 ans (nombre de décès pour 1.000 habitants de cette catégorie d'âge) dans les 80 quartiers de Paris

Taux de la mortalité	Nombre de quartiers (f)	Écarts de l'origine arbitraire (y)	yf	y^2f
2	2	— 6	— 12	72
3	3	— 5	— 15	75
4	5	— 4	— 20	80
5	6	— 3	— 18	54
6	7	— 2	— 14	28
7	9	— 1	— 9	9
8	16	0	0	0
9	8	1	8	8
10	12	2	24	48
11	4	3	12	36
12	5	4	20	80
13	»	5	»	»
14	1	6	6	36
15	»	7	»	»
16	1	8	8	64
17	1	9	9	81

$$\nu_1 = \frac{\Sigma(yf)}{n} = \frac{-1}{80} = -0{,}01,$$

$$\nu_2 = \frac{\Sigma(y^2f)}{n} = \frac{671}{80} = 8{,}3$$

Moyenne $= 8 - 0{,}01 = 7{,}99$

$$\sigma = \sqrt{8{,}3 - 0{,}01^2} = \sqrt{8{,}2999} = 2{,}88$$

Avec la même méthode, on trouve pour les 48 quartiers aisés une moyenne de la mortalité de 6,70, et un sigma de 2,75. Pour les 32 quartiers pauvres, on trouve une moyenne de 9,94 et un sigma de 1,92. Nous aurons alors :

Différence entre les moyennes : $d = 3{,}24$.

$$d = \gamma \sqrt{\frac{2 \times 2{,}75^2}{48} + \frac{2 \times 1{,}92^2}{32}}\ .$$

$$d = \gamma \sqrt{0{,}545}$$

$$3{,}24 = \gamma \times 0{,}73$$

$$\gamma = 4{,}4$$

$$\Theta(\gamma) = 0{,}9999999$$

$$D = 3\sqrt{0{,}545} = 2{,}19$$

plus petite que la différence trouvée (d) 3,24.

$$\rho = \sqrt{\frac{2{,}75^2}{48} + \frac{1{,}92^2}{32}} = 0{,}52$$

Passons maintenant à l'examen du tableau de la page suivante.

On voit d'abord (première colonne) que les *quartiers riches* ont, en comparaison des *quartiers pauvres,* un nombre plus grand de *patrons* et un chiffre plus bas d'*ouvriers,* ainsi qu'un taux plus bas de *mortalité générale,* de *mortalité de 0-1 an,* de *15-34 ans,* de *15-59 ans,* de *décès* par *phtisie* et par *diarrhée infantile.*

La distribution du *nombre des familles classifiées par nombre d'enfants,* se fait aussi d'une manière très caractéristique dans les quartiers riches et pauvres de la ville (première colonne). Les *familles ayant 0 enfant ou 1 enfant* seulement sont plus nombreuses dans les quartiers riches. Les *familles ayant 2 enfants* sont presque en nombre égal dans les quartiers riches et dans les quartiers pauvres (la différence est minime, étant de 1,6 pour les quartiers riches). *Les familles de 3, 4, 5, 6, 7 enfants* sont plus fréquentes dans les quartiers pauvres.

La première colonne du tableau nous donne ainsi les écarts entre les moyennes de deux grandes catégories de quartiers (riches et pauvres) pour chaque phénomène. Quelle est la valeur de ces écarts ? Il y en a qui sont très accentués (+ 29, — 23, — 17), il y en a qui sont assez bas (— 1, — 5, — 6, etc.), et cependant, il se pourrait très facilement qu'un petit écart ait plus d'importance qu'un écart plus large. Nous trouverons la réponse à cette question en considérant les colonnes (2), (3), (4), (5) du tableau.

Les colonnes (2) et (3) nous renseignent sur les valeurs de γ et de $\Theta(\gamma)$. Lorsque γ se rapproche de 3, $\Theta(\gamma)$ se rapproche de l'unité et il y a donc une probabilité $P = \frac{1}{2}[1 + \Theta(\gamma)]$ que les deux moyennes des deux phénomènes comparés correspondent à deux différentes moyennes abstraites. Les différences donc ne sont pas accidentelles. *On voit immédiatement que presque toutes les différences constatées entre les quartiers riches et pauvres à la colonne (1) du tableau sont des différences profondes indiquant non pas l'existence de deux moyennes empiriques qui seraient l'expression de la même moyenne abstraite, mais indiquant l'existence de deux*

COMPARAISONS entre les quartiers riches et pauvres pour les 80 quartiers de la ville de Paris	DIFFÉRENCES entre les moyennes des quartiers pauvres et celles des quartiers riches — (Les signes + ou — devant chaque différence indiquent que la moyenne est plus haute ou plus basse pour les quartiers riches). (1)	VALEURS de γ (2)	VALEURS DE $\Theta(\gamma)$ $\Theta(\gamma) = \frac{2}{\sqrt{\pi}} \int_0^{\gamma} e^{-\gamma^2} d\gamma$ (3)	VALEURS DE D $D = 3\sqrt{\frac{2\sigma_1^2}{n_1} + \frac{2\sigma_2^2}{n_2}}$ (4)	VALEURS DE ρ $\rho = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (5)
Nombre de patrons dans les quartiers riches et pauvres. . . . . (Pour 100 individus exerçant eux-mêmes une profession)	+ 17,26	7,8	0,9999999	6,6	—
Nombre d'ouvriers dans les quartiers riches et pauvres. . . . . (Pour 100 individus exerçant eux-mêmes une profession)	— 23,60	9,8	0,9999999	7,2	—
Mortalité générale dans les quartiers riches et pauvres. . . . . (Pour 1.000 habitants)	— 7,29	7,4	0,9999999	3,6	0,73
Mortalité de 0 à 1 an dans les quartiers riches et pauvres. . . . . (Pour 1.000 enfants de 0-1 an)	— 23,8	1,1134	0,8835330	60,3	13,9
Mortalité de 15 à 34 ans dans les quartiers riches et pauvres . . (Pour 1.000)	— 3,24	4,4	0,9999999	2,19	—
Mortalité de 35 à 59 ans dans les quartiers riches et pauvres. . . (Pour 1.000)	— 5,3	5,5	0,9999999	2,88	
Nombre de décès par phtisie dans les quartiers riches et pauvres . (Pour 1.000 habitants)	— 1,67	4,3	0,9999999	1,14	0,27
Nombre de décès par diarrhée infantile dans les quartiers riches et pauvres . . . . (Pour 1.000 enfants de 0-1 an)	— 27,4	2,7	0,9998657	30,3	
Nombre de familles ayant 0 enfant dans les quartiers riches et pauvres . . . . (Pour 1.000 familles)	+ 29,0	1,27	0,9275136	68,1	16,1
Nombre de familles ayant 1 enfant dans les quartiers riches et pauvres. . . . . (Pour 1.000 familles)	+ 26,8	3,07	0,9999859	25,2	5,7
Nombre de familles ayant 2 enfants dans les quartiers riches et pauvres . . . . (Pour 1 000 familles)	— 1,6	0,32	0,3491258	11,7	
Nombre de familles ayant 3 enfants dans les quartiers riches et pauvres . . . . (Pour 1.000 familles)	— 17,40	4,57	0,9999999	11,4	
Nombre de familles ayant 4 enfants dans les quartiers riches et pauvres . . . . (Pour 1.000 familles)	— 16,25	6,0	0,9999999	8,1	
Nombre de familles ayant 5 enfants dans les quartiers riches et pauvres . . . . (Pour 1.000 familles)	— 10,65	5,5	0,9999999	5,7	
Nombre de familles ayant 6 enfants dans les quartiers riches et pauvres . . . . (Pour 1.000 familles)	— 6,24	6,0	0,9999999	3,12	
Nombre de familles ayant 7 enfants dans les quartiers riches et pauvres . . . . (Pour 1.000 familles)	— 5,22	5,22	0,9999999	3	0,72

moyennes abstraites différentes l'une de l'autre. Cette inspection nous donne une idée exacte de la valeur qu'il faut accorder aux différences exposées à la colonne (1). *Car des différences qui pourraient apparaître très petites à première vue, ainsi que la différence de — 5 ou de — 6 pour le nombre de familles ayant six ou sept enfants, sont au contraire des différences très significatives et très fortes, ainsi qu'on le découvre en consultant les valeurs de γ et de $\Theta(\gamma)$; et, vice versa, des différences qui pourraient sembler très fortes, ainsi que la différence de + 29 pour les familles ayant 0 enfant, — ne constituent en réalité que des différences sans importance.*

En comparant aussi la colonne des différences (col. 1) à la colonne des valeurs de D (col. 4), indiquant la « différence maximum », on voit que les différences constatées entre les moyennes sont presque toujours plus grandes que la différence maximum D. En comparant également la colonne des différences (col. 1) aux valeurs de ρ (col. 5), on voit dans quels cas les différences entre les moyennes sont plus grandes de quatre fois et davantage de la valeur de ρ.

Toutes ces comparaisons se confirment l'une l'autre (et il ne pourrait pas en être différemment) et elles indiquent, qu'à l'exception de la mortalité de 0-1 an et du nombre de familles ayant 0 enfant, les différences trouvées entre les quartiers riches et les pauvres sont très significatives. Pour les familles ayant 2 enfants (différence = — 1,6; $\gamma = 0{,}32$; $\Theta(\gamma) = 0{,}3491258$; $D = 14{,}7$) les quartiers riches et pauvres se ressemblent sensiblement.

Ces mêmes données nous permettent d'étudier un problème d'ordre différent, mais également intéressant : la variabilité des phénomènes démographiques dans les différents quartier d'une ville et dans les différentes zones d'un pays. C'est là un sujet qui fera l'objet d'une prochaine publication.

Alfredo NICEFORO,
Professeur agrégé de statistique
à la Faculté de droit de Naples.