JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

FILIPPO VIRGILII

Sur la théorie des variations statistiques. Indices de dépendance et de corrélation

Journal de la société statistique de Paris, tome 53 (1912), p. 31-42

<http://www.numdam.org/item?id=JSFS_1912__53__31_0>

IV

SUR LA THÉORIE DES VARIATIONS STATISTIQUES

INDICES DE DÉPENDANCE ET DE CORRÉLATION

SOMMAIRE

1. L'indice de variation. — 2. L'indice de dépendance. — 3. Le coefficient de dépendance. — 4. Coefficients de dépendance entre le prix du blé et la nuptialité en Italie (1890-1908). — 5. Coefficients de dépendance entre le prix de plusieurs denrées et la nuptialité. — 6. Phénomènes corrélatifs. — 7. Coefficient de corrélation — 8. Corrélation entre le prix du blé et la nuptialité — 9. Corrélation entre le prix moyen de plusieurs denrées et la nuptialité. — 10 Comparaison entre le coefficient de dépendance et le coefficient de corrélation. — 11 Prix du blé et nuptialité en Angleterre et en France — 12 Autres applications du coefficient de corrélation.

1. — L'indice de variation. — Si nous avons une série de n éléments, qui expriment des manifestations successives du même phénomène, et si nous observons les variations entre deux données consécutives, nous pouvons nous trouver en présence de trois cas; savoir :

I. Les variations sont toutes positives; c'est-à-dire les données de la série augmentent continuellement : *série dynamique positive;*

II. Les variations sont toutes négatives; c'est-à-dire les données de la série diminuent toujours : *série dynamique négative;*

III. Les variations sont tantôt d'un sens, tantôt de l'autre : *série oscillante.*

Puisque les variations dans le premier cas, aussi bien que dans le deuxième, se trouvent toutes et toujours dans le même sens, nous n'avons que des concordances : un terme quelconque de la série est toujours plus petit que le terme immédiatement successif, dans le premier cas; toujours plus grand dans le deuxième. Si nous représentons graphiquement les deux séries, nous aurons, dans le premier cas, une courbe de croissance, dans le deuxième, une courbe qui se développe dans le sens contraire. Naturellement, nous ne nous préoccupons pour le moment que de la *tendance* du phénomène à l'augmentation ou à la diminution, sans tenir compte de la mesure ou de l'intensité de cette tendance.

Dans le troisième cas, comme les variations peuvent se trouver tantôt dans un sens, tantôt dans l'autre, nous aurons des concordances et des discordances; si le terme $r + 1$ est plus grand que r, et $r + 2$ plus grand aussi que $r + 1$, nous avons la concordance; si $r + 1 > r$ et $r + 2 < r + 1$, nous avons la discordance. En général, si deux variations consécutives ont le même signe, nous avons la *concordance,* si elles présentent des signes contraires, on a la *discordance.*

Il est très intéressant de connaître le rapport qui existe entre concordances et discordances pour évaluer, même approximativement, la marche de la série et la tendance générale du phénomène, pour embrasser, tout d'un coup, le développement de la courbe.

Il est évident, du reste, que dans une série de n éléments on aura $n - 1$ variations; il est également intuitif que la somme des concordances et des discordances donne le nombre total des variations; c'est pourquoi, en indiquant avec c les concordances, avec d les discordances, l'*indice des variations*, que nous appellerons i, sera donné par $i = \frac{c-d}{c+d}$.

On voit tout de suite que

$$\text{pour } d = 0 \qquad i = 1$$
$$c = 0 \qquad i = -1$$

c'est-à-dire : lorsque dans une série il n'y a que des concordances, quel que soit leur sens, l'indice de variation est représenté par 1 ; au contraire, si la série ne présente pas de concordance, le même indice est représenté par — 1 ; l'unité positive, donc, est le symbole de la concordance absolue ; l'unité négative, de la discordance complète.

Dans le cas de la série oscillante, on peut avoir :

$$\text{pour } c = d \qquad i = 0$$
$$c > d \qquad i = \frac{\delta}{n} \qquad 1 < \delta < n$$
$$c < d \qquad i = -\frac{\delta}{n}$$

c'est-à-dire : $i = 0$ signifie que, dans une série, les variations d'un sens égalent celles du sens contraire ; quand le nombre des concordances est plus grand que celui des discordances, l'indice de variation est une fraction positive ; quand les concordances sont moins nombreuses que les discordances, l'indice devient une fraction négative. C'est pourquoi, si i s'approche de 1, nous aurons un grand nombre de concordances ; si i s'approche de 0, nous aurons un nombre de concordances presque égal à celui des discordances ; si i s'approche de — 1, nous aurons un très grand nombre de discordances.

A quelque chose, donc, la formule est bonne : elle montre la prépondérance des variations dans l'un ou l'autre sens.

2. — L'indice de dépendance. — La même formule peut nous rendre de plus grands services quand nous allons examiner les variations qui se vérifient, non plus dans une seule série, mais dans deux séries, se rapportant à deux phénomènes observés dans les mêmes périodes, ou moments, de temps.

Soient a_1, a_2,..... a_n les valeurs du phénomène A observé dans une succession d'années ; b_1, b_2,..... b_n les valeurs du phénomène B observé dans la même succession. Comparons, maintenant, les variations successives de A avec celles de B : nous aurons des concordances et des discordances ; si à chaque variation de A correspond une variation dans le même sens de B, nous dirons que les deux séries se développent parallèlement ; si à chaque variation de A correspond une variation dans le sens contraire de B, les deux séries se développent anti-parallèlement ; dans tous les autres cas, le rapport entre la somme algébrique des variations et leur somme totale donnera le degré de dépendance entre les deux séries. Précisément, la formule

$$i = \frac{c-d}{c+d}$$

représente, si elle est appliquée aux deux phénomènes observés en même temps, l'*indice de dépendance* du premier par rapport à l'autre.

Nous aurons encore :

$$\text{pour } d = 0 \qquad i = 1$$
$$c = 0 \qquad i = -1$$

c'est-à-dire : quand à chaque variation d'une série correspond une variation dans le même sens de l'autre série, nous avons leur parallélisme (symboliquement exprimé par 1); quand à chaque variation d'une série correspond une variation dans le sens contraire de l'autre série, on a l'anti-parallélisme (symboliquement exprimé par — 1).

Dans tous les autres cas, i acquiert une valeur entre 0 et + 1, ou bien entre — 1 et 0. Si $i = 1$, on a l'expression de la dépendance parfaite entre deux phénomènes considérés; $i = 0$ exprime la complète indépendance entre eux; si i est très petit, positif ou négatif, cela signifie qu'un phénomène ne se ressent point de l'influence de l'autre; si i acquiert une valeur appréciable, et s'approche de 1, une série dépend de l'autre avec relation directe; s'il s'approche de — 1, une série dépend aussi de l'autre, mais avec relation inverse.

Comme l'indice de variation nous montre la tendance d'une série à varier dans un sens ou dans l'autre, sans tenir compte de la grandeur des variations, ainsi l'indice de dépendance nous fait connaître si deux séries ont relation entre elles ou non, et nous fait connaître aussi le sens de cette relation, mais il ne tient pas compte de l'étendue des variations comparées.

3. — Le coefficient de dépendance. — Au défaut que présente l'indice de dépendance, on supplée par la détermination d'un autre rapport, qu'on a appelé *coefficient de dépendance* (1).

Soient encore les deux séries déjà considérées; si l'on indique par x_1 l'écart de la valeur d'une donnée de la moyenne de la première série, et par x_2 l'écart de la valeur d'une donnée résultante de la moyenne de la deuxième série, s'il y a une relation de dépendance entre les deux phénomènes représentés par les deux séries, il est évident que x_1 et x_2 sont liés entre eux par une équation de la forme $x_1 = c\, x_2$, c étant un facteur constant.

Tout le problème consiste dans la détermination de c, de façon que l'on puisse donner une bonne approximation de la valeur de x_1 pour une valeur donnée de x_2.

La valeur, qui est la plus acceptable pour la propriété bien connue de la moyenne arithmétique, est celle qui rend très petite la somme des carrés des écarts, c'est-à-dire (employant les notations de Gauss) :

$$c = \frac{[x_1\, x_2]}{[x_2^2]}$$

Le signe de c indique si la dépendance entre les deux phénomènes est directe ou inverse, sa valeur nous fait connaître le degré (toujours approximativement) de la dite dépendance.

En suivant un autre chemin, on est arrivé à calculer un coefficient de dépen-

(1) L. MARCH, *Les representations graphiques et la statistique comparative*, dans notre *Journal*, déc. 1904 ; — *Comparaison numérique des courbes statistiques, Journ.*, août-sept. 1905 ; — *Essai sur un mode d exposer les principaux éléments de la théorie statistique, Journ.*, déc. 1910.

dance qui tient compte seulement des variations associées des deux séries, et qui est exprimé par cette formule (1) :

$$k = \frac{[vv']}{\sqrt{[v^2]\,[v'^2]}}$$

où v indique les variations (concordances ou discordances) de la première série, v^1 de la deuxième.

4. — Coefficient de dépendance entre le prix du blé et la nuptialité en Italie (1890-1908). — En voici quelques applications. Presque tous les statisticiens pensent que la courbe des mariages suit anti-parallèlement celle du prix du blé, c'est-à-dire qu'aux bas prix du blé correspondent des chiffres élevés de nuptialité, qu'aux prix élevés du blé correspond un petit nombre de mariages.

Vérifions le fait avec le coefficient de dépendance pour l'Italie, en rapprochant les derniers dix ans du dix-neuvième siècle aux premiers neuf ans du vingtième siècle.

Nous aurons (en adoptant des notations analogues aux notations précédentes) :

ANNÉES	PRIX DU BLÉ à l'importation	MARIAGES (pour 1 000 habitants)	v_b	v_m	v_b^2	v_m^2	$v_b\,v_m$
1890	20	7,34					
1891	23	7,48	— 3	— 0,14	9	0,0196	+ 0,42
1892	21	7,45	+ 2	+ 0,03	4	0,0009	+ 0,06
1893	17	7,39	+ 4	+ 0,06	16	0,0036	+ 0,24
1894	13,5	7,45	+ 3,5	— 0,06	12,25	0,0036	— 0,21
1895	14,5	7,29	— 1	+ 0,16	1	0,0256	— 0,16
1896	15,8	7,07	— 1,3	+ 0,22	1,69	0,0484	— 0,286
1897	19	7,22	— 3,2	— 0,15	10,24	0,0225	+ 0,48
1898	23	6,88	— 4	+ 0,34	16	0,1156	— 1,36
1899	19,2	7,33	+ 3,8	— 0,45	14,44	0,2025	— 1,71
					84,62	0,4423	— 2,526
1900	20,8	7,19					
1901	18,9	7,22	+ 1,9	— 0,03	3,61	0,0009	— 0,057
1902	17,5	7,25	+ 1,4	— 0,03	1,96	0,0009	— 0,042
1903	17,1	7,21	+ 0,4	+ 0,04	0,16	0,0016	+ 0,016
1904	17,7	7,48	— 0,6	— 0,27	0,36	0,0729	+ 0,162
1905	18	7,67	— 0,3	— 0,19	0,09	0,0361	+ 0.057
1906	18	7,77	0	— 0,10	0	0,0100	0
1907	19,1	7,70	— 1,1	+ 0,07	1,21	0,0049	— 0,077
1908	23,4	8,30	— 4,3	— 0,60	18,49	0,3600	+ 2,580
					25,88	0,4873	+ 2,639

(1) Soit m la moyenne arithmétique des variations v de la première série, m' la moyenne arithmétique des variations v' de la deuxième série; la valeur d'une concordance ou d'une discordance sera représentée avec son poids et son signe par le produit $\frac{vv'}{mm'}$. La moyenne de ces concordances ou discordances donne le *coefficient de dépendance.*

$$k = \frac{1}{n} \cdot \frac{[vv']}{mm'}$$

Pour que k soit égal à 1 (indice, comme l'on sait, de la concordance parfaite) lorsque les v et les v' sont tous égaux deux à deux du même signe, on doit avoir $m^2 = \frac{[v^2]}{n}$, $m'^2 = \frac{[v'^2]}{n}$, ce qui permet de mettre le coefficient sous la forme :

$$k = \frac{[vv']}{\sqrt{[v^2]\,[v'^2]}}$$

(*Voir : Journ. de la Soc. de Stat.,* août 1905, p. 269).

Pour la période de 1890-1899, d'après la formule déjà connue, nous avons :

$$k = \frac{-\ 2{,}526}{\sqrt{84{,}62 \times 0{,}4423}} = -\ 0{,}41$$

et pour la période 1900-1908,

$$k = \frac{+\ 2{,}639}{\sqrt{25{,}88 \times 0{,}4873}} = +\ 0{,}74$$

Ainsi, dans les derniers dix ans du dix-neuvième siècle se vérifie une dépendance inverse entre le prix du blé et la nuptialité, dépendance suffisamment appréciable, qui nous permet de conclure que, dans l'ensemble. en cette période, la courbe de la nuptialité a montré une tendance à se développer dans le sens contraire à celui des prix du blé. Au contraire, dans les premiers neuf ans du vingtième siècle, les deux phénomènes ont démontré clairement une dépendance directe : la courbe des mariages procède, avec un parallélisme presque parfait, dans le même sens que la courbe des prix du blé.

D'une très grande importance est la révélation statistique, que le coefficient de dépendance fait ressortir bien mieux que n'aurait pu faire le rapprochement de chaque chiffre concernant les deux phénomènes que l'on a pris en examen. Elle nous dit que si le prix élevé du blé pouvait encore mettre un frein au mariage vers la fin du siècle passé, ce frein vient à manquer au commencement du siècle actuel. Cela prouve que les classes ouvrière et moyenne, qui ressentent plus directement et immédiatement les conséquences du prix d'un genre alimentaire de première nécessité, tel que le blé, ont atteint dans ces dernières années une aisance qui leur a permis, et permet encore aujourd'hui, de faire face victorieusement à l'augmentation du prix du blé.

5. — Coefficients de dépendance entre plusieurs denrées et la nuptialité. — Si notre conclusion, tirée d'un seul indice, semblait trop rigide et faite trop à la hâte, nous pouvons la modifier dans le sens que, du moins, les variations dans le prix du blé n'exercent plus, à présent, sur le mouvement des mariages, la même influence, que l'on avait remarquée jusqu'à ces dernières années.

En est-il de même pour le prix moyen complexif des principaux genres que l'on emploie pour l'alimentation, l'habillement, l'habitation ?

Vérifions-le pour le prix moyen de 19 marchandises (huiles, café, sucre, couleurs, cotons, tissus, laines, peaux, fer, fonte, cuivre, machines, blés, fromages, etc...), toujours en rapport avec la nuptialité (1) :

TABLEAU.

(1) Les prix des marchandises nous sont donnés par A. Necco, *La curva dei prezzi delle merci in Italia negli anni 1881-1909*; supplément a *La Riforma Sociale*, sept -oct 1910, p. 21.

ANNÉES	PRIX DES DENRÉES (Index-Numb.)	MARIAGES (pour 1 000 habitants)	v_p	v_m	v_p^2	v_m^2	$v_p\, v_m$
1890	328	7,34					
1891	319	7,48	+ 9	— 0,14	81	0,0196	— 1,26
1892	303	7,45	+ 16	+ 0,03	256	0,0009	+ 0,48
1893	296	7,39	+ 7	+ 0,06	49	0,0036	+ 0,42
1894	271	7,45	+ 25	— 0,06	625	0,0036	— 1,50
1895	267	7,29	+ 4	+ 0,16	16	0,0256	+ 0,64
1896	270	7,07	— 3	+ 0,22	9	0,0484	— 0,66
1897	270	7,22	0	— 0,15	0	0,0225	0
1898	279	6,88	— 9	+ 0,34	81	0,1156	— 3,06
1899	294	7,33	— 15	— 0,45	225	0,2025	+ 6,75
					1.342	0,4423	+ 1,61
1900	297	7,19					
1901	289	7,22	+ 8	— 0,03	64	0,0009	— 0,24
1902	283	7,25	+ 6	— 0,03	36	0,0009	— 0,18
1903	288	7,21	— 5	+ 0,04	25	0,0016	— 0,20
1904	297	7,48	— 9	— 0,27	81	0,0729	+ 2,24
1905	299	7,67	— 2	— 0,19	4	0,0361	+ 0,38
1906	318	7,77	— 19	— 0,10	361	0,0100	+ 1,90
1907	328	7,70	— 10	+ 0,07	100	0,0049	— 0,70
1908	326	8,30	+ 2	— 0,60	4	0,3600	— 1,20
					675	0,4873	+ 2,00

Pour la période 1890-1899

$$k = \frac{1,61}{\sqrt{1342 \times 0,4423}} = +\,0,06$$

Pour la période 1900-1908.

$$k = \frac{2}{\sqrt{675 \times 0,4873}} = +\,0,11.$$

D'où l'on voit que dans cette première période d'observation, qui nous avait fait constater une dépendance inverse entre le prix du blé et les mariages, on remarque une parfaite indépendance entre la moyenne générale des prix des marchandises que l'on consomme ordinairement et la nuptialité; et, même dans la deuxième période, un phénomène ne se ressent presque pas de l'influence de l'autre.

Avant d'en tirer une conclusion qui puisse nous permettre d'apprécier dans toute son étendue les résultats obtenus, et de généraliser la signification de notre application, nous voulons voir s'il existe un lien ou une analogie, sinon une identité, entre le coefficient de dépendance et celui de corrélation.

6. — Phénomènes corrélatifs. — Si deux phénomènes se produisent de façon qu'aux variations du premier correspondent les variations de l'autre, c'est-à-dire qu'ils augmentent ou diminuent à peu près dans la même mesure, ou bien aussi qu'ils procèdent de façon qu'aux augmentations du premier correspondent d'égales diminutions de l'autre, on dit que les deux phénomènes sont corrélatifs.

C'est un sujet qui mérite l'attention critique de celui qui s'adonne à cette étude, afin qu'on ne donne pas une valeur de concordance et de corrélation à ce qui peut être une simple coincidence accidentelle. Quand Jevons présenta l'hypothèse que les modifications des taches solaires pouvaient exercer une influence sur les crises économiques, tout le monde fut grandement surpris; mais quand il établit une relation, hypothétique elle aussi, entre l'augmentation ou la diminution des taches solaires et l'état météorologique de la terre, la méditation remplaça la surprise, puisque les conditions météorologiques exercent une influence indiscutable sur les récoltes, et celles-ci sur les développements des affaires. Puis, quand après des observations soigneusement répétées, Jevons présenta la périodicité concordante des deux phénomènes, même si différents entre eux, l'hypothèse devint un rapport de corrélation (1).

Ensuite on a étendu ce procédé inductif à différents phénomènes sociaux, et l'on a découvert des corrélations importantes : dans l'anthropologie, entre la distribution des tailles et la dimension de la poitrine, entre la taille et le poids, entre la taille et l'indice céphalique, entre la couleur des cheveux et celle des yeux; dans la démographie, on a établi la distribution des mariages d'après l'âge des époux, la mortalité des enfants du même âge, par rapport à l'âge des parents, la distribution des mariages et la distribution correspondante des naissances en relation avec le prix du blé; dans l'économie, la distribution du nombre des expéditions par chemin de fer des marchandises homogènes en comparaison du nombre des kilomètres parcourus et des poids de chaque expédition, le rapport entre les rentes et les patrimoines, la relation entre le prix du blé et celui du pain; dans la statistique de la culture intellectuelle, la corrélation entre la profession du père et le choix de la faculté universitaire de la part des enfants, et ainsi de suite.

Dans ces derniers temps, du sujet de la corrélation on a fait une théorie mathématique qui a des rapports très intimes avec le calcul des variations et celui des fréquences (2).

7. — Coefficient de corrélation. — Soient deux quantités variables dépendantes A et B; si l'on indique par x_1, x_2,..... x_n les écarts de chaque valeur de A de leur moyenne arithmétique, par y_1, y_2,..... y_n les écarts de chaque valeur de B de leur moyenne arithmétique, si l'on indique aussi par σ_x et σ_y les écarts moyens respectifs, n étant le nombre des valeurs observées pour A et pour B, le coefficient de corrélation, représenté ordinairement par r, prend cette forme :

$$r = \frac{[x\,y]}{n\,\sigma_x\,\sigma_y}$$

où, par les notations déjà connues,

$$\sigma_x = \sqrt{\frac{[x^2]}{n}} \qquad \sigma_y = \sqrt{\frac{[y^2]}{n}}$$

Cette formule, qui donne la valeur de r, conseillée pour la première fois par Bravais, est à présent appuyée par Pearson, employée généralement par les statisticiens, contrairement à la méthode de Galton, qui arrive à des résultats moins exacts (3). Si r est positif, la corrélation s'appelle positive, ce qui signifie que les

(1) A. Liesse, *La Statistique*, Paris, 1905, p. 52, 53.

(2) Pour une bibliographie complète de l'argument, voir G. Undny Yule, *The applications of the method of correlation to social and economic statistics* : dans le *Bulletin de l'Inst. int. de Stat.*, XVIII, I, p. 549-51.

(3) C. Bresciani, *Sui metodi per la misura delle correlazioni*; dans le *Giorn. d. Econ.*, avril 1909.

deux variables A et B augmentent ensemble; au contraire, si r est négatif, la corrélation aussi s'appelle négative, c'est-à-dire la première variable diminue si l'autre augmente; pour $r = 0$ la corrélation est nulle, c'est-à-dire les deux variables sont indépendantes l'une de l'autre.

Si $r = 1$, la corrélation est alors parfaite (positive ou négative) et il résulte clairement de notre formule que, pour $r = 1$, on a

$$x_1 y_1 + x_2 y_2 + \ldots\ldots + x_n y_n = n \sigma_x \sigma_y$$

8. — Corrélation entre le prix du blé et la nuptialité. — Reprenons notre exemple, et calculons le coefficient de corrélation entre le prix du blé et la nuptialité en Italie dans les deux périodes 1890-1899 et 1900-1908. Nous aurons :

En 1890-1899, la moyenne des prix du blé est donnée par $M_b = 18{,}6$, la moyenne des mariages par $M_m = 7{,}29$; de là, si l'on appelle x_b et x_m les écarts des deux moyennes, on obtient :

x_b	x_m	$x_b x_m$	x_b^2	x_m^2
+ 1,4	+ 0,05	+ 0,070	1,96	0,0025
+ 4,4	+ 0,19	+ 0,836	19,36	0,0361
+ 2,4	+ 0,16	+ 0,384	5,76	0,0256
— 1,6	+ 0,10	— 0,160	2,56	0,0100
— 5,1	+ 0,16	— 0,816	26,01	0,0256
— 4,1	0	0	16,81	0
— 2,8	— 0,22	+ 0,616	7,84	0,0484
+ 0,4	— 0,07	— 0,028	0,16	0,0049
+ 4,4	— 0,41	— 1,804	19,36	0,1681
+ 0,6	+ 0,04	+ 0,024	0,36	0,0016
		— 0,878	100,18	0,3228

$$r = \frac{-0{,}878}{10\sqrt{\frac{100{,}18}{10}}\sqrt{\frac{0{,}3228}{10}}} = -0{,}15$$

De façon analogue, pour 1900-1908 nous aurons :

$$M_b = 18{,}8 \qquad M_m = 7{,}53$$

x_b	x_m	$x_b x_m$	x_b^2	x_m^2
+ 2,0	— 0,34	— 0,680	4,00	0,1156
+ 0,1	— 0,31	— 0,031	0,01	0,0961
— 1,3	— 0,28	+ 0,364	1,69	0,0784
— 1,7	— 0,32	+ 0,544	2,89	0,1024
— 1,1	— 0,05	+ 0,055	1,21	0,0025
— 0,8	+ 0,14	— 0,112	0,64	0,0196
— 0,8	+ 0,24	— 0,192	0,64	0,0576
+ 0,3	+ 0,17	+ 0,051	0,09	0,0289
+ 3,6	+ 0,77	+ 2,772	12,96	0,5929
		+ 2,771	24,13	1,0940

$$r = \frac{2{,}771}{9\sqrt{\frac{24{,}13}{9}}\sqrt{\frac{1{,}0940}{9}}} = +0{,}54$$

Le coefficient de corrélation prouve d'une façon évidente que les deux séries sont à peu près indépendantes dans la première période, tandis qu'elles indiquent une dépendance directe dans la deuxième période.

9. — Corrélation entre le prix moyen de plusieurs denrées et la nuptialité. — Passons maintenant au calcul du coefficient de corrélation entre la série des prix des différentes marchandises (denrées, habillements, habitations) et les taux des mariages pour les mêmes périodes de temps. Nous aurons :

Pour **1890-1899**. . . . . $M_p = 289,7$; $M_m = 7,29$

x_p	x_m	$x_p\, x_m$	x_p^2	x_m^2
+ 38,3	+ 0,05	+ 1,915	1.466,89	0,0025
+ 29,3	+ 0,19	+ 5,567	858,49	0,0361
+ 14,3	+ 0,16	+ 2,288	204,49	0,0256
+ 6,3	+ 0,10	+ 0,630	39,69	0,0100
— 18,7	+ 0,16	— 2,992	349,69	0,0256
— 22,7	0	0	515,29	0
— 19,7	— 0,22	+ 4,334	388,09	0,0484
— 19,7	— 0,07	+ 1,379	388,09	0,0049
— 10,7	— 0,41	+ 4,387	114,49	0,1681
+ 4,3	+ 0,04	+ 0,172	18,49	0,0016
		+ 17,680	4.343,70	0,3228

$$r = +\ 0,47$$

Pour **1900-1908**. . . . . $M_p = 302,78$; $M_m = 7,53$

x_p	x_m	$x_p\, x_m$	x_p^2	x_m^2
— 5,78	— 0,34	+ 1,9652	33,4084	0,1156
— 12,78	— 0,31	+ 3,9618	163,3284	0,0961
— 19,78	— 0,28	+ 5,5384	391,2484	0,0784
— 14,78	— 0,32	+ 4,7296	218,4484	0,1024
— 5,78	— 0,05	+ 0,2890	33,4084	0,0025
— 3,78	+ 0,14	— 0,5292	14,2884	0,0196
+ 15,22	+ 0,24	+ 3,6528	231,6484	0,0576
+ 25,22	+ 0,17	+ 4,2874	636,0484	0,0289
+ 23,22	+ 0,77	+ 17,8794	539,1684	0,5929
		+ 41,7744	2.260,9956	1,0940

$$r = +\ 0,84$$

De telle sorte que, si l'on compare le prix du blé au prix moyen de dix-neuf denrées d'usage commun, le coefficient de corrélation indique clairement une dépendance directe dans la première période, dépendance qui augmente sensiblement dans la deuxième période jusqu'à devenir presque parfaite.

10. — Comparaison entre le coefficient de dépendance et le coefficient de corrélation. — En résumé, les résultats de nos calculs nous donnent :

***A)* Prix du blé et nuptialité**

1890-1899. . . . . $k = -\ 0,41$ $\quad r = -\ 0,15$

1900-1908. . . . . $k = +\ 0,74$ $\quad r = +\ 0,54$

***B)* Prix moyen de 19 denrées et nuptialité**

1890-1899. . . . . $k = +\ 0,06$ $\quad r = +\ 0,47$

1900-1908. . . . . $k = +\ 0,11$ $\quad r = +\ 0,84$

Le coefficient de dépendance nous montre qu'il existe une dépendance inverse assez appréciable entre le prix du blé et la nuptialité dans la première période et une dépendance directe très remarquable entre les deux phénomènes dans la deuxième période; une indépendance presque parfaite entre le prix moyen des genres différents de consommation générale et la nuptialité dans les deux périodes. Le coefficient de corrélation révèle : une quasi indépendance entre le prix du blé et les mariages dans la première période et une dépendance directe très évidente dans la deuxième période; une dépendance directe toujours appréciable entre le prix moyen des différents genres et la nuptialité dans la première période, dépendance qui augmente toujours, et qui devient presque parfaite, dans la deuxième période.

L'allure du coefficient de corrélation est plus régulière et semble plus logique et plus conforme à la réalité que le coefficient de dépendance, ce qui en prouverait la supériorité. En effet, on ne peut pas comprendre que, en indiquant par des symboles graphiques la signification du coefficient de dépendance, les courbes du prix du blé et de la nuptialité passent de l'anti-parallélisme de la première période, au parallélisme de l'autre, tandis que les deux courbes du prix des genres de consommation ordinaire et de la nuptialité n'ont aucune relation entre elles dans la première aussi bien que dans la deuxième période.

Au contraire, nous sommes persuadé par le fait que la courbe de la nuptialité a une tendance à se rendre indépendante de la courbe du prix du blé dans les derniers dix ans du dix-neuvième siècle, quand en Italie aussi commence à se produire une remarquable transformation dans les conditions générales de la vie des ouvriers pour progresser ensuite, presque parallèlement, dans les premiers dix ans du nouveau siècle; et si nous remontons du prix du blé à celui des autres genres de consommation ordinaire, on comprend mieux que les deux courbes qui commencent le parallélisme, en 1890-1899, le rendent presque parfait en 1900-1908.

Cela prouve que, à mesure que l'amélioration économique s'élargit et se fortifie, le taux des mariages s'élève tout en élevant le prix des différents genres, car les salaires les plus élevés produisent une aisance générale qui permet de franchir les difficultés des prix élevés des denrées.

Cela avait déjà été remarqué par Bowley pour l'Angleterre, et nos calculs nous autorisent à répéter l'observation pour l'Italie. En effet, Bowley a trouvé que, tandis qu'avant 1870 on remarque en Angleterre une corrélation inverse entre le prix du blé et la nuptialité, après 1870 on passe au phénomène opposé, et il a remarqué aussi que, à cause de l'augmentation des appointements, une variation dans le prix du blé ne frappe assez fort qu'une toute petite partie de travailleurs (1).

En Italie, ce changement dans la loi de dépendance des deux phénomènes se vérifie trente ans après. Nous l'avons déjà constaté par le calcul du coefficient de corrélation, qui passe de l'indépendance pour 1890-1899, à la dépendance directe pour 1900-1908; mais, si nous subdivisons la période entière pour 1890-1908 en groupes de cinq ans chacun (le dernier groupe est, nécessairement, de quatre

(1) « When wheat the chief object of expenditure of the working class, its price was the chief thing for them to consider; and so when wheat rose the marriage rate fell. On the other hand, now that wheat is cheap and wages higher, a change in the price of the loaf is only of great importance to a minority; it is now the general prosperity of the country, well indicated by the condition of foreign trade, that raises the marriage-rate ». L. Bowley, *Elements of Statistics*; London, 1903, p. 175.

ans), nous voyons aussi que le changement se produit au milieu d'une quantité de crises et d'incertitudes.

En voici des résultats :

Prix du blé et nuptialité en Italie

1890-1894. . .	$r = +0,14$	**1900-1904**. . .	$r = -0,39$
1895-1899. . .	$r = -0,58$	**1905-1908**. . .	$r = +0,99$

On voit donc, d'une façon évidente, qu'à une première période, où les deux phénomènes procédaient indépendamment, suivent deux autres périodes, pendant lesquelles les deux phénomènes montrent une dépendance inverse, qui descend jusqu'à ce que, dans la dernière période, le prix du blé et la nuptialité se développent par un lien d'une parfaite dépendance directe. On doit se rappeler que 1895-1899 a été une période de profonde dépression économique pour l'Italie, et le coefficient de corrélation en donne la preuve par une valeur négative très marquée !

11. — **Prix du blé et nuptialité en Angleterre et en France.** — Nous avons légèrement effleuré les recherches de Bowley sur ce sujet, mais il sera très utile et intéressant d'en indiquer les résultats obtenus (1) :

Prix du blé et nuptialité en Angleterre

1845-1860	$r = -0,30$
1875-1890	$r = +0,47$

Avec les mêmes éléments qui ont été fournis par le statisticien anglais, Bunle, dans un récent rapport à la *Société de Statistique* de Paris, a ainsi subdivisé la série (2) ;

1845-1850. . .	$r = -0,461$	**1871-1880**. . .	$r = +0,720$
1851-1860. . .	$r = -0,504$	**1881-1890**. . .	$r = +0,623$
1861-1870. . .	$r = -0,518$	**1891-1896**. . .	$r = +0,497$

Bunle, faisant les mêmes études sur les relations entre les prix du blé et le nombre des mariages en France, est arrivé à ces conclusions numériques :

Prix du blé et nuptialité en France

1811-1820. . .	$r = -0,730$	**1861-1869** . .	$r = -0,010$
1821-1830. . .	$r = +0,294$	**1874-1880**. . .	$r = -0,273$
1831-1840. . .	$r = -0,620$	**1881-1890**. . .	$r = +0,137$
1841-1850. . .	$r = -0,596$	**1891-1900**. . .	$r = -0,326$
1851-1860. . .	$r = -0,560$		

Avant 1860 il y avait même en France une corrélation sensible et négative entre le prix du blé et la nuptialité; au contraire, après cette année-là, on ne voit

(1) Ouvr cit., p 174.
(2) *Journ. de la Soc. de Stat.*, mars 1911, p. 82.

plus aucun lien entre les deux courbes. Dans la première période de la corrélation négative on remarque, il est vrai, une déviation en 1821-1830, dont on ne peut facilement donner une explication; Bunle se borne à présenter une hypothèse d'Adolphe Bertillon, d'après laquelle « l'avilissement extrême du prix des céréales peut aussi baisser le rapport des mariages ».

Un autre statisticien anglais a calculé pour l'Angleterre ces trois coefficients de corrélation (1), qui complètent et confirment les coefficients de Bowley :

Nuptialité et prix (1865-1896)	. . . .	$r = + 0{,}79$
— commerce extérieur (1851-1890)	. . .	$r = + 0{,}90$
— et chômage (1870-1895)	. . . .	$r = - 0{,}87$

C'est-à-dire : l'élévation des prix et l'intensité du commerce extérieur, qui sont des indices de prospérité, augmentent le nombre des mariages; le chomâge procède dans le sens inverse des mariages.

12. — Autres applications du coefficient de corrélation. — Le coefficient de corrélation a été calculé par d'autres écrivains pour établir et mesurer l'éventuelle relation qui existe entre les phénomènes de la même espèce ou d'une espèce différente.

Le Dr J.-P. Norton (2) a établi la corrélation entre le rapport des réserves des banques américaines et le taux de l'escompte, entre ce taux et les variations de l'encaisse, entre le mouvement de l'encaisse et celui des prêts dans la même semaine et dans les semaines suivantes (prouvant ainsi que le maximum de l'oscillation annuelle des prêts se trouve trois semaines après le maximum de l'encaisse).

M. Niceforo a calculé plusieurs coefficients de corrélation entre le prix du logement et la mortalité, entre le même prix et les causes de mort, entre les conditions hygiéniques des maisons et la mortalité, entre les conditions économiques et la taille des conscrits (3).

M. Gini a étudié les indices de concentration et de dépendance, et il a établi des relations entre la somme des successions et le nombre des quotes-parts héréditaires, entre l'altimétrie et la population des communes, entre la productivité matrimoniale, la durée du mariage et l'âge des époux à l'époque de leur mariage et de la mort, entre la fécondité légitime et la durée du mariage, entre la fécondité matrimoniale et l'âge de la mère, etc... (4).

Il obtient l'indice de dépendance par un procédé plus compliqué que celui que nous avons indiqué, et il arrive à des formules qui ne sont que des variations de la célèbre formule de Pareto pour calculer la distribution des revenus.

Sienne, septembre 1911 Filippo Virgilii.

---

(1) C.-U. Yule, *On the changes in the marriage-and birth-rates in England and Wales during the past half Century; dans le Journ. of the Stat. Soc.*, mars 1906, p. 94-96.

(2) *Statistical studies in the New-York money market;* New-York, Mc. Millan 1902.

(3) A Niceforo, *Contribution à l'étude des corrélations*, etc., dans le *Journal de la Soc. de Stat.*, août-sept. 1911.

(4) C. Gini, *Indici di concentrazione e di dipendenza* (Biblioteca dell' Economista, serie quinta vol. XX, disp. 1-2) Torino 1910.