

REVUE DE STATISTIQUE APPLIQUÉE

JERZY HOLUBIEC

Sur une application de la régression multiple dans l'estimation des consommations partielles d'énergie électrique

Revue de statistique appliquée, tome 21, n° 2 (1973), p. 93-98

http://www.numdam.org/item?id=RSA_1973__21_2_93_0

© Société française de statistique, 1973, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

SUR UNE APPLICATION DE LA RÉGRESSION MULTIPLE DANS L'ESTIMATION DES CONSOMMATIONS PARTIELLES D'ÉNERGIE ÉLECTRIQUE

Jerzy HOLUBIEC

Institut de la Cybernétique Appliquée - Varsovie

On démontre que la condition suffisante pour obtenir une estimation correcte, par la méthode de régression multiple, de la répartition de la consommation totale d'énergie électrique parmi divers types ou groupes de récepteurs domestiques, est l'absence de corrélation entre les variables indépendantes des équations de régression multiple.

Un test de vérification de l'hypothèse de l'absence de corrélation entre les variables indépendantes retenues, est également donné.

La régression multiple est utilisée avec succès depuis de nombreuses années dans l'analyse des courbes de charge, en électricité (voir par exemple les travaux de l'UNIPÉDE).

Par contre, les tentatives d'emploi de cette méthode dans l'estimation, à partir de la consommation totale, de la répartition des consommations partielles d'énergie électrique, se soldaient jusqu'ici, en général, par des échecs. En effet, les calculs effectués pour certaines catégories de consommateurs donnaient comme résultat des valeurs négatives de la consommation partielle recherchée⁽¹⁾.

Nous établirons, dans ce qui suit, sur l'exemple des usages domestiques quelles sont les conditions qui doivent être satisfaites pour obtenir des résultats corrects dans l'estimation, à partir de la consommation totale, de la répartition des consommations partielles d'énergie avec les méthodes de régression multiple.

La consommation totale, journalière ou annuelle, d'énergie électrique pour les usages domestiques, est la somme des énergies consommées par les p groupes de récepteurs existants.

$$A = \sum_{i=1}^p A_i \quad (1)$$

(1) Cela a été le cas de trois instituts de recherche visités par l'auteur, dans trois pays.

Il est évident que

$$A, A_i \geq 0 \quad (2)$$

L'essentiel de la méthode de régression multiple, dans l'estimation des grandeurs A_i , consiste à remplacer l'équation(1) par l'équation de régression multiple

$$A = \sum_{i=1}^p B_i X_i \quad (3)$$

et ensuite, à partir d'une épreuve de n éléments (où $n > p$), à évaluer les coefficients inconnus que sont les paramètres partiels de régression B_i .

Les valeurs des variables indépendantes X_i seront déterminées à partir de l'épreuve avec toujours

$$X_i \geq 0 \quad (4)$$

Nous pouvons ainsi estimer les coefficients B_i , et partant les grandeurs A_i . Il ressort immédiatement de (2)-(4), qu'il faut que

$$B_i \geq 0 \quad (5)$$

Le système d'équations défini à partir d'une épreuve de n éléments

$$A_j = \sum_{i=1}^p B_i X_{ij}, \quad j = 1, 2, \dots, n \quad (6)$$

est résolu par la méthode des moindres carrés. On admet que chaque j -équation du système (6) renferme un facteur aléatoire ξ_j , et l'on recherche une solution pour laquelle

$$S = \sum_{j=1}^n \xi_j^2 = \min \quad (7)$$

(l'écart moyen quadratique du facteur aléatoire $\sigma_\xi < \infty$), c'est-à-dire

$$S = \sum_{j=1}^n \left(A_j - \sum_{i=1}^p B_i X_{ij} \right)^2 = \min \quad (8)$$

Si l'on annule les dérivés partielles de S par rapport à B_i , on obtient un système de p équations, dit système d'équations normales, avec p inconnues B_1, B_2, \dots, B_p :

$$B_1 \sum_{j=1}^n X_{1j}^2 + B_2 \sum_{j=1}^n X_{1j} X_{2j} + \dots + B_p \sum_{j=1}^n X_{1j} X_{pj} = \sum_{j=1}^n A_j X_{1j}$$

$$B_1 \sum_{j=1}^n X_{2j} X_{1j} + B_2 \sum_{j=1}^n X_{2j}^2 + \dots + B_p \sum_{j=1}^n X_{2j} X_{pj} = \sum_{j=1}^n A_j X_{2j} \quad (9)$$

$$B_1 \sum_{j=1}^n X_{pj} X_{1j} + B_2 \sum_{j=1}^n X_{pj} X_{2j} + \dots + B_p \sum_{j=1}^n X_{pj}^2 = \sum_{j=1}^n A_j X_{pj}$$

Remplaçons maintenant les variables X_1, X_2, \dots, X_p par les variables normalisées x_1, x_2, \dots, x_p , où

$$x_i = \frac{X_i - \bar{X}_i}{\sigma_i}, \quad i = 1, 2, \dots, p \quad (10)$$

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij} \quad (11)$$

$$\sigma_i^2 = \frac{1}{n} \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 \quad (12)$$

avec

$$a = \frac{A - \bar{A}}{\sigma_a} \quad (13)$$

$$\bar{A} = \frac{1}{n} \sum_{j=1}^n A_j \quad (14)$$

$$\sigma_a^2 = \frac{1}{n} \sum_{j=1}^n (A_j - \bar{A})^2 \quad (15)$$

Les variables normalisées x_1, x_2, \dots, x_p ont une valeur moyenne nulle et une variance égale à l'unité.

$$\bar{x}_i = 0 \quad (16)$$

$$\sigma_i^2 = 1$$

avec une covariance égale au facteur de corrélation

$$\text{cov}(x_i, x_l) = \frac{1}{n} \sum_{j=1}^n x_{ij} x_{lj} = \rho_{xi} \rho_{xl} = \rho_{il} \quad (17)$$

$$i = 1, 2, \dots, p, i \neq l$$

On a donc également

$$\text{cov}(a, x_i) = \rho_{ai} \quad (18)$$

En notation matricielle, le système d'équations (9) s'écrira

$$\begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \dots & \dots & \dots & \dots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_p \end{bmatrix} = \begin{bmatrix} \rho_{a1} \\ \rho_{a2} \\ \dots \\ \rho_{ap} \end{bmatrix} \quad (19)$$

Avec la condition

$$B_i = b_i \frac{\sigma_a}{\sigma_i} \quad (20)$$

Le système d'équations (19) est appelé système standardisé d'équations normales.

La condition (5) est remplie dans (19), si

$$\rho_{ii} = 0, i \neq 1 \quad (21)$$

c'est-à-dire que les variables x_i, x_1 – et partant les variables X_i, X_1 – sont non corrélées. On sait, en effet, que pour les variables dont l'espérance mathématique et la variance ont des valeurs finies, on a

$$\text{cov}(x_i, x_1) = \rho_{ii} = 0 \quad (22)$$

d'où l'on peut conclure, que la condition suffisante pour obtenir une estimation correcte par la méthode de régression multiple de la répartition de la consommation totale, parmi divers types ou groupes de récepteurs domestiques, est l'absence de corrélation entre les variables indépendantes. C'est dire que l'obtention de valeurs négatives de consommation d'énergie électrique, pour certains groupes de récepteurs domestiques, était dûe à la corrélation des variables indépendantes retenues.

Il n'est pas exclu que, dans les conditions polonaises, l'hypothèse suivante ne sera pas rejetée : la consommation d'énergie électrique, dans les foyers de ce groupe de villes, est indépendante (dans les limites de 3 à 5 personnes par foyer) du nombre de personnes par foyer.

C'est ainsi que, dans le cas polonais, les variables indépendantes qui pourraient être retenues, dans les foyers des villes comptant de 20 000 à 100 000 habitants seraient : la consommation d'énergie électrique dans les foyers, le nombre de personnes par foyer, le nombre d'habitants. S'il n'y a pas corrélation des variables indépendantes, la matrice ψ des coefficients du système standardisé d'équations normales (19) sera une matrice unité.

En général, les coefficients de corrélation ρ_{ii} de la matrice ψ , ne seront pas nuls, en raison du nombre relativement faible d'éléments n de l'épreuve retenue. Il est également possible de prendre des variables indépendantes corrélées. L'hypothèse $H_0 : \psi = I$ (I est une matrice unité) doit alors être vérifiée.

La vérification de l'hypothèse peut être effectuée à l'aide du test λ de Wilks [1]. On calcule, dans ce but, la grandeur χ^2 .

$$\chi^2 = -2 \ln \lambda$$

que l'on compare avec la valeur χ^2_α trouvée dans les tables pour $\nu = \frac{1}{2} p(p+1)$ degrés de liberté et un niveau de confiance α .

Dans le cas considéré des transformations appropriées de l'expression

du test λ donnent

$$\lambda = \frac{C^{1/2 n}}{n^{1/2 p n}} \quad (23)$$

où

$$C = n \psi$$

Si $\chi^2 \leq \chi_{\alpha}^2$, il n'y a pas de raisons de rejeter l'hypothèse H_0 .

Lorsque le nombre de variables indépendantes est petit (pratiquement $2 \leq p \leq 5$), il est possible de vérifier l'hypothèse que les coefficients ρ_{ij} diffèrent peu de zéro. On utilise alors le test t de Student pour (n-2) degrés de liberté.

L'utilisation des tests λ de Wilks, et t de Student, entraîne la nécessité, pour les variables indépendantes x_1, \dots, x_p , de répondre à des lois normales de distribution.

Dans l'analyse par les méthodes de régression multiple de la répartition de la consommation totale d'énergie électrique dans les foyers domestiques (suivant les groupes ou types de récepteurs), il est parfois utile d'écrire l'équation de régression sous la forme

$$A = A_0 + \sum_{i=1}^p B_i X_i \quad (24)$$

Le membre A_0 correspond, dans ce cas, à la consommation d'énergie par les "récepteurs restants", pour lesquels l'établissement d'une relation entre la consommation et une variable quelconque n'est pas fondé.

Le terme A_0 se calcule alors à partir de la formule

$$A_0 = \bar{A} - \sum_{i=1}^p B_i \bar{X}_i \quad (25)$$

Les valeurs B_i et A_i issues des calculs, sont des estimations des valeurs réelles des grandeurs en question, et peuvent être soumises à la vérification statistique [4], [7].

La conclusion exprimée ci-dessus est également valable dans le cas de l'estimation, par la méthode de régression multiple, de la répartition de la consommation totale d'énergie électrique en consommations partielles, pour les autres groupes de récepteurs d'énergie électrique.

BIBLIOGRAPHIE

- [1] ANDERSON T. : An introduction to multivariate statistical analysis, J. Wiley, New York, 1960.
- [2] FELLER W. : An introduction to probability theory and its applications, (polish translation), PWN Warszawa 1960.

- [3] FISZ M. : Rachunek prawdopodobienstwa i statystyka matematyczna *Calcul des probabilités et statistique mathématique*, Editions Scientifiques Nationales (PWN), Varsovie, 1958.
- [4] HOLUBIEC J. : Application de la régression multiple en électricité. *Przegląd Elektrotechniczny Revue d'Electrotechnique*, N° 2, 1968.
- [5] MANCZAK K. : Application de l'analyse des régressions dans l'identification des caractéristiques des installations technologiques multidimensionnelles. *Archiwum Automatyki i Telemekhaniki* T. XI, N° 1, 1968.
- [6] RUDNICKI B. : Principes d'une méthode de prévision des besoins en énergie électrique et d'établissement des courbes de charge des récepteurs domestiques et communaux. Instytut Energetyki, Varsovie. Doc. N° 5408, novembre 1966.
- [7] VEDERE E. : Méthode de regression multiple appliquée dans l'analyse des courbes de charge, UNIPEDE, Congrès de Baden-Baden, 1961, rap. VIIIa. 1.