EVALUATION DE LA PROBABILITE DE REUSSITE À L'UNIVERSITE D'ANGERS EN PREMIERE ANNEE : UNE ETUDE ECONOMETRIQUE SUR DONNEES D'ENQUÊTE

1. LA BASE DE DONNEES

1.1. L'échantillon

La base de données a été construite après une enquête réalisée durant l'année universitaire auprès des étudiants de première année dans les filières suivantes :

- Sciences Economiques Cholet,
- Sciences Economiques Angers,
- AES Cholet.

Tableau n°1 : Taille des échantillons et % de réussite

	AES1 Cholet	***SE1 Cholet***	***SE1 Angers***	***TOTAL***
Taille échantillon enquêté (nb personnes)	105	28	96	229
Taux de réussite (%)	39.04	53.57	48.95	44.97

1.2. Les déterminants de la réussite et codage des variables utilisées

Pourquoi certains étudiants réussissent-ils leurs études universitaires et d'autres pas ? Le seul critère de l'intelligence ne suffit plus aujourd'hui pour garantir un diplôme universitaire. Nous allons (essayer de) dresser une liste des conditions et attitudes internes et externes qui nous paraissent indispensables pour mener à bien des études universitaires. Les conditions et attitudes internes, c'est à dire propre à l'étudiant sont difficiles à quantifier. On peut citer par exemple le fait qu'un étudiant ait un projet, vise un objectif et se projette dans l'avenir. Dans ces conditions de motivation on trouve la pertinence du domaine d'études, la confiance en sa capacité de réussir la capacité de structurer l'avenir et l'autonomie dans la prise de décision[1]. Les conditions externes sont plus nombreuses et quantifiables. Tout d'abord on peut considérer la formation antérieure personnelle aux études universitaires (nature du bac, moyenne au bac, obtention du bac,...) et la formation des parents c'est à dire en fait l'appartenance à une classe sociale déterminée qui va fortement conditionner les conditions matérielles de l'étudiant durant ses études (bourse, travail durant l'année, lieu d'habitation, voiture,...). Le revenu des parents est dans les exemples précédents sous-jacents mais extrêmement déterminants or il est assez difficile en France de faire une enquête en demandant le salaire des personnes. C'est la raison pour laquelle nous avons remplacé la variable revenu des parents par la variable CSP des parents en distinguant la CSP de la mère de celle du père. Ensuite il faut s'intéresser au confort des études, les informations des programmes, des cours, l'intégration, les contacts entre étudiants et/ou les

[1] Remarquons que cela est peut-être utopique, la plus part des étudiants n'ont aucune idée de leur avenir.

professeurs, l’accès à la bibliothèque, au restaurant universitaire, le soutien de leur proche (parents), les conditions financières et de logement...Ces éléments qui sont bien réels sont assez difficiles à évaluer. Mentionnons aussi mais dans une catégorie à part l’âge de l’étudiant (il vaut mieux être jeune pour faire des études) et le sexe car nous savons que les filles obtiennent de meilleurs résultats que les garçons tout au long de leur parcours scolaire et de ce fait sans doute aussi universitaire. Le tableau n°2 résume les déterminants de la réussite que nous avons choisie en fonction des disponibilités humaines et temporelles.

En guise de conclusion interrogeons-nous sur les conseils d’un professeur canadien données aux nouveaux étudiants lors de la près rentrée universitaire : «Décide où tu veux aller; le reste, c'est comme un moyen de transport.».

Tableau n°2 : Variables explicatives-codage-notation

Variables	***Notations***
Sexe codé : ***1 si fille*** ***0 si garçon***	SEXE
Taille de la commune ***(en nombre de personnes)***	TAC
Lieu d'habitation codée : ***1 si chez les parents*** ***0 si cité universitaire et autres***	LHAB
Age (en années)	AGE
Obtention du bac codée : ***1 si à l'heure*** ***0 si en retard***	OBAC
Moyenne au bac [0 à 20]	MEANBAC
Nature du bac codée : ***1 si économie (ES)*** ***0 si art-langue-autres***	BACES
Nature du bac codée : ***1 si mathématique (S)*** ***0 sinon***	BACS
Etudes précédentes codées : ***0 si déjà dans le supérieur*** ***1 si dans le secondaire***	ETP
Bourse codé : ***1 si boursier*** ***0 sinon***	BOURSE
Travail durant l'année codé : ***1 si travail*** ***0 sinon***	TRAV
Nombre de personnes au foyer (nombre de personnes)	NBPF
Bureau personnel codé :	BUR

1 si bureau personnel ***0 si pas de bureau et à plusieurs***	
<u>Origine du secondaire</u> codée : ***1 si privé*** ***0 si public***	PRIV
<u>CSP de la mère</u> : ***variables dummy***	Mi (i : 1, 8)
<u>CSP du père</u> : ***variables dummy***	Pi (i : 1, 8)

Les CSP sont définies dans l'annexe n°2.

Tableau n°3 : Statistiques sur les échantillons. Par filière

Variables	***Proportions (en %)***		
	AES1 Cholet	***SC1 Cholet***	***SC1 Angers***
Sexe : ***-Fille***	75.38	40	44.82
Lieu d'habitation : ***-chez les parents***	58.46	90	44.82
Obtention du bac : ***-à l'heure***	50.76	70	82.75
Nature du BAC : ***- ES*** ***- S***	61.53 12.30	45 45	55.17 41.38
Etudes précédentes : ***-dans le secondaire***	90.76	70	86.20
Boursier	53.84	25	34.48
Travail durant l'année	9.23	20	10.34
Bureau personnel	92.30	100	100
Origine du secondaire : ***-privé***	55.38	45	48.27
CSP[2]	Père Mère	Père Mère	Père Mère
1	9.23	0	6.89

[2] Le pourcentage des CSP est différent de 100 car pour certain étudiant le père et/ou la mère sont décédés.

	6.15	0	0
2	10.76 1.53	20 5	6.89 0
3	10.76 3.07	15 10	27.58 31.03
4	9.23 15.38	20 5	24.13 20.68
5	18.46 38.48	10 50	6.89 17.24
6	27.69 13.84	25 15	10.35 13.79
7	1.53 0	5 0	0 3.44
8	3.07 16.92	0 15	0 10.34

Tableau n°3bis : Statistiques sur la totalité des filières

Variables	***Proportions en %***	***Variables***	***En %***
Sexe : -Fille	61.40	***CSP du Père***	
Lieu d'habitation : -chez les parents	60.52	***1***	7.01
Obtention du bac : -à l'heure	62.28	***2***	11.40
Nature du BAC : - ES - S	57.01 25.43	***3***	15.78
Etudes précédentes : -dans le secondaire	85.96	***4***	14.91
Boursier	43.85	***5***	14.03
Travail durant l'année	11.40	***6***	22.80
Bureau personnel	95.61	***7***	1.75
Origine du secondaire : -privé	51.75	***8***	1.75

2. COMMENT ESTIMER CETTE PROBABILITE ?
2.1. Par la méthode des MCO (ou GLS)

Considérons l'événement E ={réussite d'un individu à l'examen de fin d'année du DEUG 1}. Dans ce cas simple il n'y a que deux alternatives :

- soit cet événement se réalise pour l'individu i,
- soit cet événement ne se réalise pas.

Nous sommes en présence d'un modèle à choix binaire encore appelé en économétrie *modèle dichotomique*. La variable que nous voulons expliquer (i.e. la variable endogène) est une variable qualitative prenant dans ce cas particulier deux modalités différentes, on ne peut donc pas appliquer les méthodes de régression standard. Nous pouvons pour surmonter le problème réaliser un codage, qui est une représentation quantitative d'une variable qualitative. Le codage dans notre exemple peut être le suivant :

$Y_i = 1$ si réussite
$Y_i = 0$ si échec[3] pour l'individu i.

La variable endogène est maintenant quantitative et nous pouvons appliquer les méthodes économétriques traditionnelles. Comme l'on montré certains auteurs (notamment Mac Gillivray, 1970) *la fonction de probabilité linéaire* (estimé par OLS) ne permet plus de vérifier dans le cas d'un modèle à variable dépendante dichotomique les hypothèses relatives à la méthode des moindres carrés ordinaires. Sans faire de démonstration nous pouvons citer les points suivants :

[3] Ce codage est totalement arbitraire. A la place de 1 on peut prendre 2 ou 100.

➢Une étude graphique des observations montre que l'approximation linéaire est peu adaptée au problème. Le nuage des points, se trouvant sur deux droites parallèles (une pour Y=0, une autre pour Y=1) ne peut pas être approché par une seule droite.

➢ Comme Y ne peut prendre que deux valeurs, il en est de même des résidus. Si le modèle considéré est $Y_i = b_0 + bX_i + \varepsilon_i$ alors :

ε_i prend la valeur $1-bX_i$ avec la probabilité P_i

et la valeur $-bX_i$ avec la probabilité $1-P_i$

Dans ce cas les résidus admettent obligatoirement une loi de probabilité discrète, ce qui interdit l'hypothèse de normalité.

➢ Si nous calculons la variance des résidus nous trouvons :

$$\mathrm{Var}(\varepsilon_i) = bX_i\,(1-bX_i) = P_i\,(1-P_i)$$

qui n'est pas constante pour toutes les observations. Des observations avec une probabilité proche de 0 (ou 1) auront des variances faibles, tandis que des observations avec une probabilité proche de .5 auront de fortes variances. Il y a donc bien hétéroscédasticité. Cette dernière implique une perte d'efficience des estimateurs qui sont néanmoins toujours consistants et sans biais. Dans le cas classique la présence d'hétéroscédasticité impliquerait l'utilisation des GLS. Nous ne pouvons pas utiliser ces derniers ici car la méthode d'estimation des GLS est identique à celle des OLS.

➢ Aucune méthode linéaire n'est satisfaisante car les prévisions prédites par ces méthodes indiquent pour certaines observations des probabilités supérieures à 1 voire négatives, ce qui est aberrant.

Il est donc préférable d'estimer une fonction non linéaire prenant ses valeurs dans l'intervalle [0, 1].

2.2. Des fonctions non linéaires (le modèle probit)

Ces méthodes proviennent directement de la biologie. Elles sont employées pour déterminer la façon dont des individus tolèrent un certain produit. Pour cela on effectue plusieurs expériences où des individus de caractéristiques différentes, placés dans des conditions différentes, sont soumis à diverses doses du produit. On observe à chaque fois si l'individu a ou non bien supporté l'expérience (notée Y). On introduit aussi un seuil de tolérance qui est la dose maximale que peut supporter un individu au cours d'une expérience. Cette variable peut être considérée comme aléatoire : deux individus de mêmes caractéristiques, placés dans les mêmes conditions, n'ont pas forcément les mêmes réactions. En outre on ajoute que les observations doivent être faites sur des individus différents, sinon les résultats d'une expérience pourraient dépendre des résultats de la précédente. De ce fait les résidus sont supposés être indépendants.
En tenant compte de ces éléments, sur le plan mathématique, on peut facilement déduire une loi de probabilité de Y, puis une fonction de vraisemblance.

En économétrie des variables qualitatives les lois de probabilité utilisées sont la loi normale (modèle probit ou normit) et la loi logistique (modèle logit). La différence entre le modèle PROBIT et LOGIT peut se résumer en quelques phrases. Le PROBIT fournit des estimations où la variable endogène ne prend que deux valeurs distinctes (0,1) tandis que le LOGIT permet d'estimer des modèles conditionnels et/ou multiples (valeurs : 0,1,2...). Si un LOGIT est binaire nous obtenons les mêmes estimations qu'un PROBIT. Pour cette raison dans notre étude nous avons utilisé le modèle PROBIT.

<u>Exposé du modèle probit :</u>

On suppose qu'il existe une variable théorique Z_i qui est déterminée par un ensemble de variables exogènes X_i . La variable Z_i est une variable continue qui est aléatoire et normalement distribuée. Nous pouvons écrire alors la relation suivante :

$Z_i = \alpha + \beta X_i$

Mais pour différentier cette relation de l'économétrie classique on va supposer que les observations de la variable Z_i ne sont pas disponibles, c'est à dire non observables. Ce que nous observons en pratique c'est la dummy Y_i qui prend la valeur 1 si $Z_i > 0$ et 0 sinon. C'est à dire que nous avons des données qui nous permettent seulement de distinguer les observations qui sont dans la première catégorie (fortes valeurs de Z_i) ou dans la seconde catégorie (faibles valeurs de Z_i). Nous avons alors les relations suivantes :

$Y_i = 1$ si $\alpha + \beta X_i + \varepsilon_i > 0$
$Y_i = 0$ si $\alpha + \beta X_i + \varepsilon_i \leq 0$

En tenant compte de ces deux relations nous pouvons écrire :

$\text{Prob}(Y_i = 1) = \text{Prob}(\varepsilon_i > -\alpha - bX_i) = 1 - F(-\alpha - bX_i)$

$Prob(Y_i = 0) = Prob(\varepsilon_i \leq -\alpha - bX_i) = F(-\alpha - bX_i)$

Avec

$$F(-\alpha-\beta X_i) = P_i = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{-\alpha-\beta X_i} \exp(-\frac{t^2}{2})dt$$

et $t \sim N(0,1)$

F représente donc la fonction de distribution cumulative de la loi normale. Par construction la variable P_i se trouve bien dans l'intervalle [0, 1].

La fonction de vraisemblance est donnée par l'expression :

$$L = \prod_{Y_i=0} F(-\alpha-\beta X_i) \prod_{Y_i=1} [1-F(-\alpha-\beta X_i)]$$

Les paramètres α et β sont estimés en maximisant cette vraisemblance qui est non linéaire dans les paramètres et ne peut être estimés par les programmes de régression conventionnels mais par des méthodes non linéaires.

2.3. Les algorithmes de résolution

Lorsque nous avons des observations individuelles la technique d'estimation la plus utilisée est celle du maximum de vraisemblance. Soit la fonction de vraisemblance donnée dans le cas général :

$$L(Y;b) = \prod_{i=1}^{N} \left\{ F(x_i b)^{Y_i} [1-F(x_i b)]^{1-y_i} \right\}$$

avec Y la variable endogène, b le vecteur des paramètres à estimer et F la fonction de répartition d'une loi de probabilité. Pour trouver l'optimum il est nécessaire de différentier la fonction de vraisemblance par rapport aux paramètres inconnus, égaliser les dérivées premières à zéro puis résoudre le système. Dans les faits il est plus facile de travailler avec le logarithme népérien de L (Ln L) qu'avec L elle-même car L est toujours positive et que la fonction logarithmique est une fonction monotone :

$$\text{Ln L} = \sum_{y_i=1} \text{Ln F}(x_i b) + \sum_{y_i=0} \text{Ln}\left[1 - \text{F}(x_i b)\right]$$

Une condition suffisante pour que le maximum global de Ln L soit unique est que cette fonction soit strictement concave.

Dérivons maintenant la log-vraisemblance par rapport au vecteur b des paramètres et annulons cette dérivée, on a :

$$\frac{\partial \text{Ln L}}{\partial b} = \sum_{y_i=1} \frac{f(x_i b)}{F(x_i b)} x_i^{'} - \sum_{y_i=0} \frac{f(x_i b)}{1 - F(x_i b)} x_i^{'} = 0$$

avec f la densité associé à F et x'_i le transposé du vecteur x_i.

En résolvant ces équations nous obtenons :

$$\sum_{i=1}^{N} y_i x_i^{'} = \sum_{i=1}^{N} F(x_i b) x_i^{'}$$

Ces équations de vraisemblances associées aux modèles dichotomiques simples sont non linéaires dans les paramètres et doivent être résolues au moyen d'algorithmes. Il existe trois principaux algorithmes :

- l'algorithme de Newton-Raphson,
- la méthode du score,
- la méthode de Berndt-Hall-Hall-Hausman.

Actuellement c'est l'algorithme de Newton-Raphson qui est le plus utilisé car sa mise en oeuvre est assez facile. Le but de cet algorithme est de trouver une racine de l'équation $\frac{\partial \text{Ln L}}{\partial b} = 0$

Pour cela on se donne une valeur initiale b0 (en principe zéro) et on cherche le plan tangent en ce point à la fonction considérée c'est à dire : $d = \frac{\partial \text{Ln L}}{\partial b}$. Sur le plan mathématique ce plan a pour équation :

$$d = \frac{\partial \text{Ln L}}{\partial b}(b0) + \frac{\partial^2 \text{Ln L}}{\partial b \partial b'}(b0)[b - b0]$$

Cette dernière équation constitue une approximation de $d = \frac{\partial \text{Ln L}}{\partial b}$. Nous allons donc approcher la racine de $\frac{\partial \text{Ln L}}{\partial b} = 0$ par la racine de:

$$\frac{\partial \text{Ln L}}{\partial b}(b0) + \frac{\partial^2 \text{Ln L}}{\partial b \partial b'}(b0)[b - b0] = 0$$

c'est à dire :

$$b1 = b0 - \left[\frac{\partial^2 \text{Ln L}}{\partial b \partial b'}(b0) \right]^{-1} \frac{\partial \text{Ln L}}{\partial b}(b0)$$

Nous recommençons la démarche en prenant b1 comme valeur initiale et ainsi de suite. Nous obtenons alors une formule de récurrence de la forme :

$$b_{H+1} = b_H - \left[\frac{\partial^2 \text{Ln L}}{\partial b \partial b'}(b_H) \right]^{-1} \frac{\partial \text{Ln L}}{\partial b}(b_H)$$

On se fixe un seuil, une limite permettant d'arrêter la récurrence. Si la différence entre b_{H+1} et b_H est inférieure ou égale à ε (que l'on détermine aussi) alors l'itération cesse. On dit que la méthode converge vers une limite qui est solution des équations de vraisemblance.

Les estimations réalisées utilisent l'algorithme de Newton-Raphson.

3. LES RESULTATS ECONOMETRIQUES

Avant de présenter les résultats économétriques et leurs commentaires différentes remarques doivent être faites. En premier lieu le nombre de variables exogènes. Nous n'avons pas eu la variable « salaire », celle-ci a été remplacé par les CSP de la mère et du père or du fait du nombre de CSP retenu i.e. 16 cela aurait conduit à un nombre de variables exogènes supérieur à la taille de l'échantillon (32>20 pour l'échantillon SC1Cholet) mais aussi a des problèmes de multicolinéarité entre ces variables. Nous avons de ce fait décidé arbitrairement d'éliminer la CSP de la mère. Deuxième remarque : le problème que l'on nomme en anglais « quasi-complete separation », c'est à dire que le modèle n'est pas identifié donc pas estimable. Une ou plusieurs variables prédise parfaitement la variable endogène et leurs coefficients tendent vers ± l'infinie. Pour surmonter ce problème il faut éliminer tout simplement ces variables en espérant qu'elles ne représentent pas une grande importance. Troisième remarque : les coefficients de régression obtenus n'ont pas la même valeur que dans une régression classique par OLS, **seul le signe à une signification : une variable dotée d'un coefficient de régression positif entraîne une augmentation de la probabilité de réussite**, au contraire une variable dotée d'un coefficient de régression négatif entraîne une augmentation de la probabilité d'échec. Ce qui est pertinent de calculer c'est l'élasticité de la probabilité de réussite par rapport à chacune des variables exogènes.

3.1. Par filière

Tableau n°4 : Résultats de la procédure d'estimation par PROBIT

Variables	***AES1 Cholet***		***SE1 Cholet***		***SE1 Angers***	
	Contribution	Significativité	Contribution	Significativité	Contribution	Significativité
SEXE	+	Forte	-	Aucune	-	Aucune
TC	-	Moyenne	-	Moyenne	-	Moyenne
LHAB	+	Aucune	+	Aucune	+	Moyenne
AGE	+	Faible	-	Aucune	-	Aucune
OBAC	+	Faible	-	Aucune	-	Aucune
MEANBAC	+	Forte	+	Moyenne	+	Moyenne
BACES	+	Forte	-	Faible	-	Moyenne
BACS	-	Aucune				
ETP	-	Moyenne	-	Faible	-	Aucune
BOURSE	+	Forte	+	Aucune	+	Aucune
TRAV	+	Moyenne			-	Aucune
NBPF	+	Faible			-	Aucune
BUR	-	Forte				
PRIV	+	Aucune	-	Moyenne	-	Aucune

P1	+	Aucune				
P2	-	Forte	+	Faible	-	Aucune
P3	-	Moyenne			-	Aucune
P4	-	Moyenne			-	Forte
P5	+	Moyenne				
P6	-	Moyenne	-	Aucune	-	Aucune
P7						
P8	+	Aucune				

Tableau n°5 : Tests statistiques[4]

	AES1 Cholet	***SE1 Cholet***	***SE1 Angers***
Nombre d'observations	65	20	29
Nombre d'observations positives	32	13	16
SSR	6.23289	3.05281	3.58251
R^2	.617076	.330020	.500610
Kullback-Leibler R^2	.59	.32	.47
Log likelihood	-17.7353	-8.69082	-10.5373
% de prédictions correctes	83	75	82

[4] Le R^2 n'est pas valable ici car l'estimation n'est pas linéaire.

Tableau n°6: Elasticités de la probabilité de réussite et d'échec par rapport à chacune des variables

Variables	***AES1 Cholet***		***SE1 Angers***		***SE1 Cholet***	
	Echec	**Réussite**	**Echec**	**Réussite**	**Echec**	**Réussite**
C	5.07141	-5.07141	0.68349	-0.68349		
SEXE	-0.25116	0.25116	0.060775	-0.060775	0.13883	-0.13883
TC	4.14086D-08	-4.14086D-08	1.7451D-06	-1.7451D-06	0.00001880	-0.00001880
LHAB	-0.016450	0.016450	-0.24335	0.24335	-0.16040	0.16040
AGE	-0.050746	0.050746	0.086743	-0.086743	0.0047574	-0.0047574
OBAC	-0.12623	0.12623	0.10324	-0.10324	0.45663	-0.45663
MEANBAC	-0.33079	0.33079	-0.30868	0.30868	-0.17226	0.17226
BACES	-0.68152	0.68152	0.45073	-0.45073	0.21431	-0.21431
BACS	0.025930	-0.025930				
ETP	0.26251	-0.26251	0.29984	-0.29984	0.41289	-0.41289
BOURSE	-0.23278	0.23278	-0.12365	0.12365	-0.055856	0.055856
BUR	0.37764	-0.37764				
TRAV	-0.25627	0.25627	0.14218	-0.14218		
NBPF	-0.053794	0.053794	0.066304	-0.066304		
PRIV	-0.053277	0.053277	0.082736	-0.082736	0.85440	-0.85440
P1	-0.024599	0.024599				
P2	0.45736	-0.45736	0.35395	-0.35395	-0.69473	0.69473
P3	0.22176	-0.22176	0.032810	-0.032810		
P4	0.13524	-0.13524	0.75958	-0.75958		
P5	-0.17610	0.17610				
P6	0.15431	-0.15431	0.10509	-0.10509	0.13453	-0.13453
P8	-0.075268	0.075268				

Tableau n°7 : Contribution des variables à la probabilité de réussite par ordre décroissant

AES1 Cholet	***SE1 Angers***	***SE1 Cholet***
BACES	MEANBAC	P2
MEANBAC	LHAB	MEANBAC
TRAV	BOURSE	LHAB
SEXE	TC	BOURSE
BOURSE	P3	TC
P5	SEXE	AGE
OBAC	NBPF	P6
P8	PRIV	SEXE
NBPF	AGE	BACES
PRIV	OBAC	ETP
AGE	P6	OBAC
P1	TRAV	PRIV
LHAB	ETP	
TC	P2	
BACS	BACES	
P4	C	
P6	P4	
P3		
ETP		
BUR		
P2		
C		

Les variables soulignées représentent une contribution négative.

3.2. Ensemble des filières (AES1+SE1Cholet+SE1Angers)

Tableau n°8 : Résultats de la procédure d’estimation par PROBIT

Parameter	Estimate	Standard Error	t-statistic	P-value
C	**-10.9170**	**4.24551**	**-2.57143**	*** [.010]**
SEXE	.422219	.335147	1.25980	[.208]
TC	-.287469E-05	.264652E-05	-1.08622	[.277]
LHAB	.351989	.329194	1.06925	[.285]
AGE	.187457	.151974	1.23348	[.217]
OBAC	**.810552**	**.393885**	**2.05784**	*** [.040]**
MEANBAC	**.543791**	**.190584**	**2.85329**	**** [.004]**
BACES	**2.17916**	**.644432**	**3.38152**	**** [.001]**
BACS	**2.07269**	**.699728**	**2.96214**	**** [.003]**
ETP	-.390992	.532173	-.734708	[.463]
BOURSE	-.027453	.343688	-.079879	[.936]
TRAV	-.617225	.532031	-1.16013	[.246]
NBPF	-.167921	.181408	-.925653	[.355]
BUR	-.610541	.754328	-.809384	[.418]
PRIV	-.421786	.367710	-1.14706	[.251]
P1	**1.92839**	**.972232**	**1.98346**	*** [.047]**
P2	.549417	.706993	.777118	[.437]
P3	.495617	.663068	.747460	[.455]
P4	-.227848	.746301	-.305303	[.760]
P5	**1.42955**	**.739956**	**1.93194**	**[.053]**

P6	.379748	.606489	.626142	[.531]
P7	.734325	1.07458	.683360	[.494]
P8	.943953	1.31196	.719499	[.472]

Les variables en caractère gras contribuent significativement et positivement à la probabilité de réussite.

Tableau n°9 : Tests statistiques

	Ensemble des filières
Nombre d'observations	114
Nombre d'observations positives	61
SSR	17.5448
R^2	.381726
Kullback-Leibler R^2	.34
Log likelihood	-51.2366
% de prédictions correctes	75

Sur un échantillon plus important nous retrouvons certaines variables classiques qui contribuent à la réussite en première année de l'Université à savoir : posséder un bac ES et S, avoir une moyenne élevée au bac, avoir eu le bac à l'heure et descendre d'un père agriculteur exploitant. En terme d'élasticité nous retrouvons à peu près les mêmes constatations :

Tableau n°10 : Elasticités par ordre décroissant

BACES	0,54553
BACS	0,51888
P1	0,48275
P5	0,35787
P8	0,23631
OBAC	0,20291
P7	0,18383
P2	0,13754
MEANBAC	0,13613
P3	0,12407
SEXE	0,1057
P6	0,095066
LHAB	0,088117

AGE	0,046928
TC	0
BOURSE	-0,0068727
NBPF	-0,042037
P4	-0,057039
ETP	-0,097881
PRIV	-0,10559
BUR	-0,15284
TRAV	-0,15452
C	-2,73297

ANNEXE 1
Enquête distribuée aux étudiants

❶ Taille de la commune de résidence des parents (en personnes):

❷ Lieu de domicile de l'étudiant durant l'année universitaire:
- chez les parents
- en cité universitaire..
- autres...(préciser)..

(Cocher une seule case)

❸ Age de l'étudiant:..

❹ Profession des parents:
- Père:...
- Mère:...

(Si possible codification de l'INSEE)

❺ Age d'obtention du bac:
- en avance..
- à l'heure...
- en retard..

(Cocher une seule case)

❻ Moyenne obtenue au bac

❼ Nature du bac obtenu (es-arts-lan-s-autres):.....................

❽ Etude(s) précédente(s):
- dans le secondaire
- déjà dans le supérieur..........

(Cocher une seule case)

❾ Etes-vous boursier?.....................

(Répondre par OUI ou NON)

❶⓿ Avez-vous un travail durant l'année universitaire?.......

(Répondre par OUI ou NON)

❶❶ Note obtenue au 1er contrôle continu:..................

(Indiquer la matière)

❶❷ Nombre de personnes au foyer (y compris vous-même):

❶❸ Pour étudier avez-vous un bureau:

- personnel...........
- à plusieurs............
- sans bureau.............

(Cocher une seule case)

❶❹ Origine du secondaire:

- public:...............
- privé:...

(Cocher une seule case)

❶❺ Nom et prénom.....

❶❻ Filière:

- Sciences Economiques....
- AES...........

❶❼ Ville:

- Angers..........
- Cholet....

ANNEXE 2
Codification des CSP

La codification des CSP s'est effectuée au niveau le plus agrégé de la nomenclature élaborée par l'INSEE.

1 correspond aux agriculteurs exploitants
2 correspond aux artisans, commerçants et chefs d'entreprises
3 correspond aux cadres et professions intellectuelles supérieures
4 correspond aux professions intermédiaires
5 correspond aux employés
6 correspond aux ouvriers
7 correspond aux retraités
8 correspond aux autres personnes sans activité professionnelle