REVUE DE STATISTIQUE APPLIQUÉE

GÉRALD MARION

Colinéarité et interaction dans l'analyse de modèles linéaires : un modèle et un exemple

Revue de statistique appliquée, tome 20, nº 3 (1972), p. 57-66
<http://www.numdam.org/item?id=RSA_1972__20_3_57_0>

COLINÉARITE ET INTERACTION DANS L'ANALYSE DE MODÈLES LINÉAIRES : UN MODÈLE ET UN EXEMPLE

Gérald MARION

Université de Montréal

Dans un article publié en 1948, W.L. Stevens présentait une technique d'itération pour estimer les coefficients d'un modèle linéaire où la distribution des observations selon divers critères de classification n'est pas orthogonale. Malgré toute l'ingéniosité de l'auteur, cette méthode n'est développée et présentée qu'à l'aide d'un exemple arithmétique et ne débouche pas sur une technique générale (1).

Le présent article a pour but de définir les équations normales qui permettent de présenter la technique de Stevens sous une forme générale et d'en faire une application à l'aide d'un exemple tiré des statistiques de revenus au Canada.

1 - RAPPEL DU PROBLEME

Il est possible de mesurer l'influence d'un facteur si on peut faire varier le niveau de ce facteur sans que les autres facteurs varient concurremment. En d'autres mots, les facteurs ne doivent pas être corrélés mais indépendants. Cette condition présente dans les expériences contrôlées est plus rare dans l'analyse économique où nombre de variables présentent des collinéarités accusées (2). En termes techniques, on dit alors que la distribution des échantillons selon les facteurs est non-orthogonale. Pour fixer les idées sur cette notion, prenons un exemple. Supposons que le taux moyen de salaire dépende de l'occupation et du niveau d'instruction. Classifions les unités selon ces deux facteurs dans le tableau ci-dessous :

Scolarité / Occupation	I	II	III
1			
2		n_{ij}	
3			

----

(1) W.L. Stevens, "Statistical Analysis of a Non-Orthogonal Trifactorial Experiment", Biometrika, Vol.XXXV (1948), pp. 346-367.

(2) E. Malinvaud, Méthodes statistiques de l'Econométrie. Dunod, p. 225 et suivantes.

Appelons A, la classification selon les occupations et B la classification selon la scolarité. Chacun des carreaux de la grille contient un certain nombre d'unités égal à n_{ij}, représentant le nombre d'unités appartenant à la fois aux classes α_i et β_j.

Le point essentiel est le suivant : si tous les éléments n_{ij} sont égaux, la distribution selon les industries et la taille est dite orthogonale. Dans ce cas, le nombre d'unités dans les classes d'un facteur n'est pas relié au nombre d'unités qui apparaît dans les catégories d'un autre critère ; les facteurs sont alors indépendants les uns des autres. Il suit que le niveau d'un facteur peut changer sans que celui de l'autre facteur varie concurremment. C'est précisément cette possibilité de faire varier un facteur indépendamment des variations des autres facteurs qui permet d'observer les effets nets de ce facteur ; car dans ce contexte, c'est-à-dire lorsque les distributions sont orthogonales, l'effet obtenu par la variation d'un facteur n'est pas entaché d'une erreur provenant de variations concomitantes du second facteur. Il est alors possible, par exemple, de mesurer l'effet d'appartenance à une occupation simplement en prenant la différence entre le taux moyen de l'ensemble de l'échantillon et le taux moyen de l'occupation dont on recherche l'effet. Le niveau d'éducation n'étant pas corrélé avec la structure occupationnelle, il n'affecte pas l'effet d'occupation. Dans le cas contraire, c'est-à-dire lorsque les distributions ne sont pas orthogonales, les écarts entre la classe α_i d'un critère A et la moyenne générale résultent partiellement ou intégralement de différences dans la composition des classes par rapport à l'autre critère.

Précisons de plus que la condition de l'égalité des n_{ij} est une condition suffisante, mais non nécessaire pour que l'orthogonalité existe. Il suffit que la répartition des unités dans les classes $(\alpha_i \beta_j)$ obéisse à certaines règles de proportionnalité que nous présentons plus bas.

Une distribution des unités dans les diverses "cases" d'une classification multiple qui obéit à certaines règles d'une distribution orthogonale est donc une propriété désirable dans toute analyse factorielle. Cependant, il arrive souvent que les auteurs procèdent à des analyses de variances, par exemple, sans se soucier de la présence de ces propriétés concernant la distribution des unités. C'est d'ailleurs la même difficulté qui nous confronte lorsque les variables indépendantes d'une analyse de régression sont collinéaires.

Dans les lignes qui suivent, nous dérivons les équations normales qui nous permettent d'estimer les effets des facteurs de même que nous identifions les éléments de ces équations dont la valeur est déterminée par l'asymétrie de la distribution des unités par rapport aux critères de classement des unités. Nous présentons enfin une méthode d'itération qui pour l'estimation des facteurs ou de la valeur des paramètres des équations normales, tient compte de l'asymétrie de la distribution des unités.

II - EVALUATION DU BIAIS DU A LA NON-ORTHOGONALITE

Prenons le modèle suivant :

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijkl}$$

Dans ce modèle, y désigne le revenu moyen du groupe de travailleurs k étant à la fois dans la catégorie α_i du critère A (industrie), la classe β_j du critère B (taille) et la classe γ_k du critère C. m est une constante.

Pour l'estimation des effets α, β, γ, on cherche les paramètres a_i, b_j, c_k et m, tels que :

$$F = \sum_i \sum_j \sum_k \sum_l (y_{ijkl} - m - a_i - b_j - c_k)^2$$

soit minimum où :

$$i = 1, 2 \ldots J$$
$$j = 1, 2 \ldots J$$
$$k = 1, 2 \ldots K$$
$$l = 1, 2 \ldots n_{ijk}$$

De plus, posons :

$$n_{i..} = \sum_{jk} n_{ijk}$$
$$n_{.j.} = \sum_{ik} n_{ijk}$$
$$n_{..k} = \sum_{ij} n_{ijk}$$
$$\sum_i \sum_j \sum_k \sum_l y_{ijkl} = y_{....}$$

En prenant les dérivées partielles de F par rapport à m, a, b et c, on obtient les équations normales suivantes :

$$y_{....} = n_{...}\ m + \sum_{i=1}^{I} n_{i..}\ a_i + \sum_{j=1}^{J} n_{.j.}\ b_j + \sum_{k=1}^{K} n_{..k}\ c_k$$

$$y_{i...} = n_{i..}\ m + n_i\ a_i + \sum_{j=1}^{J} n_{ij.}\ b_j + \sum_{k=1}^{K} n_{i.k}\ c_k$$

$$y_{.j..} = n_{.j.}\ m + \sum_{i=1}^{I} n_{ij.}\ a_i + n_{.b.}\ b_j + \sum_{k=1}^{K} n_{.jk}\ c_k$$

$$y_{..k.} = n_{..k}\ m + \sum_{i=1}^{I} n_{i.k}\ a_i + \sum_{j=1}^{J} n_{.jk}\ b_j + n_{..k}\ c_k$$

D'où l'on tire :

$$\bar{y}_{....} = \frac{y_{....}}{n_{...}} = m + \sum_{i=1}^{I} \frac{n_{i..}}{n_{...}} a_i + \sum_{j=1}^{J} \frac{n_{.j.}}{n_{...}} b_j + \sum_{k=1}^{K} \frac{n_{..k}}{n_{...}} c_k$$

$$\bar{y}_{i...} = \frac{y_{i..}}{n_{i..}} = m + a_i \sum_{j=1}^{J} \frac{n_{ij.}}{n_{i..}} b_j + \sum_{k=1}^{K} \frac{n_{i.k}}{n_{i..}} c_k$$

Et :

$$\bar{y}_{i...} - y_{....} = a_i - \sum_{i=1}^{J} \frac{n_{i..}}{n_{...}} a_i + \sum_{j=1}^{J} \left(\frac{n_{ij.}}{n_{i..}} - \frac{n_{.j.}}{n_{...}}\right) b_j + \sum_{k=1}^{K} \left(\frac{n_{i.k}}{n_{i..}} - \frac{n_{..k}}{n_{...}}\right) c_k$$

Dans l'équation précédente, les expressions

$$\sum_j \left(\frac{n_{ij.}}{n_{i..}} - \frac{n_{.j.}}{n_{...}}\right) b_j \text{ et } \sum_k \left(\frac{n_{i.k}}{n_{i..}} - \frac{n_{..k}}{n_{...}}\right) c_k$$

représentent le biais introduit en utilisant $y_{i...} - y_{....}$ comme estimé de l'effet α_i.

Il est évident que ce biais est nul si

$$\frac{n_{ij.}}{n_{i..}} = \frac{n_{.j.}}{n_{...}} \quad \text{et} \quad \frac{n_{i.k}}{n_{i..}} = \frac{n_{..k}}{n_{...}}.$$

Une transformation triviale de la première expression donne

$$n_{ij.} = \frac{n_{i..}\, n_{.j.}}{n_{...}} .$$

On peut montrer que cette expression définit les règles de proportionnalité auxquelles doit obéir une distribution orthogonale.

Ajoutons enfin la condition que m soit la moyenne de revenu pour toute la population, alors

$$\sum_{i=1}^{I} n_{i..}\, a_i = 0 \; ; \quad \sum_{j=1}^{J} n_{.j.}\, b_j = 0 \quad \text{et} \quad \sum_{k=1}^{K} n_{..k}\, c_k = 0$$

et

$$\bar{y}_{i...} - \bar{y}_{....} = a_i + \sum_{j=1}^{J} \frac{n_{ij.}}{n_{i..}} b_j + \sum_{k=1}^{K} \frac{n_{i.k}}{n_{i..}} c_k$$

Rappelons que les deux dernières expressions sont le biais provenant de la non-orthogonalité des distributions et qu'elles s'annulent dès que les distributions deviennent orthogonales. Dès lors, dans le cas de distribution orthogonale, $\bar{y}_{i...} - \bar{y}_{....} = a_i$ (1).

----

(1) En effet, dans le cas d'orthogonalité des distributions

$$n_{ij.} = \frac{n_{i..}\, n_{.j.}}{n_{..}} \text{ et } n_{i.k} = \frac{n_{..k}\, n_{i..}}{n_{...}}$$

Si on substitue $n_{ij.}$ et $n_{i.k}$ par leur valeur respective dans l'équation normale $y_{i...} = (\ldots)$, on aura :

$$y_{i...} = n_{i..}\, m + n_{i..}\, a_i + \sum_{j=1}^{J} \frac{n_{i..}\, n_{.j.}}{n_{..}} b_j + \sum_{k=1}^{K} \frac{n_{..k}\, n_{i..}}{n_{...}} c_k$$

Puisque $\sum_{j=1}^{b} n_{.j.}\, b_j = o$ et $\sum_{k=1}^{K} n_{..k}\, c_k = o$,
l'équation normale se réduit à :

$$y_{i..} = n_{i..}\, m + n_{i..}\, a_i$$

Etant donné les contraintes ci-dessus, $m = \bar{y}_{....}$, ainsi :

$$n_{i..}\, a_i = y_{i...} - n_{i..}\, \bar{y}_{....}$$

$$a_i = y_{i...} / n_{i..} - \bar{y}_{....}$$

$$a_i = \bar{y}_{i...} - \bar{y}_{....}$$

etc...

On démontre de la même façon que :

$$\bar{y}_{.j..} - \bar{y}_{....} = b_j + \sum_{=1}^{J} \frac{n_{ij.}}{n_{.j.}} a_i + \sum_{k=1}^{K} \frac{n_{.jk}}{n_{.j.}} c_k$$

Et que :

$$\bar{y}_{..k.} - \bar{y}_{....} = c_k + \sum_{i=1}^{I} \frac{n_{i.k}}{n_{..k}} a_i + \sum_{j=1}^{J} \frac{n_{.jk}}{n_{..k}} b_j$$

III - INTERACTION

Le biais provenant de la collinéarité entre les facteurs dépend de la distribution des unités dans les diverses catégories des facteurs. Mais il arrive aussi que les effets des différents facteurs soient liés entre eux indépendamment de la distribution des unités. Ce phénomène se produit lorsque les effets d'un facteur dépendent du niveau d'un autre facteur. Ainsi, l'écart de revenu entre un travailleur qualifié et un non-qualifié peut ne pas être le même pour tous les groupes ethniques, mais dépendre de l'origine ethnique des salariés, étant donné que la rareté des diverses catégories de travail peut être différente d'un groupe ethnique à un autre. Pour calculer l'interaction entre deux facteurs, on doit connaître pour chaque catégorie d'un facteur les effets de ce facteur pour chacune des catégories d'un deuxième facteur. En ce sens, ces deux facteurs n'en forment qu'un seul, puisqu'ils ont un effet joint.

IV - ITERATION

1 - Sans interaction entre les facteurs

La méthode d'itération que nous utilisons consiste à éliminer graduellement les biais qui entachent la valeur des paramètres ; ce biais provient du fait que : 1) la première estimation d'un premier paramètre est elle-même biaisée, et que 2) les paramètres des autres facteurs qui, lorsque les distributions sont asymétriques, affectent la valeur du premier paramètre, sont eux-mêmes biaisés.

D'après les équations normales :

$$a_i = \bar{Y}_{i...} - \bar{Y}_{....} - \sum_{j=1}^{J} \frac{n_{ij.}}{n_{i..}} b_j - \sum_{k=1}^{K} \frac{n_{i.k}}{n_{i..}} c_k$$

$$b_j = \bar{Y}_{.j..} - \bar{Y}_{....} - \sum_{i=1}^{I} \frac{n_{ij.}}{n_{.j.}} a_i - \sum_{k=1}^{K} \frac{n_{.jk}}{n_{.j.}} c_k$$

$$c_k = \bar{Y}_{..k.} - \bar{Y}_{....} - \sum_{i=1}^{I} \frac{n_{i.k}}{n_{..k}} a_i - \sum_{j=1}^{J} \frac{n_{.jk}}{n_{..k}} b_j$$

Si on évalue d'abord la valeur des paramètres en supposant que les distributions sont orthogonales. On obtient alors à l'étape (o) :

$$a_i^{(o)} = \bar{y}_{i...} - \bar{y}_{....}$$

$$b_j^{(o)} = \bar{y}_{.j..} - \bar{y}_{....}$$

$$c_j^{(o)} = \bar{y}_{..k.} - \bar{y}_{....}$$

Par la suite, on corrige la valeur de $a^{(o)}_i$ pour éliminer les effets de b_j et c_k sur a_i. On obtient :

$$a^{(1)}_i = a^{(o)}_i - \frac{1}{n_{i..}} \sum_j n_{ij.} \; b^{(o)}_j - \frac{1}{n_{i..}} \sum_k n_{i.k} \; c^{(o)}_k$$

On fait de même pour les autres paramètres ; mais en tenant compte de la nouvelle estimation de a_i puis de b_j :

$$b^{(1)}_j = b^{(o)}_j - \frac{1}{n_{.j.}} \sum_i n_{ij.} \; a^{(1)}_i - \frac{1}{n_{.j.}} \sum_k n_{.jk} \; c^{(o)}_k$$

$$c^{(1)}_k = c^{(o)}_k - \frac{1}{n_{..k}} \sum_i n_{i.k} \; a^{(1)}_i - \frac{1}{n_{..k}} \sum_j n_{.jk} \; c^{(j)}$$

Bien que l'estimation de $a^{(1)}_i$, $b^{(1)}_j$ et $c^{(1)}_k$ présente une amélioration sur celle de $a^{(o)}_i$. $b^{(o)}_j$ et $c^{(o)}_k$, il est évident que des biais persistent, dûs au fait que les corrections ont été pratiquées à partir de paramètres eux-mêmes encore biaisés en partie.

Nous continuons donc l'itération pour obtenir :

$$a^{(2)}_i = a^{(o)}_i - \frac{1}{n_{i..}} \sum_j n_{ij.} \; b^{(1)}_j - \frac{1}{n_{i..}} \sum_k n_{i.k} \; c^{(1)}_k$$

$$b^{(2)}_j = b^{(o)}_j - \frac{1}{n_{.j.}} \sum_i n_{.j.} \; a^{(2)}_i - \frac{1}{n_{.j.}} \sum_k n_{.jk} \; c^{(1)}_k$$

$$c^{(2)}_k = c^{(o)}_k - \frac{1}{n_{..k}} \sum_i n_{i.k} \; a^{(2)}_i - \frac{1}{n_{..k}} \sum_j n_{.jk} \; b^{(2)}_j$$

. . . .

$$a^{(s)}_i = a^{(o)}_i - \frac{1}{n_{i...}} \sum_j n_{ij.} \; b^{(s-1)} - \frac{1}{n_{i..}} \sum_k n_{i.k} \; c^{(s-1)}_k$$

$$b^{(s)}_j = b^{(o)}_j - \frac{1}{n_{.j.}} \sum_j n_{ij.} \; a^{(s)}_i - \frac{1}{n_{ij.}} \sum_k n_{.jk} \; c^{(s-1)}_k$$

$$c^{(s)}_k = c^{(o)}_k - \frac{1}{n_{..k}} \sum_i n_{ik} \; a^{(s)} - \frac{1}{n_{..k}} \sum_j n_{.jk} \; b^{(s)}_j$$

1 - Avec interaction entre les facteurs

Nous avons vu précédemment que pour calculer l'interaction entre deux facteurs, on doit connaître pour chaque catégorie d'un premier facteur les effets de ce facteur pour chacune des catégories du deuxième facteur. Si par exemple, on tient pour acquis qu'il existe une interaction entre les facteurs A et B, on estimera la valeur des A_i pour chaque B_j. On cherche donc la valeur des paramètres a_{ij}. Appliquant la méthode itérative, la première étape apparaîtra alors comme étant :

$$a^{(o)}_{ij} = \bar{y}_{ij.} - \bar{y}_{...}$$

$$c^{(1)}_k = \bar{y}_{..k} - \bar{y}_{...}$$

Les autres étapes seront alors :

$$a_{ij}^{(1)} = a_{ij}^{(o)} - \frac{1}{n_{ij.}} \sum_k \quad n_{ijk}\, c_k^{(o)}$$

$$c_k^{(1)} = c_k^{(o)} - \frac{1}{n_{..k}} \sum_{kj} n_{ijk}\ a_{ij}^{(1)}$$

$$a_{ij}^{(2)} = a_{ij}^{(o)} - \frac{1}{n_{ij}} \sum_k n_{ijk}\ c_k^{(1)}$$

$$c_k^{(2)} = c_k^{(o)} - \frac{1}{n_{..k}} \sum_{kj} n_{ijk}\ a_{ij}^{(2)}$$

. . . .

$$a_{ij}^{(s)} = a_{ij}^{(o)} - \frac{1}{n_{ij}} \sum_k n_{ijk}\ c_k^{(s-1)}$$

$$c_k = c_k^{(o)} - \frac{1}{n_{..k}} \sum_{kj} n_{ijk}\ a_{ij}^{(s)}$$

La classification des unités se fait en tenant compte du fait que lorsque deux facteurs sont en interaction, à chaque catégorie d'un de ces facteurs correspond une catégorie de l'autre facteur qui est en interaction.

V - UNE APPLICATION A L'ETUDE DES DISPARITES ENTRE GROUPES ETHNIQUES SELON LES CRITERES EDUCATION ET PROFESSIONS

Le modèle précédent sera utilisé pour étudier l'importance des facteurs ethnicité, éducation et professions dans l'explication des disparités de revenus. On sait que ces trois facteurs sont fortement corrélés et qu'il est donc nécessaire d'avoir recours à une méthode itérative pour mesurer leurs effets respectifs.

Mais avant de présenter les résultats de l'itération, nous rappelons les statistiques brutes sur lesquelles cette analyse est fondée. D'abord, en ce qui concerne les revenus, nous retenons les gains annuels moyens de travail des travailleurs masculins du Grand-Montréal en 1961. Les effectifs se rapportent également aux travailleurs masculins de la région métropolitaine de Montréal en 1961.

En 1961, la moyenne générale de revenus est de $ 4,447.86. Les revenus et les quantités selon une classification simple des trois critères retenus se présentent ainsi :

Origines ethniques	Revenus moyens annuels	Nombre de personnes
Française	$ 3,998	331,734
Britannique	5,895	98,927
Autres	4,501	112,851

Niveaux de scolarité	Revenus moyens annuels	Nombre de personnes
Aucune	$ 2,731	1,845
Elémentaire	3,414	221,685
Secondaire 1 - 2	3,941	117,567
Secondaire 3 - 5	5,042	135,356
Université	7,651	67,059

Professions	Revenus moyens annuels	Nombre de personnes
Administrateurs	$ 8,085	66,339
Professionnels et techniciens	6,831	53,808
Employés de bureau	3,452	60,687
Vendeurs	4,780	41,486
Travailleurs des transports & comm.	3,588	48,724
Travailleurs des services	3,129	46,005
Ouvriers de métiers	3,668	181,623
Manoeuvres	2,483	31,697
Agriculteurs et trav. agricoles	3,418	170
Autres travailleurs primaires	2,418	3,498
Non-classés	3,836	9,475

A partir des statistiques ci-dessus, nous pouvons d'abord calculer les effets bruts des diverses catégories des trois facteurs explicatifs retenus : l'origine ethnique, le niveau d'éducation et la profession. Ces premiers résultats apparaissent à la colonne 2 du tableau ci-dessous. Ils correspondent à l'étape (o) de l'itération et s'obtiennent tout simplement en prenant la différence entre la moyenne d'une catégorie d'un facteur et la moyenne générale. En résolvant successivement les équations du modèle itératif ci-dessus, nous éliminons graduellement, de ces effets bruts, les biais causés par la distribution non symétrique des unités. Ces effets corrigés, qui sont en fait les effets nets, apparaissent à la colonne (1). La comparaison entre les colonnes (1) et (2) révèle l'importance de ces biais. Les effets de l'origine ethnique tout comme les effets du niveau d'éducation sont grossièrement surestimés lorsqu'on ne prend pas soin d'éliminer le biais causé par la présence de collinéarité entre les variables. Il demeure cependant que les diverses catégories de facteurs ont, par rapport à la moyenne générale de $ 4,447, des effets importants dont on peut mesurer l'ampleur à la lecture du tableau 1. Enfin, signalons que si les effets bruts ne sont pas additifs, en ce sens que les effets bruts d'ethnicité par exemple, qui sont amplifiés par le chevauchement des effets de scolarité ou de professions, ne peuvent pas être ajoutés à ces derniers, les effets nets eux, qui ne sont sensés contenir aucun biais de cette nature, peuvent être additionnés les uns aux autres.

Tableau 1

Effets bruts et corrigés des catégories des facteurs, mesurés par rapport à la grande moyenne 4,447.86

Groupes ethniques	Effets de l'appartenance ethnique	
	Effets corrigés (1)	Effets bruts (2)
Français	- 101,32	- 449,86
Britanniques	770,07	1,447.63
Autres	- 137,06	53.39
Niveaux d'éducation	**Effets des niveaux d'éducation**	
	Effets corrigés	**Effets bruts**
Aucune	- 810.19	- 1,717
Elémentaire	- 308.93	- 1,033.44
Secondaire 1 - 2	- 157.89	- 506.81
Secondaire 3 - 5	203.16	593.94
Université	1,453.59	3,202.70
Groupes occupationnels	**Effets des groupes occupationnels**	
	Effets corrigés	**Effets bruts**
Administrateurs	3,405.00	3,637.03
Professionnels et techniciens	2,135.42	2,383.03
Employés de bureau	- 1,172.98	- 995.95
Vendeurs	181.66	332.53
Travailleurs des services	- 907.72	- 859.87
Travailleurs des transports & comm.	- 1,427.70	- 1,318.92
Ouvriers et métiers	- 861.92	- 779.83
Manoeuvres	- 2,018.87	- 1,964.46
Agriculteurs & trav. agricoles	- 1,010.48	- 1,030.30
Autres travailleurs primaires	- 2,153.01	- 2,029.52
Non-classés	- 701.49	- 611.48

VI - INTERACTION : GROUPES ETHNIQUES/EDUCATION

Nous présentons maintenant une statistique qui, tout en corrigeant pour les biais de collinéarité, tient compte de l'interaction entre deux facteurs : l'éducation et l'appartenance ethnique. Les effets sont calculés par rapport à la grande moyenne. Au tableau 2 nous avons colligé les revenus par niveau de scolarité selon les groupes ethniques. Ces statistiques représentent les effets bruts et les effets corrigés ou nets. Les effets nets sont obtenus après dix itérations successives. La comparaison entre ces effets et les effets bruts révèle encore ici l'importance des biais causés par la présence de collinéarité entre les facteurs explicatifs. Ainsi au niveau universitaire, la différence entre les effets bruts et les effets nets est de près de $ 1,000 pour le groupe français et plus de $ 1,200 pour le groupe britannique.

Enfin, les effets d'interaction sont évidents. Ainsi, par rapport à la grande moyenne, l'effet net sur les revenus d'un diplôme universitaire est de $ 3,241, pour le groupe britannique et de $ 1,556.63, pour le groupe français. Remarquons que pour obtenir cette statistique, nous avons tenu compte de la structure occupationnelle des divers groupes. Les effets d'interaction sont donc importants. On ne peut pas calculer les effets des divers niveaux d'éducation sans tenir compte de l'origine ethnique des salariés. Il s'agit donc, en réalité, d'effet joint éducation/ethnie.

Tableau 2

Effets nets des niveaux d'éducation selon les groupes ethniques mesurés par rapport à la grande moyenne, $ 4,447.86

Niveaux d'éducation	Groupes ethniques					
	Français		Britanniques		Autres	
	bruts	corrigés	bruts	corrigés	bruts	corrigés
Aucune	- 1,929.86	- 1,367.45	- 1,334.86	- 689.94	- 1,669.86	- 1,351.37
Elémentaire	- 1,075.86	- 736.61	- 396.86	- 155.27	- 1,193.86	- 973.59
Secondaire 1 - 2	- 674.86	- 490.98	1.14	64.86	- 359.86	- 509.15
Secondaire 3 - 5	- 10.86	- 153.40	1,525.14	1,003.18	888.14	229.80
Université	2,535.14	1,556.63	4,501.14	3,241.91	2,675.14	1,499.19