

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](http://www.numdam.org/item?id=CAD_1977__2_1_73_0)

© Les cahiers de l'analyse des données, Dunod, 1977, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

## 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. Énoncé 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) ;

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

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

1.A 2°) Calculer la loi marginale  $p_I$  ;  
Calculer  $\|p_I^j - p_I\|_{p_I}^2$ , c'est-à-dire le carré de la distance entre

$p_I$  et le profil  $p_I^j$  dans la métrique du  $\chi^2$  de centre  $p_I$ . On exprimera  $\|p_I^j - p_I\|_{p_I}^2$  uniquement en fonction de  $r_j = k(j)/n$  (où  $k(j) = \sum \{k(i,j) \mid i \in I\}$ ).

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

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

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

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

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

$$\text{rep}(i, q) = q^+ ; \text{rep}(i, q) = q^- ; \text{rep}(i, q) = \emptyset,$$

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

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

On note :

$$\begin{aligned} r^+(q) &= \text{Card}\{i \mid i \in I, \text{rep}(i, q) = q^+\} / n \\ r^0(q) &= \text{Card}\{i \mid i \in I, \text{rep}(i, q) = \emptyset\} / n \\ r^-(q) &= \text{Card}\{i \mid i \in I, \text{rep}(i, q) = q^-\} / n \\ r(q) &= r^+(q) + (r^0(q)/2) \end{aligned}$$

Calculer la contribution  $CR(q^+)$  du point  $q^+$  à l'inertie du nuage  $N(J)$ . On exprimera cette contribution en fonction de  $r(q)$ , de  $r^0(q)$  et de  $\bar{Q}$ .

1.B 2°) Exprimer en fonction de  $r(q)$  et de  $r^0(q)$  la contribution  $CR(q)$  apportée à l'inertie de  $N(J)$  par l'ensemble  $J_q$  (i.e. par les deux modalités  $q^+$  et  $q^-$ ).

On exprimera cette contribution en fonction de  $r(q)$ , de  $r^0(q)$  et de  $\bar{Q}$ .

1.B 3°) Quelle est l'inertie totale  $CR(J)$  du nuage  $N(J)$  ; on exprimera cette inertie en fonction des  $r(q)$  et des  $r^0(q)$ . Les quantités  $r^0(q)$  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 :

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

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

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

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

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

Si  $\varphi^J$  est un facteur issu de  $k_{IJ}$ , relatif à la valeur propre  $\lambda$  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 :

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

alors le nouveau tableau de contingence,  $td$  sur  $J_d \times J_d$  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  $td_{J_d}^{J_d}$  :

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

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

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

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

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

2. Solution du problème

2.A 1°) :  $k = n \bar{Q}$

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

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

2.A 3°)  $CR(j) = (1 - r_j)/\bar{Q}$

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

$$CR(J) = \Sigma\{CR(q) \mid q \in Q\} = (\bar{J} - \bar{Q})/\bar{Q} = (\bar{J}/\bar{Q}) - 1 \quad (1)$$

Remarque :

On voit que si on a une subdivision de plus en plus fine des questions (cas intéressant que si l'on a un nombre infini de sujets), à la limite, quand  $\bar{J}$  tend vers l'infini,  $CR(J)$  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°)  $CR(q^+) = (1 - r(q) - r^o(q)/(4r(q)))/\bar{Q}$

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

$$CR(q) = CR(q^+) + CR(q^-) = (1 - r^o(q)/(4r(q)(1 - r(q))))/\bar{Q}$$

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

$$= 1 - (1/(4\bar{Q})) \Sigma\{r^o(q)/(r(q)(1 - r(q))) \mid q \in Q\} \quad (2)$$

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

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  $CR(J)$  vaut 1, résultat que l'on peut déduire soit de (1) où  $\bar{J} = 2\bar{Q}$ , soit de (2) où  $\forall q \in Q : r^o(q) = 0$ .

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

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

$$\forall j \in J - \{z\} : td_z^j = (1/2) t_z^j = td_z^j$$

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

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

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

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  $\varphi^{Jd}$  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

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

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  $k_{JJ}$ ).

En résumé, les  $\bar{J} + 1$  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  $1/\bar{Q}$ .
- les  $\bar{J} - 1$  facteurs  $(\varphi^{J_1}, \varphi^{z'}, \varphi^{z''})$  associés aux valeurs propres  $\lambda d = \lambda$ , et correspondant aux  $\bar{J} - 1$  facteurs non triviaux  $\varphi^J = (\varphi^{J_1}, \varphi^z)$  de  $t$ , associés aux valeurs propres  $\lambda$ .

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