Cahiers du CREAD n°70, 4ème trimestre 2004, pages 107-134.

DJAMEL FERROUKHI[*]

Le passé scolaire et sa relation à la réussite (ou l'échec) à l'examen du baccalauréat

I – Problèmatique et démarche

L'examen du baccalauréat sanctionne les études secondaires et ouvre les portes de l'université. Il porte sur l'ensemble des matières au programme officiel de la classe de terminale et les sujets revêtent un caractère régional (Est, Centre, Ouest et Sud) depuis quelques années tout en présentant, en théorie, le même niveau de difficulté.

Il est clair qu'un nombre élevé de facteurs pèsent sur la réussite (et l'échec) à cet examen et tiennent à la fois de l'environnement scolaire et familial dans lequel a évolué l'élève tout au long de sa scolarité. On se contentera, dans le cadre du présent article, de limiter notre travail à l'impact du passé scolaire de l'élève sur le phénomène de réussite au baccalauréat. Pour cela on a reconstitué le dossier pédagogique des candidats au baccalauréat de la session de juin 1995 des élèves de cinq lycées de la wilaya de Blida.

Pour chaque élève, nous avons reconstitué les performances (moyennes annuelles) à l'ensemble des matières officielles des classes de 9éme année fondamentale, de 1ère et de 3ème année secondaire et, pour ceux qui réussissent à l'examen du baccalauréat, les notes réalisées à ce test.

Seuls en effet, les notes des candidats reçus étaient saisies sur support informatique pour les besoins de l'orientation des nouveaux bacheliers à l'entrée de l'université. Il faut attendre 1997 pour qu'une première expérience soit réalisée à l'Institut National Pédagogique (IPN).

La moyenne au B.E.F, le vœu de l'élève quant au choix du tronc commun (1èreAS) et de la série du baccalauréat suivie en 2^{éme} année secondaire ainsi que le résultat du conseil d'orientation, sa situation en terminale (nouveau ou doublant) et le sexe de l'élève constituent l'histoire scolaire des candidats retenus dans l'échantillon.

Pour chacun des lycées retenus, on a consigné le collège (3^{éme} cycle) de provenance du candidat au baccalauréat ainsi que sa classe d'appartenance en dernière année de lycée (voir annexe).

Les méthodes d'estimation qui seront présentées plus loin possèdent des propriétés intéressantes (consistance, normalité, etc…) lorsque la taille de l'échantillon est suffisamment grande. Aussi, n'a-t-on retenu que les filières (séries) où les effectifs sont relativement étoffés, à savoir :

– Sciences de la nature et de la vie (SNV)
– Lettres et sciences humaines (option de la série lettres) (LSH).
Deux types d'approches sont retenues dans notre démarche :

– un modèle global (traitant l'ensemble de l'échantillon) spécifique à chaque série du baccalauréat est proposé pour tester l'hypothèse de savoir si le lycée d'appartenance a un impact sur la réussite au baccalauréat.

– un modèle propre à chaque lycée et à chaque filière du baccalauréat de sorte à pouvoir tester si le 3ème palier du fondamental de provenance du candidat influe (ou pas) sur la réussite au baccalauréat. Pour chacun des lycées, nous nous intéresserons aussi à la question de savoir si la répartition des élèves à l'entrée de la terminale (3ème AS) se fait de manière aléatoire ou, au contraire, obéit à des stratégies préalablement arrêtées par les responsables des lycées de constitution de groupes pédagogiques homogènes qui tiennent compte à la fois des performances antérieures des élèves et du profil des enseignants dont dispose l'établissement.

II – Modèle de réussite au baccalauréat

2.1 – Modèle

Les études secondaires sont sanctionnées par l'examen externe aux établissements qu'est le baccalauréat, ce dernier ouvrant l'accès à l'université. Depuis la réforme de l'enseignement secondaire de 1991, les candidats composent à l'ensemble des matières au programme de la 3ème année secondaire. La réussite à cet examen est repérée par la variable dichotomique suivante :

$$Y_{(t)} = \begin{cases} 1 \text{ si l'élève n° (t) réussit à cet examen} \\ 0 \text{ autrement} \end{cases}$$

La relation liant cette variable (Y_t) aux caractéristiques de l'élève (X_t) est donnée par la relation suivante :

$$\text{Probabilité } (y_t = 1) = F(X_t' . \beta) \qquad (\mathrm{I})_0$$

où F est une fonction de distribution.

On utilise fréquemment pour fonction de distribution F la fonction logistique en raison, d'une part, qu'elle présente une plus grande simplicité numérique et, d'autre part, le très grand nombre de situations que peut couvrir cette distribution. Dans ce dernier cas (logistique), l'expression (I) se réduit à :

$$F(X_t' . \beta) = 1 / \left(1 + exp\left(-X_t' . \beta\right)\right) \qquad (II)$$

Le modèle (I) est du type qualificatif et l'estimation du paramètre passe par des procédures itératives d'optimisation que l'on ne développera pas en détail ici. On se contentera d'esquisser, très succinctement, les différentes étapes de l'estimation du paramètre.

La première phase consiste à formuler la fonction de vraisemblance associée à un tel modèle pour l'échantillon considéré :

$$\mathrm{L}\left(\beta \;/\; X_t\right) = \prod_{t=1}^{n} \left(F\left(X_t^{'}.\beta\right)\right)^{y_t} .\left(1 - F\left(X_t^{'}.\beta\right)\right)^{1-y_t} \qquad \text{(III)}$$

On cherche alors à trouver la valeur $\hat{\beta}$ de β qui maximise cette fonction , appelée alors estimateur du maximum de vraisemblance (MLE). Un tel estimateur doit satisfaire alors à la condition nécessaire suivante:

$$\frac{\delta \log\left(L(\beta)\right)}{\delta\beta} = \sum_{t=1}^{n} \frac{Y_t - F\left(X_t^{'}.\beta\right)}{F\left(X_t^{'}.\beta\right) - (1 - \left(X_t^{'}.\beta\right)} . f\left(X_T^{'}.\beta\right) X_t = 0 \qquad \text{(IV)}$$

Le système (IV) est constitué de K équations (dimension de β) à K inconnues (β) Comme on peut le constater, ce système n'est pas linéaire par rapport à l'inconnue β. On a alors recours à des techniques d'optimisation pour cette catégorie de problèmes, la recherche d'un optimum a pour sous-bassement la méthode de base de Newton Raphson.

Ces techniques sollicitent très souvent la dérivée seconde de la fonction de vraisemblance (hessien) qui, dans notre cas (fonction **Logit**), prend la forme suivante :

$$\frac{\delta^2 \log L\left(\beta / X_t\right)}{\delta\beta\delta\beta'} = -\sum_{t=1}^{n} \left[\frac{Y_t}{F^2\left(X_t^{'}\beta\right)} + \frac{1 - Y_t}{\left(1 - F\left(X_t^{'}\beta\right)\right)^2}\right] f^2\left(X_t^{'}\,\beta\right) X_t - X_t^{'}$$

$$+ \sum_{t=1}^{n} \left[\frac{Y_t - F\left(X_t^{'}\beta\right)}{F\left(X_t^{'}\beta\right)\left(1 - F\left(X_t^{'}\beta\right)\right)}\right] f'\left(X_t^{'}\beta\right) X_t X_t^{'}$$

2.2 – Méthodes itératives d'estimation

A l'itération «m», l'estimateur du paramètre β est donné, selon la procédure d'optimisation retenue, par :

– Méthode itérative de Newton Raphson :

$$\hat{\beta}_m = \hat{\beta}_{m-1} - \left[\frac{\delta^2 \log\left(\beta / X_t\right)}{\delta\beta\delta\beta'} \right]^{-1}_{\hat{\beta}m-1} \cdot \left[\frac{\delta \log L\left(\beta / X_t\right)}{\delta\beta} \right]_{\hat{\beta}_{m-1}}$$

– Méthode du score :

$$\hat{\hat{\beta}}_m = \hat{\hat{\beta}}_{m-1} - \left[E\left[\frac{\delta^2 \log\left(L\left(\beta / X_t\right)\right)}{\delta\beta\delta\beta'} \right] \right]^{-1}_{\hat{\hat{\beta}}_{m-1}} \cdot \left[\frac{\delta \log L\left(\beta / X_t\right)}{\delta\beta} \right]_{\hat{\hat{\beta}}_{m-1}}$$

– Methode de Berndt-Hall-Hall-Haussman :

$$\hat{\hat{\beta}}_m = \hat{\hat{\beta}}_{m-1} - \left[\frac{\delta \log L\left(\beta / X_t\right)}{\delta\beta} \cdot \frac{\delta \log L\left(\beta / X_t\right)}{\delta\beta'} \right]^{-1}_{\hat{\hat{\beta}}_{m-1}} \cdot \left[\frac{\delta \log L\left(\beta / X_t\right)}{\delta\beta} \right]_{\hat{\hat{\beta}}_{m-1}}$$

Il s'agit à partir d'une valeur de départ appropriée $\hat{\hat{\beta}}_0$ de β d'initier la procédure, pour aboutir à la valeur recherchée β (maximum de vraisemblance). La convergence vers le maximum global n'est pas toujours garantie, à moins que la fonction de vraisemblance ne remplisse certaines conditions (de convexité plus exactement dans notre cas). On peut montrer que lorsque l'on retient comme hypothèse de travail une fonction de distribution F une fonction logistique, la convergence est assurée.

Toute une panoplie de logiciels est disponible sur le marché pour l'estimation du paramètre β dans le cas d'une fonction de probabilité logistique (TSP, SAS, SPSS, Eviews, etc.).

2.3 – Tests d'hypothèses et tests des variables omises

A. Tests d'hypothèses

Très souvent, on désire tester la nullité d'un ensemble de **q** coefficients, c'est à dire la non-pertinence des variables explicatives correspondantes. On peut aussi souhaiter la nullité de l'ensemble des coefficients. L'hypothèse de la nullité d'un ensemble de **q** coefficients peut être exprimée de la manière suivante :

$Q\beta=0$

avec Q(q*K) β (K*1) O(q*1)

où les éléments de Q sont tous nuls sauf les Q(i,j) (= 1) correspondant aux coefficients dont on veut tester la nullité et O est une matrice nulle.

Pour un échantillon suffisamment fourni, trois statistiques équivalentes sont utilisées à cette fin. Il s'agit des statistiques de Wald, du multiplicateur de Lagrange et du rapport du Maximum de Vraisemblance (LRT). Pour des échantillons de taille réduite, cependant, les conclusions à partir de ces statistiques peuvent être contradictoires.

Les logiciels spécialisés en statistiques et en économétrie font référence essentiellement aux tests de Wald et du rapport de vraisemblance que l'on se propose de présenter très brièvement ici.

– Test de Wald

La statistique de Wald est ainsi définie :

$$W = \left(Q\,\hat{\beta}\right)'\left(Q.Var\left(\hat{\beta}\right);Q'\right) - 1\left(Q\,\hat{\beta}\right)$$

où $\hat{\beta}$ est l'estimateur de β.

La distribution asymptotique de cette quantité est une loi variable qui suit une loi du Khi-Deux à **q** degrés de liberté et l'hypothèse Ho : $Q\beta=0$ est rejetée si la valeur de la statistique de Wald dépasse un seuil critique préalablement arrêté..

– Test du rapport de vraisemblance (LRT)

Si l'on désigne par L le logarithme de la fonction de vraisemblance associée à l'échantillon, $\hat{\beta}$ l'estimateur maximisant cette fonction et $\overline{\beta}$ la valeur de cet estimateur sous la contrainte $Q\beta=0$, on peut montrer alors que le rapport :

$$LRT = -2\left(L\left(\hat{\beta}\right) - L\left(\overline{\beta}\right)\right)$$

suit asymptotiquement une loi du Khi-Deux à **q** degrés de liberté**.**

Ce rapport va être particulièrement utile pour le test des variables omises. En effet, si l'on doit choisir entre deux modèles (disons I et II) dont l'un est une version «réduite» de l'autre :

Modèle I : les variables explicatives sont x_1, x_2, …., x_K
Modèle II : les variables explicatives sont x_1, x_2, …, x_K, x_{K+1},…,

x_{K+q} préférer le modèle I à II revient à tester l'hypothèse suivante :

$$\textbf{Ho :}\ \beta_{(k+1)}=\beta_{(k+2)}=...=\beta_{(k+q)}=0$$

Cette dernière expression peut être simplement formulée sous la forme matricielle exposée plus haut, à savoir :

$$\textbf{Ho :}\ Q\beta=0$$

Si, pour un seuil de confiance donné a, le LRT est inférieur à la valeur du Khi-2 tabulée à **q** degrés de liberté alors, l'hypothèse Ho est acceptée, c'est-à-dire que le modèle I est retenu.

B – Tests de spécification du modèle de base

Sur l'ensemble des variables relatives aux performances pédagogiques réalisées par l'élève au cours de son cheminement scolaire ($9^{ème}$AF-$3^{ème}$AS), nous avons à sélectionner les plus pertinentes d'entre-elles dans l'explication du phénomène de réussite (ou d'échec) au baccalauréat.

Il s'agit plus exactement de sélectionner, dans un premier temps, parmi les variables représentant les performances moyennes annuelles au cours successivement de la $9^{ème}$ AF, de la $1^{ère}$ AS, de la $3^{ème}$AS et à l'examen du BEF, celles qui jouent un rôle plus déterminant que le reste dans le phénomène qui nous intéresse ici, à savoir la réussite à l'examen du baccalauréat. Pour cela, nous aurons fréquemment recours au cours de ce travail à ce type de test pour juger de la pertinence (ou non) d'un groupe particulier de variables exogènes.

Les tableaux(I) et (II) donnent les résultats de ces tests respectivement pour les séries «Sciences de la nature et de la vie» (SNV) et «Lettres et sciences humaines» (LSH). La démarche consiste à introduire, à chaque étape, une variable explicative additionnelle et de tester sa pertinence par rapport au modèle précédent sur la base des critères arrêtés plus haut. Il ne peut-être fait appel au test basé sur le maximum de vraisemblance que si le modèle étudié inclut un terme constant

Pour la série SNV, l'ensemble du parcours de l'élève participe au résultat observé à l'examen de fin de cycle secondaire (Tableau (I)). En d'autres termes, la réussite ou l'échec à cet examen dépend de la performance de l'élève tout au long de son cheminement scolaire alors que pour l'autre catégorie de population (LSH), l'issue de l'épreuve du baccalauréat dépend essentiellement du score moyen réalisé en dernière classe de lycée (Tableau (II)). Ces sur la base de ces modèles spécifiques à chacune des deux séries du baccalauréat que l'on se propose d'affiner un peu plus l'analyse.

2.4 – Résultats

Les coefficients estimés des variables introduites dans les modèles (binaires ou dichotomiques) présentés au début de ce travail ne peuvent s'interpréter comme l'effet marginal sur la variable dépendante (Y) des variables explicatives correspondantes comme c'est le cas dans le modèle linéaire classique. Pour mieux fixer les idées, rappelons que :

$Probabilité\left(y_t = 1\right) = F\left(X'_t . \beta\right)$ et, de ce fait,

$E(y / x, \beta) = F(x' \beta)$

$E(y / x , \beta) = F(x'\beta)$

A partir de cette dernière expression, on peut en déduire que :

$$\partial E(y/x, \beta)/\partial x_j = \beta_j . f(x' \beta)$$

β_j est affecté, dans cette expression, du facteur $f(x'.\beta)$ (fonction de densité) qui lui-même dépend se la valeur prise par l'ensemble des variables explicatives (x).

Sachant que $f(x'\beta)>0$, une valeur positive de $\beta(j)$ aura tendance à augmenter la probabilité de réussite du candidat au baccalauréat alors qu'une valeur négative produit l'effet inverse.

A – Modèle de base

Dans un premier temps, seules les variables pédagogiques que le test précédent a retenu sont introduites dans le modèle Le modèle **probit** est retenu, pour les raisons avancées plus haut, pour mettre en relation la réussite ou pas à l'examen du baccalauréat avec cette première série de variables explicatives.

– SNV :

La troisième colonne (B) du tableau (III) révèle que si le test de spécification retient le groupe des quatre variables, la procédure d'estimation ne retient que la performance globale moyenne réalisée en dernière classe de lycée (Moygen 3) et à l'examen national de fin de cycle fondamental (MoyBEF) et rejette les scores obtenus respectivement en 9ème AF et en 1ère AS.

Ces deux variables sont toutes deux affectés d'un signe positif, ce qui se traduit par un impact positif sur la réussite à l'examen du baccalauréat avec, cependant, un poids quatre fois plus important de la Moygen3 (coefficient = 0.42 contre 0.11 pour la Moy-BEF).

– LSH :

En plus du terme constant, seule la moyenne générale obtenue en classe de terminale est retenue par le modèle, cette dernière étant affectée d'un coefficient positif égal à 0.28 (colonne A du tableau IV).

Aucun des scores intermédiaires (Moyen, Moyen 1, BEF) n'est retenu par le modèle spécifique à cette à cette série du baccalauréat.

B – Modèle avec effet «sexe» et «situation en 3ème AS»

L'évaluation des connaissances des élèves à travers l'examen du baccalauréat effectué pour la première fois en 1994, renouvelée par la suite en 1996 et en 1997, montre très nettement que la performance des filles à cet examen est supérieure à celle des garçons [(1)(2)] tant au niveau global (taux de réussite au bac) qu'au niveau des différentes matières du programme officiel à l'exception d'une ou de deux séries sur les 15 existantes au sein de l'enseignement secondaire.

D'autre part, plus d'un tiers des candidats inscrits à cet examen le passent au moins pour la seconde fois puisqu'au cours des années 90 des «classes spéciales» sont créées pour les jeunes que la réglementation n'autorise pas à redoubler la terminale.

Partant de ces constats, nous avons voulu alors savoir si effectivement les tests statistiques allaient confirmer (ou pas) l'impact de ces deux variables sur la réussite au bac. Les tableaux III et IV, dans leur colonne respective C et B, donnent le résultat des estimations.

La conclusion est négative quant à la pertinence du sexe du candidat à cet examen alors que le redoublement en classe de terminale augmente les chances de réussite des candidats (coefficient négatif accepté au seuil de 5 %) en ce qui concerne la série «Sciences de la nature et de la vie».

C – Effet «lycée d'appartenance»

Les établissements du secondaire se différencient par leur ancienneté, c'est à dire par leur date de création, ce qui peut, probablement, se traduire par un corps enseignant plus expérimenté, ne recrutent pas automatiquement au sein des mêmes strates sociales que les nouveaux lycées et les collèges qui relèvent de leur circonscription géographique de recrutement des élèves de l'enseignement fondamental sont loin, en général, de préparer de manière homogène les futurs lycéens. Pour toutes ces raisons, on peut «soupçonner», à priori, une certaine disparité dans les résultats entre les établissements à l'examen de fin de cycle secondaire. Cette hypothèse d'homogénéité des lycées face à l'examen du baccalauréat peut être testée en introduisant dans le modèle une variable d'appartenance à un établissement déterminé.

Les tableaux III et IV reprennent, dans leur dernière colonne, les résultats de ce test. En ce qui concerne la série SNV, cette hypothèse se vérifie puisque les élèves issus des lycées I (Ibnou Rouchd), II (El-Feth) et VI (Khazrouna) ont des probabilités plus élevées, relativement à deux autres établissements (Mahi et Oued-El-Alleug), de réussir l'examen du baccalauréat.

En ce qui concerne la deuxième série du bac (Tableau IV), les élèves du lycée IV (Mahi) semblent moins bien préparés que leurs camarades des autres établissements (Ibnou Rouchd, El Feth et Khazrouna) puisque le coefficient correspondant est retenu seuil de 5 % et est affecté d'un signe négatif.

D – Effet «3ème cycle fondamental» de provenance de l'élève

Chaque lycée recrute ses élèves au sein d'un ensemble bien déterminé d'établissements du 3ème cycle fondamental (Ex-collège) de sa circonscription géographique (Carte scolaire). La question qui vient alors à l'esprit est de savoir si la préparation adéquate à l'examen du baccalauréat ne se détermine pas déjà bien plus en amont, c'est à dire avant l'accès au lycée et que l'enseignement secondaire, au contraire, aura tendance à maintenir et peut-être même, dans certaines situations, à amplifier les écarts hérités du collège.

C'est là une question très délicate à traiter tant sur le plan scientifique car le processus d'acquisition des connaissances est complexe et multifactoriel et, d'autre part, qu'au niveau politique puisqu'elle risque de remettre en question l'égalité des chances devant la réussite scolaire que les pouvoirs publics considèrent comme déjà acquise.

Ces situations existent au sein même des pays les plus avancés et le danger ne se situe pas dans la reconnaissance d'une telle réalité mais plutôt d'en occulter l'existence. Tout le génie d'un système consiste justement à prendre acte d'un tel phénomène, s'il existe, et de mobiliser les efforts pour en limiter les conséquences.

Comme ce fut le cas pour les lycées, une variable d'appartenance de l'élève à un collège déterminé est introduite dans le modèle (SNV). Le tableau V (colonne A) reprend le résultat de l'estimation de ce nouveau modèle. Les collèges 8 (Bouarfa), 11 (Semiani) et 26 (El Markazia El-Djadida) sortent du lot puisque les coefficients correspondant sont dotés d'un signe positif, ce qui signifie que les élèves qui en sont issus ont une probabilité plus élevée de réussir leurs études secondaires que ceux des autres établissements du 3ème cycle fondamental de l'échantillon retenu.

Nous retiendrons qu'au sein de ce groupe particulier de collèges, les élèves d'El Markazia El-Djadida (collège 26) sont particulièrement plus brillants puisque ce dernier est doté du plus grand coefficient (+1.25). Au sein de ce même groupe, on peut inclure, au seuil de 10 %, le collège 9 (Youcef Zoukel). A l'autre extrémité, c'est à dire les moins performants, deux collèges exigent, à notre sens, une attention particulière. Il s'agit précisément des collèges 15 (Zaouia Beni-Tmou) et 27 (Lenaï Abd-Errahmane) puisque leurs coefficients respectifs sont affectés d'un signe négatif (Tableau IV).

Pour ce qui est de la série LSH, cette hiérarchie (Tableau VI, colonne A) est moins évidente que pour le cas du SNV. En effet, seul le collège 9 (Youcef Zoukel), relevant de la circonscription géographique du lycée Ibnou Rouchd, semble être en décalage au niveau du baccalauréat par rapport aux autres établissements du 3ème cycle fondamental, la variable correspondante étant dotée d'un coefficient négatif et est retenue comme pertinente (au seuil de 5 %) par le modèle explicatif du phénomène de réussite au bac.

Evidemment, il existe une étroite correspondance entre les résultats obtenus ici (variable «collège») et ceux obtenus précédemment (variable «lycée») puisque les établissements du 3ème cycle fondamental constituent un partitionnement des lycées étudiés conformément à la Carte scolaire. Le rapprochement entre ces deux dernières démarches sera opéré à l'issue de la partie relative aux modèles spécifiques aux différents lycées.

E – Modèles spécifiques aux lycées

– SNV : la seule variable explicative partagée par les trois lycées (colonnes B, C, D du tableau V) est le score moyen réalisé en classe de terminale (Moygen 3). La moyenne générale obtenue successivement à l'examen du BEF (Moy BEF) et en fin de 1ère année secondaire

(Moygen 1) ainsi que le sexe du candidat sont spécifiques au seul lycée d'Ibnou Rouchd.

L'ensemble de ces variables pédagogiques (Moygen 3, Moygen 1, Moy BEF) est doté d'un coefficient positif signifiant que leur accroissement se traduit par une probabilité plus élevée de réussir en fin de parcours et, en même temps, un avantage certain pour les filles du lycée d'Ibnou Rouchd à cet examen.

Le collège d'origine de l'élève ne joue pas un rôle déterminant dans le phénomène étudié pour les circonscriptions géographiques des lycées II et V (El Feth et Oued El-Alleug) alors que les établissements du 3ème cycle fondamental 8 (Bouarfa au seuil de 10 %) et 9 (Youcef Zoukel) semblent mieux préparés au baccalauréat au sein des candidats du lycée d'Ibnou Rouchd. Si l'on compare les modèles spécifiques aux lycées (colonnes B, C et D du Tableau V) avec le modèle général (colonne A), le lecteur peut être surpris par l'apparente contradiction des résultats obtenus.

En effet, aucun des collèges n'est retenu dans les modèles spécifiques à chacun des deux lycées II et V alors que certains d'entre eux sont considérés comme pertinents dans le modèle global (colonne A, Tableau V). C'est le cas des établissements 11 pour le lycée II et 15, 26 et 27 pour Oued El-Alleug (lycée V).

En fait, il faut rappeler que la performance du lycée El Feth est assez exceptionnelle au cours de la session du baccalauréat de Juin 1995 (44,44 % contre respectivement 31 et 15 % pour les deux autres établissements) et les collèges le composant se situent, statistiquement parlant, au même niveau de réussite (colonne C, Tableau V). Le résultat obtenu dans la colonne A de ce même tableau (V) est, somme toute, tout à fait logique du fait que le collège 11, le plus performant au sein de la circonscription géographique d'El Feth (doté du coefficient le plus petit en valeur absolue) émerge du lot lorsqu'il est mis en compétition avec l'ensemble le des collèges étudiés ici (tous lycées confondus). Oued El-Alleug, à l'inverse, est le moins performant et c'est ce qui explique la position des établissements de l'enseignement fondamental qui l'alimentent au sein de l'ensemble du groupe de collèges de l'échantillon.

– LSH : au seuil de 5 % aucun collège n'émerge dans les modèles relatifs aux lycées d'Ibnou Rouchd et d'El Feth, c'est-à-dire que la variable «3ème cycle» de provenance de l'élève n'intervient dans le modèle de réussite au bac (colonnes B et C du tableau VI). La seule situation où ce phénomène apparaît s'observe lorsqu'on arrête le risque d'erreur du type I au seuol de 10%. Dans ce dernier cas, les élèves du collège 9 (Youcef Zoukel) sont les plus handicapés face à l'examen du baccalauréat au sein du lycée d'Ibnou Rouchd (coefficient négatif égal à –1.35)

F – Effet «classe d'appartenance en terminale»

Dans cette dernière phase du travail nous avons voulu savoir si les élèves, au sein d'un même établissement (lycées) inscrits dans une même série du baccalauréat, sont répartis de manière aléatoire entre les différentes classes organisées en terminale ou, au contraire, leur répartition obéit à une autre logique, à savoir celle de répondre au souci de constituer des groupes homogènes de candidats, compte tenu de leurs performances pédagogiques antérieures et d'affecter en conséquence le personnel enseignant disponible qui permettent d'atteindre un niveau de rendement optimal de l'établissement à l'examen du baccalauréat.

Les résultats présentés ici intéressent uniquement deux établissements abritant un nombre suffisamment élevé de candidats au baccalauréat Il s'agit plus exactement des lycées Ibnou Rouchd (I) et El Feth (II) pour les deux séries du baccalauréat qui nous concernent dans cette étude. de candidats de candidats de candidats

Les tableaux VII et VIII indiquent que ce phénomène ne s'observe que dans le cas d'un seul établissement (Ibnou Rouchd) et uniquement dans le cas précis de la série SNV. Il est, cependant, délicat de conclure : dans l'état actuel des informations dont on dispose, si cette situation est le simple fruit du hasard ou relève d'une stratégie de dosage préalablement étudiée de composition homogène de groupes d'élèves.

III – Modèle de réussite au baccalauréat avec niveau de performance

3.1 – La démarche

Dans ce qui a précédé, et compte tenu du fait que l'on ne disposait pas de banques de données relatives aux scores réalisés au baccalauréat par les candidats non-reçus à cette épreuve on a tenté procédé à la mise en relation entre les caractéristique des candidats et le fait de réussir (ou pas) à cet examen. On peut, cependant, et compte tenu de la nature des données disponibles, pousser l'analyse un peu plus en profondeur. En effet, le baccalauréat est décerné avec une certaine mention selon le score réalisé à cet examen par le candidat :

– «passable» si la note se situe entre 10 et 12/20
– «assez bien» si la note est comprise entre 12 et 14/20
– «bien» si la moyenne générale se situe entre 14 et 16/20
– «très bien» si le score est supérieur à 16/20

Ainsi donc, on peut faire correspondre à chacun des candidats reçu au bac la mention reflétant son niveau de performance. Pour compléter la démarche, la mention «non-reçu» sera affectée aux candidats ayant échoué à cet examen, c'est-à-dire dont la moyenne générale est inférieure à 10/20.

On désignera par Y_t la variable endogène qui peut prendre l'une des modalités suivante :

0 si le candidat n° (t) n'est pas reçu au bac (mention «non-reçu»)
1 si la mention est «passable»
1si la mention est «assez bien»
2 si la mention est «bien»
3 si la mention est «bien»

De ce fait, chacun des candidats, à la proclamation des résultats de l'examen du baccalauréat, ne peut appartenir à la fois qu'à une des cinq catégories de classes d'élèves ainsi définies.

On est en mesure maintenant de mettre en place une relation qui permet d'expliquer le niveau de performance (mention au bac) par les antécédents pédagogiques des candidats.

3.2 – Le modèle

La variable expliquée ici (Y) est qualitative (mention) et prend plusieurs niveaux (polytomique). Elle constitue, par conséquent, une généralisation de la situation étudiée précédemment (dichotomique).

Probabilité $(Y_t = j) = F(X_t'.\beta_J)$

où j= 0, 1, 2, 3, 4 représente les modalités (mention) prises par la variable endogène Y_t et F une fonction de distribution.

Ce type de situation est généralement approché par le modèle **logit multinomial** abondamment traité par la littérature spécialisée dans ce domaine. Avant de revenir sur la situation précise qui nous intéresse directement, nous allons présenter le modèle dans un cadre plus général.

Désignons par P_1, $P_{2, \ldots}$Pm les probabilités, pour un individu (t) aux caractéristiques $X_{t,}$ d'appartenir respectivement à la catégorie (mention dans notre cas) j=1, 2, 3,…, m. L'idée est d'exprimer ces probabilités sous forme binaire simple à l'image du modèle dichotomique traité plus haut. Pour cela, posons les relations suivantes :

$P_1/(P_1+Pm) = F(Xt'\beta_1)$
$P_2/(P_2+Pm) = F(Xt'\beta_2)$
. .
.
$P_{m-1}/(P_{m-1}+Pm) = F(Xt'\beta_{m-1})$

Evidemment, comme $P_1+P_2+\ldots P_{m-1}+P_m=1$, la dernière probabilité n'a pas lieu d'être explicitée.

A partir de cette série de relations, on peut facilement établir le rapport entre les probabilités associées respectivement à la classe '' j '' (j=1,2, …,m-1) et à la dernière catégorie (''m'') :

$P_j / P_m = F(Xt'\beta_j)/(1+F(Xt'\beta_j))$

Si l'on désigne par : $G(Xt'\beta_j)$ ce rapport **(=**$F(Xt'\beta_j)$**/ (1+** $F(Xt'\beta_j)$**)** alors, la probabilité associée à ka classe ''j'' peut s'écrire simplement :

$$P_j = G(Xt'\beta_j) / (1 + \sum_{i=1}^{m-1} G(Xt'\beta_i)) \qquad j=1, 2, \ldots, m-1$$

$$\text{et} \quad P_m = 1/(1 + \sum_{i=1}^{m-1} G(Xt'\beta_i))$$

3.3 – Estimation des paramètres du modèle

L' estimation des paramètres β_j (j=1 ,2,…,m-1) du modèle va se faire par la méthode du maximum de vraisemblance étant donné que la relation les liant à la variable endogène (Y_t) n'est pas linéaire.

Pour cela, définissons la variable dichotomique suivante :

(1 si l'individu «t» appartient à la catégorie Y_{tj} = 1 si l'individu « t » appartient à la classe «j», 0 autrement (La vraisemblance associée à l'échantillon sous main s'écrit alors :

$$L = \Pi_{t=1}^{n} (P_{1t})^{Y1t} . (P_{2t})^{Y2t} \ldots (P_{mt})^{Ymt}$$

et en prenant le logarithme de cette dernière expression, on obtient :

$$Log(L) = \sum_t^n \sum_{i=1}^{m} Y_{it} . Log(P_{it})$$

Dans le cas particulier où F est une fonction **logistique,** le modèle précédent est appelé alors modèle logistique multinomial. L'expression de G s'écrit alors :

$$G(Xt'\beta_j) = e^{Xt'b_j}$$

et par conséquent celle de Pjt se simplifie pour s'écrire :

$$^{Xt'\beta_i} \; / \; (1 + \sum_{i=1}^{m-1} e^{Xt'\beta_i})$$ ce qui donne **:**

$$Log(L) = \sum_t^n (\sum_{i=1}^{m-1} Y_{it} . Log(P_{it}) + Y_{mt} P_m)$$

$$Log(L) = \sum_t^n (\sum_{i=1}^{m-1} (Y_{it} . Log(P_{it})) + Y_{mt} P_m)$$

$$Log(L) = \sum_t^n (\sum_{i=1}^{m-1} (x_t'\beta_i) - \sum_{i=1}^{m} Y_{jt} (Log(1 + \sum_{i=1}^{m-1} e^{x_t'\beta_i})))$$

La condition nécessaire pour que b1 , b2 , … , bm-1 maximisent Log(L) est que :

$$\partial Log(L)/\partial\beta j = \sum_{t=1}^{n} (Yjt - Pjt) Xt' = 0 \qquad j=1, 2, \ldots, m-1$$

et où 0 représente un vecteur nul de dimension (K*1)
De la même manière, on peut obtenir l'expression des dérivées secondes nécessaires à l'application d'un des algorithmes de recherche d'un optimum (Newton-Raphson, méthode du score, etc.) :

$$\partial 2 Log(L)/\partial\beta j \partial\beta j = -\sum tn \; Pjt \,(1 - Pjt) Xt' . Xt \qquad j=1, \ldots, m-1$$

et pour $j \neq i$:

$$\partial 2 Log(L)/\partial\beta j \partial\beta i = -\sum tn \; Pjt . Pit \; Xt' . Xt \qquad j=1, \ldots, m-1 \quad i=1, \ldots, m-1$$

3.4 – Application

Les résultats à l'examen du baccalauréat pour la session de juin 1995 en ce qui concerne les candidats des établissements retenus présentent la particularité d'être fortement concentrés sur l'intervalle (10.00, 12.00) ce qui nous oblige à adapter les modalités de la variable endogène à cette situation particulière de la manière suivante : Yt prend l'une des modalités suivante :

0 si le candidat N°(t) n'est pas reçu au bac
1 si le score réalisé au bac se situe entre 10.00 et 10.50 sur 20
2 si le score réalisé au bac se situe entre 10.50 et 11.00 sur 20
3 si le score réalisé au bac est supérieur à 11.00 sur 20

Les effets marginaux des variables explicatives Xkt se mesurent, à travers l'expression suivante :

$$\partial Pjt / \partial Xk = Pjt\ (\beta jk - \sum i=1m-1 \beta ikPit) \qquad j=1,2,\ldots,m-1=1,2,\ldots,K$$

et dépendent, comme on peut le constater, de tous les paramètres bik à la fois.

Le tableau IX ci-dessous donne le résultat concernant de l'estimation de ces paramètres pour le cas de la série SNV. La principale conclusion est que la probabilité de réaliser un score supérieur (moyenne générale se situant entre 10.50 et 11 sur 20) à l'examen du baccalauréat est fortement liée à la performance réalisée par l'élève en dernière année de lycée (Moy-3AS)

Tableau I – Tests de spécification du modèle de réussite Au BAC série – SNV-

VARIABLES	F STATISTIQUE	LOG.LIKELIHOOD RATIO
Constante + Moy gen3	_	171.5076*
Constante + Moy gen3 + Moy en	12.9477*	6.8489*
Constante + Moy gen3+ Moy BEF	18.0659*	11.1953
Constante + Moy gen3+ Moy gen1	7.6412*	3.7214**
Constante + Moy gen3+ Moy gen3 Moy gen1+ Moy BEF	6.5111*	11.654*

(*) Pertinence Acceptée de la variable explicative additionnelle au seuil de 5 %.
(**) Acceptée au seuil de 10 %.

Tableau II – Tests de spécification du modèle de réussite au bac série L.S.H

VARIABLE	F. Statistique	Log Likelihood Ratio
Constante + Moy gen3	_	60.7045
Constante + Moy gen3+ Moy gen	1.7822	1.0996
Constante + Moy gen3+ Moy BEF	0.9487	1.1084
Constante + Moy gen 3+ Moy gen 1	-0.0420	0.0052
Constante + Moy gen 3+ Moy gen + Moy BEF + Moy gen 1	0.7869	1.2197

* variable explicative additionnelle acceptée au seuil de 5 %.

Tableau III – Estimation du modèle de réussite au bac série –SNV-

VARIABLES EXPLICATIVE	MODELE A	MODELE B	MODELE C	MODELE D
Constante	-5.01* (0.44)	-5.96* (0.57)	-4.95* (0.66)	-5.71* (0.74)
Moy gen 3	0.46* (0.04)	0.42* (0.05)	0.42* (0.05)	0.52* (0.06)
Moy gen	-	0.03 (0.05)	0.02 (0.06)	-0.04 (0.06)
Moy BEF	-	0.11* (0.05)	0.11* (0.05)	0.03 (0.06)
Moy gen 1	-	-0.01 (0.06)	-0.01 (0.06)	0.05 (0.07)
Sexe	-	-	-0.28** (0.15)	-0.34* (0.16)
Sit 3AS	-	-	-0.24* (0.10)	-0.25* (0.11)
Lycée 1	-	-	-	1.52* (0.45)
Lycée 2	-	-	-	1.44* (0.46)
Lycée 5	-	-	-	-0.16 (0.47)
Lycée 6	-	-	-	1.49* (0.60)
Log Likelih ood	-200.09	-194.29	-190.04	- 165.65

Tableau IV – Estimation du modèle de réussite au bac série -L.S.H-

VARIABLE EXPLICATIVE	MODELE A	MODELE B	MODELE C
Constante	-2.79* (0.44)	-2.74* (0.63)	-3.25* (0.58)
Moy gen 3	0.26* (0.04)	0.28* (0.05)	0.28 (0.04)
Sexe	-	-0.26 (0.25)	-
Sit 3AS	-	0.14 (0.13)	-
Lycée 1	-	-	0.36 (0.34)
Lycée 2	-	-	0.31 (0.34)
Lycée 4	-	-	-0.12 (0.44)

Test de pertinence des variables additionnelles :

F.Test	1.20	1.28
Log Likelihood	3.03	2.45

* Acceptée à 5 %

Tableau V – Effet «collège» d'origine sur la réussite au bac série –SNV-

VARIABLE	TOUS LYCEES CONFONDUS A	LYCEE B	LYCEE C	LYCEE D
Constante	-5.38* (0.76)	-7.94* (1.60)	-2.19 (1.94)	-9.13* (2.22)
Moy gen	0.00 (0.65)	-0.08 (0.12)	0.01 (0.17)	-0.05 (0.15)
Moy BEF	-0.04 (0.06)	0.22* (0.11)	-0.31** (0.17)	0.07 (0.16)
Moy gen 1	0.06 (0.07)	0.25* (0.13)	0.13 (0.13)	-0.15 (0.21)
Moy gen 3	0.46* (0.07)	0.54* (0.12)	0.43* (0.11)	0.91* (0.22)
sexe	0.35* (0.17)	-0.74* (0.29)	-0.08 (0.35)	-0.05 (0.43)
Sit 3 AS	-0.18 (0.13)	-0.24 (0.25)	-0.06 (0.45)	0.18 (0.31)
Col 3	-0.30 (0.51)	-	-	-
Col 8	0.61* (0.35)	0.89** (0.48)	-	-
Col 9	0.58** (0.34)	1.21* (0.61)	-0.28 (0.64)	-
Col 10	0.27 (0.45)	0.53 (0.71)	-0.45 (0.69)	-
Col 11	0.88* (0.35)	-	-0.05 (0.51)	-
Col 15	-0.65* (0.32)	-	-	-0.19 (0.57)
Col 19	0.31 (0.30)	0.86 (0.58)	-0.27 (0.48)	-
Col 21	0.47 (0.32)	0.31 (0.43)	-0.57 (0.66)	-
Col 22	0.19 (0.34)	0.15 (0.52)	-	-
Col 23	0.41 (0.43)	0.51 (0.55)	-	-
Col 26	-1.25* (0.62)	-	-	-0.71 (0.80)
Col 27	-0.63** (0.34)	-	-	-0.24 (0.61)
Col 44	0.55 (0.62)	-	-	-
Col 45	-0.32 (0.54)	-	-	-
Col 49	-	0.72 (0.80)	-	-
Log(Likelihood)	71.85 -	-105.61 -	-44.5522 -	- 28.1203
Nb. D(observations)	463	185	99	105
Reçus au bac	143	48	44	17

Tableau VI – Estimation des modèles avec effet «collège d'origine» sur la réussite au bac série –L.S.H-

VARIABLE	MODELE GENERAL (A)	LYCEE I (B)	LYCEE II (C)
Constante	-5.89 (1.26)*	-5.47 (1.98)*	-7.63 (3.61)*
Moy gen	0.02 (0.09)	-0.11 (0.14)	0.56 (0.30)*
Moy BEF	0.002 (0.09)	0.07 (0.10)	-0.48 (0.30)
Moy gen 1	0.06 (0.11)	0.31 (0.23)	-0.12 (0.23)
Moy gen 3	0.43 (0.08)*	0.29 (0.12)*	0.57 (0.25)*
Sexe	-0.14 (0.30)	-0.30 (0.44)	0.48 (0.75)
Sit 3 AS	0.44 (0.16)*	0.48 (0.35)	-0.13 (0.83)
College 8	0.33 (0.60)	-0.08 (0.63)	-
College 9	-1.26 (0.69)**	-1.35 (0.75)**	-
College 10	1.00 (0.86)	-	1.01 (1.22)
College 11	-0.33 (0.66)	-	-8.84 (23.3)
College 19	0.59 (0.46)	-	0.85 (0.79)
College 21	-0.39 (0.73)	-0.97 (0.78)	-
College 22	0.47 (0.51)	-	0.06 (0.25)
College 23	0.47 (0.43)	-0.08 (0.48)	-
College 38	0.78 (0.58)	0.33 (0.66)	-

Log Lilihood 68.77 - 33.13 - 12.82
NB OBSERVATION 180 69 26
NB REUSSITE AU BAC 53 21 16

Tableau VII – Effet «classe d'appartenance» en terminale sur la réussite au bac série –SNV-

VARIABLE	LYCEE I	LYCEE II
CONSTANTE	-4.96* (124)	-2.73 (1.80)
Moy gen	0.13 (0.10)	0.15(0.15)
Moy BEF	0.02 (0.09)	-0.34* (0.15)
Moy gen1	0.06 (0.10)	0.03 (0.13)
Moy gen3	0.34* (0.09)	0.40* (0.10)
Sexe	-0.63* (0.24)	0.16 (0.43)
Sit 3 AS	0.09 (0.21)	-0.16 (0.38)
Classe 2	-0.61* (0.30)	0.62 (0.46)
Classe 3	-0.81* (0.37)	-0.48 (048)
Classe4	-0.16 (0.35)	-0.25 (0.48)
Classe5	-	-0.02 (0.47)

Log Likelihood - 79.3114 - 46.3969
Nb d'observations 184 98
Nb réussites au bac 51 49

Tableau VIII – Effet «classe d'appartenance» en terminale sur la réussite au bac série -L.S.H -

VARIABLE	LYCEE I	LYCEE II
Constante	-5.67* (1.15)	-2.72 (2.23)
Moy gen	0.02 (0.13)	0.33*** (0.23)
Moy BEF	0.02 (0.10)	-0.19 (0.22)
Moy gen 1	0.49* (0.22)	0.15 (0.16)
Moy gen 3	0.21** (0.11)	0.27** (0.15)
Sexe	-0.77** (0.42)	0.15 (0.68)
Sit 3 AS	0.26 (0.31)	-0.25 (0.57)
Classe 2	-0.53 (0.46)	-0.84 (0.68)
Classe 3	-0.22 (0.41)	-0.48 (064)
Classe 4	-	-1.06 (0.75)
Log Likelihood -	-35.43	22.16
Nb d'observations	67	26

Tableau IX – Résultat à l'examen du bac avec niveau de performance – Série SNV-

VARIABLE	P_1	P_2	P_3
Constante	6.71* (1.42)	5.81* (1.99)	3.20* (9.88)
Moy-3AS	0.88 (9.98)	0.007 (0.15)	0.40* (0.20)
Moy-BEF	0.20* (0.08)	0.13* (0.16)	-0.37* (0.15)
Sexe	-1.65* (0.36)	-2.65* (0.69)	2.79* (0.52)
Sit-3AS	-0.68 (1.10)	-1.44 (1.53)	-0.60 (1.82)
Lycée1	1.83* (0.49)	0.26* (0.97)	3.59* (1.15)
Lycée2	0.17 (0.55)	-0.55 (1.28)	21.49 (1.04 E-10)
Lycée3	0.0002 (0.56)	-0.07 (1.14)	0.36 (1.26)

Log Likelihood = -254.04

Annexe :

Des variables pédagogiques

Annexe : Code matière

I – Code matière 9ème année fondamentale

Techno	Technologie
Sciences	Sciences Naturelles
Lan – Arab	Langue Arabe
Lan – Etr 1	Langue Etrangère 1
Lan – Etr 2	Langue Etrangère 2
Edu – Socio	Education Socio – économique
Edu – Rel	Education Religieuse
His - Géo	Histoire Géographie
Moy – Gen	Moyenne Générale En 9ème AF
Moy – Bef	Moyenne Au BEF

II – Code matière $1^{ère}$ année secondaire (Troncs communs : technologie, sciences et sciences humaines)

La-Arab3	Langue Arabe En $3^{ème}$ AS
Math3	Mathématiques En $3^{ème}$ AS
His-Géo3	Histoire-Géographie En $3^{ème}$ AS
Sci-Rel	Sciences Islamiques En $3^{ème}$ AS
Philo	Philosophie En $3^{ème}$ AS
La-Etr13	Langue Etrangère (1) En $3^{ème}$ AS
La-Etr23	Langue Etrangère (2) En $3^{ème}$ AS
La-Etr33	Langue Etrangère (3) En $3^{ème}$ AS
Sci-Nat	Sciences Naturelles En $3^{ème}$ AS
Physiq3	Sciences Physiques En $3^{ème}$ AS
Act-Cul3	Activités Culturelles En $3^{ème}$ AS
Eps3	Education Physique En $3^{ème}$ AS
Moy-Gen3	Moyenne Générale Annuelle En $3^{ème}$ AS

III – Code matière $3^{ème}$ année secondaire

La-Arab 1	Langue arabe $1^{ère}$ AS
Maths 1	Mathématiques $1^{ère}$ AS
His- geo1	Histoire Géographie 1 ère AS
Edu-civ1	Education civique $1^{ère}$ AS
Sci-rell	Sciences Islamiques $1^{ère}$ AS
Des-tec1	Dessin technique $1^{ère}$ AS
Sci-nat1	Sciences naturelles $1^{ère}$ AS
Physiq1	Sciences physiques $1^{ère}$ AS
La-etr11	Langue Etrangère (1) $1^{ère}$ AS
La-etr21	Langue Etrangère (2) $1^{ère}$ AS
Act-cul	Activités culturelles $1^{ère}$ AS
Moy-gen1	Moyenne Générale Annuelle en $1^{ère}$ AS.

$$Sexe = \begin{cases} 1 \text{ si l'élève est de sexe masculin} \\ 0 \text{ autrement} \end{cases}$$

Vœu =

$$\begin{cases} 1 \text{ si le choix du tronc commun à l'issu de la 9ème année fondamentale} \\ \text{est respectée par la commission d'orientation} \\ 0 \text{ autrement} \end{cases}$$

$$\text{Sit\_3AS} = \begin{cases} 1 \text{ si l'élève n'est ni redoublant ni en classe spécial} \\ 0 \text{ autrement} \end{cases}$$

Notes

[*] Chercheur associé au CREAD