

ANALYSE DE LA RELATION DOSE-RÉPONSE DANS DES DONNÉES D'ENQUÊTES COMPLEXES

David Judkins, Paul Zador et Varma Nadimpalli¹

RÉSUMÉ

L'analyse de la relation dose-réponse a longtemps joué un grand rôle en toxicologie. Plus récemment, on a eu recours à ce type d'analyse pour évaluer les campagnes d'éducation populaire. Les données réunies dans de telles évaluations viendront probablement de plans types d'échantillonnage des ménages avec toute la complexité que confèrent habituellement les degrés multiples d'enquête, la stratification et les probabilités variables de sélection. Nous proposons en conséquence un nouveau test de Jonckheere-Terpstra par la méthode jackknife pour l'analyse d'une relation monotone dose-réponse. Il sera question ici des résultats d'une simulation de Monte Carlo des propriétés de ce test.

MOTS CLÉS : test de Jonckheere-Terpstra; méthode jackknife; monotonie.

1. INTRODUCTION

Le domaine usuel d'application de l'analyse de la relation dose-réponse est celui des recherches cliniques et épidémiologiques. En revanche, le contexte d'intérêt pour notre recherche est celui de l'évaluation de l'efficacité de la campagne médiatique nationale contre la toxicomanie chez les jeunes, qui est financée par le Congrès américain depuis le début de 1998 dans un but avoué de réduction et de prévention de la toxicomanie dans la jeunesse aussi bien directement qu'indirectement par l'influence sur les parents et d'autres adultes importants dans la vie des jeunes. L'instrument premier d'évaluation de l'efficacité de cette campagne dans la réalisation de ses objectifs est l'Enquête nationale auprès des parents et des jeunes (National Survey of Parents and Youth ou NSPY). C'est là une enquête nationale à domicile basée sur un plan de sondage complexe. La NSPY consiste en une collecte de données initiales et complémentaires auprès d'échantillons nationalement représentatifs de jeunes et de leurs parents. Comme cette campagne est nationale et, par conception, d'une intensité uniforme pour l'essentiel et qu'elle a démarré au même moment dans tout le pays, il a été impossible d'adopter, pour le jugement à porter sur son efficacité, les méthodes classiques d'évaluation par comparaison avant-après et/ou cas-témoins. L'équipe d'évaluation Westat-Annenberg a choisi comme solution de rechange d'estimer la relation exposition-résultat par analogie avec la relation dose-réponse. On trouvera plus de renseignements sur la campagne, l'évaluation et la NSPY dans Hornik et coll. (2001).

L'application de cette stratégie comprend quatre volets : (1) on mesure la dose, c'est-à-dire le degré d'exposition aux messages de la campagne; (2) on mesure la réponse, c'est-à-dire l'ensemble de résultats que la campagne est censée produire; (3) on contrôle les variables possibles de confusion pour la relation dose-réponse; (4) on vérifie l'hypothèse selon laquelle une relation précise exposition-résultat qui a été observée concorde avec le principe d'une juste relation où l'exposition aux messages de la campagne est liée « de la bonne façon » au résultat, c'est-à-dire produit un effet dans le sens désiré.

Il ne sera question ici que du quatrième volet de la stratégie. Comme nous l'avons déjà indiqué, notre intérêt premier était de vérifier, pour chaque association exposition-résultat observée, si la relation était monotone ou non dans la « bonne » direction. La raison en est que la constatation d'une association statistique non seulement significative mais aussi monotone est souvent considérée comme un important élément de confirmation d'un rapport de causalité.

¹ Westat, 1650, boul. Research, Rockville, Maryland, États-Unis 20850.

Nous avons opté pour le test de Jonckheere-Terpstra (Terpstra, 1952; Jonckheere, 1954; Pirie, 1983), c'est-à-dire pour un test non paramétrique où on rejette l'hypothèse nulle d'absence d'association par rapport à d'autres possibilités ordonnées. Un autre avantage de ce test est qu'il est disponible en SAS (SAS/STAT, 1996) et qu'il pouvait entrer dans un ensemble très complexe de données de traitement à programmer déjà en SAS.

Comme il a été conçu, le test JT est destiné et appliqué aux données d'échantillons aléatoires simples (EAS) et pourrait donc ne pas être valable pour les données de la NSPY. Dans la section qui suit, nous présentons une version « jackknife » du test pour des échantillons complexes comme celui de la NSPY. Nous passerons toutefois en revue d'abord les études consacrées aux tests de monotonie.

Dans une étude récente de Leuraud et Benichou (2001), on a mis en parallèle plusieurs méthodes de vérification de l'existence d'une relation monotone dose-réponse pour les *proportions tirées d'échantillons aléatoires simples*. Comme nos échantillons n'en sont pas, c'est-à-dire que les variables de réponse de notre étude peuvent être ordinales et comporter plus de deux niveaux de réponse, et comme le test JT ne figurait pas dans la comparaison de ces auteurs, les résultats de leur étude ne sont pas directement comparables aux nôtres. Leur méthodologie et leurs résultats ne nous donnent pas moins la toile de fond de notre recherche. Leuraud et Benichou ont procédé à une simulation de Monte Carlo afin d'évaluer la puissance de rejet que possèdent quatre tests contre de l'hypothèse nulle d'absence de différences entre classes pour les relations monotones, quasi monotones et strictement non monotones. Leurs résultats sont des plus complexes, mais il semblerait qu'un test par contrastes adjacents conçu par Dosemeci et Benichou (1998) acceptait une relation strictement non monotone « à pic unique » comme monotone avec la moindre fréquence, propriété à rechercher dans les analyses des données de la NSPY. Leuraud et Benichou ont enfin indiqué que ni le chi-test isotoniquement modifié de Barlow et coll. (1972) ni le test modifié de Mantel pour la tendance globale (1963) ne donnaient de bons résultats pour la relation strictement non monotone.

Simpson et Margolin (1990) ont étudié les propriétés asymptotiques d'une catégorie de tests non paramétriques de courbes dose-réponse à décroissance dans la tranche des doses élevées (ces courbes sont définies comme en croissance monotone initiale et en décroissance monotone subséquente dans la tranche supérieure des doses) pour découvrir que, dans bien des distributions paramétriques, le test JT manque de robustesse par rapport à une telle décroissance. Simpson et Margolin ont également constaté que certaines statistiques liées mais bien plus complexes étaient d'un meilleur rendement que le test JT lorsque la relation à vérifier comportait une décroissance.

2. MODIFICATION EN FONCTION D'UN PLAN D'ÉCHANTILLONNAGE COMPLEXE

La NSPY a un plan d'échantillonnage complexe à stratification, à degrés multiples et à pondération inégale. On a préparé pour cette enquête une pondération d'échantillonnage selon les probabilités de sélection et avec des corrections de non-réponse et de sous-observation. On a créé des poids de réplique pour une estimation convergente de la variance des quantités estimées de population finie.

Il aurait été possible d'élaborer un modèle paramétrique pour chaque résultat rapporté à l'exposition en tenant compte des facteurs à taux différenciés d'échantillonnage et des variables de confusion, mais le calendrier du projet n'a pas laissé le temps d'en construire pour les diverses variables de résultat. Un grand avantage de la méthode d'évaluation de propension est qu'elle peut servir à l'élaboration d'un ensemble de valeurs de pondération en projection hypothétique. Avec cette simple pondération, il est possible d'éliminer les effets de confusion de toutes les variables ayant servi à l'évaluation de propension. En joignant cette évaluation au test de Jonckheere-Terpstra, on pouvait espérer produire rapidement un grand nombre de tableaux conçus pour l'examen des effets de la campagne par une diversité de résultats dans une diversité de domaines de population. Pour conserver l'avantage d'une production rapide de tableaux analytiques, tout en tenant compte de l'informativité de l'échantillon (composition familiale et âge du logement) et en intégrant des corrections de taux de réponse différenciés et de sous-observation, on a décidé de fonder une pondération en projection hypothétique sur la pondération d'échantillonnage, puis d'effectuer les calculs du test Jonckheere-Terpstra en se reportant à cette pondération hypothétique. Il y avait alors la question de savoir comment adapter le test JT au plan d'échantillonnage complexe de la NSPY.

Les praticiens des enquêtes ont déjà considérablement réussi à établir comment analyser les tableaux de contingence issus de plans d'échantillonnage complexes. Kish et Frankel (1974) ont d'abord déterminé que, malgré que la mise en grappes a moins d'incidence sur les paramètres fixes des modèles que sur les moyennes marginales, cette incidence n'est pas négligeable pour ce qui est des grandes corrélations intraclasse. Holt et Scott (1981) et Scott et Holt (1982) ont confirmé et étendu ce travail. Rao et Scott (1981) ont examiné des travaux antérieurs pour proposer trois statistiques chi-carré corrigées pour les tableaux à deux entrées, puis par généralisation pour les tableaux à entrées multiples (Rao et Scott, 1984). Fay (1985) a avancé une méthode de test d'indépendance de diverses formes d'indépendance conditionnelle dans les tableaux de contingence au moyen d'une statistique chi-carré par la méthode jackknife. Les statistiques de Rao et Scott sont maintenant des éléments standard dans WesVar et Sudaan.

Plus récemment, Wu, Holt et Holmes (1988) ont montré la gravité de l'omission de la mise en grappes dans l'établissement d'une statistique F globale dans le cas des échantillons à grappes. Ils ont aussi montré comment effectuer la correction. Les chercheurs médicaux ont été plus lents à reconnaître ces problèmes, mais des progrès ont aussi eu lieu récemment dans ce domaine (c.f. Manda, 2002). Nous avons supposé que les problèmes constatés pour d'autres types d'analyses influeraient défavorablement sur le test JT si nous devions le calculer à partir d'une table de contingence à pondération normalisée, aussi avons-nous voulu pour ce test des corrections semblables à celles de Fay ou de Rao et Scott dans le cas des tests chi-carré d'indépendance.

Comme la statistique JT est en distribution normale asymptotique dans une hypothèse nulle d'indépendance, il nous a semblé qu'une manière simple de procéder serait de répéter le test avec chaque réplique de rééchantillonnage, puis de calculer une variance et de l'utiliser dans un test z. Plus précisément, soit J_0 la statistique JT normalisée formée sur le tableau de contingence Y-Z en pondération d'échantillon entier et soit J_r la statistique JT normalisée formée sur ce même tableau avec la r^e réplique de rééchantillonnage. Soit enfin b_r un facteur lié à la r^e réplique et à la méthode de création des poids de réplique. Le test Jonckheere-Terpstra par la méthode jackknife est alors :

$$JJT = \frac{J_0}{\sqrt{\sum_r b_r (J_r - J_0)^2}}$$

À noter que nous donnons ici un sens plus général au terme « jackknife » que de faire voir que les poids de réplique doivent être créés par la méthode jackknife. Les valeurs en question peuvent s'établir par la méthode de « balanced repeated replication » (BRR), une méthode bootstrap ou un plan de rééchantillonnage parmi plusieurs.

On applique une formule tout à fait analogue dans WesVar pour estimer la variance des coefficients de régression. Au lieu de J_0 et J_r , il y a β_0 , coefficient estimé de régression calculé en pondération d'échantillon entier, et β_r , coefficient estimé de régression calculé par le r^e ensemble de poids de réplique. Des recherches antérieures démontrent que cette technique donne généralement de bons résultats, bien qu'une pondération par la méthode jackknife fonctionne bien mieux qu'une pondération par BRR ou la méthode bootstrap pour au moins certaines distributions des erreurs. Connaissant bien l'application de la méthode jackknife aux coefficients de régression et sachant que la statistique Jonckheere-Terpstra ressemble à de tels coefficients, nous avons décidé d'employer le test JJT (test JT par la méthode jackknife) pour les données de la NSPY. Le calendrier de publication ne nous a pas laissé le temps de développer la théorie, pas même des études de simulation. Nous n'avons toujours pas essayé d'élaborer cette théorie, mais nous avons mené une étude de simulation qui est l'objet même du présent exposé.

3. PARAMÈTRES DE L'ÉTUDE DE SIMULATION

Nous avons choisi de simuler les seules caractéristiques du plan d'échantillonnage de la NSPY qui paraissent devoir le plus influencer sur le rendement du test JJT, à savoir les grappes au niveau UPE, la variation de taille des grappes et 100 valeurs de poids de réplique pour l'estimation de variance. Nous n'avons pas voulu mettre la stratification en simulation, ni une pondération différentielle quelconque. Nous avons fixé les niveaux de corrélation

intraclasse au niveau UPE comme ils auraient pu exister si nous avions fait intervenir la stratification. La variation de la taille des grappes offre de l'intérêt, parce que le rendement de la technique de sélection varie naturellement selon les UPE et que la NSPY est une enquête de fin de décennie. Précisons que chaque recensement décennal aux États-Unis sert à établir les probabilités de sélection des UPE de la plupart des enquêtes. En fait, on sélectionne les UPE tôt dans la décennie et celles-ci entrent dans une diversité d'enquêtes au cours de la décennie. Le jeu optimal de probabilités de sélection serait proportionnel à l'ensemble de la population admissible au moment de la collecte de données d'enquête. À mesure que s'écoule la décennie, la construction neuve, l'accroissement naturel, l'immigration et la migration interne concourent à une dégradation de la qualité des probabilités de sélection, ce qui fait varier la taille d'échantillon des grappes.

La variation de cette taille influe sur la variance des estimations d'enquête, qui est plus difficile à projeter que l'incidence de la corrélation intraclasse. Elle dépend assez étroitement de la nature des analyses effectuées. Une règle empirique générale est qu'un surcroît de conditionnement réduit probablement cette incidence. Il reste que, outre l'effet qu'elle produit sur la variance, la variation de la taille d'échantillon, des grappes en échantillon agit sur l'application du théorème central limite aux estimations d'enquête. L'éléphant de Basu se situe à l'extrémité de la variation de taille des grappes, aussi avons-nous décidé d'intégrer à notre simulation une variation plutôt ample de la taille d'échantillonnage des grappes.

Nous avons jugé bon de mettre en simulation 100 UPE de l'échantillon et 100 poids de réplique, car c'est là un plan d'échantillonnage et une stratégie d'estimation de variance courants à Westat. Nous avons produit les tailles d'échantillon UPE comme iid $n_i \sim \Gamma(7,5,7) + \Gamma(0,3,33,3)$, en arrondis au nombre naturel le plus proche. Pour une réplique, nous avons obtenu une moyenne de 50 unités (personnes ou ménages) par UPE avec un écart-type de 24, une asymétrie de 1,9, un minimum de 12 et un maximum de 202. L'ordre de grandeur approximatif était de 5 000 dans l'ensemble, mais la taille d'échantillon totale a été laissée comme variable aléatoire.

Nous avons produit des poids de réplique par la méthode jackknife. C'est ainsi que chacune des 100 UPE a été retranchée à son tour pour une réplique. Nous avons corrigé d'un facteur de 100/99 les poids de toutes les UPE restantes d'une réplique. Nous avons égalé à une constante tous les poids d'échantillon entier. Dans ce plan de répétition, les facteurs de réplique sont $b_r = 99/100$.

Variable d'exposition. Pour simuler la variable ordinaire Y d'exposition ou de dose, nous nous sommes reportés à une distribution normale double et avons ainsi simulé une cote latente au niveau des personnes, puis l'avons notée 0, 1, 2, 3, les seuils étant sélectionnés comme quartiles de la distribution latente. Au niveau des UPE, nous avons laissé la cote d'exposition latente dépendre de la taille des grappes selon la distribution

$$\mu_i \sim N \left(1 - \left(\frac{4}{n_i} \right)^{0.3}, \sigma_{y1}^2 \right).$$

Au niveau des personnes, nous avons simulé les cotes Y d'exposition latente par $\mu_{ij} \sim N(\mu_i, \sigma_{y2}^2)$, où nous avons varié σ_{y1}^2 et σ_{y2}^2 pour créer différents niveaux de corrélation intraclasse,

$$\rho_y = \frac{\sigma_{y1}^2}{\sigma_{y1}^2 + \sigma_{y2}^2},$$

tout en gardant fixe la variance totale $\sigma_{y1}^2 + \sigma_{y2}^2 = c$.

Nous avons ensuite traduit les cotes d'exposition latente Y en cotes ordinaires manifestes représentant les quartiles de μ .

Variable de résultat. Pour la variable de résultat Z , nous avons élaboré un modèle avec une double perturbation de la moyenne marginale et de l'écart de la relation exposition-résultat au niveau des UPE et une relation exposition-résultat au niveau des personnes. Nous avons simulé la perturbation de moyenne marginale d'UPE par $\xi_i \sim N(0, \sigma_{z1}^2)$ et la perturbation de l'écart de relation UPE par $\theta_i \sim N(\alpha, \sigma_\theta^2)$. On peut voir dans α une pente et dans σ_θ^2 une mesure de la variance entre UPE de l'écart de la relation dose-réponse.

Au niveau des personnes, nous avons simulé les cotes d'exposition latente Z par $\xi_{ij} \sim N(\xi_i + f_\lambda(y_{ij}), \sigma_{z2}^2)$, où nous avons sélectionné 8 formes pour la dépendance de la cote latente Z à l'égard de la cote manifeste Y et varié une fois de plus σ_{z1}^2 et σ_{z2}^2 pour restituer la diversité des corrélations intraclasse. Voici les huit configurations ainsi définies :

Planéité :	$f_A(y_{ij}) = 1.5\theta_i / 3$
Linéarité :	$f_B(y_{ij}) = \theta_i y_{ij} / 3$
Quadraticité :	$f_C(y_{ij}) = \theta_i \sqrt{y_{ij} / 3}$
Quatrième puissance :	$f_D(y_{ij}) = \theta_i (y_{ij} / 3)^4$
Bond prochain :	$f_E(y_{ij}) = \theta_i [\text{if } y_{ij} > 0]$ et 0 dans les autres cas
Bond lointain :	$f_F(y_{ij}) = \theta_i [\text{if } y_{ij} = 3]$ et 0 dans les autres cas
Butte centrale :	$f_G(y_{ij}) = \theta_i [\text{if } y_{ij} = 1 \text{ or } 2]$ et 0 dans les autres cas
Pointe prochaine :	$f_H(y_{ij}) = \theta_i [\text{if } y_{ij} = 1]$ et 0 dans les autres cas

La figure 1 indique le rapport entre la cote latente moyenne Z (axe vertical) et la cote manifeste Y (axe horizontal) pour chacune des 8 formes et la moyenne de 10 tirages aléatoires de la codistribution avec $\alpha=0,66$ et $\sigma_\theta^2=0$. Les formes B à F dénotent une relation monotone, la forme A, l'indépendance, et les formes G et H, une dépendance non monotone. Nous avons varié le niveau de α pour créer des formes d'une constatation plus ou moins facile.

On notera que la corrélation intraclasse de Z est plutôt complexe. Pour la planéité, elle est

$$\rho_z = \frac{\sigma_{z1}^2 + \sigma_\theta^2 / 4}{\sigma_{z1}^2 + \sigma_\theta^2 / 4 + \sigma_{z2}^2}.$$

Nous avons ensuite mis les cotes latentes Z au niveau des personnes sous la forme de la variable ordinale manifeste Z en divisant ξ à ses quartiles.

Échantillon. Nous avons tiré un échantillon de 2 000 de cette distribution pour chaque forme de dépendance et chaque niveau de corrélation intraclasse. Nous espérons que l'hypothèse nulle d'indépendance serait rejetée seulement 5 % des fois pour les formes A, G et H et que la puissance de rejet serait assez grande pour les formes B à F. Dans cette simulation, nous avons seulement comparé le test JJT au test ordinaire JT. Il serait intéressant de mettre en comparaison les rendements d'autres tests comme celui que proposent Leuraud et Benichou, mais nous ne nous sommes pas engagés dans cette voie. Nous prévoyions que le test ordinaire JT serait plus puissant pour toutes les formes, ce qui vaut, bien sûr, uniquement pour les formes B à F.

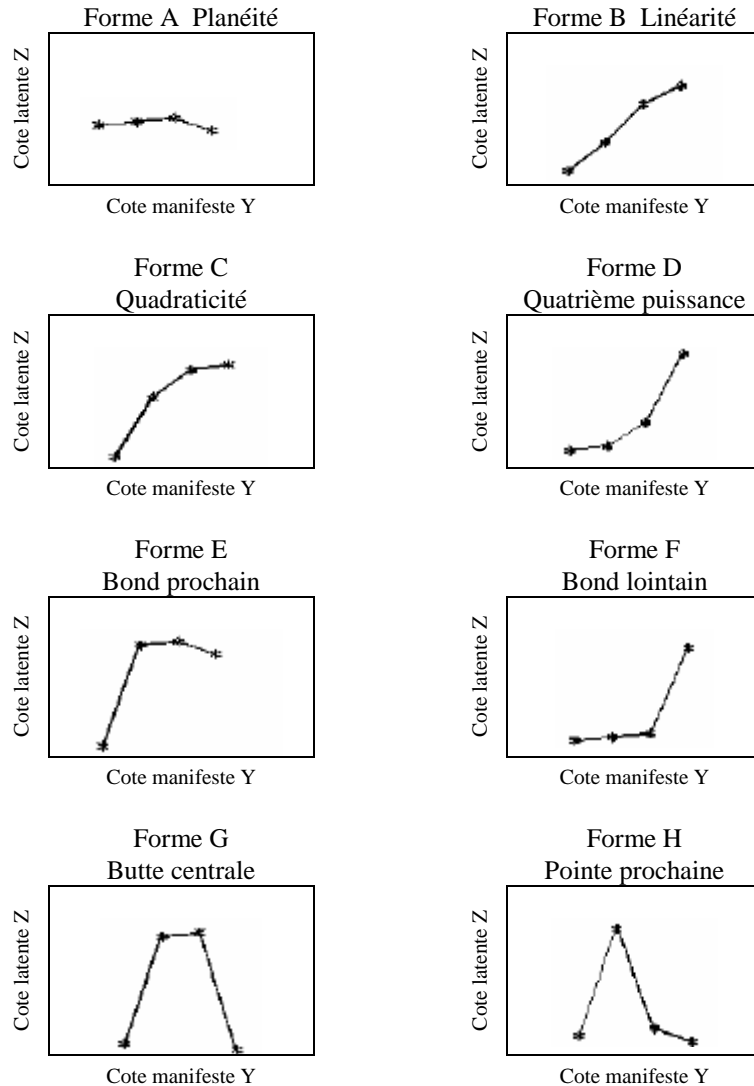


Figure 1. Huit formes testées en simulation

4. RÉSULTATS

La principale statistique d'intérêt vise la taille réelle du test, c'est-à-dire le pourcentage de tirages à l'égard desquels l'hypothèse nulle a été rejetée alors qu'elle se vérifiait. D'un intérêt secondaire est la puissance du test sous diverses autres hypothèses et d'autres encore qui, à proprement parler, ne sont ni nulles ni alternatives.

La figure 2 indique la taille réelle du test en fonction de la corrélation intraclasse. Nous avons appliqué un test bilatéral d'une taille nominale de 0,05. Toute valeur de plus de 0,05 est l'indice que le test est trop « mordant », c'est-à-dire rejette à un taux supérieur à la taille nominale. Nous nous sommes reportés à la même corrélation intraclasse pour les variables tant d'exposition que de résultat. À noter que le test JT ordinaire est d'un très bon rendement pour une corrélation intraclasse qui varie de faible à moyenne. Dans la plupart des enquêtes, on s'attendrait à une corrélation d'au plus 0,005 au niveau des UPE. Ce n'est que lorsque la corrélation intraclasse est d'environ 0,05 ou plus que le test ordinaire ne donne plus le niveau nominal de signification. Si cette corrélation est d'environ 0,1, il devient beaucoup trop « mordant ». Le test JJT (par la méthode jackknife) préserve le niveau de

signification à tous les niveaux de corrélation intraclasse. Il convient de noter que, avec 2 000 tirages, l'intervalle de confiance à 95 % pour la puissance estimée est de plus ou moins un point (en pourcentage).

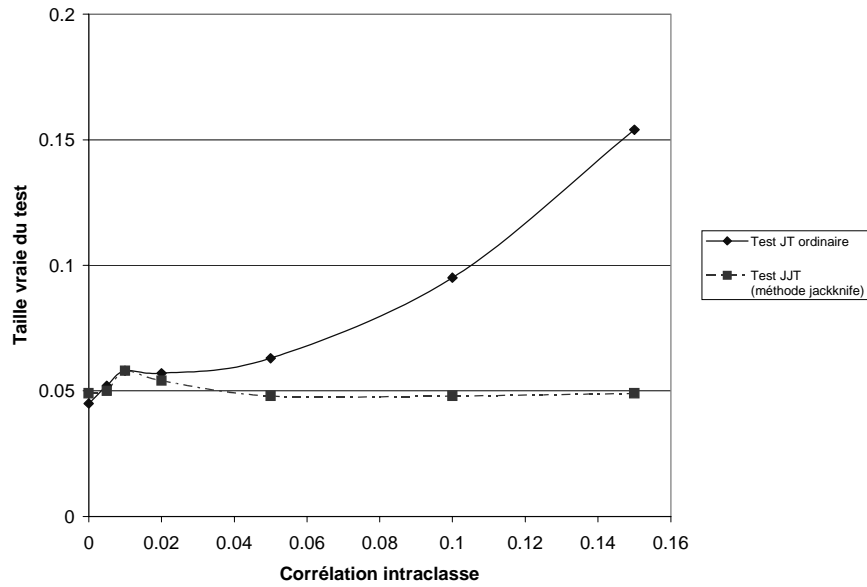


Figure 2. Probabilité de rejet de l'hypothèse nulle de planéité où la relation est en effet plate (même corrélation intraclasse et variance totale de 35 pour Y et Z)

La figure 3 décrit la puissance du test JJT pour des formes monotones avec $\alpha=0,66$ et $\sigma_{\theta}^2=0$. Nous pouvons voir ici que la puissance est toujours optimale pour la linéarité. Suit en optimalité la puissance des formes lisses non linéaires. La puissance est la moindre pour les formes non lisses et non linéaires où toute la variation du résultat est liée à une seule des transitions d'exposition. On peut enfin voir que la puissance décroît sur l'ensemble des formes à mesure que s'accroît la corrélation intraclasse, ce à quoi on pouvait s'attendre.

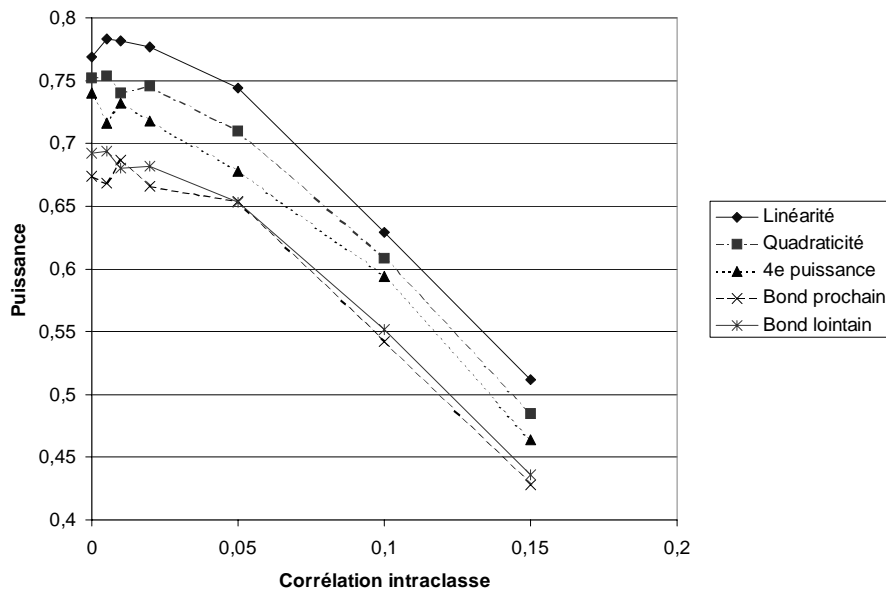


Figure 3. Puissance du test JJT pour une diversité de pentes et de formes monotones

La figure 4 indique la puissance pour les formes non monotones de dépendance. Ces formes jugées illogiques ont été écartées au départ dans l'application du test JT. Comme nous l'avons dit, l'hypothèse nulle est que les deux variables sont indépendantes et l'hypothèse alternative, qu'il existe une relation monotone dose-réponse. Les formes non monotones se situent hors de l'espace paramétrique. Des situations sont néanmoins possibles où se présenteront des formes non monotones pour des raisons complexes. Nous avons appliqué le test JT en espérant un taux de rejet de l'hypothèse nulle ne dépassant pas 5 % pour les formes non monotones. Ce souhait est comblé par la forme de la butte centrale, mais en partie déçu par la forme de la pointe prochaine. Il reste que, comme le test JJT est généralement moins puissant que le test JT ordinaire, il crée moins de fausses constatations de tendances monotones même avec la forme de la pointe prochaine.

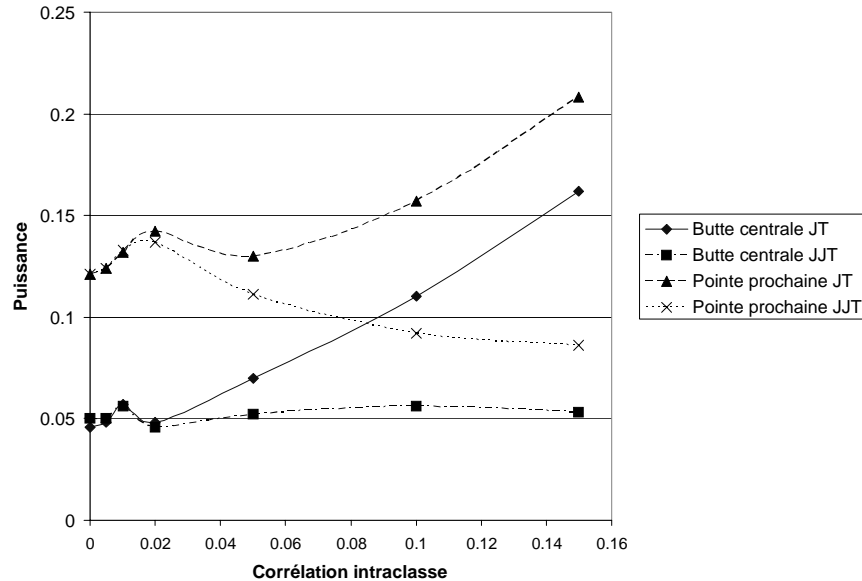


Figure 4. Puissance des tests JT ordinaires et JJT (méthode jackknife) pour une diversité de pentes et de formes non monotones de dépendance

La figure 5 indique comment la puissance du test JJT dépend de l'origine de la corrélation intraclasse pour la variable de résultat. Plus précisément, elle décrit l'effet de corrélation intraclasse engendré en tout ou en partie par la variabilité de l'écart de la relation entre les grappes par opposition à la variabilité des tendances sous-jacentes des résultats. À la figure 3, nous avons simulé les valeurs de $\rho_z = 0.1$ avec $\sigma_{z1}^2 = 0.1 \times 35$, $\sigma_{z2}^2 = 0.9 \times 35$ et $\sigma_{\theta}^2 = 0$. Lorsque nous avons simulé $\rho_z = 0.1$ avec $\sigma_{z1}^2 = \frac{0.1 \times 35}{2}$, $\sigma_{z2}^2 = 0.9 \times 35$ et $\sigma_{\theta}^2 = 7$ pour que la moitié de la corrélation intraclasse soit due à des pentes variables pour le résultat, la puissance est tombée de la fourchette 54-63 % à la fourchette 32-37 %. Lorsque nous avons poussé la chose et simulé $\rho_z = 0.1$ avec $\sigma_{z1}^2 = 0$, $\sigma_{z2}^2 = 0.9 \times 35$ et $\sigma_{\theta}^2 = 14$ pour que toute la corrélation intraclasse soit due à des pentes variables pour le résultat, la puissance n'était plus que de 26 % à 28 %. Cette « perte » de puissance peut être trompeuse; le test JT est censé vérifier une relation monotone dose-réponse qui est universelle. Si l'écart de la relation varie dans une large mesure à cause d'une interaction avec des covariables inconnues, il peut se révéler indésirable de conclure à l'universalité de la relation monotone.

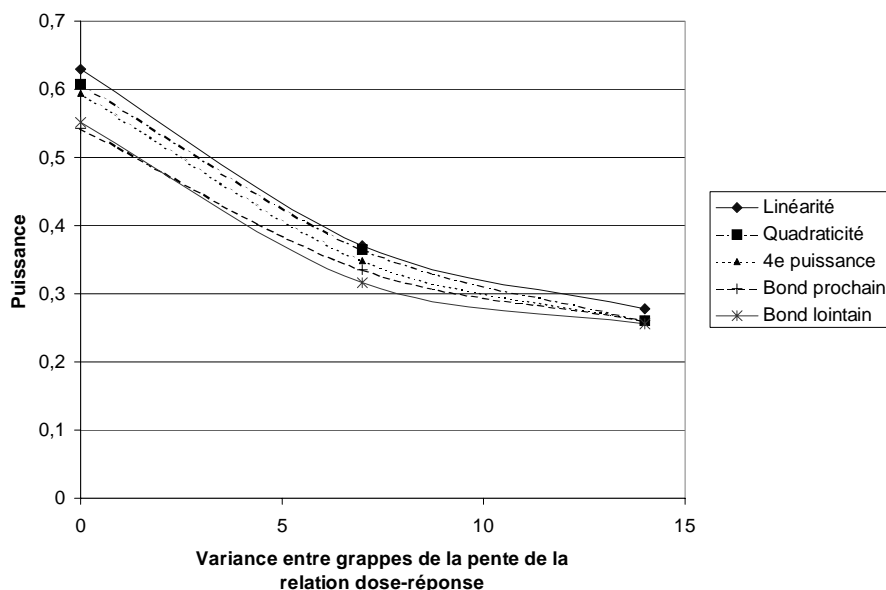


Figure 5. Puissance du test JJT (méthode jackknife) selon la source de corrélation intraclasse de la variable de résultat

5. RECOMMANDATIONS

Nous avons démontré que le test JT par la méthode jackknife (test JJT) est d'un rendement acceptable dans le cas des relations monotones dose-réponse pour une structure par grappes d'une enquête par échantillonnage complexe, un plan à mesures répétées ou un plan de mise en grappes aléatoire. Il accepte l'hypothèse nulle au taux désiré lorsque la forme vraie est plate ou symétrique (forme de la butte centrale, par exemple). Il rejette cette même hypothèse avec une puissance qui n'est que légèrement inférieure à celle du test JT ordinaire lorsque la taille réelle du JT nominal est proche de sa taille nominale. Le test JJT accepte l'hypothèse alternative d'une relation monotone dose-réponse plus souvent qu'on ne le désire pour une forme vraie de dépendance asymétrique et non monotone (forme de la pointe prochaine, par exemple), mais il la rejette aussi plus souvent que non et se comporte bien mieux que le test JT ordinaire s'il existe une forte structure par grappes.

Les avantages du test JJT se remarquent d'emblée à des niveaux de corrélation intraclasse qui peuvent ne pas souvent se réaliser. Toutefois, comme la perte de puissance est infime dans les cas où la correction est inutile, nous recommandons de toujours appliquer cette méthode à des données en grappes sans égard au degré prévu de corrélation intraclasse.

RÉFÉRENCES

- Barlow, R. E., Bartholmew, D. J., Bremner, J. M., et Brunk, H. D. (1972), *Statistical Inference under Order Restrictions. The Theory and Application of Isotonic Regression*, New York:Wiley.
- Collings, B. J., Margolin, B. H., et Oehlert, G. W. (1981), "Analyses for Binomial Data, with Applications to the Fluctuation Test for Mutagenicity", *Biometrics* 37, pp. 775-794.
- Dosemeci M., et Benichou J. (1998), "An alternative test for trend in exposure-response analysis", *Journal of Exposure Analysis and Environmental Epidemiology*, 8 (1), pp. 9-15.

- Fay, R. E. (1985), "A jackknifed chi-squared test for complex samples", *Journal of the American Statistical Association*, 80, pp. 148-157.
- Holt, D., et Scott, A. J. (1981), "Regression analysis using survey data", *The Statistician*, 30, pp. 169-178.
- Hornik, R., Maklan, D., Orwin, R., Cadell, D., Judkins, D., Barmada, C., Yanovitzky, I., Moser, M., Zador, P., Southwell, B., Baskin, R., Morin, C., Jacobsohn, L., Prado, A., et Steele, D., (2001). Evaluation of the National Youth Anti-Drug Media Campaign: Third Semi-Annual Report of Findings - October 2001. Rockville, Maryland: Westat.
- Imbens, G. W. (1999), "The role of propensity score in estimating dose-response functions". Technical Working Paper 237 (<http://www.nber.org/papers/T0237>). Cambridge, MA:NBER.
- Jonckheere, A. R. (1954), "Distribution-free k-sample test against ordered alternatives", *Biometrika*, 7, pp. 93-100.
- Kish, L., et Frankel, M. R. (1974), "Inference from complex samples (with discussion)", *Journal of the Royal Statistical Society, Series B*, 36, pp. 1-37.
- Leraud, K., et Benichou J. (2001), "A comparison of several methods to test for the existence of monotone dose-response relationship in clinical and epidemiological studies", *Statist. Med.*, 20, pp. 3335-3351.
- Manda, S. O. M. (2002), "A Bayesian ordinal model for heterogeneity in a multi-centre myocardial infarction clinical trial", *Statistics in Medicine*, 21, pp. 3011-3022.
- Mantel N. (1963), "Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure", *Journal of the American Statistical Association*, 58, pp. 690-700.
- Pirie, W. (1983), "Jonckheere tests for ordered alternatives", in Kotz, S., and Johnson, N. L (eds.) *Encyclopedia of Statistical Sciences*, New York:Wiley, vol. 4, pp. 315-318.
- Rao, J. N. K., et Scott, A. J. (1981), "The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables", *Journal of the American Statistical Association*, 76, pp. 221-230.
- Rao, J. N. K., et Scott, A. J. (1984), "On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data", *The Annals of Statistics*, 12, pp. 46-60.
- Rosenbaum P. R. (1995), *Observational Studies*, New York:Springer-Verlag.
- Rosenbaum P., et Rubin, R. (1983), "The central role of propensity score in observational studies for causal effects", *Biometrika*, 70, pp. 41-55.
- SAS/STAT Software: Changes and Enhancements for Release 6.12. (1996), Cary, N.C.:SAS Institute, Inc.
- Scott, A. J., et Holt, D. (1982), "The effects of two-stage sampling on ordinary least squares methods", *Journal of the American Statistical Association*, 77, pp. 848-854.
- Simpson D. G., et Margolin B. H. (1990), "Nonparametric testing for dose-response curves subject to downturns: asymptotic power considerations", *The Annals of Statistics*, 18, No 1, pp. 372-390.
- Terpstra, T. J. (1952), "The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking", *Indag. Mat.*, 14, pp. 327-333.
- Wu, C. F. J., Holt, D., et Holmes, D. J. (1988), "The effect of two-stage sampling on the F statistic", *Journal of the American Statistical Association*, 83, pp. 150-159.