JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

SERBAN GHEORGHIU

Essai sur les finances d'après-guerre de la Roumanie en rapport avec les mouvements économiques européens

Journal de la société statistique de Paris, tome 67 (1926), p. 355-372
<http://www.numdam.org/item?id=JSFS_1926__67__355_0>

II

ESSAI SUR LES FINANCES D'APRÈS-GUERRE

DE LA ROUMANIE

EN RAPPORT

AVEC LES MOUVEMENTS ÉCONOMIQUES EUROPÉENS

Thèse soutenue pour l'obtention du diplôme de statisticien de l'Institut de Statistique de l'Université de Paris le 19 juin 1925

Suite (1)

DEUXIÈME PARTIE (2)

L'ÉNONCÉ DU PROBLÈME DE LA CORRÉLATION

§ 1. — Étant donnés deux phénomènes, il est toujours intéressant de connaître s'il y a ou non une dépendance entre ces phénomènes et dans le cas affirmatif, quelle est la nature de cette dépendance

Analytiquement, ce problème conduit à envisager les deux séries de grandeur statistiques.

$$\begin{array}{l} x_1, x_2, \ldots, x_n \\ y_1, y_2, \ldots, y_n \end{array}$$

qui mesurent les phénomènes donnés et à rechercher une fonction $f(x)$ telle, que l'on ait :

$$y_i = f(x_i). \text{ — (I) } (i = 1, 2, \ldots, n).$$

La fonction $f(x)$ sera alors la traduction analytique de la dépendance des phénomènes donnés.

Or, en général il existe une infinité de fonctions pareilles. Pour s'assurer qu'il en est ainsi il suffit d'interpoler, non pas sur les deux séries données, mais sur les deux séries suivantes :

$$\begin{array}{l} x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_{n+m} \\ y_1, y_2, \ldots, y_n, y_{n+1}, \ldots, y_{n+m} \end{array}$$

(1) Voir numéro d'octobre 1926.

(2) Dans une note présentee à la séance du 2 août 1926 de l'Académie des Sciences de Paris, nous avons résumé les résultats que nous exposons dans cette partie de notre travail.

où les $2m$ quantités $x_{n+1}, \ldots, x_{n+m}, y_{n+1}, \ldots, y_{n+m}$ sont des paramètres choisis arbitrairement. En d'autres termes, si on définit la fonction $f(x)$, par la seule condition de satisfaire au système d'égalités (1), elle est indéterminée. Pourtant, on ne peut nullement conclure de là, qu'étant donnés deux phénomènes, il existe toujours une dépendance entre ces phénomènes et que cette dépendance peut être exprimée analytiquement d'une infinité de manières différentes. Car, il est évident que les fonctions $f(x)$ ainsi déterminées ne sont que des expressions analytiques n'ayant que des rapports très lointains avec les phénomènes considérés.

La question qui se pose est, en définitive, de savoir non pas si entre les x et les y donnés il y a une relation quelconque, car en général il y une infinité de pareilles relations, mais de savoir, si entre ces quantités il y a une dépendance, dont la forme est connue *a priori*. Se donner *a priori*, une forme déterminée pour la loi de dépendance n'est autre chose, qu'exprimer en langage mathématique les différentes hypothèses qu'on admet au début de l'étude de tout phénomène physique ou social. Ainsi, pour prendre un exemple concret, supposer par hypothèse que le taux instantané, de natalité d'une certaine population est proportionnel à la grandeur de cette population, c'est admettre *a priori*, que la grandeur d'une population est susceptible d'être représentée, en fonction du temps, par une expression exponentielle.

§ 2. — Nous nous proposons donc, d'examiner dans cette partie de notre mémoire, le problème de la corrélation, énoncé de la manière suivante :

Étant données deux séries de grandeurs statistiques :

$$\begin{matrix} x_1, x_2, \ldots, x_n \\ y_1, y_2, \ldots, y_n \end{matrix} \qquad \text{(S)}$$

et une loi de dépendance :

$$y = f(x, a_0, a_1, \ldots, a_p) \qquad \text{(L)}$$

où $a_0, a_1, \ldots\ldots, a_p$ sont des constantes arbitraires, — dont le nombre est inférieur à n, — quand peut-on affirmer que les y donnés sont liés aux x donnés, par la loi de dépendance donnée?

Il va sans dire, qu'en général, il sera impossible de déterminer les constantes $a_0, a_1, \ldots\ldots, a_p$ de manière à avoir :

$$y_i = f(x_i, a_0, \ldots, a_p) \; (i = 1, 2, \ldots, n).$$

Par conséquent, la loi de dépendance ne sera, en général, vérifiée que d'une manière approchée. Il y a donc lieu de se demander comment peut-on reconnaître, *a priori*, si l'approximation qu'on est ainsi conduit à envisager est satisfaisante ou non.

Nous nous proposons de montrer que la considération de certains coefficients de corrélation, peut permettre de donner une réponse à cette question.

Le premier chapitre a pour objet la formation de ces coefficients, le second chapitre est destiné à faire voir l'intérêt que présentent certains, parmi ces coefficients pour le problème que nous nous sommes posé.

LES COEFFICIENTS DE CORRÉLATION

§ 3. — Étant données les deux séries de grandeurs statistiques (S) et la loi de dépendance (L), nous appellerons coefficient de corrélation relatif à la loi de dépendance (L) une fonction :

$$r = F(x_1, x_2, \ldots, x_n, y_1, \ldots, y_n)$$

qui remplit les conditions suivantes :

1°. La fonction F est homogène et de degré zéro par rapport aux variables $x_1, x_2, \ldots.., x_n$ d'un côté et $y_1, y_2, \ldots.., y_n$ de l'autre. Ou, autrement dit, la valeur de r reste invariable si on change les unités de mesure des x ou des y;

2° Pour toute valeur réelle des variables, on a :

$$-1 \leqslant r \leqslant +1. -$$

De plus, si les valeurs des grandeurs (S) sont telles qu'il est possible de déter miner les constantes, $a_0, a_1, \ldots.., a_p$ de manière à avoir :

$$y_i = f(x_i, a_0, a_1, \ldots, a_p) \quad (i = 1, 2, \ldots, n)$$

on a aussi :

$$r^2 = 1. -$$

Et, réciproquement.

La définition même des coefficients de corrélation nous suggère comment on peut les former. Écrivons les n relations suivantes :

$$(2) \qquad y_i = f(x_i, a_0, \ldots, a_p) \quad (i = 1, 2, \ldots, n).$$

et éliminons entre ces équations les $p + 1$ constantes arbitraires $a_0, a_1, \ldots.., a_p$. Soient :

$$\left.\begin{array}{l} G_1(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n) = 0 \\ G_2(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n) = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ G_{n-p-1}(x_1, x_2, \ldots x_n, y_1, y_2, \ldots, y_n) = 0 \end{array}\right\} \quad (3)$$

les n-p-1 relations qui expriment le résultat de cette élimination. Ces équations ne sont autre chose que les conditions nécessaires et suffisantes pour que le système (2), dont les inconnues sont $a_0, a_1, \ldots.., a_p$ soit compatible. Si la loi de dépendance, et nous supposons qu'il en est ainsi, ne change pas, pour un changement d'unités, il est évident que les fonctions G sont homogènes.

Pour former le coefficient de corrélation cherché, il suffit de choisir arbitrairement une fonction homogène :

$$H(x_1, x_2, \ldots x_n, y_1, y_2, \ldots y_n)$$

telle que si l'on a :

$$H = 0$$

on ait aussi :

$$G_1 = G_2 = \ldots = G_{n-p-1} = 0$$

et réciproquement. Il est clair qu'il existe une infinité de fonctions H, qui remplissent ces conditions. Pour le montrer, il suffit, de donner à la fonction H la forme suivante :

$$(4) \quad H = \mu_1^2 G_1^2 + \mu_2^2 G_2^2 + \ldots + \mu^2_{n-p-1} G^2_{n-p-1} + K^2(x_1, x_2, \ldots x_n, y_1, y_2, \ldots, y_n)$$

où les μ sont des fonctions homogènes, toujours positives et ne s'annulant pour aucune valeur réelle des variables, sauf, peut-être pour :

$$x_1 = x_2 = \ldots = x_n = y_1 = . \quad y_n = 0. -$$

Ainsi, on peut prendre pour les μ des constantes ou des polynômes homogènes de degré pair, dont tous les coefficients sont positifs et tous les exposants pairs. La fonction K est une fonction homogène, du même degré d'homogénéité que les μ G et qui s'annule si l'on a :

$$G_1 = G_2 = \ldots = G_{n-p-1} = 0. -$$

Il est évident qu'on peut remplacer dans (4) l'exposant 2 des G, par tout autre entier, pair et positif sans que la fonction H cesse de remplir les conditions imposées.

La fonction H, étant choisie, on le mettra sous la forme :

$$H = (M_1 N_1 + M_2 N_2 + \ldots + M_g N_g)^2 - (M_1^2 + \ldots + M_g^2)(N_1^2 + \ldots + N_g^2)$$

les M et les N étant à leur tour, des fonctions homogènes. Ces fonctions doivent être déterminées pour chaque forme de H. Si leur détermination est possible, l'expression :

$$r = \frac{M_1 N_1 + M_2 N_2 + . + M_g N_g}{\sqrt{(M_1^2 + M_2^2 + \ldots + M_g^2)(N_1^2 + N_2^2 + \ldots + N_g^2)}}$$

est un coefficient de corrélation relatif à la loi de dépendance (L).

Il résulte aussi, de la méthode employée, que pour toute loi de dépendance donnée, on peut concevoir une infinité de coefficients de corrélation.

De plus, tout coefficient r, ainsi obtenu, est caractéristique pour la loi (L) correspondante, dans ce sens qu'une même fonction :

$$r = F(x_1, x_2, \ldots x_n, y_1, \ldots, y_n)$$

ne peut pas être à la fois coefficient de corrélation pour deux — et, à plus forte raison pour plus de deux — lois de dépendance différentes :

$$y = f(x, a_0, a_1, \ldots, a_p) \text{ et } y = g(x, a_0, a_1, \ldots, a_g).$$

En effet, pour que la fonction F soit coefficient de corrélation à la fois pour f et pour g, il faudrait que les systèmes (3) relatifs à ces lois ne soient pas distinctes l'un de l'autre, ce qui est impossible.

*
* *

§ 4. — Pour mieux préciser les généralités qui précèdent, considérons le cas particulier, remarquablement intéressant ,où la loi de dépendance est linéaire, soit :

$$f(x, a_0, a_1) = a_0 + a_1 x$$

a_0 et a_1 étant des constantes arbitraires. En désignant par $\Lambda_{\alpha,\beta,\gamma}$ le déterminant :

$$A_{\alpha,\beta,\gamma} = \begin{vmatrix} 1 & 1 & 1 \\ x_\alpha & x_\beta & x_\gamma \\ y_\alpha & y_\beta & y_\gamma \end{vmatrix}$$

les fonctions G sont dans ce cas particulier :

$$G_1 = A_{1,2,3},\ G_2 = A_{2,3,4}, \ldots,\ G_{n-2} = A_{n-2,n-1,n}. —$$

Les coefficients de corrélation de MM. March et Pearson sont des coefficients de corrélation par la loi linéaire, dans le sens que nous venons d'indiquer.

En effet, posons :

$$k = F(x_1, x_2, \ldots x_n, y_1, y_2 \ldots, y_n) = \frac{\sum_{i=1}^{n-1} (x_i - x_{i+1})(y_i - y_{i+1})}{\sqrt{\left[\sum_{i=1}^{n-1} (x_i - x_{i+1})^2\right]\left[\sum_{i=1}^{n-1} (y_i - y_{i+1})^2\right]}}. —$$

k est le coefficient de corrélation introduit par M. March. Il est évident que ce coefficient satisfait à la condition d'homogénéité. De plus, si les x et les y sont tels, qu'il est possible de déterminer les constantes a_0 et a_1, de manière à avoir :

$$y_i = a_0 + a_1 x_i \qquad (i = 1, 2, \ldots, n)$$

on a aussi :

$$k^2 = 1. —$$

Et réciproquement, de cette relation on déduit facilement :

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{y_2 - y_3}{x_2 - x_3} = \ldots = \frac{y_{n-1} - y_n}{x_{n-1} - x_n}. \tag{5}$$

Désignons par M_i le point qui a, par rapport à un système d'axes rectangulaires, les coordonnées x_i et y_i. Le système (5) exprime l'égalité des coefficients angulaires des droites $M_1 M_2$, $M_2 M_3$, ..., $M_{n-1} M_n$. Donc, si l'on a $k^2 = 1$, les points M_1, M_2, ..., M_n sont sur une même droite et par conséquent les y_i sont liés aux x_i par une relation linéaire. Notre affirmation en ce qui concerne le coefficient k se trouve ainsi justifiée.

Il est facile de voir que la fonction H, qui correspond au coefficient k, a bien la forme indiquée par l'égalité (4). Considérons à ce sujet la somme :

$$H = \sum_{\alpha=1}^{n-1} \sum_{\beta=1}^{n-1} [(x_\alpha - x_{\alpha+1})(y_\beta - y_{\beta+1}) - (x_\beta - x_{\beta+1})(y_\alpha - y_{\alpha+1})]^2.$$

Soient :

$$S = \sum_{\sigma=1}^{n-2} [(x_\alpha - x_{\alpha+1})(y_{\alpha+1} - y_{\alpha+2}) - (x_{\alpha+1} - x_{\alpha+2})(y_\alpha - y_{\alpha+1})]^2$$

et K^2 la somme des termes de H qui n'appartiennent pas à S. Il est évident que l'on a :

$$S = \sum_{\alpha=1}^{n-2} A_{\sigma,\sigma+1,\sigma+2} = \sum_{\alpha=1}^{n-2} G_\alpha^2. —$$

Quant à la somme K^2, elle est composée par des termes de la forme :

$$[(x_\alpha - x_{\alpha+1})(y_\beta - y_{\beta+1}) - (x_\beta - x_{\beta+1})(y_\alpha - y_{\alpha+1})]^2$$

où $\beta\text{-}\alpha \geqslant 2$ et on voit aisément, que chacun de ces termes s'annule si l'on a :

$$G_1 = G_2 = \ldots = G_{n-2} = 0. —$$

Les μ_i, qui figurent dans (4) sont tous égaux à l'unité. De même, dans ce cas, les fonctions M_i et N_i sont respectivement :

$$M_i = x_i - x_{i+1}$$
$$N_i = y_i - y_{i+1} \qquad (i = 1, 2, \ldots, n - 1). —$$

Passons maintenant au coefficient r :

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \frac{x_1+x_2+\ldots+x_n}{n}\right)\left(y_i - \frac{y_1+y_2+\ldots+y_n}{n}\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(x_i - \frac{x_1+x_2+\ldots+x_n}{n}\right)^2\right]\left[\sum_{i=1}^{n}\left(y_i - \frac{y_1+y_2+\ldots+y_n}{n}\right)^2\right]}}$$

de M. Pearson. On peut facilement démontrer, par une méthode absolument analogue à celle employée pour le coefficient k, que si les points M_i ($i = 1, 2, \ldots, n$) sont en ligne droite, on a :

$$r^2 = 1$$

et réciproquement.

La fonction H d'où l'on peut déduire r, est :

$$H = \left[\sum_{i=1}^{n}\left(x_i - \frac{x_1+x_2+\ldots+x_n}{n}\right)^2\right]\left[\sum_{i=1}^{n}\left(y_i - \frac{y_1+y_2+\ldots+y_n}{n}\right)^2\right]$$
$$\left[\Sigma\left(x_i - \frac{x_1+x_2+\ldots+x_n}{n}\right)\left(y_i - \frac{y_1+y_2+\ldots+y_n}{n}\right)\right]^2$$

On trouve facilement :

$$H = \frac{1}{6n}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sum_{\gamma=1}^{n} A^2_{\alpha,\beta,\gamma}$$

d'ou l'on conclut que H est encore de la forme indiquée par (4).

Les fonctions M_i et N_i deviennent maintenant :

$$M_i = x_i - \frac{x_1+x_2+\ldots+x_n}{n}$$
$$N_i = y_i - \frac{y_1+y_2+\ldots+y_n}{n} \quad (i = 1, 2, \ldots, n).$$

§ 5. — On peut montrer sans aucune difficulté, que pour la loi de dépendance linéaire, il existe une infinité de coefficients de corrélation. Il suffit pour cela de décomposer les poincs M_i, dans un certain nombre de groupes de points et d'appliquer soit le coefficient k, soit le coefficient r, non pas aux points M_i eux-mêmes, mais à des points du groupe qui jouissent de la propriété que, s'ils sont en ligne droite, tous les points du groupe le sont. Ainsi, on obtient un autre coefficient de corrélation si on détermine le coefficient r, ou k, pour les milieux des côtés des triangles $M_{i-1}\, M_i\, M_{i+1}$, c'est-à-dire si on opère non pas sur les x_i et y_i donnés, mais sur les quantités :

$$\frac{x_i + x_{i+1}}{2}, \frac{x_{i+1} + x_{i-1}}{2}, \frac{x_{i-1} + x_i}{2}$$
$$\frac{y_i + y_{i+1}}{2}, \frac{y_{i+1} + y_{i-1}}{2}, \frac{y_{i-1} + y_i}{2} \quad (i = 2, 3, \ldots, n-1).$$

***

§ 6. — Supposons maintenant, que la loi de dépendance (L) ait une forme telle que des égalités (2), du paragraphe 3, on puisse déduire :

$$g_i\,(x_1, x_2, \ldots x_n, y_1, \ldots y_n) = A_0 + A_1\, h_i\,(x_1, x_2, \ldots, x_n, y_1, \ldots, y_n)\ (i = 1, 2, \ldots, t)$$

et réciproquement. Dans cette dernière égalité les g_i et les h_i sont des fonctions convenablement choisies, tandis que A_0 et A_1 sont des constantes. Sous ces conditions, il est clair que tout coefficient de corrélation relatif à la loi de dépendance linéaire et calculé pour les deux séries de variables :

$$g_i\,(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$$
$$h_i\,(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)\ (i = 1, 2, \ldots, t)$$

est coefficient de corrélation pour la loi de dépendance donnée (L) entre les variables données (S).

Considérons comme application la loi de dépendance parabolique :

$$f\,(x) = a_0 + a_1\,x + a_2\,x^2. -$$

Le système (2) d'égalités devient dans ce cas :

$$(2') \qquad y_i = a_0 + a_1\,x_i + a_2\,x_i^2\ (i = 1, 2, \ldots, n)$$

d'où l'on déduit :

$$\frac{y_i - y_1}{x_i - x_1} = a_1 + a_2\,(x_i + x_1). - (i = 2, \ldots, n)\cdot$$

Réciproquement, si les x et les y donnés, sont tels, qu'il est possible de déterminer les constantes a_1 et a_2 de manière que ce dernier système d'égalités soit satisfait, on peut déterminer a_0 de manière que le système (2') le soit aussi. Donc, tout coefficient de corrélation relatif à la loi linéaire, calculé pour les grandeurs :

$$x_2 + x_1,\ x_3 + x_1, \ldots, x_n + x_1$$
$$\frac{y_2 - y_1}{x_2 - x_1}, \frac{y_3 - y_1}{x_3 - x_1}, \ldots, \frac{y_n - y_1}{x_n - x_1}$$

est coefficient de corrélation relatif à la loi de dépendance parabolique entre les grandeurs données :

$$x_1, x_2, \ldots, x_n$$
$$y_1, y_2, \ldots, y_n. -$$

§ 7. — Un artifice absolument analogue réussit si on l'applique à la loi de Makeham. Nous rappelons qu'en vertu de cette loi, le nombre des survivants d'âge x serait :

$$v_x = k\,s^x.\,g^{c^x}$$

k, s, g et c étant des constantes, ou encore :

$$v_x = e^{\alpha + \beta\,x + \gamma\,e^{x \log. c}}$$

α, β, γ, et c étant aussi des constantes. Supposons connus certains nombres de survivants :

$$V_{x\,1}, V_{x,\,2}, \ldots, V_{x,\,n}$$

d'âges respectivement :

$$x_1, x_2, \ldots, x_n. -$$

Posons :

$$y_i = \log V_{x\,i} \qquad (i = 1, 2 \ldots n)$$

On sait, qu'étant donnés les nombres $V_{x,i}$, on peut facilement en déduire la valeur de c (Voir, Poterin du Motel, *Théorie des assurances sur la vie*, page 141). Donc, on a :

$$\log c = \varphi (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n). —$$

On conclut de là, que tout coefficient corrélation relatif à la loi linéaire, calculé pour les quantités :

$$\frac{e^{x_2 \varphi (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)} - e^{x_1 \varphi}}{x_2 - x_1}, \frac{e^{x_3 \varphi} - e^{x_1 \varphi}}{x_3 - x_1}, \ldots, \frac{e^{x_n \varphi} - e^{x_1 \varphi}}{x_n - x_1}$$

$$\frac{y_2 - y_1}{x_2 - x_1}, \frac{y_3 - y_1}{x_3 - x_1}, \ldots, \frac{y_n - y_1}{x_n - x_1}$$

est coefficient de corrélation, relatif à la loi de Makeham, pour les quantités :

$$x_1, x_2, \ldots, x_n$$
$$V_{x,1}, V_{x,2}, V_{x,n}. —$$

LE PROBLÈME DE LA CORRÉLATION

§ 8. — Si théoriquement on peut envisager, pour chaque loi de dépendance, donnée *a priori*, une infinité de coefficients de corrélation, il est certain que pratiquement tous ces coefficients ne présentent pas le même intérêt. Dans les applications, on choisira donc, celui ou ceux des coefficients dont l'emploi est imposé par les circonstances spéciales du problème qu'on étudie.

Pourtant, il y a pour chaque loi de dépendance certains coefficients de corrélation, qui méritent une attention spéciale. Ce sont les coefficients susceptibles de montrer l'approximation avec laquelle est réalisée la dépendance entre les grandeurs données. Cette dernière proposition doit, elle aussi, être précisée. Pour le faire et pour éviter des complications d'écriture, nous allons nous borner à un cas particulier, d'ailleurs assez étendu, du problème que nous nous sommes posé au paragraphe 2. La méthode et les conclusions que nous obtiendrons, sur ce cas particulier, pourront du reste être utilisées dans tout autre problème de la même nature.

Supposons, qu'étant données les deux séries de grandeurs :

$$x_1, x_2, \ldots, x_n$$
$$y_1, y_2, \ldots, y_n$$

l'on veuille savoir s'il existe ou non entre ces grandeurs une dépendance de la forme :

$$y = a_0 + a_1 x + \ldots + a_p x^p = f(x)^{(p)}.$$

Pour résoudre ce problème on emploie d'habitude la méthode des moindres carrés. On détermine ainsi les constantes a_0, a_1, ..., a_p de manière à rendre minimum l'expression suivante :

$$E_1 = \sum_{i=1}^{n} [y_i - f(x_i)]^2. —$$

Mais, il est évident qu'on peut procéder différemment. On n'a pour cela qu'à remplacer l'expression E_1, par :

$$E_k = \sum_{i=1}^{n} [y_i - f(x_i)]^{2k}$$

k étant un entier positif, arbitraire et à déterminer les constantes $a_0, a_1, \ldots, a_p$ de manière que cette dernière expression soit minimum.

Nous remarquerons maintenant que, en général, cette détermination est possible d'une seule manière. En effet, l'expression E_k est positive, pour toute valeur réelle des a. Elle admet donc un minimum positif ou nul. Par conséquent, le système d'équations :

$$\frac{\partial E_k}{\partial a_j} = 0 \quad (6) \qquad (j = 0, 1, \ldots, p)$$

dont les inconnues sont $a_0, a_1, \ldots, a_p$ est vérifié, pour un système — au moins — de valeurs réelles :

$$\alpha_0, \alpha_1, \ldots, \alpha_p. —$$

Soit $e_{k,p}$ la valeur de E_k correspondant à ce système de solutions des équations (6). En général, il n'existe pas un second système de valeurs réelles :

$$\beta_0, \beta_1, \ldots, \beta_p$$

vérifiant le système (6). Car, si on désigne par e_β la valeur correspondante de E_k, on devrait avoir — comme il est facile de le démontrer, en faisant intervenir les dériéves du second ordre — à la fois :

$$e_\beta - e_{k,p} > 0 \text{ et } e_{k,p} - e_\beta > 0. —$$

Il résulte, en définitive, que $e_{k,p}$ est le minimum unique de E_k.

§ 9 — Il est à remarquer qu'on peut calculer $e_{k,p}$ sans avoir, au préalable, déterminé effectivement les quantités $\alpha_0, \alpha_1, \ldots, \alpha_p$. Posons, pour cela :

$$f_1(x) = \alpha_0 + \alpha_1 x + \ldots + \alpha_p x^p. —$$

On a :

$$e_{k,p} = \sum_{i=1}^{n} y_i [y_i - f_1(x_i)]^{2k-1}$$

Donc, pour avoir $e_{k,p}$ il suffit d'éliminer $a_0, a_1, \ldots, a_p$ entre les $p + 2$, équations algébriques fournies par le système (6) et l'équation :

$$e_{k,p} = \sum_{i=1}^{n} y_i [y_i - f(x_i)]^{2k-1}$$

Le résultat de cette élimination sera de la forme :

$$\Phi(e_{k,p}, x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n) = 0. —$$

Étant donnée la signification de $e_{k,p}$ il est clair que l'équation, en $e_{k,p}$:

$$\Phi = 0$$

aura, en général, une seule racine réelle et positive $e_{k,p}$.

§ 10. — De toute expression E_k on peut déduire un coefficient de corrélation relatif à la loi parabolique (P).

Il est évident que la valeur de $e_{k,p}$ est inférieure ou au plus égale à la valeur qu'on obtient, pour E_k, en donnant aux constantes a_0, a_1 ,....., a_p des valeurs arbitraires. En particulier, en faisant dans E_k :

$$a_0 = \gamma_k \qquad a_1 = a_2 = \ldots = a_p = 0$$

on a :

$$e_{k,p} \leqslant \sum_{i=1}^{n} (y_i - \gamma_k)^{2k} = \sum_{i=1}^{n} y_i \, (y_i - \gamma_k)^{2k-1}$$

où γ_k représente la racine réelle de l'équation :

$$\sum_{i=1}^{n} (y_i - \gamma_k)^{2k-1} = 0. —$$

Cette équation n'admet d'ailleurs qu'une seule racine réelle, comme on le démontre aisément, en observant que la dérivée de son premier membre, par rapport à γ_k, garde un signe constant. De plus, si l'on a :

$$\alpha_0 = \gamma_k \quad \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$$

où nous rappelons que les α sont les solutions réelles du système (6) — l'on a aussi :

$$e_{k,p} = \sum_{i=1}^{n} (y_i - \gamma_k)^{2k} . —$$

Et réciproquement. En définitive, on obtient que si l'on pose :

$$e_{k,p} = \sum_{i=1}^{n} (y_i - \gamma_k)^{2k} - z_{k,p}$$

la quantité $Z_{k,p}$ est positive ou nulle.

Le rapport :

$$r_{k,p} = \sqrt{\frac{z_{k,p}}{\sum_{i=1}^{n} (y_i - \gamma_k)^{2k}}}$$

est un coefficient de corrélation relatif à la loi de dépendance parabolique (P).

On remarque d'abord, qu'il satisfait aux conditions d'homogénéité. En effet, si on multiplie tous les x par un même facteur λ, le système (6) se transforme en lui-même, à condition d'y remplacer les a_0, a_1, ..., a_p respectivement par a_0, $\frac{a_1}{\lambda}$, ..., $\frac{a_p}{\lambda^p}$. Le minimum $e_{k,p}$ et par conséquent $Z_{k,p}$ sont des fonctions homogènes et de degré zéro par rapport aux x. D'autre part, si l'on multiplie les y et les a par un même facteur, le système (6) ne change pas. Il s'ensuit que $e_{k,p}$ et $Z_{k,p}$ sont des fonctions homogènes de degré $2\,k$, par rapport aux y.

De plus, si l'on a $r^2_{k,p} = 1$, on a aussi $e_{k,p} = 0$, donc :

$$y_i = f_1(x_i) = \alpha_0 + \alpha_1 x_i + \ldots + \alpha_p x_i^p \; (i = 1, 2, \ldots, n)$$

et réciproquement. Donc, tous les $r_{k,p}$ sont des coefficients de corrélation pour la loi parabolique (P).

§ 11. — Dans le cas particulier, où $k = 1$ et $p = 1$ le coefficient $r_{k,p}$ n'est autre chose que le coefficient r, de M. Pearson.

De même, on peut aussi rattacher le coefficient k, de M. March, à un problème de minimum. Afin d'éviter toute ambiguité, dans les notations, nous désignerons par ρ ce coefficient, dont il a été question au paragraphe 4.

Déterminons à cet effet α de manière que l'expression :

$$E = \sum_{i=1}^{n-1} [y_i - y_{i+1} - \alpha (x_i - x_{i+1})]^2$$

soit minimum. On trouve :

$$\alpha = \frac{\sum_{i=1}^{n-1} (x_i - x_{i+1}) (y_i - y_{i+1})}{\sum_{i=1}^{n-1} (x_i - x_{i+1})^2} = \rho \sqrt{\frac{\sum_{i=1}^{n-1} (y_i - y_{i+1})^2}{\sum_{i=1}^{n-1} (x_i - x_{i+1})^2}}$$

La valeur du minimum de E correspondant, à cette valeur de α, est :

$$E_{min} = \sum_{i=1}^{n-1} (y_i - y_{i+1}) [y_i - y_{i+1} - \alpha (x_i - x_{i+1})]$$

ou encore :

$$E_{min} = (1 - \rho^2) \sum_{i=1}^{n-1} (y_i - y_{i+1})^2. —$$

§ 12. — Il n'est peut-être pas sans intérêt de chercher l'interprétation qu'on peut donner à l'égalité $r_{k,p} = 0$.

D'après la théorie classique de la corrélation, on déduit de l'égalité $r = r_{1,1} = 0$ qu'il n'y a pas de corrélation entre les phénomènes mesurés par les grandeurs x_i et y_i, ou encore, qu'il y a indépendance entre ces phénomènes. Cette conclusion est traduite analytiquement par le fait que, si l'on cherche, conformément à la méthode des moindres carrés les valeurs de a_0 et a_1 qui rendent minimum l'expression :

$$\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2,$$

on trouve pour $r = 0 : a_1 = 0$. La méthode des moindres carrés conduit donc, dans ce cas, à attribuer aux y une valeur constante : $a_0 = m_y$, m_y étant la moyenne des y.

Le même fait a lieu en ce qui concerne les coefficients $r_{k,p}$. S'il y a indépendance entre les x et les y donnés, ce qui analytiquement se traduit par les équations :

$$\alpha_0 = \gamma_k \quad \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0,$$

on a aussi :

$$r_{k,p} = 0.$$

Et réciproquement, de l'égalité :

$$r_{k,p} = 0$$

on en déduit :

$$\alpha_0 = \gamma_k \quad \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0.$$

Cela tient au fait, que le système (6) n'admet qu'un seul système de solutions réelles.

§ 13. — Nous rappelons maintenant que les coefficients $r_{k,p}$ ont été définis de manière à avoir :

$$e_{k,p} = (1 - r^2_{k,p}) \Sigma (y_i - \gamma_k)^{2k}.$$

D'autre part, la valeur moyenne des différences

$$y_i - \alpha_0 - \alpha_1 x_i - \dots - \alpha_p x_i^p \ (i = 1, 2, \dots, n)$$

est de l'ordre de :

$$\sqrt[2k]{\frac{e_{k,p}}{n}}.$$

Admettons donc, par hypothèse, ce qui est assez légitime, que l'expression :

$$\varepsilon_{k,p} = \sqrt[2k]{\frac{e_{k,p}}{n}}$$

mesure l'approximation moyenne avec laquelle les x et les y vérifient la loi de dépendance (P). On voit que la détermination de $\varepsilon_{k,p}$ et de $r_{k,p}$ sont deux problèmes équivalents. Ainsi, de la valeur de $\varepsilon_{k,p}$ on peut tirer $r_{k,p}$ et réciproquement. On voit aussi, que dans ces questions, la somme :

$$\sum_{i=1}^{n} (y_i - \gamma_k)^{2k} \tag{7}$$

joue un rôle remarquable. L'importance de cette expression est justifiée, du reste, par le fait qu'elle constitue la généralisation pour l'exposant $k > 1$ de la fluctuation. On obtient cette dernière en faisant dans (7) $k = 1$.

* * *

§ 14. — Malgré l'intérêt que les différentes expressions E_k, peuvent présenter, leur introduction dans les applications soulève des difficultés dont les développements qui précèdent laissent entrevoir l'étendue. Pour les éviter, il est donc préférable de se borner dans les applications, à l'emploi de la méthode des moindres carrés. Cette méthode est imposée, sinon par des considérations qui découlent de la loi de Gauss, du moins, et ce n'est pas peu dire, par la simplicité des calculs qu'elle exige. Ces calculs peuvent être rendus encore plus simples si on emploie la méthode d'interpolation de Tchebycheff sous la forme qui a été indiquée récemment par M. Romanowsky (Voir Romanowsky, « Sur une méthode d'interpolation de Tchebycheff », *Comptes rendus de l'Académie des Sciences de Paris*, page 595, tome 181, 1925).

A cause de l'importance que présente cette méthode et afin de nous en servir pour expliciter la valeur du coefficient $r_{1,p}$, nous allons la reproduire rapidement.

Nous commencerons par orthogonaliser, sur la base (x_i), la suite des fonctions :

$$1, x, \dots, x^p$$

à l'aide des polynômes :

$$\varphi_0 (x), \varphi_1 (x), \dots, \varphi_p (x).$$

Posons :

$$(u) = \sum_{i=1}^{n} u(x_i)$$

et soit :

$$\varphi_o(x) = 1 \ (1).$$

On prend pour $\varphi_1(x)$ un polynôme de la forme :

$$\varphi_1(x) = a\,\varphi_o(x) + x$$

où a est déterminé par la condition :

$$(\varphi_o\,\varphi_1) = a\,(\varphi_o\,\varphi_o) + (x\,\varphi_o) = 0$$

ou encore :

$$\varphi_1(x) = -\frac{(x\,\varphi_o)}{(\varphi_o\,\varphi_o)}\,\varphi_o(x) + x.$$

On définit de la même manière $\varphi_2(x)$:

$$\varphi_2(x) = a\,\varphi_o(x) + h\,\varphi_1(x) + x^2$$

où a et b sont choisis de telle sorte que l'on ait :

$$(\varphi_2\,\varphi_o) = (\varphi_2\,\varphi_1) = 0.$$

On obtient :

$$\varphi_2(x) = -\frac{(x^2\,\varphi_o)}{(\varphi_o\,\varphi_o)}\,\varphi_o(x) - \frac{(x^2\,\varphi_1)}{(\varphi_1\,\varphi_1)}\,\varphi_1(x) + x^2$$

et en général :

$$\varphi_k(x) = -\sum_{i=0}^{k-1} \frac{(x^k\,\varphi_i)}{(\varphi_i\,\varphi_i)}\,\varphi_i(x) + x^k.$$

Le polynôme d'interpolation :

$$f(x) = a_o + a_1 x + \dots + a_p x^p$$

devient avec les nouvelles notations :

$$f(x) = A_o\,\varphi_o(x) + A_1\,\varphi_1(x) + \dots + A_p\,\varphi_p(x)$$

A_0, A_1, ..., A_p étant des constantes. On détermine leurs valeurs de manière à rendre minimum l'expression :

$$E_1 = \sum_{j=1}^{n} \left[y_j - \sum_{i=1}^{p} A_i\,\varphi_i(x_j) \right]^2$$

On obtient :

$$A_k = \frac{(y\,\varphi_k)}{(\varphi_k\,\varphi_k)}.$$

La valeur correspondante $e_{1,p}$ du minimum de E_1 est :

$$e_{1,p} = (yy) - \sum_{i=0}^{p} A_i\,(y\,\varphi_i).$$

---

(1) Pour des raisons de symétrie, $\varphi_o(x)$ continuera à figurer dans ce qui suit.

Le coefficient de corrélation $r_{1,p}$ devient maintenant :

$$r_{1,p} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - m_y)^2 - (y\,y) + \sum_{1}^{p} A\,(y\,\varphi_i)}{\Sigma\,(y_i - m_y)^2}}$$

$$r_{1,p} = \sqrt{\frac{\sum_{i=1}^{p} A_i\,(y\,\varphi_i)}{\sum_{i=1}^{n} (y_i - m_y)^2}}$$

m_y désignant comme d'habitude la moyenne des y. On a ,de même :

$$e_{1,p} = (1 - r_{1,p}^2) \sum_{i=1}^{n} (y_i - m_y)^2.$$

L'erreur moyenne devient maintenant :

$$\varepsilon_{1,p} = \sqrt{1 - r_{1,p}^2}\cdot \sigma_y$$

σ_y désignant la déviation type des y.

§ 15. — On voit ainsi que la valeur de la quantité $\varepsilon_{1,p}$, quantité qui est susceptible de mesurer la corrélation, par rapport à la loi parabolique, dépend à la fois du coefficient de corrélation et de la déviation type. Donc, une étude de la corrélation qui se bornerait à considérer seulement le coefficient de corrélation, indépendamment de la variabilité serait sinon erronée, du moins incomplète.

Les quantités $\frac{\varepsilon_{1,p}}{m_y}$ sont comparables entre elles, dans ce sens que, si l'on considère deux ou plusieurs couples de séries statistiques de la valeur plus ou moins grande du rapport $\frac{\varepsilon_{1,p}}{m_y}$ correspondant à chacun de ces couples, on peut déduire qu'il y a moins ou plus de corrélation, par rapport à la loi parabolique, entre les deux séries qui constituent ce couple.

La valeur de $\varepsilon_{1,p}$ étant déterminée, on peut répondre à la question que nous nous sommes posée plus haut, savoir s'il existe ou non, une dépendance de la forme (P) entre les x_i et les y_i donnés. On dira qu'il existe une pareille dépendance s'il est admissible de commettre dans l'évaluation des y une erreur relative de l'ordre $\frac{\varepsilon_{1,p}}{m_y}$ sans que le phénomène auquel ces quantités se rattachent en soit modifié. Ainsi, pour les phénomènes physiques on déduira une dépendance (P), toutes les fois que le rapport $\frac{\varepsilon_{1,p}}{m_y}$ sera inférieur à l'erreur des instruments d'observation ; pour les phénomènes économiques, il suffira, en général, que ce rapport soit de l'ordre des dixièmes, etc.

§ 16. — Considérons, comme application les trois séries statistiques suivantes :

Circulation fiduciaire (janvier 1922-décembre 1925).
Indice du dollar (janvier 1922-décembre 1925).
Indice du coût de la vie (janvier 1922-décembre 1925).

que nous avons précédemment reproduites. Multiplions d'abord chaque terme de ces séries par :

$$\frac{100}{M}$$

M étant la moyenne de la série à laquelle appartient le terme multiplié. De cette manière, la moyenne de chacune des nouvelles séries obtenues devient 100. Les séries ainsi transformées présentent un intérêt particulier, au point de vue statistique. Les différents calculs se simplifient et les coefficients de corrélation, calculés pour les nouvelles séries sont les mêmes que ceux relatifs aux séries données. Nous désignerons dans la suite par x la nouvelle série qui correspond à la circulation fiduciaire, et par y celle qui correspond à l'indice du coût de la vie.

Le détail des calculs est indiqué dans le tableau qui se trouve à la fin du mémoire.

On trouve :

Coefficients k (M. March)

Circulation fiduciaire — indice du dollar $k_{1,2} = \dfrac{27}{\sqrt{164 \cdot 1535}} = 0{,}053.$

Indice du dollar — indice du coût de la vie $k_{2,3} = \dfrac{-294}{\sqrt{1535 \cdot 849}} = -0{,}026.$

Indice du coût de la vie. Circulation fiduciaire $k_{3,1} = \dfrac{17}{\sqrt{164 \cdot 849}} = 0{,}046.$

Coefficients de variabilité.

Circulation fiduciaire $\eta_x = 12{,}2$
Indice du dollar $= 13{,}9$
Indice du coût de la vie $\eta_y = 23{,}9$

Coefficients de corrélation r (M. Pearson)

Circulation fiduciaire — indice du dollar $r_{1,2} = \dfrac{6092}{\sqrt{7126 \cdot 9322}} = 0.7474.$

Indice du dollar — indice du coût de la vie $r_{2,3} = \dfrac{11880}{\sqrt{9322 \cdot 27599}} = 0{,}7407.$

indice du coût de la vie — circulation fiduciaire $r_{3,1} = \dfrac{13438}{\sqrt{7126 \cdot 27599}} = 0{,}9580.$

Il est à remarquer que les coefficients k et r relatifs à la circulation fiduciaire et au coût de la vie ont des valeurs plus grandes que celles des autres coefficients. De plus, la valeur $\varepsilon_{1,1}$ correspondante à ces deux séries est :

$$\varepsilon_{1,1} = \sqrt{1 - 0{,}958^2} \,.\, 24 = 6{,}859 \backsim 7.$$

Le rapport $\dfrac{\varepsilon_{1,1}}{100}$ étant de l'ordre des dixièmes, on déduit qu'il existe une dépendance linéaire entre la circulation fiduciaire et le prix de la vie. Pour déterminer cette dépendance, employons la méthode d'interpolation parabolique de Tchebycheff. On trouve :

$$\varphi_1(x) = -\frac{(x\varphi_0)}{(\varphi_0\varphi_0)}\varphi_0'(x) + x = x - m_x = x - 100$$

$$A_0 = \frac{(y\varphi_0)}{(\varphi_0\varphi_0)} = m_y = 100 \quad A_1 = \frac{(y\varphi_1)}{(\varphi_1\varphi_1)} = 1{,}886$$

$$y = A_0\varphi_0(x) + A_1 y_1(x) = 100 + 1{,}886\,(x - 100)$$

$$y = 1{,}886\,x - 88{,}6.$$

CIRCULATION FIDUCIAIRE							INDICES DU DOLLAR				
V_1	δ_1	δ_1^2	Δ_1	Δ_1^2	Δ_1^3	Δ_1^4	V_2	δ_2	δ_2^2	Δ_2	Δ_2^2
1922											
79	+ 1	1	— 21	441	— 9.261	194.481	69	+ 1	1	— 31	961
78	— 2	4	— 22	484	— 10.648	234.256	68	— 3	9	— 32	1.024
80	— 1	1	— 20	400	— 8.000	160.000	71	— 1	1	— 29	841
81	— 1	1	— 19	361	— 6.859	130.321	72	— 2	4	— 28	784
82	0	0	— 18	324	— 5.832	104.976	74	— 5	25	— 26	676
82	— 1	1	— 18	324	— 5.832	104.976	79	— 8	64	— 21	441
83	— 1	1	— 17	289	— 4.913	83.521	87	+ 6	36	— 13	169
84	— 4	16	— 16	256	— 4.096	65.536	81	0	0	— 19	361
88	— 1	1	— 12	144	— 1.728	20.736	81	— 3	9	— 19	361
89	0	0	— 11	121	— 1.331	14.641	84	0	0	— 16	256
89	+ 1	1	— 11	121	— 1.331	14.641	84	— 3	9	— 16	256
88	0	0	— 12	144	— 1.728	20.736	87	— 19	361	— 13	169
1.003		27		3.409	— 61.560	1.148.821	937		519		6.299
1923											
88	— 1	1	— 12	144	— 1.728	20.736	106	— 3	9	+ 6	36
89	— 1	1	— 11	121	— 1.331	14.641	109	+ 5	25	+ 9	81
90	— 1	1	— 10	100	— 1.000	10.000	104	— 7	49	+ 4	16
91	— 1	1	— 9	81	— 729	6.561	111	+ 7	49	+ 11	121
92	— 4	16	— 8	64	— 512	4.096	104	+ 1	1	+ 4	16
94	— 2	4	— 6	36	— 216	1.296	103	— 1	1	+ 3	9
96	— 2	4	— 4	16	— 64	256	104	— 10	100	+ 4	16
98	— 3	9	— 2	4	— 8	16	114	+ 3	9	+ 14	196
101	— 0	0	+ 1	1	+ 1	1	111	— 3	9	+ 11	121
101	— 3	9	+ 1	1	+ 1	1	114	+ 3	9	+ 14	196
104	— 0	0	+ 4	16	+ 64	256	111	+ 9	81	+ 11	121
104	+ 2	4	+ 4	16	+ 64	256	102	— 5	25	+ 2	4
1.148		50		600	— 5.458	58.016	1.293		367		933
1924											
102	— 1	1	+ 2	4	+ 8	16	107	+ 6	36	+ 7	49
103	0	0	+ 3	9	+ 27	81	101	0	0	+ 1	1
103	+ 1	1	+ 3	9	+ 27	81	101	0	0	+ 1	1
102	0	0	+ 2	4	+ 8	16	101	— 6	36	+ 1	1
102	— 6	36	+ 2	4	+ 8	16	107	— 11	121	+ 7	49
108	+ 3	9	+ 8	64	+ 512	4.096	118	0	0	+ 18	324
105	— 3	9	+ 5	25	+ 125	625	118	+ 4	16	+ 18	324
108	— 1	1	+ 8	64	+ 512	4.096	114	+ 13	169	+ 14	196
109	— 2	4	+ 9	81	+ 729	6.561	101	+ 4	16	+ 1	1
111	0	0	+ 11	121	+ 1.331	14.641	97	0	0	— 3	9
111	— 1	1	+ 11	121	+ 1.331	14.641	97	— 10	100	— 3	9
112	+ 4	1	+ 12	144	+ 1.728	20.736	107	+ 6	36	+ 7	49
1.276		63		650	6.346	65.606	1.269		494		1.013
1925											
111	0	0	+ 11	121	+ 1.331	14.641	101	— 2	4	+ 1	1
111	+ 1	1	+ 11	121	+ 1.331	14.641	103	— 4	16	+ 3	9
110	0	0	+ 10	100	+ 1.000	10.000	107	— 7	49	+ 7	49
110	— 1	1	+ 10	100	+ 1.000	10.000	114	+ 3	9	+ 14	196
111	— 1	1	+ 11	121	+ 1.331	14.641	111	0	0	+ 11	121
112	— 1	1	+ 12	144	+ 1.728	20.736	111	+ 4	16	+ 11	121
113	— 3	9	+ 13	169	+ 2.197	28.561	107	+ 4	16	+ 7	49
116	— 3	9	+ 16	256	+ 4.096	65.536	103	— 4	16	+ 3	9
119	0	0	+ 19	361	+ 6.859	130.321	107	— 2	4	+ 7	49
119	+ 1	1	+ 19	361	+ 6.859	130.321	109	— 5	25	+ 9	81
118	+ 1	1	+ 18	324	+ 5.832	104.976	114	0	0	+ 14	196
117			+ 17	289	+ 4.913	83.521	114			+ 14	196
1.367		24		2.467	38.477	627.895	1.301		155		1.077
TOTAL GÉNÉRAL . . .		164		7.126	— 22.195	1.900.338	4.800		1.535		9.232

Sur ce tableau les V_1, V_2, V_3 représentent les valeurs respectivement de la circulation fiduciaire, de l'indice des V_1, V_2, V_3 soit 100. —

δ_i $(i = 1, 2, 3)$ désigne la différence des deux valeurs consécutives des V_i.

Δ_i désigne l'écart à la moyenne.

A LA PÉRIODE 1922-1925

INDICES DU COUT DE LA VIE					DÉTERMINATION DES k			DÉTERMINATION DES r			Determination de φ^2
V_3	δ_3	δ_3^2	Δ_3	Δ_3^2	$\delta_1\ \delta_2$	$\delta_2\ \delta_3$	$\delta_3\ \delta_1$	$\Delta_1\ \Delta_2$	$\Delta_2\ \Delta_3$	$\Delta_3\ \Delta_1$	$(\Delta_1^2,\ \Delta_3)$
64	0	0	— 36	1.296	+ 1	0	0	+ 651	+ 1.116	+ 756	— 15.876
64	0	0	— 36	1.296	+ 6	0	0	+ 704	+ 1.152	+ 792	+ 17.434
64	0	0	— 36	1.296	+ 1	0	0	+ 580	+ 1.044	+ 720	— 14.400
64	0	0	— 36	1.296	+ 2	0	0	+ 532	+ 1.008	+ 684	— 12.996
64	+ 1	1	— 36	1.296	0	— 5	0	+ 468	+ 936	+ 648	— 11.664
63	— 1	1	— 37	1.369	+ 8	+ 8	+ 1	+ 378	+ 777	+ 666	— 11.988
64	+ 1	1	— 36	1.296	— 6	+ 6	— 1	+ 221	+ 468	+ 612	— 10.644
63	— 3	9	— 37	1.369	0	0	+ 12	+ 304	+ 703	+ 592	— 9.472
66	— 3	9	— 34	1.156	+ 3	+ 9	+ 3	+ 228	+ 646	+ 408	— 4.896
69	— 6	36	— 31	961	0	0	0	+ 176	+ 496	+ 341	— 3.751
75	— 4	16	— 25	625	— 3	+ 9	— 4	+ 176	+ 400	+ 275	— 3.025
79	— 2	14	— 21	441	0	0	0	+ 142	+ 273	+ 252	— 3.024
802		77		13.647	+ 12	+ 27	+ 11	+ 4.560	+ 9.019	+ 6.746	—119.162
81	— 4	16	— 19	361	+ 3	+ 12	+ 4	— 72	— 114	+ 228	— 2.736
85	— 7	49	— 15	225	— 5	— 35	+ 7	— 99	— 135	+ 165	— 1.815
92	+ 10	100	— 8	64	+ 7	— 70	— 10	— 40	— 32	+ 80	— 800
82	— 15	225	— 18	324	— 7	—105	+ 15	— 99	— 198	+ 162	— 1.458
97	+ 2	4	— 3	9	— 4	+ 2	— 8	— 32	— 12	+ 24	— 192
95	+ 1	1	— 5	25	+ 2	— 1	— 2	— 18	— 15	+ 30	— 180
94	+ 2	4	— 6	36	+ 20	— 20	— 4	— 16	— 24	+ 24	— 96
92	— 3	9	— 8	64	— 9	— 9	+ 9	— 28	— 112	+ 16	— 32
95	— 5	25	— 5	25	0	+ 15	0	+ 11	— 55	— 5	— 5
100	— 1	1	0	0	— 9	— 3	+ 3	+ 14	0	0	0
101	0	0	+ 1	1	0	0	0	+ 44	+ 11	+ 4	+ 16
101	— 4	16	+ 1	1	— 10	+ 20	— 8	+ 8	+ 2	+ 4	+ 16
1.115		450		1.135	— 12	—194	+ 6	— 327	— 684	+ 732	— 7.282
105	— 2	4	+ 5	25	— 6	— 12	+ 2	+ 14	+ 35	+ 10	+ 20
107	+ 1	1	+ 7	49	0	0	0	+ 3	+ 7	+ 21	+ 63
106	+ 3	9	+ 6	36	0	0	+ 3	+ 3	+ 6	+ 18	+ 54
103	— 1	1	+ 3	9	0	+ 6	0	+ 2	+ 3	+ 6	+ 12
104	+ 1	1	+ 4	16	+ 66	— 11	— 6	+ 14	+ 28	+ 8	+ 16
103	— 1	1	+ 3	9	0	0	— 3	+ 144	+ 54	+ 24	+ 192
104	— 7	4 9	+ 4	16	— 12	— 28	+ 21	+ 90	+ 72	+ 20	+ 100
111	— 2	4	+ 11	121	— 13	— 26	+ 2	+ 112	+ 154	+ 88	+ 704
113	— 3	9	+ 13	169	— 8	— 12	+ 6	+ 9	+ 13	+ 117	+ 1.053
116	— 2	4	+ 16	256	0	0	0	— 33	— 48	+ 176	+ 1.936
118	— 3	9	+ 18	324	+ 10	+ 30	+ 3	— 33	— 54	+ 198	+ 2.178
121	— 5	25	+ 21	441	+ 6	— 30	— 5	+ 84	+ 147	+ 252	+ 3.031
1.311		117		1.361	+ 43	— 83	+ 23	+ 419	+ 417	+ 938	+ 9.359
126	— 6	36	+ 26	676	0	+ 12	0	+ 11	+ 26	+ 286	+ 3.146
132	— 3	9	+ 32	1.024	— 4	+ 12	— 3	+ 33	+ 96	+ 352	+ 3.872
135	+ 2	4	+ 35	1.225	0	— 14	0	+ 70	+ 245	+ 350	+ 3.500
133	— 3	9	+ 33	1.089	— 3	— 9	+ 3	+ 140	+ 462	+ 330	+ 3.300
136	+ 1	1	+ 36	1.296	0	0	— 1	+ 121	+ 396	+ 396	+ 4.356
135	+ 6	36	+ 35	1.225	— 4	+ 24	— 6	+ 132	+ 385	+ 420	+ 5.040
129	— 4	16	+ 29	841	— 12	— 16	+ 12	+ 91	+ 203	+ 377	+ 4.901
133	+ 8	64	+ 33	1.089	+ 12	— 32	— 24	+ 48	+ 99	+ 528	+ 8.448
125	— 2	4	+ 25	625	0	+ 4	0	+ 133	+ 175	+ 475	+ 9.025
127	+ 1	1	+ 27	729	— 5	— 5	— 1	+ 171	+ 243	+ 513	+ 9.747
126	— 5	25	+ 26	676	0	0	— 5	+ 252	+ 364	+ 468	+ 8.424
131			+ 31	961				+ 238	+ 434	+ 527	+ 8.959
1.568		205		11.456	— 16	— 24	— 23	+ 1.440	+ 3.128	+ 5.022	+ 72.718
4.796		849		27.599	+ 27	—294	+ 17	+ 6.092	+ 11.880	+ 13.438	— 44.367

du dollar et de l'indice du coût de la vie, multipliées par des facteurs convenablement choisis, de manière que la moyenne

Voici maintenant les valeurs qu'on obtient pour la série y en interpolant d'après cette dernière égalité :

ANNÉES	VALEURS données	VALEURS interpolées	ANNÉES	VALEURS donnés	VALEURS interpolées
1922 :			1924 :		
Janvier.	64	61,4	Janvier.	105	104,5
Février.	64	60,6	Février.	107	105,9
Mars.	64	62,7	Mars.	106	105,8
Avril.	64	63,8	Avril.	103	103,9
Mai.	64	65,7	Mai.	104	104,9
Juin	63	65,8	Juin	103	115,2
Juillet	64	67,1	Juillet	104	109,1
Août.	63	70,5	Août.	111	115,2
Septembre.	66	76,9	Septembre.	113	117,5
Octobre.	69	79,1	Octobre.	116	120,7
Novembre.	75	79,0	Novembre.	118	121,0
Décembre.	79	76,5	Décembre.	121	122,9
1923 :			1925 :		
Janvier.	81	78,0	Janvier.	126	120,1
Février.	85	79,4	Février.	132	119,8
Mars.	92	80,3	Mars.	135	119,6
Avril.	82	82,9	Avril.	133	119,6
Mai.	87	84,6	Mai.	136	119,8
Juin	95	88,7	Juin	135	122,9
Juillet	94	92,1	Juillet	129	124,0
Août.	92	96,8	Août.	133	131,1
Septembre.	95	101,6	Septembre.	125	136,5
Octobre.	100	102,4	Octobre.	127	136,9
Novembre.	101	108,0	Novembre.	126	133,4
Décembre.	101	107,0	Décembre.	131	131,6

§ 17. — Bien qu'il soit exagéré de chercher à vérifier un phénomène économique à plus d'un dixième près, nous avons — à titre d'exemple — calculé un terme de plus pour notre interpolation. On obtient :

$$\frac{(x^2 \varphi_0)}{(\varphi_0 \varphi_0)} = 17.126 \qquad \frac{(x^2 \varphi_1)}{(\varphi_1 \varphi_1)} = \frac{1.403.005}{7.126} = 196,9$$

$$\therefore \varphi_2(x) = (x - 100)^2 + 3,1\,(x - 100) - 148,5$$

$$\therefore (y\,\varphi_2) = -2.709.2 \quad (\varphi_2\,\varphi_2) = 2.606.466,5$$

$$\therefore A_2 = -0,00104.$$

Le coefficient $r_{1,2}$ et l'erreur $\varepsilon_{1,2}$ sont respectivement :

$$r_{1,2}\sqrt{\frac{1,886 \times 13.438 + 0,00104 \times 2.709,2}{27.599}} = 0,9583$$

$$\varepsilon_{1,2} = \sqrt{1 - r_{1,2}^2} \qquad \eta_y = 6,849.$$

Donc, il faudrait, pour tenir compte des termes qui proviennent du polynôme $\varphi_2(x)$, corriger les valeurs des y déjà obtenues par la première interpolation, d'une erreur relative de l'ordre de :

$$\frac{\varepsilon_{1,1} - \varepsilon_{1,2}}{100} = \frac{6,859 - 6,849}{100} = 0,0001.$$

Or, une telle correction serait dépourvue de tout sens, étant donné que les x et les y sur lesquels on a effectué l'interpolation ont été calculés avec une erreur relative de l'ordre des centièmes. Cette simple remarque constitue à elle seule la vérification de l'existence d'une dépendance linéaire entre le coût de la vie et la circulation fiduciaire.

Serban GHEORGHIU.