

REVUE DE STATISTIQUE APPLIQUÉE

E. MORICE

Confection, biométrie et statistique

Revue de statistique appliquée, tome 8, n° 3 (1960), p. 87-94

[<http://www.numdam.org/item?id=RSA_1960__8_3_87_0>](http://www.numdam.org/item?id=RSA_1960__8_3_87_0)

© Société française de statistique, 1960, 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/>

CONFECTION, BIOMÉTRIE ET STATISTIQUE

E. MORICE

Inspecteur Général à l'Institut National de la Statistique et des
Études Économiques

Au cours des vingt dernières années, diverses études ont été faites (aux États-Unis sur une très vaste échelle, et plus récemment en Hollande) pour déterminer à partir de mesures précises, des "standards" de confection correspondant au mieux aux dimensions moyennes de la population à laquelle ces vêtements étaient destinés.

La réalisation, en France, d'une enquête analogue avait été étudiée dès le début de 1947 : une "Commission Nationale des Mensurations du corps humain", groupant des représentants du Centre d'Études Techniques des Industries de l'Habillement, et des principales maisons de confection masculine en grande série, ainsi que des biométriciens et des statisticiens, avait établi un projet d'enquête dont la réalisation avait dû être différée faute de moyens matériels suffisants.

Cette étude a été reprise et réalisée en 1959 pour la confection féminine par le Centre d'Études Techniques des Industries de l'Habillement avec la collaboration du Centre de Formation aux Applications Industrielles de la Statistique (voir article M. Férignac, ci-après).

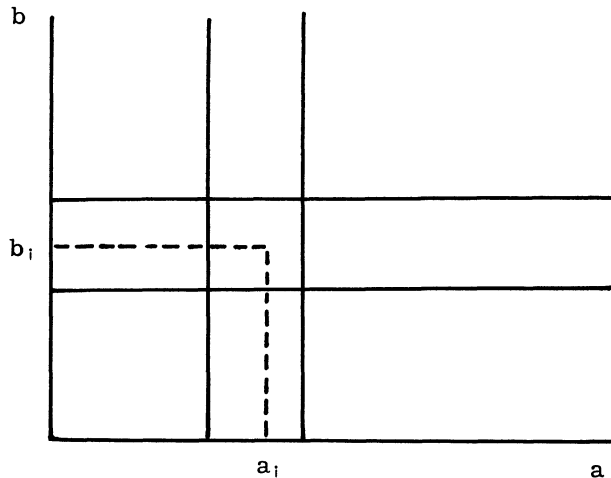
Étude rationnelle de la confection

Étant donnée la nécessité, au point de vue économique, de limiter de manière acceptable le nombre de tailles confectionnées en série, le problème essentiel consiste donc à déterminer quelles sont, parmi les différentes mesures (longueurs, périmètres, poids) auxquelles peut donner lieu le corps humain, celles (en petit nombre, deux ou trois au maximum) qui décrivent le mieux toutes les autres, c'est-à-dire à partir desquelles on peut espérer estimer les autres avec le minimum de dispersion résiduelle.

Ces mesures ayant été définies, au nombre de deux par exemple, a et b , et un certain nombre de classes ayant été choisi pour chacune d'elles de manière à définir des intervalles considérés comme acceptables, on détermine ainsi un certain nombre de cases, chacune correspondant à un "patron" de confection dont toutes les dimensions peuvent être estimées à partir de (a_i, b_i) par les équations de régression déduites de l'ensemble des observations.

Au point de vue statistique, il s'agit donc :

1/ d'un problème d'échantillonnage : des considérations pratiques, financières ou autres, limitant, en général, assez rapidement l'effectif soumis aux



mensurations (notons en passant qu'aux Etats-Unis, l'enquête organisée en 1936 par le "U. S. Department of Agriculture" et la "Works Progress Administration" et réalisée de février 1937 à juin 1939 dans seize Etats, avait porté sur 76 744 garçons et 70 344 filles de 4 à 17 ans.

2/ d'un problème d'observation, c'est-à-dire ici de mensurations dont il faut souligner toute l'importance, nécessitant la formation d'un corps d'enquêteurs, capables de définir d'une manière précise les repères limitant les mensurations prises ou fixant une position et utilisant de la même manière les mêmes instruments de mesure.

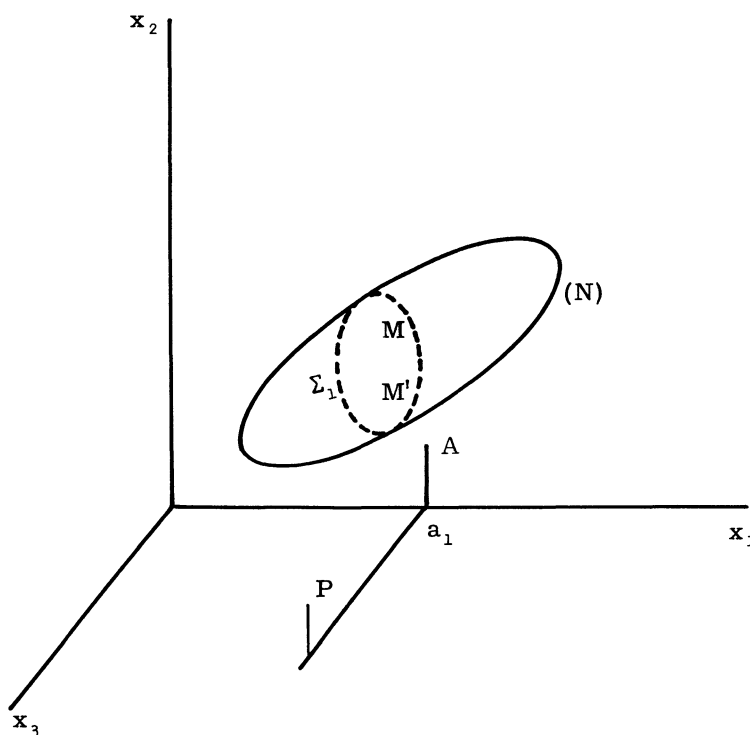
3/ après le dépouillement, vient le problème d'analyse statistique : étude des distributions observées, calculs de corrélations et enfin le problème spécifique posé ci-dessus, c'est-à-dire l'analyse de ces corrélations (analyse factorielle), en vue de déterminer, parmi les multiples facteurs observés, quels sont les composantes principales de cet ensemble, les facteurs de groupe expliquant au mieux tel ou tel groupe de mesures et dont la simple considération permettra d'estimer au mieux tous les autres, avec une précision qu'il sera utile de connaître en calculant les variances résiduelles.

Le calcul des coefficients de corrélation est déjà assez lourd : $\left(\frac{n(n-1)}{2}\right)$ coefficients à calculer si on a conservé dans l'étude n mesures), mais ce calcul peut être assez facilement réalisé au moyen d'un équipement mécanographique de type usuel.

Reste le problème d'analyse factorielle proprement dit qui représente une masse considérable de calculs si le nombre n des mesures est important.

On peut schématiser le problème de la manière suivante :

Supposons que trois mesures x_1, x_2, x_3 suffisent pour dessiner un patron (ce qui n'est évidemment pas le cas), et proposons-nous de construire des patrons entièrement déterminés par une mesure $x_1 = a_1$, les deux autres étant estimées à partir de celle-là à l'aide des équations de régression déduites de l'ensemble des mesures (x_1, x_2, x_3) . Laquelle de ces trois variables convient-il de choisir comme variable permettant d'estimer au mieux les deux autres.



Géométriquement, on peut faire correspondre, à chacune des trois mesures faites sur un individu de l'échantillon étudié, un point $M(x_1, x_2, x_3)$. L'ensemble de ces points détermine un nuage (N) .

Se donner $x_1 = a_1$, c'est couper ce nuage par un plan P perpendiculaire à Ox_1 , au point A , d'abscisse x_1 . Le sous-échantillon $(x_1 = a_1)$ est alors représenté par le nuage plan Σ_1 . Dans ce nuage, le patron est représenté par le point $M'(x_1, y_2, y_3)$ y_2 et y_3 étant les valeurs estimées de x_2, x_3 .

La solution choisie sera d'autant meilleure que le point M' sera moins éloigné de l'ensemble des points M du nuage Σ_1 , ceci étant vrai pour les différentes valeurs de $x_i = a_i$. Schématiquement, ceci revient à chercher quelles sont les sections $x_i = \text{constante}$ ($i = 1, 2$ ou 3) qui, dans l'ensemble, sont les plus concentrées.

Plus généralement, si parmi les n variables mesurées $x_1, \dots, x_n$, on a choisi s variables explicatives $x_1, x_2, \dots, x_s$, à partir desquelles on estime les $t = n - s$ variables restantes par $y_{s+1}, \dots, y_n$, on aura à envisager dans l'espace à n dimensions les sections du nuage total, obtenues en considérant des valeurs fixées $x_1 = a_1, \dots, x_s = a_s$ du groupe des variables explicatives, c'est-à-dire à étudier les variances résiduelles obtenues en associant les points échantillons $M(x_1, \dots, x_s, \dots, x_n)$ au point $M'(x_1, \dots, x_s, y_{s+1}, \dots, y_n)$ correspondant aux valeurs fixées :

$$x_1 = a_1, \quad x_2 = a_2, \quad \dots, \quad x_s = a_s$$

Il restera ensuite à choisir parmi les combinaisons possibles de s variables celle pour laquelle cette variance résiduelle est suffisamment petite.

Sans entrer dans le détail des calculs, on se rend compte que ceci exigera la manipulation de nombreuses équations de régression, c'est-à-dire le calcul - à partir des corrélations déjà obtenues - de très nombreux coefficients de corrélation partielle ou multiple nécessaires pour obtenir et ensuite comparer les variances résiduelles correspondant aux différents groupes envisagés de variables descriptives.

Une étude théorique complète du problème est extrêmement lourde, sinon matériellement irréalisable. Néanmoins, dans la pratique divers éléments peuvent limiter la masse des calculs.

Si on se place au point de vue strict du confectionneur, il est tout d'abord évident que certaines mesures biométriques, bien que correspondant à des parties du corps couvertes par le vêtement sont pratiquement sans intérêt.

D'autre part, le problème ne peut avoir de solution pratiquement utilisable que si s est petit (2 ou 3 par exemple), sinon, on serait conduit à un nombre trop grand de patrons différents, satisfaisants peut-être du point de vue de l'ajustement, mais économiquement sans intérêt du point de vue du confectionneur.

Il y a lieu aussi de tenir compte des résultats acquis par l'expérience de longues années, du fait que certaines mesures qui peuvent paraître théoriquement intéressantes à partir d'un échantillon de mesures soigneusement étudié sont, dans la pratique courante, trop difficiles à obtenir de façon précise avec le matériel couramment employé.

Enfin, il faut noter que le schéma utilisé aura un pouvoir de discrimination d'autant plus grand que les variables explicatives retenues auront un champ de variation plus large.

Cependant, même en tenant compte de ces éléments, il restera un nombre appréciable de variables explicatives auxquelles il conviendrait théoriquement d'appliquer les méthodes de l'analyse factorielle, seules capables de donner une solution absolument correcte du problème.

Une méthode graphique fort simple, inspirée des profils psychologiques utilisés en psychotechnique, semble devoir jeter quelque lumière sur le problème envisagé.

Considérons pour chaque individu un certain nombre de grandeurs $A, B, \dots, K, L$, caractérisées par leurs mesures $a, b, \dots, k, l$, et soient calculés les coefficients de corrélation :

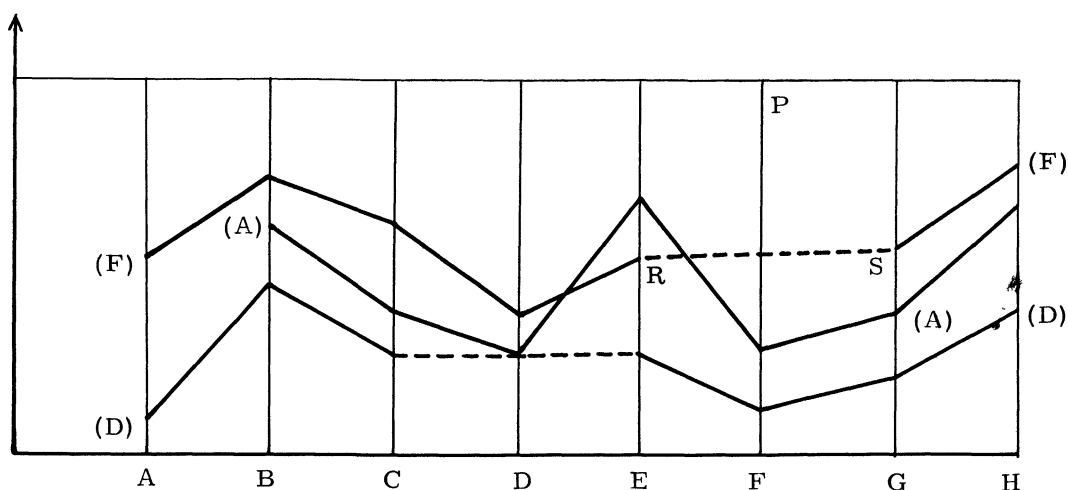
$$r_{ab}, r_{ac}, \dots, r_{bc}, \dots, r_{kl}$$

Portons sur l'axe des abscisses, à intervalles égaux (en principe dans un ordre quelconque), les points $A, B, \dots, K, L$, représentatifs des grandeurs considérées, et traçons pour chacune de ces grandeurs la ligne polygonale représentative des variations du coefficient de corrélation de l'une de ces grandeurs, F par exemple, avec toutes les autres.

On a évidemment :

$$r_{ff} = 1$$

mais, pour faciliter les comparaisons qui vont suivre, il n'y a pas intérêt à inclure ce point P dans le graphique : il vaut mieux, simplement pour faciliter l'examen de la ligne (F), joindre en pointillé les deux points R et S qui l'encadrent.



On obtiendra ainsi un réseau plus ou moins entrelacé de lignes polygonales (A), (B), , (L).

Si toutes les variables étudiées jouent chacune, dans l'explication globale du phénomène étudié, un rôle individuel sans lien avec les autres, le graphique se présentera comme un treillis indéchiffrable. Mais il peut arriver que certaines grandeurs F, D, A, par exemple, jouent chacune des rôles comparables dans la description (ou l'estimation) des autres. Dans ce cas, les graphiques (F), (D), (A); se présenteront comme des lignes polygonales ayant sensiblement la même forme.

Si de plus, l'une d'entre elles (F), par exemple, tend à se placer systématiquement au-dessus du réseau partiel ainsi déterminé, cela voudra dire que si F, D, A fournissent des explications relativement comparables de l'ensemble, l'explication (l'estimation) fournie par (F) est meilleure que les autres.

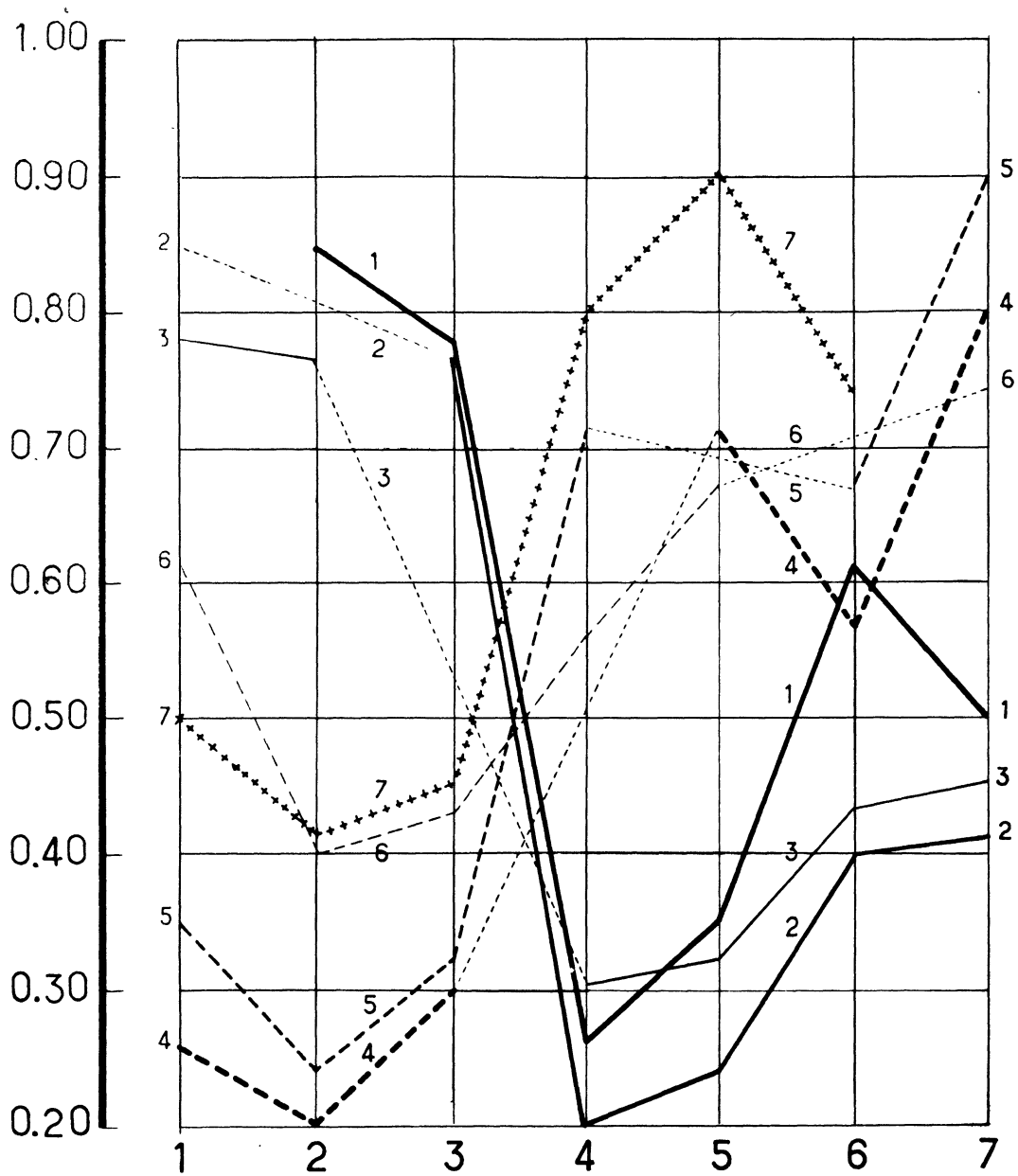
En un mot, pour employer le langage de l'analyse factorielle, on aura mis en évidence un groupe de facteurs F, D, A, le facteur F pouvant être considéré comme facteur caractéristique du groupe.

Un exemple numérique précisera la signification et la portée du procédé.

Soient 7 variables, A, B, C, D, E, F, G, dont les coefficients de corrélation deux à deux sont donnés par le tableau :

	A 1	B 6	C 5	D 2	E 4	F 7	G 3
A 1		0,61	0,35	0,85	0,26	0,50	0,78
B 6			0,67	0,40	0,56	0,74	0,43
C 5				0,24	0,71	0,90	0,32
D 2					0,20	0,41	0,77
E 4						0,80	0,30
F 7							0,45
G 3							

PROFIL DE CORRÉLATIONS



On en déduit le graphique ci-dessus (fig. 2).

L'examen de ce graphique montre l'existence de deux groupes :

1/ B, C, E, F (numéros 6, 5, 4, 7)

2. A, D, G (numéros 1, 2, 3).

Dans le premier groupe, les points du graphique F sont, en général au-dessus des autres, tandis que dans le deuxième groupe ce sont, en général, les points A.

Si l'on se propose de décrire l'ensemble des 7 variables, A, B, C, D, E, F, G, à l'aide de deux d'entre elles, ce sont manifestement les deux variables A et F qu'il conviendra de choisir pour servir de base à l'estimation des autres.

Les 7 variables considérées dans l'exemple ci-dessus sont sept caractéristiques biométriques mesurées sur 3 329 jeunes filles américaines de 16 ans⁽¹⁾, ces mesures étant classées ci-dessous dans un ordre absolument arbitraire sans se soucier des liaisons que leur signification physique aurait pu suggérer à l'avance :

A : Taille

B : Périmètre vertical du tronc (épaule, entrejambes)

C : Tour de hanches

D : Hauteur des hanches

E : Tour de poitrine (aisselle)

F : Poids

G : Longueur totale postérieure du bras.

Les deux groupes mis en évidence par cette analyse graphique que l'on pourrait appeler profil de corrélation sont donc :

I - LE GROUPE DES LONGUEURS -

1/ A Taille

2/ D Hauteur des hanches

3/ G Longueur totale postérieure du bras

II - LE GROUPE DES PERIMETRES AUQUEL S'AJOUTE LA VARIABLE POIDS

4/ E Tour de poitrine

5/ C Tour de hanches

6/ B Périmètre vertical du tronc

7/ F Poids.

En ce qui concerne le partage entre longueurs simples et périmètres, on pouvait évidemment s'y attendre encore que certaines idées simples conçues a priori peuvent quelquefois se trouver contredites par les observations.

Mais de plus le graphique montre :

(1) "Body measurement of american boys and girls" - USA - Dept. of agriculture - N° 366 - July 1941.

1/ Que le poids se comporte - en tant que variable explicative - tout à fait comme les périmètres et pas du tout comme les longueurs.

2/ Que si l'on considère deux groupes de facteurs, les deux meilleures variables explicatives sont, d'une part, la taille et le poids (chose à laquelle on n'aurait vraisemblablement pas pensé a priori).

3/ Et que si l'on admet que les six mesures numérotées de 1 à 6 suffisent à définir un patron, la meilleure estimation des dimensions 2, 3, 4, 5 et 6, c'est-à-dire (d, g, e, c, b) se ferait à partir d'équations d'estimations (ou de régression) de la forme :

$$\begin{array}{l} d = F_2(a, f) \\ g = F_3(a, f) \\ \text{-----} \\ b = F_6(a, f) \end{array}$$

Pour une jeune fille du groupe considéré, la détermination du "patron" lui convenant le mieux serait simplement basée sur la connaissance de sa taille et de son poids.

Il est évident que les dimensions nécessaires pour dessiner un patron - variables d'ailleurs avec le type de vêtement - ne sont pas celles qui figurent dans l'exemple ci-dessus, mais les résultats acquis pour des dimensions caractéristiques du corps humain restent valables pour celles des éléments de surfaces développables que le confectionneur veut appliquer sur ce corps.

Ce sont là les résultats - valables d'ailleurs pour tout le groupe d'âge de 6 à 17 ans - auxquels a conduit l'enquête américaine citée ci-dessus, après de laborieux calculs d'analyse factorielle basés sur la méthode de recherche des composantes principales proposée par Hotelling, en 1933 (1).

Seules, les considérations d'ordre psychologique (il est peut-être trop inhabituel de demander à une jeune fille quel est son poids quand on se préoccupe simplement de la vêtir), ont conduit les confectionneurs américains à baser leur série de confection féminine sur la taille et le tour de hanches, cette dernière mesure étant, après le poids, le meilleur facteur descriptif du groupe II pour l'ensemble de la population étudiée (6 à 17 ans), après élimination du périmètre vertical du tronc, de mesure difficile à réaliser dans les conditions habituelles (sujet habillé).

(1) H. Hotelling - Analysis of a complex of Statistical variables into Principal Components - Journal Ed. Psychol. 1933.