A présent on se pose la question: quelle sera la formule de la liaison entre x_1 et y_1, entre x_2 et y_2, entre x_3 et y_3 . . . entre x_i et y_i, etc. si

1-o — cette liaison (dépendance) peut s'exprimer par toute une fonction rationnelle du premier degré (fonction linéaire).

2-o — cette fonction reste la même pour toutes les N couples de termes des deux séries, et

3-o — à notre cas peut s'appliquer la méthode dite „des moindres carrés".

Autrement dit, on suppose que la liaison entre le terme arbitraire, d'ordre i, de la première série x_i et le terme correspondant de la seconde série y_i (i pouvant être égal à 1, 2, 3, 4 . . . . N), s'exprime par la formule: $y_i = a_1 + b_1 x_i$, ou bien par la formule:

$$x_i = a_2 + b_2 y_i.$$

Ici, les coefficients a_1, a_2, b_1 et b_2 sont constants („paramètres") de l'équation et on suppose qu'ils restent invariables pour tous les N couples observés. Les valeurs de tous les termes des deux séries étant connues, on dispose, pour le calcul des 4 paramètres inconnus a_1, b_1, a_2, b_2, de 2 N équations: N équations pour le calcul de a_1 et b_1 et N équations pour celui de a_2 et b_2. Ces équations paraîtront, dans la plupart des cas, réciproquement *incompatibles*, chaque groupe de 4 équations donnant aux a_1, b_1, a_2 et b_2 des valeurs numériques autres que celles qu'on cherche. Pour éviter cette contradiction interne, on a recours à la „méthode des moindres carrés". On raisonne alors de la manière suivante:

Supposons que dans l'équation $y_i = a_1 + b_1 x_i$ les grandeurs a_1 et b_1 aient des valeurs *arbitraires* a_1' et b_1'. Dans ce cas, naturellement, la partie droite de l'équation ne peut être égale à la partie gauche. Si on désigne leur différence par e_i, il est évident que

$$y_i - a_1' - b_1' x_i = e_i,$$

et que e_i peut être considéré comme erreur de la détermination de a_1 et b_1

On cherche à présent de telles valeurs de a'_1 et b'_1, afin que la *somme des carrés* des erreurs de toutes les N équations soit *la moindre* (d'où la dénomination: „méthode des moindres carrés"), c'est-à-dire,

$$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_1 - a_1' - b_1' x_i)^2 = \text{minimum *)}$$

La question ainsi posée, le problème de trouver les valeurs de a_1' et b_1' devient mathématiquement tout à fait précisé. Conformément aux règles du calcul différentiel (réduction à zéro des dérivées par a_1', par b_1', etc.), on obtient aisément:

$$a_1' = 0; \qquad b_1' = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

De la même manière, de l'équation conditionnelle $\sum_{i=1}^{N} (e_1')^2 = \sum_{i=1}^{N} (x_i - a_2' - b_2' y_i)^2 = $ minimum on obtient:

$$a_2' = 0; \qquad b_2' = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} y_i^2}$$

Les valeurs a_1', a_2', b_1', b_2' ne sont pas les véritables valeurs des paramètres de la fonction qui lie, *peut-être*, en réalité x_i et y_i, mais seulement les valeurs approchées de ces paramètres, obtenues par la méthode des moindres carrés. Si nous avions posé comme condition: $\sum_{i=1}^{N} e_i^3 =$ minimum, ou $\sum_{i=1}^{N} E_i^4 =$ minimum, nous aurions obtenu des valeurs tout différentes.

En désignant la moyenne géométrique des deux paramètres b'_1 et b'_2 par r_{12}', on aura:

$$r_{12}' = \sqrt{b_1' \, b_2'} = \frac{\sum_{i=1}^{N} x_i y_i}{\sqrt{\sum_{i=1}^{N} x_i^2 \; \sum_{i=1}^{N} y_i^2}} \qquad [1]$$

Considérant que les Anglais appellent „déviation-type" ou „écart-type" les grandeurs

$$\sigma_1' = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}} \text{ et } \sigma_2' = \sqrt{\frac{\sum_{i=1}^{N} y_i^2}{N}} \text{ *)} \qquad [2]$$

nous pouvons écrire:

$$r'_{12} = \frac{\sum_{i=1}^{N} x_i y_i}{N \sigma_1' \sigma_2'} \qquad [3]$$

D'autre part, en désignant b_1' par b_{12}' et b_2' par b_{21}', nous avons:

$$b_{12}' = \frac{\sigma'_1}{\sigma'_2} r_{12}'; \; b_{21}' = \frac{\sigma_2'}{\sigma_1'} r_{21}' \text{ (puisque } r_{12}' = r_{21}') \qquad [4]$$

Le signe $\sum_{i=1}^{N}$ indique qu'on doit prendre la somme qui embrasse toutes les grandeurs à indices 1, 2, 3, 4, etc., en se terminant par l'indice N. Par exemple:

$$\sum_{i=1}^{N} e_i^2 = e_1^2 + e_2^2 + e_3^2 + \ldots . e^{N2}$$

*) Il serait plus exact de mettre au dénominateur de la grandeur sous racine non pas N, mais (N—1); Or, si N est quelque peu signifiant, cette correction n'a presque aucune importance.