P. CAZES

Problème sur l'analyse des questionnaires

Les cahiers de l'analyse des données, tome 2, n° 1 (1977), p. 73-77

<http://www.numdam.org/item?id=CAD_1977__2_1_73_0>

Les Cahiers de l'Analyse des Données
Vol. II - 1977 - n° 1 - p. 73-77

PROBLÈME SUR L'ANALYSE DES QUESTIONNAIRES

par P. Cazes [1]

Le corrigé du problème présenté ici donne un formulaire sur les contributions d'une modalité, d'une question à la trace de l'analyse des correspondances du tableau disjonctif complet associé à un questionnaire ; ensuite l'on voit comment ces contributions varient dans un questionnaire binaire, quand il y a des non réponses, ces non réponses étant codées (1/2, 1/2) ; enfin dans la dernière partie on montre que la subdivision d'une modalité en deux sous modalités peut conduire à un facteur sans intérêt, car opposant ces deux sous modalités, et nul sur le reste du questionnaire.

1. *Enoncé du problème*

N.B. : La partie C peut être traitée immédiatement après A, indépendamment de B.

Un questionnaire est constitué par un ensemble Q de questions dont chacune, q, admet un ensemble J_q de modalités de réponse. On note :

$$J = \cup \{J_q \mid q \in Q\} ;$$

$$\forall j \in J : q(j) = q \Leftrightarrow j \in J_q ,$$

(i.e. $q(j)$ nous dit à quelle question particulière q, correspond la modalité j) ;

$$\text{Card } J = \overline{\overline{J}} \; ; \; \text{Card } Q = \overline{\overline{Q}} \; ; \; \text{Card } J_q = \overline{\overline{J}}_q \; .$$

Un ensemble I de sujets (Card I = n) répond à ce questionnaire. On suppose dans la partie A que les réponses de tous les sujets sont complètes, i.e. qu'à chaque question q, chaque sujet i fournit une modalité de réponse :

$$\text{rep}(i,q) \in J_q \subset J$$

A partir de ces réponses on constitue le tableau de contingence k_{IJ} :

$$\begin{aligned} k(i,j) &= 1 \text{ si } \text{rep}(i,q(j)) = j \\ &= 0 \text{ si } \text{rep}(i,q(j)) \neq j \end{aligned}$$

(*) *Problème proposé aux étudiants du D.E.A. de Statistique de l'Université Pierre et Marie Curie, à la session de Juin 1974.*

(1) *I.S.U.P. - Laboratoire de Statistique - Université Pierre et Marie Curie.*

1.A 1°) Calculer $k = \Sigma\{k(i,j) \mid i \in I,\ j \in J\}$.
On considère désormais le tableau $p_{IJ} = \{k(i,j)/k\}$.

1.A 2°) Calculer la loi marginale p_I ;
Calculer $\|p_I^j - p_I\|^2_{p_I}$, c'est-à-dire le carré de la distance entre p_I et le profil p_I^j dans la métrique du χ^2 de centre p_I . On exprimera $\|p_I^j - p_I\|^2_{p_I}$ uniquement en fonction de $r_j = k(j)/n$ (où $k(j) = \Sigma\{k(i,j) \mid i \in I\}$).

1.A 3°) Exprimer en fonction de r_j et de $\bar{\bar{Q}}$ la contribution $CR(j)$ de j à l'inertie du nuage $\mathcal{N}(J)$ associé à la correspondance p_{IJ}.

1.A 4°) Quelle est la contribution $CR(q)$ apportée à l'inertie de $\mathcal{N}(J)$ par l'ensemble J_q des modalités de réponse j à la question q ? Quelle est l'inertie totale $CR(J)$ du nuage $\mathcal{N}(J)$. On exprimera cette inertie totale en fonction de $\bar{\bar{J}}$ et $\bar{\bar{Q}}$.

1.B 1°) On suppose désormais dans la partie B que le questionnaire Q ne comprend que des questions à deux modalités de réponses :

$$\forall\, q \in Q : J_q = \{q^+, q^-\} \ ; \ \mathrm{Card}\, J_q = 2.$$

On admet la possibilité de non réponse ; en sorte que l'on pourra avoir :

$$rep(i,q) = q^+ \quad ; \quad rep(i,q) = q^- \quad ; \quad rep(i,q) = \emptyset ,$$

mais les non réponses sont codées par (1/2), (1/2) sur les deux colonnes q^+, q^-, suivant la règle :

$$\begin{aligned} k(i,q^+) &= 1, \text{ si } rep(i,q) = q^+ \\ &= 1/2, \text{ si } rep(i,q) = \emptyset \\ &= 0, \text{ si } rep(i,q) = q^- \\ k(i,q^-) &= 0, \text{ si } rep(i,q) = q^+ \\ &= 1/2, \text{ si } rep(i,q) = \emptyset \\ &= 1, \text{ si } rep(i,q) = q^-. \end{aligned}$$

On note :

$$\begin{aligned} r^+(q) &= \mathrm{Card}\{i \mid i \in I,\ rep(i,q) = q^+\}/n \\ r^\circ(q) &= \mathrm{Card}\{i \mid i \in I,\ rep(i,q) = \emptyset\}/n \\ r^-(q) &= \mathrm{Card}\{i \mid i \in I,\ rep(i,q) = q^-\}/n \\ r(q) &= r^+(q) + (r^\circ(q)/2) \end{aligned}$$

Calculer la contribution $CR(q^+)$ du point q^+ à l'inertie du nuage $\mathcal{N}(J)$. On exprimera cette contribution en fonction de $r(q)$, de $r^\circ(q)$ et de $\bar{\bar{Q}}$.

1.B 2°) Exprimer en fonction de $r(q)$ et de $r^\circ(q)$ la contribution $CR(q)$ apportée à l'inertie de $\mathcal{N}(J)$ par l'ensemble J_q (i.e. par les deux modalités q^+ et q^-).

On exprimera cette contribution en fonction de $r(q)$, de $r^\circ(q)$ et de $\bar{\bar{Q}}$.

1.B 3°) Quelle est l'inertie totale $CR(J)$ du nuage $\mathcal{N}(J)$; on exprimera cette inertie en fonction des $r(q)$ et des $r^\circ(q)$. Les quantités $r^\circ(q)$ étant supposées fixées, quelle est la valeur maximale possible de l'inertie totale? Pour quelles valeurs des $r(q)$ est-elle atteinte.

1.C L'objet de cette dernière partie est d'étudier l'influence de la subdivision d'une modalité dans un tableau mis sous forme disjonctive complète. On rappelle les notations usuelles :

$$\forall\, j, j' \in J : t(j,j') = \Sigma\{k(i,j)\ k(i,j') \mid i \in I\}$$

(on suppose comme dans la partie A que le tableau k ne comporte aucune omission)

$$t_{j'}^{j} = t(j,j')/(\bar{\bar{Q}}\, t(j,j)) ;$$

$$t_J^J = \{t_{j'}^{j} \mid j,j' \in J\}$$

$$t_J^J = p_J^I \circ p_I^J$$

Si φ^J est un facteur issu de k_{IJ} , relatif à la valeur propre λ on a :

$$\varphi^J \circ t_J^J = \lambda\, \varphi^J ;$$

dans la partie C, on utilisera cette définition des facteurs.

Ceci posé, on suppose qu'une modalité de réponse $z \in J$ est subdivisée en deux z', z" ; et que le partage des réponses entre z' et z" se fait comme suit. Notons :

$$Jd = (J - \{z\}) \cup \{z', z''\},$$

alors le nouveau tableau de contingence, td sur $Jd \times Jd$ est :

$$\forall\, j, j' \in J - \{z\} \ : td(j,j') = t(j,j').$$

$$\forall\, j \in J - \{z\} : td(j,z') = td(j,z'') = t(j,z)/2 .$$

$$td(z',j) = td(z'',j) = t(z,j)/2 = t(j,z)/2.$$

$$td(z', z'') = td(z'', z') = 0$$

$$td(z', z') = td(z'', z'') = t(z,z)/2$$

A partir de t_d on définit la transition td_{Jd}^{Jd} :

$$\forall\, j, j' \in Jd : td_{j'}^{j} = td(j,j')/(\bar{\bar{Q}}\ td(j,j))$$

et on étudie les facteurs φ^{Jd} tels que $\varphi^{Jd} \circ td_{Jd}^{Jd} = \lambda\, \varphi^{Jd}$

1.C 1°) Montrer que td admet un facteur φ^{Jd} nul en dehors de {z', z"} ; quel est ce facteur ? A quelle valeur propre est-il relatif ?

1.C 2°) Déduire des résultats de l'analyse de t_J^J les autres facteurs et valeurs propres issus de td_{Jd}^{Jd} .

1.C 3°) Vérifier que la différence entre la somme des valeurs propres de td_{Jd}^{Jd} et celle de t_J^J est conforme aux résultats de A 4°.

2. *Solution du problème*

2.A 1°) : $k = n\,\bar{\bar{Q}}$

2.A 2°) $\forall\, i \in I : p_i = 1/n$

$$\forall\, j \in J : \|p_I^j - p_I\|^2_{p_I} = (1/r_j) - 1$$

2.A 3°) $C\,R(j) = (1 - r_j)/\bar{\bar{Q}}$

2.A 4°) $C\,R(q) = \Sigma\{C\,R(j) \mid j \in J_q\} = (\bar{\bar{J}}_q - 1)/\bar{\bar{Q}}$

$$C\,R(J) = \Sigma\{C\,R(q) \mid q \in Q\} = (\bar{\bar{J}} - \bar{\bar{Q}})/\bar{\bar{Q}} = (\bar{\bar{J}}/\bar{\bar{Q}}) - 1 \qquad (1)$$

Remarque :

On voit que si on a une subdivision de plus en plus fine des questions (cas interessant que si l'on a un nombre infini de sujets), à la limite, quand $\bar{\bar{J}}$ tend vers l'infini, $C\,R(J)$ est infini, et donc tous les pourcentages d'inertie sont nuls ; c'est la raison pour laquelle il vaut mieux considérer les valeurs propres (qui sont les carrés des valeurs propres associées à k_{IJ}) la trace et les pourcentages d'inertie issus du tableau de Burt construit à partir de k_{IJ} (cf [Bin. Mult.], ce cahier, § 2).

2.B 1°) $C\,R(q^+) = (1 - r(q) - r°(q)/(4\,r(q)))/\bar{\bar{Q}}$

2.B 2°) La contribution $C\,R(q^-)$ de q^- à l'inertie s'obtenant en remplaçant dans l'expression de $C\,R(q^+)$, $r(q)$ par $1 - r(q)$, la contribution totale apportée par la question q vaut :

$$C\,R(q) = C\,R(q^+) + C\,R(q^-) = (1 - r°(q)/(4\,r(q)\,(1 - r(q))))/\bar{\bar{Q}}$$

2.B 3°) $C\,R(J) = \Sigma\{C\,R(q) \mid q \in Q\}$

$$= 1 - (1/(4\bar{\bar{Q}}))\,\Sigma\{r°(q)/(r(q)\,(1 - r(q))) \mid q \in Q\} \qquad (2)$$

Les quantités $r°(q)$ étant fixées, $C\,R(J)$ sera maximum si $\forall\, q \in Q$: $r(q) = 1 - r(q) = 1/2$, la valeur maximale de $C\,R(J)$ étant égale à $1 - \Sigma\{r°(q) \mid q \in Q\}/\bar{\bar{Q}}$.

Cette inertie maximale est donc plus faible que celle que l'on aurait obtenue s'il n'y avait pas eu de non réponses, auquel cas $C\,R(J)$ vaut 1, résultat que l'on peut déduire soit de (1) où $\bar{\bar{J}} = 2\,\bar{\bar{Q}}$, soit de (2) où $\forall\, q \in Q : r°(q) = 0$.

2.C 1°) De la définition de td, et de td_{Jd}^{Jd}, l'on déduit :

$$\forall\, j, j' \in J - \{z\} : td_{j'}^{j} = t_{j'}^{j}$$

$$\forall\, j \in J - \{z\} : td_{z'}^{j} = (1/2)\;t_z^j = td_{z''}^{j}$$

$$td_j^{z'} = t_j^z = td_j^{z''}$$

$$td_{z''}^{z'} = td_{z'}^{z''} = 0$$

$$td_{z'}^{z'} = td_{z''}^{z''} = t_z^z = 1/\bar{\bar{Q}}$$

Il est immédiat de vérifier que td_{Jd}^{Jd} est le tableau des lois conditionnelles associé à td, ce qui justifie l'équation des facteurs de td donnée dans l'énoncé.

Si φ^{Jd} est l'un de ces facteurs, et si l'on pose $J_1 = J - \{z\}$, $\varphi^{Jd} = (\varphi^{J_1}, \varphi^{z'}, \varphi^{z''})$, on a donc :

$$\left.\begin{aligned}
&\varphi^{J_1} \circ t_{J_1}^{J_1} + (1/2)\ (\varphi^{z'} + \varphi^{z''})\ t_z^{J_1} &&= \lambda d\ \varphi^{J_1}\\
&\varphi^{J_1} \circ t_{J_1}^{z} + \varphi^{z'}\ t_z^{z} &&= \lambda d\ \varphi^{z'}\\
&\varphi^{J_1} \circ t_{J_1}^{z} + \varphi^{z''}\ t_z^{z} &&= \lambda d\ \varphi^{z''}
\end{aligned}\right\} (3)$$

On en déduit que $\varphi^{Jd} = (O^{J_1}, 1, -1)$ où O^{J_1} désigne la fonction nulle sur J_1, est facteur de td relatif à la valeur propre

$$\lambda d = t_z^z = 1/\bar{\bar{Q}}.$$

2.C 2°) Les autres facteurs de td étant non corrélés au facteur trouvé ci-dessus sont de la forme, puisque td(z',z') = td(z",z"): $\varphi^{Jd} = (\varphi^{J_1}, \varphi^z, \varphi^z)$.

Posant $\varphi^J = (\varphi^{J_1}, \varphi^z)$, ces facteurs vérifient d'après (3) les équations :

$$\varphi^J \circ t_J^J = \lambda d\ \varphi^J$$

On retrouve ainsi les facteurs de t (ou de k_{IJ}).

En résumé, les $\bar{\bar{J}} + 1$ facteurs de td se décomposent de la façon suivante :

- le facteur trivial
- le facteur $\varphi^{Jd} = (O^{J_1}, 1, -1)$ associé à la valeur propre $1/\bar{\bar{Q}}$.
- les $\bar{\bar{J}} - 1$ facteurs $(\varphi^{J_1}, \varphi^z, \varphi^z)$ associés aux valeurs propres $\lambda d = \lambda$, et correspondant aux $\bar{\bar{J}} - 1$ facteurs non triviaux $\varphi^J = (\varphi^{J_1}, \varphi^z)$ de t, associés aux valeurs propres λ.

2.C 3°) La différence entre la somme des valeurs propres de td_{Jd}^{Jd} et celle de t_J^J est égale à $1/\bar{\bar{Q}}$, ce qui correspond bien à la formule (1) du § 2.A 4°, où le passage de t à td augmente $\bar{\bar{J}}$ de 1, $\bar{\bar{Q}}$ restant inchangé.