Techniques d'enquête, décembre 2005
Vol. 31, N° 2, pp. 185-200
Statistique Canada, N° 12-001-XPB au catalogue

Fonctions de variance-covariance pour les moyennes de domaine des questions avec valeurs ordonnées

Alistair James O'Malley et Alan Mark Zaslavsky [1]

Résumé

De nombreuses analyses statistiques, particulièrement l'analyse multiniveaux, requièrent l'estimation d'une matrice des variances-covariances d'échantillonnage. Dans le cas de problèmes univariés, des fonctions reliant la variance à la moyenne ont été utilisées pour obtenir des estimations de la variance, en regroupant l'information sur l'ensemble des unités ou des variables. Nous présentons des fonctions de variance et de corrélation pour des moyennes multivariées de questions d'enquête avec valeurs ordonnées, pour des données complètes, ainsi que pour des données avec non-réponse structurée. Nous élaborons aussi des méthodes permettant d'évaluer l'ajustement du modèle et de calculer des estimateurs composites qui combinent des prédictions directes et fondées sur un modèle. Nous utilisons des données d'enquête provenant de la Consumer Assessments of Health Plans Study (CAHPS®) pour illustrer l'application de la méthodologie.

Mots clés : Fonction de variance; fonction de corrélation; modèle hiérarchique; réponse ordonnée, non-réponse; enchaînement de questions.

1. Introduction

Les données d'enquête sont souvent utilisées pour obtenir des mesures permettant de faire des comparaisons entre domaines d'estimation. Dans notre exemple pratique, des enquêtes sont réalisées pour recueillir des déclarations sur les expériences vécues en ce qui concerne les régimes d'assurance-maladie (entités qui administrent les soins de santé) auprès des membres inscrits; de même, une enquête pourrait être conçue pour évaluer les écoles en faisant passer des tests à un échantillon d'élèves.

Une part essentielle de l'analyse des données d'enquête est le calcul des variances d'échantillonnage ou de la matrice des covariances d'échantillonnage d'un estimateur multivarié. L'approche type en échantillonnage consiste à calculer les variances directement pour chaque estimateur dans chaque domaine. Les estimations directes de la variance peuvent être instables si le nombre de répondants à une question est faible parce que la taille de l'échantillon pour un domaine est petite, vu que la question s'applique seulement à une fraction des répondants (comme les utilisateurs d'équipement spécialisé dans les enquêtes sur la santé), ou parce que nous souhaitons calculer les moyennes pour un petit sous-groupe (comme les personnes atteintes de maladie chronique).

En modélisant les estimations de la variance sous forme de fonctions des moyennes d'unité (domaine), nous pouvons regrouper l'information sur l'ensemble des unités pour obtenir des estimations plus stables. Bien que la modélisation puisse introduire un biais, pour les petites unités, ce problème est compensé par la réduction de la variation d'échantillonnage. On peut aussi envisager de généraliser les estimations de la variance sur l'ensemble des questions en plus des domaines, ou à la place de ceux-ci. Cette approche convient lorsqu'il existe des groupes de questions pour lesquelles il est probable que la même relation moyenne-variance soit vérifiée. Cependant, si le nombre de domaines est beaucoup plus grand que celui des questions, l'amélioration éventuelle la plus importante s'obtient en généralisant sur l'ensemble des domaines plutôt que sur l'ensemble des questions.

Une *fonction de variance généralisée* (FVG) est un modèle mathématique qui décrit la relation entre la variance ou la variance relative d'un estimateur et son espérance. Si plusieurs estimations sont produites d'après le même échantillon, Wolter (1985, chapitre 5) propose le modèle

$$V / M^2 = \theta_0 + \theta_1 / M,$$

où M et V représentent la valeur prévue et la variance de l'estimateur, respectivement. Une forme de ce genre pourrait convenir pour des variables, comme le revenu ou la richesse, pour lesquelles un coefficient de variation quasi constant serait plausible, parce que la moyenne et l'écart-type sont proportionnels à la longueur de la période de référence. La modélisation du coefficient de variation est donc pertinente surtout dans les situations où les variables ont un contenu semblable, mais des échelles différentes d'étendue non restreinte (par exemple, données sur le revenu recueillies mensuellement et annuellement). Dans notre problème, les questions ont des valeurs ordonnées, si bien qu'un modèle du coefficient de variation n'est pas un

1. Alistair James O'Malley et Alan Mark Zaslavsky, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115-5899, États-Unis. Courriel : omalley@hcp.med.harvard.edu et zaslavsk@hcp.med.harvard.edu.

choix naturel. D'autres FVG proposées ont également une forme simple (Woodruff 1992; Otto et Bell 1995).

Trouver une FVG appropriée peut simplifier les calculs et rendre les estimations de la variance plus stables. En outre, résumer les estimations de la variance d'échantillonnage sous la forme d'une fonction facilite la présentation de grandes quantités de statistiques (Wolter 1985, pages 201–202). Enfin, modéliser les variances sous forme de fonctions des moyennes facilite la réestimation itérative des variances d'échantillonnage en cas de modélisation hiérarchique. En pratique, la décision d'utiliser des fonctions de variance dans un contexte de modélisation hiérarchique dépend de la qualité de l'ajustement de la FVG; l'utilisation de cette dernière ne vaut la peine que si l'ajustement est suffisamment bon.

Les études antérieures sur les FVG sont assez rares. Wolter (1985, chapitre 5) donne un aperçu, mais ne fournit que quelques références, comme le font aussi Valliant, Dorfman et Royall (2000, pages 344 à 348). Valliant (1992a, 1992b) utilise des FVG pour lisser des indices variables en fonction du temps dans les analyses de séries chronologiques. Woodruff (1992) utilise des FVG pour estimer la variance de la variation de l'emploi dans la Current Employment Survey, et Wolter (1985, pages 208 à 217) illustre l'utilisation des FVG au moyen de données provenant de la Current Population Survey. Des FVG sont également utilisées dans la National Health Interview Survey (Valliant et coll. 2000, page 344).

Huff, Eltinge et Gershunskaya (2002), ainsi que Cho, Eltinge, Gershunskaya et Huff (2002) considèrent l'utilisation de FVG pour la Current Employment Survey et la Consumer Expenditure Survey réalisées aux États-Unis. Eltinge (2002) utilise des FVG pour estimer une matrice des covariances d'échantillonnage complète, lorsque les échantillons sont trop petits pour produire des estimations stables pour toutes les régions, et estime les composantes de l'erreur quadratique moyenne (MSE pour *mean squared error*) du modèle de FVG. Otto et Bell (1995) ajustent des FVG au revenu médian, au revenu par habitant et au taux de pauvreté selon le groupe d'âge dans la Current Population Survey, en supposant que l'interdépendance des taux au cours du temps est autorégressive et que les matrices des covariances d'échantillonnage suivent une loi de Wishart.

Notre étude prolonge les travaux antérieurs sur les FVG dans quatre directions. En premier lieu, nous utilisons la FVG pour faire une généralisation sur l'ensemble des domaines plutôt que sur l'ensemble des questions. Donc, nous ne supposons pas que les diverses questions ont la même FVG, bien qu'il pourrait être raisonnable d'ajuster des modèles de la même forme pour les questions ayant des catégories de réponses semblables. En deuxième lieu, nous élaborons des FVG pour la matrice des covariances complète, qui doivent être estimées pour une inférence conjointe sur plusieurs résultats. En troisième lieu, nous nous concentrons sur la relation entre les moyennes et les variances des questions à format de réponse ordonné souvent utilisées dans les questionnaires d'enquête, plutôt que sur des réponses continues homoscédastiques. Enfin, nous tenons compte explicitement des profils de non-réponse due à des enchaînements de questions structurés. Alors qu'on peut ignorer la non-réponse partielle structurée (sauf son effet sur la taille d'échantillon) dans le cas de l'estimation univariée, il faut en tenir compte explicitement pour modéliser les relations bivariées, parce qu'elle a une incidence sur la covariance d'échantillonnage des moyennes des questions. De surcroît, comme le nombre de réponses varie selon la question, nous ne pouvons modéliser les covariances d'échantillonnage au moyen de la loi de Wishart, qui ne possède qu'un seul paramètre pour la taille d'échantillon.

Nous commençons par décrire l'estimation directe des variances et des covariances, y compris le cas où des données manquent à cause d'enchaînements de questions. À la section 3, nous présentons des modèles pour les fonctions de variance et de covariance généralisées (FVCG) et nous exposons nos stratégies d'ajustement et d'évaluation de modèles, et de combinaison d'estimations directes et de prédictions par modèle. À la section 4, nous appliquons nos méthodes à une grande enquête sur les soins de santé. À la section 5, pour conclure, nous décrivons des applications et des extensions de nos méthodes.

2. Estimations directes des variances d'échantillonnage des moyennes de domaine

Nous indiçons les observations par domaine (indice h), par question (indices i et j) et par répondant (indices k et l); $y_{h,ik}$ et $r_{h,ik}$ représentent le résultat et l'indicateur de réponse du sujet k dans le domaine h pour la question i. Nous supprimons l'indice inférieur de question quand nous faisons référence à l'ensemble des questions pour un répondant ou un domaine, et nous n'avons pas besoin d'utiliser d'indice inférieur de répondant quand nous discutons des moyennes, des variances et des corrélations de questions.

L'estimation directe de la matrice des covariances d'échantillonnage des moyennes de domaine (donc, « estimation de variance ») débute par l'expression des moyennes sous forme de fonctions des totaux des résultats et des indicateurs de réponse. Nous remplaçons $y_{h,ik}$ par 0 pour les observations manquantes, de sorte que les totaux soient définis en présence d'enchaînements de questions. Suivant la notation de Särndal, Swenson et Wretman (1992, pages 24 à 28; 36 à 42), soit U_h et S_h la population et

l'échantillon, respectivement, pour le h^{e} domaine, $Y_{h,i} = \sum_{U_h} y_{h,ik}$, $R_{h,i} = \sum_{U_h} r_{h,ik}$, $\hat{Y}_{h,i} = \sum_{S_h} \breve{y}_{h,ik}$, et $\hat{R}_{h,i} = \sum_{S_h} \breve{r}_{h,ik}$, où $\breve{y}_{h,ik} = y_{h,ik} / \pi_{h,k}$, $\breve{r}_{h,ik} = r_{h,ik} / \pi_{h,k}$, et $\pi_{h,k} = \text{pr}(k \in S_h)$.

Le vecteur des résultats moyens pour la population d'éléments compris dans le domaine *h* est

$$M_h = f(Y_h, R_h) = \left(\frac{Y_{h,1}}{R_{h,1}}, \ldots, \frac{Y_{h,I}}{R_{h,I}} \right),$$

où $Y_h = (Y_{h,1}, \ldots, Y_{h,I})$ et $R_h = (R_{h,1}, \ldots, R_{h,I})$. Un estimateur est donné par

$$f(\hat{Y}_h, \hat{R}_h) = \left(\frac{\hat{Y}_{h,1}}{\hat{R}_{h,1}}, \ldots, \frac{\hat{Y}_{h,I}}{\hat{R}_{h,I}} \right).$$

Un développement en série de Taylor de premier ordre de $f(\hat{Y}_h, \hat{R}_h)$ autour de $f(Y_h, R_h)$ produit l'approximation

$$\text{var}(f(\hat{Y}_h, \hat{R}_h)) \approx V_h = f'(Y_h, R_h)\, \text{var}(\hat{Y}_h, \hat{R}_h)\, f'(Y_h, R_h)^T,$$

où $f'(Y_h, R_h)$ est le Jacobien de $f(Y_h, R_h)$. Souvent, il est informatiquement plus facile de commencer par calculer $u_{h,k} = f'(Y_h, R_h) z_{h,k}$, où $z_{h,k} = (y_{h,k}, r_{h,k})$, puis d'évaluer la variance sous la forme

$$\begin{aligned} V_h &= \text{var}\left(\sum_{S_h} \breve{u}_{h,k} \right) \\ &= \text{var}\left(\sum_{U_h} \breve{u}_{h,k} I_{h,k} \right) \\ &= \sum_{k,l \in U_h} \Delta_{h,kl} \breve{u}_{h,k} \breve{u}_{h,l}^T, \end{aligned}$$

où $I_{h,k} = 1$ si $k \in S_h$ (indiquant que le k^{e} membre du domaine *h* est échantillonné) et 0 autrement, $\Delta_{h,kl} = \pi_{h,kl} - \pi_{h,k} \pi_{h,l}$, et $\pi_{h,kl} = \text{pr}(k, l \in S_h)$. Un estimateur de V_h est

$$\hat{V}_h = \sum_{k,l \in S_h} \breve{\Delta}_{h,kl} \breve{u}_{h,k} \breve{u}_{h,l}^T, \tag{1}$$

où $\breve{\Delta}_{h,kl} = \Delta_{h,kl} / \pi_{h,kl}$.

Pour décrire l'évaluation de $\hat{V}_h$, nous ne devons considérer qu'un seul élément diagonal (c'est-à-dire, variance) et un seul élément hors diagonale (c'est-à-dire, covariance). La sous-matrice du Jacobien formée par les i^{e} et j^{e} questions est donnée par

$$f'(Y_h, R_h) = \begin{pmatrix} \dfrac{1}{R_{h,i}} & 0 & -\dfrac{Y_{h,i}}{R_{h,i}^2} & 0 \\ 0 & \dfrac{1}{R_{h,j}} & 0 & -\dfrac{Y_{h,j}}{R_{h,j}^2} \end{pmatrix}.$$

Pour $z_{h,k} = (y_{h,ik}, y_{h,jk}, r_{h,ik}, r_{h,jk})$, il s'ensuit que

$$u_{h,k} = f'(Y_h, R_h) z_{h,k} = \begin{pmatrix} \dfrac{1}{R_{h,i}} (y_{h,ik} - M_{h,i} r_{h,ik}) \\ \dfrac{1}{R_{h,j}} (y_{h,jk} - M_{h,j} r_{h,jk}) \end{pmatrix},$$

où $M_{h,i} = Y_{h,i} / R_{h,i}$ est le résultat moyen de la i^{e} question dans le domaine *h*. Donc,

$$\hat{V}_{h,ii} = \frac{1}{R_{h,i}^2} \sum_{k,l \in S_h} \breve{\Delta}_{h,kl} (\breve{y}_{h,ik} - M_{h,i} \breve{r}_{h,ik})(\breve{y}_{h,il} - M_{h,i} \breve{r}_{h,il}) \tag{2}$$

et

$$\hat{V}_{h,ij} = \frac{1}{R_{h,i} R_{h,j}} \sum_{k,l \in S_h} \breve{\Delta}_{h,kl} (\breve{y}_{h,ik} - M_{h,i} \breve{r}_{h,ik}) \times (\breve{y}_{h,jl} - M_{h,j} \breve{r}_{h,jl}). \tag{3}$$

Pour évaluer (2) et (3), nous faisons une autre approximation en substituant $\hat{R}_{h,i} = \sum_{S_h} \breve{r}_{h,ik}$ et $\hat{M}_{h,i} = \sum_{S_h} \breve{y}_{h,ik} / (\sum_{S_h} \breve{r}_{h,ik})$ à $R_{h,i}$ et $M_{h,i}$.

Si les taux d'échantillonnage sont faibles ou que nous souhaitons faire des prédictions pour une grande superpopulation (par exemple, tous les participants possibles à un régime d'assurance-maladie plutôt que ceux couramment inscrits seulement), $\breve{\Delta}_{h,kl} = 1 - \pi_{h,k} \approx 1$ si $k = l$, $\breve{\Delta}_{h,kl} \approx 0$ si $k \neq 1$, et le plan d'échantillonnage s'approche de l'échantillonnage avec remise. Sous le plan d'échantillonnage avec remise, les estimateurs approximativement sans biais sont

$$\hat{V}_{h,ii} = \frac{1}{\hat{R}_{h,i}^2} \sum_{k \in S_h} (\breve{y}_{h,ik} - \hat{M}_{h,i} \breve{r}_{h,ik})^2 \tag{4}$$

et

$$\hat{V}_{h,ij} = \frac{1}{\hat{R}_{h,i} \hat{R}_{h,j}} \sum_{k \in S_h} (\breve{y}_{h,ik} - \hat{M}_{h,i} \breve{r}_{h,ik})(\breve{y}_{h,jk} - \hat{M}_{h,j} \breve{r}_{h,jk}). \tag{5}$$

Ces estimateurs peuvent être généralisés pour prendre en compte la mise en grappes.

En cas d'échantillonnage avec probabilités égales dans les domaines, (4) et (5) se réduisent à

$$\hat{V}_{h,ii} = \frac{1}{\hat{R}_{S_{h,i}}^2} \sum_{k \in S_h} (y_{h,ik} - \hat{M}_{h,i} r_{h,ik})^2 \tag{6}$$

et

$$\hat{V}_{h,ij} = \frac{1}{\hat{R}_{S_{h,i}} \hat{R}_{S_{h,j}}} \sum_{k \in S_h} (y_{h,ik} - \hat{M}_{h,i} r_{h,ik})(y_{h,jk} - \hat{M}_{h,j} r_{h,jk}), \tag{7}$$

où $\hat{R}_{S_{h,i}}$ est le nombre de répondants à la question *i* dans le domaine *h*.

3. Modèles pour les fonctions de variance

À la présente section, nous proposons des spécifications de modèle pour les variances et pour les corrélations d'échantillon pour les réponses complètes ou pour celles avec enchaînements de questions structurés. Puis, nous discutons des stratégies d'ajustement et d'évaluation des modèles. Nous supposons que ces domaines sont des strates non chevauchantes, de sorte que les erreurs d'échantillonnage pour divers domaines soient indépendantes.

Nous transformons les évaluations ordonnées pour les amener à l'intervalle $[0,1]$ par la transformation $p_{h,i} = (B_{h,i} - M_{h,i})/(B_{h,i} - A_{h,i})$, où $A_{h,i}$ et $B_{h,i}$ sont les catégories de réponse minimale et maximale pour la question *i* dans le domaine *h*, respectivement. Nous nous concentrons sur la modélisation des variances pour les grandes valeurs de $M_{h,i}$ (petites valeurs de $p_{h,i}$) parce que, dans l'exemple que nous avons choisi, les résultats moyens sont habituellement proches de l'extrémité supérieure de l'échelle.

3.1 Fonctions de variance

Afin de tenir compte du nombre variable de répondants sur les domaines et les questions, et des différences d'échelle, nous normalisons les estimateurs de variance donnés par (6) pour la taille d'échantillon et faisons un rééchelonnement :

$$\tilde{V}_{h,ii} = \frac{\hat{R}_{S_{h,i}} \hat{V}_{h,ii}}{(B_{h,i} - A_{h,i})^2}.$$

Sous échantillonnage avec probabilités inégales dans les domaines, nous pourrions utiliser un facteur de normalisation qui tient compte des pondérations. Une normalisation possible consiste à multiplier $\hat{V}_{h,ii}$ par $\hat{R}^*_{S_{h,i}} = (\sum \breve{r}_{h,ik})^2 / (\sum \breve{r}^2_{h,ik})$, où $\breve{r}_{h,ik}$ est l'indicateur de réponse à la question *i* pour le k^{e} sujet dans le h^{e} domaine, à la place de $\hat{R}_{S_{h,i}}$. Cette approximation, proposée par Kish (1965), possède une justification fondée sur un modèle (Gabler, Haeder et Lahiri 1999). Elle donne de bons résultats si les probabilités d'échantillonnage varient moyennement dans l'échantillon, mais peut être inefficace si la variation est excessive (Korn et Graubard 1999, page 173; Spencer 2000).

Comme, dans notre exemple, les questions ont des valeurs ordonnées, la variance doit tendre vers 0 quand $p_{h,i} \to 0$ ou $p_{h,i} \to 1$. Un prédicteur ayant manifestement cette propriété est la fonction de variance de la loi de Bernoulli, $p_{h,i}(1 - p_{h,i})$. Celle-ci est vérifiée exactement pour les questions dichotomiques et pourrait être une approximation utile pour les questions comportant au moins trois catégories de réponse.

Comme autres solutions que le modèle de variance de Bernoulli, nous considérons des modèles contenant diverses fonctions polynomiales et autres des moyennes comme prédicteur. De tous les modèles envisagés, la famille de modèles quadratiques a donné des résultats d'ajustement aussi bons que n'importe quelle autre. Nous nous concentrons sur les modèles quadratiques qui suivent.

$$\text{Modèle V1:} \quad \tilde{V}_{h,ii} = \beta_{1i} p_{h,i}, \tag{8}$$

$$\text{Modèle V2:} \quad \tilde{V}_{h,ii} = \beta_{2i} p_{h,i}(1 - p_{h,i}), \tag{9}$$

$$\text{Modèle V3:} \quad \tilde{V}_{h,ii} = \beta_{1i} p_{h,i} + \beta_{2i} p_{h,i}(1 - p_{h,i}). \tag{10}$$

Donc, nous considérons un modèle de variance linéaire V1, un modèle de type binomial V2 et un modèle de variance quadratique général V3. Tous ces modèles assurent correctement que $\tilde{V}_{h,ii} = 0$ quand $p_{h,i} = 0$, mais seul V2 assure que $\tilde{V}_{h,ii} = 0$ quand $p_{h,i} = 1$. La logique qui sous-tend V1 est que les relations sont souvent approximativement linéaires sur de petits intervalles. Aussi bien V1 que V2 sont des sous-modèles du modèle quadratique à deux paramètres V3. Nous avons également considéré des modèles pour $\log(\tilde{V}_{h,ii})$, mais ceux-ci n'ont pas donné un aussi bon ajustement.

Le modèle V3 est équivalent au modèle proposé par Wolter (1985, chapitre 5); l'équivalence se voit en exprimant le deuxième membre de V3 en fonction de $p_{h,i}$ et de $p^2_{h,i}$, puis en divisant les deux membres par $p^2_{h,i}$ pour obtenir la variance relative. Cependant, les estimations des paramètres obtenues par ajustement des deux formes du modèle peuvent différer selon les hypothèses de modélisation utilisées.

3.2 Fonctions de corrélation avec données complètes

Comme les corrélations sont indépendantes de l'échelle des données, nous les modélisons et nous dérivons les covariances d'échantillonnage, au lieu de modéliser directement les covariances. Nous modélisons les corrélations d'échantillon

$$\hat{\rho}_{h,ij} = \frac{\hat{V}_{h,ij}}{(\hat{V}_{h,ii} \hat{V}_{h,jj})^{1/2}},$$

par la voie des valeurs transformées non contraintes $Z_{h,ij} = \log\{(1 + \hat{\rho}_{h,ij})/(1 - \hat{\rho}_{h,ij})\}$. Contrairement aux modèles de variance, les modèles de corrélation peuvent inclure une ordonnée à l'origine non contrainte, puisque la corrélation n'est sujette à aucune contrainte naturelle quand $p_{h,i}$ ou $p_{h,j}$ s'approche de 0 ou de 1.

Puisque $\hat{\rho}_{h,ij}$ est une fonction des premier et deuxième moments des questions *i* et *j*, il semble raisonnable de se concentrer d'abord sur les modèles linéaires et quadratiques pour $Z_{h,ij}$. Comme pour les fonctions de variance, nous constatons qu'une gamme plus étendue de modèles

(par exemple, les modèles avec logarithme des moyennes comme prédicteur) n'améliorent pas considérablement l'ajustement. En dernière analyse, nous retenons la série de modèles emboîtés qui suit.

$$\text{Modèle C1:} \quad Z_{h,ij} = \alpha_{0ij}, \tag{11}$$

$$\text{Modèle C2:} \quad Z_{h,ij} = \alpha_{0ij} + \alpha_{3ij}\ p_{h,i}\ p_{h,j}, \tag{12}$$

$$\text{Modèle C3:} \quad Z_{h,ij} = \alpha_{0ij} + \alpha_{1ij}(p_{h,i} + p_{h,j}) + \alpha_{3ij}\ p_{h,i}\ p_{h,j}, \tag{13}$$

$$\text{Modèle C4:} \quad Z_{h,ij} = \alpha_{0ij} + \alpha_{1ij} p_{h,i} + \alpha_{2ij} p_{h,j} + \alpha_{3ij}\ p_{h,i}\ p_{h,j}, \tag{14}$$

$$\text{Modèle C5:} \quad Z_{h,ij} = \alpha_{0ij} + \alpha_{1ij} p_{h,i} + \alpha_{2ij} p_{h,j} + \alpha_{3ij}\ p_{h,i}\ p_{h,j} + \alpha_{4ij}\ p_{h,i}^2 + \alpha_{5ij}\ p_{h,j}^2, \tag{15}$$

Le modèle C3 est le modèle C4 avec la contrainte $\alpha_{1ij} = \alpha_{2ij}$.

3.3 Prédiction des covariances avec données manquantes structurées

Lorsque les données comportent des enchaînements de questions, les corrélations d'échantillon des évaluations pour l'ensemble des répondants qui ont répondu aux deux questions peuvent être modélisées au moyen des modèles (11) à (15), comme dans le cas de la réponse complète. Il est facile d'estimer les covariances d'échantillon correspondantes en utilisant les fonctions de variance ajustées pour rééchelonner les corrélations prévues. Cependant, comme la covariance d'échantillonnage reflète la variabilité dans l'ensemble du processus d'échantillonnage, et non simplement la variabilité dans la sous-population de répondants qui ont répondu aux deux questions, la relation entre la covariance d'échantillon et la covariance d'échantillonnage est plus complexe que si les données étaient complètes. À la présente section, nous dérivons la relation entre la covariance d'échantillon pour l'ensemble de répondants qui ont répondu aux deux questions et la covariance d'échantillonnage. Cela nous permet d'appliquer des modèles de corrélation tels que (11) à (15) à des données avec enchaînements de questions.

Pour toute paire de questions, il existe quatre schémas de données, à savoir une réponse aux deux questions, une réponse et une question sautée (deux schémas), et deux questions sautées. Nous étendons notre notation en introduisant un indice supérieur représentant la situation de réponse à une deuxième question. Soit $\hat{Y}^1_{h,ij} = \sum_{S_h} \breve{y}_{h,ik}\breve{r}_{h,jk}$, $\hat{Y}^0_{h,ij} = \sum_{S_h} \breve{y}_{h,ik}(1-\breve{r}_{h,jk})$, $\hat{R}^1_{h,ij} = \sum_{S_h} \breve{r}_{h,ik}\breve{r}_{h,jk}$, $\hat{R}^0_{h,ij} = \sum_{S_h} \breve{r}_{h,ik}(1-\breve{r}_{h,jk})$, $\hat{M}^1_{h,ij} = \hat{Y}^1_{h,ij} / \hat{R}^1_{h,ij}$, $\hat{M}^0_{h,ij} = \hat{Y}^0_{h,ij} / \hat{R}^0_{h,ij}$. Alors

$$\hat{M}_{h,i} = \frac{\hat{R}^1_{h,ij}\ \hat{M}^1_{h,ij} + \hat{R}^0_{h,ij}\ \hat{M}^0_{h,ij}}{\hat{R}_{h,i}}.$$

Dans le cas de l'échantillonnage avec probabilités égales, le remplacement de $\hat{M}_{h,i}$ par l'expression susmentionnée dans (7) donne

$$\tilde{V}_{h,ij} = \frac{\hat{R}^1_{h,ij}}{\hat{R}_{h,i}\ \hat{R}_{h,j}} \left\{ \hat{C}^1_{h,ij} + \frac{\hat{R}^0_{h,ij}\ \hat{D}_{h,ij}\ \hat{R}^0_{h,ji}\ \hat{D}_{h,ji}}{\hat{R}_{h,i}\ \hat{R}_{h,j}} \right\}, \tag{16}$$

où $\hat{D}_{h,ij} = \hat{M}^1_{h,ij} - \hat{M}^0_{h,ij}$. Ici, $\hat{C}^1_{h,ij} = \sum_S (\breve{y}_{h,ik} - \hat{M}^1_{h,ij}\ \breve{r}_{h,ik})(\breve{y}_{h,ik} - \hat{M}^1_{h,ji}\ \breve{r}_{h,jk}) / \hat{R}^1_{h,ij}$ est la covariance d'échantillon normalisée des évaluations pour l'ensemble des répondants qui ont répondu aux deux questions (que l'on peut prédire en utilisant les fonctions de corrélation et de variance et, dans le cas de l'échantillonnage avec probabilités inégales, en appliquant un facteur de normalisation). Lorsque les probabilités d'échantillonnage ne sont pas égales, l'équation (16) n'est vérifiée exactement que si $\sum_S \breve{r}_{h,jk}(\breve{y}_{h,ik} - \hat{M}^1_{h,ik}\ \breve{r}_{h,ik}) = 0$. Par conséquent, nous pouvons nous attendre à ce que (16) donne une bonne approximation si les probabilités d'échantillonnage pour une question ne sont pas fortement corrélées aux résidus d'une autre question. En général, il convient de vérifier s'il est approprié d'utiliser (16) pour les plans d'échantillonnage avec probabilités inégales.

Les différences estimées entre les moyennes $\hat{D}_{h,ij}$ déterminent la contribution du schéma de réponse à la covariance d'échantillonnage. Nous pouvons modéliser $\hat{D}_{h,ij}$ ou $\hat{D}_{h,ij}\ \hat{D}_{h,ji}$ dans le processus d'obtention d'estimations lissées de $\tilde{V}_{h,ij}$. Dans notre application, les $\hat{D}_{h,ij}$ sont généralement petites. Comme le deuxième terme de (16) est un produit de deux facteurs de petite taille ($\hat{D}_{h,ij}$ et $\hat{D}_{h,ji}$), la contribution de $\hat{D}_{h,ij}$ à (16) est faible et il suffit d'utiliser un modèle simple pour $\hat{D}_{h,ij}$, comme une constante pour chaque paire de questions. Cependant, une constante propre devrait être estimée pour chaque paire de questions.

3.4 Ajustement et évaluation des modèles

Nous estimons les paramètres de la fonction de variance ou de la fonction de corrélation par régression par les moindres carrés repondérée itérativement. La pondération est importante quand le nombre de réponses varie considérablement d'un domaine à l'autre, comme dans le cas de notre exemple.

À la présente section, nous utilisons un indice pour les domaines (h) et pour les répondants (k), mais non pour les questions, car la même méthodologie s'applique à chaque modèle de variance et de corrélation. Des calculs exacts sont faits pour le cas de l'échantillonnage avec probabilités

égales et des approximations sont notées pour le cas de l'échantillonnage avec probabilités inégales. De façon générique, les estimateurs directs $\tilde{f}_h$, les valeurs réelles f_h et les prédictions de modèle $\hat{f}_h$ sont reliés au moyen du modèle hiérarchique

$$\text{Niveau I}: \quad \tilde{f}_h = f_h + \in_h, \tag{17}$$

$$\text{Niveau II}: \quad f_h = \hat{f}_h + e_h, \tag{18}$$

où $\in_h \sim [0, \sigma_h^2 / \hat{R}_{S_h}]$, $e_h \sim [0, \tau^2]$, et $[\mu, \sigma^2]$ indiquent une loi d'espérance μ et de variance σ^2, mais de forme non précisée. Dans le cas de l'échantillonnage avec probabilités inégales, nous remplaçons $\hat{R}_{S_h}$ par $\hat{R}^*_{S_h}$. Ici, $\in_h$ représente l'erreur d'échantillonnage et e_h, l'erreur de modélisation. Marginalement, $\tilde{f}_h = \hat{f}_h + e_h + \in_h$, de sorte que, dans la régression, nous pondérons l'observation pour le domaine h par $w_h = (\tau^2 + \sigma_h^2 / \hat{R}_{S_h})^{-1}$, qui est l'inverse de la variance marginale. Sous échantillonnage avec probabilités égales, la variance de l'estimation directe de $\sigma_h^2 = E[\tilde{f}_h - f_h]^2$ est donnée par

$$\hat{\sigma}_h^2(\tilde{f}_h) = \frac{1}{\hat{R}_{S_h} - 1}\left\{\frac{1}{\hat{R}_{S_h}}\sum_{k\in S}(y_{h,k} - \hat{M}_h r_{h,k})^4 - (1 - \frac{3}{\hat{R}_{S_h}})\tilde{f}_h^2\right\}$$
$$\text{si } f \text{ est une variance} \tag{19}$$

et par

$$\hat{\sigma}_h^2(\tilde{f}_h) = \frac{4}{\hat{R}_{S_h} - 3} \text{ si } f \text{ est une corrélation transformée.} \tag{20}$$

Dans le cas de l'échantillonnage avec probabilités égales, l'équation (19) est exacte et ne dépend pas des hypothèses paramétriques (Seber 1977, page 14). L'approximation asymptotique (20) de la variance de la corrélation transformée Z_h (Freund et Walpole 1987, page 477) se détériore à mesure que diminuent les tailles d'échantillon et elle échoue entièrement pour $\hat{R}_{S_h} \le 3$. Cependant, les domaines pour lesquels les échantillons sont de petite taille ont peu d'effet sur les modèles ajustés; nous excluons donc les domaines pour lesquels $\hat{R}_{S_h} \le 3$ de la modélisation des corrélations.

Lorsque les probabilités d'échantillonnage ne sont pas égales, nous pouvons utiliser la grande contrepartie d'échantillon de (19), donnée par

$$\hat{\sigma}_h^2(\tilde{f}_h) = \sum_{k\in S}\left\{\begin{array}{l}\dfrac{(\breve{y}_{h,k} - \hat{M}_h \breve{r}_{h,k})^2}{\sum_{l\in S}\breve{r}_{h,l}^2} - \dfrac{2w_h}{\sum_{l\in S}\breve{r}_{h,l}} \\ \times(\breve{y}_{h,k} - \hat{M}_h \breve{r}_{h,k}) - \dfrac{\tilde{f}_h^2}{\sum_{l\in S}\breve{r}_{h,l}^2}\breve{r}_{h,k}^2\end{array}\right\}^2,$$

où $w_h = (\sum_S \breve{y}_{h,l}\breve{r}_{h,l}) / \sum_S \breve{r}_{h,l}^2) - \hat{M}_h$, En cas d'échantillonnage avec probabilités égales, $w_h = 0$ et l'expression susmentionnée se réduit à une version non corrigée pour le biais de (19). Si les probabilités d'échantillonnage ne sont pas égales, nous proposons de remplacer (20) par l'estimateur corrigé pour l'effet de plan.

$$\hat{\sigma}_h^2(\tilde{f}_h) = \frac{4}{\hat{R}^*_{S_h} - 3}.$$

La variance de l'erreur de modélisation τ^2 est estimée par :

$$\hat{\tau}^2 = \max\left\{0, \hat{\text{MSE}} - \frac{\sum \hat{R}_{S_h}\hat{\sigma}_h^2(\tilde{f}_h)}{\sum \hat{R}_{S_h}}\right\},$$

où $\hat{\text{MSE}} = \sum_h q_h(\tilde{V}_h - \hat{f}_h)^2$, $q_h = N\,\hat{R}_{S_h} / \sum_h \hat{R}_{S_h}$, et $N = \sum_h I(\hat{R}_{S_h} > 0)$. Nous réestimons alors les poids selon l'expression $\hat{w}_h = (\tau^2 + \hat{\sigma}_h^2(\tilde{f}_h)/\hat{R}_{S_h})^{-1}$ et nous rajustons les modèles de FVCG par itération jusqu'à la convergence. Nous suggérons de nouveau de remplacer $\hat{R}_{S_h}$ par $\hat{R}^*_{S_h}$ si les probabilités d'échantillonnage ne sont pas égales.

Nous avons comparé l'exactitude prédictive des modèles en utilisant $R^2 = 1 - \hat{\text{MSE}} / \hat{\text{MSV}}$, où $\hat{\text{MSE}}$ est l'erreur quadratique moyenne de la régression et $\hat{\text{MSV}}$ est la moyenne, pondérée par la taille de l'échantillon, des variances d'échantillonnage des estimateurs directs (variances ou corrélations transformées) pour chaque domaine. Notons que nous pourrions avoir $R^2 < 0$ pour un modèle dont l'ajustement est très médiocre.

3.5 Estimateurs combinés

Pour les domaines dont l'échantillon est petit, les estimations directes de la variance de sondage sont trop imprécises pour être utiles, tandis que, pour les domaines plus grands dans la même étude, les estimations peuvent être assez fiables. Fay et Herriot (1979), ainsi que Ghosh et Rao (1994) ont montré que réduire les estimations directes vers une valeur lissée fondée sur un modèle peut améliorer considérablement la précision. Ils proposent des estimateurs bayésiens composites ou empiriques qui sont des moyennes pondérées des estimateurs directs et d'estimateurs fondés sur un modèle. Autrement dit, au lieu d'utiliser les estimations directes ou celles obtenues par modélisation généralisée de la variance/covariance, nous utilisons une moyenne pondérée des deux estimateurs pour éventuellement obtenir d'encore meilleures estimations.

Nous pouvons construire de tels estimateurs pondérés pour les variances de domaine en utilisant le modèle spécifié en (17) et (18). Une approche naturelle consiste à appliquer aux estimateurs directs fondés sur un modèle une pondération inversement proportionnelle aux variances d'échantillonnage et d'erreur de modélisation correspondantes, respectivement (notées σ_h^2 et τ^2, respectivement, pour le

domaine h). L'estimateur résultant pour le domaine h (pour les variances et les corrélations transformées) est :

$$\tilde{f}_h = \frac{\hat{\tau}^2 \tilde{f}_h^{\text{dir}} + \hat{\sigma}_h^2 \tilde{f}_h^{\text{mod}}}{\hat{\tau}^2 + \hat{\sigma}_h^2} = \tilde{f}_h^{\text{dir}} + \frac{\hat{\sigma}_h^2}{\hat{\tau}^2 + \hat{\sigma}_h^2}\left(\tilde{f}_h^{\text{mod}} - \tilde{f}_h^{\text{dir}}\right),$$

où $\tilde{f}_h^{\text{dir}}$ et $\tilde{f}_h^{\text{mod}}$ représentent les estimateurs directs et fondé sur un modèle. Cette formule générique s'applique aux estimations de la variance pour l'ensemble des questions et aux estimations des corrélations pour toutes les paires de questions. L'expression la plus à droite a la forme d'un estimateur bayésien empirique.

Si l'estimateur de variance direct et celui fondé sur un modèle sont indépendants, la variance de l'estimateur combiné résultant est $\tau^2\sigma_h^2/(\tau^2+\sigma_h^2) \le \min\{\tau^2, \sigma_h^2\}$. Donc, l'estimateur composite est, au moins, aussi précis que l'un ou l'autre des deux estimateurs qui le constituent, donc représente une amélioration par rapport au choix ponctuel entre la prédiction directe et celle fondée sur un modèle. Cette stratégie est utile, surtout quand les prédictions fondées sur un modèle sont meilleures que les estimations directes pour certains domaines, mais non tous.

4. Exemple : Ensemble de données de la CAHPS®

La Consumer Assessments of Health Plans Study (CAHPS®) (Goldstein, Cleary, Langwell, Zaslavsky et Heller, 2001) a été conçue principalement pour recueillir les évaluations et les déclarations des consommateurs au sujet des régimes d'assurance-maladie. Les scores moyens des régimes (peut-être après recodage) pour les diverses questions du sondage sont calculés et communiqués aux consommateurs, aux régimes d'assurance-maladie et aux acheteurs. Chaque domaine d'analyse comprend les personnes inscrites à un régime d'assurance-maladie (ou une partie géographiquement définie d'un régime) durant une année particulière; la plupart des régimes sont échantillonnés pour plusieurs années. La strate est l'unité déclarante (régime ou partie de celui-ci) durant une année donnée; les unités déclarantes correspondent à des régimes sauf dans le cas de quelques régimes très importants comptant plusieurs unités déclarantes. Par conséquent, le nombre d'unités pour l'estimation des fonctions de variance et de covariance est grand.

Nous illustrons notre méthode au moyen d'un ensemble de données de la CAHPS pour les bénéficiaires des régimes de soins américains gérés par Medicare, c'est-à-dire un système d'entités privées, mais financées par les deniers publics qui ont desservi de 5,7 à 6,9 millions de bénéficiaires âgés ou handicapés chaque année de la période couverte par l'étude (1997 à 2001). Nos données représentent 381 domaines déclarants, chacun échantillonné dans une à cinq années, pour aboutir à un total de 932 domaines distincts correspondant à une unité déclarante par année, avec 705 848 réponses. Comme les échantillons sont tirés indépendamment chaque année, les patients peuvent être échantillonnés pour plusieurs années. Cependant, l'échantillonnage répété est rare et nous pouvons omettre d'en tenir compte dans notre analyse. Par conséquent, les domaines sont des strates dans chacune desquelles est effectué un échantillonnage d'éléments avec probabilités égales. Il convient de souligner que, dans les analyses de la CAHPS, aucune correction n'est faite pour l'échantillonnage en population finie, puisque les données sont recueillies pour orienter les choix faits lors d'années ultérieures plutôt que pour enregistrer les expériences vécues par la population sondée durant une année particulière.

Les questions de la CAHPS comportent divers formats de réponse ordonnée comportant 11, 4, 3 ou 2 options de réponse. L'évaluation globale des médecins, des spécialistes, des soins et des régimes est faite au moyen d'une échelle de 0 à 10 allant du « pire possible » au « meilleur possible ». D'autres questions comportent une échelle de « fréquence » ordonnée de quatre points (jamais/parfois/habituellement/toujours), ou une échelle d'« intensité du problème » ordonnée de trois points (pas un problème/plus ou moins un problème/un gros problème), ou sont dichotomiques (non/oui). La réponse à de nombreuses questions est fournie uniquement par les répondants utilisant des services particuliers ou ayant des besoins particuliers, qui sont dépistés au moyen des questions filtres. Par exemple, une question visant à savoir si la personne a réussi à obtenir des conseils par téléphone n'est posée qu'aux répondants qui ont d'abord déclaré qu'ils avaient essayé d'obtenir des conseils de cette façon.

4.1 Statistiques descriptives

Le tableau 1 présente la répartition des réponses et celle des moyennes de domaine selon le type de question. Les observations manquantes à cause d'un enchaînement de questions structuré surviennent souvent en blocs, le nombre de questions sautées pouvant aller jusqu'à 11 sur la base d'une seule question filtre. La proportion de non-réponses n'étant pas due à un enchaînement de questions structuré est très faible (moins de 2 % pour presque toutes les questions). Dans la présente analyse, nous traitons tous les types de non-réponse de la même façon.

Les taux les plus faibles de réponse à une question (aussi faibles que 4 %) sont ceux enregistrés pour les questions à échelle d'intensité de problème, dont plusieurs ont trait à des services spécialisés, comme des traitements ou des soins de santé à domicile dont ont besoin un assez petit nombre de répondants seulement. Certaines questions à échelle de fréquence et à réponse oui/non produisent des taux de

réponse assez faibles. La variation la plus forte des proportions de questions sautées s'observe pour les questions à réponse oui/non, le taux allant de 96,7 % pour « une plainte ou un problème concernant le régime » à 12,5 % pour « l'obtention d'une prescription par l'entremise du régime ». Les moyennes de domaine sont généralement concentrées vers l'extrémité supérieure des échelles respectives, indiquant que la plupart des réponses sont favorables.

Tableau 1
Distribution des réponses et des évaluations pour des questions de même type (n = 705 848 répondants)

Statistique	Numérique	Fréquence	Problème	Oui/non
Nombre de questions	4	11	11	9
Pourcentage de réponses				
Moyenne	74,97	62,56	30,32	57,26
Minimum	50,90	27,70	4,00	12,50
Maximum	95,00	74,50	64,40	96,70
Moyenne de questions				
Moyenne	8,76	3,57	2,70	1,78
Minimum	8,57	3,09	2,49	1,62
Maximum	8,88	3,84	2,86	1,97
Distribution des évaluations (entre les questions de même type)				
0	0,5			
1	0,4	2,0	5,7	19,5
2	0,4	6,3	12,1	80,5
3	0,7	23,9	82,2	
4	0,9	67,8		
5	4,6			
6	3,0			
7	6,2			
8	16,1			
9	17,8			
10	49,5			

Les questions comportent une échelle numérique 0–10 allant de « pire possible » à « meilleur possible », une échelle de « fréquence » ordonnée de quatre points 1–4 (jamais/parfois/habituellement/ toujours), une échelle d'« intensité de problème » ordonnée de trois points 1–3 (pas un problème/plus ou moins un problème/un gros problème) ou sont dichotomiques 1–2 (non/oui).

Nous présentons aussi dans le tableau 1 la moyenne, le minimum et le maximum de domaine pour l'ensemble des questions de même type. Ces données montrent que les questions à échelle 0–10 sont celles pour lesquelles la variation totale est la plus faible (après rééchelonnement pour produire la fourchette 0–1 commune), tandis que les questions à échelle 1–2 sont celles dont la variation totale est la plus importante entre les domaines et entre les questions. Ces observations se dégagent aussi de la figure 1, où nous observons que la distribution des questions à échelle 1–2 varie considérablement d'une question à l'autre, tandis que celle des questions à échelle 0–10 est plus homogène.

Le tableau 2 donne les statistiques sommaires pour les moyennes et les écarts-types des évaluations moyennes de domaine, calculées sur l'ensemble des questions de même type. Ces données complètent celles de la figure 1 en résumant les différences entre les distributions des questions pour une échelle donnée. Les questions comportant un plus grand nombre de catégories de réponse sont concentrées vers l'extrémité supérieure de l'échelle et ont donc une variance plus faible. Par exemple, l'écart-type moyen pour les questions à échelle 1–2 (0,36) est égal au double de celui des questions à échelle 0–10 rééchelonnées (0,172). À part les questions à échelle 0–10, les distributions des évaluations moyennes de domaine varient fortement entre les questions de même type. Par exemple, l'écart-type des moyennes des questions à échelle 1–2 sur l'ensemble des questions est de 0,30 comparativement à un écart-type rééchelonné de 0,03 pour les questions à échelle 0–10.

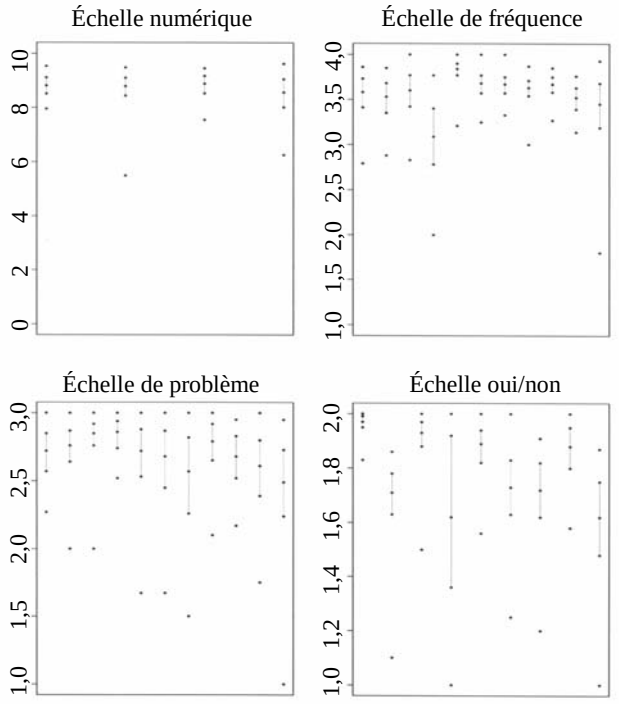

Figure 1. Sommaire de cinq points des moyennes d'échantillon de domaine pour chaque type de question. Le sommaire de cinq points comprend le minimum, le 10e centile, la moyenne, le 90e centile et le maximum.

Tableau 2
Statistiques sommaires des moyennes et des écarts-types de domaine évalués sur les domaines et sur les questions

Type	Statistiques sommaires pour : Moyennes des questions				É.-T. des questions	
	Min	Max	Moy.	É.-T.	Moy.	É.-T.
Numérique 0 –10	6,82	9,52	8,76	0,30	1,72	0,26
Fréquence 1– 4	2,86	3,90	3,57	0,12	0,66	0,09
Problème 1–3	1,88	2,99	2,70	0,14	0,57	0,13
Oui/non 1–2	1,34	1,96	1,78	0,08	0,36	0,06

Nota : Les colonnes 2 à 5 donnent le minimum, le maximum, la moyenne et l'écart-type des moyennes des questions de domaine sur l'ensemble des questions d'un type donné. Les colonnes 6 et 7 donnent la moyenne et l'écart-type des écarts-types des questions de domaine sur l'ensemble des questions d'un type donné.

Les corrélations d'échantillon varient également beaucoup d'une paire de questions à l'autre (figure 2), quoique la plupart soient positives. Le plus souvent, les corrélations entre questions de même type sont plus fortes que celles entre questions de types différents. Les évaluations numériques à échelle 0–10 sont celles dont les corrélations sont les plus importantes (moyenne = 0,49) et, en général, les évaluations comportant un grand nombre de catégories ont tendance à produire des corrélations plus fortes que celles comportant un moins grand nombre de catégories. Bien que la plupart des paires de questions à échelle 1–4 donnent des corrélations moyennes s'approchant de 0,5, l'une des questions est négativement corrélée aux autres (révélé par la grappe de corrélations moyennes inférieures à 0); ce résultat est dû au codage inverse d'une question dont la moyenne globale d'échantillon ne se situait pas dans la moitié supérieure de l'échelle. Les distributions des corrélations des paires de questions dichotomiques 1–2 sont centrées autour de 0, ce qui signifie que ces corrélations sont souvent négatives. L'énoncé complet des questions et des statistiques sommaires supplémentaires figure dans Zaslavsky, Beaulieu, Landon et Cleary (2000) et dans Zaslavsky et Cleary (2002).

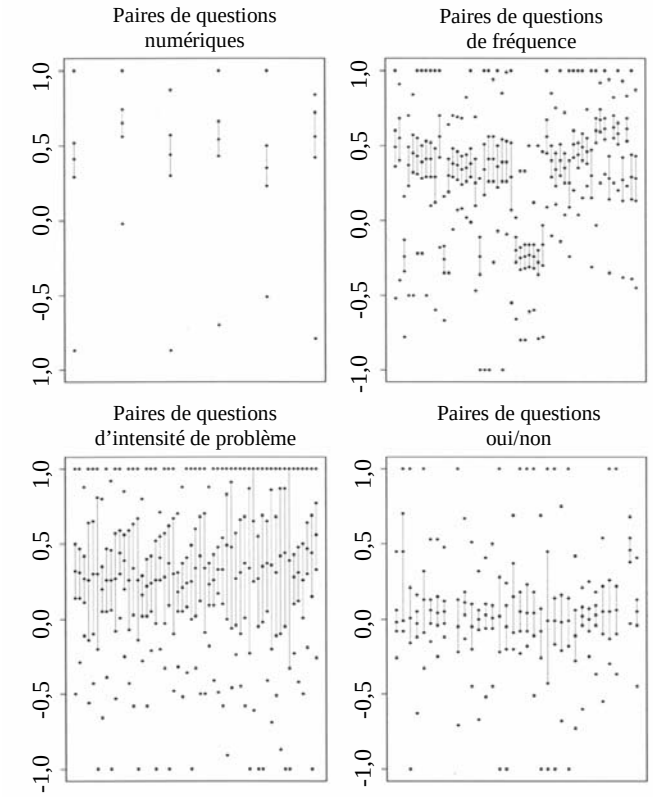

Figure 2. Sommaire de cinq points des corrélations d'échantillon de domaine entre les questions de même type. Le sommaire de cinq points comprend le minimum, le 10e centile, la moyenne, 90e centile et le maximum.

Nous présentons les modèles ajustés aux variances et aux corrélations dans la suite de la section. Des vérifications approfondies des modèles les mieux ajustés ont montré que les résidus ne présentaient aucune régularité discernable.

4.2 Fonctions de variance

Lors de travaux préliminaires non présentés ici, nous avons ajusté deux modèles pour les groupes de questions ayant la même échelle de réponse, l'un avec les mêmes paramètres de régression pour toutes les questions et l'autre avec des paramètres de régression différents pour chaque question, à l'ensemble de données comprenant toutes les questions. Les comparaisons des ajustements globaux des modèles (au moyen de critères tels que le C_p de Mallow, le R^2 et le R^2 corrigé) et les tests de signification des interactions effet-question ont montré que permettre aux paramètres de varier selon la question améliore de façon significative l'ajustement du modèle. Par exemple, pour les évaluations numériques rééchelonnées, pondérées par la taille d'échantillon du domaine, les racines des erreurs quadratiques moyennes des deux modèles étaient de 0,446 contre 0,402, et les valeurs de R^2 étaient de 0,783 contre 0,825. D'après ces résultats, nous avons décidé d'ajuster des modèles distincts pour chaque question.

Nous avons ajusté les fonctions de variance (8) à (10) à chaque question, sauf celles à réponse oui/non, qui suivent la fonction de variance binomiale dans le cas de l'échantillonnage avec probabilités égales. La procédure itérative décrite à la section 3.4 a convergé presque précisément en exactement deux itérations. Ce résultat tient au fait que les poids des observations ne varient qu'en fonction de l'estimation de τ^2, de sorte que leur variation est très faible après la première itération.

Le tableau 3 donne la variation d'échantillonnage moyenne, la variation moyenne de l'erreur de modélisation (ModErr) et R^2 pour la valeur moyenne de chaque modèle calculés sur l'ensemble des questions correspondant à chaque type d'échelle. La variation d'échantillonnage, calculée selon (19), ne dépend pas du modèle.

Pour les questions ne comportant qu'un petit nombre de catégories (celles qui ressemblent le plus à la loi binomiale), la composante quadratique de la fonction de variance a tendance à dominer la composante linéaire, ce qui produit un meilleur ajustement des modèles V2 et V3 que du modèle V1. Comme V2 impose une contrainte en un point très en dehors de la fourchette de valeurs des moyennes de domaine, il n'est pas aussi bien ajusté aux données quand le nombre de catégories augmente et que les données s'écartent par conséquent davantage de la loi binomiale. Les réponses aux questions à échelle 0–10 sont moins dispersées que celles aux questions à échelle 1–4 ou 1–3, de sorte que le modèle linéaire est mieux ajusté. Les valeurs de

R^2 pour le modèle V3 s'approchent de 0,75 pour les questions à échelle numérique (0–10), de 0,85 pour les questions à échelle de fréquence (1–4) et de 0,95 pour les questions à échelle d'intensité de problème (1–3).

Tableau 3
Statistiques de qualité d'ajustement du modèle pour les fonctions de variance

Échelle d'évaluation	0–10		1–4		1–3	
Variation d'échantillonnage	0,1460		0,3511		3,1703	
	ModErr	R^2	ModErr	R^2	ModErr	R^2
Modèle V1	0,020	0,741	0,066	0,824	0,069	0,916
Modèle V2	0,043	0,710	0,036	0,835	0,000	0,940
Modèle V3	0,016	0,750	0,024	0,847	0,000	0,947
	Prob(ModErr < Variation d'échantillonnage)					
Modèle V1	0,968		0,916		0,996	
Modèle V2	0,858		0,967		0,996	
Modèle V3	0,981		0,983		0,996	

ModErr est la composante de la variance due au manque d'ajustement, R^2 est la statistique définie à la section 3.4, Prob(ModErr < Variation d'échantillonnage) est la proportion de domaines pour lesquels l'erreur de modélisation est plus faible que la variation d'échantillonnage. Toutes les évaluations sont rééchelonnées sur une échelle de 0–1 et les erreurs de modélisation sont multipliées par 10^4.

La partie inférieure du tableau 3 donne, pour chaque question, la proportion de domaines (parmi ceux comptant au moins deux réponses à la question considérée) pour lesquels la variation d'échantillonnage est supérieure à la variation de l'erreur de modélisation. Pour plus de 90 % des domaines, la variation de l'erreur de modélisation est plus faible que la variation d'échantillonnage de l'estimation directe de la variance.

La figure 3 illustre l'ajustement de V3 pour deux questions de chacun des groupes 0–10, 1–4 et 1–3. Les illustrations pour les autres questions sont semblables, mais ne sont pas présentées ici faute d'espace. Les courbes ajustées sont contraintes de passer par la valeur 0 à l'évaluation maximale. Pour évaluer l'effet de cette contrainte sur la fonction de variance ajustée, nous ajustons également une fonction de variance quadratique (à trois paramètres) non contrainte. Celle-ci atteint une valeur très proche de 0 à l'évaluation maximale et s'approche de très près de la courbe ajustée d'après les modèles avec contraintes, ce qui appuie encore davantage le modèle V3.

Les estimations moyennes des paramètres et de leurs écarts-types sur l'ensemble des questions de même type sont présentées au tableau 4. La valeur des paramètres varie considérablement selon la question, ce qui soutient la décision d'estimer des coefficients de régression distincts.

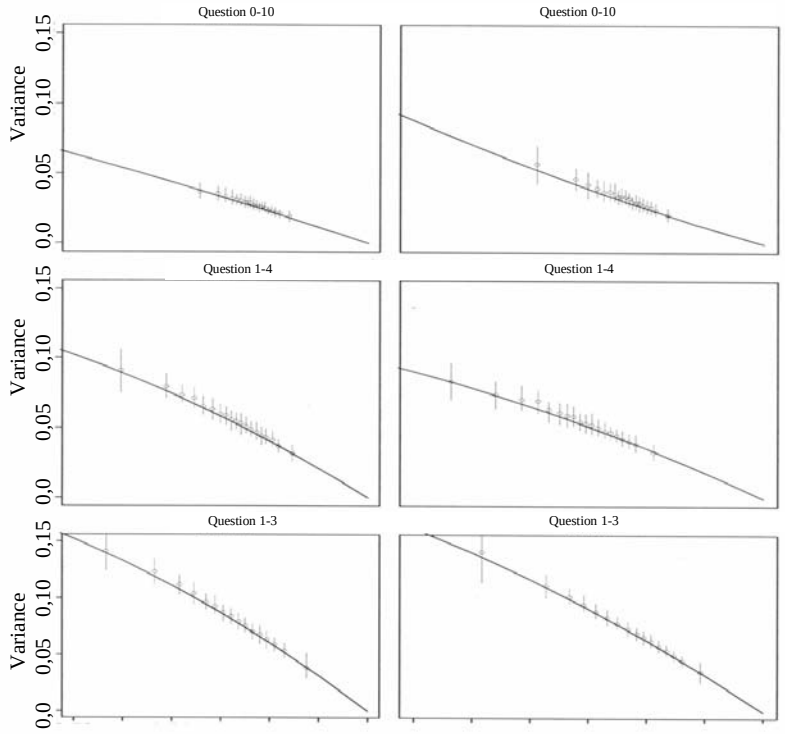

Figure 3. Fonction de variance quadratique (V3) de deux questions pour chaque type d'évaluation. Chaque point est la moyenne de 60 domaines. Les lignes verticales joignent les 10e et 90e centiles de la distribution des variances. Pour ces tracés et les suivants, la direction de l'axe horizontal a été inversée afin qu'elle concorde avec celle des variables originales.

Dans la plupart des cas, les coefficients pour les termes $p_{h,i}$ ainsi que $p_{h,i}(1-p_{h,i})$ de V3 sont significatifs, ce qui indique que ces termes sont nécessaires pour la modélisation généralisée de la variance. Dans certains cas (particulièrement pour les questions à échelle 0–10), le coefficient du terme $p_{h,i}(1-p_{h,i})$ est négatif, ce qui donne une fonction de variance estimée convexe plutôt que concave (forme de la fonction de variance binomiale). Cette situation peut se produire si les moyennes d'échantillon des évaluations sont concentrées sur une petite partie de l'échelle de réponse, sur laquelle le terme linéaire explique une grande part de la variation des données. Comme nous l'avons mentionné plus haut, l'ajout de fonctions polynomiales ou logarithmiques d'ordre plus élevé de $p_{h,i}$ n'améliore pas significativement l'ajustement du modèle.

Tableau 4
Estimations moyennes des paramètres de la fonction de variance pour chaque type de question et écart-type entre les questions (entre parenthèses)

Modèle	Type de question					
	0–10		1–4		1–3	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
V1	0,236	–	0,354	–	0,569	–
	(0,016)	–	(0,039)	–	(0,068)	–
V2	–	0,271	–	0,421	–	0,711
	–	(0,020)	–	(0,034)	–	(0,069)
V3	0,334	– 0,114	0,151	0,241	0,239	0,420
	(0,143)	(0,155)	(0,104)	(0,132)	(0,112)	(0,110)

Voir le tableau 1 pour une description des questions de type 0–10, 1–4 et 1–3.

4.3 Fonctions de corrélation

Les modèles sont classés du plus simple (C1, modèle constant) au plus complexe (C5, contenant tous les termes linéaires et quadratiques). Comme pour les modèles de variance, les tests statistiques indiquent des effets d'interaction entre questions hautement significatifs, qui sous-entendent que des modèles distincts devraient être ajustés pour chaque paire. Nous ne nous attendions pas à ce que toutes les paires de questions présentent les mêmes corrélations, puisque, intentionnellement, les questions sont réparties en groupes intérieurement cohérents, qui mesurent chacun un aspect distinct des expériences vécues par les patients, comme les interactions avec le médecin ou celles avec les agents des services à la clientèle (Hays, Shaul, Williams, Lubalin, Harris-Kojetin, Sweeny et Cleary 1999).

Les ajustements des modèles de corrélation pour les paires de questions de même type sont résumés au tableau 5. Pour la gamme de modèles considérée, les améliorations les plus importantes de l'efficacité des modèles (mesurées par R^2) sont celles observées entre C1 et C2, et entre C3 et C4. Par exemple, le R^2 moyen pour les évaluations numériques à échelle 1–10 dans les modèles C3 à C5 est de 0,0391, 0,1494 et 0,1508, respectivement, et le R^2 moyen pour les évaluations à échelle 1–4 pour les modèles C1 à C3 est de 0, 0,0700 et 0,0789, respectivement. Ces résultats donnent à penser que C2 et C4 sont les meilleurs modèles pour différentes paires de questions, affirmation qui est appuyée par les tests de vérification d'hypothèse concernant la signification des améliorations marginales de l'ajustement des modèles.

La variation d'échantillonnage la plus élevée s'observe pour les échelles d'évaluation 1–3, du moins en partie parce que les taux élevés de non-réponses dues à des enchaînements de questions ont réduit les tailles d'échantillon. L'erreur de modélisation et le R^2 des modèles de corrélation pour les questions de types différents sont semblables à ceux des modèles pour les questions de même type.

Les valeurs de R^2 des modèles de corrélation sont comprises entre 0,029 et 0,15 pour toutes les paires de questions. Bien qu'il n'existe aucune preuve que C4 soit un modèle inapproprié pour les corrélations, ces résultats indiquent qu'une variation importante des corrélations ne peut être expliquée par les moyennes de question.

Les variances d'échantillonnage des estimations directes sont souvent inférieures aux variances de l'erreur de modélisation correspondante (partie inférieure des tableaux 5 et 6, particulièrement pour les questions à échelle 0–10. Sous C4, les variances de l'erreur de modélisation ne sont plus faibles que pour 13 % des domaines pour les évaluations de type 0–10, 40 % des domaines pour les évaluations de type 1–4 et environ 81 % des domaines pour les évaluations de type 1–3 ou 1–2.

La figure 4 donne les corrélations observées et la fonction ajustée C4 pour un exemple de paire de questions pour chacune des dix combinaisons de type de question, représentant les 595 paires de questions distinctes. Pour illustrer les modèles de corrélation ajustés, nous rajustons les corrélations observées et ajustées sur la moyenne de l'une des questions et représentons graphiquement les valeurs résultantes dans un espace bidimensionnel. Nous répétons le processus pour l'autre question, ce qui nous donne deux tracés pour chaque corrélation.

La figure 4 illustre la relation généralement faible entre les corrélations et les moyennes des questions observées aux tableaux 5 et 6. L'analyse de ces deux tableaux révèle que la relation entre la corrélation et le résultat moyen est plus faible pour les questions comportant un petit nombre de catégories et pour les corrélations de questions de différents types. En particulier, les évaluations numériques à échelle 0–10 sont le seul groupe pour lequel il existe une relation claire corrélation-moyenne.

Bien que les courbes ajustées pour les fonctions de corrélation soient presque plates, la variation des estimations des paramètres sous le modèle C4 pour α_4 sont grandes et évoquent une instabilité. La très forte variabilité des estimations des paramètres est une conséquence de la

colinéarité entre les prédicteurs dans le modèle C4. Dans de nombreux cas, la valeur estimée de α_4 contrebalance les estimations des paramètres des prédicteurs linéaires, ce qui donne une courbe ajustée pratiquement plate.

4.4 Fonctions de différence de moyennes

La différence $\hat{D}_{h,ij}$ semble ne dépendre ni de la moyenne marginale ni de son carré, ce qui implique qu'un modèle analogue à V3 pourrait être approprié. Cependant, comme $\hat{D}_{h,ij}$ est habituellement suffisamment faible pour que $\hat{D}_{h,ij}\ \hat{D}_{h,ji}$ ait un effet minime sur (16), nous ajustons un modèle constant.

4.5 Estimateur composite

Le tableau 7 donne les valeurs moyennes, calculées sur l'ensemble des questions (ou paires de questions) de même type, des quantiles de la distribution des poids $\sigma_h^2 / (\tau^2 + \sigma_h^2)$ pour l'estimation fondée sur un modèle utilisée dans l'estimateur composite de la section 3.5. La proportion de domaines pour lesquels l'erreur-type des prédictions fondées sur un modèle est plus faible que celle des estimations directes est également présentée. Comme nous l'avons mentionné plus haut, les prédictions fondées sur un modèle ont plus de poids dans les estimations composites de la variance que dans celles des corrélations. La médiane moyenne (sur les questions ou les paires de questions) des poids de l'estimateur fondé sur un modèle varie de 0,892 à 1,000 pour les variances, de 0,256 à 0,709 pour les corrélations de questions de même type, et de 0,468 à 0,738 pour les corrélations de questions de types différents. En outre, tant pour les variances que pour les corrélations, le poids des prédictions fondées sur un modèle est plus important pour les questions comportant un petit nombre de catégories de réponses. Par exemple, les poids médians de l'estimateur fondé sur un modèle sont de 0,256, 0,468, 0,540 et 0,647 sur les estimations composites des corrélations quand les évaluations numériques de type 0–10 sont appariées aux évaluations de type 0–10, 1–4, 1–3 et 1–2, respectivement. Cependant, même pour les paires d'évaluation numérique à échelle 0–10, pour lesquelles l'erreur d'échantillonnage de l'estimateur direct n'est supérieur à l'erreur de modélisation que pour 3,81 % de domaines, ces résultats indiquent que le poids médian de l'estimateur fondé sur un modèle est de 0,256, valeur non triviale.

Tableau 5
Diagnostics d'ajustement de modèle pour les fonctions de corrélation pour les questions de même type, moyenne sur les paires de questions de même type

Échelle d'évaluation	0 – 10		1–4		1–3		1–2	
Variation d'échantillonnage	0,0124		0,0178		0,1482		0,0325	
	ModErr	R^2	ModErr	R^2	ModErr	R^2	ModErr	R^2
Modèle C1	0,060	0,000	0,028	0,000	0,112	0,000	0,018	0,000
Modèle C2	0,060	0,013	0,025	0,070	0,103	0,048	0,017	0,014
Modèle C3	0,057	0,039	0,024	0,079	0,102	0,054	0,017	0,018
Modèle C4	0,047	0,150	0,023	0,100	0,100	0,068	0,016	0,029
Modèle C5	0,044	0,151	0,023	0,105	0,096	0,080	0,015	0,034
	Prob(ModErr < Variation d'échantillonnage)							
Modèle C1	0,033		0,339		0,461		0,788	
Modèle C2	0,033		0,400		0,498		0,795	
Modèle C3	0,034		0,411		0,502		0,796	
Modèle C4	0,038		0,435		0,516		0,799	
Modèle C5	0,065		0,440		0,530		0,802	

Voir le tableau 1 pour une description des questions de type 0–10, 1–4, 1–3 et 1–2, et le tableau 3 pour une explication des en-têtes de colonne.

Tableau 6
Diagnostics d'ajustement de modèle pour les fonctions de corrélation pour C4 selon le type de question. Moyenne sur les questions de même type.

Type	0–10		1–4		1–3		1–2	
	ModErr	R^2	ModErr	R^2	ModErr	R^2	ModErr	R^2
0–10	0,047	0,149	0,021	0,104	0,040	0,094	0,013	0,059
1–4			0,023	0,100	0,038	0,076	0,013	0,039
1–3					0,100	0,068	0,028	0,031
1–2							0,016	0,029
Prob(ModErr < Variation d'échantillonnage)								
0–10	0,038		0,358		0,523		0,784	
1–4			0,435		0,605		0,790	
1–3					0,516		0,827	
1–2							0,799	

Voir le tableau 1 pour une description des questions de type 0–10, 1–4, 1–3 et 1–2, et le tableau 3 pour une explication des en-têtes de colonne.

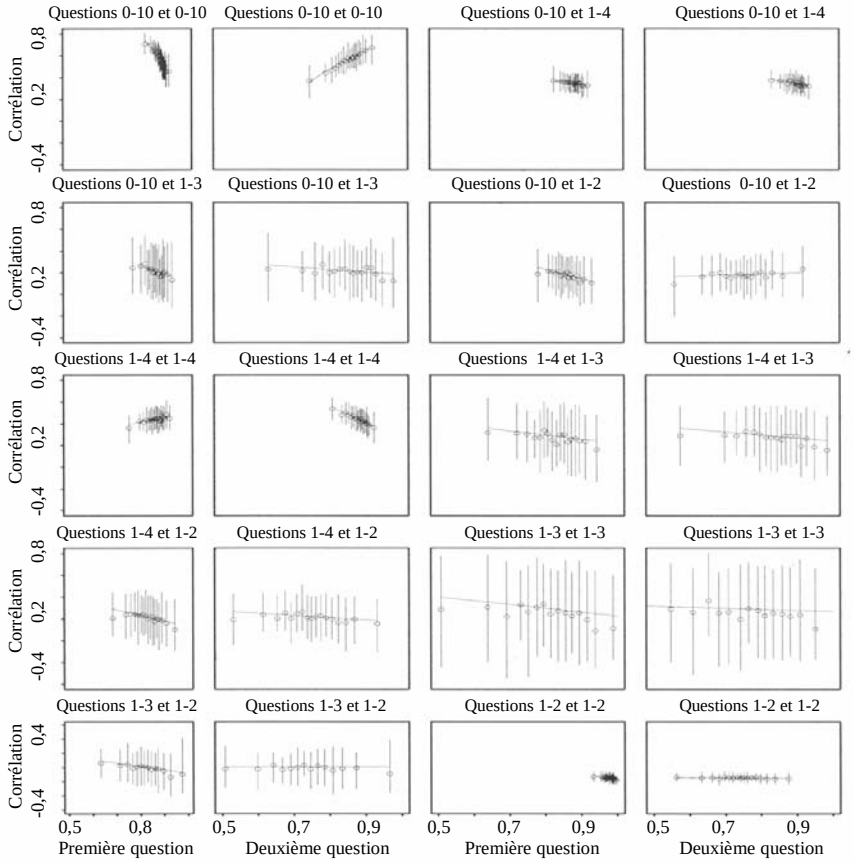

Figure 4. Fonctions de corrélation pour une paire de questions pour chaque combinaison d'échelle d'évaluation.

Nota : Les tracés pour chaque question intervenant dans la corrélation sont côte à côte. Consulter la figure 3 pour une description du contenu et des axes du tracé.

Tableau 7
Distribution des poids pour la composante fondée sur un modèle de l'estimateur composite, moyenne sur les questions de même type

Modèle	Type de question		Prob(ModErr <	Quantiles		
	1	2	Variation d'échantillonnage)	10 %	Médiane	90 %
Variance	0 – 10	–	0,981	0,778	0,892	0,948
	1 – 4	–	0,983	0,948	0,966	0,974
	1 – 3	–	0,996	1,000	1,000	1,000
Corrélation	0 – 10	0 – 10	0,038	0,141	0,256	0,335
	0 – 10	1 – 4	0,358	0,301	0,468	0,562
	0 – 10	1 – 3	0,523	0,357	0,540	0,654
	0 – 10	1 – 2	0,784	0,531	0,695	0,767
	1 – 4	1 – 4	0,435	0,324	0,497	0,591
	1 – 4	1 – 3	0,605	0,404	0,587	0,699
	1 – 4	1 – 2	0,853	0,584	0,738	0,805
	1 – 3	1 – 3	0,516	0,349	0,540	0,675
	1 – 3	1 – 2	0,827	0,584	0,737	0,817
	1 – 2	1 – 2	0,799	0,541	0,709	0,780

La distribution des poids est résumée au moyen des 10e, 50e et 90e centiles. Voir le tableau 3 pour la définition de ModErr.

4.6 Prédiction conjointe

Comme nous avons modélisé indépendamment les corrélations pour chaque question, nos matrices de corrélations ajustées ne satisfont pas nécessairement la contrainte de définie positive, qui peut être importante pour les inférences multivariées. Dans le cadre de travaux supplémentaires, nous avons déterminé qu'à condition de limiter l'analyse multivariée aux questions de même type, les corrélations ajustées d'après les modèles C2 et C4 donnent des estimations définies positives des matrices de corrélation pour presque tous les domaines. Cependant, pour les analyses portant sur des questions de types différents (par exemple, les questions à échelle numérique 0–10 et les questions à réponse oui/non 1–2), les prédictions fondées sur C4 donnent des matrices de corrélations qui sont indéfinies pour de nombreux domaines, tandis que celles fondées sur C2 sont plus stables et donnent presque systématiquement des matrices définies positives. Ceci donne à penser que, si C4 peut être légèrement supérieur en ce qui concerne l'ajustement du modèle univarié, C2 pourrait être plus approprié pour l'inférence multivariée.

Un moyen de contourner le problème des matrices de corrélations prédites indéfinies consiste à utiliser une moyenne pondérée de la matrice de corrélations prédite pour un domaine et de la matrice de corrélations moyenne estimée (MCME) sur l'ensemble de domaines. Nous pouvons construire la MCME par pondération des estimations directes (chacune étant au moins semi-définie positive) par la taille totale de l'échantillon pour chaque domaine. Puis, nous remplaçons toute matrice de corrélations prédite indéfinie par la moyenne pondérée de la matrice de corrélations prédite et de la MCME, en accroissant le poids utilisé pour chaque domaine jusqu'à ce que nous obtenions une matrice définie positive. Comme pour un estimateur bayésien empirique, ce processus stabilise les estimations en réduisant effectivement les coefficients du modèle vers ceux d'un modèle plus simple (constant).

Lors de l'analyse simultanée des 35 questions de la CAHPS, la MCME avait un poids moyen sur l'ensemble des domaines de 0,65 pour le modèle C4, mais de 0,01 seulement pour le modèle C2, puisque les corrélations prédites sous C2 sont habituellement définies positives. Lors de l'analyse des questions de type 0–10, 1–4 ou 1–3 seulement, la MCME avait un poids moyen de 0,28 et de 0,00 pour C4 et C2, respectivement, tandis que lors de l'analyse des questions de type 0–10 ou 1–4 seulement, les poids moyens correspondants étaient de 0,06 et de 0,00. Lors de l'analyse de divers types de questions séparément, le poids moyen de la MCME avec le modèle C4 était de 0,00 pour les questions de type 0–10 ou 1–4, de 0,01 pour les questions de type 1–3 et 0,17 pour les questions de type 1-2. La MCME n'est donc pas nécessaire pour analyser les questions de type 0–10 ou 1–4, parce que les matrices de corrélations prédites sont définies positives pour chaque domaine.

5. Conclusion

Nous présentons la méthodologie pour estimer les fonctions de variance et de covariance pour les moyennes de domaine de questions d'enquête ordonnées. Notre méthodologie peut également s'appliquer aux questions d'enquête comportant une échelle de mesure continue. Nous présentons une décomposition de l'erreur de modélisation qui permet de séparer la variation due à l'échantillonnage de celle due à l'ajustement du modèle. La décomposition permet aussi d'éviter le surajustement du modèle, parce qu'elle estime la proportion de la variation des données qui peut être modélisée et, donc, le moment où les prédicteurs courants suffisent.

La procédure d'ajustement des modèles de variance et de corrélation est la même que les données contiennent ou non des enchaînements structurés de questions. L'exposé analytique de la section 3.3 montre que, s'il existe des enchaînements de questions, il faut connaître les différences moyennes entre les questions selon la situation de réponse à d'autres questions afin d'estimer la covariance d'échantillonnage. Cependant, nous soutenons que ces quantités ont vraisemblablement un effet minime sur les résultats et que, par conséquent, on pourrait utiliser un modèle constant, argument qui est appuyé par nos résultats empiriques.

Une fonction de variance quadratique dont la valeur est contrainte d'être nulle pour l'évaluation maximale et un modèle pour les corrélations transformées comportant le produit, mais non le carré des moyennes sont les meilleurs prédicteurs des estimations directes dans l'application que nous avons choisie comme exemple. En général, les erreurs-types des estimations modélisées de la variance sont beaucoup plus faibles que celles des estimations directes; toutefois, il n'en est pas ainsi des estimations des corrélations. Il est intéressant et rassurant de constater que notre fonction de variance quadratique peut être exprimée sous la forme du modèle de variance relative très répandu de Wolter (1985).

Pour nos données ordonnées, les évaluations moyennes de domaine contiennent des informations minimales concernant la corrélation entre les évaluations. Donc, la relation moyenne-covariance est principalement un artefact de la relation moyenne-variance. Cependant, pour les questions comportant un grand nombre de catégories de réponse, l'association entre les corrélations et les résultats moyens pour les questions de même type est plus forte, surtout pour les paires de questions à échelle 0–10. À part les évaluations de type 0–10 et, éventuellement, celles de

type 1–4, les corrélations pourraient tout aussi bien être modélisées par des constantes, ce qui permet aussi de garantir plus facilement le caractère défini positif de la matrice de corrélations prédite. Cependant, il est important que les paramètres du modèle de corrélation puissent varier selon la paire de questions.

Un estimateur composite, résultant de la combinaison pondérée des estimateurs direct et fondé sur un modèle proportionnellement à leur précision, a une variance plus faible que l'un et l'autre estimateur pris individuellement, surtout quand les composantes ont des poids presque égaux. L'estimateur fondé sur un modèle est celui dont l'influence sur les estimations pour les petits domaines pour lesquels ont dispose de peu d'information est la plus forte. L'influence de l'estimateur fondé sur un modèle est, par ordre décroissant d'importance, la plus forte sur les estimations de la variance, les corrélations des questions de même type et, enfin, les corrélations des questions de types différents. Les estimateurs fondés sur un modèle et les estimateurs composites peuvent les uns et les autres êtres calés (ajustement par le quotient) de sorte que les moyennes sur l'ensemble des domaines concordent avec les estimations directes, bien que cela n'ait pas été nécessaire dans notre exemple.

Les fonctions de variance et de covariance généralisées (FVCG) ont plusieurs applications dans nos travaux de recherche en cours. Nous élaborons des méthodes fondées sur la quasi-vraisemblance d'estimation des matrices de covariance pour les moyennes de domaines des questions ordonnées d'enquête, avec représentation de la covariance de deuxième niveau (structurelle) au moyen d'un modèle hiérarchique (O'Malley et Zaslavsky 2004). Les modèles de FVCG sont nécessaires pour obtenir des estimations des variances et des covariances d'échantillonnage, ainsi que pour modifier ces estimations lorsque les moyennes sont réestimées durant la procédure d'ajustement de modèle. Si la variabilité d'échantillonnage des estimations des FVCG est minimale parce que le nombre de domaines est grand, les variances et les covariances prévues des FVCG peuvent être considérées comme étant connues. Cependant, si l'erreur d'échantillonnage des estimations fondées sur les FVCG est importante, il convient d'utiliser un modèle permettant à ces erreurs de se propager tout au long de l'analyse. Dans le cadre de travaux connexes, Fay et Train (1997) ont utilisé un modèle binomial avec un effet de plan pour chaque domaine dans l'estimation bayésienne empirique des taux binomiaux. Notre étude étend cette approche à l'estimation multivariée et à des formats de réponse plus généraux.

Une autre application des FVCG est le calcul des estimations de la variance pour les combinaisons linéaires de moyennes de question afin de faciliter l'estimation de la variance de scores composites, comme ceux utilisés dans la publication des résultats de la CAHPS. Les méthodes décrites à la section 2 sont applicables à l'estimation de la variance de toute fonction de totaux, y compris les fonctions de moyennes, d'autres ratios ou des coefficients de régression.

La méthodologie des FVCG peut être étendue de plusieurs façons. Outre les mesures sommaires des résultats, les fonctions de variance et covariance généralisées (FVCG) peuvent dépendre d'autres variables indépendantes, en particulier celles qui prédiraient mieux les corrélations. Nous avons considéré des variables résumant les profils de réponse, comme la proportion de répondants dans un domaine, mais celles-ci n'ont pas amélioré le modèle. Les FVCG pourraient aussi être étendues à l'échantillonnage à plusieurs degrés.

Remerciements

Les présents travaux ont été financés par la U.S. Agency for Healthcare Research and Quality par la voie de la Consumer Assessments of Health Plans Study (subvention U18 HS09205-06) et par les U.S. Centers for Medicare and Medicaid Services (contrat 500-95-007). Nous remercions Paul D. Cleary de son appui continu durant les travaux, Matt Cioffi de la gestion des données, ainsi qu'Elizabeth Goldstein et Amy Heller des Centers for Medicare and Medicaid Services (CMS), et les autres membres de l'équipe de mise en œuvre de l'Enquête CAHPS-MMC.

Bibliographie

Cho, M.J., Eltinge, J.L., Gershunskaya, J. et Huff, L.L. (2002). Evaluation of generalized variance function estimators for the U.S. Current Employment Survey. Dans *Proceedings of the Joint Statistical Meetings* [CDROM]. Alexandra, VA: American Statistical Association, 534-539.

Eltinge, J. (2002). Use of generalized variance functions in multivariate analysis. Dans *Proceedings of the Section on Survey Research Methods*, Alexandra, VA: American Statistical Association, 904-913.

Fay, R.E., et Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

Fay, R.E., et Train, G.F. (1997). Small domain methodology for estimating income and poverty characteristics for states in 1993. Dans *Proceedings of the Social Statistics Section*, Alexandra, VA: American Statistical Association, 183-188.

Freund, J.E., et Walpole, R.E. (1987). *Mathematical Statistics*. New Jersey: Prentice-Hall, Inc., 4ième Édition.

Gabler, S., Haeder, S. et Lahiri, P. (1999). Justification à base de modèle de la formule de Kish pour les effets de plan de sondage lies à la pondération et à l'effet de grappe. *Techniques d'enquête*, 25, 119-120.

Goldstein, E., Cleary, P.D., Langwell, K.M. Zaslavsky, A.M. et Heller, A. (2001). Medicare Managed Care CAHPS: A tool for performance improvement. *Health Care Financing Review*, 22, 101-107.

Ghosh, M., et Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55-76.

Hays, R.D., Shaul, J.A., Williams, V.S.L., Lubalin, J.S., Harris-Kojetin, L.D., Sweeny, S.F. et Cleary, P.D. (1999). Psychometric properies of the CAHPS 1.0 survey measures. *Medical Care*, 37 (Supplément), 22-31.

Huff, L.L., Eltinge, J.L. et Gershunskaya, J. (2002). Exploratory analysis of generalized variance function models for the U.S. Current Employment Survey. Dans *Proceedings of the Joint Statistical Meetings* [CDROM], Alexandra, VA: American Statistical Association, 1519-1524.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Korn, E.L., et Graubard, B.I. (1999). *Analysis of Health Surveys*. New York: John Wiley & Sons, Inc.

O'Malley, A.J. et Zaslavsky, A.M. (2004). Implementation of cluster-level covariance analysis for survey data with structured nonresponse. Dans *Proceedings of the Joint Statistical Meetings*, American Statistical Association, 1907-1914.

Otto, M.C., et Bell, W.R. (1995). Sampling error modeling of poverty and income statistics for states. Dans *Proceedings of the Section on Government Statistics*, American Statistical Association, 160-165.

Särndal, C.-E., Swensson, B. et Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Seber, G.A.F. (1977). *Linear Regression Analysis*. New York: John Wiley & Sons, Inc.

Spencer, B.D. (2000). Un effet de plan de sondage approximatif pour une pondération inégale en cas de corrélation possible entre les mesures et les probabilités de sélection. *Techniques d'enquête*, 26, 153-155.

Valliant, R. (1992a). Longitudinal smoothing of price index variances. Dans *Statistics Canada Symposium*. Ottawa: Statistique Canada. 113-120.

Valliant, R. (1992b). Smoothing variance estimates for price indexes over time. *Journal of Official Statistics*, 8, 433-444.

Valliant, R., Dorfman, A.H. et Royall, R.M. (2000). *Finite Population Sampling and Inference*. New York: John Wiley & Sons, Inc.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Woodruff, S. (1992). Variance estimation for estimates of employment change in the Current Employment Statistics Survey. Dans *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Alexandria, VA: American Statistical Association, 626-631.

Zaslavsky, A.M., Beaulieu, N.D., Landon, B.E. et Cleary, P.D. (2000). Dimensions of consumer-assessed quality of Medicare managed-care health plans. *Medical Care*, 38, 162-174.

Zaslavsky, A.M., et Cleary, P.D. (2002). Dimensions of plan performance for sick and healthy members on the Consumer Assessments of Health Plans Study 2.0 Survey. *Medical Care*, 40, 951-964.