

La réussite universitaire : caractéristiques de la population de première année de bachelier aux FUCaM et étude des facteurs explicatifs

Promoteur : Prof. Franz Jomaux
Co-promotrice : Prof. Bernadette Noël

Mémoire présenté par :
Virginie VANDENBULCKE
En vue de l'obtention du diplôme
de Master en Ingéniorat de Gestion

Année académique 2010-2011

Remerciements

Je tiens à remercier tout particulièrement les Professeurs Franz Jomaux et Bernadette Noël, promoteur et co-promotrice de ce travail de fin d'étude, pour leur suivi, leurs conseils éclairés ainsi que pour leur constante disponibilité et leur patience.

Un merci est également adressé à Mademoiselle A. Schatteman pour son aide, son soutien et pour l'intérêt qu'elle a porté à la réalisation de ce travail ainsi qu'à Mademoiselle C. Ducarroz pour ses encouragements.

Je voudrais également remercier Monsieur D. Franzini, Mesdames S. Goemare, R. Testa et A-S Hazebroucq, pour leur collaboration dans la récolte des informations qui m'ont permis de construire ma base de données et aussi Mesdames N. Meskens et E. Horge pour leurs conseils.

Enfin, je tiens à remercier mes parents, mes sœurs, ma grand-mère, Timothée, Guillaume et mes proches pour leur soutien durant ces cinq années d'études.

Résumé

Aujourd'hui, réussir à l'université peut sembler un réel défi. En effet, les taux de réussite sont très bas : seuls 40% des inscrits réussissent leur première année universitaire. Cette problématique a d'ailleurs des conséquences financières, psychologiques, sociales pour l'étudiant, ses parents et la société. Elle entraîne donc un réel intérêt de recherches.

Sur base de ce constat, nous avons réalisé des tests pour tenter d'identifier des facteurs potentiellement explicatifs de la réussite ou de l'échec. Pour cela, un aperçu de la littérature sur ce thème a été réalisé puis, les actions déjà menées au sein des Facultés Universitaires Catholiques de Mons (FUCaM) ont été recensées. Ensuite, une collecte des données concernant les étudiants de BAC1 des FUCaM pour les années 2008-2009 et 2009-2010 a été effectuée en vue de la réalisation de tests et régressions statistiques.

Ainsi, les résultats suite aux tests de comparaisons de moyennes, des régressions linéaires, tests du Chi-carré et régressions logistiques semblent montrer que les facteurs les plus significatifs de la réussite sont : le nombre d'heures de mathématique suivies les deux dernières années du secondaire, la participation aux cours d'été, les résultats au passeport mathématique organisé en début de BAC1, le pourcentage obtenu en fin de secondaire et le fait d'avoir doublé durant les études secondaires. D'autres facteurs tels que le genre de la personne ou son âge interviennent aussi mais de façon moins marquée dans les résultats.

Cependant, il faut nuancer ces conclusions car d'autres critères pourraient être ajoutés et étudiés et la base de données utilisée pourrait être étoffée. Ainsi, de nouvelles pistes d'aides peuvent être envisagées pour élargir le panel de solutions déjà proposées.

Table des matières

Remerciements.....	1
Résumé.....	2
Table des matières.....	3
Liste des figures	6
Liste des tableaux	7
Introduction générale.....	9
Premier chapitre : Aperçu de la littérature sur le thème de la réussite universitaire	11
Section 1 : Introduction	11
Section 2 : Les facteurs liés à l'échec ou à la réussite à l'université.....	13
1. <i>Le profil de l'étudiant.....</i>	13
1.1 Le milieu social et économique.....	13
1.2 Le diplôme des parents	14
1.3 Le genre	15
1.4 L'âge.....	15
1.5 Le lieu de résidence lors des études.....	16
2. <i>Le parcours scolaire de l'étudiant</i>	16
2.1 Le parcours suivi avant l'entrée à l'université.....	16
2.2 Le passage vers l'université	18
3. <i>La difficulté du choix d'orientation</i>	19
3.1 L'orientation	19
3.2 La réorientation	21
4. <i>La perception et la motivation de l'étudiant</i>	22
5. <i>L'intégration sociale et académique de l'étudiant.....</i>	23
6. <i>Les cours d'été.....</i>	23
7. <i>Les pratiques d'enseignement et leur efficacité.....</i>	24
8. <i>Les tests formatifs</i>	25
9. <i>Les monitorats</i>	25
10. <i>La réussite en janvier.....</i>	26
12. <i>La déclaration de Bologne : une nouvelle organisation</i>	27
Section 3 : Conclusion.....	27
Deuxième chapitre : Projets mis en place au sein des FUCaM.....	29
Section 1 : Introduction	29
Section 2 : Les projets mis en œuvre aux FUCaM	30

1. Les matinées portes ouvertes et salons SIEP.....	30
2. Le projet boussole	31
3. Les cours d'été (cours propédeutiques).....	31
4. Les tests organisés en début d'année académique.....	32
5. Les tests formatifs de novembre.....	32
6. Les monitorats	33
7. Les entretiens individuels	34
Section 3 : Conclusion	34
Troisième chapitre : Modèles, méthodes et tests statistiques	35
Section 1 : Introduction	35
Section 2 : Rappels de quelques concepts de base	36
1. La nature des variables	36
2. Les hypothèses	37
3. Le risque d'erreur.....	38
4. La significativité (Sig.) ou p-value et les tests d'hypothèses.....	39
Section 3 : Modèles, méthodes et tests statistiques	40
1. Modèles étudiant le lien avec une variable à expliquer quantitative.....	41
1.1 Variables explicatives binaires : comparaisons de moyennes	41
1.2 Variables explicatives quantitatives : corrélations linéaires.....	51
1.3 Variables explicatives qualitatives et quantitatives : régressions linéaires.....	52
2. Modèles étudiant le lien avec une variable à expliquer qualitative	62
2.1 Variables explicatives qualitatives : test du Chi-carré de Pearson	62
2.2 Variables explicatives quantitatives : comparaisons de moyennes.....	65
2.3 Variables explicatives qualitatives et quantitatives : régressions logistiques	66
Section 4 : Conclusion	77
Quatrième Chapitre : Présentation des données	78
Section 1 : Construction de la base de données	78
1. Profil de l'étudiant	79
2. Parcours scolaire de l'étudiant avant l'entrée à l'université.....	80
3. Parcours scolaire de l'étudiant universitaire.....	82
Section 2 : Présentation descriptive des données.....	83
1. Pour les sciences et ingénierat de gestion (Orient = 1)	84
1.1 Nombre de filles et de garçons	84
1.2 Age	84
1.3 Bisseur	85
1.4 Cours d'été.....	85
1.5 Heures de mathématique	86
2. Pour les sciences humaines, sciences politiques et information et communication (Orient = 2)	86
2.1 Nombre de filles et de garçons	86

2.2	Age	87
2.3	Bisseur	87
2.4	Cours d'été.....	88
Cinquième chapitre : Résultats des analyses statistiques		89
Section 1 : Pour les sciences et ingénierat de gestion (Orient = 1)		89
1.	<i>Facteurs influençant la moyenne globale obtenue en fin de BAC1 (Moyenne_global_100)</i>	<i>89</i>
1.1	Comparaison des moyennes	89
1.2	Corrélation linéaire	99
1.3	Régression linéaire multiple : méthode « pas à pas »	100
2.	<i>Facteurs influençant la réussite ou non de la première année de bachelier (Reussite_annee).....</i>	<i>103</i>
2.1	Test du Chi-carré de Pearson.....	103
2.2	Régression logistique	109
Section 2 : Pour les sciences humaines, sciences politiques et information et communication (Orient = 2)		113
1.	<i>Facteurs influençant la moyenne globale obtenue en fin de BAC1 (Moyenne_global_100)</i>	<i>113</i>
1.1	Comparaison des moyennes	113
1.2	Corrélation linéaire	114
1.3	Régression linéaire multiple	115
2.	<i>Facteurs influençant la réussite ou non de la première année de bachelier (Reussite_annee).....</i>	<i>117</i>
2.1	Test du Chi-carré de Pearson.....	117
2.2	Régression logistique	118
Section 3 : Conclusion des analyses statistiques		120
Pistes et actions concrètes proposées comme solutions à la problématique		124
Conclusion générale		127
Perspectives		131
Références bibliographiques		132

Liste des figures

FIGURE 1 : GRAPHIQUE REPRESENTANT LES MOYENNES DES NOTES DE LA SESSION DE JANVIER (SUR 20) ET LES GRADES OBTENUS FIN D'ANNEE PAR LES ETUDIANTS DES FUCAM.....	26
FIGURE 2 : REPRESENTATION SCHEMATIQUE DU CLASSEMENT DES VARIABLES.....	37
FIGURE 3 : GRAPHIQUE SPSS REPRESENTANT LES VARIABLES MOYENNE_GLOBAL_100 ET PASSEPORT_MOY_20 POUR L'ANNEE 2008-2009 ET POUR ORIENT = 1.....	51
FIGURE 4 : GRAPHIQUE DE LA PROPORTION D'ETUDIANTS QUI ONT REUSSI LEUR BAC1 EN FONCTION DES POINTS AU PASSEPORT MATHEMATIQUE	67
FIGURE 5 : GRAPHIQUE D'UNE COURBE EN « S »	67
FIGURE 6 : EVOLUTION DU FACTEUR ORIENTATION.....	83
FIGURE 7 : EVOLUTION DU FACTEUR SEXE	84
FIGURE 8 : EVOLUTION DU FACTEUR AGE.....	84
FIGURE 9 : EVOLUTION DU FACTEUR BISSEUR	85
FIGURE 10 : EVOLUTION DU FACTEUR COURS D'ETE	85
FIGURE 11 : EVOLUTION DU FACTEUR HEURES DE MATHEMATIQUE.....	86
FIGURE 12 : EVOLUTION DU FACTEUR SEXE	86
FIGURE 13 : EVOLUTION DU FACTEUR AGE.....	87
FIGURE 14 : EVOLUTION DU FACTEUR BISSEUR	87
FIGURE 15 : EVOLUTION DU FACTEUR COURS D'ETE	88

Liste des tableaux

TABEAU 1 : TAUX DE REUSSITE TOTAL ET PAR REGROUPEMENT D'ORIENTATIONS POUR LES ANNEES 2008-2009 ET 2009-2010 DES ETUDIANTS DES FUCAM	29
TABEAU 2 : TABLEAU DES TYPES D'ERREURS	38
TABEAU 3 : TABLEAU SPSS – COEFFICIENTS DE LA REGRESSION LINEAIRE.....	39
TABEAU 4 : TABLEAU SPSS – STATISTIQUES DESCRIPTIVES POUR LA VARIABLE MOYENNE_GLOBAL_100	42
TABEAU 5 : TABLEAU SPSS – TESTS D'HOMOGENEITE DES VARIANCES ET D'EGALITE DES MOYENNES POUR LA VARIABLE MOYENNE_GLOBAL_100.....	43
TABEAU 6 : TABLEAU SPSS – ANALYSE DESCRIPTIVE DE LA VARIABLE MOYENNE_GLOBAL_100 REPARTIE EN 3 SOUS-GROUPES	47
TABEAU 7 : TABLEAU SPSS – TEST D'HOMOGENEITE DES VARIANCES POUR LA VARIABLE MOYENNE_GLOBAL_100.....	48
TABEAU 8 : TABLEAU SPSS – ANALYSE ANOVA POUR LA VARIABLE MOYENNE_GLOBAL_100	48
TABEAU 9 : TABLEAU SPSS – COMPARAISONS MULTIPLES DE MOYENNES POUR LA VARIABLE MOYENNE_GLOBAL_100.....	50
TABEAU 10 : TABLEAU SPSS – TABLEAU DE CORRELATIONS ENTRE LES VARIABLES MOYENNE_GLOBAL_100 ET PASSEPORT_MOY_20	52
TABEAU 11 : TABLEAU SPSS – COEFFICIENTS DE LA REGRESSION LINEAIRE.....	54
TABEAU 12 : TABLEAU SPSS – ANALYSE DESCRIPTIVE POUR LA REGRESSION LINEAIRE	55
TABEAU 13 : TABLEAU SPSS – COEFFICIENTS DE CORRELATION POUR LA REGRESSION LINEAIRE SIMPLE	57
TABEAU 14 : TABLEAU SPSS – TABLEAU ANOVA DE LA REGRESSION LINEAIRE SIMPLE	58
TABEAU 15 : TABLEAU SPSS – REGRESSION LINEAIRE MULTIPLE	60
TABEAU 16 : TABLEAU SPSS – TABLEAU DES COEFFICIENTS DE CORRELATION POUR LA REGRESSION LINEAIRE MULTIPLE	61
TABEAU 17 : TABLEAU SPSS – TABLEAU ANOVA POUR LA REGRESSION LINEAIRE MULTIPLE	61
TABEAU 18 : TABLEAU SPSS – TABLEAU DES EFFECTIFS OBSERVES D'UN TEST DU CHI-CARRE DE PEARSON POUR LES VARIABLES REUSSITE_ANNEE ET MATH6H.....	62
TABEAU 19 : TABLEAU SPSS – TABLEAU DES EFFECTIFS THEORIQUES D'UN TEST DU CHI-CARRE DE PEARSON POUR LES VARIABLES REUSSITE_ANNEE ET MATH6H	63
TABEAU 20 : TABLEAU SPSS – TEST DU CHI-CARRE DE PEARSON POUR LES VARIABLES REUSSITE_ANNEE ET MATH6H	63
TABEAU 21 : TABLEAU SPSS – COEFFICIENTS DES VARIABLES DE LA REGRESSION LOGISTIQUE SIMPLE.....	70
TABEAU 22 : TABLEAU SPSS – TABLEAU RESUME POUR LE TEST DU RAPPORT DE VRAISEMBLANCE POUR LA REGRESSION LOGISTIQUE SIMPLE.....	71
TABEAU 23 : TABLEAU SPSS – TEST DU RAPPORT DE VRAISEMBLANCE POUR LA REGRESSION LOGISTIQUE SIMPLE	72
TABEAU 24 : TABLEAU SPSS - TEST DU RAPPORT DE VRAISEMBLANCE POUR LA REGRESSION LOGISTIQUE SIMPLE	72
TABEAU 25 : TABLEAU SPSS – VRAISEMBLANCE	73
TABEAU 26 : TABLEAU SPSS – R ² DE COX&SNELL.....	73
TABEAU 27 : TABLEAU SPSS – TABLEAU DE CLASSIFICATION POUR LA REGRESSION LOGISTIQUE SIMPLE.....	74
TABEAU 28 : TABLEAU SPSS – COEFFICIENTS DE LA REGRESSION LOGISTIQUE MULTIPLE.....	76
TABEAU 29 : TABLEAU SPSS – TABLEAU RESUME DE LA REGRESSION LOGISTIQUE MULTIPLE	77
TABEAU 30 : TABLEAU DE PRESENTATION DES DONNEES – PROFIL DE L'ETUDIANT	80
TABEAU 31 : TABLEAU DE PRESENTATION DES DONNEES – PARCOURS SCOLAIRE DE L'ETUDIANT AVANT SON ENTREE A L'UNIVERSITE	81
TABEAU 32 : TABLEAU DE PRESENTATION DES DONNEES – PARCOURS SCOLAIRE DE L'ETUDIANT UNIVERSITAIRE	83
TABEAU 33 : TABLEAU SPSS – STATISTIQUES DESCRIPTIVES DE LA VARIABLE MOYENNE_GLOBAL_100 POUR L'ANNEE 2008-2009.....	90
TABEAU 34 : TABLEAU SPSS - TEST D'INDEPENDANCE DES MOYENNES POUR LA VARIABLE MOYENNE_GLOBAL_100 POUR L'ANNEE 2008-2009	90
TABEAU 35 : TABLEAU RECAPITULATIF DES FACTEURS INFLUENÇANT LA MOYENNE GLOBALE EN FIN DE BAC1 POUR LES ANNEES 2008-2009 ET 2009-2010	98
TABEAU 36 : TABLEAU SPSS – TABLEAU DE CORRELATIONS LINEAIRES POUR L'ANNEE 2008-2009	99

TABLEAU 37 : TABLEAU SPSS – TABLEAU DE CORRELATIONS LINEAIRES POUR L’ANNEE 2009-2010	100
TABLEAU 38 : TABLEAU SPSS – TABLEAU DES COEFFICIENTS POUR LA REGRESSION LINEAIRE MULTIPLE	101
TABLEAU 39 : TABLEAU SPSS – TABLEAU DE RESUME DU MODELE POUR LA REGRESSION LINEAIRE MULTIPLE .	102
TABLEAU 40 : TABLEAU SPSS – TABLEAU DES COEFFICIENTS POUR LA REGRESSION LINEAIRE MULTIPLE	102
TABLEAU 41 : TABLEAU SPSS – TABLEAU CROISE DES EFFECTIFS OBSERVES POUR LES VARIABLES REUSSITE_ANNEE ET SEXE	104
TABLEAU 42 : TABLEAU SPSS – TEST DU CHI-CARRE DE PEARSON	104
TABLEAU 43 : TABLEAU RECAPITULATIF – TEST DU CHI-CARRE DE PEARSON	109
TABLEAU 44 : TABLEAU SPSS – COEFFICIENTS DE LA REGRESSION LOGISTIQUE	111
TABLEAU 45 : TABLEAU SPSS – TABLEAU RESUME DU MODELE POUR LA REGRESSION LOGISTIQUE	111
TABLEAU 46 : TABLEAU SPSS – COEFFICIENTS DE LA REGRESSION LOGISTIQUE	112
TABLEAU 47 : TABLEAU SPSS – TABLEAU RECAPITULATIF DES COMPARAISONS DE MOYENNES	114
TABLEAU 48 : TABLEAU SPSS – TABLEAU DE CORRELATION LINEAIRE	114
TABLEAU 49 : TABLEAU SPSS – TABLEAU DE CORRELATION LINEAIRE	115
TABLEAU 50 : TABLEAU SPSS – TABLEAU DES COEFFICIENTS POUR LA REGRESSION LINEAIRE MULTIPLE	116
TABLEAU 51 : TABLEAU SPSS – TABLEAU DES COEFFICIENTS POUR LA REGRESSION LINEAIRE MULTIPLE	116
TABLEAU 52 : TABLEAU SPSS – TABLEAU RECAPITULATIF DU MODELE POUR LA REGRESSION LINEAIRE MULTIPLE	117
TABLEAU 53 : TABLEAU RECAPITULATIF DU TEST DU CHI-CARRE DE PEARSON	118
TABLEAU 54 : TABLEAU SPSS – COEFFICIENTS DE LA REGRESSION LOGISTIQUE	118
TABLEAU 55 : TABLEAU SPSS - COEFFICIENTS DE LA REGRESSION LOGISTIQUE	119
TABLEAU 56 : TABLEAU RECAPITULATIF DES FACTEURS EXPLICATIFS PAR TEST UTILISE	121



Introduction générale

Aujourd'hui, de nombreuses recherches s'intéressent à l'enseignement, l'éducation et la formation. Parmi celles-ci, certaines étudient plus particulièrement la réussite universitaire. En effet, le taux de réussite des études supérieures est relativement bas. Ainsi, en Belgique, ce taux est seulement de 40% pour les universités wallonnes.

Cette constatation du taux d'échec élevé a amené les chercheurs à réfléchir aux causes et aux facteurs explicatifs de celui-ci. En effet, cette problématique concerne de nombreux étudiants, enseignants, parents d'étudiants et futures entreprises à la recherche d'employés. Elle a aussi diverses conséquences pour la personne et pour la société : financières, sociales, psychologiques...

Ce travail de fin d'étude est dans la continuité de recherches qui ont déjà été menées dans ce domaine et plus particulièrement à la suite du mémoire d'A. Schatteman (2009) : « La réussite universitaire : Etude de la liaison entre les cotes du passeport mathématique et les résultats finaux des étudiants de première bachelier ».

L'objectif de ce travail de fin d'étude est donc d'identifier les critères significatifs qui influencent la réussite universitaire dans le but d'amener des pistes de réflexion et d'actions pour améliorer celle-ci.

Dans le *premier chapitre* de ce mémoire est présenté un aperçu de la littérature sur des facteurs liés à la réussite. Ceux-ci sont repris en diverses catégories qui vont du profil de l'étudiant (milieu socio-économique, diplôme des parents, âge...) à son parcours scolaire (redoublement, option...) en citant aussi la difficulté du choix d'orientation, la motivation, l'intégration de l'étudiant dans son nouvel environnement...

Le *deuxième chapitre* énumère les projets et actions déjà mis en place au sein des FUCaM pour soutenir et encourager l'étudiant vers la réussite de sa première année de bachelier.



Le *troisième chapitre* est consacré à l'étude des modèles, méthodes et tests statistiques réalisés par la suite sur les bases de données construites. Après un rappel des concepts de base, l'explication didactique des modèles et théories permet au lecteur de comprendre les développements et les analyses des tableaux fournis par le logiciel SPSS. Cette partie fait une distinction entre les modèles où la variable à expliquer est quantitative et ceux où celle-ci est qualitative.

Le *quatrième chapitre* présente les deux bases de données pour les années 2008-2009 et 2009-2010 des étudiants de BAC1 aux FUCaM. Celles-ci sont divisées en trois catégories : le profil de l'étudiant, son parcours scolaire avant son entrée à l'université et son parcours universitaire. Par la suite, une partie consacrée aux statistiques descriptives offre une vision de l'évolution des critères sur deux ou trois ans.

Le *cinquième chapitre* s'intéresse aux résultats des analyses statistiques. Ceux-ci sont énoncés séparément pour l'orientation sciences de gestion et ingénierat de gestion et pour les orientations regroupées de sciences politiques, sciences humaines et information et communication.

Finalement, des pistes de solutions et d'actions sont proposées pour soutenir, aider et encourager l'étudiant vers la réussite de sa première année de bachelier.



Premier chapitre : Aperçu de la littérature sur le thème de la réussite universitaire

La première partie de ce travail de fin d'étude est consacrée aux travaux et recherches existants dans la littérature sur le thème de la réussite universitaire. Elle consistera d'abord en une brève introduction du contexte, suivie d'une présentation de facteurs liés à l'échec ou à la réussite qui permettront ainsi au lecteur de mieux les appréhender.

Section 1 : Introduction

Depuis les années 90, les universités connaissent une augmentation des effectifs étudiants mais aussi des changements organisationnels comme ceux amenés par la déclaration de Bologne (Parmentier, 2011). Cette dernière vise à harmoniser le système d'enseignement supérieur européen avec, par exemple, la mise en place d'un système commun de crédits pour décrire les programmes d'études. Cependant, alors que le nombre d'étudiants a augmenté et que la population étudiante s'est diversifiée (origines sociale, économique, culturelle...), nous pouvons remarquer une stabilité dans le taux d'échec élevé (égal à environ 60%) depuis ces trente dernières années (Nils & Lambert, 2011, p.11). Dès lors, une question fondamentale émerge : comment comprendre et appréhender l'échec à l'université pour mieux le prévenir ? (Galand & al., 2005).

Selon Leclercq et Parmentier (2011, p.6), il est facile de définir la réussite comme « l'autorisation de passage à l'année suivante » et l'obtention, au final, d'un diplôme. Néanmoins, dans cette définition, les auteurs estiment que l'idée « d'année de transition » secondaire-supérieur que peut représenter la première année universitaire est oubliée. L'échec est, quant à lui, difficile à définir car il peut faire référence à de nombreuses notions. Dès lors, nous pouvons mieux comprendre ce concept au travers de ses diverses conséquences comme un arrêt d'étude et donc une sortie du système scolaire (sans l'obtention d'un diplôme in fine),



un redoublement (et donc le retard futur dans la vie active) ou une réorientation vers d'autres études. « L'échec peut aussi être partiel et impliquer la présentation de plusieurs sessions d'examens » (Droesbeke & Vermandelde, 2005, p.23).

Pour un étudiant, l'échec peut avoir des répercussions aussi bien financières que psychologiques : une remise en question de lui-même, de ses capacités, une démotivation, une perte de confiance en soi, un obstacle à ses aspirations et à certaines professions, une perte de certains privilèges...

Dans une étude réalisée en 2004 par Vandamme, Meskens et Superby (cités par Vandamme & al., 2006, p.43), les variables les plus corrélées avec les performances universitaires sont : « la présence aux cours, les chances de réussite estimées [par l'étudiant lui-même], le bagage scolaire antérieur [principalement les heures de mathématique] et la manière d'étudier ». Ces variables sont donc constituées à la fois de facteurs structurels (des données propres à l'histoire de l'étudiant) et processuels, changeants (par exemple, « l'implication de l'étudiant durant son cursus ou le sentiment d'être soutenu par ses proches ») (Vandamme & al., 2006, p.43).

Le but des recherches présentées ci-après est d'aller plus loin que le simple constat « statistique » de certains facteurs liés à l'échec ou à la réussite. Comme le souligne d'ailleurs Galand, un intérêt particulier des recherches menées dans ce domaine est de susciter des réactions positives et motivées ainsi que des réflexions en termes « d'orientation, de parcours de formation, d'accompagnement et de remédiation, d'évaluations des apprentissages, de formations des enseignants, de luttes contre les inégalités, de modes de financement des universités, d'accès au marché de l'emploi... » (Galand, 2005, p.4). Dans ce travail de fin d'étude, nous essayerons donc, à partir de ce constat, d'enrichir ces recherches.



Section 2 : Les facteurs liés à l'échec ou à la réussite à l'université

1. Le profil de l'étudiant

Avant de décrire les facteurs qui ont trait au profil de l'étudiant, une première question se pose : l'échec ainsi que la réussite seraient-ils liés aux variables économiques et socio-culturelles du milieu familial de l'étudiant ? Plusieurs études empiriques telles que celle de Romainville (Romainville, 2000, cité par Galand, Neuville & Frenay, 2005) mettent en évidence une corrélation entre la réussite de l'étudiant et son milieu socio-économique (qui comprend, par exemple, le dernier diplôme obtenu par ses parents). Les étudiants ne seraient-ils donc pas tous égaux face à la réussite ?

N'y aurait-il donc pas de réelle démocratisation de l'enseignement supérieur ? Prost distingue deux sortes de démocratisation. La première, quantitative, consiste en l'augmentation du nombre d'étudiants inscrits dans le supérieur. La deuxième, qualitative, renvoie, « quant à elle, aux différences d'accès entre les groupes sociaux à un niveau scolaire spécifique » (Prost, cité par Nils & Lambert, 2011, p.10). La première, quantitative, est réelle car le nombre d'inscriptions dans les hautes écoles et les universités ne cesse d'augmenter. Ainsi, plus d'élèves sortant du secondaire se lancent dans des études supérieures. La deuxième démocratisation n'est pas clairement prouvée car les étudiants ne semblent pas tous avoir les mêmes chances d'accès et de réussite à l'université.

1.1 Le milieu social et économique

Certaines études (entre autres, Arulampalam, Hayor & Smith (2004), Galand, Frenay & Bourgeois (2004), Romainville (2002), cités par Galand, 2005, p.6) étudient le milieu socio-économique et tendent à montrer que les étudiants universitaires sont plus souvent issus d'un milieu privilégié. En outre, des recherches, parmi lesquelles celle de Van Campenhoudt & al. (2008), consacrées à l'impact du diplôme des parents et de leur profession sur la réussite



de l'étudiant, affirment que l'université est accessible essentiellement à des groupes sociaux favorisés.

Il est cependant difficile de définir correctement le « milieu socio-économique ». Il faut donc rester conscient que les différentes variables pouvant refléter celui-ci sont nombreuses et parfois difficiles à appréhender. Par exemple, ne prendre en compte que la profession, le salaire ou le diplôme des parents semble restreindre la richesse de ce facteur. D'autres variables pourraient être étudiées telles que le capital culturel des parents, les revenus du ménage, l'ambition et le soutien des parents... (Van Campenhoudt & al., 2008).

Le caractère « économique » du milieu est difficile à définir car, par exemple, les personnes sont peu enclines à divulguer leur salaire ou leurs conditions de vie. Cela est dû aux phénomènes de désirabilité sociale et de tradition européenne concernant certains « tabous » entretenus par rapport aux revenus. Mais, nous pouvons noter que la probabilité pour qu'un jeune s'inscrive à l'université est d'autant plus faible qu'il provient d'une famille à « revenus modestes » (Van Campenhoudt & al., 2008). D'un point de vue « culturel », les résultats vont dans le même sens : plus le jeune est issu d'un milieu faiblement ouvert à la culture, moins il aura de chance d'entamer des études à l'université. (Van Campenhoudt & al., 2008).

1.2 Le diplôme des parents

En ce qui concerne le facteur « dernier diplôme des parents », l'étude « Eurostudent 2000 » (citée par Van Campenhoudt & al., 2008, p.14), montre une surreprésentation de diplômés universitaires parmi les pères des étudiants universitaires. Il est également à noter que plus le diplôme de la mère est élevé plus les chances de réussite universitaire sont grandes. De plus, la réussite des filles serait davantage liée au niveau d'études de leur mère que celle des garçons (De Meulemeester, 2001, cité par Van Campenhoudt & al., 2008, p.26).



1.3 Le genre

Le nombre de femmes commençant des études universitaires a fortement augmenté en vingt ans (avec une hausse de 70% par rapport à 1987) (Nils & Lambert, 2011, p.11). En outre, depuis les années 2000, en Belgique, le pourcentage de femmes à l'université est supérieur à celui des hommes. L'augmentation globale du nombre d'inscriptions à l'université pourrait, dès lors, être attribué au fait que, depuis 30 ans, plus de femmes se lancent dans des études universitaires (Droesbeke, Hecquet & Wattelar, 2001, cités par Van Campenhoudt & al., 2008).

Pour faire un lien entre le milieu social et le genre de la personne, une étude réalisée à l'ULB en 2003 (Van Campenhoudt & al., 2008) postule que cette augmentation du nombre de filles dans les études supérieures serait à l'origine de l'ouverture des universités aux milieux moins privilégiés.

D'après Vermandele, Plaigin, Dupriez, Maroy, Van Campenhoudt et Lafontaine (2010), il y aurait plus d'« hommes de milieux favorisés dans les sciences de l'ingénieur, les sciences de gestion et les sciences économiques et d'un autre côté, une population plus féminine issue de milieux plus modestes en sciences psychologiques et en information et communication » (Vermandele & al., 2010, cités par Nils & Lambert, 2011, p.11). Ainsi, le genre pourrait aussi être lié avec le facteur du choix d'études.

En outre, Droesbeke & al. (2001), lors d'une étude réalisée par l'UCL-ULB affirment que le taux de réussite des femmes est supérieur de 5,9% par rapport à celui des hommes.

1.4 L'âge

Selon le Ministère de la Communauté française de Belgique (2006, cité par Van Campenhoudt & al., 2008, p.27), plus l'âge d'entrée à l'université augmente, plus le taux moyen de réussite diminue. Ainsi, les chiffres pour l'année 2003-2004 sont les suivants : pour un étudiant ayant l'âge « normal » (une personne qui n'a pas doublé et n'a pas été avancée dans son cursus scolaire), le taux moyen de réussite en première année universitaire est de



51,9% ; pour ceux ayant un an de « retard » le taux est de 31,2% ; ceux avec plus d'un an de retard n'ont, en moyenne, que 21,4% de chances de réussite. Cependant, à mon sens, un étudiant réellement motivé qui recommence son année (avec des dispenses, par exemple) peut réussir avec plus de facilité.

1.5 Le lieu de résidence lors des études

Le fait d'avoir un kot ou d'habiter à une distance proche de son établissement scolaire peut jouer un rôle dans la réussite. Ainsi, le gain de temps pour étudier (non perdu dans les transports) peut influencer la réussite (Derbousse, Lambotte & Sermeus, 2000, in Schatteman, 2009).

2. Le parcours scolaire de l'étudiant

L'échec est-il lié à l'acquisition des compétences et aux performances antérieures de l'étudiant ? Les étudiants qui n'ont pas doublé ou qui avaient de bons résultats durant leurs études secondaires ont-ils une plus grande chance de réussir à l'université ? Des études telles que celle de Galand & al. (2005) semblent le prouver. L'option suivie durant le cursus secondaire pourrait également être corrélée avec l'échec ou la réussite. Selon les mêmes auteurs, limiter le redoublement ainsi que réduire les inégalités entre options, voire entre filières pourraient être des solutions envisageables afin de réduire ce taux d'échec (Galand & al., 2005).

2.1 Le parcours suivi avant l'entrée à l'université

2.1.1 La filière

En Belgique, tout élève détenant un certificat de l'enseignement secondaire supérieur, général ou technique (voire même professionnel, s'il réalise une septième année) peut s'inscrire à l'université (Nils & Lambert, 2011, p.10). Le lien entre la filière d'enseignement



secondaire et la réussite est réel. Ainsi, les chiffres de l'année 2003-2004 du Ministère de la Communauté française (2006, cité par Van Campenhoudt & al., 2008, p.27) le montrent : les étudiants ayant suivi la filière d'enseignement secondaire de type général ont un taux de réussite moyen en première année universitaire de 45,3%, ceux sortant de technique de transition un taux de 12% et enfin, ceux issus de technique de qualification un taux de 7,4%.

Droesbeke & al. (2001) déclarent que les élèves sortant de l'option latin-math (option d'enseignement général) en secondaire ont un taux de réussite supérieur de 42% à ceux sortant de l'enseignement technique.

2.1.2 L'option

Les sections fortes (latin-mathématique, latin-sciences, scientifiques fortes, mathématiques fortes...) sembleraient « mieux préparer les étudiants à la réussite » (Droesbeke & al., 2001 et Leclercq & al., 2003, cités par Van Campenhoudt & al., 2008, p.28). Ainsi, pour Droesbeke & al. (2001), les élèves sortant de l'option latin-math en secondaire ont un taux de réussite supérieur d'environ 35% aux options basées plus spécialement sur les sciences humaines.

2.1.3 Le redoublement en secondaire

Doubler pendant les études secondaires est-il lié à l'échec à l'université ? L'échec entraîne-t-il l'échec ? Selon Galand (2005), non seulement l'échec dans le secondaire amène un risque d'échec dans les études supérieures mais, aussi, un étudiant ayant échoué lors de sa première année à l'université a plus de risque d'échouer à nouveau. En outre, il faut souligner le coût financier de l'échec pour l'étudiant (ou ses parents) mais également pour la société. En effet, l'échec augmente d'un an minimum la scolarité de l'étudiant ce qui retarde son entrée dans la vie professionnelle (Rapport MOC-UCL, 2002, cité par Galand & al., 2005, p.12).



Néanmoins, Lambert et de Kerchove (1996, cités par Van Campenhoudt et al., 2008, p.28) « constatent une surreprésentation, à l'université, des étudiants n'ayant pas échoué au cours de leurs études secondaires ».

2.2 Le passage vers l'université

En Belgique, les acquis et compétences des écoles secondaires ne sont pas standardisés (Romainville, 2002). En effet, ils dépendent de l'établissement fréquenté, de la région dans laquelle celui-ci se situe, de la filière (le réseau d'enseignement), de l'option suivie, des critères d'évaluation, de l'équipe enseignante... Selon Romainville (2005, p.18), il peut donc déjà y avoir des inégalités sociales durant le parcours scolaire antérieur. Par exemple, aucune épreuve externe ne clôture les études secondaires. Il y a donc une diversité au sein des étudiants vis-à-vis de leurs acquis et compétences et aucun moyen mis en place, au niveau de la Communauté française, pour mesurer de façon identique la maîtrise des compétences.

Une solution proposée à cette diversité consiste en des examens d'entrée dans les universités. En effet, les facultés de sciences appliquées (ingénieur civil), seules en Belgique à imposer un tel examen ont un taux de réussite plus élevé que la moyenne allant jusque 60% de réussite (Nils & Lambert, 2011, p.11). Mais cet examen peut aussi être vu comme une façon de renforcer les inégalités sociales en ce qui concerne l'accès aux études et d'autres questions peuvent alors apparaître comme les critères d'évaluation et de sélection, l'uniformité des tests pour toutes les universités... (Galand & al., 2005). De plus, un étudiant réputé « faible » à son entrée à l'université peut développer une meilleure méthode de travail et mieux réussir que certains étudiants qui ont réussi les tests d'entrée mais sont par exemple, moins assidus au travail. Il s'agit donc de ne pas exclure directement ces étudiants. En bref, l'idée d'imposer un examen d'entrée ne fait pas l'unanimité.

Le passage des études secondaires au supérieur en Belgique est donc caractérisé, pour la majorité des universités, par un « libre accès » (Romainville, 2005, p.18). Il n'y a ni examen d'entrée, ni numerus clausus, ni test informatif, ni année préparatoire requise, ni sélection en fonction des résultats obtenus dans le secondaire ou en fonction de l'option ou de



la filière choisie précédemment. Sans sélection préalable, le tri en première année des études supérieures serait donc dû, entre autres, à ce libre accès (Galand, 2005).

Par ailleurs, le libre accès aux études supérieures montre une « certaine efficacité globale ». En effet, Droesbeke et al. (2001, cités par Romainville, 2005, p.18) estiment « que sur 100 jeunes qui entrent dans le supérieur, 62 finiront par y décrocher un diplôme, même s'il ne s'agit pas toujours de celui qu'ils convoitaient au départ et/ou qu'ils y mettront parfois plus de temps que prévu ». La réussite ou l'échec ne doivent donc pas être vus de façon déterministe. (Galand, 2005).

Selon Tinto (1997) et Prevatt & Kelly (2003) (cités par Galand & al., 2005, p.11), une autre alternative pour remédier à ce problème d'échec serait de proposer de la remédiation, des activités d'accompagnement aux étudiants, des tutorats par les pairs, l'apprentissage coopératif... Mais d'autres questions restent en suspens : faut-il, par exemple, des actions collectives ou individualisées, facultatives ou obligatoires ?

En outre, d'après Tuton (1997, cité par Romainville, 2005, p.19), il faut cependant nuancer le lien entre le parcours scolaire de l'étudiant et la réussite. En effet, si ce lien est démontré pour la première année d'étude universitaire, il n'est plus réel quand nous regardons la relation entre le parcours scolaire (dans le secondaire) et la maîtrise des compétences professionnelles.

3. La difficulté du choix d'orientation

3.1 L'orientation

Un autre facteur explicatif de la réussite est la démarche d'orientation des étudiants, l'intérêt et la motivation dans leurs projets futurs. En effet, d'après Galand, Neuville & Frenay (2005), les étudiants sûrs de leur choix d'étude ou ayant entrepris une démarche informative sur les études possibles ont plus de chance d'obtenir de bons résultats. Devrait-on alors imposer un projet d'étude dans les écoles secondaires pour aider les étudiants à choisir leurs



études futures, à mieux se connaître et à avoir davantage confiance dans leurs choix et engagements ? Devrait-on mettre plus en avant les années préparatoires ?

Cependant, d'autres études telles que celle de Biémar & al., 2003, Galand, Bentein, Frenay & Bourgeois, 2003 et Bentein, Frenay & Bourgeois, 2003 (cités par Galand & al., 2005) indiquent qu'avoir des projets trop figés peut être un désavantage (par exemple, si les objectifs de l'étudiant sont trop éloignés de la formation proposée). L'important est donc que l'étudiant trouve un sens dans son choix et dans sa formation (Biémar & al., 2003, cités par Galand & al., 2005, p.7).

En outre, selon le service de pédagogie universitaire de la FUNDP (Service de Pédagogie Universitaire (SPU), revue réseau n°49, 2001, p.2), les critères de choix d'études supérieures sont au nombre de cinq :

- « un intérêt intellectuel pour des domaines plus ou moins précis, en rapport avec le contenu de la formation (goût, envie, passions...) ;
- des atouts personnels ou des faiblesses, essentiellement en termes de compétences (sentiment de capacité ou d'incapacité, compétences estimées développées ou non, connaissances acquises ou lacunaires...) ;
- les caractéristiques de l'orientation d'étude (type d'études, longueur, investissement requis, programme...) ;
- les avantages socio-économiques escomptés de la poursuite des études, grâce à l'obtention du diplôme ;
- le champ professionnel auquel donne accès le diplôme, ce champ pouvant être envisagé sous trois angles : le statut social, les actes professionnels, les valeurs sociales ».

Selon Nils (2011, p.14), trois acteurs semblent importants pour l'orientation de l'élève : les centres PMS (psycho-médico-sociaux), les écoles secondaires elles-mêmes et les universités. En outre, il existe d'autres actions mises en place telles que les salons SIEP, les sites web des institutions, les soirées d'informations... car la « bonne » orientation de l'élève à la fin de sa rhéto est un critère important pour sa réussite future mais celle-ci est souvent sous-estimée par l'élève et son entourage. En effet, d'après Moutte (2008, cité par Nils, 2011,



p.16), « un tiers des étudiants ne fait aucune démarche de recherches d'informations avant de s'inscrire à l'université ».

3.2 La réorientation

Selon Galand, Neuville et Frenay (2005), une initiative envisageable durant la première année de bachelier est la mise en place d'un projet d'orientation progressive en cours de cursus. En effet, certains étudiants prennent conscience que leurs études ne leur conviennent pas en tout début d'année et ils souhaitent donc se réorienter. Comment dès lors les encourager dans ce sens et éviter toute démotivation ? Comment faire en sorte que l'étudiant culpabilise moins suite à son échec et ne dramatise pas celui-ci ? La déclaration de Bologne et la nouvelle organisation qu'elle implique peuvent faciliter les réorientations. L'étudiant pourrait, en quelque sorte, essayer de « tirer profit » de cette situation d'échec. En effet, l'étudiant en échec a un an de plus de maturité. Il a acquis certaines connaissances et compétences durant l'année échouée. Le système des ECTS (crédits) peut d'ailleurs être une aide à la valorisation de certains acquis (Droesbeke & al., 2001, cités par Galand, 2005, p.13).

Selon Romainville (2005, p.21), « l'introduction d'une plus grande souplesse dans les parcours nécessiterait d'importantes modifications structurelles, telles que :

- la suppression de la notion d'année ;
- une plus grande collaboration entre universités et hautes écoles ;
- la confection de programmes largement polyvalents, avec détermination progressive de la spécialisation de l'étudiant ».

Mais ces considérations sont peut-être extrêmes en ce qui concerne, par exemple, la suppression de la notion d'année.



Il existe également des systèmes tels que la formation « Rebond »¹ en province de Namur ou la « Formation-Relais » à Louvain-la-Neuve². Ces programmes sont centrés sur le projet d'étude mais aussi de vie de l'étudiant dans le cadre de sa réorientation. (Galand & al., 2005, p.13).

4. La perception et la motivation de l'étudiant

Les étudiants qui échouent n'ont-ils pas assez confiance en leur capacité de réussir ? « Ne voient-ils pas l'intérêt des cours qu'ils suivent » (Galand & al., 2005, p.7) ? Cherchent-ils juste à obtenir un diplôme ? D'après Galand (2005, p.7), « plus l'étudiant croit en ses capacités à réussir et plus il a des objectifs élevés, plus il a de chance de réussir ». L'intérêt pour la matière enseignée joue également un rôle dans la motivation de l'étudiant.

En psychologie cognitive, des auteurs tels que Bandura ou Viau ont étudié la motivation scolaire. Ainsi, une de leurs conclusions est que « l'élève construit sa motivation scolaire à partir de ses expériences, de ses réussites et de ses échecs ; il en tire des conclusions, il en extrait des règles, des lois » (Bandura, 1997, cité par Van Campenhoudt & al., 2008, p.31). L'influence des représentations des étudiants sur la réussite à l'université peut dès lors être un facteur explicatif de la réussite ou de l'échec. Un étudiant issu d'un milieu favorisé a d'ailleurs une meilleure confiance en lui et en ses capacités. Il aura donc une plus grande probabilité de réussite. De plus, les étudiants de milieux moins favorisés avec un parcours scolaire difficile « tendent à percevoir plus que les autres un écart entre les exigences académiques et leurs capacités et entre leurs projets de vie et les débouchés » (Lambert & de Kerchove, 1996, cités par Van Campenhoudt & al., 2008, p.31). Ils s'inscriraient donc à l'université avec une image plus négative d'eux-mêmes et une moindre confiance en eux. Ce

¹ La formation REBOND à Namur est un semestre (de février à juin) de réorientation et de formation se voulant une aide pour les étudiants en décrochage axé sur « la maîtrise des savoirs de base, le développement d'une méthode de travail et la réflexion sur le projet personnel » (Galand & al., 2005, p.13)

² Le CIO (centre d'information et d'orientation) de Louvain-la-Neuve propose la « formation Relais » aux étudiants qui veulent se réorienter. Celle-ci comprend des cours et des activités pour les aider à orienter leurs choix. Il y a trois sessions organisées (de novembre à juin, de février à juin et d'avril à juin). (<http://www.uclouvain.be/326547.html?idQuestion=58>, page consultée le 01/01/2011, html)



manque de motivation et cette perception fausse de soi pourraient dès lors les entraîner moins facilement vers la réussite.

En outre, la persévérance que les étudiants peuvent avoir dès l'entrée à l'université (et qui peut être comprise comme l'engagement à continuer ses études) est corrélée avec la réussite de la première année de BAC. (Noël & Frenay, 2011, p.42).

5. L'intégration sociale et académique de l'étudiant

Pour Tinto (1982, cité par Galand & al., 2005, p.7), le facteur de l'intégration sociale et académique de l'étudiant reprend « la perception subjective du sujet de trouver sa place, de façon harmonieuse, dans la vie académique et sociale de l'institution d'enseignement » mais également les interactions qui lui permettront de s'intégrer. Par exemple, le fait de ne pas se sentir isolé socialement serait un atout sur le chemin de la réussite (Beghin & al. 2002 et Bodson, 1999, cités par Galand & al, 2005, p.7). Passer du statut d'élève à celui d'étudiant engendre des changements. D'abord une adaptation à l'environnement est nécessaire afin d'acquérir une certaine autonomie et gestion du temps et des matières. Ensuite, il y a des changements dans la relation pédagogique; par exemple, les cours en auditoriums et les séances d'exercices en groupes restreints. L'étudiant devra aussi s'adapter à un contenu de matières plus vaste. Enfin, les méthodes et les manières d'apprendre sont aussi différentes; par exemple, la prise de notes, les résumés et les syllabus (Noël B., « *Comment préparer son entrée à l'université, calendrier 2010-2011 des activités d'information et d'orientation pour les élèves du secondaire* », 2010, p.3). En outre, la réussite est corrélée avec un travail régulier de l'étudiant (Parmentier, 1994, Galand & al, 2005, p.8).

6. Les cours d'été

Y a-t-il un lien entre la participation à des journées d'intégration et la réussite universitaire ? De nombreuses universités (FUCaM, Ulg, UCL, FUNDP...) organisent des « cours d'été » fin août ou début du mois de septembre. Ces cours sont destinés aux étudiants inscrits en première année de bachelier en vue de les préparer à leur nouvelle vie universitaire.



(Delhaxhe, Houart & Pollet, 2011). Selon le rapport du CIUF (Parmentier, 2011), il y a généralement une corrélation positive entre la réussite et la participation à une semaine propédeutique mais celle-ci ne se retrouve pas dans toutes les études.

7. Les pratiques d'enseignement et leur efficacité

Les pratiques d'enseignement regroupant l'encadrement, les aides financières, l'accès aux études, la façon d'évaluer... sont des critères assez peu étudiés dans la littérature et dont le lien avec la réussite ou l'échec n'est pas clairement établi (Galand & al, 2005, p.8). Néanmoins, certaines études telles que celle de Lizzio, Wilson & Simons (2002, cités par Galand & al., 2005, p.8), affirment que « le contexte d'apprentissage proposé à l'étudiant, façonné notamment par les pratiques d'enseignement, a un effet sur la performance ». Il ne faut donc pas considérer l'échec uniquement comme un problème propre à l'étudiant. Les pratiques d'enseignement dépendent fortement de l'équipe d'enseignants, de la filière, de la faculté... Ainsi, par exemple, les critères d'évaluation (examens, corrections, notations, délibérations...) peuvent avoir un lien avec la réussite (Parmentier, 2005, p.27). En outre, Galand, Neuville & Frenay (2005, p.12) mettent en avant des questions telles que « les évaluations portent-elles bien sur les compétences que les enseignants souhaitent développer chez leurs étudiants ? L'objet de ces évaluations est-il pertinent pour le développement ou la vie future des étudiants ? Ces évaluations aident-elles les étudiants à démontrer leurs acquis ? ».

En ce qui concerne l'efficacité des méthodes d'apprentissage, il ressort que des facteurs tels que « le climat de discipline et de travail en classe, le temps réel d'apprentissage, les opportunités d'apprentissage, la qualité de l'instruction, le caractère structuré de l'apprentissage, les feedbacks donnés aux étudiants et des attentes élevées des enseignants sont des paramètres favorables aux apprentissages » (Campbell, Kyriakides, Muijs & Robinson, 2004, cités par Vause, Dupriez & Dumay, 2008, p.4). Le caractère coopératif de l'apprentissage est aussi souligné par Noël et Frenay (2011). Dans ce cas, les étudiants ne sont pas mis en compétition, ils apprennent plutôt à travailler ensemble et à se faire mutuellement confiance sans perdre d'énergie à se faire concurrence.



8. Les tests formatifs

Il existe plusieurs sortes de tests organisés par les universités. Par exemple, le projet « passeport pour le bac », organisé dès la première semaine de cours, a été développé par quatre universités partenaires : FUNDP, UCL, FUCaM et FUSL. A l’UCL³, il y a des passeports en biologie, chimie, compréhension d’un texte de droit, compréhension d’un texte en sciences humaines, physique et mathématique. Par la suite, des séances de correction, remédiation et renforcement sont proposées par l’université pour que l’étudiant puisse évaluer son niveau de connaissances et comprendre ses erreurs.

La plupart des universités organisent ces tests formatifs durant les mois d’octobre ou novembre. Les étudiants sont alors confrontés pour la première fois aux exigences universitaires et ils peuvent se rendre compte de leurs lacunes et ainsi, par exemple, adapter leur méthode de travail. (Closset & Delforge, 2011, p.25).

Aux Facultés Universitaires Notre-Dame de la Paix (FUNDP) à Namur, il y a une forte participation des étudiants de BAC1 à ces tests. Même si certains ne participent aux tests que pour voir le type de questions que le professeur peut poser. Les résultats d’une enquête (Messian, 2008, cité par Closset & Delforge, 2011, p.27), montrent une forte corrélation entre les résultats aux tests formatifs et la réussite en fin d’année académique (et cela sur plusieurs années : de 2003 à 2008).

9. Les monitorats

Les séances de remédiation peuvent aider les étudiants si celles-ci sont organisées assez tôt dans l’année et donc si les étudiants ont pris conscience de leurs lacunes assez vite. Parmentier (2005, p.28) soulève le problème de la participation à ces activités. Est-ce les étudiants qui en ont le plus besoin qui y prennent part ou les plus motivés ? Aux FUCaM, par

³ <http://www.uclouvain.be/passeport>, page consultée le 01/02/2011, html



exemple, il a été constaté que les monitorats visent et touchent vraiment un public ciblé : des étudiants ayant une chance de réussite mais conscients qu'ils ont besoin d'aide. (Noël B, 2011, Rapport bisannuel).

10. La réussite en janvier

Une étude réalisée aux FUCaM (Vandamme & al., 2006) sur les étudiants de l'année 2003-2004 a essayé de classer ceux-ci en trois groupes : low risk, medium risk et high risk. Ces groupes sont basés sur la probabilité d'échouer leur année (risque faible, risque moyen et haut risque d'être en échec en fin de BAC1). Le but était de cibler les étudiants ayant besoin d'actions ou d'accompagnement pour lutter contre l'échec scolaire (medium risk) en gardant une chance de réussite.

Cette étude montre, par exemple, un lien fort entre les résultats obtenus à la fin de la session de janvier, ceux de juin et la réussite de l'année. Par exemple sur le graphique ci-dessous, nous voyons le lien entre les moyennes des notes obtenues en janvier et le grade de fin d'année. Il y a deux groupes d'étudiants opposés. Les « high risk » qui ont obtenu une note moyenne en janvier de moins de 9 sur 20 ont tous échoué leur année (sauf deux étudiants). Les « low risk » ayant une moyenne supérieure à 14 sur 20 en janvier ont tous réussi.

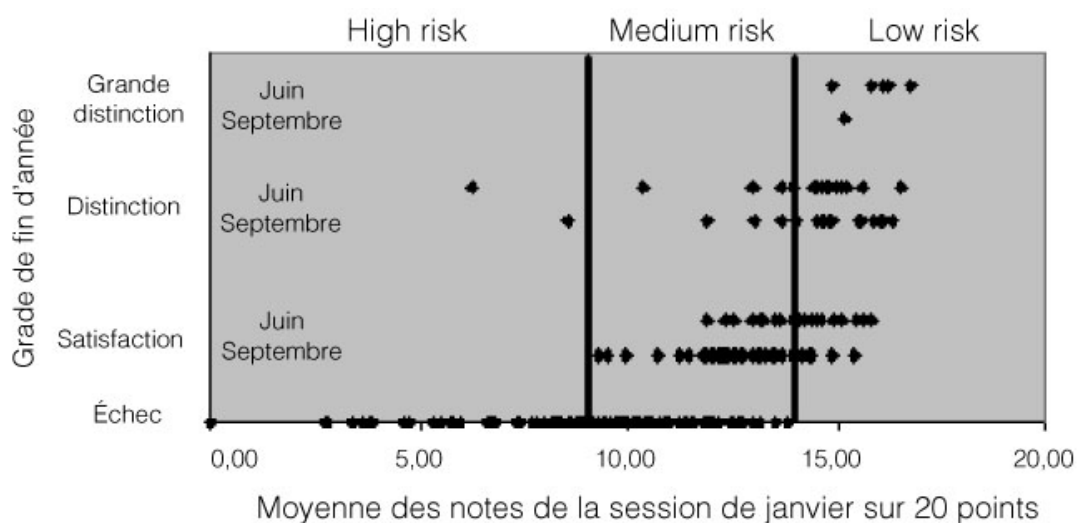


Figure 1 : Graphique représentant les moyennes des notes de la session de janvier (sur 20) et les grades obtenus fin d'année par les étudiants des FUCaM



Cependant, il faut rester prudent quant à ces conclusions liées à un outil prédictif axé sur les premières performances de l'étudiant (Romainville, 2005). En effet, ces liens reposent uniquement sur une note attribuée (performance) et non sur les compétences réellement acquises.

Néanmoins, B. Noël, professeur et responsable de l'équipe d'accompagnement des FUCaM retrouve également cette liaison forte entre les résultats de janvier des étudiants de BAC1 et leurs points en fin d'année (Noël B., 2011, Rapport bisannuel).

12. La déclaration de Bologne : une nouvelle organisation

Le programme d'une année académique est constitué de deux quadrimestres (de septembre à décembre et de janvier à juin). Chacun de ceux-ci est clôturé par une session d'examens. Depuis la déclaration de Bologne, l'étudiant de BAC1 a la possibilité de représenter en juin et en août les examens de janvier qu'il n'a pas réussis. Il y a également un système de report des notes où l'étudiant qui a obtenu 10/20 ou 11/20 peut conserver sa cote d'une session à l'autre (Leclercq & Parmentier, 2011, p.7). Une autre nouvelle mesure est la possibilité de passer à l'année suivante avec 48 crédits au lieu de 60 pour une année complète (ceci impliquant de devoir réussir en BAC2 les cours de BAC1 non crédités). (Leclercq & Parmentier, 2011, p.7). Ces nouvelles mesures veulent donc aider les étudiants sur le chemin de la réussite.

Section 3 : Conclusion

Cette première partie qui reprend des recherches tirées de la littérature sur ce sujet n'est pas exhaustive et des facteurs complémentaires pourraient bien sûr être pris en compte pour expliquer la réussite universitaire. Ainsi, d'autres critères pourraient apporter des précisions. Par exemple :

- L'étudiant a-t-il des activités extra-scolaires ? Pratique-t-il un sport ?



- Quelle place occupe-t-il dans l'auditoire ? S'assied-t-il plutôt en avant, tout derrière ou sur le côté ?
- A-t-il un frère ou une sœur qui suit des études universitaires ? Ou qui se trouve dans la même université ?
- ...

En outre, Van Campenhoudt & al. (2008, p.33) déclarent « qu'il convient d'aller au-delà des approches descriptives [...] et que dans des études ultérieures, il sera, à cet égard, important de veiller à appréhender simultanément les effets spécifiques des facteurs ».

Il est aussi important de tenir compte du fait que les données et leurs modélisations permettent d'exprimer des taux de réussite ou d'échec qui concernent des groupes d'étudiants et non des prévisions individuelles. Les conclusions devraient donc plutôt être considérées comme des mesures de « facteurs de risques rendant l'échec plus ou moins probable » plutôt que comme des réels facteurs d'échec. (Droesbeke & Vermandele, 2005, p.24).

Comme nous le voyons, les causes de l'échec ou de la réussite à l'université peuvent donc être multiples (et parfois également mal identifiées). Romainville (2005) met, par exemple, en évidence trois catégories de facteurs : l'histoire personnelle de l'étudiant (qui reprend le profil de l'étudiant), les perceptions de l'étudiant face à son nouveau mode de vie et son adaptation (par exemple, sa motivation et son intégration) et les pratiques d'enseignements (programmes de cours, évaluations...). Dans ce travail de fin d'étude, pour nos analyses, nous ne prendrons en compte que le premier facteur (le deuxième faisant appel surtout à des facteurs motivationnels et le dernier dépendant plus particulièrement de l'institution). Nous tenterons donc de démontrer si nos résultats sont en lien avec la littérature existante.



Deuxième chapitre : Projets mis en place au sein des FUCaM

Cette deuxième partie s'intéresse aux éléments déjà mis en place au sein des FUCaM, université dans laquelle sont ensuite étudiés les facteurs potentiellement explicatifs de la réussite. Il est important de se rendre compte du travail et des actions qui y sont déjà menés. En effet, cette problématique est au cœur des préoccupations des FUCaM depuis des années.

Section 1 : Introduction

Comme d'autres universités, les FUCaM organisent des activités ayant pour but de favoriser la réussite des étudiants car, en moyenne, le taux de réussite est assez bas et approche les 35-40% de réussite. Il y a donc une certaine déperdition de la population de BAC1. Au sein des FUCaM, les taux de réussite des étudiants de BAC1 sont d'ailleurs repris dans le tableau suivant, d'une part, calculés avec tous les étudiants inscrits pour la rentrée de septembre et, d'autre part, lorsqu'on retire les abandons durant l'année (pourcentages calculés par madame S. Goemaere, membre de l'équipe d'accompagnement) :

	<u>2008-2009</u>		<u>2009-2010</u>	
	Tous les étudiants	Si on retire les abandons en cours d'année	Tous les étudiants	Si on retire les abandons en cours d'année
Total	34,8 %	64,4 %	34,5 %	58,6 %
Orient = 1 (sciences de gestion & ingénierat de gestion)	38,3 %	66 %	38,1 %	60 %
Orient = 2 (sciences humaines, sciences politiques & information et communication)	29,7 %	61,4 %	29,6 %	56 ,45 %

Tableau 1 : Taux de réussite total et par regroupement d'orientations pour les années 2008-2009 et 2009-2010 des étudiants des FUCaM



Comme les auteurs Parmentier et Leclercq (2011, p.6) le rappellent, il faut donc cependant nuancer ce taux de réussite très bas (35-40%). Ainsi, par exemple, si nous comparons ces taux avec les statistiques de réussite en primaire et secondaire, ces dernières sont plus élevées mais il ne faut pas oublier que l'enseignement y est obligatoire. Les auteurs nous disent qu'« à l'université, en première année, il en va souvent tout autrement. En effet, la statistique du taux de réussite à l'université est calculée par la division du nombre de réussites au numérateur et du total des inscrits au dénominateur. Or combien abandonnent, ne se rendent jamais aux cours ou ne se présentent pas aux examens ? Il faudrait les retrancher du dénominateur qui ne devrait comporter que les étudiants assidus » (voir deuxième et quatrième colonnes du tableau 1).

Section 2 : Les projets mis en œuvre aux FUCaM

Les FUCaM ont déjà mis en place différentes actions pour aider les étudiants : ces activités sont menées aussi bien avant l'entrée à l'université que pendant le cursus de l'étudiant.

1. Les matinées portes ouvertes et salons SIEP

Comme explicité précédemment, l'orientation de l'étudiant et le choix de ses études sont importants en matière de motivation de l'étudiant. Les FUCaM organisent des matinées portes ouvertes pour les élèves qui finissent leurs études secondaires. Un accueil privilégié est prévu ainsi qu'« une présentation des programmes de cours et une visite du campus : laboratoires de langues, salles informatiques, logements... » (Facultés Universitaires Catholiques de Mons, « *Comment préparer son entrée à l'université, calendrier 2010-2011 des activités d'information et d'orientation pour les élèves du secondaire* », 2010, p.5). De plus, des professeurs, étudiants et membres du personnel administratif répondent aux questions des élèves. Les FUCaM sont également présentes aux différents salons SIEP (service d'information sur les études et les professions).



2. Le projet boussole

Chaque année, les FUCaM organisent une journée « projet boussole » durant les congés scolaires des élèves de secondaire (« congé de carnaval »). Son but premier est « d'améliorer l'orientation des élèves en fin de secondaire et de clarifier leur démarche d'élaboration de projet personnel et professionnel avant l'entrée à l'université » (Noël, 2011, Rapport bisannuel, p.21). Ils peuvent y découvrir la vie universitaire aux FUCaM, visiter les bâtiments, feuilleter des syllabus, participer à un cours magistral, rencontrer des étudiants moniteurs, poser des questions... En 2008, 60 élèves y ont participé et en 2009 ils étaient 69 inscrits (Noël, 2011, Rapport bisannuel, p.21).

En outre, pendant l'année, les élèves de 6^e secondaire peuvent demander à l'administration des études pour venir suivre l'un ou l'autre cours « ouvert ».⁴

3. Les cours d'été (cours propédeutiques)

Les cours préparatoires, cours d'été ou propédeutiques consistent en une semaine de préparation à la première année d'étude universitaire (ils comprennent des exemples de mini-cours, séances de révision en mathématique, ateliers de méthodologie et de prise de notes, activités linguistiques au centre des langues, visites des bâtiments, contacts avec des professeurs et assistants...). C'est aussi l'occasion pour le futur étudiant de se familiariser avec son nouvel environnement et de connaître d'autres futurs étudiants (Facultés Universitaires Catholiques de Mons, « *Comment préparer son entrée à l'université, calendrier 2010-2011 des activités d'information et d'orientation pour les élèves du secondaire* », 2010, p.8). Par exemple, un temps de midi est consacré à la présentation des activités organisées sur le campus et des associations étudiantes. Cette semaine existe aux FUCaM depuis 1999 et depuis 2003, elle accueille chaque année environ cent cinquante étudiants (Noël, 2011, Rapport bisannuel, p.22).

⁴ En 2010, un cours de prise de notes en contexte a également été organisé lors du cours d'histoire économique et sociale. Celui-ci a réuni des étudiants de BAC1 mais également des élèves de 6^e secondaire. Un feedback a été donné à la fin du cours. (Noël, 2011, Rapport bisannuel).



4. Les tests organisés en début d'année académique

De 1993 à 2008, les étudiants passaient un test informatif de français dont ils recevaient ensuite un feedback personnalisé lors d'un entretien individuel avec l'équipe d'accompagnement. Par la suite, pour s'adapter, entre autres, au projet « passeport pour le bac » de l'Académie Louvain, le test a été abandonné au profit d'un test plus quantitatif des prérequis et mieux adapté au public des FUCaM : « le passeport mathématique ». L'objectif de ce test reste le même que celui du test de français, c'est-à-dire « favoriser un premier contact par un entretien, très tôt dans l'année, ainsi que déceler d'éventuelles difficultés d'intégration ou d'adaptation à l'université » (Noël, 2011, Rapport bisannuel, p.23).

En effet, la maîtrise et l'acquisition de prérequis sont des facteurs importants de réussite. La définition de cette notion de prérequis donnée par Wathelet et Vieillevoys est la suivante : « Toutes connaissances ou compétences qui s'avèrent cruciales pour la maîtrise d'un cours, d'une discipline, d'un programme et qui sont considérées par les enseignants comme devant être acquises préalablement à l'entrée à l'université »⁵. Il est donc important d'identifier les lacunes de chaque étudiant concernant les prérequis en début d'année pour ainsi, par la suite, mettre en place des séances de monitorats⁶ ou d'aides personnalisées. D'autres objectifs du « passeport pour le bac » sont de mieux identifier les prérequis et l'adéquation qui doit être réalisée entre le secondaire et le supérieur mais aussi de donner au jeune une indication sur ses connaissances et compétences réelles grâce à un feedback rapide et constructif. (Romainville & Slosse, 2011, p.23).

5. Les tests formatifs de novembre

Les tests formatifs datent de 2005. Au départ, ils portaient sur toutes les matières vues depuis la rentrée. Depuis 2008, les tests formatifs portent sur quatre cours maximum.

⁵ http://www.fundp.ac.be/det/actualites/reussir-transition-secondaire-universite/maitrisedes_prerequis-et-reussite-universite/view, page consultée le 25/03/2011, pdf

⁶ Depuis 2010, des sessions appelées « coup de pouce en math » sont organisées aux FUCaM dès le début d'année.



Selon Noël (2011, Rapport bisannuel, p.25), leurs objectifs sont « de permettre aux étudiants de faire un diagnostic précoce de leurs connaissances et, en corollaire, de leur méthode de travail, de mieux percevoir les exigences des enseignants et d’être confrontés à une simulation proche des examens de janvier ».

De plus, il faut savoir que ces tests reposent sur différentes modalités d’organisation :

- « ils sont obligatoires et cotés mais non dispensatoires (non valorisés⁷) ;
- l’étudiant peut s’inscrire à deux tests minimum et quatre au maximum [...] ;
- les tests sont courts, construits de manière à permettre à l’étudiant qui maîtrise la matière d’y répondre en 45 minutes environ ;
- un feedback rapide est prévu (dans les 10 jours) [...] ;
- en cas de résultats insatisfaisants, les étudiants peuvent bénéficier de monitorats dans les branches quantitatives » (Noël, 2011, Rapport bisannuel, p.25).

Par exemple, en 2008, 96,5% des étudiants avaient pris part aux tests. 56% d’entre eux s’étaient inscrits à quatre tests, 23% à trois tests et 21% à deux tests. (Noël, 2011, Rapport bisannuel, p.25)

6. Les monitorats

Les monitorats, séances de remédiation en groupes, sont de deux types pour les étudiants de BAC1 aux FUCaM :

- « le monitorat durant l’année, créé en 2002, assumé à partir de 2008 par un enseignant extérieur aux FUCaM ;
- le monitorat dit « d’été » destiné à favoriser la réussite des étudiants (en échec en juin) à la session d’août et assuré par des étudiants plus âgés » (Noël, 2011, Rapport bisannuel, p.34).

⁷ Cependant, depuis septembre 2010, ces tests sont valorisés. Cette décision a été prise pour encourager et motiver un plus grand nombre d’étudiants à y participer de façon sérieuse.



7. Les entretiens individuels

L'équipe d'accompagnement des FUCaM reçoit les étudiants désireux d'un rendez-vous individuel pour les aider, les écouter, les conseiller... En outre, une adresse e-mail « réussir » a été mise en place pour encourager les étudiants à se manifester et à demander de l'aide. (Noël, 2011, Rapport bisannuel, p.36). Ainsi, cette équipe peut aider et promulguer des conseils si l'étudiant a des difficultés dans la gestion du temps, l'organisation de son travail, la prise de notes, le fait de pouvoir résumer ou synthétiser un cours mais elle peut aussi écouter le jeune qui a des difficultés d'ordre parfois plus personnel.

Section 3 : Conclusion

Nous voyons à la lecture de ces projets que les FUCaM tentent, tout comme d'autres universités, d'aider et d'accompagner au mieux l'étudiant vers la réussite tant par des cours de remédiation que grâce à l'équipe d'accompagnement des BAC1. Cette problématique est également le sujet de nombreuses recherches et mémoires au sein des FUCaM (Jomaux F., Meskens N., Noël B., Goemaere S., Lust T. ...)



Troisième chapitre : Modèles, méthodes et tests statistiques

La troisième partie de ce travail est consacrée à l'étude des modèles, méthodes et tests statistiques réalisés par la suite sur les bases de données. Après une brève introduction, la deuxième section est consacrée à des rappels d'ordre théorique et la troisième aux explications des développements permettant au lecteur de suivre les modèles utilisés tout au long de ce travail.

Les différents résultats ont été obtenus à l'aide du logiciel **SPSS** (Statistical Package for the Social Sciences), nouvellement renommé **PASW** (Predictive Analytics Software). Ce logiciel a été choisi car il est simple à utiliser et qu'il offre, entre autres, les possibilités d'analyser des données, de générer des tableaux de résultats, des diagrammes de distributions et de tendances, des statistiques descriptives et des analyses statistiques (régressions linéaires, logistiques...).

Section 1 : Introduction

Notre but est d'analyser les données en vue de sélectionner les facteurs les plus pertinents pour expliquer la réussite et par la suite, définir les implications pédagogiques et actions possibles qui en découlent pour les FUCaM. Différents tests statistiques, mathématiques et économétriques ont donc été utilisés pour analyser nos bases de données.

La réalisation de ce chapitre (particulièrement les parties théoriques) se base principalement sur la consultation des deux ouvrages de référence suivants :

- KINNEAR P.R., GRAY C.D. (2005), « *SPSS facile appliqué à la psychologie et aux sciences sociales* », trad. HUET N., Paris, Ed. De Boeck
- HOWELL D.C. (2008), « *Méthodes statistiques en sciences humaines* », trad. ROGIER M., YZ ERBYT V., BESTGEN Y., Paris, Ed. De Boeck, (2^e éd.)



Section 2 : Rappels de quelques concepts de base

1. La nature des variables

Les variables peuvent être de quatre types : nominales, ordinales, d'intervalles et à échelle de rapport. Ces quatre types peuvent être aussi classés en deux catégories : les variables non métriques, qualitatives (nominales et ordinales) et les variables métriques, quantitatives (d'intervalles et à échelle de rapport).

Les variables nominales prennent des valeurs dans une liste finie, par exemple des couleurs (« vert, bleu, rouge,... »). Parmi celles-ci, il existe les variables binaires (variables prenant seulement les valeurs 0 ou 1), par exemple : le sexe d'une personne (homme = 0, femme = 1) ou la réussite d'un examen (réussi = 0, échec = 1). Les variables ordinales se différencient des premières par le fait que l'ordre des valeurs est important. Ainsi, un classement des valeurs est possible. Les variables d'intervalles sont des mesures continues. Les différences entre ces variables peuvent être comparées, par exemple : la température en degrés Celsius. Enfin, les variables à échelle de rapport permettent d'identifier, de classer, de comparer des différences, par exemple : l'âge, le revenu ou les points obtenus. Il est à noter qu'avec ce type de variables la valeur « 0 » représente une absence de phénomène ce qui les différencie des variables d'intervalles pour lesquelles la valeur « 0 » est une valeur qui existe (par exemple : 0°C).

Le graphique ci-dessous est un résumé schématique du classement des variables.

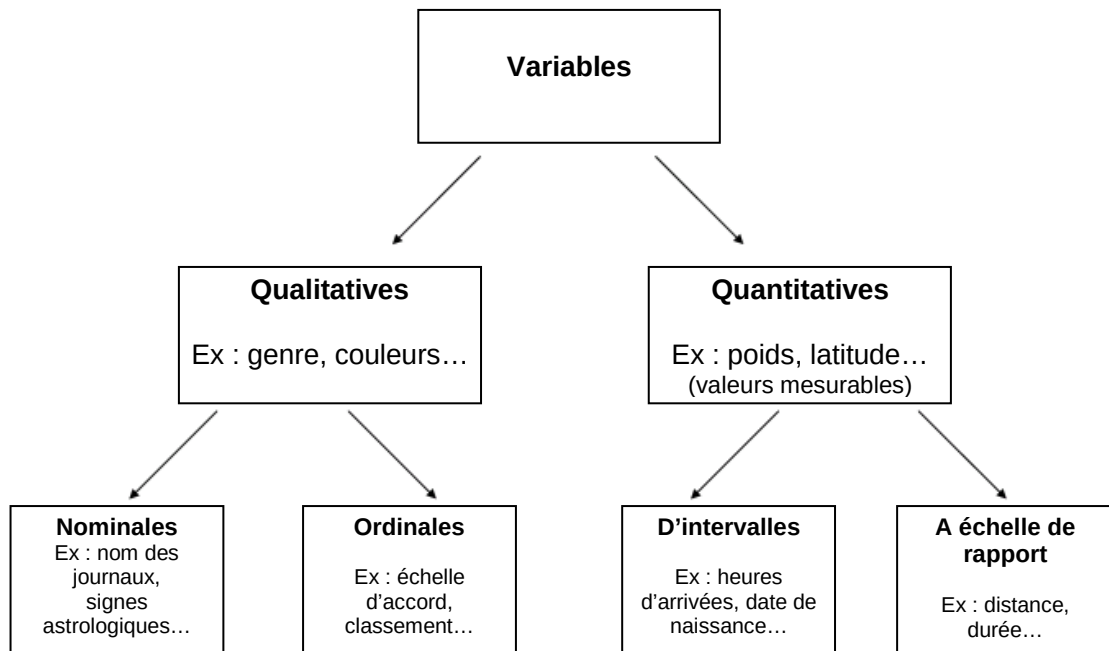


Figure 2 : Représentation schématique du classement des variables

2. Les hypothèses

Les **hypothèses statistiques** sont posées à partir des caractéristiques de la distribution de probabilités d'une population. Généralement, une hypothèse « établit qu'il y a une relation causale entre deux ou plusieurs variables » (Kinnear & Gray, 2005, p.11). Il existe deux hypothèses : celle de départ, l'hypothèse nulle (H_0) qui est l'hypothèse « attendue » et l'hypothèse alternative (H_1) différant de H_0 .

Par exemple, pour un test de comparaison de moyennes, les hypothèses de départ sur la population sont les suivantes :

- $H_0 : \mu_1 = \mu_2$ (égalité des moyennes des deux populations)
- $H_1 : \mu_1 \neq \mu_2$

Par la suite, les hypothèses sont acceptées ou rejetées.



3. Le risque d'erreur

Le **risque d'erreur** est appelé alpha. Celui-ci peut être défini de cette façon :

$$\alpha = P(\text{rejeter } H_0 \text{ quand } H_0 \text{ est vraie})$$

Alpha peut aussi être appelé « seuil de rejet » ou « seuil de signification » de l'hypothèse H_0 . Dans nos analyses, nous avons choisi de façon arbitraire un risque d'erreur α fixé à 5%. Il est important de souligner que ce seuil sera régulièrement pris comme valeur de référence dans ce travail. Il sera ainsi comparé à une significativité donnée (souvent reprise dans les tableaux du logiciel SPSS par **Sig.**). Si $\alpha = 5\%$, nous avons alors 5% de chances de rejeter l'hypothèse nulle à tort et ainsi de se tromper. En outre, il est possible de diminuer ou d'augmenter ce taux suivant le risque d'erreur que l'on est prêt à encourir.

En outre, il existe deux **types d'erreurs** possibles lors de tests d'hypothèses. Si nous regardons dans le tableau 2, nous voyons que quand nous rejetons H_0 alors qu'elle est vraie, nous prenons le risque α de faire une erreur (de type I). Si nous ne la rejetons pas, il y a un risque β de se tromper.

Le bêta est d'ailleurs exprimé comme suit :

$$\beta = P(\text{ne pas rejeter } H_0 \text{ quand } H_0 \text{ est fausse})$$

Le tableau des types d'erreurs est le suivant :

		<u>Réalité</u>	
		<u>H0 vraie</u>	<u>H1 vraie (H0 fausse)</u>
<u>Décision</u>	<u>Ne pas rejeter H0</u>	Décision correcte	Erreur de type II (probabilité β)
	<u>Rejeter H0</u> (supposer H1 vraie)	Erreur de type I (probabilité α)	Décision correcte

Tableau 2 : Tableau des types d'erreurs



4. La significativité (Sig.) ou p-value et les tests d'hypothèses

Lors de la réalisation des tests d'hypothèses est calculée une probabilité qui sera, par la suite, comparée à un seuil critique. Par exemple, pour un test d'hypothèses sur le coefficient β_1 de la droite de régression $y = \beta_0 + \beta_1 x + \varepsilon$, les hypothèses sont les suivantes :

- $H_0 : \beta_1 = 0$
- $H_1 : \beta_1 \neq 0$

Prenons un exemple concret pour expliquer la significativité dans le cas d'une régression linéaire simple avec la variable à expliquer **Moyenne_global_100** (qui reprend les points obtenus (sur 100) par l'étudiant en fin de BAC1) et la variable explicative **Passeport_moy_20** (à savoir, les points de l'étudiant (sur 20) au passeport mathématique passé en début d'année). Soit H_0 , l'hypothèse : pas de liaison linéaire entre les deux variables (donc $\beta_1 = 0$). Avec la base de données 2008-2009 et les étudiants de sciences de gestion et ingénierat de gestion, nous obtenons avec le logiciel SPSS le tableau suivant :

Model	Unstandardized Coefficients		t	Sig.
	B	Std. Error		
1 (Constant)	-17,135	8,076	-2,122	,036
Passeport_moy_20	4,284	,600	7,138	,000

Tableau 3 : Tableau SPSS – Coefficients de la régression linéaire

La statistique utilisée pour le test d'hypothèses sur le coefficient β_1 est une statistique en t qui suit une loi de probabilité de Student à $n - 2$ degrés de liberté :

$$t_{b_1} = \frac{b_1}{s_{b_1}} = \frac{4,284}{0,6} = 7,14$$



Ainsi, nous rejetons H_0 (dans le cadre du test sur β_1) lorsque :

$$|t_{b_1}| \geq t_{1-\frac{\alpha}{2}}(n-2)$$

où $t_{1-\frac{\alpha}{2}}(n-2)$ est le seuil critique.

Dès lors, si nous prenons un risque $\alpha = 5\%$ et que nous savons que $n = 163$:

$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(163-2) = 1,96$$

La valeur de 1,96 est trouvée dans la table de Student. Ici, après comparaison des statistiques et seuils critiques, nous rejetons donc l'hypothèse H_0 et nous concluons à l'existence d'une liaison linéaire entre **Moyenne_global_100** et **Passeport_moy_20**.

Le niveau de signification de la statistique t (représenté par Sig.) peut être calculé de la manière suivante (dans le cadre du test d'hypothèses sur β_1) :

$$2 * \text{Prob}(t(n-2) \geq |t_{b_1}|) =$$

$$2 * \text{Prob}(t(161) \geq 7,14) = 0$$

Ainsi, comme nous avons une valeur nulle pour la significativité (Sig. = 0), la liaison entre les deux variables est très significative. Cette notion de significativité (faisant référence aux valeurs trouvées dans des tables statistiques) sera régulièrement citée et utilisée dans la suite de ce travail de fin d'étude.

Section 3 : Modèles, méthodes et tests statistiques

Cette section est divisée en deux grandes parties. La première présente les tests réalisés pour des modèles étudiant le lien entre une variable à expliquer quantitative et :

- des variables explicatives binaires : **tests de comparaisons de moyennes** ;
- des variables explicatives quantitatives : **corrélations linéaires** ;
- des variables explicatives qualitatives et quantitatives : **régressions linéaires**.



La deuxième section du chapitre étudie des modèles montrant le lien entre une variable à expliquer qualitative et :

- des variables explicatives qualitatives : **test du Chi-carré de Pearson** ;
- des variables explicatives quantitatives : **comparaisons de moyennes** ;
- des variables explicatives qualitatives et quantitatives : **régressions logistiques**.

Pour expliquer chacun des tests de façon didactique, nous présenterons à chaque fois un exemple chiffré illustrant le modèle avec les tableaux réalisés avec le logiciel SPSS.

1. Modèles étudiant le lien avec une variable à expliquer quantitative

1.1 Variables explicatives binaires : comparaisons de moyennes

Nous avons choisi de prendre comme exemple une comparaison des moyennes des résultats de fin de 1^{ère} année (variable : **Moyenne_global_100**) en fonction de la variable **math6h** qui est une variable binaire prenant la valeur « 1 » si l'étudiant a suivi six heures de mathématique ou plus durant ses études secondaires et la valeur « 0 » s'il a suivi moins de six heures de mathématique. Ici, la base de données utilisée est celle de l'année 2008-2009 pour les sciences de gestion et ingénieurs de gestion (**Orient** = 1).

Pour effectuer cette comparaison de moyennes, nous pouvons effectuer un **test de Student**. En effet, le test-t de Student a pour but de déterminer si les moyennes des deux groupes, à savoir d'une part, les étudiants ayant suivi moins de 6 heures de mathématique dans le secondaire et d'autre part, ceux ayant suivi au moins 6 heures de mathématique sont significativement différentes.

Dans le logiciel SPSS, nous obtenons tout d'abord un tableau de statistiques où sont reprises les valeurs des effectifs (N), des moyennes, des écarts-types et des erreurs standard moyennes de chacun des groupes.



Group Statistics

	Math6h	N	Mean	Std. Deviation	Std. Error Mean
Moyenne_global_100	0	115	31,92	28,517	2,659
	1	68	48,54	28,526	3,459

Tableau 4 : Tableau SPSS – Statistiques descriptives pour la variable Moyenne_global_100

Sur base de ce tableau, nous voyons que la moyenne des étudiants ayant eu moins de 6 heures de mathématique dans le secondaire (31,92) est nettement inférieure à celle de l'autre groupe (48,54).

Pour détailler la première ligne du tableau 4 :

- La **moyenne** (Mean) de l'échantillon peut être calculée avec la formule suivante :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{115} (3671,27) = 31,92$$

Remarque relative aux notations : le logiciel SPSS nous parle de « N » dans ses tableaux pour les effectifs mais dans nos formules, nous utilisons « n » (« N » étant, pour nous, réservé aux effectifs de la population).

- L'**écart-type** (Standard Deviation) :

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{114} (92708,0817)} = 28,517$$

- L'**erreur standard moyenne** (Standard Error Mean) :

$$\frac{s}{\sqrt{n}} = \frac{28,517}{\sqrt{115}} = 2,659$$

Cette dernière est un indice de dispersion tout comme l'écart-type et est utilisée pour comparer des groupes ne comportant pas le même nombre d'observations.



Le tableau 5 (ci-dessous) reprend les résultats de deux tests : celui d'homogénéité des variances et celui d'égalité des moyennes.

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means							
								95% Confidence Interval of the Difference		
	F	Sig.	T	Df	Sig. (2- tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
Moyenne_global_100	Equal variances assumed	1,863	,174	-3,809	181	,000	-16,617	4,363	-25,225	-8,008
	Equal variances not assumed			-3,808	140,7	,000	-16,617	4,363	-25,243	-7,991

Tableau 5 : Tableau SPSS – Tests d'homogénéité des variances et d'égalité des moyennes pour la variable Moyenne_global_100

En effet, le test de Student de comparaison de moyennes requiert la validation de deux hypothèses : l'homogénéité des variances et la normalité des distributions. L'hypothèse de normalité pourrait poser problème pour notre échantillon mais le test en t de Student est relativement robuste aux violations de cette hypothèse surtout pour les échantillons où n (le nombre total d'observations) est supérieur à 30.

Pour l'**homogénéité des variances**, un test d'égalité des variances doit être effectué. Celui-ci pose comme hypothèses :

- H_0 : homogénéité des variances (par exemple : $\sigma_1^2 = \sigma_2^2$)
- H_1 : non homogénéité des variances

Avec le logiciel SPSS, cette hypothèse H_0 est vérifiée grâce au **test de Levene** qui se situe dans les 2 premières colonnes du tableau 5 : la statistique F de Levene vaut 1,863 avec une probabilité critique associée de 0,174. Cette probabilité est supérieure au seuil critique α de 5% (= 0,05). Le calcul de la significativité a été expliqué ci-avant (section 2 : Rappels de quelques concepts de base - point 5 : La significativité (Sig.) ou p-value et les tests d'hypothèses). Donc, nous acceptons l'hypothèse nulle d'égalité des variances.



A partir des conclusions du test de Levene, nous pouvons analyser les résultats au test t de Student (suivant l'égalité ou non des variances, nous choisissons la première ou deuxième ligne du tableau et les valeurs du test en t correspondantes). Donc, dans notre cas, nous considérons la première ligne puisque l'hypothèse des variances égales est acceptée.

Le **test de Student**, quant à lui, pose les hypothèses suivantes :

- $H_0 : \mu_1 = \mu_2$ (égalité des moyennes des deux populations)
- $H_1 : \mu_1 \neq \mu_2$

Dans l'exemple, la statistique t du test d'égalité des moyennes vaut -3,809 avec une probabilité critique associée de 0,000 qui est inférieure aux 5% fixés. L'hypothèse H_0 est rejetée. Il existe donc une différence significative entre les moyennes.

La formule pour calculer le **t de Student** est la suivante :

- Si les variances de la population sont égales (ceci étant montré grâce au test de Levene) :

- Il faut d'abord calculer l'écart-type estimé (s) :

$$s = \frac{(n_1 - 1) * s_1 + (n_2 - 1) * s_2}{(n_1 + n_2 - 2)} = \frac{(114 * 28,517 + 67 * 28,526)}{(115 + 68 - 2)} = 28,52063$$

où n_1 est le nombre d'observations du groupe 1 et n_2 le nombre d'observations du groupe 2.

- Puis, la valeur de la statistique t :

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left(\frac{s^2}{n_1}\right) + \left(\frac{s^2}{n_2}\right)}} = \frac{|31,9241 - 48,5409|}{\sqrt{\left(\frac{28,52063^2}{115}\right) + \left(\frac{28,52063^2}{68}\right)}} = 3,809$$



- Si les variances de la population ne sont pas égales, la valeur de la statistique t est :

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}} = \frac{|31,9241 - 48,5409|}{\sqrt{\left(\frac{28,51717^2}{115}\right) + \left(\frac{28,52654^2}{68}\right)}} = 3,808$$

Pour détailler le tableau 5 :

- La « **Mean Difference** », la différence entre les moyennes est de :

$$-16,617 = 31,9241 - 48,5409 = \bar{x}_1 - \bar{x}_2$$

- Si les variances sont égales, la « **Std. Error Difference** », l'erreur standard de la différence entre les moyennes est de 4,36296, trouvée à partir de la formule :

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\left(\frac{28,52063^2}{115}\right) + \left(\frac{28,52063^2}{68}\right)} = 4,3629$$

- La « **Std. Error Difference** » si les variances ne sont pas homogènes est :

$$4,363 = \sqrt{\left(\frac{28,51717^2}{115}\right) + \left(\frac{28,52654^2}{68}\right)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Le paragraphe suivant détaille les **degrés de liberté** (« Df » dans le tableau 5). En effet, quand les variances sont égales (suite au test de Levene), le nombre de degrés de liberté vaut $(n_1 + n_2 - 1 - 1)$ (avec n_1 = la taille du groupe 1 ; n_2 = la taille du groupe 2). Ainsi, les degrés de liberté sont ici égaux à 181 [= $(115 + 68) - 1 - 1$]. En effet, il y avait, dans notre exemple 115 étudiants ayant suivi moins de 6 heures de mathématiques dans le secondaire et 68 ayant suivi au moins 6 heures.

En ce qui concerne le nombre de degrés de liberté pour l'hypothèse des variances inégales, la manière dont nous calculons le t observé fait intervenir séparément la variance du groupe 1 et la variance du groupe 2. Le nombre de degrés de liberté devient donc une valeur approchée (par l'approximation de Welch-Satterthwaite). Cette valeur est comprise entre :

- le plus petit des nombres entre $n_1 - 1$ et $n_2 - 1$ (entre 114 et 67)
- et le nombre de degrés de liberté décrit auparavant, à savoir $n_1 + n_2 - 2 = 181$ (qui est la somme de $(n_1 - 1)$ et de $(n_2 - 1)$).



Dans Tenenhaus (2000), nous trouvons la formule permettant d'approcher le nombre de **degrés de liberté** dans ce cas :

- dl = degré de liberté ;
- $1 / dl = [c^2 / (n_1 - 1)] + [(1 - c)^2 / (n_2 - 1)]$;
- $c = [\sigma_1^2 / n_1] / [(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)]$.

Ainsi, dans l'approximation du nombre de degrés de liberté, nous utilisons un certain poids, défini par la variable « c ». Nous mesurons donc le poids de chaque groupe en terme de variance et utilisons ce poids pour calculer les degrés de liberté. Donc, si la variance du groupe 2 est beaucoup plus grande que celle du groupe 1, l'approximation tient compte essentiellement du nombre de degrés de liberté du groupe 2, à savoir $(n_2 - 1)$. (Dagnelie, 2006).

L'intervalle de confiance (tableau 5) est de $[-25,22560 ; -8,00799]$ quand les variances sont égales. Cet intervalle est construit sur base d'une probabilité :

$$P(-25,22560 < (\mu_1 - \mu_2) < -8,00799) = 0,95$$

Comme la valeur « 0 » n'appartient pas à l'intervalle, nous rejetons l'hypothèse H_0 d'égalité des moyennes.

- Si les variances sont égales, l'intervalle de confiance est donné par la formule :

$$\begin{aligned} IC_{0,95} &= (\bar{x}_1 - \bar{x}_2) \pm t_{0,5} \hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = (31,9241 - 48,5409) \pm 1,97 * \sqrt{\left(\frac{28,52063^2}{115}\right) + \left(\frac{28,52063^2}{68}\right)} \\ &= -16,6168 \pm 1,97 * 4,36296 \end{aligned}$$

- Si les variances ne sont pas égales, nous utilisons la même formule mais avec la valeur de la variance de chaque groupe.

L'**analyse Anova** (analyse de variances) est également un test de comparaison de moyennes mais il peut être réalisé sur deux ou plus de deux variables. De ce fait, le test en t de Student est un cas particulier de l'Anova.



Par exemple, l'analyse Anova permet de réaliser une comparaison multiple de moyennes des résultats de fin de 1^{ère} année de bachelier en fonction des résultats des étudiants en fin de secondaire répartis selon trois catégories (variable **Pourc**) :

- 1 : moins de 65% ;
- 2 : entre 65% et 75% ;
- 3 : plus de 75%.

Le tableau 6 (ci-dessous) reprend les effectifs, moyennes, écarts-types et erreurs standard moyennes des trois groupes. Il comprend également les intervalles de confiance, les valeurs minimums et maximums.

Descriptives

Moyenne_global_100

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	35	35,643	27,520	4,652	26,189	45,096	,00	74,42
2	102	37,350	28,655	2,837	31,722	42,979	,00	81,00
3	27	51,953	32,944	6,340	38,921	64,986	,00	88,00
Total	164	39,390	29,523	2,305	34,838	43,942	,00	88,00

Tableau 6 : Tableau SPSS – Analyse descriptive de la variable Moyenne_global_100
répartie en 3 sous-groupes

Pour notre exemple, la première ligne du tableau reprend la « **Standard Error** » ou l'erreur standard moyenne :

$$\frac{s}{\sqrt{n}} = \frac{27,52022}{\sqrt{35}} = 4,652$$

- L'**intervalle de confiance** est de :

$$[26,1890 ; 45,0961] = [35,6426 - 2,03 * 4,652 ; 35,6426 + 2,03 * 4,652]$$



La première condition à remplir avant de réaliser l'analyse Anova est le test d'homogénéité des variances.

Test of Homogeneity of Variances

Moyenne_global_100

Levene Statistic	df1	Df2	Sig.
,468	2	161	,627

Tableau 7 : Tableau SPSS – Test d'homogénéité des variances pour la variable Moyenne_global_100

Dans le tableau ci-dessus, $df\ 1 = m - 1 = 3 - 1 = 2$ et $df\ 2 = N - m = 164 - 3 = 161$ où m est le nombre de groupes et N est le nombre total d'observations. La **statistique de Levene** est de 0,468 avec une significativité de 0,627. Donc l'hypothèse d'homogénéité est acceptée et nous pouvons par conséquent poursuivre avec l'analyse Anova.

ANOVA

Moyenne_global_100

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	5177,44	2	2588,721	3,044	,050
Within Groups	136902,76	161	850,328		
Total	142080,21	163			

Tableau 8 : Tableau SPSS – Analyse ANOVA pour la variable Moyenne_global_100

Dans le tableau 8 (ci-dessus) est représentée l'analyse Anova. Les écarts « **Between Groups** » représentent les écarts existants entre les différents échantillons (= variation factorielle, entre groupes) et les écarts « **Within Groups** » sont les écarts existants à l'intérieur des échantillons (= variation résiduelle, au sein des groupes).

Pour détailler ce tableau, il faut tout d'abord calculer la moyenne :

- La **moyenne** :

$$\bar{x}_{..} = [35,6426 * (35/164) + 37,3504 * (102/164) + 51,9533 * (27/164)] = 39,3901$$



Et ensuite,

- La **somme des carrés des écarts factoriels** (Between Groups) :

$$SCE_F = \sum_{j=1}^r n_j (\bar{x}_{.j} - \bar{x}_{..})^2 = 35 * (35,6426 - 39,3901)^2 + 102 * (37,3504 - 39,3901)^2 + 27 * (51,9533 - 39,3901)^2 = 5177,441$$

- La **somme des carrés des écarts résiduels** (Within Groups) :

$$SCE_R = \sum_{j=1}^r \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2 = \sum_{i=1}^{n_1} (x_{i1} - \bar{x}_{.1})^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_{.2})^2 + \sum_{i=1}^{n_3} (x_{i3} - \bar{x}_{.3})^2 = 25750,3215 + 82934,214 + 28218,2298 = 136902,765$$

- La **somme des carrés des écarts totaux** :

$$SCE_T = SCE_F + SCE_R = 136902,765 + 5177,441 = 142080,206$$

Les **degrés de liberté** sont de $N - 1$ ($164 - 1 = 163$) pour les écarts totaux; $m - 1$ ($3 - 1 = 2$) pour les écarts factoriels (où m est le nombre de groupes) et le nombre de degrés de liberté des écarts résiduels est $N - m$ ($164 - 3 = 161$).

La troisième colonne du tableau 8 reprend les « **Mean Square** », les carrés moyens. Elle est obtenue en divisant la colonne des sommes des carrés des écarts (colonne 1) par les degrés de liberté (colonne 2).

Par exemple : les carrés moyens factoriels valent $2588,721 = 5177,441 / 2$.

Les quatrième et cinquième colonnes reprennent la statistique F qui vaut 3,044 et sa probabilité critique de 0,05. L'hypothèse nulle affirmant qu'il n'existe aucune différence entre les moyennes est donc acceptée à un seuil de 1% et rejetée à un seuil de 10%. Cette **statistique F** est calculée par les carrés moyens factoriels divisés par les carrés moyens résiduels $= 2588,721 / 850,328 = 3,044$.



Pour déterminer entre quels groupes les différences sont statistiquement valables, nous pouvons utiliser le test « post hoc » de **Scheffé**. En effet, ce test utilise des comparaisons de moyennes par paires.

Multiple Comparisons

Moyenne_global_100

Scheffe

(I) Pourc	(J) Pourc	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-1,708	5,7124	,956	-15,821	12,406
	3	-16,311	7,4692	,095	-34,765	2,143
2	1	1,7078	5,7124	,956	-12,406	15,821
	3	-14,603	6,3111	,072	-30,196	,9899
3	1	16,311	7,4692	,095	-2,143	34,765
	2	14,603	6,3111	,072	-,9899	30,196

Tableau 9 : Tableau SPSS – Comparaisons multiples de moyennes pour la variable Moyenne_global_100

La première colonne reprend la **différence des moyennes** ; par exemple, à la première ligne : $35,6426 - 37,3504 = -1,7078$

L'**intervalle de confiance**, pour la première ligne, est de :

$$[-15,8215 ; 12,4058] = [-1,70782 - 2,47 * 5,71241 ; -1,70782 + 2,47 * 5,71241]$$

Ce test admet qu'une différence entre les moyennes des résultats est significative si la probabilité associée aux regroupements de la variable **Pourc** est inférieure à 5%. Ici, à un seuil de 5%, aucune différence de moyennes n'est significative. A un seuil de 10%, nous voyons que les étudiants ayant obtenu moins de 65% en fin de secondaire (**Pourc** = 1) ont une moyenne en fin de BAC1 différente des étudiants qui ont plus de 75% à la fin de leur secondaire (**Pourc** = 3). Les étudiants ayant obtenu entre 65% et 75% en fin de secondaire (**Pourc** = 2) ont également une moyenne en fin de BAC1 différente des étudiants qui ont plus de 75% à la fin de leur secondaire (**Pourc** = 3). Le fait d'avoir eu plus de 75% en fin de secondaire est donc pour la population étudiée, un facteur favorable à la réussite en fin de BAC1.



1.2 Variables explicatives quantitatives : corrélations linéaires

La **corrélation linéaire** mesure le degré de liaison entre deux variables dans une relation linéaire entre celles-ci. Le coefficient de corrélation (r) entre une variable x et une variable y varie de -1 à 1. S'il n'y a aucune corrélation, le coefficient vaut 0.

« Il y a corrélation entre deux variables observées sur les éléments d'une même population lorsque les variations des deux variables se produisent dans le même sens (corrélacion positive) ou lorsque les variations sont de sens contraire (corrélacion négative) » (Baillargeon, 2003, p.508).

Dans un premier temps, la représentation du diagramme de dispersion permet de se faire une idée intuitive sur le sens de la relation. Chaque observation est représentée par un point. Par exemple, si nous prenons les variables **Moyenne_global_100** et **passport_moy_20** pour l'année 2008-2009 et pour les orientations sciences de gestion et ingénieur de gestion (**Orient =1**), le graphique est le suivant :

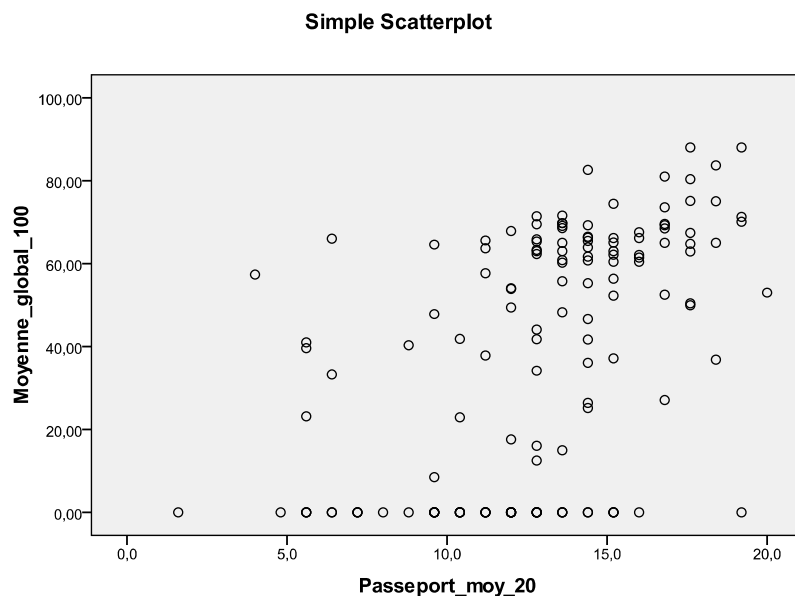


Figure 3 : Graphique SPSS représentant les variables **Moyenne_global_100** et **Passeport_moy_20** pour l'année 2008-2009 et pour Orient = 1



Sur le graphique, nous voyons intuitivement le sens de la relation entre les deux variables. Nous pouvons voir que la corrélation sera positive.

Dans le tableau 10 (ci-dessous), pour la première ligne, la corrélation entre la variable **Moyenne_global_100** et **Passeport_moy_20** est de 0,515 avec une significativité de 0,000. Le coefficient r est calculé avec la formule suivante :

$$r = \frac{\text{cov}(x, y)}{s_x * s_y} = \frac{55,89}{30,14 * 3,62} = 0,51$$

La diagonale du tableau affiche logiquement la valeur « 1 » car le programme SPSS réalise la corrélation entre une variable et elle-même.

Correlations

		Moyenne_global_100	Passeport_moy_20
Moyenne_global_100	Pearson Correlation	1	,515 ^{**}
	Sig. (2-tailed)		,000
	N	183	143
Passeport_moy_20	Pearson Correlation	,515 ^{**}	1
	Sig. (2-tailed)	,000	
	N	143	143

** . Correlation is significant at the 0.01 level (2-tailed).

Tableau 10 : Tableau SPSS – Tableau de corrélations entre les variables Moyenne_global_100 et Passeport_moy_20

Comme r vaut 0,515, nous pouvons conclure que la relation entre les deux variables est d'intensité moyenne.

1.3 Variables explicatives qualitatives et quantitatives : régressions linéaires

La théorie de cette partie sur les régressions linéaires est essentiellement tirée de l'ouvrage de référence suivant :

- TENENHAUS M. (2007), « *Méthodes pour décrire, expliquer et prévoir* », Paris, Ed. Dunod



1.3.1 Ajustement linéaire et calcul de la droite de régression linéaire simple

La **régression linéaire** est une méthode qui tente de répondre à la question : existe-t-il une relation linéaire entre la variable à expliquer y et une ou des variable(s) explicative(s) ? Cette méthode peut s'appréhender de deux façons :

- Du point de vue descriptif, une droite permettant de prédire au mieux la variable « y » à partir de la ou des variables explicatives est construite.
- Du point de vue inférentiel, des tests permettent de vérifier l'existence de cette liaison linéaire.

En ce qui concerne l'**analyse descriptive**, un intérêt est de rechercher la droite qui permet de prédire au mieux la variable y à partir de x . Pour une population, l'équation linéaire pour la régression simple est la suivante:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

où

- y est la variable à expliquer ;
- β_0 et β_1 sont les paramètres à estimer ; β_0 représentant l'ordonnée à l'origine et β_1 la pente de la droite ;
- ε est le terme d'erreur.

Pour l'échantillon, nous recherchons une **estimation de y** (nommée $\hat{y}$) et les **estimateurs des paramètres β_0 et β_1** (appelés b_0 et b_1). Nous obtenons donc cette équation :

$$\hat{y}_i = b_0 + b_1 x_i$$

Pour ajuster au mieux la droite au nuage de points, nous pouvons utiliser la méthode des moindres carrés ordinaires. La première étape pour obtenir les valeurs des estimateurs b_0 et b_1 est de minimiser la somme des carrés des écarts entre y_i et $\hat{y}_i$.

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$



Les erreurs de prédiction ou résidus sont les écarts entre les valeurs observées (y_i) et les valeurs prédites ($\hat{y}_i$).

A partir de cette formule, nous obtenons les valeurs b_1 et b_0 :

$$b_1 = \frac{\text{cov}(x, y)}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Prenons un exemple concret de régression avec comme variable à expliquer **Moyenne_global_100** et comme variable explicative **Passeport_moy_20**. Ici, nous avons utilisé la base de données 2008-2009 pour les étudiants en sciences de gestion et ingénierat de gestion (**Orient** = 1). Le logiciel SPSS nous donne le tableau suivant :

Coefficients ^a					
Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
1 (Constant)	-17,135	8,076		-2,122	,036
Passeport_moy_20	4,284	,600	,515	7,138	,000

a. Dependent Variable: Moyenne_global_100

Tableau 11 : Tableau SPSS – Coefficients de la régression linéaire

En effet, à partir des formules, nous obtenons :

$$b_1 = \frac{\text{cov}(x, y)}{s_x^2} = \frac{56,2851}{(3,6247)^2} = 4,28$$



Nous trouvons dans le tableau 12, les moyennes pour x et y ainsi que les écarts-types qui permettent de calculer $b_0 = \bar{y} - b_1 \bar{x}$:

Descriptive Statistics			
	Mean	Std. Deviation	N
Moyenne_global_100	38,3952	30,14017	143
Passeport_moy_20	12,962	3,6247	143

Tableau 12 : Tableau SPSS – Analyse descriptive pour la régression linéaire

Nous obtenons ainsi :

$$b_0 = \bar{y} - b_1 \bar{x} = 38,3952 - 4,284 * 12,962 = -17,13$$

Dès lors, nous avons $\hat{y} = b_0 + b_1 x = -17,135 + 4,284 * x$. Ainsi, pour un étudiant qui a obtenu 16/20 au passeport mathématique : $\hat{y} = -17,135 + 4,284 * 16 = 51,409$; l'étudiant devrait donc obtenir théoriquement 51,4% de moyenne.

Pour expliciter la 3^e colonne du tableau 11, nous allons tout d'abord estimer la **variance σ^2** :

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - b_0 - b_1 * x)^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2 = \frac{1}{141} (94756,6735) = 672,033145$$

$$s = \sqrt{s^2} = \sqrt{672,033145} = 25,9236021$$

L'**écart-type de l'estimateur b_0** est estimé par :

$$s_{b_0} = \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{25,9236021}{43,1934728} = 0,6$$



Et celui de l'estimateur b_1 par :

$$s_{b_1} = s * \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = 25,9236021 *$$

$$\sqrt{\frac{1}{143} + \frac{12,962^2}{1865,67609}} = 8,076$$

Le **bêta** est le coefficient b pondéré. Il est standardisé. Il ne dépend pas des unités de mesure et est dès lors plus facile à interpréter. Il montre un changement dans la variable dépendante (exprimé par des unités d'écart-type) qui serait produit par une augmentation positive d'un écart-type pour la variable indépendante. Par exemple, ici, le coefficient de 0,515 signifie que, si la variable **Passeport_moy_20** (variable explicative) augmente d'un écart-type, la variable **Moyenne_global_100** augmente de 0,515 écart-type. Les coefficients ne dépendant pas des unités de mesures peuvent être plus facilement comparables.

$$beta = \frac{b_1 * s_{Passeport_moy_20}}{s_{Moyenne_global_100}} = \frac{4,28 * 3,6247}{30,14017} = 0,515$$

où $s_{Passeport_moy_20}$ et $s_{Moyenne_global_100}$ sont trouvés dans le tableau 12.

D'un point de vue **inférentiel**, le t (dans le tableau 11) est un test de Student qui permet de vérifier la significativité du coefficient de régression. Il est calculé de la manière suivante :

$$t = \frac{b_1}{s_{b_1}} = \frac{4,284}{0,6} = 7,14 \text{ (pour la variable } \mathbf{Passeport_moy_20})$$

Ici, le test est réalisé sur la pente de la droite et part de l'hypothèse suivante :

- $H_0 : \beta_1 = 0$
- $H_1 : \beta_1 \neq 0$

Nous voulons donc voir s'il y a une relation entre les variables x et y. Vu la valeur du t de 7,14 et sa significativité proche de 0, nous refusons H_0 . Il y a donc bien une relation linéaire entre les deux variables.



Dans le tableau 13 (ci-dessous), le r , **coefficient de corrélation** a déjà été calculé dans la section concernant la corrélation linéaire :

$$r = \frac{\text{cov}(x, y)}{s_x * s_y} = \frac{56,2851}{3,6247 * 30,14017} = 0,515$$

Un r supérieur à 0,5 montre que l'ajustement du modèle est relativement bon. En outre, le coefficient est positif donc la relation est croissante.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,515 ^a	,265	,260	25,92360

a. Predictors: (Constant), Passeport_moy_20

Tableau 13 : Tableau SPSS – Coefficients de corrélation pour la régression linéaire simple

- Le **R-Square** est de :

$$R^2 = (0,515)^2 = 0,265$$

Le R^2 mesure la force de liaison entre les variables. Ainsi, un R^2 de 0,265 signifie que 26,5% de la variation de la **Moyenne_global_100** est expliquée par la variable **Passeport_moy_20**.

- Et l'**Adjusted R-Square** est de :

$$1 - \frac{(1 - r^2) * (N - 1)}{(N - 2)} = 1 - \frac{(1 - 0,265) * (142)}{141} = 1 - 0,74 = 0,26$$

- La **Standard Error of the Estimate** (tableau 13) est égale à :

$$\sqrt{\frac{\sum (y - \hat{y})^2}{N - 2}} = \sqrt{\frac{94756,6735}{141}} = 25,924$$

où le dénominateur est aussi appelé la somme des carrés résiduelle ($SC_{\text{résiduelle}}$).



Pour calculer la **statistique de F**, nous devons d'abord trouver :

- la **somme des carrés de la régression** ($SC_{régression}$) : $\sum (\hat{y} - \bar{y})^2 = 34\,240,348$;
- la **somme des carrés résiduelle** ($SC_{résiduelle}$) qui est égale à $\sum (y - \hat{y})^2 = 94\,756,673$;
- et la **somme des carrés totaux** (SC_{totaux}) : $\sum (y - \bar{y})^2 = 128\,997,021$.

La somme des carrés totaux peut également se retrouver en additionnant la somme des carrés résiduelle et la somme des carrés de la régression ($128997,021 = 34240,348 + 94756,6735$). Ces trois éléments se retrouvent dans la 2^e colonne du tableau ci-dessous :

ANOVA ^b						
Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	34240,348	1	34240,348	50,950	,000 ^a
	Residual	94756,674	141	672,033		
	Total	128997,021	142			

a. Predictors: (Constant), Passeport_moy_20

b. Dependent Variable: Moyenne_global_100

Tableau 14 : Tableau SPSS – Tableau Anova de la régression linéaire simple

Les **degrés de liberté** (Df) sont égaux à m-1, N-m et N-1 où m est le nombre de groupes et N le nombre d'observations. Les « **mean square** » sont calculés de la façon suivante :

$$34\,240,348 = \frac{(SC_{régression})}{Df} = \frac{34\,240,348}{1}$$

$$672,033 = \frac{(SC_{résiduelle})}{Df} = \frac{94\,756,674}{141}$$

- La **statistique F** correspond à :

$$\frac{MeanSquare_{régression}}{MeanSquare_{résiduelle}} = \frac{34\,240,348}{672,033} = 50,95$$



Comme pour le test de Student, nous voulons voir si l'ajustement est globalement valable. Ce test d'hypothèses en F part des hypothèses :

- $H_0 : \beta_1 = 0$
- $H_1 : \beta_1 \neq 0$

Nous pouvons d'ailleurs montrer que $F = t_{b_1}^2$

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (N - 2)}$$

Or,

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$s = \sqrt{\frac{1}{N - 2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$s_{b_1} = \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Donc,

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (N - 2)} = \frac{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{s^2} = \frac{b_1^2}{s_{b_1}^2} = t_{b_1}^2$$

Ici, nous avons donc $F = 50,95 = (7,138)^2 = t_{b_1}^2$. Comme la statistique du F vaut 50,95 avec une probabilité critique de 0,000, l'hypothèse nulle est rejetée.



1.3.2 Régression linéaire multiple

La **régression linéaire multiple** est semblable à la régression simple avec plusieurs variables indépendantes.

Pour la population, l'équation peut s'écrire :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

où $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ sont des paramètres et ε est un terme aléatoire suivant une loi normale $N(0, \sigma)$.

Dans cet exemple, nous avons pris comme variable dépendante **Moyenne_global_100** et comme variables indépendantes **Passeport_moy_20** et **Sexe**. La variable **Sexe** est une variable binaire définie comme suit : « 1 » pour les garçons et « 2 » pour les filles. Nous obtenons à nouveau des coefficients qui nous permettent d'écrire l'équation de régression :

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-32,518	9,833		-3,307	,001
	Passeport_moy_20	4,196	,589	,506	7,128	,000
	Sexe	12,352	4,453	,197	2,774	,006

Tableau 15 : Tableau SPSS – Régression linéaire multiple

$$\hat{y} = -32,518 + 4,196 * x_1 + 12,352 * x_2$$

Ainsi, si la variable **Sexe** augmente d'une unité (donc si elle passe de 1 à 2), la **Moyenne_global_100** augmentera de 12,352 unités.

Si nous prenons un autre exemple, un étudiant qui a 15/20 au passeport mathématique et qui est une fille devrait théoriquement obtenir, une moyenne en fin de BAC1 de 55,13%, grâce à la formule suivante :

$$\hat{y} = -32,518 + 4,196 * 15 + 12,352 * 2 = 55,13$$



Le R (du tableau 16), **coefficient de corrélation multiple** se définit comme « la corrélation entre la variable y et la meilleure combinaison linéaire des prédicteurs. En tant que tel, R n'est en réalité rien d'autre que $r_{y\hat{y}}$ » (Howell, 2008, p.521).

$$r_{y\hat{y}} = \frac{\text{cov}(y, \hat{y})}{s_y * s_{\hat{y}}} = \frac{275,999}{30,072 * 16,62} = 0,55$$

Le R^2 est égal à $(0,55)^2 = 0,3$. Nous pouvons ainsi dire que 30% de la variation de y peut être prédite sur base des deux variables **Passeport_moy_20** et **Sexe**. Avec une valeur de 30%, nous pouvons considérer que la relation entre les variables est d'intensité moyenne. Quand nous regardons le R^2 de la régression linéaire simple (de 0,265), nous voyons que celui-ci est plus élevé quand nous ajoutons la variable **Sexe** au modèle (0,3). Le modèle a donc été amélioré.

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,548 ^a	,300	,290	25,33729

a. Predictors: (Constant), Sexe, Passeport_moy_20

b. Dependent Variable: Moyenne_global_100

Tableau 16 : Tableau SPSS – Tableau des coefficients de corrélation pour la régression linéaire multiple

Le tableau suivant a déjà été détaillé pour la régression linéaire simple précédemment.

ANOVA^b

Model	Sum of Squares	Df	Mean Square	F	Sig.
1 Regression	34999,509	2	17499,755	26,138	,000 ^a
Residual	91722,190	137	669,505		
Total	126721,699	139			

a. Predictors: (Constant), Passeport_moy_20, heures_math

b. Dependent Variable: Moyenne_global_100

Tableau 17 : Tableau SPSS – Tableau Anova pour la régression linéaire multiple



2. Modèles étudiant le lien avec une variable à expliquer qualitative

2.1 Variables explicatives qualitatives : test du Chi-carré de Pearson

Le **test d'indépendance du Chi-carré** a pour but de savoir s'il existe un lien entre deux variables qualitatives. Par exemple, y a-t-il un lien entre d'une part le genre de l'étudiant et, d'autre part, la réussite de la première année de bachelier ?

Il se base sur une comparaison des effectifs observés et des effectifs théoriques (si H_0 est vrai). Il mesure ensuite l'écart entre ces deux types d'effectifs et calcule une valeur du χ^2 . Celle-ci sera alors comparée au seuil critique pour savoir si nous acceptons ou rejetons l'hypothèse nulle d'indépendance des variables. Ainsi, si nous refusons H_0 , les deux caractères seront dépendants l'un de l'autre.

Prenons un exemple concret dans notre base de données de 2008-2009 pour les étudiants en sciences ou ingénierat de gestion : y a-t-il dépendance ou indépendance entre la réussite en BAC1 (**Reussite_annee** = 1) ou non (**Reussite_annee** = 0) et le fait d'avoir suivi moins de 6 heures de mathématique dans le secondaire (**math6h** = 0) ou au moins 6 heures (**math6h** = 1) ?

Voici le tableau des **effectifs observés** :

Reussite_annee * math6h Crosstabulation

Count

		math6h		Total
		0	1	
Reussite_annee	0	82	31	113
	1	33	37	70
Total		115	68	183

Tableau 18 : Tableau SPSS – Tableau des effectifs observés d'un test du Chi-carré de Pearson pour les variables Reussite_annee et math6h



Nous pouvons également calculer la matrice des **effectifs théoriques** (attendus) :

Reussite_annee * math6h Crosstabulation

Count

		math6h		Total
		0	1	
Reussite_annee	0	71,01	41,99	113
	1	43,99	26,01	70
Total		115	68	183

Tableau 19 : Tableau SPSS – Tableau des effectifs théoriques d'un test du Chi-carré de Pearson pour les variables Reussite_annee et math6h

La manière d'obtenir, par exemple, les effectifs théoriques pour un étudiant qui a eu moins de 6 heures de mathématique et qui a raté est la suivante :

$$\frac{115 * 113}{183} = 71,01$$

où

- 115 est le nombre total d'étudiants ayant eu moins de 6 heures de mathématique en secondaire (math6h = 0) ;
- 113 est le nombre total d'étudiants ayant raté leur BAC1 (Reussite_annee = 0);
- 183 est l'effectif total.

Le tableau ci-dessous (tableau 20), nous donne la valeur du **Chi-carré de Pearson** et sa significativité.

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11,964 ^a	1	,001
Continuity Correction ^b	10,900	1	,001
Likelihood Ratio	11,892	1	,001
N of Valid Cases	183		

Tableau 20 : Tableau SPSS – Test du Chi-carré de Pearson pour les variables Reussite_annee et math6h



Le Chi-carré se calcule sur base de la formule suivante :

$$\chi^2 = \sum \frac{(\text{effectifs}_{\text{observés}} - \text{effectifs}_{\text{attendus}})^2}{\text{effectifs}_{\text{attendus}}} =$$

$$\frac{(82 - 71,01)^2}{71,01} + \frac{(31 - 41,99)^2}{41,99} + \frac{(33 - 43,99)^2}{43,99} + \frac{(37 - 26,01)^2}{26,01} = 11,96$$

Nous devons alors comparer cette valeur à la valeur trouvée dans la table de la distribution du χ^2 .

Le **nombre de degrés de liberté** est donné par la formule suivante :

$$Df = (L-1)*(C-1) = (2-1)*(2-1) = 1$$

où L = le nombre de lignes du tableau 18 et C = le nombre de colonnes du tableau 18.

Ici, nous voyons que la significativité est de 0,001, elle est donc inférieure au seuil α des 5%. Sous forme mathématique, la probabilité peut s'écrire : $0,001 = P(\chi^2_{df=1} > 11,96)$ [le calcul de cette statistique étant d'ailleurs détaillé au tout début de ce chapitre (section 2 : Rappels de quelques concepts de base - point 5 : La significativité (Sig.) ou p-value et les tests d'hypothèses)]. Ici, nous rejetons l'hypothèse H_0 d'indépendance des variables avec α égal à 5%. Il y a donc un lien entre les deux variables : réussite du BAC1 et le fait d'avoir suivi plus de 6 heures de mathématique dans le secondaire.

La deuxième ligne du tableau 20 représente la **correction de la continuité de Yates**. « Cette correction consiste uniquement à réduire la valeur absolue de chaque numérateur de 0,5 unité avant d'élever au carré » (Howell, 2008, p.147). Ainsi, nous avons :

$$\frac{(|82 - 71,01| - 0,5)^2}{71,01} + \frac{(|31 - 41,99| - 0,5)^2}{41,99} + \frac{(|33 - 43,99| - 0,5)^2}{43,99} + \frac{(|37 - 26,01| - 0,5)^2}{26,01} = 10,90$$



Le **likelihood ratio** ou χ^2 **de vraisemblance maximum** découle d'un test qui calcule la probabilité (ou vraisemblance d'occurrence) des données quand l'hypothèse nulle est vraie ou quand l'hypothèse alternative est vraie. « Si les données sont beaucoup plus vraisemblables pour l'hypothèse alternative que pour H_0 , nous serons tentés de rejeter H_0 . Par contre, si les données sont presque aussi vraisemblables selon H_0 que selon l'hypothèse alternative, nous serons tentés de maintenir H_0 . Le rapport de vraisemblance sert donc de base à l'évaluation de l'hypothèse nulle. » (Howell, 2008, p.155). Quand il n'y a qu'une dimension, la formule est la suivante :

$$\chi^2_{(C-1)} = 2 \sum \text{effectifs}_{\text{observés}_i} \ln\left(\frac{\text{effectifs}_{\text{observés}_i}}{\text{effectifs}_{\text{attendus}_i}}\right)$$

« Cette valeur peut être évaluée à l'aide de la table standard du χ^2 avec C-1 degrés de liberté. » (Howell, 2008, p.155).

Pour un tableau avec des **effectifs** n_{ij} , la formule est légèrement modifiée :

$$\chi^2_{(L-1)*(C-1)} = 2 \sum \text{effectifs}_{\text{observés}_{ji}} \ln\left(\frac{\text{effectifs}_{\text{observés}_{ji}}}{\text{effectifs}_{\text{attendus}_{ji}}}\right) =$$

$$2 * [82 * \ln\left(\frac{82}{71,01}\right) + 31 * \ln\left(\frac{31}{41,99}\right) + 33 * \ln\left(\frac{33}{43,99}\right) + 37 * \ln\left(\frac{37}{26,01}\right)] = 11,89$$

En outre, différentes conditions d'utilisation doivent être remplies avant d'effectuer un test du χ^2 . En effet, l'échantillon doit être aléatoire, la population de laquelle il provient doit être infinie ou pratiquement infinie et les effectifs attendus doivent tous être égaux au moins à 5.

2.2 Variables explicatives quantitatives : comparaisons de moyennes

Le principe est le même qu'énoncé précédemment (dans le paragraphe 1.1 variables explicatives binaires : comparaisons de moyennes) excepté que les deux groupes dont la moyenne est à comparer sont définis en fonction des deux valeurs prises par la variable dépendante binaire (et non plus par la variable explicative).



2.3 Variables explicatives qualitatives et quantitatives : régressions logistiques

Il peut y avoir différentes finalités dans l'utilisation de la régression logistique. Par exemple, un but **explicatif** qui pousse à déterminer l'influence des facteurs explicatifs sur la variable à expliquer ou une finalité **prédictive** qui a pour but de prédire l'appartenance à un groupe.

La régression logistique est une méthode servant à expliquer un lien de causalité entre une variable à expliquer binaire et une ou un ensemble de variables explicatives. Elle veut donc estimer la probabilité d'occurrence d'un événement ou d'une appartenance à un groupe.

En outre, elle permet de résoudre certains problèmes qui pourraient se poser lors de l'utilisation de la régression linéaire avec une variable dépendante binaire. Premièrement, la régression linéaire dont la variable dépendante est binaire peut fournir des valeurs de probabilité inférieure à zéro ou supérieure à un. Ce qui n'a pas de sens d'un point de vue mathématique. De plus, dans le modèle linéaire, tout accroissement de la variable indépendante a un effet constant sur la variable dépendante (nous pouvons d'ailleurs le voir intuitivement grâce à sa représentation graphique qui est une droite).

Si nous représentons graphiquement, avec les données de l'année 2008-2009 pour les sciences de gestion et ingénierat de gestion, la proportion des étudiants ayant réussi leur première année de bachelier en fonction des points obtenus au passeport mathématique, nous voyons se dessiner la forme d'une courbe en « S » (excepté la valeur de 5/20 où un étudiant sur les 4 que comporte le groupe (25%) a réussi son BAC1, ce qui n'est pas représentatif).

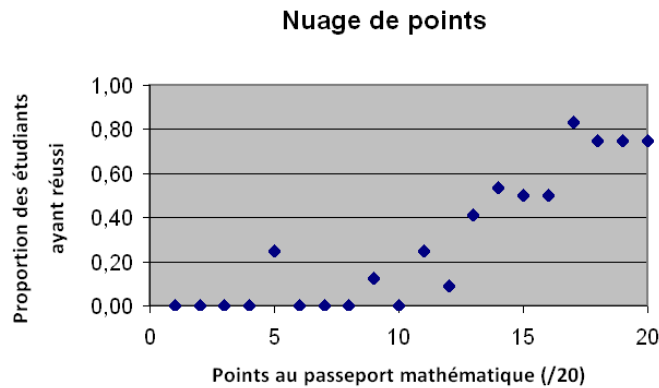


Figure 4 : Graphique de la proportion d'étudiants qui ont réussi leur BAC1 en fonction des points au passeport mathématique

En effet, le modèle de la courbe logistique présente une **fonction « S-shape »** (courbe en S-sigmoïde) plutôt qu'une droite comme dans le modèle de régression linéaire. Cette courbe en S permet de n'obtenir que des valeurs comprises entre 0 et 1 car elle est limitée par deux asymptotes horizontales. Ainsi toutes les valeurs ont un sens.

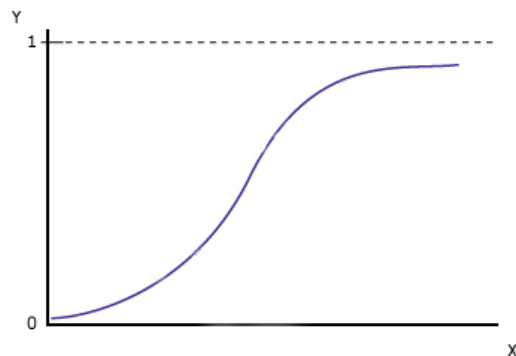


Figure 5 : Graphique d'une courbe en « S »

Les conditions d'utilisation du modèle sont les suivantes : les variables à expliquer doivent faire partie de classes exhaustives et mutuellement exclusives et une absence de multicollinéarité entre les variables explicatives est nécessaire.

Le modèle **Logit** (de régression logistique) possède néanmoins quelques désavantages. Ce n'est pas un modèle de régression linéaire donc nous ne pouvons pas l'estimer par la



méthode des moindres carrés ordinaires. En outre, les paramètres du modèle Logit sont difficiles à interpréter de façon directe en comparaison avec ceux d'une régression linéaire.

Ici, pour nos analyses, la méthode de régression logistique choisie est la **Forward Stepwise** (Likelihood Ratio). Avec cette méthode de sélection progressive ascendante pas à pas (ou régression par sélection), à chaque étape du modèle, le logiciel SPSS fait entrer dans l'équation la variable la plus significative (la plus corrélée avec la variable dépendante).

En outre, une variante du Logit est le modèle **Probit** (ou **Normit**). Cette méthode est très proche du Logit mais utilise une fonction de répartition de la loi normale centrée réduite pour calculer les probabilités.

2.3.1 Régression logistique simple

En plus des deux autres ouvrages de référence cités précédemment, la théorie de cette partie s'inspire de :

- TENENHAUS M. (2007), « *Méthodes pour décrire, expliquer et prévoir* », Paris, Ed. Dunod

Tout d'abord, nous posons $P(y = 1) = \pi(x)$: la probabilité d'occurrence de l'évènement à expliquer (comprise entre 0 et 1). Pour la régression logistique simple cette probabilité est définie par la fonction suivante :

$$\pi(x) = P(y = 1) = \frac{1}{e^{-(\beta_0 + \beta_1 x)} + 1} = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

De là, découle que :

$$1 - \pi(x) = \frac{1}{1 + e^{(\beta_0 + \beta_1 x)}}$$

et

$$\frac{\pi(x)}{1 - \pi(x)} = e^{\beta_0 + \beta_1 x}$$



$\frac{\pi(x)}{1 - \pi(x)}$ est appelé le **odds ratio** qui est le ratio de la réalisation de l'évènement par la non réalisation de l'évènement.

Le modèle prend ensuite le logarithme népérien du odds ratio (log odds = **logit**) :

$$\log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x$$

Ainsi la relation entre la probabilité $\pi(x)$ et les variables indépendantes est non linéaire tandis que la relation entre le logit et les variables indépendantes est linéaire.

a. Estimation des paramètres

Le modèle Logit peut s'estimer grâce à la méthode du **maximum de vraisemblance**. Cette méthode consiste à chercher b_0 et b_1 en maximisant la vraisemblance des paramètres. Ainsi, nous définissons tout d'abord une **vraisemblance** V :

$$V = \prod_{i=1}^n \pi(x_i)^{y_i} * [1 - \pi(x_i)]^{(1-y_i)}$$

Ensuite, nous prenons le logarithme de la vraisemblance que nous désirons maximiser. La maximisation du logarithme de vraisemblance (log likelihood) est reprise dans la formule :

$$\text{Max } \ln(V) = \sum_{i=1}^n [y_i * \ln(\pi(x_i)) + (1 - y_i) * \ln(1 - \pi(x_i))]$$

Nous annulons les dérivées partielles de la fonction pour la maximiser et ainsi estimer β_0 et β_1 . Les conditions sont les suivantes :

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \text{ et } \sum_{i=1}^n [x_i (y_i - \pi(x_i))] = 0$$

Il faut également considérer les conditions de second ordre. Dans notre cas, la dérivée doit être négative pour obtenir un maximum.



Utilisons un exemple chiffré pour expliquer le modèle logistique simple. Dans celui-ci, la variable dépendante binaire est **Reussite_annee** et la variable indépendante est **Passeport_moy_20**. Réalisons cette régression pour les étudiants de sciences de gestion et ingénierat de gestion (**Orient** = 1) avec la base de données de l'année 2008-2009.

Le tableau ci-dessous, nous donne les coefficients de la constante et de la variable explicative.

Variables in the Equation						
		B	S.E.	Wald	Df	Sig.
Step 1 ^a	Passeport_moy_20	,365	,075	23,436	1	,000
	Constant	-5,325	1,062	25,143	1	,000

a. Variable(s) entered on step 1: Passeport_moy_20.

Tableau 21 : Tableau SPSS – Coefficients des variables de la régression logistique simple

Dès lors, nous pouvons écrire le logit :

$$\log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x = -5,325 + 0,365 * x =$$

$$-5,325 + 0,365 * (\text{Passeport_moy_20})$$

Par exemple, si l'étudiant a obtenu 15/20 au passeport mathématique :

$$\log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = -5,325 + 0,365 * 15 = 0,15$$

$$e^{0,15} = 1,161834243$$

$$\pi(x) = P(y = 1) = 1,161834243 / (1 + 1,161834243) = 0,5374$$

La probabilité que l'évènement se réalise (que cet étudiant réussisse) est de 0,5374.

b. Significativité des paramètres estimés

Les hypothèses que nous cherchons à accepter ou réfuter à partir de ces tests sont les suivantes :

- $H_0 : \beta_1 = 0$
- $H_1 : \beta_1 \neq 0$



Deux tests ont été réalisés sur les paramètres : le test de Wald et celui du rapport de vraisemblance.

- Test de Wald

Le **test de Wald** (dans le tableau 21) - test sur les paramètres β_0 et β_1 (comme les tests de Student) - se calcule avec la formule suivante :

$$\frac{b_1^2}{s_{b_1}^2} = \frac{(0,365)^2}{(0,075)^2} = 23,68$$

La statistique de Wald a pour but de tester la contribution individuelle des différentes variables. Cette statistique est supérieure à la valeur du $\chi^2_{1-\alpha, \nu=1}$ trouvée dans la table (ν étant le nombre de degrés de liberté). En outre, nous voyons que la significativité est de 0,000, inférieure au seuil des 5%. Le paramètre β_1 est donc bien différent de 0.

- Test du rapport de vraisemblance

Le **test du rapport de vraisemblance** est également un test sur le paramètre β_1 . Le premier tableau (tableau 22) reprend la valeur de -2 Log Likelihood pour le modèle complet qui comprend la constante et la variable **Passeport_moy_20** tandis que le tableau 23 présente la valeur du modèle incluant uniquement la constante.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	156,777 ^a	,220	,298

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Tableau 22 : Tableau SPSS – Tableau résumé pour le test du rapport de vraisemblance pour la régression logistique simple



Model Summary

Iteration		-2 Log likelihood
Step 0	1	192,319

Tableau 23 : Tableau SPSS – Test du rapport de vraisemblance pour la régression logistique simple

La formule du **ratio de vraisemblance** (Likelihood ratio) est la suivante :

$$-2 * \log(\text{vraisemblance}) = -2LL$$

$$\chi^2 = [-2\log(V(\beta_0, \beta_1)) - [-2\log(V(\beta_0))] = -156,777 - (-192,319) = 35,54$$

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Model	35,541	1	,000

Tableau 24 : Tableau SPSS - Test du rapport de vraisemblance pour la régression logistique simple

Ici, nous testons l'hypothèse $H_0 : \beta_1 = 0$. La significativité étant inférieure à 5%, nous pouvons conclure que le β_1 est différent de 0, ce qui confirme le test de Wald.

Pour expliquer la valeur 192,319 (tableau 23) qui est la vraisemblance quand nous considérons le modèle uniquement avec la constante, nous pouvons utiliser la formule suivante :

$$n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n) = 86 * \ln(86) + 57 * \ln(57) - 143 * \ln(143) = -96,1589$$

et

$$-2 * \log(V(\beta_0)) = 2 * 96,1589 = 192,31$$



Nous retrouvons également la valeur de - 96,15 dans un tableau SPSS calculant le Log Likelihood (log de vraisemblance) uniquement pour la constante.

Model if Term Removed		
Variable		Model Log Likelihood
Step 1	Passeport_moy_20	-96,159

Tableau 25 : Tableau SPSS – Vraisemblance

c. Force de liaison

Pour mesurer la force de liaison des variables, nous pouvons utiliser différentes méthodes comme par exemple le R^2 de Cox & Snell ou le R^2 de Nagelkerke.

- R^2 de Cox & Snell

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	156,777 ^a	,220	,298

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Tableau 26 : Tableau SPSS – R^2 de Cox&Snell

La formule du **R^2 de Cox&Snell** est la suivante :

$$R^2_{Cox\&Snell} = 1 - \left[\frac{V(\beta_0)}{V(\beta_0, \beta_1)} \right]^{\frac{2}{n}} = 1 - e^{\left[\left(\frac{2}{143} \right) * \left(\frac{192,319}{-2} \right) - \left(\frac{2}{143} \right) * \left(\frac{156,777}{-2} \right) \right]} = 0,220$$

Cette valeur est donc une mesure de la force de liaison entre les variables y et x. Plus cette valeur est élevée plus l'ajustement est bon.

- R^2 de Nagelkerke

$$R^2_{Nagel\ ker\ ke} = \frac{R^2_{Cox\&Snell}}{Max(R^2_{Cox\&Snell})} = \frac{0,22}{1 - \left[V(\beta_0) \right]^{\frac{2}{n}}} = \frac{0,22}{1 - e^{\left[\left(\frac{2}{143} \right) * \left(\frac{192,319}{-2} \right) \right]}} = 0,298$$



La signification du **R² de Nagelkerke** est proche de celle du coefficient de détermination (R²) (voir le paragraphe sur la régression linéaire). Par exemple, un R² de Nagelkerke de 0,298 signifie que 29,8% de la variation de la réussite est expliquée par les facteurs conservés à la dernière étape du modèle (ici, les points au passeport mathématique).

d. Pouvoir discriminant du modèle

Le **tableau de classification** de la variable à expliquer est le suivant :

Classification Table ^a					
Observed			Predicted		
			Reussite_annee		Percentage Correct
			0	1	
Step 1	Reussite_annee	0	72	14	83,7
		1	28	29	50,9
Overall Percentage					70,6

a. The cut value is ,500

Tableau 27 : Tableau SPSS – Tableau de classification pour la régression logistique simple

Ce tableau reprend les taux de classement. En effet, les prédictions peuvent être différentes de la réalité de notre base de données. Ici, pour notre exemple, 72 étudiants ont été prédits en échec et sont réellement en échec. Donc 72 étudiants ont correctement été prédits en échec. Ainsi, parmi tous ceux qui ont raté, 83,7% [72 / (72+14)] des cas ont bien été prédits et nous pouvons noter que ce taux est élevé. Au niveau de la réussite, le taux de prédiction est un peu plus bas (50,9%). Cela signifie que sur les 57 étudiants (28 + 29) qui ont réellement réussi, nous en avons bien prédit 29. Dans l'annexe 3 sont reprises les 143 données de la base de données et nous avons détaillé la construction pas à pas du Logit et ensuite des probabilités observées.



2.3.2 Régression logistique multiple

La régression logistique multiple est semblable à la régression logistique simple. La seule différence est le nombre de variables explicatives supérieur ou égal à 2. Le modèle peut s'écrire :

$$\log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K$$

Avec une probabilité d'occurrence de l'évènement à expliquer (comprise entre 0 et 1) qui s'écrit :

$$\pi(x) = P(y = 1) = \frac{1}{e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K)} + 1} = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K)}}$$

Le **odds ratio** vaut donc :

$$\frac{\pi(x)}{1 - \pi(x)} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K}$$

La manière d'estimer les paramètres et les tests que nous pouvons ensuite effectuer sont les mêmes que ceux expliqués précédemment pour la régression linéaire simple.

Nous avons pris comme exemple la même variable à expliquer binaire **Reussite_annee** et comme variables explicatives **Passeport_moy_20** et **Pourc**, le pourcentage obtenu par l'étudiant en fin de secondaire réparti selon trois catégories :

- 1 : moins de 65% ;
- 2 : entre 65% et 75% ;
- 3 : plus de 75%.

Nous voyons dans le tableau suivant qu'il y a deux étapes. A la première étape, la variable la plus significative est **Passeport_moy_20** et ensuite la variable **Pourc** entre également dans le modèle.



Nous obtenons les estimations suivantes pour les paramètres :

		Variables in the Equation					
		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Passeport_moy_20	,354	,075	22,254	1	,000	1,425
	Constant	-5,190	1,061	23,921	1	,000	,006
Step 2 ^b	Passeport_moy_20	,337	,076	19,608	1	,000	1,400
	Pourc	,878	,352	6,217	1	,013	2,406
	Constant	-6,724	1,287	27,286	1	,000	,001

a. Variable(s) entered on step 1: Passeport_moy_20.

b. Variable(s) entered on step 2: Pourc.

Tableau 28 : Tableau SPSS – Coefficients de la régression logistique multiple

Ainsi, à la deuxième étape, nous voyons que les deux variables **Pourc** et **Passeport_moy_20** sont significatives. L'équation peut donc s'écrire :

$$\log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = -6,724 + 0,337 * x_1 + 0,878 * x_2$$

Prenons un exemple, si un étudiant a 16/20 au passeport mathématique et entre 65% et 75% à la fin des secondaires (Pourc = 2) :

$$\log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = -6,724 + 0,337 * 16 + 0,878 * 2 = 0,424$$

$$e^{0,424} = 1,52806$$

$$\pi(x) = P(y = 1) = 1,52806 / (1 + 1,52806) = 0,6044$$

La probabilité que l'évènement se réalise (que cet étudiant réussisse) est de 60,44%.

Nous constatons aussi une amélioration dans l'ajustement du modèle. Par exemple, le R^2 de Nagelkerke à la deuxième étape est meilleur que celui de la régression logistique simple qui était de 0,298.



Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	154,979 ^a	,213	,287
2	148,238 ^a	,250	,337

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Tableau 29 : Tableau SPSS – Tableau résumé de la régression logistique multiple

Section 4 : Conclusion

Cette partie plus théorique s'est donc basée sur des exemples didactiques pour expliquer chaque tableau réalisé par le logiciel SPSS. Ainsi, les résultats trouvés par la suite auront une signification et le lecteur pourra revenir à cette partie plus théorique pour comprendre l'origine de chaque résultat.



Quatrième Chapitre : Présentation des données

Cette partie assez courte a pour but, tout d'abord, la présentation des données et de leur collecte ainsi que l'organisation pratique de la base de données. Ensuite, une analyse descriptive des données est proposée. Le lecteur peut ainsi se faire une idée de l'évolution des données sur deux ou trois ans.

Section 1 : Construction de la base de données

La base de données peut être divisée en deux parties [voir Annexe 1 et Annexe 2]. L'une reprend les étudiants des FUCaM qui étaient en première année de bachelier au cours de l'année académique **2008-2009** et l'autre contenant les étudiants en première année de bachelier durant l'année académique **2009-2010**. Ces données proviennent de différentes sources. La première est une fiche sociologique remplie début octobre lors du passeport mathématique⁸. Cette fiche contient des informations sur l'âge, l'orientation, le fait d'avoir doublé une année en secondaire, le nombre d'heures de mathématique suivies en secondaire. Tous les étudiants de première année de bachelier (donc les 5 orientations proposées aux FUCaM) ont passé ce passeport. Une seconde source rassemble les données concernant les personnes présentes aux cours d'été. Ces cours sont « destinés à préparer l'étudiant aux études universitaires. Le programme de ces séances consiste en une révision des connaissances en mathématique, en une formation aux techniques de prise de notes et de gestion du temps ainsi qu'en deux cours de BAC1⁹ ». Une troisième source provient des résultats des sessions de juin et/ou de septembre et de données transmises par l'administration des études.

⁸ Le passeport mathématique a pour but d'évaluer les prérequis en mathématique et de mettre en place, par la suite, des séances de monitorat si cela s'avère nécessaire.

<http://www.fucam.ac.be/view.php3?include=100466&pere=21376&print=1>, page consultée le 17/03/2011, html

⁹ <http://www.fucam.ac.be/view.php3?include=1890&pere=20955&print=0>, page consultée le 17/03/2011, html



Les variables de la base de données peuvent être réparties en trois catégories : le profil de l'étudiant, le parcours scolaire de l'étudiant avant son entrée à l'université et le parcours scolaire de l'étudiant universitaire.

Les différentes variables utilisées lors des analyses statistiques sont présentées dans les tableaux ci-dessous.

1. Profil de l'étudiant

Sexe	Genre de l'étudiant	Variable qualitative nominale (binaire)	1 : Masculin 2 : Féminin
Annee_naissance	Année de naissance	Variable quantitative échelle de rapport	
Age	Age de l'étudiant	Variable quantitative échelle de rapport	
Mere	Dernier diplôme obtenu par la mère de l'étudiant	Variable qualitative nominale	1 : Néant 2 : Primaire 3 : Secondaire inférieur 4 : Secondaire supérieur 5 : Supérieur 6 : Universitaire
Mere_regroup	Regroupement des catégories	Variable qualitative nominale	1 : Regroupement des catégories néant, primaire et secondaire inférieur 2 : Catégorie secondaire supérieur 3 : Regroupement des catégories supérieur et universitaire

Pere	Dernier diplôme obtenu par le père de l'étudiant	Variable qualitative nominale	1 : Néant 2 : Primaire 3 : Secondaire inférieur 4 : Secondaire supérieur 5 : Supérieur 6 : Universitaire
Pere_regroup	Regroupement des catégories	Variable qualitative nominale	1 : Regroupement des catégories néant, primaire et secondaire inférieur 2 : Catégorie secondaire supérieur 3 : Regroupement des catégories supérieur et universitaire
Bourse	L'étudiant a-t-il une bourse ?	Variable qualitative nominale (binaire)	0 : Non 1 : Oui
Kot	L'étudiant a-t-il un kot sur le campus ?	Variable qualitative nominale (binaire)	0 : Non 1 : Oui

Tableau 30 : Tableau de présentation des données – Profil de l'étudiant

2. Parcours scolaire de l'étudiant avant l'entrée à l'université

Filiere	Filière que l'étudiant a suivie durant ses études secondaires	Variable qualitative nominale (binaire)	1 : Général 2 : Autre (technique ou artistique)
----------------	---	---	--

Heures_math	Heures de cours de mathématique suivies en secondaires par semaine	Variable quantitative échelle de rapport	0 : 0h 1 : 1h 2 : 2h ... 10 : > 9h
math6h	L'étudiant a-t-il suivi au moins 6h de mathématique durant les secondaires ?	Variable qualitative nominale (binaire)	0 : Non 1 : Oui
Pourc	Pourcentage obtenu par l'étudiant à la sortie des études secondaires	Variable qualitative ordinaire	1 : < 65% 2 : 65% à 75% 3 : > 75%
Sec	Année où l'étudiant a terminé ses études secondaires	Variable qualitative nominale	1 : n (année d'inscription en BAC1 = année de fin de secondaire) 2 : n - 1 3 : n - 2 4 : n - 3 5 : Avant (n - 3) 6 : Pas diplômé
Redoubl	L'étudiant a doublé une année dans le secondaire	Variable qualitative nominale (binaire)	1 : Non 2 : Oui
Cours_ete	L'étudiant a-t-il suivi les cours d'été proposés aux FUCaM ?	Variable qualitative nominale (binaire)	0 : Non 1 : Oui

Tableau 31 : Tableau de présentation des données – parcours scolaire de l'étudiant avant son entrée à l'université



3. Parcours scolaire de l'étudiant universitaire

Orient	Orientation choisie par l'étudiant	Variable qualitative nominale	21 : Ingénieur de gestion 22 : Sciences de gestion 23 : Sciences politiques 24 : Sciences humaines et sociales 25 : Information et communication
Option	Regroupement des orientations	Variable qualitative nominale (binaire)	1 : Regroupement ingénieur de gestion et sciences de gestion 2 : Regroupement sciences politiques, sciences humaines et sociales et information et communication
Bisseur	L'étudiant recommence (bisse) son année	Variable qualitative nominale (binaire)	1 : Non 2 : Oui
Passeport_moy_20	Cote obtenue par l'étudiant lors du passeport mathématique (/20)	Variable quantitative échelle de rapport	
Reussite_annee	Réussite de l'étudiant de sa première année de bachelier	Variable qualitative nominale (binaire)	0 : Non 1 : Oui



Moyenne_global_100	Cote maximale obtenue par l'étudiant soit en juin soit en septembre (/100)	Variable quantitative échelle de rapport	
---------------------------	--	--	--

Tableau 32 : Tableau de présentation des données – Parcours scolaire de l'étudiant universitaire

Section 2 : Présentation descriptive des données

Dans cette partie sont repris, sur graphiques et par tableaux, des fréquences et résultats qui montrent l'évolution de différentes variables sur trois ans. Cette analyse statistique permet, entre autres, d'avoir une idée de l'évolution des données.

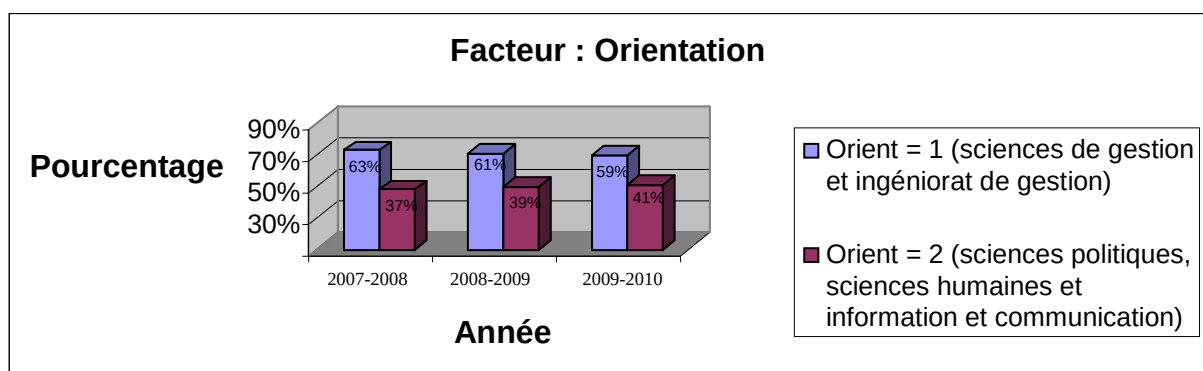


Figure 6 : Evolution du facteur Orientation

Sur ce graphique, nous voyons la répartition du nombre d'étudiants par orientation. En moyenne, 60% des étudiants prennent les options sciences de gestion et ingénierat de gestion, 20% sciences politiques et 20% sciences humaines ou information et communication.

1. Pour les sciences et ingéniorat de gestion (Orient = 1)

1.1 Nombre de filles et de garçons

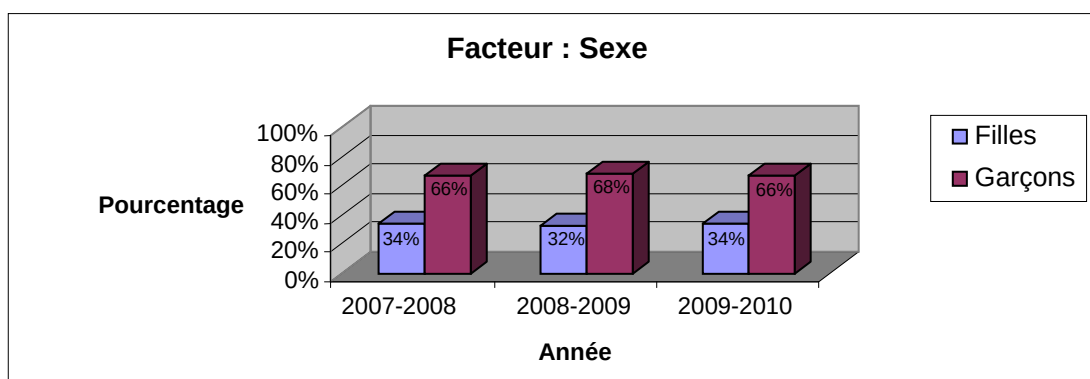


Figure 7 : Evolution du facteur Sexe

Sur ce graphique, nous voyons clairement qu'il y a un nombre supérieur de garçons inscrits en BAC1 pour l'orientation sciences de gestion et ingéniorat de gestion et cela, pour les trois années.

1.2 Age

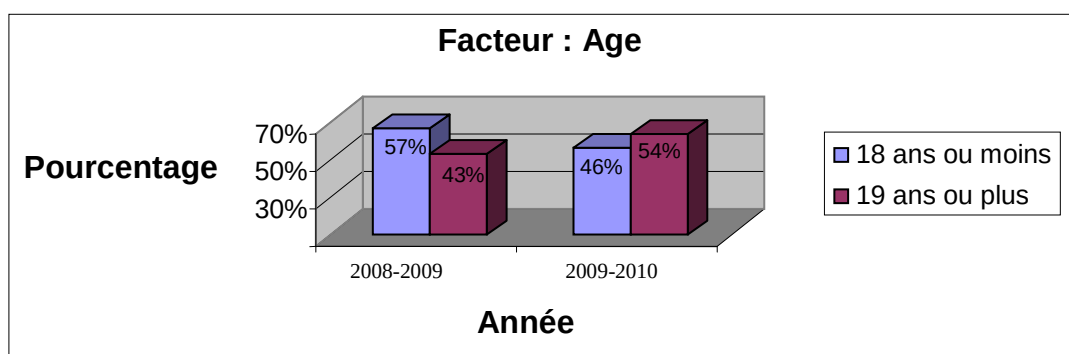


Figure 8 : Evolution du facteur Age

Nous voyons qu'en 2009-2010, un peu plus de la moitié des étudiants avaient plus de 18 ans en BAC1.

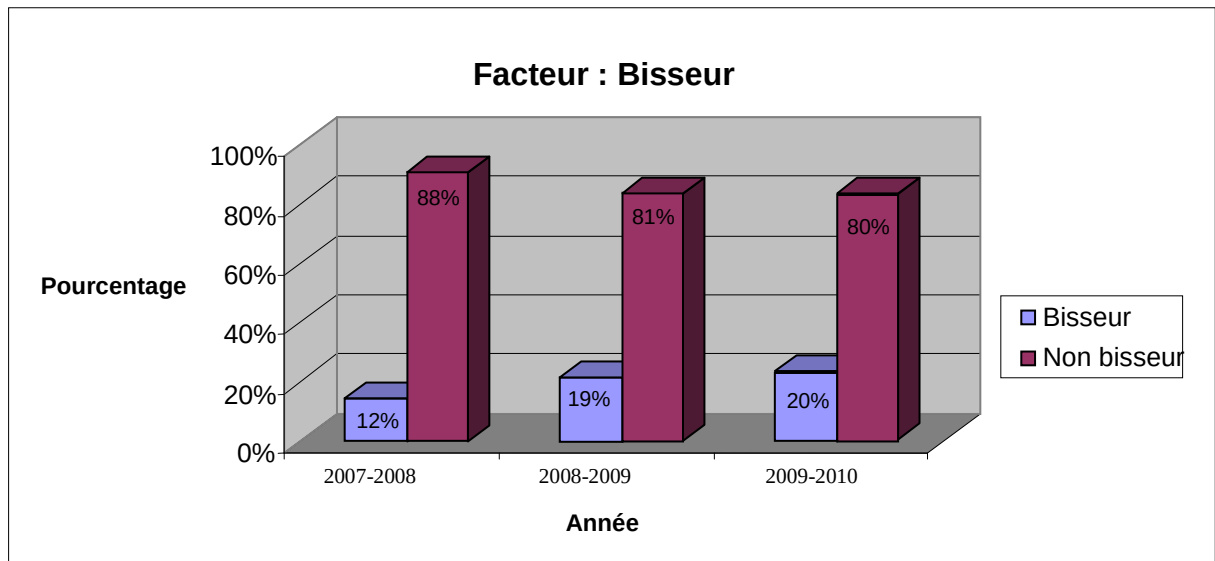


Figure 9 : Evolution du facteur Bisseur

En 2009-2010, il y avait 20% d'étudiants qui recommençaient leur première année de bachelier.

1.4 Cours d'été

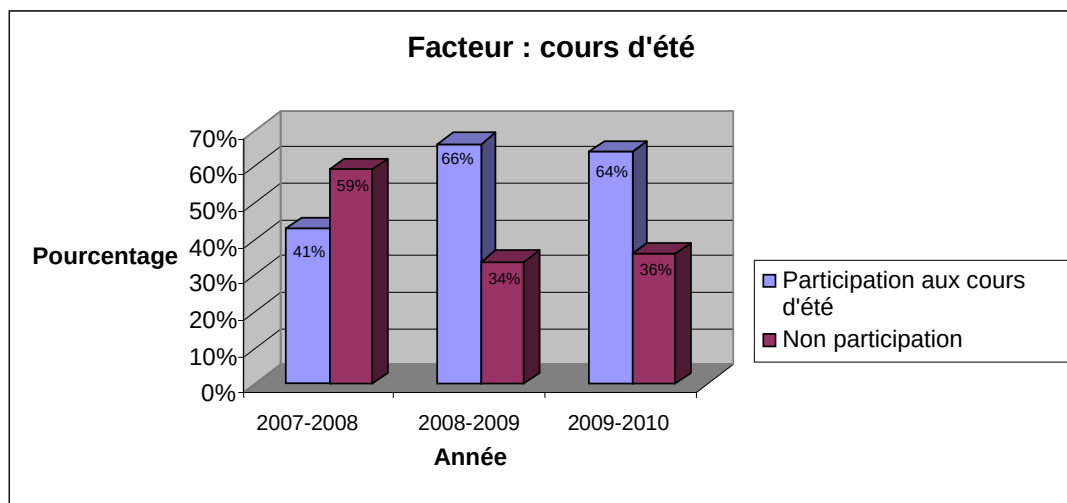


Figure 10 : Evolution du facteur cours d'été



La participation aux cours préparatoires d'été a fortement évolué : la participation est passée de 40% des étudiants à, en moyenne, 65% de participation pour les années 2008-2009 et 2009-2010.

1.5 Heures de mathématique

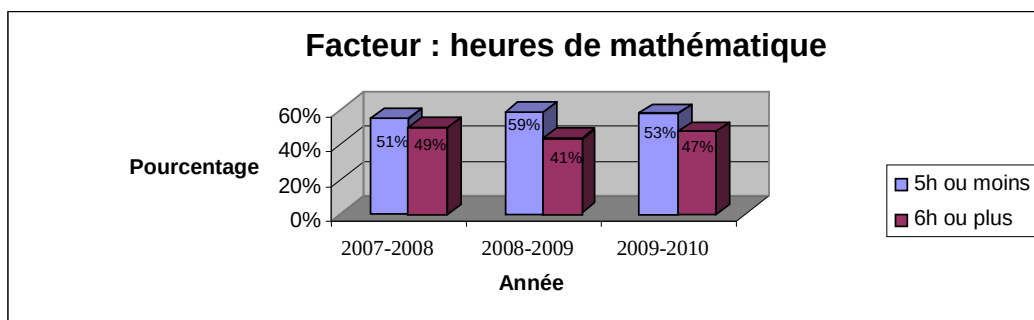


Figure 11 : Evolution du facteur heures de mathématique

La proportion n'évolue pas sensiblement sur les trois années.

2. Pour les sciences humaines, sciences politiques et information et communication (Orient = 2)

2.1 Nombre de filles et de garçons

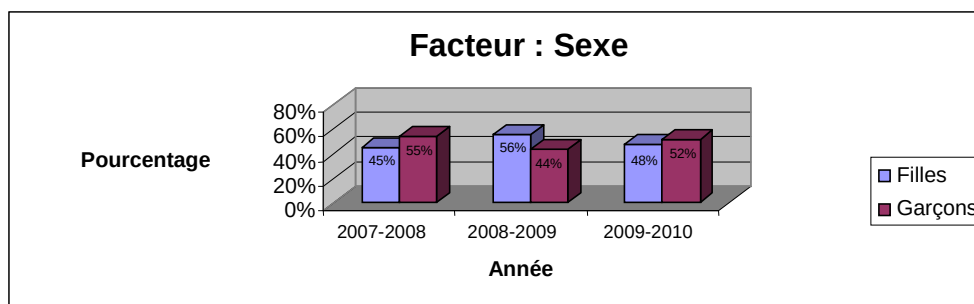


Figure 12 : Evolution du facteur Sexe

Pour les années 2007-2008 et 2009-2010, il y a plus de garçons que de filles pour les options sciences politiques, sciences humaines et information et communication même si la proportion filles-garçons reste quand même proche des 50-50.

2.2 Age

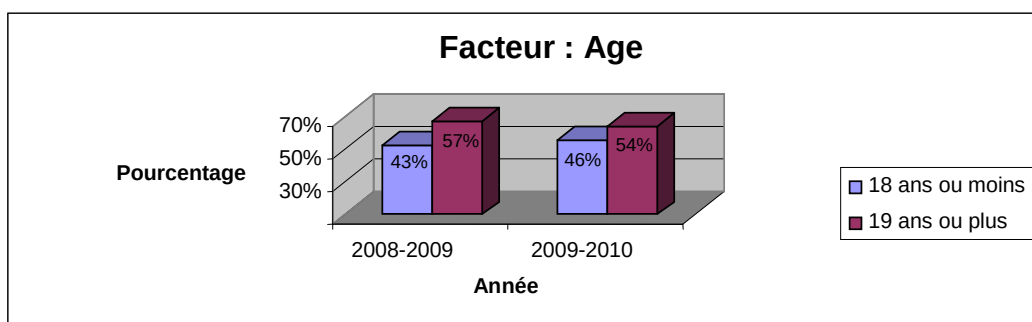


Figure 13 : Evolution du facteur Age

Nous voyons qu'il y a plus d'étudiants qui ont plus de 19 ans que d'étudiants ayant 18 ans ou moins et cela, pour les deux années.

2.3 Bisseur

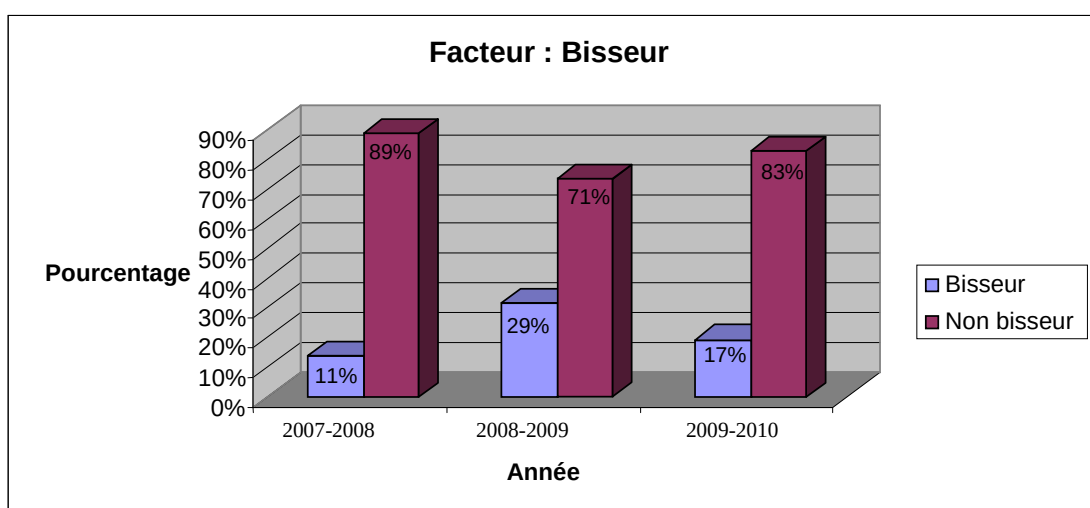


Figure 14 : Evolution du facteur Bisseur

Le pourcentage d'étudiants qui bissent leur BAC1 est faible : même s'il passe de 11% à 29% entre 2007-2008 et 2008-2009, il retombe à 17% en 2009-2010.



2.4 Cours d'été

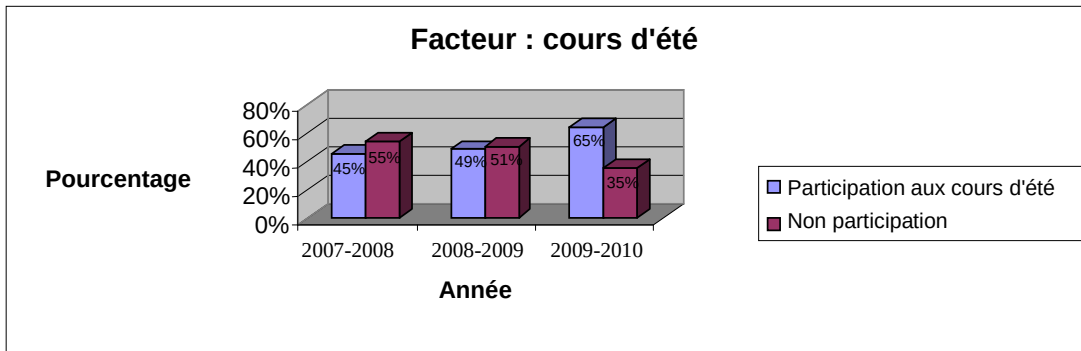


Figure 15 : Evolution du facteur cours d'été

Tout comme pour les sciences de gestion et ingénieur de gestion, le nombre de participants aux cours d'été augmente avec les années.

Pour conclure cette partie, nous voyons des évolutions dans les données sans pouvoir conclure à un changement radical d'une tendance. Peut-être le facteur « cours d'été » est-il le critère le plus encourageant dans son changement car nous voyons un réel intérêt des étudiants pour cette activité.



Cinquième chapitre : Résultats des analyses statistiques

Maintenant que les différentes méthodes d'investigations statistiques sont connues grâce au chapitre 3 et que la base de données a été précisée dans le chapitre 4, il nous reste à appliquer ces méthodes à notre base de données pour l'analyse pratique des résultats. Dans cette cinquième partie, nous présenterons donc les résultats en ayant soin de diviser ceux-ci en deux selon l'orientation choisie par l'étudiant. Ainsi, nous aurons deux subdivisions : les résultats pour les sciences et ingénierat de gestion et ceux pour les sciences politiques, humaines et information et communication. Tous les tableaux ayant permis d'amener aux résultats et conclusions sont repris dans les annexes par orientation, titre des paragraphes et années.

Section 1 : Pour les sciences et ingénierat de gestion (Orient = 1)

1. Facteurs influençant la moyenne globale obtenue en fin de BAC1 (Moyenne_global_100)

La variable à expliquer **Moyenne_global_100** reprend la cote maximale obtenue par l'étudiant en juin ou en septembre (/100).

1.1 Comparaison des moyennes

1.1.1 Genre de l'étudiant (Sexe)

Une comparaison des moyennes des résultats de fin de 1^{ère} année (**Moyenne_global_100**) en fonction du genre de l'étudiant (**Sexe**) est effectuée grâce à un test en T de Student.



- Pour 2008-2009 [Annexe 4]

Group Statistics

	Sexe	N	Mean	Std. Deviation	Std. Error Mean
Moyenne_global_100	1	124	34,736	29,458	2,645
	2	58	45,944	28,365	3,724

Tableau 33 : Tableau SPSS – Statistiques descriptives de la variable
Moyenne_global_100 pour l'année 2008-2009

A partir de ce tableau où la variable **Sexe** = 1 pour les garçons et **Sexe** = 2 pour les filles, nous constatons que les moyennes sont éloignées selon le genre de l'étudiant. De façon intuitive, il semble donc y avoir une différence entre les résultats de fin de premier bachelier des filles et des garçons en faveur des filles. Mais cela doit être vérifié grâce au test en t de Student (test d'égalité des moyennes).

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Moyenne_global_100	Equal variances assumed	2,751	,099	-2,42	180	,017	-11,208	4,632	-20,348	-2,068
	Equal variances not assumed			-2,453	115,41	,016	-11,208	4,568	-20,257	-2,159

Tableau 34 : Tableau SPSS - Test d'indépendance des moyennes pour la variable
Moyenne_global_100 pour l'année 2008-2009

La **statistique T** du test d'égalité des moyennes vaut -2,420 avec une probabilité critique associée de 0,017 qui est inférieure aux 5% fixés.



Ainsi, pour l'année 2008-2009 et pour les options sciences de gestion et ingénierat de gestion, l'hypothèse H_0 d'égalité des moyennes est rejetée. Il existe donc une différence statistiquement significative entre les résultats en fin de première année de bachelier chez les filles et chez les garçons en faveur des filles.

- Pour 2009-2010 [Annexe 5]

Pour l'année 2009-2010, les moyennes sont de 36,595% pour les garçons et de 46,339% pour les filles. A première vue, la différence semble significative. Grâce au **test de Levene**, nous acceptons l'égalité des variances et voyons que la statistique de Student est de -2,196 avec une probabilité critique de 0,029.

Pour l'année 2009-2010, il existe donc une différence statistiquement significative entre les résultats en fin de première année de bachelier chez les filles et chez les garçons.

- Conclusion

Le facteur « **genre de l'étudiant** » est un critère significatif pour expliquer la moyenne globale des points obtenus en fin de première année de bachelier.

1.1.2 Nombre d'heures de mathématique suivies pendant les deux dernières années du secondaire (math6h)

Nous avons réalisé une comparaison des moyennes des résultats de fin de 1^{ère} année (**Moyenne_global_100**) en fonction de la variable **math6h**.

- Pour 2008-2009 [Annexe 6]

Grâce aux tests réalisés avec le logiciel SPSS, nous pouvons dire qu'il existe une différence significative entre les moyennes des résultats globaux de BAC1 des étudiants ayant



suivi six heures de mathématique ou plus dans le secondaire et les autres. Cette différence étant, en faveur des étudiants ayant un plus grand nombre d'heures de mathématique.

Ainsi, pour l'année 2008-2009, il existe une différence statistiquement significative entre les résultats en fin de première année de bachelier en faveur des étudiants ayant suivi six heures de mathématique ou plus dans le secondaire.

- Pour 2009-2010 [Annexe 7]

En ce qui concerne l'année 2009-2010, les résultats vont dans le même sens qu'en 2008-2009.

- Conclusion

Tout comme, le genre de l'étudiant, le facteur « **math6h** » est un critère significatif pour expliquer la moyenne globale des points obtenus en fin de première année de bachelier.

1.1.3 Etudiant bisseur (Bisseur)

La variable « **Bisseur** » est une variable binaire qui prend la valeur 2 si l'étudiant recommence sa première année de bachelier et sinon la valeur 1. Ce test de comparaison des moyennes a pour but de comparer les moyennes des résultats de fin de 1^{ère} année (**Moyenne_global_100**) pour les étudiants qui recommencent leur année (**Bisseur**) ou ceux qui suivent les cours pour la première fois (non bisseur).

- Pour 2008-2009 [Annexe 8]

Alors que la revue de la littérature nous dit que les étudiants qui doublaient leur année en 2003-2004 dans la même orientation réussissaient mieux que la moyenne (Van Campendhoudt & al., 2008), les résultats des tests pour notre base de données pour l'année



2008-2009 montrent le contraire. Il n'y aurait pas de différence significative entre les moyennes des résultats de fin de BAC1 des étudiants « bisseurs » et des « non bisseurs ».

Ainsi, pour l'année 2008-2009, il n'existe pas de différence significative entre les résultats en fin de première année de bachelier des étudiants redoublants et de ceux suivant les cours pour la première fois.

- Pour 2009-2010 [Annexe 9]

Pour l'année 2009-2010 par contre, la différence des moyennes des résultats en fin de BAC1 est significative. Ce qui rejoint les conclusions de la revue de la littérature.

Pour l'année 2009-2010 et les options sciences de gestion et ingénierat de gestion, l'hypothèse H_0 d'égalité des moyennes est rejetée. Il existe donc une différence statistiquement significative entre les résultats en fin de première année de bachelier des étudiants ayant doublé et des autres. Cette différence est en faveur des étudiants « bisseurs ».

1.1.4 Participation aux cours d'été (Cours_ete)

La moyenne des résultats globaux en fin de BAC1 est-elle différente si l'étudiant a participé aux cours d'été proposés par les FUCaM début septembre de celle d'un étudiant qui n'y aurait pas pris part ?

- Pour 2008-2009 et 2009-2010 [Annexes 10 et 11]

Dans les résultats du logiciel SPSS, nous voyons qu'il y a une différence dans les moyennes en faveur des étudiants qui ont participé aux cours d'été.

Pour les années 2008-2009 et 2009-2010 et pour les options sciences de gestion et ingénierat de gestion, il existe donc une différence significative entre les résultats en fin de première année de bachelier des étudiants ayant participé aux cours d'été et les autres.



- Conclusion

Le critère « cours d'été », à savoir la participation des étudiants début septembre à la semaine propédeutique pourrait être un critère significatif pour expliquer la moyenne globale des points obtenus en fin de première année de bachelier. Selon moi, cela pourrait s'expliquer également par la motivation que les étudiants qui suivent les cours d'été ont, avant même l'entrée à l'université (cela étant détaillé plus amplement dans l'aperçu de la littérature).

1.1.5 Redoublement durant les années secondaires (Redoubl)

Le fait d'avoir doublé durant les années secondaires aurait-il des conséquences sur la moyenne des étudiants en fin de BAC1 ? Le test-t de Student compare les moyennes de deux groupes d'étudiants.

- Pour 2008-2009 [Annexe 12]

Nos résultats vont dans le même sens que ceux de la revue de la littérature. En effet, pour notre base de données reprenant l'année 2008-2009, il y a une différence significative dans les moyennes des résultats fin de BAC1 en faveur des étudiants n'ayant pas doublé d'année durant leurs études secondaires.

Ainsi, pour l'année 2008-2009 et pour les options sciences de gestion et ingénierat de gestion, nous pouvons voir qu'il existe une différence significative entre les résultats en fin de première année de bachelier des étudiants ayant doublé durant leurs études secondaires et les autres. Cette différence est en faveur des étudiants n'ayant pas doublé.

- Pour 2009-2010 [Annexe 13]

Les conclusions vont dans le même sens que l'année 2008-2009. Le critère du redoublement durant les années de secondaire est un facteur significatif pour expliquer la



moyenne globale des points obtenus en fin de première année de bachelier et cela en faveur des étudiants n'ayant pas doublé.

1.1.6 Pourcentage en fin de secondaire (Pourc)

Ce critère « **Pourc** » est divisé en trois catégories : les étudiants ayant moins de 65% (**Pourc** = 1), ceux ayant obtenu entre 65% et 75% (**Pourc** = 2) et ceux ayant plus de 75% (**Pourc** = 3). Nous utilisons dès lors une analyse de type Anova pour réaliser un test d'indépendance entre la variable résultats de fin de 1^{ère} année (**Moyenne_global_100**) et les résultats en fin de secondaire (**Pourc**).

- Pour 2008-2009 [Annexe 14]

Pour l'année 2008-2009, le test de l'analyse ANOVA est détaillé précédemment dans la partie théorique (voir le paragraphe sur les comparaisons de moyennes).

Ainsi, pour l'année 2008-2009, nous pouvons conclure qu'il existe des différences significatives entre les résultats obtenus en fin de première année de bachelier des étudiants ayant des pourcentages différents en fin de secondaire. Les étudiants ayant plus de 75% en fin de secondaire ont une moyenne supérieure en fin de BAC1 par rapport aux autres.

- Pour 2009-2010 [Annexe 15]

Dans l'analyse de la base de données 2009-2010. Le test de Levene assume l'égalité des variances. L'analyse Anova révèle qu'il y a des différences entre les résultats obtenus à la fin de la première année de bachelier en fonction du pourcentage obtenu à la fin des études secondaires. En outre, le test post-hoc de Scheffe montre que seuls les étudiants ayant un pourcentage compris entre 65% et 75% et ceux ayant obtenu plus de 75% en fin de secondaire ont une différence de moyenne significative.



Pour l'année 2009-2010 et pour les options sciences de gestion et ingéniorat de gestion, les étudiants ayant plus de 75% en fin de secondaire ont une moyenne significativement supérieure en fin de BAC1 par rapport à celle des étudiants ayant des résultats compris entre 65% et 75% en fin de secondaire.

1.1.7 Age

- Pour 2008-2009 [Annexe 16]

Ce test comprend deux groupes, les étudiants ayant 18 ans ou moins et les étudiants ayant 19 ans ou plus. Grâce aux tests SPSS, nous acceptons l'hypothèse des variances égales pour un α de 5% et nous rejetons l'hypothèse nulle d'égalité des moyennes.

Ainsi, l'hypothèse H_0 d'égalité des moyennes est rejetée. Il existe donc une différence statistiquement significative entre les résultats en fin de première année de bachelier pour les étudiants ayant 18 ans ou moins et ceux de 19 ans ou plus. Cette différence est en faveur de ceux ayant 18 ans ou moins.

- Pour 2009-2010 [Annexe 17]

Les conclusions sont semblables à celle de l'année 2008-2009.

1.1.8 Bourse

[Annexes 18 et 19]

Pour les années 2008-2009 et 2009-2010 et pour les options sciences de gestion et ingéniorat de gestion, il n'existe pas de différence significative entre les résultats en fin de première année de bachelier pour les étudiants ayant une bourse et les autres.



1.1.9 Kot

- Pour 2008-2009 [Annexe 20]

Pour l'année 2008-2009, l'hypothèse H_0 d'égalité des moyennes est rejetée. Il existe donc une différence statistiquement significative entre les résultats en fin de première année de bachelier pour les étudiants ayant un kot sur le campus et ceux qui n'en n'ont pas. Cette différence est en faveur de ceux ayant un kot.

- Pour 2009-2010 [Annexe 21]

Pour les années 2009-2010, il n'existe pas de différence significative entre les résultats en fin de première année de bachelier pour les étudiants ayant kot sur le campus et ceux qui n'en n'ont pas.

1.1.10 Tableau récapitulatif

Tableau des facteurs influençant la moyenne globale en fin de BAC1 pour les sciences et ingénierat de gestion :

	<u>2007-2008</u> (mémoire d'A. Schattelman)	<u>2008-2009</u>	<u>2009-2010</u>
Genre de l'étudiant	Différence en faveur des filles	Différence en faveur des filles	Différence en faveur des filles
Heures de mathématique en secondaire	Différence en faveur de ceux qui ont suivi au moins 6h	Différence en faveur de ceux qui ont suivi au moins 6h	Différence en faveur de ceux qui ont suivi au moins 6h
Cours d'été	Différence en faveur de ceux qui y ont participé	Différence en faveur de ceux qui y ont participé	Différence en faveur de ceux qui y ont participé



« Bisseur »	Pas de différence	Pas de différence	Différence en faveur des bisseurs
Redoublement dans le secondaire (Redoubl)	Non disponible	Différence en faveur de ceux qui n'ont pas doublé	Différence en faveur de ceux qui n'ont pas doublé
Pourcentage en fin de secondaire	Différence en faveur de ceux qui ont plus de 75% comparés aux autres	Différence en faveur de ceux qui ont plus de 75% comparés aux autres	Différence en faveur de ceux qui ont plus de 75% comparés à ceux qui ont entre 65% et 75%
Kot	Différence en faveur de ceux qui ont un kot	Différence en faveur de ceux qui ont un kot sur le campus	Pas de différence
Bourse	Pas de différence	Pas de différence	Pas de différence
Age	Non disponible	Différence en faveur de ceux qui ont 18 ans ou moins	Différence en faveur de ceux qui ont 18 ans ou moins

Tableau 35 : Tableau récapitulatif des facteurs influençant la moyenne globale en fin de BAC1 pour les années 2008-2009 et 2009-2010

La première colonne de ce tableau a été reprise du mémoire d'A. Schatteman. De ce tableau, nous pouvons conclure que les facteurs les plus liés à la moyenne en fin de BAC1 pour les sciences de gestion et ingénieur de gestion sont : le genre de l'étudiant, les heures de mathématique dans le secondaire, la participation aux cours d'été, le fait de ne pas avoir doublé durant le secondaire, le pourcentage en fin de secondaire et l'âge de l'étudiant. Le fait de posséder un kot sur le campus pourrait aussi être considéré comme relativement favorable aux résultats obtenus en fin de BAC1.



1.2 Corrélation linéaire

- Pour 2008-2009

La corrélation permet de voir si deux variables quantitatives sont liées l'une avec l'autre. Il faut cependant rester conscient que la corrélation entre deux variables ne démontre pas la présence d'un lien de causalité entre ces variables (Galand, 2005).

Correlations		Moyenne_global_100	Passeport_moy_20	heures_math
Moyenne_global_100	Pearson Correlation	1	,515**	,268**
	Sig. (2-tailed)		,000	,001
	N	183	143	165
Passeport_moy_20	Pearson Correlation	,515**	1	,486**
	Sig. (2-tailed)	,000		,000
	N	143	143	140
heures_math	Pearson Correlation	,268**	,486**	1
	Sig. (2-tailed)	,001	,000	
	N	165	140	167

** .Correlation is significant at the 0.01 level (2-tailed).

Tableau 36 : Tableau SPSS – Tableau de corrélations linéaires pour l'année 2008-2009

Dans le tableau ci-dessus, une relation positive forte ($0,5 < r < 0,8$) est montrée pour les variables **Passeport_moy_20** et **Moyenne_global_100** (0,515) et quasiment forte pour **Passeport_moy_20** et **heures_math** (0,486).



- Pour 2009-2010

Correlations		Moyenne_global_100	Passeport_moy_20	heures_math
Moyenne_global_100	Pearson Correlation	1	,507**	,237**
	Sig. (2-tailed)		,000	,003
	N	168	130	155
Passeport_moy_20	Pearson Correlation	,507**	1	,583**
	Sig. (2-tailed)	,000		,000
	N	130	130	124
heures_math	Pearson Correlation	,237**	,583**	1
	Sig. (2-tailed)	,003	,000	
	N	155	124	155

** . Correlation is significant at the 0.01 level (2-tailed).

Tableau 37 : Tableau SPSS – Tableau de corrélations linéaires pour l'année 2009-2010

Nous pouvons noter une relation positive forte ($0,5 < r < 0,8$) pour les variables **Passeport_moy_20** et **heures_math** (0,583) et pour la paire **Passeport_moy_20** et **Moyenne_globale_100** (0,507).

1.3 Régression linéaire multiple : méthode « pas à pas »

- 2008-2009

Cette méthode, pas à pas, permet de garder à chaque étape la variable la plus significative du modèle et d'ainsi, expliquer le mieux possible la variable dépendante. La méthode est itérative et elle s'arrête lorsque plus aucune variable n'est significative au seuil des 5%. Les variables introduites à la première étape sont : **Sexe**, **Filiere**, **Redoubl**, **Bisseur**, **Mere_regroup**, **Pere_regroup**, **Age**, **Pourc**, **heures_math**.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,237	8,624		,491	,624
	Heures_math	6,63	1,601	,338	4,14	,000
2	(Constant)	30,505	12,25		2,49	,014
	Heures_math	5,504	1,604	,280	3,432	,001
	Redoubl	-15,6	5,303	-,240	-2,94	,004
3	(Constant)	12,274	14,311		,858	,393
	Heures_math	5,301	1,579	,27	3,357	,001
	Redoubl	-14,814	5,224	-,228	-2,836	,005
	Pourc	9,124	3,869	,185	2,358	,020

a. Dependent Variable: Moyenne_global_100

Tableau 38 : Tableau SPSS – Tableau des coefficients pour la régression linéaire multiple

A la troisième étape, nous voyons que les trois variables choisies sont : **heures_math**, **Redoubl** et **Pourc**. Nous pouvons, dès lors, écrire l'équation linéaire de régression :

$$\hat{y} = 12,274 + 5,301 * x_1 + (-14,814) * x_2 + 9,124 * x_3$$

Par exemple, un étudiant qui a eu 6 heures de mathématique durant ses études secondaires, qui n'a pas doublé (**Redoubl** = 1) et qui a eu, en fin de secondaire, entre 65% et 75% (**Pourc** = 2) aura théoriquement une moyenne en fin de première BAC de :

$$\hat{y} = 12,274 + 5,301 * 6 + (-14,814) * 1 + 9,124 * 2 = 47,514\%$$



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,338 ^a	,114	,108	28,485
2	,411 ^b	,169	,156	27,699
3	,450 ^c	,203	,184	27,233

a. Predictors: (Constant), heures_math

b. Predictors: (Constant), heures_math, Redoubl

c. Predictors: (Constant), heures_math, Redoubl, Pourc

Tableau 39 : Tableau SPSS – Tableau de résumé du modèle pour la régression linéaire multiple

Le R^2 , à la troisième étape, est de 0,203. Nous pouvons dire que la variable **Moyenne_global_100** est expliquée à 20,3% par les trois variables explicatives.

- 2009-2010

Dans le tableau ci-dessous, nous voyons qu'à la quatrième étape, les variables **Redoubl, Bisseur, heures_math, Pourc** sont les plus significatives.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	63,135	6,634		9,517	,000
	Redoubl	-17,087	4,627	-,295	-3,693	,000
2	(Constant)	49,056	9,154		5,359	,000
	Redoubl	-17,248	4,567	-,298	-3,777	,000
	Bisseur	11,846	5,384	,174	2,200	,029
3	(Constant)	16,483	15,073		1,094	,276
	Redoubl	-14,013	4,629	-,242	-3,027	,003
	Bisseur	15,536	5,445	,228	2,853	,005
	heures_math	3,994	1,486	,222	2,687	,008
4	(Constant)	-4,692	17,749		-,264	,792
	Redoubl	-13,074	4,589	-,226	-2,849	,005
	Bisseur	17,582	5,455	,258	3,223	,002
	heures_math	4,189	1,470	,233	2,851	,005
	Pourc	8,370	3,827	,170	2,187	,030

a. Dependent Variable: Moyenne_global_100

Tableau 40 : Tableau SPSS – Tableau des coefficients pour la régression linéaire multiple



Dès lors, l'équation peut s'écrire :

$$\hat{y} = - 4,692 + (- 13,074) * x_1 + 17,582 * x_2 + 4,189 * x_3 + 8,370 * x_4$$

Ainsi, un étudiant qui n'a pas doublé d'année dans le secondaire (**Redoubl** = 1) mais qui recommence sa première année de bachelier (**Bisseur** = 2), qui a suivi 6 heures de mathématique dans le secondaire (**heures_math** = 6) et avait entre 65% et 75% (**Pourc** = 2) aura théoriquement une moyenne globale à la fin de BAC1 de :

$$\hat{y} = - 4,692 + (- 13,074) * 1 + 17,582 * 2 + 4,189 * 6 + 8,370 * 2 = 59,272\%$$

2. Facteurs influençant la réussite ou non de la première année de bachelier (**Reussite_annee**)

La variable à expliquer « **Reussite_annee** » est binaire. Elle prend la valeur 1 lorsque l'étudiant a réussi son année et 0 sinon.

2.1 Test du Chi-carré de Pearson

2.1.1 Genre de l'étudiant (**Sexe**)

- Pour 2008-2009

Pour voir s'il existe un lien entre deux variables qualitatives, nous pouvons utiliser le test d'indépendance du Chi-carré. Par exemple, ici, nous cherchons à savoir s'il existe un lien entre, d'une part, le genre de l'étudiant (**Sexe** = 1 pour les garçons et **Sexe** = 2 pour les filles) et, d'autre part, la réussite de la première année de bachelier (**Reussite_annee** = 0 en cas d'échec et égale à « 1 » en cas de réussite) pour la base de données 2008-2009.



Reussite_annee * Sexe Crosstabulation

Count

		Sexe		Total
		1	2	
Reussite_annee	0	82	30	112
	1	42	28	70
Total		124	58	182

Tableau 41 : Tableau SPSS – Tableau croisé des effectifs observés pour les variables Reussite_annee et Sexe

Le tableau précédent montre la répartition homme-femme des effectifs observés. Celui-ci compte 124 garçons pour 58 filles pour les options sciences de gestion et ingénierat de gestion. Ainsi, 82 garçons sur 124 au total ont échoué (= 0,6613) et 30 filles sur 58 au total ont échoué (= 0,5172). En outre, 112 personnes (tout sexe confondu) ont échoué sur 182 personnes au total (= 0,6154).

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3,464 ^a	1	,063		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 22,31.

Tableau 42 : Tableau SPSS – Test du Chi-carré de Pearson

Le test part de l'hypothèse H0 que les deux variables (**Sexe** et **Reussite_annee**) sont indépendantes, qu'elles n'ont pas de lien l'une avec l'autre. Dès lors, l'hypothèse H1 est une hypothèse de dépendance (il y a un lien entre les deux variables).

Le tableau ci-dessus montre une valeur du Chi Carré de Pearson de 3,464 avec une probabilité associée de 0,063. L'hypothèse H0 est donc acceptée à 5%. Néanmoins, il faut rester conscient qu'à un seuil d'acceptation de 10%, l'hypothèse aurait été rejetée. Il faut donc être critique face à ce résultat et le nuancer.



Ainsi, pour l'année 2008-2009 et pour les options sciences de gestion et ingénierat de gestion, nous acceptons l'hypothèse H_0 d'indépendance à un seuil de 5%. Dès lors, à 5% (mais pas à 10%), les variables sont indépendantes l'une de l'autre, le fait de réussir ou non n'a pas de lien avec le genre de la personne.

- Pour 2009-2010 [Annexe 22]

Pour l'année 2009-2010, la valeur du Chi-carré de Pearson est de 5,977 avec une probabilité critique de 0,014. Celle-ci est inférieure au seuil de 5% et donc l'hypothèse H_0 d'indépendance des variables est rejetée.

Pour l'année 2009-2010 et les options sciences de gestion et ingénierat de gestion, nous pouvons conclure qu'il y a un lien de dépendance entre la réussite en fin de BAC1 et le genre de l'étudiant.

2.1.2 Nombre d'heures de mathématique suivies pendant les deux dernières années du secondaire (math6h)

- Pour 2008-2009 et 2009-2010 [Annexes 23 et 24]

Pour les années 2008-2009 et 2009-2010, il y a un lien de dépendance entre la réussite en fin de BAC1 et le fait d'avoir suivi au moins 6 heures de mathématique dans le secondaire.

- Conclusion

Dans les deux années, il y a donc un lien de dépendance entre les variables **Reussite_annee** et **math6h**.



2.1.3 Etudiant bisseur (Bisseur)

- Pour 2008-2009 et 2009-2010 [Annexe 25 et 26]

Pour les années 2008-2009 et 2009-2010 et les options sciences de gestion et ingéniorat de gestion, il n'y a pas de lien de dépendance entre la réussite en fin de BAC1 et le fait de recommencer son année.

2.1.4 Dernier diplôme des parents

a. Diplôme du père (Pere_regroup)

[Annexes 27 et 28]

Pour les années 2008-2009 et 2009-2010 et les options sciences de gestion et ingéniorat de gestion, il n'y a pas de lien de dépendance entre la réussite en fin de BAC1 et le dernier diplôme obtenu par le père de l'étudiant.

b. Diplôme de la mère (Mere_regroup)

- Pour 2008-2009 [Annexe 29]

Pour l'année 2008-2009, il y a un lien de dépendance entre la réussite en fin de BAC1 et le dernier diplôme obtenu par la mère de l'étudiant.

- Pour 2009-2010 [Annexe 30]

Pour l'année 2009-2010, il n'y a pas de lien de dépendance entre la réussite en fin de BAC1 et le dernier diplôme obtenu par la mère de l'étudiant.



2.1.5 Participation aux cours d'été (Cours_ete)

- Pour 2008-2009 [Annexe 31]

Il y a un lien de dépendance entre la réussite en fin de BAC1 et la participation aux cours d'été proposés par les FUCaM début septembre.

- Pour 2009-2010 [Annexe 32]

Pour l'année 2009-2010 et les options sciences de gestion et ingénierat de gestion, il n'y a pas de lien de dépendance entre la réussite en fin de BAC1 et la participation aux cours d'été proposés par les FUCaM début septembre si nous prenons un taux de risque α de 5%. Par contre, avec un risque de 10%, il y a un lien. Il faut donc rester critique quant à la présence ou non d'un lien.

2.1.6 Age (Age_regroup)

- Pour 2008-2009 et 2009-2010 [Annexes 33 et 34]

Pour les années 2008-2009 et 2009-2010 et les options sciences de gestion et ingénierat de gestion, nous pouvons conclure qu'il y a un lien de dépendance entre la réussite en fin de BAC1 et l'âge de l'étudiant.

2.1.7 Bourse

- Pour 2008-2009 et 2009-2010 [Annexes 35 et 36]

Pour l'année 2008-2009, la valeur du Chi-carré de Pearson est de 1,560 avec une probabilité critique de 0,212. Celle-ci est supérieure au seuil de 5% et donc l'hypothèse H_0



d'indépendance des variables est acceptée. Cette hypothèse est également acceptée pour l'année 2009-2010.

Ainsi, pour les années 2008-2009 et 2009-2010, et les options sciences de gestion et ingénierat de gestion, nous pouvons conclure qu'il n'y a pas de lien de dépendance entre la réussite en fin de BAC1 et le fait d'avoir une bourse.

2.1.8 Kot

- Pour 2008-2009 [Annexe 37]

Pour l'année 2009-2010 et les options sciences de gestion et ingénierat de gestion, il y a un lien de dépendance entre la réussite en fin de BAC1 et le fait d'avoir un kot sur le campus.

- Pour 2009-2010 [Annexe 38]

Pour l'année 2009-2010 et les options sciences de gestion et ingénierat de gestion, nous pouvons conclure qu'il n'y a pas de lien de dépendance entre la réussite en fin de BAC1 et le fait d'avoir un kot sur le campus.

2.1.9 Tableau récapitulatif

Tableau des facteurs influençant la réussite en fin de BAC1 pour les sciences et ingénierat de gestion

	<u>2008-2009</u>	<u>2009-2010</u>
1) Genre de l'étudiant	* (mais dépendance à 10%)	Dépendance



2) Heures de mathématique en secondaire	Dépendance	Dépendance
3) « Bisseur »	*	*
4.1) Diplôme du père	*	*
4.2) Diplôme de la mère	Dépendance	*
5) Cours d'été	Dépendance	* (mais dépendance à 10%)
6) Age	Dépendance	Dépendance
7) Bourse	*	*
8) Kot	Dépendance	*

Tableau 43 : Tableau récapitulatif – Test du Chi-carré de Pearson

* indépendance à 5%

Grâce à ce tableau récapitulatif, nous voyons que les facteurs les plus liés à la réussite pour les sciences de gestion et ingénieurs de gestion sont : le genre de l'étudiant, les heures de mathématique dans le secondaire, l'âge, la participation aux cours d'été.

2.2 Régression logistique

Cette méthode a pour but de déterminer les facteurs influençant la réussite de la première année de BAC. Nous avons choisi comme variables potentielles explicatives : **Sexe**, **Filiere**, **Redoubl**, **Mere_regroup**, **Pere_regroup**, **Pourc**, **heures_math**, **Bourse**, **Kot**, **Cours_ete** et **Passeport_moy_20**.

- Régression logistique réalisée avec la base de données 2008-2009 pour Orient = 1 (sciences de gestion et ingénieurs de gestion)

Grâce au logiciel SPSS, nous obtenons une sélection de 4 variables, les plus significatives : **Redoubl**, **Pourc**, **Kot** et **Passeport_moy_20**. Ici, nous avons donc utilisé la méthode « forward-LR » qui choisit étape par étape la variable la plus significative pour expliquer la réussite en BAC1.



Il y a différentes façons d'interpréter les coefficients du tableau 44 (ci-dessous) [ne sont repris dans ce tableau que les coefficients de la 4^e et dernière étape de la régression]. En général, ils représentent l'impact d'une unité de plus sur le ratio des probabilités. Par exemple, si la cote du passeport mathématique (/20) augmente d'un point, le ratio des probabilités augmenterait de 0,335. Une autre manière d'interpréter les coefficients est d'utiliser les **exponentielles**. Ainsi, $e^{0,335} = 1,39$; ce qui signifie que si la cote au passeport mathématique augmente de un, les chances de réussite de l'étudiant sont multipliées par 1,39.

Nous pouvons également calculer le **Logit**, ainsi :

$$\text{Log} \frac{\pi(x)}{1 - \pi(x)} = -5,428 + (-1,622) * (\text{Redoubl}) + 1,016 * (\text{Pourc}) \\ + 1,427 * (\text{Kot}) + 0,335 * (\text{Passeport_moy_20})$$

Ainsi, pour un étudiant ayant eu 14/20 au passeport mathématique, ayant un kot sur le campus (**Kot** = 1), ayant eu entre 65% et 75% comme résultat en fin de secondaire (**Pourc** = 2) et n'ayant pas redoublé dans le secondaire (**Redoubl** = 1) :

$$\text{Log} \frac{\pi(x)}{1 - \pi(x)} = -5,428 + (-1,622) * (1) + 1,016 * (2) + 1,427 * (1) \\ + 0,335 * (14) = 1,099$$

$$\frac{\pi(x)}{1 - \pi(x)} = e^{1,099} = 3,00116$$

$$P(y = 1) = 0,75$$

Cet étudiant aura 75% de chance de réussite pour sa première année de BAC.



Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 4 ^a	Redoubl	-1,622	,567	8,183	1	,004	,197
	Pourc	1,016	,389	6,811	1	,009	2,762
	Kot	1,427	,497	8,234	1	,004	4,167
	Passeport_moy_20	,335	,087	14,713	1	,000	1,399
	Constant	-5,428	1,562	12,078	1	,001	,004

a. Variable(s) entered on step 4: Pourc.

Tableau 44 : Tableau SPSS – Coefficients de la régression logistique

Le R^2 du modèle à l'étape 4 est de 0,469. Cela signifie que 47% de la variation de la réussite est expliquée par les facteurs conservés à la dernière étape du modèle.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
4	123,572	,347	,469

Tableau 45 : Tableau SPSS – Tableau résumé du modèle pour la régression logistique

- Régression logistique réalisée avec la base de données 2009-2010 pour Orient = 1 (sciences de gestion et ingénieurs de gestion)

Nous avons également demandé à SPSS une régression logistique avec les mêmes variables d'entrée pour l'année 2009-2010.



		Variables in the Equation					
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Passeport_moy_20	,386	,092	17,671	1	,000	1,471
	Constant	-5,737	1,315	19,035	1	,000	,003
Step 2 ^b	Sexe	1,494	,498	9,017	1	,003	4,457
	Passeport_moy_20	,401	,094	18,000	1	,000	1,493
	Constant	-7,958	1,618	24,188	1	,000	,000
Step 3 ^c	Sexe	1,147	,530	4,676	1	,031	3,149
	Pourc	,991	,467	4,503	1	,034	2,693
	Passeport_moy_20	,413	,100	16,941	1	,000	1,512
	Constant	-9,635	1,927	24,988	1	,000	,000

a. Variable(s) entered on step 1: Passeport_moy_20.

b. Variable(s) entered on step 2: Sexe.

c. Variable(s) entered on step 3: Pourc.

Tableau 46 : Tableau SPSS – Coefficients de la régression logistique

Nous pouvons dès lors écrire le **Logit** de cette manière :

$$\text{Log} \frac{\pi(x)}{1-\pi(x)} = -9,635 + 0,413 * (\text{Passeport_moy_20}) + 0,991 * (\text{Pourc}) + 1,147 * (\text{Sexe})$$

Ainsi, pour un étudiant ayant eu 14/20 au passeport mathématique, ayant eu entre 65% et 75% comme résultat en fin de secondaire et étant une fille :

$$\text{Log} \frac{\pi(x)}{1-\pi(x)} = -9,635 + 0,413 * (14) + 0,991 * (2) + 1,147 * (2) = 0,423$$

$$\frac{\pi(x)}{1-\pi(x)} = e^{0,423} = 1,5265343$$

$$P(y = 1) = 0,6042$$

Cette étudiante aura 60,42% de chance de réussite pour sa première année de BAC.



Section 2 : Pour les sciences humaines, sciences politiques et information et communication (Orient = 2)

1. Facteurs influençant la moyenne globale obtenue en fin de BAC1 (Moyenne_global_100)

1.1 Comparaison des moyennes

Voici le tableau récapitulatif des facteurs influençant la moyenne globale en fin de BAC1 pour les sciences humaines, sciences politiques et information et communication. [Annexes 39-56]

	<u>2008-2009</u>	<u>2009-2010</u>
1) Genre de l'étudiant	Pas de différence	Pas de différence
2) Heures de mathématique en secondaire	Différence en faveur de ceux qui ont suivi au moins 6h	Pas de différence
3) « Bisseur »	Pas de différence	Pas de différence
4) Cours d'été	Différence en faveur de ceux qui y ont participé	Pas de différence
5) Redoublement en secondaire	Différence en faveur de ceux qui n'ont pas doublé à un seuil de 10% mais pas à 5%	Différence en faveur de ceux qui n'ont pas doublé à un seuil de 10% mais pas à 5%
6) Pourcentage en fin de secondaire	Il y a des différences . Les étudiants qui ont moins de 65% en fin de secondaire ont une moyenne différente et plus basse que celle des étudiants qui ont entre 65% et 75% ou plus de 75%	Pas de différence
7) Age	Pas de différence	Pas de différence



8) Bourse	Pas de différence	Pas de différence
9) Kot	Pas de différence	Pas de différence

Tableau 47 : Tableau SPSS – Tableau récapitulatif des comparaisons de moyennes

Dans ce tableau, nous voyons que le facteur le plus lié à la moyenne en fin de BAC 1 pour les sciences politiques, humaines et information et communication est : le fait d'avoir doublé durant ses études secondaires. Le pourcentage en fin de secondaire, les heures de mathématique et la participation aux cours d'été peuvent également jouer un rôle. Les données le montrent pour l'année 2008-2009.

1.2 Corrélation linéaire

- Pour 2008-2009

Correlations		Moyenne_global_100	Passeport_moy_20	heures_math
Moyenne_global_100	Pearson Correlation	1	,404**	,303**
	Sig. (2-tailed)		,000	,001
	N	118	86	108
Passeport_moy_20	Pearson Correlation	,404**	1	,443**
	Sig. (2-tailed)	,000		,000
	N	86	86	81
heures_math	Pearson Correlation	,303**	,443**	1
	Sig. (2-tailed)	,001	,000	
	N	108	81	108

** . Correlation is significant at the 0.01 level (2-tailed).

Tableau 48 : Tableau SPSS – Tableau de corrélation linéaire

Nous pouvons noter un lien positif d'intensité moyenne entre **Passeport_moy_20** et **heures_math** (0,443) et **Moyenne_globale_100** et **Passeport_moy_20** (0,404).



- Pour 2009-2010

Correlations		Moyenne_global_100	Passeport_moy_20	heures_math
Moyenne_global_100	Pearson Correlation	1	,408**	,319**
	Sig. (2-tailed)		,000	,001
	N	115	86	97
Passeport_moy_20	Pearson Correlation	,408**	1	,372**
	Sig. (2-tailed)	,000		,001
	N	86	86	80
heures_math	Pearson Correlation	,319**	,372**	1
	Sig. (2-tailed)	,001	,001	
	N	97	80	97

** . Correlation is significant at the 0.01 level (2-tailed).

Tableau 49 : Tableau SPSS – Tableau de corrélation linéaire

Pour l'année 2009-2010, il y a une corrélation positive d'intensité moyenne entre **Moyenne_globale_100** et **Passeport_moy_20** (0,408) et entre les variables **heures_math** et **Passeport_moy_20** (0,372).

1.3 Régression linéaire multiple

- 2008-2009

Nous utilisons donc la méthode « Stepwise » (pas à pas) proposée par le logiciel SPSS. A la deuxième étape (voir tableau 50, ci-après), les variables **heures_math** (x_1) et **Pourc** (x_2) sont les plus significatives. Le modèle peut dès lors s'écrire :

$$\hat{y} = -23,792 + 8,408 * x_1 + (11,860) * x_2$$

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-1,795	11,479		-,156	,876
Heures_math	8,608	2,692	,342	3,198	,002
2 (Constant)	-23,792	14,022		-1,697	,094
Heures_math	8,408	2,600	,334	3,233	,002
Pourc	11,860	4,630	,265	2,562	,012

a. Dependent Variable: Moyenne_global_100

Tableau 50 : Tableau SPSS – Tableau des coefficients pour la régression linéaire multiple

- 2009-2010

Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
1 (Constant)	4,029	11,159		,361	,719
heures_math	6,884	2,252	,307	3,056	,003
2 (Constant)	-22,872	17,274		-1,324	,189
heures_math	8,475	2,351	,378	3,604	,001
Bisseur	16,586	8,226	,211	2,016	,047
3 (Constant)	-47,963	19,840		-2,418	,018
heures_math	8,642	2,293	,385	3,769	,000
Bisseur	19,820	8,130	,252	2,438	,017
Pourc	10,795	4,517	,233	2,390	,019

Tableau 51 : Tableau SPSS – Tableau des coefficients pour la régression linéaire multiple

A la troisième étape, les variables **heures_math**, **Bisseur** et **Pourc** sont les plus significatives. Ce modèle nous donne un R^2 de 0,186 à la troisième étape.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,307 ^a	,094	,084	27,9202	,094	9,341	1	90	,003
2	,366 ^b	,134	,114	27,4565	,040	4,066	1	89	,047
3	,432 ^c	,186	,159	26,7572	,053	5,712	1	88	,019

a. Predictors: (Constant), heures_math

b. Predictors: (Constant), heures_math, Bisseur

c. Predictors: (Constant), heures_math, Bisseur, Pourc

Tableau 52 : Tableau SPSS – Tableau récapitulatif du modèle pour la régression linéaire multiple

2. Facteurs influençant la réussite ou non de la première année de bachelier (Reussite_annee)

2.1 Test du Chi-carré de Pearson

Voici le tableau récapitulatif des facteurs influençant la réussite en fin de BAC1 pour les sciences humaines, sciences politiques et information et communication. [Annexes 57-70]

	<u>2008-2009</u>	<u>2009-2010</u>
1) Genre de l'étudiant	*	*
2) Heures de mathématique en secondaire	Dépendance	Non disponible
3) « Bisseur »	*	* (mais dépendance à 10%)
4.1) Diplôme du père	Non disponible	*
4.2) Diplôme de la mère	*	Non disponible
5) Cours d'été	Dépendance	*



6) Age	*	*
7) Bourse	*	*
8) Kot	*	Non disponible

Tableau 53 : Tableau récapitulatif du test du Chi-carré de Pearson

* indépendance à 5%

De ce tableau récapitulatif, nous pouvons conclure que les critères les plus liés à la réussite pour les sciences politiques, humaines et information et communication sont : le nombre d'heures de mathématique durant les deux dernières années du secondaire et la participation aux cours d'été. Nous voyons cependant que pour l'année 2009-2010, aucune de ces variables n'est statistiquement liée à la réussite.

2.2 Régression logistique

- 2008-2009

La variable explicative est **Reussite_annee**, variable binaire qui prend la valeur 1 si l'étudiant a réussi son BAC1 et 0 si l'étudiant a échoué. Nous avons choisi comme variables explicatives : **Sexe**, **Filiere**, **Redoubl**, **Mere_regroup**, **Pere_regroup**, **Pourc**, **heures_math**, **Bourse**, **Kot**, **Cours_ete** et **Passeport_moy_20**.

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
Passeport_moy_20	,326	,095	11,810	1	,001	1,385
Constant	-4,278	1,082	15,620	1	,000	,014

a. Variable(s) entered on step 1: Passeport_moy_20.

Tableau 54 : Tableau SPSS – Coefficients de la régression logistique

Seule une variable ressort de cette analyse : **Passeport_moy_20**. Nous pouvons calculer le **Logit** :



$$\text{Log} \frac{\pi(x)}{1 - \pi(x)} = -4,278 + (0,326) * (\text{Passeport_moy_20})$$

Dès lors, pour un étudiant qui a eu 15/20 au passeport math :

$$\text{Log} \frac{\pi(x)}{1 - \pi(x)} = -4,278 + (0,326) * (15) = 0,612$$

$$\frac{\pi(x)}{1 - \pi(x)} = e^{0,612} = 1,844$$

$$P(y = 1) = 0,65$$

Cet étudiant aura 65% de chance de réussite pour sa première année de BAC.

- 2009-2010

Les mêmes variables ont été utilisées.

Variables in the Equation						
		B	S.E.	Wald	Df	Sig.
Step 1 ^a	Pourc	1,827	,545	11,229	1	,001
	Constant	-4,704	1,207	15,183	1	,000

a. Variable(s) entered on step 1: Pourc.

Tableau 55 : Tableau SPSS - Coefficients de la régression logistique

Seule une variable ressort de cette analyse : **Pourc**. Nous pouvons calculer le Logit :

$$\text{Log} \frac{\pi(x)}{1 - \pi(x)} = -4,704 + (1,827) * (\text{Pourc})$$

Dès lors, pour un étudiant qui a eu plus de 75% comme résultat en fin de secondaire (**Pourc** = 3) :



$$\text{Log} \frac{\pi(x)}{1 - \pi(x)} = -4,704 + (1,827) * (3) = 0,777$$

$$\frac{\pi(x)}{1 - \pi(x)} = e^{0,777} = 2,1749$$

$$P(y = 1) = 0,69$$

Cet étudiant aura 69% de chance de réussite pour sa première année de BAC.

Section 3 : Conclusion des analyses statistiques

Les différents modèles, méthodes et tests nous apportent des résultats variés sur les liens entre les facteurs. Parfois, ces résultats renforcent clairement une tendance connue mais certains sont aussi contre-intuitifs. Néanmoins, il faut rester conscient que ces résultats ressortent d'une analyse statistique de données faite uniquement sur certaines variables choisies sur base de la littérature existante et des données qu'il était possible de recueillir.

La différence faite entre les deux subdivisions d'orientation était importante à souligner car ce n'est pas le même type d'étudiants qui choisit l'une ou l'autre orientation.



Ainsi, pour chaque test sont repris dans le tableau suivant les facteurs explicatifs les plus significatifs de la réussite du BAC1 ou de la moyenne globale en fin d'année.

<u>Tests</u>	<u>Orient = 1</u> (<u>Sciences et Ingéniorat de</u> <u>Gestion</u>)	<u>Orient = 2</u> (<u>Sciences humaines, politiques et</u> <u>information et communication</u>)
Comparaison de moyennes	• Sexe • Math6h • Cours_ete • Redoubl • Pourc • Age	• Math6h • Cours_ete • Redoubl • Pourc
Régression linéaire	• Heures_math • Redoubl • Pourc • Passeport_moy_20 • Bisseur	• Heures_math • Pourc • Passeport_moy_20
Test du Chi-carré	• Sexe • Math6h • Cours_ete • Age	• Math6h • Cours_ete
Régression logistique	• Redoubl • Pourc • Passeport_moy_20 • Sexe • Kot	• Passeport_moy_20 • Pourc

Tableau 56 : Tableau récapitulatif des facteurs explicatifs par test utilisé

Les deux premiers tests (comparaison de moyennes et régression linéaire) sont réalisés avec la variable explicative **Moyenne_global_100** et les deux suivants (test du Chi-carré et



régression logistique) avec **Reussite_annee**. Les résultats sont légèrement différents même si de nombreux facteurs apparaissent à plusieurs reprises.

Avec nos bases de données et les tests que nous avons réalisés, il est également à noter qu'il y a moins de variables significatives pour **Orient** = 2, à savoir le regroupement des sciences humaines, politiques et information et communication. Nous pouvons aussi souligner que les facteurs tels les diplômes obtenus par les parents, le fait de recommencer son BAC1 et le fait d'avoir une bourse ou un kot sur le campus n'ont pas d'influence sur la moyenne obtenue en fin de BAC1. Les étudiants sont donc égaux sur ces facteurs face à la réussite universitaire.

Les actions menées pour aider l'étudiant doivent donc être plus ciblées sur des facteurs significatifs tels que les heures de mathématique ou les cours d'été. Ainsi, les activités entreprises actuellement par les FUCaM (monitorats, semaine propédeutique...) doivent être maintenues, voire renforcées et améliorées au fil des années.

Nous voyons que la variable **Passeport_moy_20** qui reprend les résultats au « passeport mathématique » est une variable importante. Celle-ci rappelle le fait que certains prérequis sont importants pour entamer des études universitaires. Néanmoins, un étudiant qui a raté ce test doit garder à l'esprit qu'il a encore toutes ses chances pour réussir son année. Le test se déroule au tout début d'année ce qui lui permet de prendre conscience de ses limites et de réagir.

Nous pouvons également souligner que, logiquement, nous ne pouvons pas intervenir sur la variable **Sexe**. En ce qui concerne la variable **Age** (fortement corrélée avec le fait d'avoir doublé en secondaire), elle peut remettre en question le fait d'obliger l'étudiant à recommencer une année déjà en secondaire et donc à entrer dans la spirale de l'échec. Il existe d'ailleurs, dans l'enseignement secondaire des alternatives, telles que le changement de filière sans redoubler, une année complémentaire considérée comme une remédiation et non comme un redoublement, l'obligation de ne pouvoir doubler qu'une seule fois par degré (à savoir que le premier degré reprend la 1^{ère} et 2^e année du secondaire).



En outre, les variables **Math6h**, **Cours_ete** et **Redoubl** apparaissent comme significatifs pour les deux orientations.

Ainsi, ces facteurs explicatifs donnés par les tests statistiques (à partir des bases de données) confortent notre revue de la littérature. Nos résultats vont globalement dans le même sens que les recherches précédentes.



Pistes et actions concrètes proposées comme solutions à la problématique

Dans cette partie, nous élargirons notre problématique qui était de comprendre et d'appréhender à partir de la littérature et de nos bases de données les facteurs explicatifs de la réussite universitaire en BAC1. Nous présenterons donc, ici, d'autres actions, pistes et solutions.

Nous pouvons souligner qu'il est important de continuer à poursuivre les actions déjà menées dans les universités et plus particulièrement aux FUCaM. Celles-ci sont d'ailleurs adaptées et améliorées au fil des années. Ainsi, avant toute chose, une considération à énoncer est l'idée que les universités ne veulent pas niveler leur niveau par le bas mais bien améliorer le taux de réussite grâce à des actions de soutien, d'aides, de conseils...

Une première catégorie d'aides proposées est celle liée au « passeport pour le bac ». Celui-ci permet, entre autres, à l'étudiant de se remettre en question et de mieux se situer dans sa connaissance de la matière. Il reçoit d'ailleurs, par la suite, un feedback rapide et la possibilité d'une remédiation. Aux FUCaM, le passeport n'est proposé aux étudiants que pour les mathématiques. Peut-être celui-ci pourrait-il être ouvert à d'autres branches demandant des prérequis comme par exemple : la compréhension de textes¹⁰, l'utilisation des outils informatiques de base (comme un traitement de textes)...

Une deuxième forme d'aide, également déjà présente aux FUCaM est assurée par l'équipe d'accompagnement qui propose des rendez-vous et un « coaching » individuel pour l'étudiant. A Liège, cette équipe a mis en place un système de « coaching par courrier

¹⁰ Les FUCaM organisaient ce test de compréhension de textes mais il a été abandonné en 2008 pour deux raisons : les étudiants avaient déjà d'autres tests qui leur prenaient du temps de passation et il n'y avait pas clairement de remédiation prévue par la suite pour pallier les insuffisances dans la maîtrise de la langue française. Néanmoins, il est possible que ce test soit réinstauré d'ici septembre 2011 avec de nouveaux textes plus adaptés au public des FUCaM.



électronique » (moins individualisé) qui envoie, entre autres, à l'étudiant des conseils méthodologiques. En faculté de médecine à Liège, une newsletter accompagne les étudiants durant leurs blocus et examens (Renotte, Delforge, Houart & Slosse, 2011). Selon moi, cette aide de l'équipe d'accompagnement pourrait être renforcée par une sorte de parrainage où chaque nouvel étudiant se verrait attribuer un parrain (complètement indépendant des baptêmes estudiantins), étudiant d'une année supérieure. Cet étudiant de BAC1 pourrait contacter celui-ci et le rencontrer en cas de difficultés. Cela pourrait être une collaboration et expérience « win-win » pour le nouvel étudiant soutenu dans son cursus mais aussi pour l'étudiant plus âgé qui, plus tard, dans sa vie professionnelle devra pouvoir gérer les relations humaines (par exemple, lors d'un projet de groupe, être capable de remotiver les membres d'une équipe).

Une troisième assistance pourrait être liée aux TIC (technologies de l'information et de la communication). Par exemple, les exercices de e-learning (apprentissage en ligne grâce à une pédagogie basée sur l'internet) « prennent en compte la diversité des rythmes et des styles d'apprentissage » (Renotte & al., 2011, p.35). Ces exercices peuvent être réalisés par l'étudiant en dehors des heures de cours. L'étudiant peut s'entraîner en recommençant plusieurs fois les exercices à son rythme et en recevant, par la suite, un feedback rapide. D'autres activités liées aux TIC pouvant être développées ou soutenues sont les blogs, forums, outils d'échanges et de discussions, réseaux sociaux (qui peuvent renforcer les liens au sein des groupes d'étudiants)...(Renotte & al., 2011). Un autre exemple serait celui des facultés polytechniques de Mons où des e-tests comprenant une dizaine de questions à choix multiples sont mis en place et permettent à l'étudiant de s'auto-évaluer. Nous voyons qu'il existe donc de nombreux moyens de tirer profit des richesses des technologies multimédia si l'université accepte d'allouer des nouvelles ressources humaines à ce type d'encadrement.

Une autre proposition, demandant certes aussi des moyens pour sa mise en place, serait l'organisation de séminaires ayant pour but de guider l'étudiant vers la réussite. Ceux-ci auraient pour thèmes¹¹ des sujets tels que : la gestion du temps, la gestion de la quantité de matière à restituer, la prise de notes, la lecture d'un énoncé... Des fiches sur ces mêmes

¹¹ Aux FUCaM, la Professeur B. Noël assure chaque année début décembre une séance du cours de psychologie sur le thème de la gestion du blocus.



thématiques : blocus, emploi du temps, cours, mémorisation, prise de notes...pourraient être réalisées et distribuées aux étudiants ou mises en ligne sur le site internet de l'université. Là aussi, une aide supplémentaire devrait être prévue.

Une toute autre piste serait l'ouverture en Belgique d'un master « formant à l'orientation scolaire et professionnelle » (Nils, 2011, p15). En parallèle, renforcer la formation des assistants par des cours liés à la pédagogie universitaire pourrait les aider à mieux encadrer et guider les étudiants vers la réussite de leur année. En effet, les méthodes et pratiques d'enseignement et leur amélioration sont souvent oubliées lorsqu'est abordé le problème du taux d'échec élevé en première année à l'université.

En outre, pour aider à l'orientation et au choix de l'élève de 6^e secondaire, une plus grande collaboration entre les universités et écoles du secondaire pourrait être instaurée même si celle-ci demande beaucoup de moyens et de temps. Aux FUCaM, cette démarche est d'ailleurs déjà mise en route.

Enfin, pour trouver de nouvelles idées et continuer à vouloir amener un maximum d'étudiants vers la réussite, certaines questions peuvent nous aider :

- « Quels types d'activités (permanentes, régulières, ponctuelles) mettre en place ?
- Par quelles approches (individuelle, collective) ?
- Par qui (généralistes, experts, professeurs, assistants, services de gestion) ?
- Par quels moyens (démonstration, simulation, discussion, débat) ?
- Quel temps leur consacrer ?
- Comment mesurer leurs effets ? » (Delhaxhe et al., 2011).

Pour conclure, toutes les institutions universitaires ont mis en place des actions pour aider les étudiants et plus particulièrement ceux de BAC1. Ces actions sont diversifiées et nombreuses mais parfois peu connues des étudiants eux-mêmes. Il est donc bon de continuer à améliorer la diffusion et l'efficacité de celles-ci tout en cherchant de nouvelles pistes qui répondent à un besoin réel des étudiants.



Conclusion générale

L'objectif de ce travail de fin d'étude a été d'identifier les critères significatifs qui influencent la réussite universitaire et ensuite, de les analyser à partir de bases de données reprenant les étudiants de BAC1 des FUCaM pour les années 2008-2009 et 2009-2010. Pour réaliser les différents tests statistiques sur les données, nous avons utilisé le logiciel SPSS.

L'aperçu de la littérature existante sur le thème de la réussite et les actions déjà mises en place au sein des FUCaM démontrent un réel besoin d'aide, de soutien et d'actions en vue d'améliorer le taux de réussite des étudiants en BAC1. De même, de nombreuses recherches sur cette thématique ont déjà été menées (comme par exemple, celles de l'Académie Louvain concernant le « passeport pour le bac »). Les deux premiers chapitres de ce mémoire nous donnent une vision de facteurs potentiellement explicatifs et de projets déjà mis en œuvre.

Dans le troisième chapitre, chaque test utilisé a été expliqué afin de permettre au lecteur de comprendre les détails des développements et chaque nombre délivré par le logiciel SPSS. Cette section a été divisée en deux grandes parties. La première a présenté les tests réalisés pour des modèles étudiant le lien entre une variable à expliquer quantitative et :

- des variables explicatives binaires : tests de comparaisons de moyennes ;
- des variables explicatives quantitatives : corrélations linéaires ;
- des variables explicatives qualitatives et quantitatives : régressions linéaires.

Tandis que la deuxième section du chapitre a étudié des modèles montrant le lien entre une variable à expliquer qualitative et :

- des variables explicatives qualitatives : test du Chi-carré de Pearson ;
- des variables explicatives quantitatives : comparaisons de moyennes ;
- des variables explicatives qualitatives et quantitatives : régressions logistiques.

D'autres tests et méthodes auraient pu être développés et utilisés. Nous pensons notamment à la méthode des arbres de décisions ou à l'analyse discriminante. En outre, notre base de données ne concernait que deux années d'étudiants de BAC1 et certaines sections y étaient sous-représentées. De futures recherches pourraient donc étoffer cette base de données



avec les étudiants de BAC1 des prochaines années ou même l'élargir en suivant les étudiants actuels tout au long de leurs cinq années d'études (voire même en commençant le suivi dès le secondaire).

Dans le quatrième chapitre sont présentées les données et une analyse descriptive de celles-ci. Les facteurs, qui nous ont paru les plus pertinents, ont été choisis d'après la littérature sur le thème de la réussite mais aussi d'après les données déjà disponibles sur le terrain. Cette liste n'est donc pas exhaustive. D'autres informations telles que l'option choisie, la langue de l'étudiant, les résultats obtenus aux tests formatifs de novembre ou en janvier auraient apporté des résultats supplémentaires. Car, nous savons que les résultats de janvier sont fortement corrélés avec ceux obtenus en juin. Cela montre d'ailleurs que les actions d'aide aux étudiants doivent être menées assez tôt dans l'année, voire dès le début de celle-ci.

La dernière partie de ce travail de fin d'étude a été consacrée aux conclusions découlant des tests réalisés sur notre base de données. Les résultats sont présentés séparément selon deux regroupements d'orientations :

- Orient = 1 : Sciences de gestion et ingénierat de gestion ;
- Orient = 2 : Sciences humaines, sciences politiques et information et communication.

Les tests nous ont montré que le nombre d'heures de mathématique suivies durant les deux dernières années du secondaire ainsi que les résultats au passeport mathématique (réalisé en début de BAC1) ont une influence sur la réussite. Cela met en évidence que la maîtrise des prérequis est importante pour entamer des études universitaires. Les cours de remédiation proposés aux étudiants à la suite du passeport mathématique doivent donc cibler leurs difficultés et porter sur les compétences les moins acquises. Néanmoins, ce lien entre les heures de mathématique et la réussite est à nuancer. Comme le dit A. Schatteman (2009) dans son mémoire, le choix des options mathématiques fortes ou non peut simplement être dû à l'envie d'apprendre une seconde langue, d'étudier plus particulièrement la littérature ou les sciences économiques et non dû au fait de vouloir éviter les cours de mathématique et d'en prendre un minimum d'heures.



Le pourcentage obtenu en fin de secondaire ressort également comme facteur significatif sur la route vers la réussite. L'université a peu d'influence directe sur ce critère si ce n'est de pouvoir favoriser et encourager le dialogue entre écoles secondaires, hautes écoles et universités pour une meilleure préparation des élèves aux études supérieures, ce que font déjà actuellement les FUCaM.

En outre, la participation des étudiants aux cours d'été semble également une variable significative pour favoriser la réussite en fin de BAC1. Cependant, ce lien peut être nuancé par le fait que ces étudiants sont sans doute également les plus motivés par leur nouveau projet scolaire. Nous pouvons également souligner que, même si logiquement, nous n'avons aucune influence sur la variable **Sexe**, cette variable est significativement liée à la réussite et à la moyenne obtenue en fin d'année dans deux de nos tests pour les sciences et ingénieurs de gestion. La variable **Age** (fortement corrélée avec le fait d'avoir doublé en secondaire) montre aussi une corrélation avec la réussite. La question du redoublement, de sa suppression ou de méthodes alternatives reste donc d'actualité.

De plus, il est à noter que des facteurs portant plus spécialement sur le profil de l'étudiant tels que les diplômes obtenus par les parents, le fait d'avoir une bourse ou un kot et le fait de recommencer son BAC1 n'ont pas d'influence sur la moyenne obtenue en fin de BAC1 d'après les tests réalisés sur nos données. Les étudiants sont donc égaux sur ces facteurs face à la réussite universitaire.

Ainsi, ce mémoire a eu pour but d'identifier des facteurs explicatifs de la réussite universitaire. Ceux-ci avaient déjà été repris dans la revue de la littérature. Ils viennent dès lors corroborer cette dernière.

Finalement, une section de ce mémoire est spécifiquement destinée à de nouvelles pistes et actions proposées aux universités ou écoles secondaires en vue d'essayer de diminuer le taux d'échec en première année de BAC.



Nous espérons que ce mémoire contribuera aux recherches déjà menées sur la réussite universitaire, qu'il pourra corroborer certaines tendances et théories mais aussi apporter des nouveaux constats et idées pour des actions et pistes concrètes.

Pour conclure, ce mémoire m'aura permis d'atteindre les objectifs décrits dans le guide du mémoire¹² c'est-à-dire « mettre en avant mon autonomie, ma capacité de synthèse, mon sens critique et mes compétences personnelles ». En outre, je pense avoir pu développer une certaine rigueur dans mon travail grâce aux conseils de mes promoteurs. Je tiens donc une fois encore à les remercier. Je voudrais également adresser un merci particulier aux FUCaM, aux membres du personnel et à tous mes professeurs pour leurs suivis et leur aide pendant mes cinq années d'études.

¹² <http://www.fucam.ac.be/engine/index.php3?pere=21366>, page consultée le 20/12/2010, html



Perspectives

Ce mémoire à la suite ou en parallèle à de nombreuses autres recherches et mémoires nous montre que cette thématique est pleine de richesses. Nous avons dû limiter notre étude mais celle-ci ouvre de nouvelles perspectives à envisager.

Nous pensons notamment à une étude qui pourrait suivre les étudiants de BAC1 des FUCaM des prochaines années et ainsi venir étoffer nos bases de données. Une autre idée serait de suivre les étudiants sur cinq années (voire même en commençant le suivi dès le secondaire).

L'identification des prérequis pourrait aussi être affinée car nous n'avons pris en compte que les variables heures_math et Passeport_moy_20 pour les identifier. Ceux-ci pourraient être certainement plus développés et analysés.

D'autres facteurs tels que la motivation de l'étudiant, son intérêt pour les cours, la perception de son environnement et de son intégration dans la vie académique et sociale... pourraient être étudiés et faire partie d'une nouvelle enquête plus ciblée sur des facteurs psychologiques. Par exemple : en analysant le taux de présence des étudiants aux cours, le fait d'avoir relu son cours avant la séance suivante, de prendre des notes ou d'avoir une attitude plus active durant le cours...

Ce sont, ici encore, quelques perspectives de recherches dans une liste non exhaustive qui nous montrent combien le sujet est important, riche et encore trop souvent sous-estimé.



Références bibliographiques

BAILLARGEON G. (2003), « *STATISTIQUE APPLIQUEE pour les sciences de la gestion et les sciences économiques* », Ed SMG

BOXUS E. (1993), « *Rapport du groupe de travail Réussite du CIUF* », Rapport du Conseil Interuniversitaire de la Communauté française, Bruxelles

CLOSSET J-L., DELFORGE M. (2011), « *Comment tirer parti d'une évaluation formative ?* » In PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF

DAGNELIE P. (2006), « *Statistique théorique et appliquée. (2, Inférence statistique à une et à deux dimensions)* », Bruxelles, Ed. De Boeck et Larcier, (2^e éd.)

DE KETELE J.M., ROGIERS X. (1993), « *Méthodologie du recueil d'informations* », Bruxelles, De Boeck Université

DELHAXHE M., HOUART M., POLLET M-C. (2011), « *Comment développer des compétences méthodologiques et organisationnelles ?* » In PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF

DELLA GUISTINA L. & DUFRASNE V. (2010), « *Facteurs de réussite en 1^{ère} année de bachelier et lien entre les résultats des étudiants aux passeports présentés en début d'année et leurs résultats académiques : étude statistique* », mémoire de master en ingénieur de gestion, Mons, FUCaM



DERBOUSSE X., LAMBOTTE F., SERMEUS G. (2000), « *dis-moi comment est ton école...* » In SCHATTEMAN A. (2009), « *La réussite universitaire : Etude de la liaison entre les cotes du passeport mathématique et les résultats finaux des étudiants de première bachelier* », mémoire de master en ingénieur de gestion, Mons, FUCaM

DE WINNE R., MAJOIS C (2009), « *Cours d'Inférence Statistique* », Cours de bachelier en ingénieur/sciences de gestion, Mons, FUCaM

DROESBEKE J-J., HECQUET I., WATTELAR C., (2001) « *La population étudiante. Description, évolution, perspectives* », Éditions de l'Université de Bruxelles

DROESBEKE J-J, VERMANDELE C., (2005), « *L'échec à l'université : comprendre pour mieux prévenir – Quelques réflexions sur la démarche de compréhension* » In GALAND B. (2005), « *L'échec à l'université en Communauté française de Belgique* », Cahier de Recherche en Education et Formation, n°39

DUCARROZ C. (2009), « *Cours de Méthodes et Modèles en Marketing* », Cours de master en ingénieur/sciences de gestion, Mons, FUCaM

FACULTES UNIVERSITAIRES CATHOLIQUES DE MONS (2010), « *Comment préparer son entrée à l'université, calendrier 2010-2011 des activités d'information et d'orientation pour les élèves du secondaire* », Mons, FUCaM

FIELD A. (2009), *Discovering statistics using SPSS*, Londres, Ed. Sage, (third ed.)

GALAND B. (2005), « *L'échec à l'université en Communauté française de Belgique* », Cahier de Recherche en Education et Formation, n°39

GALAND B., NEUVILLE S., FRENAY M. (2005), « *L'échec à l'université en communauté française de Belgique : Comprendre pour mieux prévenir ?* » In GALAND B. (2005), « *L'échec à l'université en Communauté française de Belgique* », Cahier de Recherche en Education et Formation, n°39



GOFFIN T. (2010), « *La réussite universitaire : L'influence de la confiance en soi en 1^{ère} Bachelier* », mémoire de master en ingénieur de gestion, Mons, FUCaM

GOVAERTS B. (2009), « *Statistique appliquée à la psychologie II* », cours de troisième bachelier en psychologie et sciences de l'éducation, Louvain-la-Neuve, UCL

HOWELL D.C. (2008), « *Méthodes statistiques en sciences humaines* », trad. ROGIER M., YZERBYT V., BESTGEN Y., Paris, Ed. De Boeck, (2^e ed.)

IRVING B., BLOXOM C., « *Predicting adolescent delinquent behaviour and criminal conviction by age 30 Evidence from the British Birth Cohort* », (page consultée le 29/08/2010), [Pdf, en ligne],

<http://www.police-foundation.org.uk/files/police0001/publications/juvenile%20delinquency%20&%20post%20adolescent%20offending.pdf>

JOMAUX F. (2006), « *Statistique descriptive* », cours de premier bachelier en sciences de gestion et ingéniorat de gestion, en sciences politiques et sciences humaines et sociales, Mons, FUCaM

KINNEAR P.R., GRAY C.D. (2000), « *SPSS for Windows Made Simple Release 10* », Hove, UK, Ed. Psychology Press Ltd

KINNEAR P.R., GRAY C.D. (2005), « *SPSS facile appliqué à la psychologie et aux sciences sociales* », trad. HUET N., Paris, Ed. De Boeck

LECLERCQ D., PARMENTIER P. (2011), « *Qu'est ce que la réussite à l'université d'un étudiant ?* » In PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF



NILS F., LAMBERT J-P, « *Quelle est l'ampleur de l'échec en première année à l'université ? Tout est-il joué d'avance ?* » In PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF

NILS F. (2011), « *Comment favoriser une bonne orientation lors du choix des études universitaires ?* » In PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF

NOEL B. (2010), « *Comment préparer son entrée à l'université, calendrier 2010-2011 des activités d'information et d'orientation pour les élèves du secondaire* », Mons, FUCaM

NOEL B. (2011), « *Rapport bisannuel sur la lutte contre l'échec dans le premier cycle de l'enseignement universitaire* », année 2008-2009 et 2009-2010, Mons, FUCaM

NOEL B., FRENAY M. (2011), « *Comment soutenir la persévérance dans les études ?* », In PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF

PARMENTIER P. (1994), « *La réussite des études universitaires: Facteurs structurels et processuels de la performance académique en première année de médecine* », thèse à l'Université Catholique de Louvain en Faculté de Psychologie et des Sciences de l'Education, 1994

PARMENTIER P. (2005), *Quelques réactions au texte « L'échec à l'université : Comprendre pour mieux prévenir »* In GALAND B. (2005), « *L'échec à l'université en Communauté française de Belgique* », Cahier de Recherche en Education et Formation, n°39



PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF

RENOTTE C., DELFORGE M., HOUART M., SLOSSE P. (2011), « *Comment les TICE peuvent-elles participer à la réussite des étudiants débutants ?* » In PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF

ROMAINVILLE M. (Décembre 2002), « *L'évaluation des acquis des étudiants dans l'enseignement universitaire* », rapport établi à la demande du Haut Conseil de l'évaluation de l'école

ROMAINVILLE M. (2005), « *Quelques interrogations sur l'échec à l'Université* » In GALAND B. (2005), « *L'échec à l'université en Communauté française de Belgique* », Cahier de Recherche en Education et Formation, n°39

ROMAINVILLE M., SLOSSE P. (2011), « *Comment établir un diagnostic des prérequis précoce et impliquant pour l'étudiant ?* », In PARMENTIER P. (2011), « *Recherches et actions en faveur de la réussite en première année universitaire. Vingt ans de collaboration dans la Commission « Réussite » du Conseil interuniversitaire de la Communauté française de Belgique* », Bruxelles, CIUF

SCHATTEMAN A. (2009), « *La réussite universitaire : Etude de la liaison entre les cotes du passeport mathématique et les résultats finaux des étudiants de première bachelier* », mémoire de master en ingénieur de gestion, Mons, FUCaM

SERVICE DE PEDAGOGIE UNIVERSITAIRE (SPU) et les Facultés Universitaires Notre-Dame de Namur (2001), « *Le projet personnel de l'étudiant: un facteur de réussite ?* », revue Réseau n°49, décembre 2001



TENENHAUS M. (2007), « *Méthodes pour décrire, expliquer et prévoir* », Paris, Ed. Dunod

TENENHAUS M. (2000), « *Méthodes statistiques en gestion* », Paris, Ed. Dunod

VAN CAMPENHOUDT M., DELL' AQUILA F., DUPRIEZ V. (2008), « *La démocratisation de l'enseignement supérieur en Communauté française de Belgique : état des lieux* », Cahier de Recherche en Education et Formation, n°65, décembre 2008

VANDAMME J-P., MESKENS N., SUPERBY J-F. (2006), « *Evaluation du risque d'échec des étudiants de première année universitaire selon leur profil* », Reflets et perspectives de la vie économique 2/2006 (Tome XLV), pp.39-46

VAUSE A., DUPRIEZ V., DUMAY X. (2008), « *L'efficacité des pratiques pédagogiques : la nécessité de prendre en compte l'environnement social* », Cahier de Recherche en Education et Formation, n°63, août 2008

<http://www.uclouvain.be/326547.html?idQuestion=58>, page consultée le 01/02/2011, html

<http://www.cio.ucl.ac.be/old/intro.php?espace=formations&porteUrl=4&idQuestion=64>, page consultée le 28/01/2011, html

<http://www.uclouvain.be/passeport>, page consultée le 01/02/2011, html

<http://www.fucam.ac.be/view.php3?include=1890&pere=20955&print=0>, page consultée le 21/03/2011, html

<http://www.fundp.ac.be/det/actualites/reussir-transition-secondaire-universite/maitrisedes-prerequis-et-reussite-universite/view>, page consultée le 25/03/2011, pdf