



HAL
open science

Méthodes statistiques globales et locales d'analyse d'un tableau de contingence par les tailles d'effet et leurs intervalles de confiance

Brice Lefevre, Stéphane Champely

► To cite this version:

Brice Lefevre, Stéphane Champely. Méthodes statistiques globales et locales d'analyse d'un tableau de contingence par les tailles d'effet et leurs intervalles de confiance. Bulletin de Méthodologie Sociologique / Bulletin of Sociological Methodology, 2009, 103 (1), pp.50-65. hal-01713159

HAL Id: hal-01713159

<https://insep.hal.science/hal-01713159v1>

Submitted on 20 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TITRE : Méthodes statistiques globales et locales d'analyse d'un tableau de contingence par les tailles d'effet et leurs intervalles de confiance

AUTEURS : Brice Lefèvre (1) et Stéphane Champely (2)

(1) Institut national du sport (INSEP) – Département des sciences, laboratoire de sociologie, 11 avenue du Tremblay, 75012 Paris, France. Tél. : 01 41 74 44 69. e-mail : brice.lefevre@insep.fr

(2) Université de Lyon – Université Lyon I, EA 647, Centre de Recherche et d'Innovation sur le Sport (CRIS), 27-29 Boulevard du 11 novembre 1918, 69622 Villeurbanne cedex, France. Tél. : 04 72 43 28 45. e-mail : champely@univ-lyon1.fr

Publié dans : Bulletin de méthodologie sociologique, 2009, n°103, pp. 50-65

Résumé : Cet article propose une méthode d'analyse globale et locale d'un tableau de contingence suivant les « bonnes pratiques statistiques » comme définies par Wilkinson et al. (1999) c'est-à-dire en ayant principalement recours à des mesures de taille d'effet, des intervalles de confiance et des graphiques plutôt qu'à des tests de significativité. Au niveau global, la relation entre les deux variables qualitatives est mesurée par le V de Cramér. Un intervalle de confiance basé sur la loi du chi-carré décentrée peut être calculé par inversion du test de significativité. La méthode du bootstrap est également utilisée. Au niveau local, le PEM de Cibois (Pourcentage d'Ecart Maximum à l'indépendance) constitue une mesure de taille d'effet permettant d'estimer l'écart de chaque case du tableau de contingence à l'hypothèse d'indépendance. La procédure de ré-échantillonnage bootstrap en donne un intervalle de confiance. Un graphique original résumant ces informations pour l'ensemble du tableau est proposé.

Abstract : This paper presents global and local analyses of a contingency table along the lines of Wilkinson et al. (1999) defining good statistical practices as using effect sizes, confidence intervals and plots instead of tests of significance. On one hand, the global relationship between the two qualitative variables is measured by Cramer's V. A confidence interval may be calculated using the decentred chi-squared distribution by inverting the corresponding significance test. A bootstrap confidence interval is also proposed. On the other hand, Cibois's PEM (percentage of maximum deviation from independence) is a local effect size that can be computed for every cell of the contingency table giving its own distance from the independence hypothesis. A corresponding confidence interval is calculated by the previous resampling bootstrap technique. An original plot summarizing these pieces of information for the whole contingency table is proposed.

Mots-clefs : Tableau de contingence, V de Cramér, taille d'effet, intervalle de confiance, bootstrap, PEM.

INTRODUCTION

Dans les sciences sociales et humaines, les enquêtes par questionnaire, mais parfois aussi les observations, conduisent généralement à des données qualitatives. Dans un premier temps, le sociologue analyse à l'aide d'un logiciel *ad hoc* les variables qualitatives une par une et dans un second temps deux par deux pour affiner la description de l'objet étudié. Cette opération de tri croisé produit un tableau à double entrée autrement appelé tableau de contingence : « *Le tableau croisé étant l'outil de base quantitatif du sociologue, il importe qu'il dispose d'une technique de lecture rapide pour examiner les nombreux résultats de croisement qui peuvent lui être fournis par un logiciel de dépouillement d'enquête.* » (Cibois, 1990, p.13). Ces deux étapes peuvent ou non être complétées par des analyses multivariées descriptives avec notamment l'analyse factorielle des correspondances et l'enchaînement analyse factorielle des correspondances multiples/classification ou encore multivariées inférentielles avec le modèle log-linéaire.

Ainsi, dans le cadre d'une étude générale sur les activités physiques et sportives (APS) des Français, on étudie dans la sous-population des seniors franciliens la relation entre l'âge (3 classes d'âge) et la fréquence de pratiques d'une APS (3 modalités également : non pratiquant, pratiquant moins d'une fois par semaine, pratiquant une fois par semaine ou plus). Le croisement de ces deux informations et le calcul correspondant des effectifs est résumé dans le Tableau 1 de contingence.

Tableau 1

Le mode usuel d'analyse statistique d'un tel tableau est d'employer le test de significativité du chi-carré d'indépendance de Pearson afin de s'assurer de la liaison globale entre les deux variables c'est-à-dire entre l'ensemble des lignes et des colonnes du tableau. Si cette liaison est significative, elle est détaillée localement en observant quelle ligne est surtout reliée à quelle colonne au moyen de fréquences conditionnelles voire de leur représentation graphique (Martin, 2005). A la suite des propositions méthodologiques de Wilkinson et al. (1999), d'Ely (1999) et surtout de Thompson (2002) pour de « bonnes pratiques statistiques », il semble intéressant plutôt que de se focaliser sur les tests de significativité de s'intéresser à la force de la liaison observée ce que mesure précisément la notion de *taille d'effet*. Cela permet d'effectuer des comparaisons entre tableaux de contingence au sein de la même étude ou même entre études dans le cadre d'une méta-analyse. Quant à l'incertitude concernant la force de cette liaison, elle peut être mieux appréhendée par des intervalles de confiance de ces mêmes tailles d'effet.

Nous allons donc dans une première partie étudier la mesure de taille d'effet la plus classique de liaison globale pour une table de contingence, le V de Cramér, et montrer comment elle peut faire l'objet d'un calcul d'intervalle de confiance. Deux méthodes seront présentées, l'une basée sur des lois de probabilité décentrées (Smithson, 2003) et l'autre reposant sur le *bootstrap* qui est une méthode « moderne » d'inférence statistique qui utilise un algorithme général simple basé sur la puissance de calcul de l'ordinateur. Dans une seconde partie, nous reprendrons la proposition de Cibois (1993) : le *Pourcentage de l'Ecart Maximum à l'indépendance* (PEM) *local* qui autorise, dans l'esprit du V de Cramér, la mesure de la liaison au niveau local. La technique du *bootstrap* permettra là encore de calculer des intervalles de confiance du PEM local et autorisera une représentation graphique inédite et prometteuse d'un tableau de contingence.

PARTIE 1 LIAISON GLOBALE :

Bien qu'il soit régulièrement utilisé par les sociologues, nous commencerons par un rappel détaillé du test du chi-carré d'indépendance de Pearson. Il convient d'en avoir le principe à l'esprit pour bien saisir les développements qui suivent.

RAPPELS SUR LE TEST D'INDEPENDANCE DE PEARSON

On notera $L(=3)$ le nombre de lignes du tableau et $C(=3)$ le nombre de ses colonnes. L'effectif correspondant à la ligne i et la colonne j sera noté n_{ij} . Ainsi, l'effectif $n_{31}=74$ indique le nombre de non pratiquants chez les hommes âgés de 60 à 65 ans. Afin d'illustrer plus concrètement nos propos, nous nous focaliserons sur cette case particulière du tableau dans la suite des analyses.

L'effectif total de la ligne i sera noté n_{i+} . On a $n_{3+}=74+8+32=114$ hommes âgés de 60 à 65 ans. De la même façon, l'effectif total de la colonne j sera noté n_{+j} . On a $n_{+1}=56+35+74=165$ seniors non pratiquants. Enfin, l'effectif total sera noté $n_{++}=293$.

La situation d'*indépendance statistique* signifie que les deux variables considérées (âge et fréquence de pratique des APS) ne sont pas liées. Pour la case (3,1) que nous avons décidé d'étudier, cela signifie que parmi les hommes de 60 à 65 ans qui sont au nombre de $n_{3+}=114$, on observe la même proportion de non pratiquants que dans l'ensemble de l'échantillon soit $n_{+1}/n_{++}=165/293=56.3\%$. On s'attend donc, si cette hypothèse d'indépendance statistique est vérifiée, à observer $114*56.3\%=64.2$ unités dans cette case. Cette quantité est appelée *effectif théorique* et sera notée t_{31} . De façon générale, la formule de calcul des effectifs théoriques est donc :

$$t_{ij}=n_{i+}*n_{+j}/n_{++}.$$

Les calculs d'effectifs théoriques sont présentés dans le coin supérieur gauche de chaque case du Tableau 2.

Tableau 2

Afin de déterminer dans quelle mesure les deux variables sont liées, il faut donc s'intéresser à l'écart entre l'effectif réellement observé (n_{ij}) et l'effectif théorique correspondant (t_{ij}). Cet *écart à l'indépendance* sera noté $e_{ij}=n_{ij}-t_{ij}$. Nous obtenons pour la case de référence $e_{31}=74-64.2=9.8$, un nombre positif, qui signifie qu'il y a plus de non pratiquants dans cette tranche d'âge de 60 à 65 ans que dans l'ensemble de l'échantillon. Les écarts à l'indépendance sont présentés dans le coin supérieur droit de chaque case du Tableau 2.

Un écart de 9.8 unités comme celui observé n'a toutefois pas tout à fait la même signification suivant que la case possède un petit ou un grand effectif. C'est pourquoi, afin d'en quelque sorte « standardiser » (1) ces écarts, sont calculés les *résidus de Pearson* : $r_{ij}=e_{ij}/\sqrt{t_{ij}}$. Nous avons par exemple : $r_{31}=9.8/\sqrt{64.2}=1.22$. Pour résumer l'ensemble des écarts à l'indépendance, on fait la somme des carrés des résidus de Pearson, ce qui donne la *statistique du chi-carré d'indépendance de Pearson* notée X^2 . Plus cette quantité est élevée, plus la liaison observée est importante. On obtient en l'espèce $X^2=10.1$. Cette valeur est comparée à une table de la loi de probabilité du chi-carré avec $df=(L-1)*(C-1)=(3-1)*(3-1)=4$ degrés de liberté. Les logiciels statistiques indiquent le plus souvent une probabilité critique (p value), qui est la probabilité que la loi du chi-carré soit supérieure à la valeur observée X^2 . Elle est ici de $p=0.04$. Lorsque cette probabilité est faible, le seuil de décision conventionnel étant de 5%,

on dit que l'on rejette l'hypothèse d'indépendance. C'est bien le cas ici (2). La liaison entre l'âge et la fréquence de pratique est par conséquent déclarée statistiquement significative.

UNE MESURE DE TAILLE D'EFFET POUR UN TABLEAU DE CONTINGENCE : LE V DE CRAMER

La statistique X^2 est intéressante car elle est nulle en situation (observée) parfaite d'indépendance et elle indique dans quelle mesure l'échantillon diverge de cette hypothèse. Malheureusement, elle dépend à la fois de la taille d'échantillon (n_{++}) et du nombre de modalités (L et C) ce qui la rend difficilement comparable d'une étude à l'autre. Cramér a en effet montré que la valeur maximum de X^2 lorsque la relation est parfaitement dépendante est $m = \max(X^2) = (\min(L, C) - 1) * n_{++}$. D'où l'idée de « standardiser » la valeur du X^2 par cette quantité maximum, comme le fait la statistique V de Cramér : $V = \sqrt{X^2/m}$, qui prend donc ses valeurs entre 0 (indépendance parfaite) et 1 (dépendance parfaite).

On obtient en l'espèce $m = (\min(3, 3) - 1) * 293 = 586$ et donc $V = \sqrt{10.1/586} = 0.13$. A quelle aune juger cette quantité ? Cohen (1988) a proposé des mesures conventionnelles pour une taille d'effet w qui est identique à la statistique V de Cramér lorsque le tableau de contingence n'a que 2 lignes et 2 colonnes. En revanche, la taille d'effet de Cohen, contrairement à celle de Cramér, présente l'inconvénient de ne pas tenir compte de la taille du tableau, c'est-à-dire de L et C. On va reprendre en première approximation (3) ses propositions. Ainsi $V = 0.1$ correspondrait à un effet faible, $V = 0.3$ à un effet moyen et $V = 0.5$ à un effet fort. Par rapport à cette grille de lecture, nous observerions un effet classé comme plutôt faible. La taille d'échantillon ($n_{++} = 293$) a permis de justesse de détecter statistiquement cet effet.

UN INTERVALLE DE CONFIANCE « CLASSIQUE » DU V DE CRAMER

Si la taille d'effet V apporte une information essentielle en mesurant la significativité pratique de la liaison et donc en permettant des comparaisons d'une étude à l'autre, il convient de mesurer l'incertitude qui entoure cette information. Obtenir un intervalle de confiance pour des tailles d'effet standardisées est relativement complexe mais peut parfois se faire en utilisant des lois de probabilités décentrées (Cummings et Finch, 2001).

Lorsque l'hypothèse d'indépendance est vérifiée, la statistique X^2 suit effectivement une loi du chi-carré (χ^2). Mais lorsque cette hypothèse est fautive, elle suit une loi dite du chi-carré décentrée indicée par le degré de liberté df de la table et surtout un *paramètre de centralité* (Δ). La taille d'effet théorique v , que nous cherchons à estimer par la statistique V est fonction de ce paramètre : $v = \sqrt{\Delta/m}$. La procédure décrite dans Smithson (2003) (4) consiste à rechercher de façon numérique par ordinateur les valeurs du paramètre de centralité compatibles avec la valeur observée de X^2 à un niveau de confiance donné, d'où un intervalle de confiance du paramètre de non centralité. Cet intervalle est ensuite transformé en intervalle de confiance de la taille d'effet v .

Cette procédure (5) a été programmée à l'aide du logiciel statistique libre de distribution R (R Development Core Team, 2007). Elle donne un intervalle de confiance à 90% de $0.02 \leq v \leq 0.18$. On voit donc que la borne supérieure de l'intervalle de confiance ne semble pas compatible avec un effet moyen au sens de la grille de Cohen mais que nous sommes tout proches de zéro ce qu'indiquait le résultat du test du chi-carré. Notons qu'au cas où le test à 5% n'est pas significatif, l'intervalle à 90% contient la valeur zéro.

UN INTERVALLE DE CONFIANCE DU V DE CRAMER PAR BOOTSTRAP

La procédure précédente de calcul de l'intervalle de confiance du V de Cramér est complexe. Elle fait appel à la théorie des lois de probabilité décentrées, à celle de l'inversion des tests de significativité et, en pratique, à une optimisation numérique par ordinateur. Il est possible de procéder tout autrement avec une méthode qui permet de nombreuses extensions : le bootstrap. Le bootstrap (Efron et Tibshirani, 1993 ; Jonhson, 2001) est une méthode de *ré-échantillonnage informatique* qui permet de réaliser de l'inférence statistique non plus en se basant sur les lois de probabilités habituelles (avec leur cortège d'hypothèses restrictives) et des formules mathématiques à chaque fois différentes mais en passant par un algorithme informatique toujours semblable.

La grande question de la statistique inférentielle est la suivante : « Que devient la statistique (V) si on répète l'échantillonnage (sur 293 autres seniors de la même population) ? » Le bootstrap y répond de façon étonnamment simple en disant que l'échantillon est ce qui peut le mieux remplacer la population de seniors, il suffit par conséquent de tirer avec remplacement le même nombre d'individus dans l'échantillon - et non pas dans la population - et de calculer la statistique V sur ce nouvel échantillon dit échantillon bootstrap. Rien n'empêche alors de recommencer ce tirage et donc d'obtenir une idée concrète, lorsqu'on répète le ré-échantillonnage, de la distribution dite distribution bootstrap de la statistique V. On peut en particulier construire un histogramme de cette distribution bootstrap (Graphique 1), calculer son écart-type qui est donc l'erreur standard de la statistique V et déterminer des intervalles de confiance du paramètre d'intérêt.

En l'espèce, nous avons donc un échantillon de départ de 293 unités statistiques sur lesquelles nous disposons de deux mesures (âge et fréquence de pratique). Nous tirons au hasard et avec remplacement de cet échantillon 293 *nouvelles* unités ce qui permet de construire une nouvelle table de contingence sur laquelle la statistique V est calculée. Cette procédure a été répétée B=2000 fois. Nous obtenons un écart-type de SE=0.04 et un intervalle de confiance à 90% par la méthode BCa de $0.05 \leq v \leq 0.18$. La cohérence avec le résultat précédent est bonne (6).

Graphique 1

Le bootstrap est une procédure statistique un peu déroutante mais qui présente des avantages indéniables : 1. elle est plus simple conceptuellement que l'arsenal probabiliste habituel, 2. elle ne demande pas d'employer des lois de probabilités dépendant d'hypothèses fortes toujours complexes à vérifier et 3. elle peut être utilisée dans des situations où nous ignorons quelle peut être la distribution de la statistique qui nous intéresse ce qui va être le cas dans la seconde partie.

PARTIE 2 LIAISON LOCALE :

APPROCHES DESCRIPTIVES ET INFÉRENTIELLES « CLASSIQUES » DE LA LIAISON LOCALE

Lorsque la liaison paraît globalement intéressante, il convient ensuite de décrire précisément sa forme. C'est l'objectif de l'analyse locale d'un tableau de contingence que de voir quelle colonne est plus particulièrement reliée à quelle ligne. Pour ce faire, l'approche la plus classique est descriptive et consiste à calculer des fréquences conditionnelles (en ligne le plus souvent). Ainsi pour la case (3,1) de référence, parmi les 114 personnes de la tranche de 60 à 65 ans, 74 sont des non pratiquants ce qui correspond à une fréquence conditionnelle de $74/114=65\%$ que l'on peut comparer aux $56/97=58\%$ de la tranche des 50-54 ans et $35/82=43\%$ de la tranche des 55-59 ans. On peut résumer graphiquement les fréquences conditionnelles soit par une collection de graphiques en barres (Graphique 2 à gauche) soit par un graphique en mosaïque (Graphique 2 à droite).

Graphique 2

Une autre approche consiste à s'intéresser aux écarts à l'indépendance, c'est ce que fait le graphique des écarts à l'indépendance de Cibois (2004). On peut y lire (Graphique 3) plus clairement la sur-représentation des pratiquants de 55-59 ans et des non pratiquants de 60-65 ans.

Graphique 3

De façon plus inférentielle, on peut véritablement « standardiser » les écarts en calculant à partir des résidus de Pearson, les *résidus standardisés de Pearson* (Agresti, 2007) grâce à la formule :

$$z_{ij} = n_{++} * r_{ij} / \sqrt{((n_{++} - n_{i+})(n_{++} - n_{+j}))}$$

ce qui autorise une comparaison avec la loi normale. Pour la case de référence, nous obtenons : $z_{31} = 293 * 1.22 / \sqrt{((293 - 114)(293 - 165))} = 2.4$ qui est une valeur exceptionnelle (seuil de 1,96 pour 95% de niveau de confiance) pour la loi normale (7). L'ensemble de ces calculs figure dans le coin inférieur droit des cases du Tableau 2.

Enfin, lorsque l'on a *a priori* une idée précise des lignes et colonnes qu'il nous plait de comparer, il existe des méthodes de partition du chi-carré (Agresti, 1990), qui permettent de décomposer cette statistique à la manière des *contrastes* en analyse de variance et d'en tester chaque composante. C'est toutefois rarement le cas et la démarche de recherche de liaison locale est plus dans l'esprit des *tests post-hoc* de l'analyse de variance, où l'ensemble des comparaisons est fait afin de n'en retenir que les plus importantes.

LE POURCENTAGE D'ECART MAXIMUM A L'INDEPENDANCE LOCAL (PEM LOCAL) : UNE TAILLE D'EFFET DE LA LIAISON LOCALE

L'écart à l'indépendance est une mesure absolue de taille d'effet : nulle si l'indépendance est exactement respectée et augmentant avec la divergence à cette hypothèse. Pour la « standardiser », Cibois (1993) a proposé une autre voie que les résidus standardisés de

Pearson en s'inspirant du mode de construction de la statistique V de Cramér. Cette méthode consiste à se demander quel peut être l'écart maximum à l'indépendance dans cette case et à utiliser cette quantité comme diviseur afin d'obtenir une nouvelle taille d'effet, appelée *Pourcentage d'Ecart Maximum à l'indépendance*.

En reprenant la case (3,1) de référence, l'effectif observé était de $n_{31}=74$, le total de la ligne correspondante $n_{3+}=114$, celui de la colonne $n_{+1}=165$ et l'effectif général de $n_{++}=293$. L'effectif théorique était de $t_{31}=64.2$ et l'écart à l'indépendance de $e_{31}=9.8$. Cet écart est positif (8). Quel aurait pu être l'écart maximum dans cette case ? Considérant que le total de la ligne 3 est de 114 et 165 celui de la colonne 1, l'effectif de la case aurait pu au plus être de 114 ce qui conduirait à un écart maximum de $114-64.2=49.8$. On obtient donc un rapport de PEM local $e_{31}/49.8=19.7\%$. Les calculs correspondant aux autres cases sont résumés dans le Tableau 3.

Cibois (2004) propose sur la base de son expérience de sociologue de considérer qu'un PEM local $<5\%$ constitue un effet négligeable, que 10% constitue un seuil au dessus duquel l'effet est généralement intéressant et qu'au-delà de 50% il est exceptionnel. La lecture du Tableau 3 montre donc qu'il y a attraction (PEM local positif et $>10\%$) entre la tranche de 55-59 ans et la pratique la plus régulière (une fois par semaine et plus) et répulsion (PEM local négatif et $>10\%$) avec l'absence de pratique. Le résultat est inversé pour la tranche des 60-65 ans.

Tableau 3

UN INTERVALLE DE CONFIANCE DU PEM LOCAL PAR BOOTSTRAP

Le PEM local semble donc une mesure de taille d'effet standardisée ($-100\% \leq \text{PEM local} \leq +100\%$) intéressante mais il convient, comme pour le V de Cramér, de mesurer l'incertitude qui l'entoure par un intervalle de confiance. A notre connaissance, il n'existe pas de résultat probabiliste concernant cette statistique mais on peut utiliser la technique du bootstrap qui permet de suppléer informatiquement à cette absence théorique.

En ce qui concerne la case de référence, nous avons observé PEM local $e_{31}=19.7\%$ et un intervalle de confiance (technique BCa) obtenu après 2000 ré-échantillonnages du tableau de contingence donne un intervalle de confiance à 95% de $2\% \leq \text{PEM local } e_{31} \leq 35\%$. On voit que la largeur de cet intervalle est considérable, qu'il exclut la valeur de l'indépendance (PEM local = 0%) mais pas la possibilité que l'effet soit négligeable (PEM local $<10\%$). En définitive si cette attraction a peu de chances de se produire « par hasard », elle ne s'avère pas forcément importante.

L'ensemble des résultats se lit plus facilement de façon graphique (Graphique 4), particulièrement si les cases sont ordonnées en fonction de la valeur des PEM locaux. On valide sur ce graphique que les attractions et répulsions observées sont statistiquement significatives en regardant simplement si leur intervalle de confiance coupe la ligne horizontale de hauteur zéro. Les deux répulsions significatives (à gauche du graphique : 55-59 ans/non pratiquants et 60-65 ans/une fois et plus) et les deux attractions significatives (à droite du graphique : 60-65 ans/non pratiquant et 55-59 ans/une fois et plus) précédemment décrites sont immédiatement identifiées.

Graphique 4

DISCUSSION

Comme le dit Cibois (2004), les techniques de liaison globale et locale prennent tout leur sens lorsque l'on considère non pas un tableau de contingence isolé mais une série de tableaux de contingence partageant les mêmes lignes pour établir un profil de ces différentes lignes. En l'espèce, pour chacune des tranches d'âge, on aurait pu non seulement s'intéresser à leur fréquence de pratique mais aussi à leur ancienneté de pratique, au type d'APS en jeu, au lieu de cette pratique, à sa durée... Dans un premier temps, au niveau global, un « balayage » sur la base du V de Cramér permet de sélectionner les variables les plus liées à celle définissant les profils. Un graphique similaire à celui du Graphique 4 mais comportant pour les différentes variables les intervalles de confiance du V de Cramér permet alors une détection visuelle des liaisons majeures. Dans un second temps, au niveau local, le PEM local constitue une quantité comparable d'un tableau de contingence à l'autre. En se concentrant sur les valeurs les plus importantes, il est possible de définir ce qui caractérise un profil par les attractions et les répulsions les plus fortes ce que Cibois appelle « le profil de modalité ». Nos propositions prolongent bien cette démarche, rien n'empêchant un graphique tel que celui du Graphique 4 de représenter et d'ordonner le croisement des tranches d'âge avec plusieurs variables qui définissent la pratique. On propose là une méthode de dépouillement d'enquête graphique en deux temps efficace qui s'insère dans la stratégie de Cibois (2007) en produisant des intervalles de confiance.

Il reste bien entendu à disposer des moyens informatiques adéquats. L'ensemble des procédures a été programmé à l'aide du logiciel statistique libre de distribution R utilisable sur toute plate-forme : Windows, Unix et Mac (R Development Core Team, 2007) et un document, disponible par simple demande auprès des auteurs, contient le mode d'emploi de ces fonctions.

Sur un plan inférentiel, le bootstrap réalise-t-il des miracles ? Nous avons vu son intérêt sur cet exemple afin de calculer l'incertitude entourant notre connaissance du PEM local alors que peu de résultats théoriques étaient disponibles. Plusieurs travaux ont montré comment de nouvelles statistiques (plus robustes en particulier) sur lesquelles l'inférence n'était pas jusqu'alors possible peuvent à présent être employées pour mieux décrire les données recueillies. Le bootstrap a également prouvé son efficacité dans des situations théoriques bien connues en levant des hypothèses (de normalité en particulier) trop contraignantes. Notons pour le V de Cramér que si la borne supérieure de l'intervalle a donné le résultat attendu, la borne inférieure était en revanche légèrement surestimée (9). On peut donc s'attendre à de bonnes performances de cette procédure, dont le principe est rappelons-le de remplacer la population par l'échantillon observé, à deux conditions toutefois : 1. que l'échantillon soit bien représentatif, et théoriquement seules les méthodes d'échantillonnage aléatoire permettent d'avoir quelque assurance à ce sujet, et 2. que l'échantillon soit de taille raisonnable et, pour des variables qualitatives, notre échantillon est sans doute un peu juste à ce sujet. Ce dernier point transparaît dans la largeur des intervalles de confiance des PEM locaux. Remarquons de surcroît que ces intervalles auraient été encore plus larges si une correction avait été appliquée pour tenir compte de leur multiplicité. Alors que la procédure surprenante de ré-échantillonnage pourrait le laisser penser, il faut bien comprendre que le bootstrap ne crée pas de l'information, il utilise en revanche au mieux celle qui est disponible pour, en particulier, décrire l'incertitude qui l'entoure.

Sur un plan plus général, il est proposé de remplacer les classiques tests de significativité statistique par des mesures de taille d'effet et surtout des intervalles de confiance de ces tailles d'effet. Toutefois, l'expérience montre (Fidler et al., 2004) que même une politique volontariste des éditeurs de revue scientifique ne garantit pas que ce changement rencontre un grand succès car il peut être superficiellement suivi par les auteurs (qui n'ont pas le choix) mais ne pas être intégré véritablement au cœur de la discussion scientifique. Il reste donc sans

doute à mener une réflexion plus systématique sur un nouveau mode de présentation des résultats de recherche qui s'appuierait sur les intervalles de confiance afin de montrer comment les calculer, les présenter et les discuter. A l'instar de Thompson (2002), nous inclinons à penser que des représentations graphiques comme celle du Graphique 4 sont sans doute un bon point de départ dans ce changement.

NOTES

1. De *véritables* résidus standardisés seront présentés plus avant.
2. Cet exemple est intéressant car le résultat est tout juste significatif, une légère variation des effectifs du tableau de contingence aurait conduit à un résultat diamétralement opposé si on interprète trop strictement la notion de seuil de décision.
3. Il faut bien prendre les valeurs de Cohen pour ce qu'elles sont, de simples propositions, et comme il le dit lui-même essayer de définir à moyen terme pour chaque champ de recherche ce qu'on peut considérer comme des effets faible, moyen ou fort. D'après notre expérience et comme l'écrit Cibois, ces valeurs sont « un peu sévères » en ce qui concerne la sociologie. Cibois parle de « technique imparfaite car elle exagère le maximum du khi-deux » proposant une alternative avec le PEM global. Cette technique ne correspond toutefois pas à l'approche de cet article car la comparaison du PEM global se fait à marges fixées dans le tableau de contingence.
4. Notons que Smithson définit de façon un peu différente la taille d'effet théorique : $\sqrt{((df+\Delta)/m)}$.
5. L'ensemble des procédures informatiques est disponible auprès des deux auteurs.
6. Un peu moins pour la borne inférieure de l'intervalle. Nous y reviendrons dans la discussion.
7. On note ainsi la correction importante qui a été effectuée par rapport au résidu de Pearson non standardisé dont la valeur ($r_{31}=1.2$) n'avait rien d'extraordinaire.
8. On a donc $PEM=(n_{ij}-t_{ij})/(\min(n_{i+}; n_{+j})-t_{ij})$ pour les écarts positifs et $PEM=-(n_{ij}-t_{ij})/(\max(0; n_{i+}+n_{+j}-n_{++})-t_{ij})$ pour les écarts négatifs (se reporter à l'annexe de Cibois, 1993).
9. La statistique V est un estimateur biaisé de la taille d'effet théorique v . On peut calculer une approximation linéaire de ce biais qui est $(\max(L,C)-1)/(2n_{++}v)$. Autrement dit, plus le tableau est grand, plus petit est l'échantillon et plus petit est l'effet, plus le biais grandit. Or, l'échantillon recueilli n'est pas si grand pour des variables qualitatives et nous avons estimé que la taille d'effet était plutôt petite, nous sommes donc précisément avec cet exemple dans une situation délicate. Si la méthode BCa corrige correctement le biais pour la partie supérieure, on voit d'ailleurs sur le graphique 1 la différence considérable avec le percentile à 95% de la distribution bootstrap, elle ne parvient pas complètement à corriger le percentile à 5% pour la borne inférieure. Concrètement, ce qui nous intéresse n'est pas de tester si l'effet est exactement nul (il ne l'est sans doute pas), mais d'appréhender son importance, ce que l'intervalle de confiance, même légèrement biaisé, autorise.

REFERENCES

- Agresti, A. (1990). *Categorical data analysis*, John Wiley & Sons, New-York.
- Agresti, A., & Finlay B. (2007). *Statistical methods for the social sciences*, Pearson/Prentice Hall, Upper Saddle River.
- Cibois, P. (1990). *L'analyse de données en sociologie*, PUF, Paris.
- Cibois, P. (1993). Le PEM, pourcentage de l'écart maximum: un indice de liaison entre modalités d'un tableau de contingence, *Bulletin de Méthodologie Sociologique*, 40, 43-63.
- Cibois, P. (2004). Les écarts à l'indépendance. Techniques simples pour analyser des données d'enquête, Collection en ligne « Méthodes quantitatives pour les sciences sociales » de *Sciences Humaines*.
- Cibois, P. (2007). *Les méthodes d'analyse d'enquêtes*, PUF, Paris.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaum Associates publishers, Hillsdale.
- Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions, *Educational and Psychological Measurement*, 61, 532-575.
- Efron, B., & Tibshirani, R.J. (1993). *An introduction to the bootstrap*, Chapman & Hall, London.
- Ely, M. (1999). The importance of estimates and confidence intervals rather than P values, *Sociology*, 33, 185-190.
- Fidler, F., Thomason, N., Cumming, G. Finch, S., & Leeman, J. (2004). Editors can lead researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine, *Psychological Science*, 15 (2), 119-126.
- Johnson, R.W. (2001). An introduction to the bootstrap, *Teaching Statistics*, 23 (1), 49-54.
- Kline, R.B. (2004). *Beyond significance testing*, American Psychological Association, Washington D.C.
- Martin, O. (2005). *L'analyse des données quantitatives*, Armand Colin, Paris.
- R Development Core Team (2007). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Smithson, M. (2003). *Confidence intervals*, Sage Publications, Thousand Oaks, California.
- Thompson, B. (2002). What future quantitative social science research could look like: confidence intervals for effect sizes, *Educational Researcher*, 31(3), 25-32.
- Wilkinson, L., & the Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.

TABLEAUX ET GRAPHIQUES

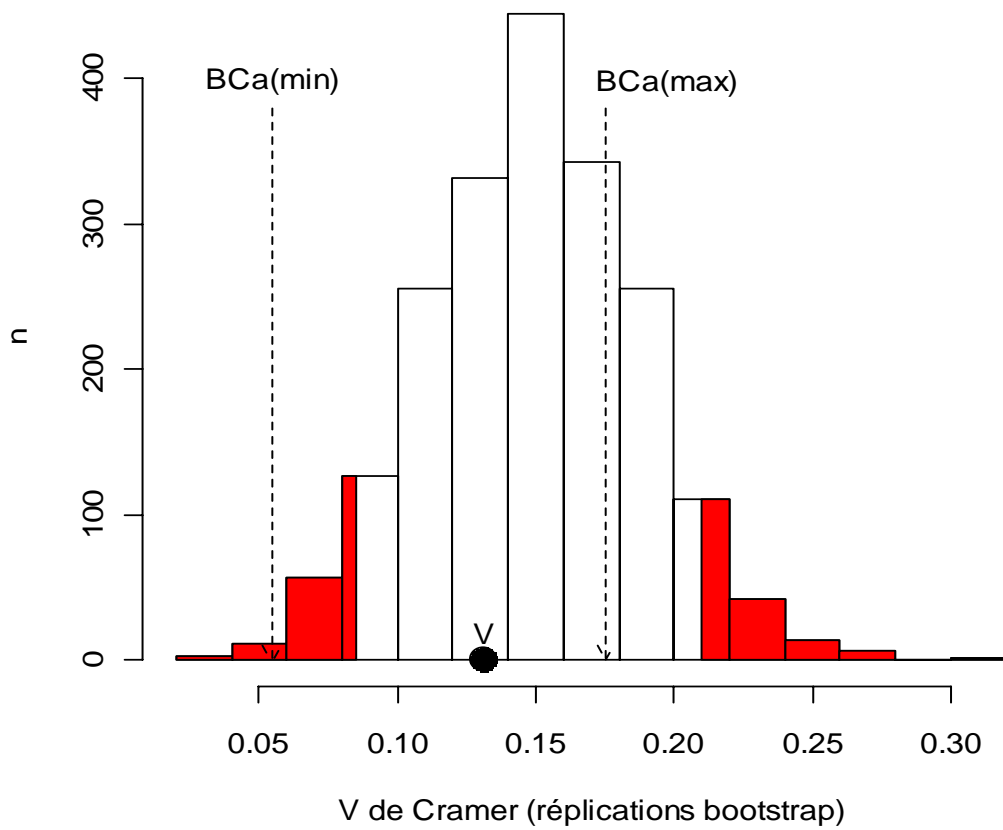
Tableau 1 : Tableau de contingence issu du croisement entre l'âge et la fréquence de pratique d'APS chez les seniors franciliens (source : Observatoire du sport FPS/IPSOS, 2004).

Effectifs	Non pratiquant	Moins d'une fois par semaine	Une fois par semaine et plus
50-54 ans	56	7	34
55-59 ans	35	7	40
60-65 ans	74	8	32

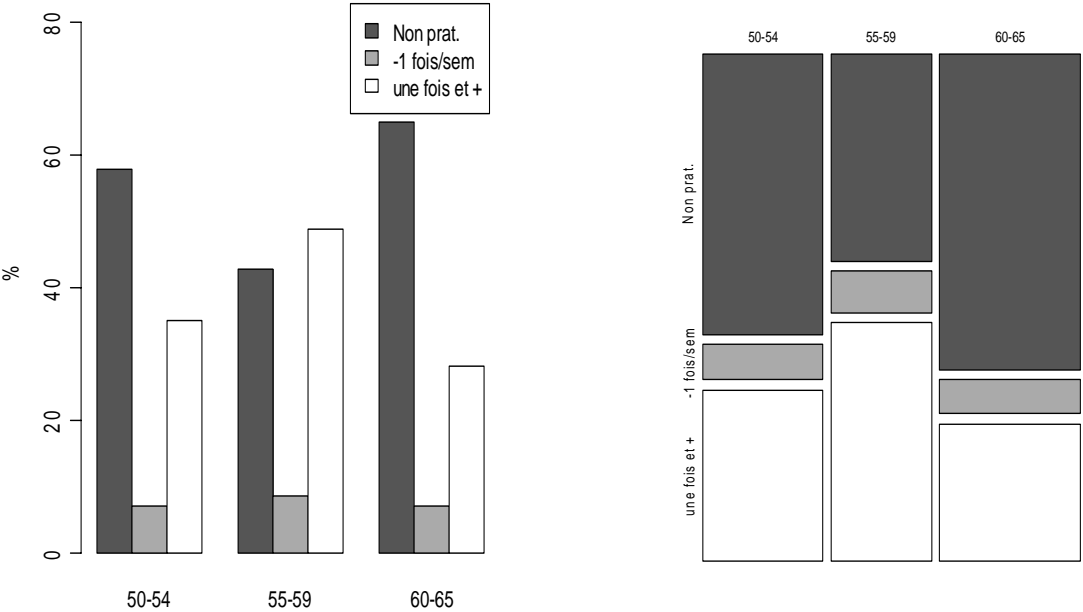
Tableau 2 : Effectifs théoriques t_{ij} (haut-gauche), écarts à l'indépendance e_{ij} (haut-droit), résidus de Pearson r_{ij} (bas-gauche) et résidus standardisés de Pearson z_{ij} (bas-droit) correspondant au tableau 1 dont les valeurs sont rappelées au centre de la case correspondante.

t_{ij} n_{ij}	e_{ij} z_{ij}	Non pratiquant	Moins d'une fois par semaine	Une fois par semaine et plus
50-54 ans	56	54,6 1,4 0,2 0,3	7 7,3 -0,3 -0,1 -0,1	34 35,1 -1,1 -0,2 -0,3
55-59 ans	35	46,2 -11,2 -1,6 -2,9	7 6,2 0,8 0,3 0,4	40 29,7 10,3 1,9 2,8
60-65 ans	74	64,2 9,8 1,2 2,4	8 8,6 -0,6 -0,2 -0,3	32 41,2 -9,2 -1,4 -2,3

Graphique 1 : Histogramme de 2000 réplifications bootstrap de la statistique V de Cramér sur l'exemple des seniors. Les bornes inférieures et supérieures de l'intervalle de confiance (méthode BCa) sont également indiquées sur le graphique. La valeur observée de la statistique V est signalée par un point noir. Les zones foncées comprennent chacune 5% de la distribution bootstrap et en indiquent donc l'emplacement des percentiles à 5% et 95%.



Graphique 2 : A gauche, collection de graphiques en barres décrivant les fréquences conditionnelles de pratique APS selon la tranche d'âge des seniors. A droite, la version par graphique en mosaïque.



Graphique 3 : Graphique des écarts à l'indépendance décrivant les liaisons locales entre l'âge et la pratique d'APS chez les seniors

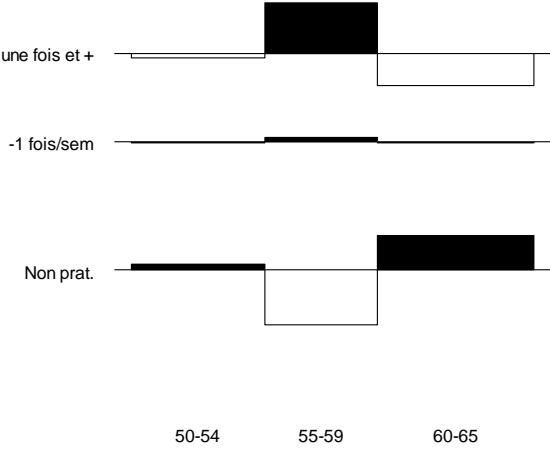


Tableau 3 : PEM locaux correspondant au tableau 1 de contingence

Effectifs	Non pratiquant	Moins d'une fois par semaine	Une fois par semaine et plus
50-54 ans	3.2	-3.9	-3.1
55-59 ans	-24.2	5.3	19.7
60-65 ans	19.7	-6.5	-22.4

Graphique 4 : PEM locaux correspondant à la liaison entre l'âge et la fréquence de pratique chez les seniors. Les PEM locaux sont rangés par ordre croissant et leurs intervalles de confiance sont indiqués.

