

Philippe Cibois

Laboratoire Printemps (CNRS) – Université de Versailles-St-Quentin

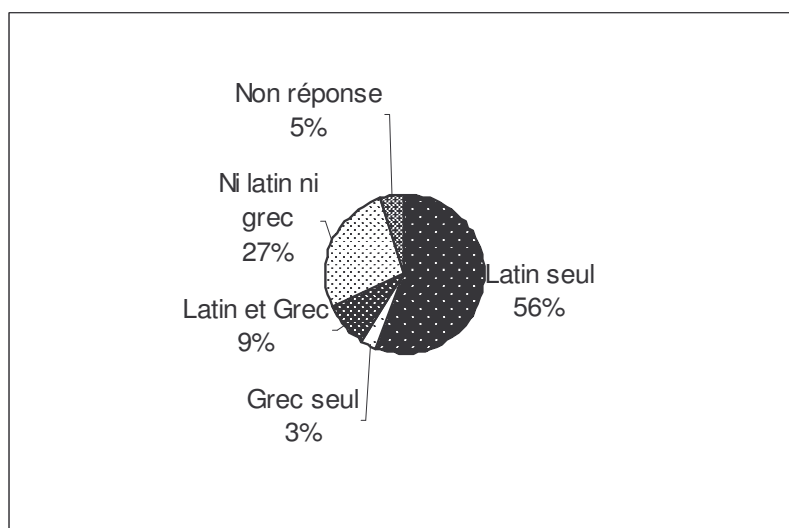
Analyse tabulaire et régression logistique (et linéaire)

Le but de cette communication est de montrer qu'il existe une manière simple et facilement compréhensible d'examiner, sur des données catégorisées, la notion d'effets toutes choses égales par ailleurs.

A cette fin, nous utilisons une enquête faite à l'université Paris V en 1994 sur une population d'un millier de répondants issus de classe moyenne supérieure en ce qui concerne leur désir de faire étudier une langue ancienne à leurs enfants.

Actuellement, environ un quart de la classe d'âge commence une initiation au latin, mais cette proportion est plus importante dans les classes moyennes supérieures et c'est pour cette raison que ce sont elles qui ont été enquêtés. On a obtenu la distribution suivante :

Projet pour l'enfant	Latin seul	Grec seul	Latin et Grec	Ni latin ni grec	Non réponse	Total
Effectif	590	34	99	284	51	1058
%	55,8	3,2	9,4	26,8	4,8	100



On cherche dans la suite à examiner "l'effet toutes choses égales par ailleurs" de l'affiliation religieuse des parents (catholique / non catholique) et de leur niveau d'étude (inférieur, égal ou supérieur au niveau du bac).

On fait un tri de profondeur 3 : religion et bac d'une part pour le choix du latin seul, du reste, de l'ensemble. On donne ensuite le pourcentage de latin seul pour chaque situation.

LatinSeul	InfBac	NivBac	SupBac	Total
Cath	44	117	152	313
Non Cath	42	99	168	309
Total	86	216	320	622

Reste	InfBac	NivBac	SupBac	Total
Cath	30	53	43	126
Non Cath	41	66	100	207
Total	71	119	143	333

Total	InfBac	NivBac	SupBac	Total
Cath	74	170	195	439
Non Cath	83	165	268	516
Total	157	335	463	955

	%LatSeul	Ref			Effet InfBac Effet SupBac	
		InfBac	NivBac	SupBac	InfB - NivB	SupB - NivB
Ref	Cath	59,5	68,8	77,9	-9,4	9,1
	Non Cath	50,6	60,0	62,7	-9,4	2,7
					-9,4	5,9
					Moyenne des effets	
Effet Cath	Cath-NonCat	8,9	8,8	15,3	11,0	Moyenne des effets

Pour voir l'effet de l'affiliation catholique, il faut l'examiner par rapport à la situation non catholique qui sert de référence. Par exemple on voit que pour la situation de diplôme inférieure au bac, le taux de latin seul passe de 50,6% à 59,5% soit un effet de 8,9%. Pour parler d'effet "toutes choses égales par ailleurs", il faut précisément examiner les autres situations que le niveau inférieur et donc aussi le niveau bac (effet de 8,8%) et le niveau supérieur (15,3%). Parler **d'un seul** effet de **l'effet** de l'affiliation catholique (et non des différents effets observés) consiste à dire que dans les trois cas on a un seul effet qui est une approximation de chacun des effets observés et que l'on trouvera simplement en prenant la moyenne des trois effets.

$$(8,9 + 8,8 + 15,3)/3 = 11,0$$

Pour ce qui est de l'effet du diplôme, comme il y a trois situations, il faut choisir l'une d'elle comme référence des effets : on prendra par exemple l'effet du niveau inférieur qui fait baisser le niveau du bac de 9,4% (pour les catholiques) et également de 9,4% chez les non catholiques, ce qui donne évidemment un effet moyen de -9,4%.

L'effet du fort diplôme par rapport au diplôme moyen est de 9,1% (77,9 – 68,8) pour les catholiques et de 2,7 pour les autres et donc un effet moyen de 5,9%

Le choix du latin est influencé positivement par l'affiliation catholique et le diplôme.

Ici cependant, la moyenne est simple : on peut tenir compte des effectifs qui provoquent les sous effets. Par exemple, le sous effet de 8,9% (niv Inf) correspond à un effectif de 157 pour un total de 955. On pondère ainsi la moyenne

$$8,9 \times 157/955 + 8,8 \times 335/955 + 15,3 \times 463/955 = 12,0$$

Dans Trideux : les résultats sont présentés de la façon suivante pour l'analyse tabulaire avec une présentation en ordre lexicographique de toutes les situations (données de base).

Pour la régression linéaire on a

Données de base

Nombre d'individus = 955

Tous les sous-effets possibles sont attestés

		Eff.Prop.
Cath	InfB	74 59.5
Cath	NivB R	170 68.8
Cath	SupB	195 77.9
NonC R	InfB	83 50.6
NonC R	NivB R	165 60.0
NonC R	SupB	268 62.7

Population concernée= 955 soit 100.0% de l'effectif total

Modalité à expliquer : LatinSeul

Situation de référence :

NonC NivB = 60.0 effectif associé = 165

Effet Cath par rapport à Ref (NonC) 3 sous-effets (s-e)

Sous-population

	s-e	Eff	Tot
InfB	8.9	157	955
NivB	8.8	335	955
SupB	15.3	463	955

moyenne pondérée des sous-effets = 12.0

Ecart type pondéré en pourcentage = 3.2

Effet InfB par rapport à Ref (NivB) 2 sous-effets (s-e)

Sous-population

	s-e	Eff	Tot	Khi2	Test
Cath	-9.4	244	492	2.0	ns
NonC	-9.4	248	492	2.0	ns

moyenne pondérée des sous-effets = -9.4

Ecart type pondéré en pourcentage = 0.0

Effet SupB par rapport à Ref (NivB) 2 sous-effets (s-e)

Sous-population

	s-e	Eff	Tot	Khi2	Test
Cath	9.1	365	798	3.9	**
NonC	2.7	433	798	0.3	ns

moyenne pondérée des sous-effets = 5.6

Ecart type pondéré en pourcentage = 3.2

Résumé des paramètres de l'analyse tabulaire

Modalité à expliquer : LatinSeul

Situation de référence	Intervalle de confiance à 1 écart-type	
NonC NivB		
60.0%	56.2%	63.8%
Effets marginaux	Intervalle de confiance à 1 écart-type	
Cath 12.0%	8.7%	15.2%
InfB -9.4%	-9.4%	-9.4%
SupB 5.6%	2.4%	8.8%

On peut donner les résultats à la manière de la régression linéaire :

$$\%LatinSeul = 12,0 \times Cath - 9,4 \times InfB + 5,6 \times SupB + 60,0$$

Pour la régression linéaire on a

Paramètres de la régression en pourcentages

Régression linéaire
 Modalité à expliquer : LaSeul
 Situation de référence NonC NivB
 0.5818 58.2 Khideux de Vrais.

Effets marginaux
 Cath 0.1213 12.1 28.2288 0.000 ***
 InfB -0.0917 -9.2 5.7013 0.016 **
 SupB 0.0600 6.0 7.4835 0.006 ***

Log-Vraisembl. -604.4978
 Cst seulement 1323.9111
 Modèle complet 1208.9957
 Khideux Vrais. 114.9154
 Degré liberté 3 Prob.=0.000 ***
 Nb d'itérations= 183

$$\%LatinSeul = 12,1 \times Cath - 9,2 \times InfB + 6,0 \times SupB + 58,2$$

Et pour la régression logistique :

Paramètres de la régression en pourcentages
 Régression logistique
 Modalité à expliquer : LaSeul
 Situation de référence NonC NivB ChancesRef % Khideux de Vr.
 0.3334 1.3957 58.3

Effets marginaux Odds-ratio
 Cath 0.5383 1.71 12.3 27.4877 0.000 ***
 InfB -0.3919 0.68 -9.7 5.7602 0.016 **
 SupB 0.2591 1.30 6.1 6.7441 0.009 ***

Log-Vraisembl. -604.7735
 Cst seulement 1323.9111
 Modèle complet 1209.5471
 Khideux Vrais. 114.3640
 Degré liberté 3 Prob.=0.000 ***
 Nb d'itérations= 158

D'une manière multiplicative on a

ChancesRef

$$\text{PropLatinSeul} / (1 - \text{PropLatinSeul}) = 1,71\text{Cath} \times 0,68\text{InfB} \times 1,30\text{SupB} \times 1,3957$$

Connaissant les chances de la référence $ch = (p / 1-p)$ on peut en déduire p

$p = Ch / (1 + Ch)$ et retrouver l'effet en pourcentage.

La comparaison des trois méthodes montre leur similitude qui traduit leur nature semblable profonde.

Comparaison des méthodes

	Tabulaire	Linéaire	Logistique
Cat	12,0	12,1	12,3
InfB	-9,4	-9,2	-9,7
SupB	5,6	6,0	6,1
Ref.	60,0	58,2	58,3

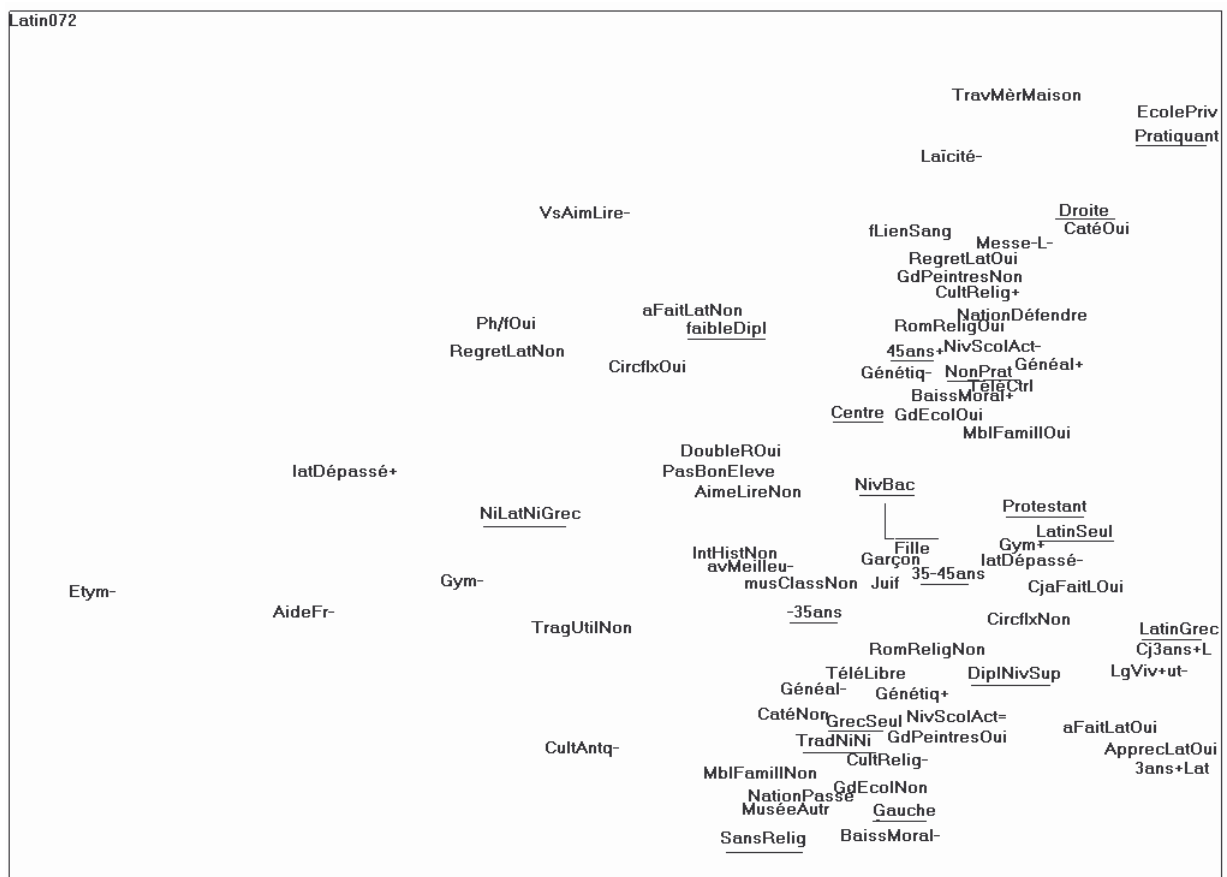
Une analyse factorielle de l'ensemble des données (supplémentaires soulignées) montre, en haut du graphique, du côté catholique (pratiquant), la proximité avec l'affiliation de droite, le choix de l'école privée, le choix que la mère reste à la maison (plutôt qu'elle travaille), le peu d'importance apporté à la laïcité, l'importance donnée à la culture religieuse, le sentiment que l'idée de nation est à

défendre, que le niveau scolaire baisse, un souci de généalogie et une défiance envers la procréation médicalement assistée, le désir d'un contrôle parental sur la télévision.

Inversement en bas, la gauche, sans religion ou de tradition catholique mais ni croyant ni pratiquant, prend des positions inverses symétriques.

Par contre les niveaux de diplôme se situent en intermédiaire avec les deux types d'affiliations et en liaison avec le choix du latin (pour les plus diplômés) ou avec son refus (pour les moins diplômés). La particularité de la situation réside dans le fait que l'opposition au latin de gauche est contrecarrée par le fort niveau de diplôme. Inversement, le faible diplôme est contrecarré par l'affiliation de droite catholique.

Le plus lié au fort diplôme est le choix du latin et du grec, la gauche évite le conflit en choisissant le grec seul, emblème de la démocratie alors que le latin a été très lié à l'Eglise.



Site pour chargement de Trideux (et textes de référence)

<http://perso.wanadoo.fr/cibois/SitePhCibois.htm>

(ou <http://www.printemps.uvsg.fr/> rubrique logiciels)

Bibliographie : Philippe Cibois, Les méthodes d'analyse d'enquêtes, PUF, 2007