Méthode informationnelle pour vérifier la signification des résultats de l'analyse des données - application
pour un indicateur « synthétique substitutif »

Sabina Popescu-Spineni[1], Justinian Popescu-Spineni[2]
1) Université de Médicine et Pharmacie "Carol Davila";
2) Institut de Biologie, Bucarest, Roumanie;
spopescu@ispb.ro
indicateur agrégé, analyse factorielle, statistique informationnelle

Résumé

Ont été proposés beaucoup de méthodes pour édifier un indicateur global synthétique, substitutif, consécutif aux processus de décision complexes, basé sur une relation d'ordre. Etablir une relation d'ordre, ou quelques « priorités », dans une politique économique, c'est une démarche usuelle pour l'efficacité de la décision. Si le système analysé est sous l'influence d'un large group d'indicateurs, la décision doit être basée sur l'analyse vectorielle. Un indicateur synthétique c'est théorique inacceptable (le théorème d'Arrow), mais il est encore possible l'existence d'un substitut. Ils sont plusieurs techniques pour réduire les éléments d'un vecteur basé sur une construction scalaire de la décision, en utilisant un « substitut global synthétique ». [2]

Les auteurs Dinescu et all. (1986) ont proposés un substitut aléatoire synthétique pour remplacer plusieurs indicateurs d'un système économique analysé. Ils ont réduit la demande de la représentativité du « substitut synthétique » à la demande de la dépendance maximale entre les groups des variables aléatoires et leur propre variable – « *substitut synthétique* », ceci construit comme une fonction des indicateurs du group analysé. Une mesure appropriée pour ce concept de *substitut synthétique*, a été construite pour décrire la dépendance entre une variable aléatoire et un vecteur aléatoire. [2]

Par exemple, ils ont proposée une extension du concept du coefficient de corrélation, en utilisant la méthode des *multiplicateurs du Lagrange*, donc les solutions λ du système global, comme les valeurs propres de la matrice S, où : $|S - \lambda E| = 0$. Pour augmenter la possibilité de déterminer la variable aléatoire normée pour un indicateur synthétique maximal, ils ont utilisés le concept du coefficient de corrélation, dans le cas de dépendance entre une variable aléatoire et un vecteur aléatoire. Les auteurs ont proposés aussi plusieurs applications pour leur indicateur synthétique. [2]

Dans nôtre travail, nous voulons présenter une méthode informationnelle pour tester la signification statistique de l'indicateur justement décrit - « *indicateur synthétique substitutif* » (Dinescu et all.), à l'aide d'une des applications présentés par les auteurs. Mais le teste de signification est basé sur l'indicateur *synergique informationnel - SYNSPS* (S. Popescu- Spineni, 1998), élaboré pour l'évaluation globale de la qualité de l'analyse factorielle et pour réduire le nombre des facteurs sans pêrte d'information, en utilisant un concept de la statistique informationnelle, avec des références théoriques appropriées. [4] [5]

Rezumat

S-au propus diferite metode pentru construirea unui indicator sintetic global, substitutiv, consecutiv unor procese de decizie complexe, bazate pe o relație de

ordine. Stabilirea unei relații deordine sau a unor priorități, în cayul unor politici economice, este o cerință usuală pentru eficientizarea deciziilor. Dacă sistemul analizat este sub influența unui mare grup de indicatori, decizia trebuie să se bazeze pe o analiză vectorială. Un indicator sintetic este teoretic inacceptabil (teorema lui Arrow), dar este posibilă construirea unui substitut. Există tehnici pentru reducerea elementelor unui vector pentru o construcție scalară a deciziei, utilizând un « substitut sintetic global » (Dinescu et all.).

Lucrarea noastră reprezintă încercarea de a recomanda o metodă informațională, verificată pentru testarea calității reprezentării în cazul analizei factoriale a datelor, bazată pe crearea unui indicator informațional agregat (S. Popescu-Spineni, 1998), pentru a fi, eventual, folosită şi la testarea semnificației statistice a unui indicator de tip substituent sintetic global.

1. Introduction

Dans les sciences médico-sociales, aussi que dans le management sanitaire, l'analyse multivariée des données a été imposée par des besoins opérationnels. Mais l'analyse des données utilise beaucoup de méthodes d'optimisation qui proposent des algorithmes très rigoureuses pour établir la partition d'un set n d'objets caractérisés par k variables, qui caractérise une population ou un group de référence. L'analyse multidimensionnelle des données (MDA) peut inclure deux principales méthodes d'analyse: l'analyse cluster (AC) et l'analyse linéaire des données (composantes principales, canonique, factorielle, analyse des correspondances), simples ou multiples. C'est une représentation synthétique sans avoir manque d'information. [1]

On a beaucoup de techniques empiriques pour réduire le nombre des axes factorielles (Kaiser, Cattel), mais dans cet étude on présente un bon critère pour trouver et tester un indicateur informationnel agrégé (*SYNSPS* - Popescu- Spineni, 1998), celui-ci calculé par l'aide d'un concept statistique informationnel, sans avoir manque d'information. Sont bien proposés beaucoup d'exemples, pour appliquer et tester cet indicateur informationnel agrégé, avec des références théoriques appropriées. [4] [5]

2. L'aproximation de la distribution des valeures propres

La réduction du nombre des facteurs dans l'analyse des donnees, factorielle ou des correspondances (Benzécri, 1980), dépend de l'analyse de la distribution théorique des valeurs propres (λ_i). Dans un tableau de contingence, sous l'hypothèse de l'indépendence des lignes et des colonnes, cette distribution s'approche d'une lois de distribution connue d'une matrice Wishart (Lebart, Morineau, Piron, 1995). On se souligne que la lois de distribution des valeurs propres résultées de l'analyse a été erroné appréciée, aussi que l'inertie totale du nuage des points, comme de type χ^2, pendant que différentes simulations ont démontré au contraire, en restant un problème théorique encore. L'utilisation de la rate de l'inertie (pourcentage de la variance) pour la "qualité de la représentation", proposée de Benzécri (1979) est très difficile d'être globalisée. Enfin, pour mesurer la variation de l'information, Kullback (1959) a proposé la théorie de Shannon-Wiener pour utiliser un indice de divergence dans les problèmes d'inférence pour la statistique multivariée (Lebart et autres, 1995). Basée sur le théorème de Bayes, on mesure la distance $J(H_1{:}H_2)$ entre les hypothèses H_1 versus H_2 (alternatives), en tenant compte de la matrice de covariance (Γ) et des valeurs propres (λ_j) : $J(I,\Gamma) = \frac{1}{2}\left(\sum_{i=1}^{p}\lambda i + \sum_{i=1}^{p}\frac{1}{\lambda i}\right) - p > 0,$

où I(1: 2) est l'information moyenne après la discrimination dans les deux espaces des deux échantillons, sous les deux hypothèses contraires (Lebart et autres, 1995). [3]
Si les deux inerties théoriques totales sont égales sous les deux hypothèses H_1 et H_2,

on a: $$\sum_{i=1}^{p} \lambda i = \mathrm{p}$$

Donc, le seul tèrme qui peut avoir une variation est: $\sum_{i=1}^{p} \frac{1}{\lambda_i}$,

le fait qui pose en question les petites valeurs propres, en temps que l'analyse factorielle ne retient en général que les plus grandes (Lebart et autres, 1995). D'après Lebart et autres, on peut aussi construir, basé sur la théorie de Kullback (1959), un domaine critique: $$\{ \left| \sum_{\alpha=1}^{p} \frac{1}{\lambda_\alpha} - \sum_{\alpha=1}^{p} \lambda_\alpha \right| \} \geq \mathrm{C}.$$

On peut obsérver que la divérgence entre les deux hypothèses augmente dans le cas où les valeurs propres de la matrice Γ se rapprochent de zéro. Conformement au théorie de l'information, les valeurs propres infiniment petites aurront un impact plus grand par rapport aux celles qui peuvent éxpliquer près de 80% de l'inertie totale du souséspace des deux facteurs mis en correspondence, avec une pêrte considérable d'information (Lebart et autres, 1995). [3]

3.Indicateur informationnel aggrégé *SYNSPS*:

En général, dans l'analyse factorielle: $\tau_j = \lambda_j / \sum_{j=1}^{p} \lambda_j$, avec: $\sum_{j=1}^{p} \tau_j = 1$, $(1 < j < k)$, représente l'indicateur de la "qualité de la représentation" (Benzécri, 1979). En tenant compte de ces considérations, aussi que des théories informationnelles (de Kullback, 1959 et de Onicescu, 1963), pour une évaluation globale de l'analyse factorielle, il a été proposé dans nos travaux l'utilisation d'un indicateur informationnel aggrégé (SYNSPS- Popescu-Spineni, 1998), d'omogénité / hétérogénité, construit en partant des deux rapports (informationnels), construites dans ce but, à l'aide du concept *d'énergie informationnelle* (O. Onicescu, 1963): [3] [4] [5]

$$(1) \quad \mathrm{E}(\tau) = \sum_{j=1}^{K} \frac{\lambda_j^2}{(\sum_{j=1}^{K} \lambda_j)^2}, \quad \text{et} \qquad (2) \quad \mathrm{E}(\tau') = \sum_{j=1}^{K} \frac{\frac{1}{\lambda_j^2}}{(\sum_{j=1}^{K} \frac{1}{\lambda_j})^2},$$

Ces rapports ont conduit à élaborer le *SYNSPS* - "*indicateur synergique*" (ou domaine critique), étant en mesure d'évaluer la signification globale de la repré-sentation, par comparaison avec le nombre des variables retenues (1/ k, avec k < p):

$$\text{(SYNSPS):} \qquad \{|\mathrm{E}(\tau) - \mathrm{E}(\tau')|\} \geq \frac{1}{\mathrm{K}}, \qquad (3)$$

L'indicateur $\mathrm{E}(\tau)$, est mis en balance (dans *SYNSPS*) avec l'indicateur, $\mathrm{E}(\tau')$, en tenant compte qu'ils ont une variation informationnelle, stricte entre 1/K et 1, car la constante C = 1/ K est précisement établie, sans avoir besoin d'une simulation et sans avoir manque d'information.

Interprétation : dans le cas où la valeur de ISPS > 1/K, on a une signification statistique de l'analyse (H_0 acceptée), au contraire, pour SYNSPS $\leq$ 1/K, l'analyse globale n'est pas significative (H_0 rejetée). [4] [5]

4.Application pour l'analyse factorielle des données

4.1. Pour le commencement, on peut reprendre l'exemple d'utilisation comparative des deux indicateurs informationnels $E(\tau_j)$ et $E(\tau'_j)$, aussi que de *SYNSPS* (*l'indicateur synergique*) sur les valeurs propres sorties d'une application éffectuée de A. Rizzi ("*Analisi dei Dati*", Roma, 1989, p.188), sur un tableau "4x10": [6] [7]

Exemple No.1:

Ex.1	λ_j	τ_j	cum.	cum.	$1/\lambda_j$	τ'j
1	0.0443	0.532	-	1.000	22.573	0.004
2	0.0202	0.242	0.532	0.996	49.506	0.008
3	0.0088	0.106	0.772	0.998	113.636	0.018
4	0.0054	0.065	0.887	0.970	185.785	0.029
5	0.0023	0.028	0.942	0.941	434.783	0.069
6	0.0021	0.025	0.970	0.872	476.191	0.076
7	0.0002	0.003	0.997	0.796	5000.000	0.796
	0.0833	1.000	1.000	–	6282.474	1.000

En calculant: (1) $E(\tau_j)=\sum_{j=1}^{K}\frac{\lambda_j^2}{(\sum_{j=1}\lambda_j)^2}$; et (2) $E(\tau'_j)=\sum_{j=1}^{K}\frac{1/\lambda_j^2}{(\sum_{j=1}1/\lambda_j)^2}$;

En calculant aussi l'indicateur *SYNSPS*: $\{|E(\tau)-E(\tau')|\}\geq\frac{1}{K}$, (3)

on obtient un résultat global significatif:

$E(\tau_j)$ = 0.357; $E(\tau'_j)$ = 0.645; où: k = 7;

$\{|E(\tau)-E(\tau')|\}=0.645-0.357=0.288>0.143=\frac{1}{7}$.

En conclusion, il faut rejeter H_0, car le tableau n'est pas équilibré. [4] [5]

5. Application sur le coefficient synthétique substitutif

Explication : Le coefficient synthétique substitutif ξ a été construit sur la base des indicateurs q'on dispose, comme fonction des indicateurs du groupe synthétisé, en supposant pour le commencement que la fonction est linéaire.

Soit les variables aléatoires $X_1, X_2, \ldots, X_n$; soit la fonction du substituent synthétique :

$$Y=\sum_{i=1}^{n}a_iX_i\ .$$

En principal, le problème de construire le substituent synthétique se réduit au problème de trouver les valeurs des coefficients de pondération $a_1, a_2, \ldots, a_n$ pour lesquelles la dépendance de la variable aléatoire Y par rapport au vecteur aléatoire X = $(X_1, X_2, \ldots, X_n)$ est maximale .Mais, pour mesurer la dépendance on ne peut pas utiliser le coefficient de corrélation linéaire multiple qui peut banaliser le problème .

Les auteurs (Dinescu et all.) ont construit, en première étape, une mesure, adéquate au concept de substituent synthétique, pour exprimer la dépendance entre une variable aléatoire et un vecteur aléatoire. Ils supposent que les variables aléatoires qui interviennent sont normées. [2]

Pour α l'espace des variables aléatoires normées définies sur un champ de probabilité donné ($(\forall)$ X $X\in\alpha$, on a donc $M(X) = 0$, $D^2(X) = 1$), on a :

Définition : Soit le vecteur aléatoire $X = (X_1, X_2, \ldots, X_n)$ avec la propriété: $(\forall)$i = 1,2, …,n , on a : $X_i \in\alpha$; $Y\in\alpha$. Les auteurs ont défini le coefficient synthétique de dépendance entre la variable Y et le vecteur aléatoire X comme :

$$SRO_1(X,Y) = \left[\frac{1}{n}\cdot\sum_{i=1}^{n}\mathrm{cov}^2(X_i;Y)\right]^{1/2}.$$

On a pour n =1, $SRO_1(X_1, Y) = |\mathrm{cov}(X_1, Y)|$, donc SRO_1 représente une extension de la notion de coefficient de corrélation, pour le cas de la dépendance variable aléatoire – vecteur aléatoire. Dans le cas où :

$$Y = \sum_{j=1}^{n} \alpha_j X_j \ (\alpha_j \in R, (\forall)\ j), \text{ on obtient:}$$

$$SRO_1^2(X,Y) = \frac{1}{n}\cdot\sum_{i=1}^{n}\mathrm{cov}^2\left(X_i;\sum_{j=1}^{n}\alpha_j X_j\right) = \frac{1}{n}\cdot\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\alpha_j\cdot\mathrm{cov}(X_i,X_j)\right)^2 \qquad (5.1)$$

Avec la notation: $\mathrm{cov}(X_i, X_j) = \sigma_{ij}$ (i, j = 1,2, …,n), la formule (1.1.1) devient:

$$SRO_1^2(X,Y) = \frac{1}{n}\cdot\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\alpha_j\sigma_{ij}\right)^2. \qquad (5.2)$$

La condition que la variable aléatoire $Y = \sum_{j=1}^{n}\alpha_j X_j$ soit normée, se réduit à :

$$D^2(Y) = 1 \iff \sum_{i=1}^{n}\alpha_i^2 + 2\cdot\sum_{\substack{i,j=1\\ i<j}}^{n}\alpha_i\alpha_j\cdot\sigma_{ij} = 1. \qquad (5.3)$$

Le problème de déterminer les coefficients: $\alpha_1, \alpha_2, \ldots, \alpha_n$ optimales devient:

$$(\text{Maxim})SRO_1\left(X;\sum_{j=1}^{n}\alpha_j\cdot X_j\right). \qquad (5.4)$$

Avec : $D^2\left(\sum_{j=1}^{n}\alpha_j X_j\right) = 1$, en ayant pour (1.1.2) et (1.1.3) aussi :

$$(\text{Maxim})\ \mathrm{F}(\alpha_1,\alpha_2,\cdots,\alpha_n) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\alpha_i\sigma_{ij}\right)^2 \qquad (5.5.)$$

Pour résoudre le problème d'extremum, ils ont utilisés la méthode des multiplicateurs du Lagrange, avec la fonction auxiliaire :

$$\Phi(\alpha_1\alpha_2,\ldots,\alpha_n,\lambda) = F(\alpha_1,\alpha_2\ldots,\alpha_n) - \lambda\cdot\left(\sum_{i=1}^{n}\alpha_i^2 + 2\sum_{i<j}\alpha_i\alpha_j\sigma_{ij} - 1\right).$$

On obtient le système:

$$\begin{cases} S^2A - \lambda\cdot SA = 0, & \text{(a)}\\ A'SA = 1. & \text{(b)}\end{cases} \qquad (5.6)$$

Ils ont supposé que la matrice S n'est pas singulière; donc, en amplifiant à gauche (a) avec S^{-1}, on a : $(S-\lambda\mathrm{E})\cdot\mathrm{A} = 0.$ (5.7)

Comme, avec (b), on a $A \neq 0$, on se déduit que : $|S-\lambda\mathrm{E}| = 0$;

Théorème 5.1. Les solutions λ du système (1.1.6) sont les valeurs propres de la matrice S : $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1}$. [2]

Avec $X=(X_1,X_2,\cdots,X_n)$ un vecteur aléatoire, en ayant les propriétés :

$$\mathrm{M}(X_i)=0, D^2(\mathrm{X}_i)=1, (\forall) i=1,2,\ldots,n, \text{ et la matrice:}$$

$$S=\|\sigma_{ij}\|_{i,j=1,2,\ldots,n,}\ , \quad \text{où : } \sigma_{ij}=\mathrm{cov}(X_i,X_j),$$

avec S non singulière, pour déterminer la variable aléatoire normée :
$\hat{\xi}=\hat{\mathrm{A}}'\mathrm{X}, \hat{\mathrm{A}}\in R^*$, pour laquelle le coefficient synthétique de dépendance $SRO_1(X,\xi)$ este *maxime*, les auteurs ont déterminé: [2]

-la plus grande valeur propre de la matrice S, notée λ_0;

-la solution $\hat{\mathrm{A}}$ du système : $\begin{cases}(S-\lambda_Q\mathrm{E}) \cdot \mathrm{A}=0\\ \mathrm{A}'\cdot S\cdot A=1;\end{cases}$;

-donc, on a: $\xi=\hat{\mathrm{A}}'\cdot x; et F(\hat{\mathrm{A}})=\lambda_0$,

et pour $(\forall)$ A $\in R^n$, on a : $\mathrm{SRO}_1(\mathrm{X},\hat{}\zeta)\geq \mathrm{SRO}_1(\mathrm{X},\mathrm{A'X})$. [2]

Les autres valeurs propres de la matrice S admettent des interprétations analogues.

Pour $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1}$ les valeurs propres de la matrice S, le $\hat{}\xi=\xi_0$ est le « substituent synthétique » déduit comme dans le Théorème 5.2. Ils ont trouvé la variable aléatoire ξ =A' X, normée, pour laquelle SRO_2 [ξ, X] est *maxime*. [2]

5.1. Les auteurs ont montré un premier exemple pour la technique de construire le substituent synthétique global, dans le cas d'une variable aléatoire tri-dimensionnelle discrète. Soit X = (X_1,X_2,X_3), le vecteur aléatoire ayant la répartition donnée : [2]

Table 5.1.

Z	X / Y	0	1	2	3	4
1	1	1/115	5/115	0	1/115	1/115
	2	2/115	3/115	1/115	2/115	5/115
	3	1/115	2/115	4/115	5/115	2/115
2	1	3/115	3/115	4/115	1/115	2/115
	2	4/115	2/115	3/115	6/115	1/115
	3	3/115	3/115	2/115	3/115	3/115
3	1	5/115	2/115	3/115	4/115	1/115
	2	1/115	1/115	2/115	1/115	4/115
	3	1/115	5/115	1/115	1/115	6/115

La répartition commune des variables X_1, X_2, se déduit du Table 5.1. Sont calculés aussi les répartitions communes des variables aléatoires X_1,X_3 et X_2 ,X_3 .

Comme: |S| = 0.3433, la matrice S de la corrélation totale est non singulière.

En final, on obtient la matrice de la corrélation totale S et l'équation: $|S-\lambda\cdot E|=0$:

$$S=\begin{pmatrix}1.000 & 0.184 & -0.072\\ 0.184 & 1.000 & 0.670\\ -0.082 & 0.770 & 1.000\end{pmatrix}; \qquad \begin{vmatrix}1-\lambda & 0.184 & -0.082\\ 0.184 & 1-\lambda & 0.770\\ -0.082 & 0.770 & 1-\lambda\end{vmatrix}=0.$$

En calculant la valeur du déterminant, ils ont obtenu l'équation, qui a comme racines les *valeurs propres* de la matrice S:

$$(1-\lambda)^3-0.63348\cdot(1-\lambda)-0.02324=0.$$

Les valeurs approximatives des racines de cette équation sont :

$$\lambda_1 = -0.07769;\ \lambda_2 = -0.03677;\ \lambda_3 = 0.18634.$$

La plus grande valeur propre de la matrice S est donc $\lambda_0 = 0.18634$,

avec : $$\text{SRO}_2\text{'} = \frac{1}{3} \cdot 0.18634 = 0.062\ ;$$

qui montre une faible corrélation ; en final, le «substituent synthétique » est:

$$\xi = 0.12912 \cdot \hat{X}_1 - 0.6181 \cdot \hat{X}_2 + 1.41068 \cdot \hat{X}_3;$$

où : $\hat{X}_1, \hat{X}_2, \hat{X}_3$ sont les variables normées correspondantes aux variables X_1, X_2, X_3 ; ces variables ont les répartitions connues. Ils en résultent les valeurs du substituent synthétique ξ . [2]

Les auteurs ont construit aussi un substituent synthétique pour une mesure non symétrique, $\varphi(X,Y) = \theta(X,Y)$, avec une meilleure corrélation SRO_2'.

Pour l'équation : $\left|S^{-1} \cdot W' \cdot W - \lambda E\right| = 0$, en calculant les valeurs du déterminant, on a l'équation: $$\lambda^3 - 3.105 \cdot \lambda^2 - 0.333 \cdot \lambda + 2.333 = 0\ ;$$

avec les valeurs propres : $\lambda_1 = -0.815;\ \lambda_2 = 0.970;\ \lambda_3 = 2.950$;

La valeur propre la plus grande est : $\lambda_0 = 2.95$, donc : $\text{SRO}_2\text{'} = \frac{1}{3} \cdot \lambda_0 = 0.983$;

ce qui montre une plus grande concordance du nouveau « substituent synthétique » ξ'.

La valeur de ξ' est donnée de l'équation :

$$\xi' = 0.12912 \cdot \hat{X}_1 - 0.6181 \cdot \hat{X}_2 + 1.41068 \cdot \hat{X}_3.$$

Dans notre travail, on assume q'on peut appliquer l'indicateur *SYNSPS* pour les valeurs propres, en module, sorties du modèle pour le «substituent synthétique », mais seulement en tenant compte que *la trace* des valeurs propres est positive (S. Camiz, 2002), car il y a aussi des valeurs propres négatives (acceptables seulement dans un sousespace euclidien). Nous avons utilisé un des plusieurs exemples des auteurs .[2]

D'ailleurs, les auteurs (Dinescu et all., 1986) ont parti d'un espace S des phases, associé à un phénomène économique, avec un système de *n* indicateurs qui caractérisent ce phénomène. Ce système d'indicateurs induit une correspondance entre les éléments de l'espace S et l'espace euclidien n- dimensionnel R^n .

Ils ont utilisé une fonction de synthèse $\psi : R^n \to R$; fonction composée avec l'application précédente, qui peut induire une application entre l'espace des phases S et l'axe réelle R, en réalisant, d'après les auteurs, « un substituent synthétique » aléatoire pour le group des indicateurs aléatoires initiales, sous condition de dépendance maximale. Pour les valeurs propres de la matrice $S^{-1} \cdot W' \cdot W$ on a : [2]

Exemple No. 2:

Ex.3	λ_j	τ_j	cum.	cum.	$1/\lambda_j$	$\tau'j$
1	0.0815	0.1720	-	1.0000	1.2270	0.4720
2	0.0970	0.2050	0.1720	0.5280	1.0310	0.3970
3	2.9500	0.6230	0.3770	0.1310	0.3390	0.1310
	4.7350	1.0000	1.0000	-	2.5970	1.0000

En calculant les indicateurs (1) $E(\tau_j)$ et (2) $E(\tau'_j)$, avec $1< j < 3$, on a :

$$E(\tau_j) = 0.4595 \text{ et } E(\tau'_j) = 0.3980; \text{ où: } k = 3;$$

En calculant aussi l'indicateur *SYNSPS*, on obtient un résultat non-significatif:

$$\{|E(\tau)–E(\tau')|\} = 0.4595 – 0.3990 = 0.0615 < 0.333 = 1/3.$$

Donc, il faut accepter H_0, car l'indicateur synthétique substuitutif calculé ne diffère pas significativement des n indicateurs initiales. [4] [5] [6]

Références

1. Benzécri, J.P. – *L'Analyse des données*, Dunod, Paris, 1980;
2. Dinescu, C., Savulescu, B., Vasiliu, D. – *Metode matemetice pentru fundamentarea deciziilor in productie*; Editura Tehnica, Bucuresti, 1986;
3. Lebart, L., Morineau, A., Piron, M.– *Statistique exploratoire multidimensionnelle*, Dunod, Paris, 1995;
4. Popescu-Spineni Sabina – *Statistical Methods for Multidimensional Data Analysis and Classification*; Editura Universitara "Carol Davila", 2000, (145 pages);
5. Popescu-Spineni Sabina - *Hierarchy Techniques of Multidimensional Data Analysis (MDA) in Social Medicine Research*; Studies in Classification, Data Analysis and Knowledge Organization, Springer Verlag (A. Rizzi, M. Vichi, H.H. Bock Editors), 641-646, 1998;
6. Popescu-Spineni Sabina - *Indicateur informationnel aggrégé pour évaluer la qualité de l'analyse des données sans avoir manque d'information*; Conférence de ALASS, Milano, Aoùt 2006;
7. Rizzi, A., editore – *Some Relation between Matrices and Structures of Multidimensional Data Analysis*; Giardini Editori è Stampatori, Pisa, 1995.