



المندوبية السامية للتخطيط
HAUT-COMMISSARIAT AU PLAN

ROYAUME DU MAROC
..*.*.*
HAUT COMMISSARIAT AU PLAN
..*.*.*
INSTITUT NATIONAL
DE STATISTIQUE ET D'ECONOMIE APPLIQUEE



INSEA

Projet de Fin d'Etudes

Modèles internes de notation des entreprises

Préparé par : Mme Olayya BENCHOUAF

Sous la direction de : Mr Khalil SAID (INSEA)

Mr Younes BOURFOUNE (Crédit Agricole du Maroc)

Soutenu publiquement comme exigence partielle en vue de l'obtention du

Diplôme d'Ingénieur d'Etat

Filière : Actuariat-Finance

Devant le jury composé de :

- **Mr. Khalil SAID (INSEA)**
- **Mr. Fouad MARRI (INSEA)**
- **Mr. Younes BOURFOUNE (CAM)**

Septembre 2020/ PFE N°4

Résumé et mots clés

Les modèles de scoring sont employés dans l'estimation de la probabilité de défaut et dans la prédiction de la défaillance ou de la non défaillance du client, au niveau du processus d'octroi de crédit. Cette probabilité est également indispensable au calcul de la perte de crédit attendue, une valeur que la banque prend soin de provisionner. Nous nous sommes intéressés, dans ce projet, aux méthodes de scoring les plus utilisées dans le domaine bancaire, à savoir la régression logistique et l'analyse factorielle discriminante, en plus de la méthode DISQUAL et de certains modèles de machine learning. Dans le but de maîtriser ces techniques, une recherche sur les concepts, sur lesquels elles s'appuient, sur leurs hypothèses d'application et sur leurs caractéristiques était nécessaire. Ensuite, grâce au Crédit Agricole du Maroc nous avons eu accès à une base de données rassemblant des informations comptables sur plus de 1980 TPME. Le but étant d'implémenter les modèles et de les appliquer sur les données dont nous disposons, il fallait tout d'abord effectuer le traitement de la base, en repérant les valeurs manquantes et aberrantes, afin de les imputer, par la suite, par la moyenne. L'étape qui a suivi la préparation de la base de données, a été le calcul et le choix des ratios à travers l'étude des corrélations et la sélection par méthode backward. Les résultats et les prédictions des modèles, taux de bons classements et l'AUC, ont montré que la régression logistique était la méthode la plus concluante.

Mots clés :scoring, processus d'octroi de crédit, perte de crédit attendue, prédiction.

ملخص وكلمات رئيسية

تستخدم نماذج التسجيل في تقدير احتمالية التعثر والتنبؤ بتعثر الزبون أو عدم تعثره في عملية منح الائتمان. وهذا الاحتمال وارد ومن المرجح قد يؤدي خسارة في حساب الائتمان المتوقعة ، وهي قيمة هامة تستلزم على البنك الاعتناء بتوفيرها.

وقد ركزنا اهتمامنا في هذا البحث ، على أساليب تقنية تقوم على تسجيل الدرجات الأكثر استخدامًا في القطاع المصرفي ، وندرجها كالتالي : الانحدار اللوجستي، وتحليل العوامل التمييزية، بالإضافة إلى طريقة الاستبعاد وبعض نماذج التعلم الآلي.

من أجل استعمال هذه التقنيات ، كان من الضروري البحث أولاً عن المفاهيم التي تستند إليها ، وفرضيات تطبيقها ، ومعرفة خصائصها. وبفضل مؤسسة البنك القرض الفلاحي بالمغرب Crédit Agricole du Maroc ، تمكنا بالفعل من الوصول إلى قاعدة بيانات تقوم على جمع المعلومات المحاسبية لأكثر من 1980 من الشركات الصغيرة والمتوسطة. الهدف من هذه العملية هو تنفيذ النماذج وتطبيقها على البيانات الموجودة تحت تصرفنا ، وكان من الضروري إتباع الإجراءات التالية:

أولاً: إجراء معالجة لقاعدة البيانات ، من خلال تحديد القيم المفقودة والقيم المتطرفة ، من أجل تقييدها لاحقاً ، عن طريق نسب متوسطة.

ثانياً: إعداد قاعدة البيانات، وتتم هذه العملية بعد إجراء خدمات محاسبية واختيار النسب من خلال دراسة الارتباطات والاختيار بالطريقة العكسية. وقد أظهرت نتائج وتوقعات النماذج ومعدلات الترتيب الجيدة والجامعة الأمريكية بالقاهرة أن الانحدار اللوجستي كان أكثر الطرق نجاحاً.

الكلمات المفتاحية: التصنيف ، عملية منح الائتمان ، خسارة الائتمان المتوقعة ، التنبؤ.

Dédicaces

Je dédie ce mémoire

A mes chers parents pour leur amour et sacrifices

A ma chère sœur

A mes amis (es) pour leur soutien et amitié sincère

A tous les membres de ma famille pour leur encouragement et motivation

A tous mes professeurs qui m'ont accompagnée tout au long de mes études.

Remerciements

Au terme de ce travail, ma sincère reconnaissance est particulièrement adressée à M. K. SAID, professeur de l'Institut National de Statistique et d'Economie Appliquée, pour avoir accepté de m'encadrer à l'occasion de mon projet de fin d'études ainsi que pour ces précieux conseils, sa disponibilité et à l'attention bienveillante qu'il a portée à l'égard de ce travail.

J'exprime ma profonde gratitude envers M. Y. BOURFOUNE, responsable du Département Modélisation Risque Crédit au Crédit Agricole du Maroc, pour ses directives, sa patience, et sa contribution dans ma formation.

Je tiens aussi à exprimer mes remerciements à M. F. MARRI, professeur à l'Institut National de Statistique et d'Economie Appliquée, pour avoir accepté d'examiner ce modeste travail.

Enfin, un grand merci au corps enseignant et administratif de l'Institut National de Statistique et d'Economie Appliquée.

Table des matières

Résumé et mots clés	2
Dédicaces	4
Remerciements.....	5
Table des matières.....	6
Liste des abréviations	8
Liste des figures et tableaux	9
Introduction.....	11
Chapitre 1 : Contexte et problématique	12
1. Groupe Crédit Agricole.....	12
2. Crédit Agricole du Maroc	12
1.1 Services proposés par le CAM	14
2.1 Processus d’octroi de crédit.....	14
3. Problématique :.....	16
Chapitre 2 : Cadre théorique.....	17
1. Revue littéraire.....	17
2. Scoring.....	18
3. Régression logistique :.....	19
3.1 Hypothèses du modèle :.....	19
3.2 Choix des variables et du modèle :.....	19
3.3 Adéquation du modèle :.....	20
3.4 Qualité d’ajustement du modèle :.....	20
4. L’analyse factorielle discriminante	21
4.1 Notion de distance	22
4.2 Conditions d’application	22
4.3 Estimation de la probabilité de défaut	23
4.4 Qualité du modèle	23
5. La méthode DISQUAL	23
5.1 L’analyse de correspondance multiple :.....	24
5.2 L’analyse factorielle discriminante	25

Table des matières

6. Les arbres de décisions : CART	25
7. Comparaison des modèles	27
Chapitre 3 : Modélisation et implémentation	29
1. Traitement de la base de données et calcul des ratios	29
1.1 Statistiques descriptives sur les variables	30
1.2 Traitement de la base	40
1.3 Calcul des ratios financiers	46
2. Régression logistique	55
3. L'analyse discriminante	59
4. Méthode DISQUAL	61
1.1 L'ACM appliquée au jeu de données	61
1.2 L'analyse discriminante appliquée aux axes factoriels	63
5. Méthode CART.....	64
6. Comparaison des modèles	66
1.1 Matrice de confusion.....	67
1.2 Courbe ROC	69
Conclusion :	73
خاتمة	73
Annexe 1 : Statistiques descriptives sur les variables :	75
1. Matrice de corrélation de Pearson	75
2. Statistiques descriptives sur les variables	89
3. Statistiques descriptives des indicateurs les plus importants.....	90
Bibliographie :	95

Liste des abréviations

FP	: Faux Positifs
VP	: Vrais Positifs
FN	: Faux Négatifs
VN	: VraisNégatifs
ROC	: Receiver Operating Characteristic
CART	: Classification And RegressionTrees
CA	: Chiffre d'affaire
EBE	: Excédent Brut d'Exploitation
CAF	: Capacité d'Autofinancement
BFR	: Besoin en Fonds de Roulement
CAM	: Crédit Agricole du Maroc
PDV	: Point De Vente
AUC	: Area Under Curve
DISQUAL	: Discrimination sur variables Qualitatives

Liste des figures et tableaux

<i>Figure 1: Organigramme général du CAM</i>	13
<i>Figure 2: Organigramme du Domaine Gestion du Risque Crédit</i>	13
<i>Figure 3: Processus d'octroi de crédit</i>	15
<i>Figure 4: Courbe ROC</i>	21
<i>Figure 5 : Utilisation de la distance de Mahalanobis dans la classification</i>	22
<i>Figure 6: Arbre de décision</i>	25
<i>Figure 7: Camembert illustrant les nombres d'entreprises défaillantes et non défaillantes</i>	30
<i>Figure 8: Histogramme du Chiffre d'affaire des entreprises non défaillantes</i> ...	31
<i>Figure 9: Histogramme du Chiffre d'affaire des entreprises défaillantes</i>	32
<i>Figure 10: Histogramme de la variable dette à long terme pour les entreprises non défaillantes</i>	33
<i>Figure 11: Histogramme de la variable dette à long terme des entreprises défaillantes</i>	33
<i>Figure 12: Histogramme de la variable trésorerie nette des entreprises non défaillantes</i>	34
<i>Figure 13: Histogramme de la variable trésorerie nette des entreprises défaillantes</i>	35
<i>Figure 14: Nuage de points CA*EBE</i>	37
<i>Figure 15: Nuage de points CA*CAF</i>	38
<i>Figure 16: Nuage de points EBE*CAF</i>	38
<i>Figure 17: Nuage de points TRESORERIENETTE*BFR</i>	39
<i>Figure 18: Boxplot de la variable CA avant transformation</i>	42
<i>Figure 19: Boxplot de la variable Log(CA)</i>	44
<i>Figure 20: Segmentation de la variable CA à l'aide de SAS Enterprise Miner</i> .	45
<i>Figure 21: Boxplot de la variable chiffre d'affaire de l'année N par segment</i> ..	46
<i>Figure 23: Arbre de décision CART</i>	64
<i>Figure 24: Courbe ROC de la régression logistique</i>	69
<i>Figure 25: Courbe ROC de l'analyse factorielle discriminante</i>	70
<i>Figure 26: Courbe ROC de la méthode DISQUAL</i>	70
<i>Figure 27: Courbe ROC de l'arbre CART</i>	71
<i>Tableau 1: Matrice de confusion</i>	20
<i>Tableau 2: Exemple de tableau disjonctif complet</i>	24
<i>Tableau 3: Dictionnaire des variables</i>	29

Liste des figures et tableaux

Tableau 4: tableaux récapitulatifs des valeurs manquantes de l'ensemble des variables	42
Tableau 5: Ratios financiers et formules de calcul	49
Tableau 6: Tableau récapitulatif des coefficients de Spearman les plus élevés.	50
Tableau 7: Tableau récapitulatif des valeurs de rho de Pearson et de rho de Spearman	51
Tableau 8: Définitions des ratios retenus	52
Tableau 9: Tableau des p-value des tests de corrélation de Pearson.....	53
Tableau 10: Tests de normalité de la variable r	54
Tableau 11: Tests de normalité de la variable t	54
Tableau 12: Tests de normalité de la variable rb	54
Tableau 13: test de Levene pour la variable r	55
Tableau 14: Test de Levene pour la variable t	55
Tableau 15: Test de Levene pour la variable rb	55
Tableau 16: Estimation des coefficients de la régression	56
Tableau 17: Test sur le vecteur des paramètres.....	57
Tableau 18: Résultat du test de Hosmer-Lemeshow	58
Tableau 19: Test de normalité sur les résidus de Pearson.....	58
Tableau 20: Probabilité à priori estimée par SAS	59
Tableau 21: Fonctions discriminantes.....	59
Tableau 22: Statistiques de la procédure STEPDISC	60
Tableau 23: Tableau de la décomposition de l'inertie	61
Tableau 24: Coordonnées des modalités sur les axes factoriels	62
Tableau 25: Résultat de la méthode forward sur les axes factoriels	63
Tableau 26: Fonction discriminante linéaire de l'AFD sur les axes	63
Tableau 27: Tableau de distribution des scores	65
Tableau 28; Tableau d'importance des variables	66
Tableau 29: Tests d'ajustement sur la méthode CART	66
Tableau 30: Matrice de confusion de la régression logistique.....	67
Tableau 31: Matrice de confusion de l'analyse discriminante	67
Tableau 32: Matrice de confusion de la méthode DISQUAL.....	68
Tableau 33: Matrice de confusion de la méthode CART.....	68
Tableau 34: Taux de bons et de mauvais classements des méthodes.....	69
Tableau 35: Matrice de corrélation de Pearson	80
Tableau 36: Matrice de corrélation de Spearman	88
Tableau 37: Statistiques descriptives générales	89

Introduction

Le recyclage des dépôts bancaires en crédits, est l'une des principales sources de revenu des institutions bancaires. Lors de ce processus, ces dernières s'exposent à plusieurs risques dont celui de contrepartie.

Le management du risque crédit permet d'identifier et de maîtriser les pertes futures liées à ce risque. Ce thème fut le principal sujet des accords de Bale II, puisqu'ils ont introduit des exigences réglementaires en termes d'efficacité du contrôle interne. En effet, la banque se trouve dans l'obligation de suivre l'une des trois méthodes de mesure de risque crédit, la méthode standard, la méthode de notation interne et la méthode de notation interne avancée. Au niveau du Crédit agricole du Maroc, le choix s'est porté sur la troisième méthode qui consiste à effectuer une estimation interne de l'ensemble des éléments du risque crédit : la probabilité de défaut, la perte encourue en cas de défaut et l'exposition en cas de défaut.

Il existe plusieurs modèles qui permettent l'estimation de la probabilité de défaut, des méthodes paramétriques fondées sur des hypothèses statistiques et des méthodes non paramétriques, de machine learning, qui ne nécessitent aucune condition sur les variables. Dans le but de déterminer les avantages et inconvénients de ces modèles, plusieurs études ont été consacrées à leur analyse et à leur comparaison.

Ce mémoire vise à reproduire les étapes de ces recherches en implémentant le modèle de régression logistique, d'analyse factorielle discriminante, de la méthode DISQUAL et de la méthode CART.

Chapitre 1 : Contexte et problématique

1. Groupe Crédit Agricole

La caisse nationale du crédit agricole a été créée en 1894. Voici quelques dates clés de l'évolution du groupe :

- 1966 : la caisse nationale du crédit agricole est devenue autonome financièrement et s'est élargie sur le territoire français à travers l'ouverture de nombreuses agences et a développé ses activités, en diversifiant ses clients cibles (accord de prêts aux PME et PMI).
- 1979 : cet organisme a été classé parmi les meilleures banques au niveau international par la revue « The Banker ».
- 2001 : autonomie totale du groupe et introduction en bourse sous le nom de « Crédit Agricole SA ».

Depuis, le groupe n'a cessé de progresser et propose aujourd'hui ses services dans 70 pays du monde, dont le Maroc. Il compte plusieurs filiales notamment Predica (1986) pour l'assurance vie, Pacifica (1990) pour l'assurance dommage, Indosuez (1996) banque de financement d'investissement, Sofinco (1999) et Finaref (2003) pour le crédit de consommation et Amundi (2010) pour la gestion d'actifs.

2. Crédit Agricole du Maroc

Le Crédit Agricole du Maroc est l'un des principaux premiers piliers du système financier du Maroc. Sa genèse a été lancée au vingtième siècle, sous le nom de « Caisse Nationale de Crédit Agricole », avec comme principale mission la promotion et l'accompagnement du développement du système agricole national. L'organigramme général de l'organisme est représenté dans la figure 1 comme suit :

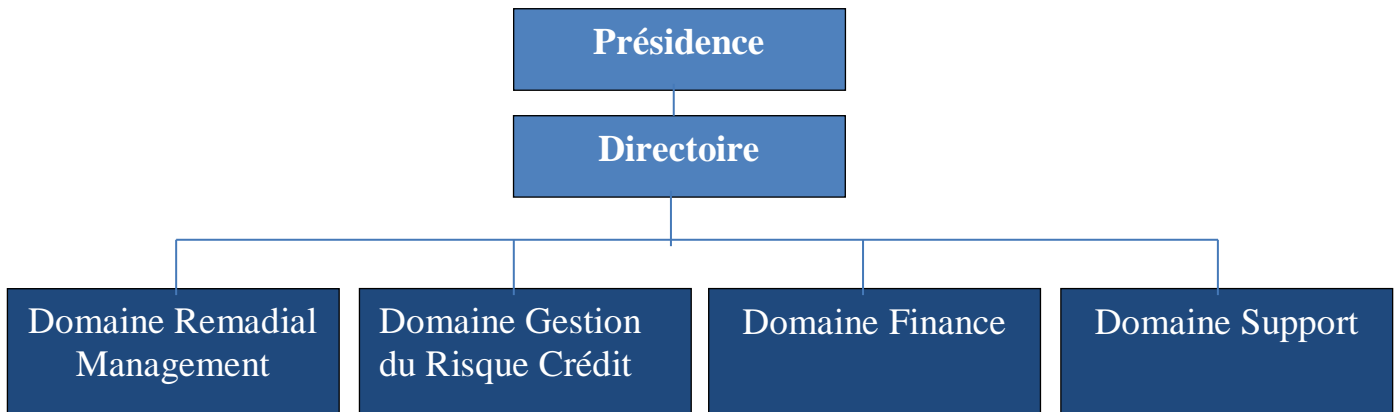


Figure 1: Organigramme général du CAM

Le stage de projet de fin d'étude, s'est déroulé au sein du **Domaine Gestion du Risque Crédit**, et il compte quatre pôles :

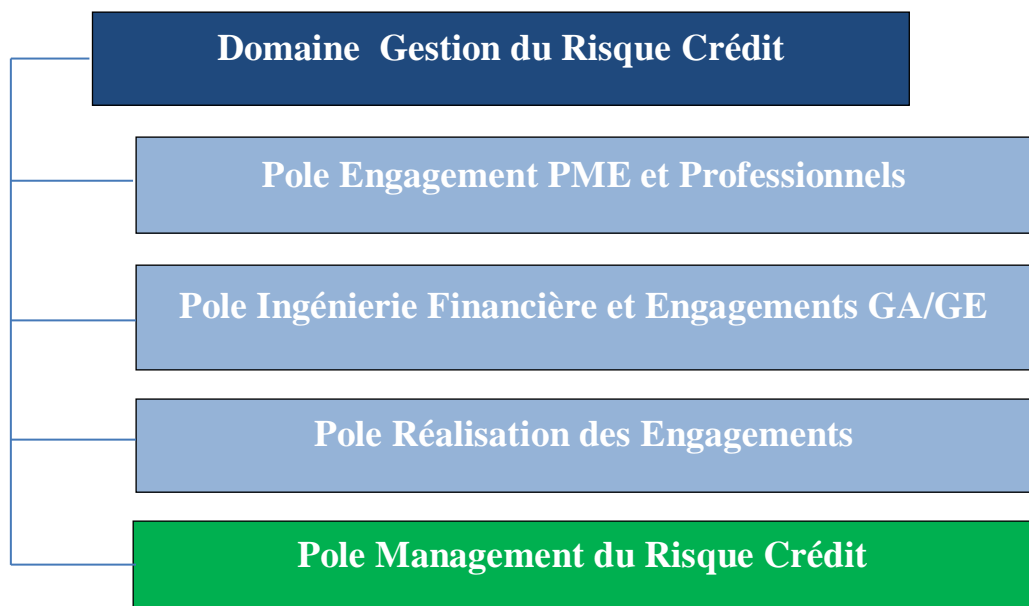


Figure 2: Organigramme du Domaine Gestion du Risque Crédit

Cette direction a pour mission l'identification, la prévoyance et la maîtrise du risque crédit. Il contrôle également le respect des directives en matière de gestion du risque crédit.

1.1 Services proposés par le CAM

Depuis sa création, le Crédit Agricole du Maroc s'est consacré au secteur de l'agriculture (loi 15-99). Comme conséquence des aléas et de l'incertitude liée au climat (sécheresse et distribution non uniforme des pluies sur le territoire), la banque a subi des pertes importantes. Encadrée par la loi bancaire 34-03 relative aux établissements de crédit, elle avait l'obligation de générer des revenus positifs et d'entretenir sa santé financière.

Dans l'optique de rétablir la situation, elle a établi une stratégie de diversification, à travers laquelle, elle a dédié 50% de son portefeuille à l'agriculture et les 50% restantes aux secteurs des particuliers et professionnels, du commerce et service, de l'industrie ainsi que d'autres secteurs d'activités.

La banque met à disposition de ses clients le pack « Tamwil El Fellah », destiné à accompagner et soutenir les entreprises fragiles, financièrement, et qui n'arrivent pas à obtenir un crédit auprès des autres banques. De plus, elle participe au programme « Tamwil Invest » dont les cibles sont les entreprises industrielles.

Elle accompagne également les TPME, du secteur des commerces, spécialisés dans les opérations de la grande et moyenne distribution, du commerce de proximité ou de réseau, du commerce indépendant et la franchise.

Cependant, les clients dont les activités s'inscrivent dans le domaine du textile, de l'hôtellerie ou de la métallurgie sont directement pris en charge par la CCG, Crédit-Capital-Garantie.

2.1 Processus d'octroi de crédit

Le processus d'octroi de crédit est composé de trois sous processus qui traitent le dossier du client et les données y afférentes pour permettre la prise de décision d'octroi ou de rejet de la demande :

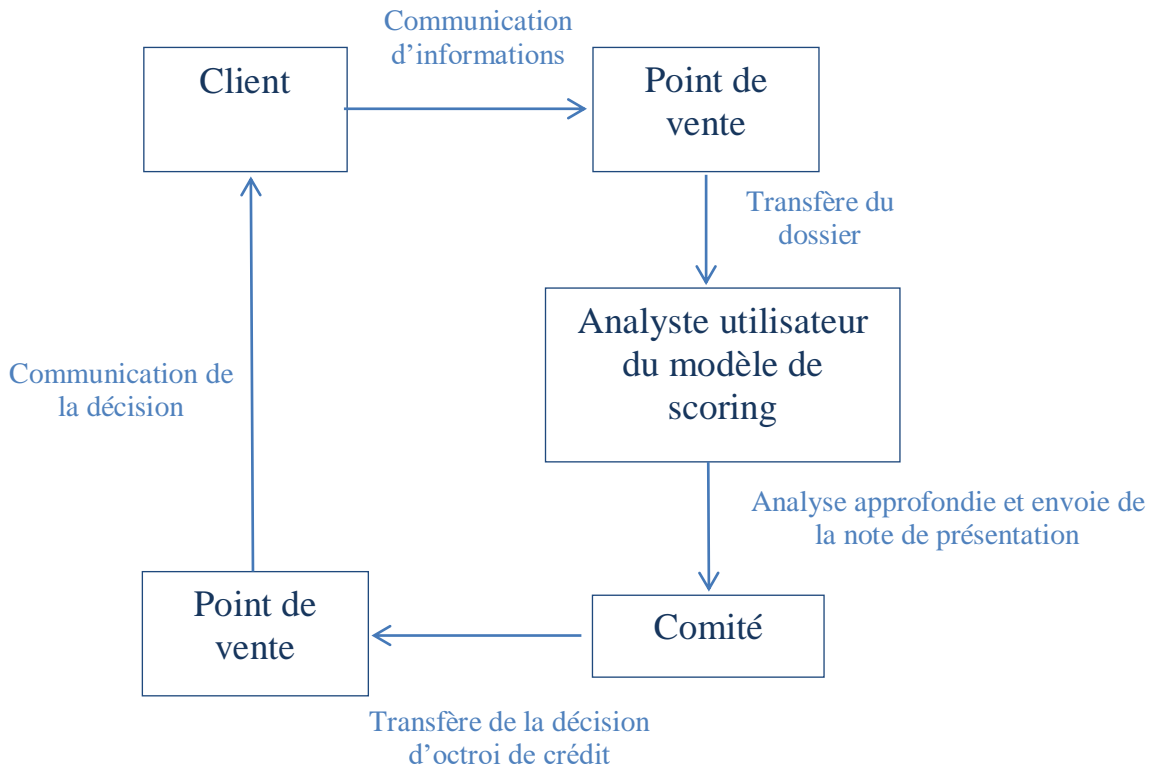


Figure 3: Processus d'octroi de crédit

- L'instruction du dossier de crédit :

Le client doit déposer une demande d'octroi de crédit auprès d'un point de vente. Ce dernier, s'entretient aussitôt avec lui afin de rassembler les informations nécessaires et les papiers indispensables pour la constitution du dossier. Le PDV prépare la note de présentation et l'envoie à l'instance concernée.

- Analyse du dossier de crédit :

Un spécialiste procède à une analyse approfondie de l'ensemble du dossier et utilise des modèles, construits à partir d'historique et de benchmarks, dans le but d'attribuer une note au client (scoring). Ses appréciations sont rédigées au sein de la note de présentation et sont transmises au comité compétent pour décision.

- **Prise de décision de crédit :**

Le comité étudie la note et veille à la cohérence des éléments qui constituent le dossier. Il signe la décision d'octroi de crédit, et la transfère au PDV.

3. Problématique :

La direction notation interne et méthodes du Pôle Management du Risque Crédit, est chargée de l'évaluation chiffrée et automatisée du dossier déposé par l'emprunteur. Au niveau de la banque, chaque client suivant des critères d'évaluation.

En étant affilié au département Modélisation Risque Crédit de la direction, nous sommes chargés de construire une grille de score robuste et fiable que l'on pourra exploiter, au niveau de l' « Analyse du dossier de crédit », à partir d'une base de données historique, sur des entreprises clientes du Crédit Agricole.

Chapitre 2 : Cadre théorique

1. Revue littéraire

Le CréditScoring n'est apparu qu'il y a environ soixante ans et il a été introduit comme étant l'ensemble des modèles de décision et des techniques sous-jacentes qui aident les prêteurs dans la décision d'octroi des crédits de consommation d'après Thomas, Edelman and Crook (2002)

La notion de score a vu le jour et a été utilisée par les banques dans l'optique d'attribuer une notation à chaque client, « Information collected by banks and/or financial institutions of a credit applicant is used to develop a numerical score for each applicant » (Thomas et al. 2002).

En Crédit Scoring, une attention particulière est portée au choix des ratios financiers retenus et employés dans l'élaboration de la grille de score. En effet, il faut s'assurer de leur fiabilité statistique pour assurer la qualité des modèles obtenus (Cf. Bardos (2001) chapitre 5 sur la préparation des données et la sélection des variables)

L'étude des méthodes de scoring a fait l'objet de plusieurs recherches. Cela a permis de découvrir plusieurs caractéristiques qui leur sont propres, tel qu'il est le cas pour l'analyse discriminante de Fisher qui n'exige pas d'hypothèses paramétriques, mais qui nécessite la régularité de la forme du nuage des points (Saporta 1990).

D'autres méthodes sont apparues, notamment la méthode DISQUAL, créée par Saporta, qui traite spécifiquement les variables qualitatives. Des travaux, comme ceux menés par L. Lelogeais en 2003, ont en fait l'application. D'autres statisticiens et spécialistes en économie ont abordé le thème de scoring, tels que

Baesens et al. (2003), Thiria (1997), Gnanadesikan, Kettenring et Tsao (1995) et McLachlan (1992).

2. Scoring

Le scoring est une méthode de notation qui permet d'affecter, à chaque individu, un score dans le but de constituer des groupes d'unités statistiques de mêmes caractéristiques. Ce projet de fin d'études est dédié aux méthodes de prédiction du non remboursement des prêts bancaires, on parle alors du *creditscoring*. Cet outil de détection de risque a pour but général le calcul de la perte globale espérée, *Expected Loss*, grâce à l'estimation de la probabilité de défaut.

Cependant, des questions se posent sur l'homogénéité des groupes formés et sur la fiabilité des estimations retrouvées. Par conséquent, nous soumettrons les modèles retenus à des tests statistiques et nous constituerons une base test afin de former une matrice de confusion (vrais positifs, faux positifs, vrais négatifs et faux négatifs).

Il existe plusieurs méthodes d'élaboration de modèles :

- Des méthodes de type linéaire, telles que l'analyse factorielle discriminante linéaire et la régression logistique.
- D'autres qui prennent en compte la non-linéarité des variables telles que l'analyse discriminante quadratiques et la régression polynomiale.
- Des méthodes non paramétriques, graphiques, dont les arbres de décisions et réseau de neurones.

Nous nous focaliserons sur les méthodes les plus utilisées dans le secteur bancaire, soient l'analyse factorielle discriminante, la régression logistique, ainsi que la méthode des arbres de décisions CART.

3. Régression logistique :

Lorsque la variable à expliquer est binaire, la relation $Y = \beta_0 + \sum_{i=1}^{\infty} x_i \beta_i$ ne prend plus de sens. Puisque la variable à expliquer prend les valeurs 0 ou 1 et que les variables explicatives prennent des valeurs réelles, adopter le modèle Logit reste plus approprié.

A travers la régression logistique, nous n'expliquons pas la variable aléatoire Y mais la probabilité qu'elle soit égale à l'unité. Le modèle s'écrit :

$$P(Y=1) = \frac{e^{\beta_0 + \sum_{i=1}^{\infty} x_i \beta_i}}{1 + e^{\beta_0 + \sum_{i=1}^{\infty} x_i \beta_i}}$$

Nous admettons la règle de décision ci-dessous :

$$\left\{ \begin{array}{l} \text{Si } \frac{P(y=1)}{P(y=0)} > 1 \text{ alors } y=1 \\ \text{Sinon } y=0 \end{array} \right.$$

3.1 Hypothèses du modèle :

Contrairement à la régression linéaire, la régression logistique n'exige pas l'homoscédasticité ($V(\varepsilon) = \text{cst}$) ou la normalité des erreurs, puisqu'elles suivent également une loi de Bernoulli. Cependant, il nécessite l'absence de colinéarité entre les x_i .

3.2 Choix des variables et du modèle :

Le modèle retrouvé, après estimation des β_i , peut contenir des variables non significatives. Les tests de Wald et de Vraisemblance sont utilisés pour les retrouver. Il est à noter que les statistiques de ces deux tests suivent une loi de Khi-deux dont le nombre des degrés de liberté est égale au nombre des paramètres du modèle testés.

En plus des variables non significatives, d'autres seront écartées grâce à la méthode backward et au test de déviance entre modèles emboîtés.

3.3 Adéquation du modèle :

Nous parvenons à tester l'adéquation du modèle choisi en :

- Effectuant le test de Hosmer-Lemeshow, dont les hypothèses testées sont H_0 : le modèle est adéquat contre H_1 : le modèle n'est pas adéquat
- En effectuant des tests de normalités sur les résidus de Pearson $r_{pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$ ou les résidus de déviance $r_{Di} = \text{signe}(y_i - \hat{\mu}_i) \sqrt{2l(y_i) - l(\hat{\mu}_i)}$. Nous concluons que le modèle est adéquat lorsque le test est positif.

3.4 Qualité d'ajustement du modèle :

Dans le but de déterminer si le modèle s'ajuste parfaitement aux données, nous poserons y_i la valeur observée et \hat{y}_i la valeur estimée par le modèle. Ensuite, nous établirons la matrice de confusion :

\hat{y}_i	0	1	Total
y_i			
0	N_{VN}	N_{FP}	Négatifs
1	N_{FN}	N_{VP}	Positifs

Tableau 1: Matrice de confusion

Avec :

- N_{VN} le nombre des Vrais Négatifs ($y_i=0$ et $\hat{y}_i=0$)
- N_{FN} le nombre des Faux Négatifs ($y_i=0$ et $\hat{y}_i=1$)

- N_{FP} le nombre de Faux Positifs ($y_i=1$ et $\hat{y}_i=0$)
- N_{VP} le nombre de Vrais Positifs ($y_i=1$ et $\hat{y}_i=1$)

Nous pouvons apprécier la qualité du modèle en calculant le coefficient de détermination $R^2 = \frac{N_{VP} + N_{VN}}{N}$. Plus il est élevé, plus les prédictions du modèle sont bonnes.

A partir de la matrice de confusion nous calculons :

- Le critère de spécificité = $\frac{FP}{\text{Négatifs}}$
- Le critère de sensibilité = $\frac{VP}{\text{Positifs}}$

Ces critères constituent la courbe ROC :

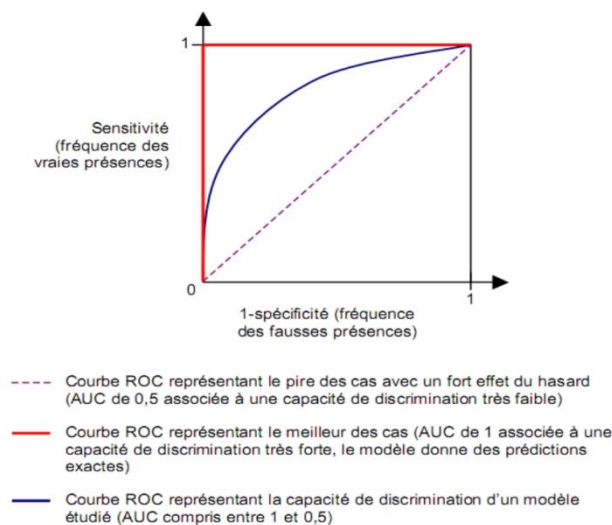


Figure 4: Courbe ROC

Comme l'indique le schéma ci-dessus, plus la surface entre la courbe et la première bissectrice est grande, plus le modèle s'ajuste bien.

4. L'analyse factorielle discriminante

L'analyse factorielle discriminante joue un rôle important dans la description des caractéristiques des variables (corrélations, existence de caractères

globaux,...etc) mais également dans la prédiction. En effet, elle permet de construire des règles de classification à partir de données dont nous disposons. En d'autres mots, nous serons capables d'assimiler de nouveaux sujets à une classe à travers des variables indépendantes, tout en minimisant le nombre des prédictions erronées.

4.1 Notion de distance

Cette technique de classification affecte une unité au groupe dont le centroïde est le plus proche. La distance utilisée est la celle de Mahalanobis.

Cette distance généralisée, est utilisée lorsque nous manipulons des variables inter corrélées de variances différentes.

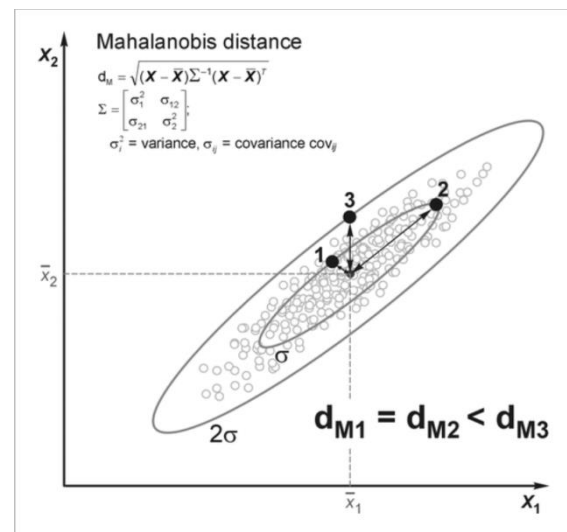


Figure 5 : Utilisation de la distance de Mahalanobis dans la classification

$$D^2 = (x_i - \bar{x}_k)' \Sigma^{-1} (x_i - \bar{x}_k)$$

Σ : Matrice de variances-covariances

4.2 Conditions d'application

- Y doit être une variable à modalités.
- Les variables explicatives sont indépendantes.
- La taille de l'échantillon doit être importante.
- Les matrices de covariances des populations doivent être homogènes.
- La distribution des variables explicatives doit être multinomiale.

Remarque :

Si les conditions de multinormalité et d'homoscédasticité ne sont pas respectées, nous pouvons quand même appliquer l'AFD sous condition que les tests de significativité soient fiables.

4.3 Estimation de la probabilité de défaut

La probabilité de défaut, recherché à travers le scoring, est la probabilité à postériori. Cette dernière peut être estimée à partir du théorème de Bayes qui met en relation les probabilités à priori, d'appartenance aux groupes de la variable à expliquer, et les probabilités conditionnelles $P(x/G_i)$. La formule de calcul est :

$$P(G_i / x) = \frac{P(x/G_i)}{\sum_{i=1}^k P(x/G_i).P(G_i)}$$

Avec

- k le nombre de modalité de la variable à expliquer
- G_i le groupe de la modalité i de la variable à expliquer
- $P(G_i)$ les probabilités à priori d'appartenance au groupe i
- $P(x/G_i)$ les probabilités conditionnelles

4.4 Qualité du modèle

Dans l'optique de mesurer la qualité du modèle, nous procédons au calcul du lambda de Wilks. Ce dernier permet de définir si les groupes formés par l'AFD ont des moyennes significativement différentes. En effet, cette statistique contribue à l'appréciation du pouvoir discriminant des variables sélectionnées par l'analyse discriminante.

5. La méthode DISQUAL

La méthode DISQUAL est une méthode qui se base sur l'analyse factorielle discriminante. En effet, elle consiste à appliquer une analyse de correspondance multiple sur les données et de récupérer les axes factoriels les plus discriminants. Ensuite, il suffit de faire une analyse factorielle discriminante avec comme nouveau jeu de données les coordonnées des individus sur les axes, nouvelles variables explicatives du modèle.

5.1 L'analyse de correspondance multiple :

L'analyse de correspondance multiple consiste à constituer des combinaisons linéaires, de variables qualitatives. L'utilisation de cette méthode requiert la transformation du tableau de données initiales en tableau de données disjonctif complet.

	m_1			m_2		m_3		
<i>Ind</i>	1	2	3	1	2	1	2	3
1	0	1	0	0	1	1	0	0
2	1	0	0	1	0	0	1	0
...								
...								
<i>i</i>	0	0	1	0	1	0	0	1
<i>N</i>	0	0	1	1	0	0	0	1

Tableau 2: Exemple de tableau disjonctif complet

Parmi les axes générés nous retenons les n premiers qui captent un minimum de 70% de la variance totale. Chaque axe A_i est défini par les variables qui lui sont le plus corrélés.

On écrit :

$A_i = \sum_j^{\infty} \alpha_j v_j$ avec v_j : variables qui définissent l'axe et α_j leurs coordonnées sur les axes.

5.2 L'analyse factorielle discriminante

La deuxième étape de la méthode DISQUAL consiste à appliquer une AFD sur une nouvelle base de données comportant les axes factoriels retenus de l'ACM.

6. Les arbres de décisions : CART

La méthode CART est l'une des nombreuses méthodes des arbres de décisions. C'est une méthode graphique simple à manipuler avec laquelle nous parvenons à appuyer les résultats des méthodes statistiques.

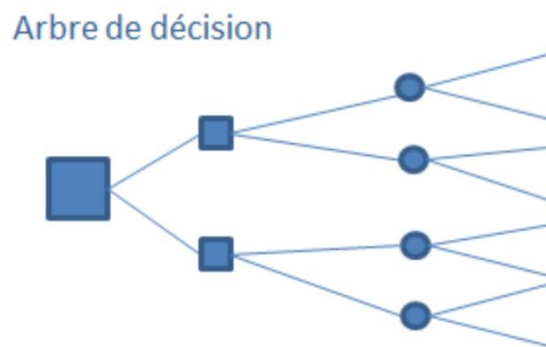


Figure 6: Arbre de décision

Un arbre de décision est constitué de nœuds, de branches et de feuilles. Un nœud correspond à une des variables explicatives retenues et donne lieu à une division qui génère des nœuds fils. Mais quelle variable choisir et quel est le critère de division ? L'indice de Gini permet de répondre à toutes ces questions.

A la racine :

A cette étape, l'algorithme détermine la variable qui constituera le premier nœud de l'arbre. Pour cela, il calcule l'indice de Gini avant séparation comme suit :

$$\mathbf{Gini(avant\ séparation)} = 1 - \sum_1^2 f_i^2$$

avec f_i les fréquences des valeurs 0 et 1 prises par la variable à expliquer

Puisqu'elles sont segmentées, l'algorithme calcule l'indice de Gini pour chaque variable :

$$\mathbf{Gini(mod}_j) = 1 - \sum_1^2 f_i^2$$

avec f_i les fréquences de 0 et de 1 pour les observations de la modalité j

$$\mathbf{Gini(var)} = \mathbf{Gini(avant\ séparation)} - \sum_1^n \mathbf{Gini(mod}_j)$$

Parmi toutes ces divisions possibles, CART sélectionne celle qui maximise l'indice de Gini.

Au niveau des autres nœuds :

Cette fonction se réitère jusqu'à ce qu'il n'y ait plus de divisions admissibles. Les derniers éléments de l'arbre sont appelés « feuilles » et sont attribués à une modalité de Y , et le chemin qui lie la racine à chaque feuille correspond aux « branches ».

Commentaire :

La méthode des arbres de décisions CART est apparue en 1984. Suite aux inconvénients de la méthode, plusieurs de ses variantes sont apparues par la suite, dont Bagging en 1996 et Random Forest en 2001. En effet, les résultats de la classification par la méthode CART et ses prédictions dépendent fortement

de l'ordre des variables d'entrée. Par conséquent, cette méthode est qualifiée d'instable.

L'algorithme Random Forest réduit la variance des résultats en moyennant les prévisions de plusieurs arbres de décisions, formés aléatoirement et entraînés sur des jeux de données différents. On en déduit, que cette méthode se base sur des choix aléatoires des variables explicatives à couper, ce qui la rend un des classifieurs les plus efficaces.

7. Comparaison des modèles

La probabilité de défaut ne sert pas seulement à noter les clients mais également à estimer la perte de crédit attendue. Cette dernière représente les pertes, auxquelles peut faire face une banque, lorsque la contrepartie n'honore pas ses engagements sur un crédit. Il n'est pas intéressant d'étudier la perte attendue au niveau d'un seul client mais au niveau d'un portefeuille. La perte attendue totale est définie comme étant la somme des pertes attendues individuelles.

Trois composantes interviennent dans le calcul de la perte attendue : la probabilité de défaut, l'exposition en cas de défaut et la perte encourue en cas de défaut. La relation utilisée dans le processus d'estimation est la suivante :

$$\mathbf{EL = PD \times EAD \times LGD}$$

Avec :

EL : Expected Loss

PD : Probability of Default

EAD : Exposure At Default

LGD : LossGiven Default

La probabilité de défaut, à un horizon d'un an, reflète les chances qu'un client ne rembourse pas la totalité de ses dettes. L'exposition en cas de défaut est le montant que la banque pourrait perdre lorsque le client est défaillant, tandis que la perte encourue en cas de défaut représente le pourcentage de ce montant qui n'a pas été récupéré grâce aux garanties et saisies d'actifs.

Afin de faire face à cette perte, la banque doit constituer des provisions et ce sous les critères de l'IFRS 9, entrée en vigueur dès le 1 janvier 2018. On en déduit que pour obtenir une estimation se rapprochant le plus de la perte attendue réelle, une estimation, la plus exacte possible de la probabilité de défaut est nécessaire. Cette dernière est réalisée par des modèles robustes et fiables, d'où l'importance du contrôle du pouvoir prédictif des modèles, qui s'affirment, et l'intérêt de la comparaison entre eux.

Comparer les modèles de scoring obtenus, se fera à l'aide de la matrice de confusion et de la courbe ROC comme indiqué dans le paragraphe « qualité d'ajustement » de la partie dédiée à la régression logistique.

Les matrices de confusion contribuent aux calculs des taux de bons classements et de mauvais classements de l'ensemble des techniques. Tandis que les courbes ROC, permettent de construire un tableau récapitulatif des AUC afin de les comparer par la suite. Le choix du modèle optimale se fait donc sur la base des taux de classements et des surfaces AUC.

Chapitre 3 : Modélisation et implémentation

1. Traitement de la base de données et calcul des ratios

Le portefeuille qu'a mis à notre disposition le Crédit Agricole comprend 1982 entreprises et nous détenons 22 variables comptables reflétant la situation financière de ces dernières. Nous avons également la possibilité de savoir si chacune d'elles a remboursé, ou pas son crédit auprès de la banque par le biais de la variable *Défaut*.

Variable	Description
NF_CHIFFREAFFAIRE_N	Chiffres d'affaires année N
NF_CHIFFREAFFAIRE_N_1	Chiffres d'affaires année N-1
NF_DETTECOURTERME	Dettes Court terme
NF_DETTELONGTERME	Dettes long terme
NF_ACTIFCIRCULANT	Actif circulant
NF_EBE	Excedent Brut d'Exploitation
NF_TRESORERIE NETTE	Trésorerie nette
NF_TOTALBILAN	Total Bilan
NF_RESULTATNET_N	Résultat Net
NF_RESULTAT_NET_N_1	Résultat Net N-1
NF_FPNETS	Fonds Propres Nets
NF_KPROPRES	Capitaux propres
NF_FINANPERMANENT	Financements permanents
NF_FDR	Fond de roulement
NF_BFR	Besoin en Fond de Roulement
RESULTAT_COURANT_N	Résultat courant
RESULTAT_COURANT_N_1	Résultat courant N-1
RESULTAT_EXP_N	Résultat d'exploitation
RESULTAT_EXP_N_1	Résultat d'exploitation N-1
CAPACITE_AUTO_FIN	Capacité d'Autofinancement
TRESORERIE_ACTIF	Trésorerie actif
PASSIF_CIRCULANT	Passif circulant
DEFAULT	Défaillance du client

Tableau 3: Dictionnaire des variables

Remarque :

Il aurait été plus intéressant pour nous de faire des analyses par secteur d'activité mais nous ne détenons pas cette information.

Les entreprises défaillantes et non défaillantes sont réparties de la sorte :

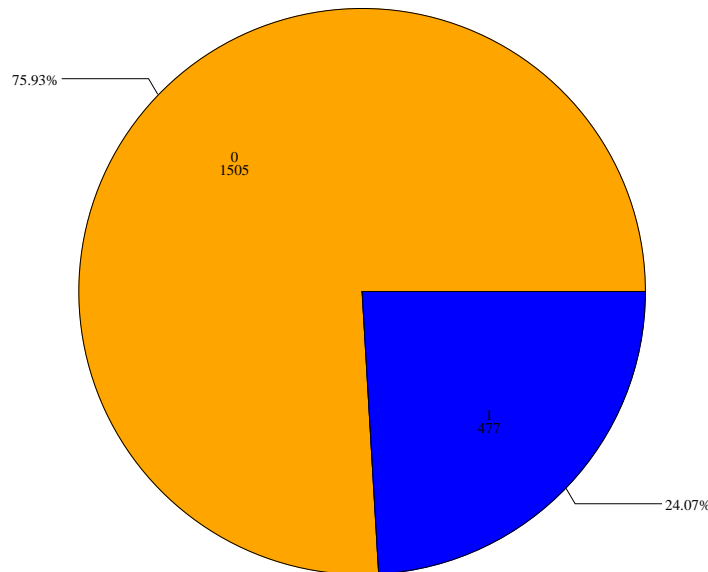


Figure 7: Camembert illustrant les nombres d'entreprises défaillantes et non défaillantes

1.1 Statistiques descriptives sur les variables

Avant d'effectuer n'importe quel traitement sur la base de données, il est impératif de générer des statistiques descriptives. Elles nous permettront de mieux analyser les variables, dont nous disposons, et d'observer les relations qui existent entre elles. Pour cela, nous distinguons entre les statistiques descriptives univariées, qui traitent variable par variable, et les statistiques descriptives bivariées qui traitent les variables deux à deux.

Statistiques univariées :

Les premiers indicateurs auxquels nous prêterons attention sont: le Kurtosis, mesure d'épaisseur de la queue de distribution, et le Skewness, mesure de l'asymétrie de la loi de distribution. Nous exposerons les statistiques des variables les plus pertinentes.

Chiffre d'affaire de l'année N

Nos observations nous permettent de constater que le Skewness prend une valeur positive et que le Kurtosis est très élevé. De plus, le minimum est égal à 0 et le maximum est égal à 23677899.30. Sachant que 50% des entreprises ont un chiffre d'affaire inférieur à 9585.28 et que la variance se rapproche de $4.5 \cdot 10^{11}$, on conclut facilement que les valeurs prises par cette variable varient énormément et qu'il y a un grand écart entre les scores les plus petits et les plus élevés.

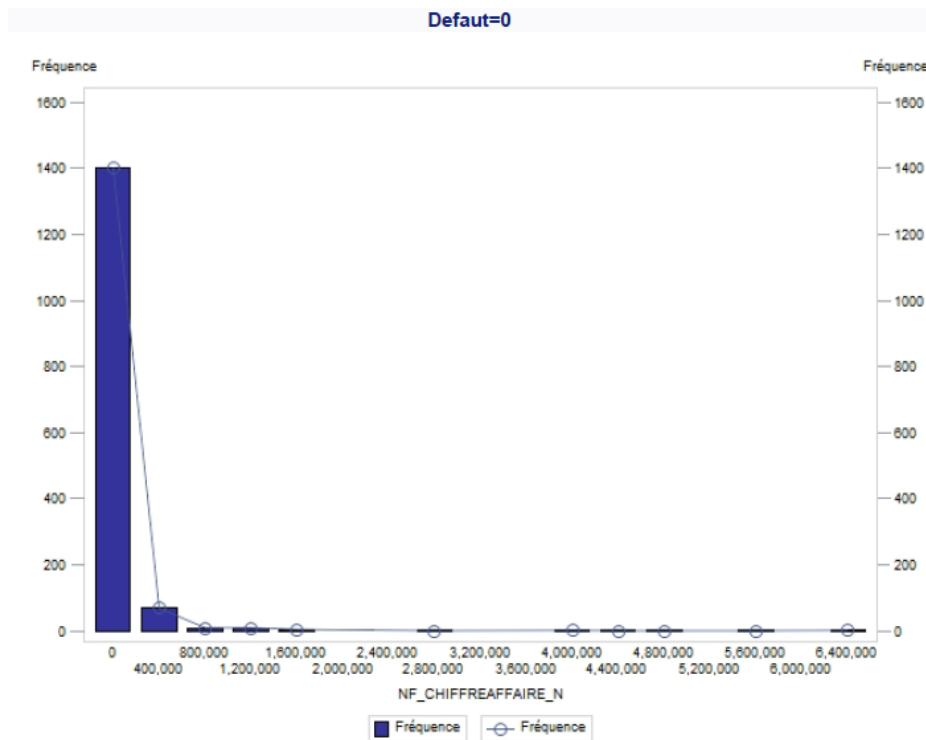


Figure 8: Histogramme du Chiffre d'affaire des entreprises non défaillantes

L'histogramme du chiffre d'affaire de l'année N pour les entreprises non défaillantes, montre qu'à peu près 1400 d'entre elles ont un chiffre d'affaire

inférieur à 200 000DH. Nous déduisons également, à partir du graphe que peu d'entreprises ont un chiffre d'affaire supérieur à 2.400.000DH.

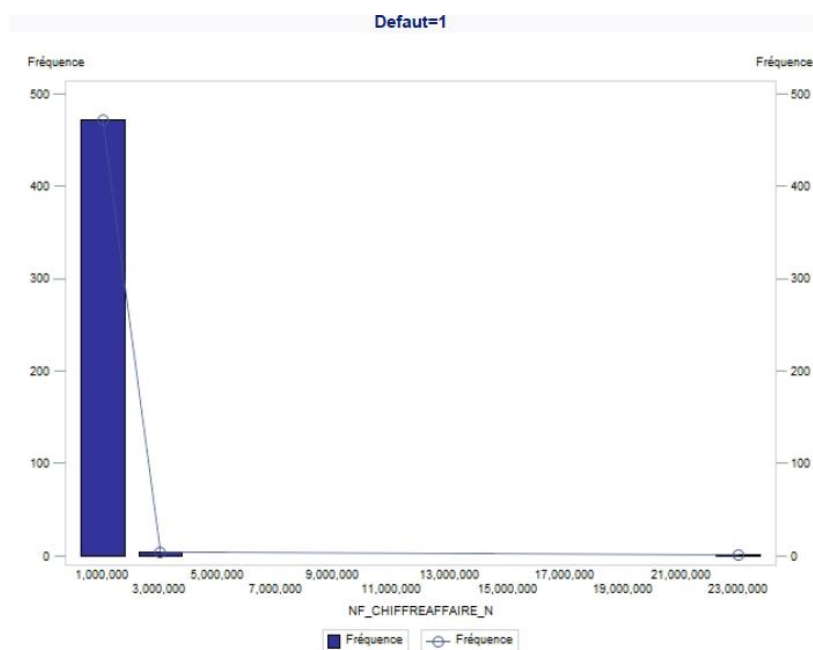


Figure 9: Histogramme du Chiffre d'affaire des entreprises défaillantes

En ce qui concerne les entreprises qui ont fait défaut, plus ou moins 480 entreprises détiennent un chiffre d'affaire inférieur à 2 000 000DH avec une observation qui se démarquent des autres et dont le chiffre d'affaire est supérieur à 23.000.000DH. A partir des résultats des deux histogrammes, nous soupçonnons la présence de valeurs aberrantes.

Dette à long terme

Dans le cas de la variable dette à long terme, le Kurtosis est de 1890.38259 et le Skewness est de 43.0160022. En portant attention aux quantiles, nous nous apercevons que la moitié des observations prennent des valeurs inférieures ou égales à 0. En effet, parmi 1982 observations, 1062 sont nulles.

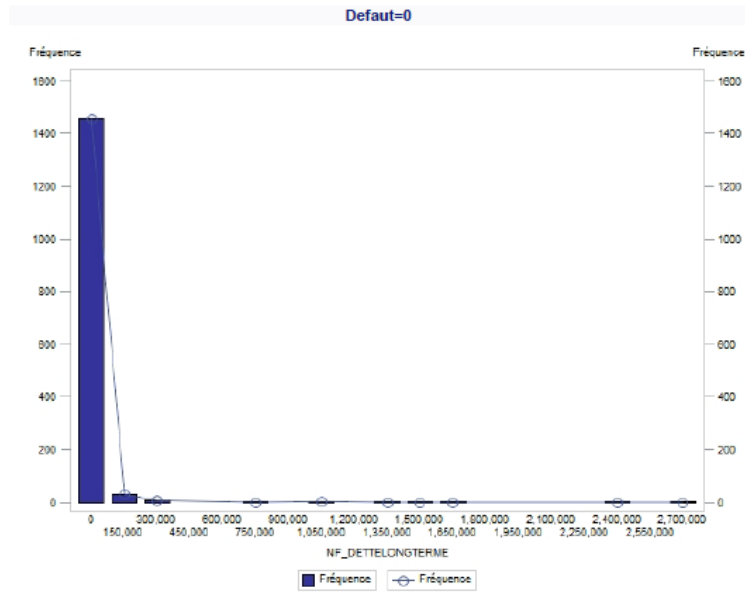


Figure 10: Histogramme de la variable dette à long terme pour les entreprises non défaillantes

Un peu plus de 1420 entreprises non défaillantes ont des dettes à long terme inférieur à 75.000DH. En ce qui concerne les autres entreprises, la dette maximale enregistrée est supérieure à 2.700.000DH.

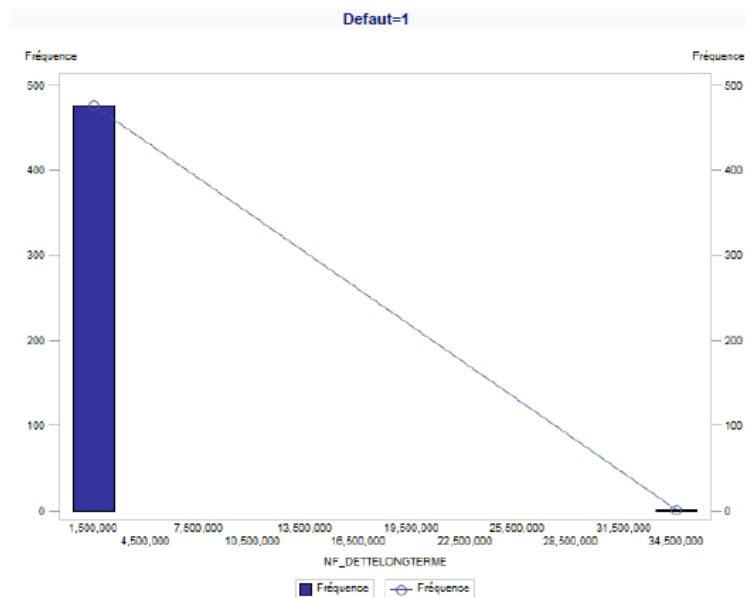


Figure 11: Histogramme de la variable dette à long terme des entreprises défaillantes

Seulement très peu des entreprises défailtantes ont une dette à long terme supérieure à 34.500.000DH. Le même constat, fait pour la variable chiffre d'affaire de l'année N, s'impose pour la variable dette à long terme.

Trésorerie nette

Les valeurs que prend cette variable sont comprises entre -49.169 et $2.70861 \cdot 10^6$ et son coefficient de variation est 528.753966. De plus, trente pourcent de la population, soit 599 entreprises, ont une trésorerie nette nulle. Par conséquent, nous déduisons que leurs fonds suffisent à peine à répondre aux dépenses, générées par leur cycle d'exploitation, et à leurs autres besoins.

En termes de Kurtosis et de Skewness, nous faisons les mêmes remarques que pour les variables précédentes.

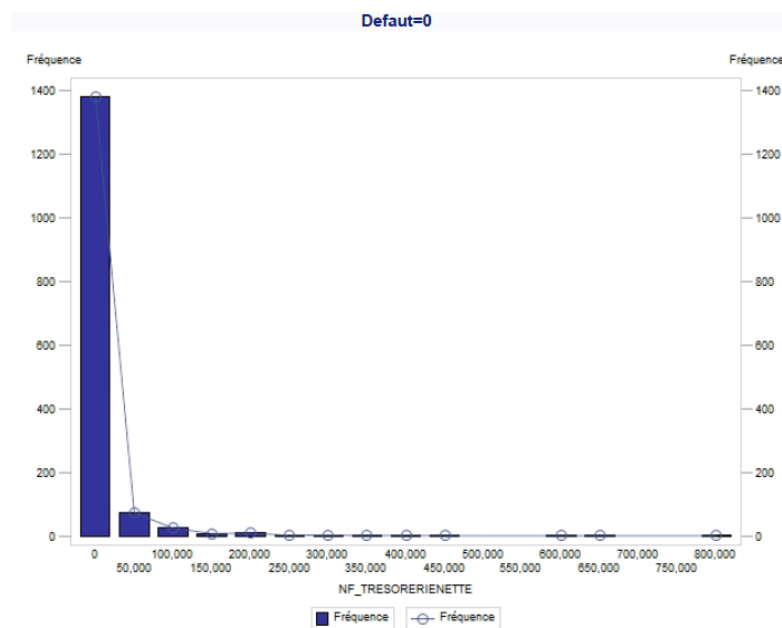


Figure 12: Histogramme de la variable trésorerie nette des entreprises non défailtantes

Plus de 90% des entreprises non défailtantes (1390 entreprises) cumule une trésorerie nette inférieure à 25.000DH.

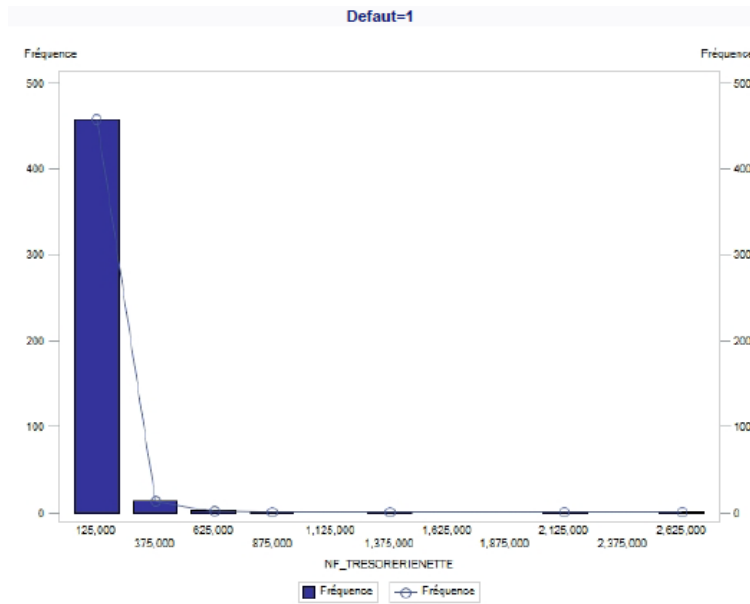


Figure 13: Histogramme de la variable trésorerie nette des entreprises défaillantes

Malgré leur défaillance, certaines entreprises cumulent une trésorerie nette supérieure à 2.626.000DH, ceci est signe d'une mauvaise gestion. Le reste des entreprises ont une trésorerie nette inférieure à 150.500DH. Nous suspectons encore la présence de valeurs aberrantes.

Statistiques bivariées :

Les indicateurs comptables les plus observés en entreprise sont la marge brute, l'excédent brut d'exploitation, la capacité d'autofinancement, la trésorerie nette et le besoin en fonds de roulement. Dans le but de déterminer les relations possibles entre ces variables, nous allons tout d'abord les définir et étudier de près les corrélations par la suite.

La marge brute est égale à la différence entre le chiffre d'affaires hors taxes et le coût d'achat des marchandises. Elle permet le calcul du **Taux de marge**

$\left(\frac{\text{marge brute}}{\text{CA HT}}\right)$ qui est le gain de l'entreprise par rapport à son chiffre d'affaires.

Puisque nous ne disposons pas d'assez d'informations pour pouvoir la calculer, nous allons analyser le **chiffre d'affaires**. Cette grandeur reflète la part d'une entreprise du marché et renseigne sur sa taille. De plus, sa variation donne une indication sur la santé de l'activité de cette dernière.

L'EBE renvoie une image sur la capacité de l'entreprise à générer des ressources du seul fait de son cycle d'exploitation.

La CAF rassemble les ressources que lui procure l'ensemble de son activité, sans prendre en considération ses autres modes de financement.

Le BFR exprime les besoins en fonds de l'entreprise lors des périodes de décalage de trésorerie.

La trésorerie nette représente la différence entre la trésorerie active (Actifs circulants) et la trésorerie passive (Dettes financières à court terme) et correspond à l'argent disponible à un instant T, et dont dispose l'entreprise. Le suivi régulier de cet indicateur garantit une gestion efficace de cette dernière.

Après avoir défini l'ensemble des variables qui nous intéressent, nous avons examiné la matrice de corrélation de Pearson et de Spearman. Le coefficient de corrélation de Pearson est un outil pertinent pour déterminer s'il existe une relation linéaire entre deux variables quantitatives.

En revanche, le coefficient de corrélation de Spearman détecte également les relations non linéaires.

EBE et CA

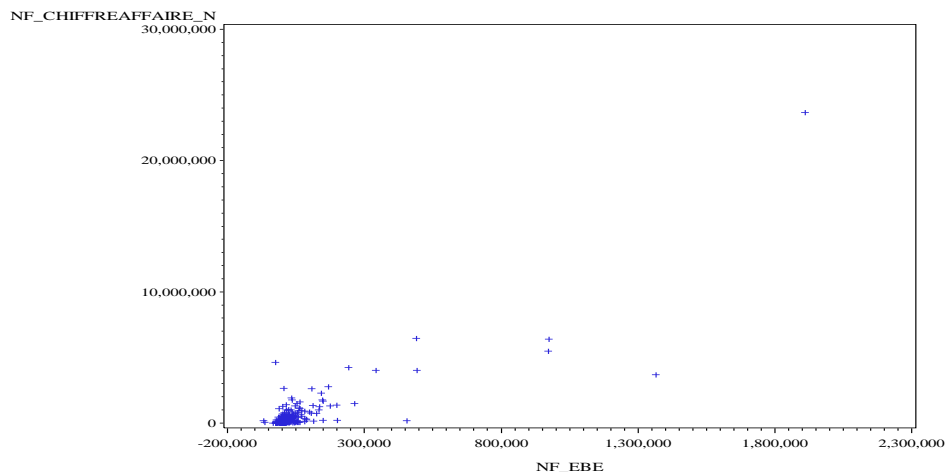


Figure 14: Nuage de points CA*EBE

$$\rho_P=0.85519 \quad \rho_S=0.73433$$

Ces deux variables sont positivement corrélées (le coefficient de corrélation est 0.85519). Plus l'excédent brute d'exploitation augmente plus le chiffre d'affaires augmente. En effet, ceci s'explique par le biais de la relation comptable suivante :

CA = EBE + achat de marchandises et de matières premières + services extérieurs + services provenant des tiers + impôts et autres taxes + salaires et charges sociale

CA et CAF

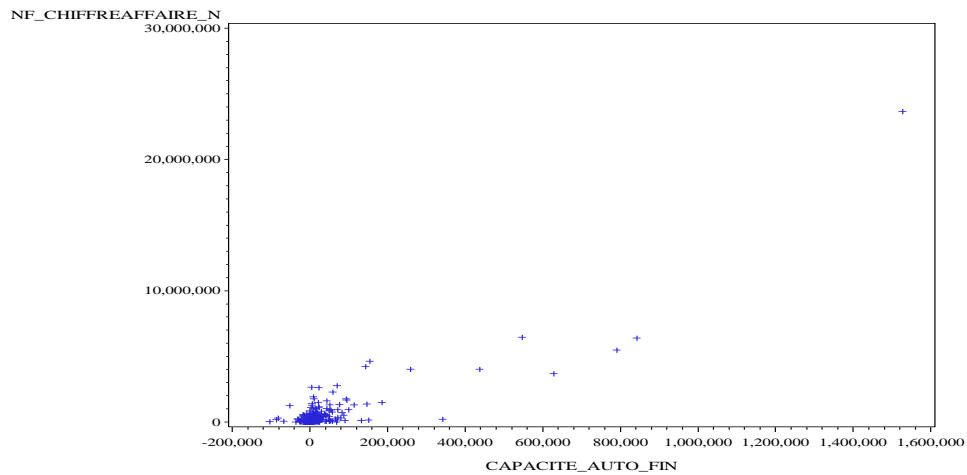


Figure 15: Nuage de points CA*CAF

$$\rho_P = 0.88877 \quad \rho_S = 0.62998$$

Au sein de l'entreprise, la capacité d'autofinancement aide à financer une partie du cycle d'exploitation. Par conséquent, plus elle augmente plus le nombre des unités produites augmente et plus le chiffre d'affaire s'accroît.

EBE et CAF

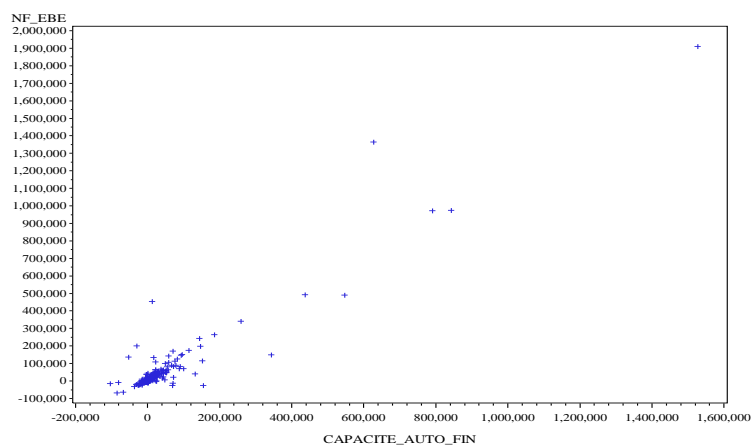


Figure 16: Nuage de points EBE*CAF

$$\rho_P = 0.94393 \quad \rho_S = 0.86337$$

Il existe plusieurs méthodes qui permettent le calcul de la capacité d'autofinancement. L'une d'entre elles est la méthode soustractive :

CAF = EBE- autres charges (correspondant à des décaissements à l'exception des charges calculées (dotations) et des valeurs comptables des immobilisations cédées) + autres produits (correspondant à des encaissements à l'exception des produits calculés et des produits de cession d'immobilisations).

La capacité d'autofinancement et l'excédent brut d'exploitation sont fortement corrélés. En effet, il apparaît clairement sur le nuage des points l'existence d'une relation linéaire entre ces deux variables.

Trésorerie nette et BFR

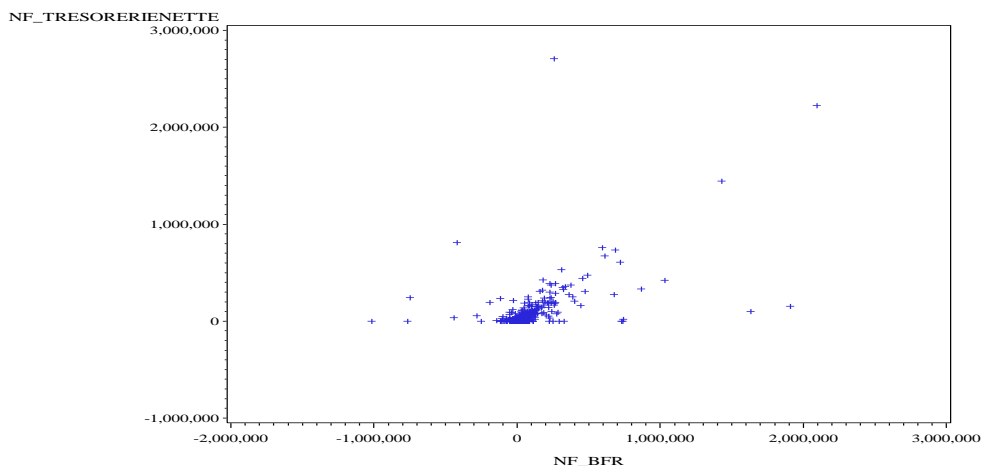


Figure 17: Nuage de points TRESORERIENETTE*BFR

$$\rho_P = 0.56394 \rho_S = 0.50667$$

La trésorerie nette est la différence entre les fonds de roulement et le besoin en fonds de roulement. Cependant, ces deux grandeurs ne sont pas fortement corrélées.

Nous remarquons de manière générale, que l'indicateur de Skewness est différent de zéro et que l'indicateur de Kurtosis est différent de 3. Il est clair que nous ne pouvons modéliser aucune variable par la loi normale. Au contraire, les scores très élevés retenus pour le Kurtosis, et les scores positifs retenus pour le Skewness, indiquent une distribution asymétrique à queue épaisse avec aplatissement faible. Ceci est signe de présence de valeurs extrêmes.

Nous notons également que la variable « Trésorerie Nette » est celle qui varie le moins, alors que la variable « Dette à Long Terme » est celle qui varie le plus, selon le coefficient de variation.

Les nuages de points et les coefficients de corrélation montrent que les indicateurs clés sont positivement corrélés, et qu'il existe des valeurs aberrantes parmi les données.

1.2 Traitement de la base

Le traitement de la base de données permet de repérer les valeurs manquantes et aberrantes/extrêmes de celles-ci par des méthodes statistiques et graphiques. En plus, il envisage une imputation de ces valeurs par la moyenne, médiane, ...etc.

Les valeurs manquantes :

Nous définissons « valeurs manquantes » comme étant toute case de la base de données comportant un vide ou toute observation non cohérente, exemple : retrouver un nombre pour la variable « Identifiant » à la place d'une chaîne de caractères ou retrouver dans la base une entreprise dont le chiffre d'affaires est négatif.

Defaut	Identifiant	NF_CHIFFREAFF AIRE_N	NF_CHIFFREAFFA IRE_N_1
0	0	0	0

NF_DETTECOUR TTERME	NF_DETTELONG TERME	NF_ACTIFCIRC ULANT	PASSIF_CIRC ULANT
0	0	0	0

NF_EBE	NF_TRESORER IENETTE	NF_TOTALBILA N	NF_RESULTAT NET_N
0	0	0	0

NF_RESULTAT_ NET_N_1	NF_FPNETS	NF_KPROPRE S	NF_FINANPERM ANENT
0	0	0	0

NF_FDR	NF_BFR	RESULTAT_COU RANT_N	RESULTAT_COU RANT_N_1
0	0	0	0

RESULTAT_EX P_N	RESULTAT_EXP _N_1	CAPACITE_AU TO_FIN	TRESORERIE_ ACTIF
0	0	0	0

Tableau 4: tableaux récapitulatifs des valeurs manquantes de l'ensemble des variables

Les sorties ci-dessus montrent qu'il n'y a aucune valeur manquante. Nous ne constatons également aucune observation incohérente.

Les valeurs aberrantes :

Dans le but de retrouver les valeurs aberrantes, nous userons de méthodes graphiques : boxplots et histogrammes. En ne considérant que les données brutes, les tracés obtenus ne sont pas concluants. Exemple :

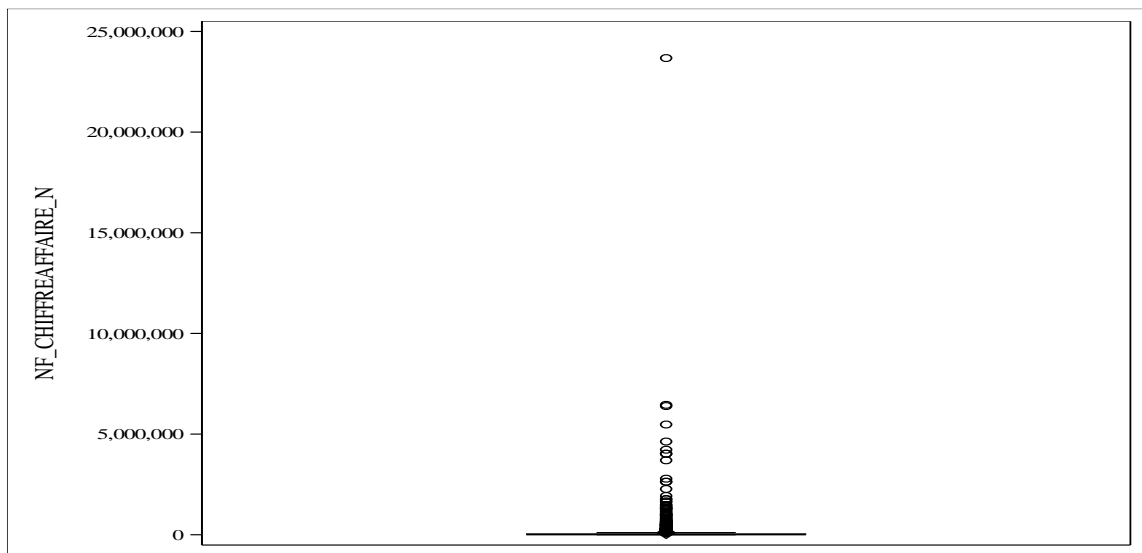


Figure 18: Boxplot de la variable CA avant transformation

Nous avons donc songé à deux alternatives afin de palier à ce problème :

- Appliquer le logarithme sur toutes les variables comptables pour faire diminuer les écarts entre les observations.
- Segmenter chaque variable grâce à SAS Enterprise Miner et tracer le boxplot par groupe.

Remarque :

- Nous appliquerons le logarithme sur la valeur absolue des variables car quelques-unes contiennent des valeurs négatives.
- La segmentation au niveau des variables ne servira qu'à la détection des valeurs aberrantes. Une autre segmentation est envisagée après le calcul des ratios financiers.

Transformation par Logarithme

Pour réaliser cette transformation, plusieurs étapes sont à suivre :

- Appliquer le logarithme sur la valeur absolue des variables.
- Former une nouvelle base à partir de l'initiale et remplacer les valeurs manquantes par la moyenne.
- Engendrer les boxplot et calculer les seuils: $\min(\max, Q_3 + 1.5(Q_3 - Q_1))$ et $\max(\min, Q_3 - 1.5(Q_3 - Q_1))$.
- Remplacer dans la base initiale toutes les valeurs supérieures au premier seuil et inférieures au deuxième seuil par la moyenne des valeurs non aberrantes.
- Employer la fonction inverse du logarithme.

Pour appliquer ces étapes sur la variable CHIFFRE_AFFAIRE_N, nous avons suivi cinq démarches :

- Nous avons appliqué le logarithme sur la variable NF_CHIFFREAFFAIRE_N.

- Selon la base « str » nous avons remplacé les valeurs manquantes par 4.0972076723.
- $Q_1 = 3.56$ $Q_3 = 4.63$

$$\min(\max, Q_3 + 1.5(Q_3 - Q_1)) = 4.63 + 1.5(4.63 - 3.56) = 6.235$$

$$\max(\min, Q_1 - 1.5(Q_3 - Q_1)) = 3.56 - 1.5(4.63 - 3.56) = 1.955$$

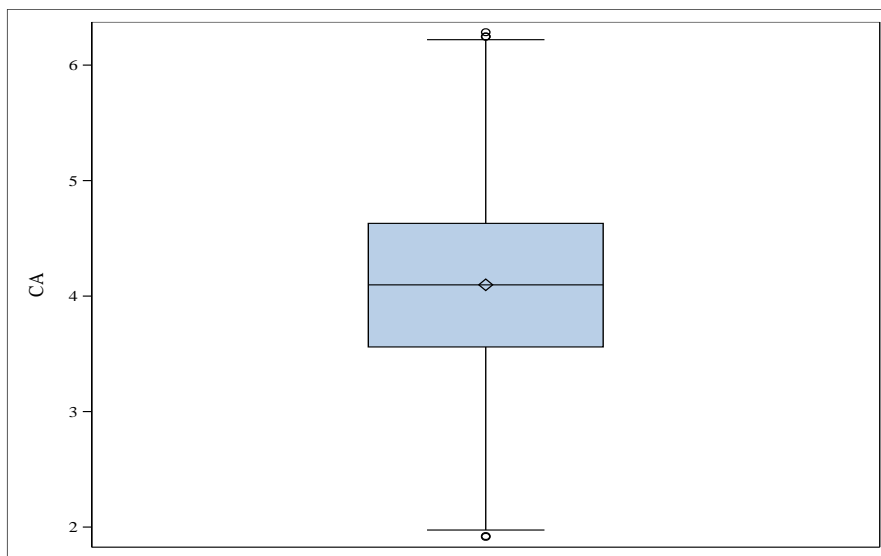


Figure 19: Boxplot de la variable Log(CA)

- Nous avons remplacé les valeurs aberrantes par 4.0977697419.
- Nous avons basculé vers les valeurs initiales en calculant $10^{\text{Log(CA)}}$.

Traitement après segmentation par SAS Enterprise Miner

Comme précédemment, plusieurs étapes s'imposent :

- Segmenter les variables avec SAS Enterprise Miner en entrant comme variable à expliquer la variable Défaut et comme variable explicative la variable comptable à traiter.
- Constituer les groupes.
- Générer les boxplots par segment.

- Calculer les seuils.
- Imputer les valeurs aberrantes par la moyenne des valeurs non aberrantes.

Appliquons les étapes précédentes sur la variable CHIFFRE_AFFAIRE_N :

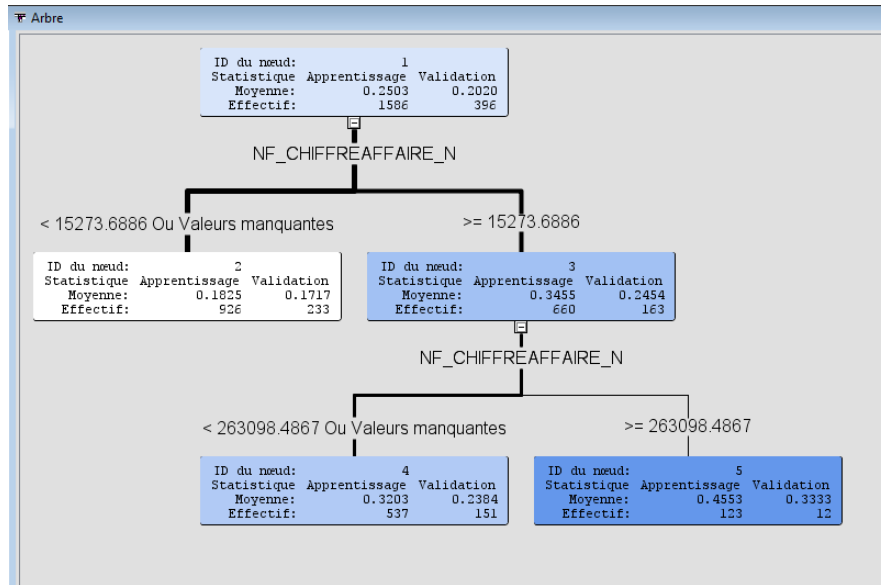


Figure 20: Segmentation de la variable CA à l'aide de SAS Enterprise Miner

Le premier groupe rassemble les entreprises dont le chiffre d'affaire est inférieur strictement à 15273.6886.

Le deuxième est constitué des entreprises dont le chiffre d'affaire est compris entre 15273.6886 et 263098.4867.

Tandis que le troisième comprend les entreprises dont le chiffre d'affaire est supérieur strictement à 263098.4867.

- nous avons généré les boxplots par segments après avoir trier la base selon le groupe.

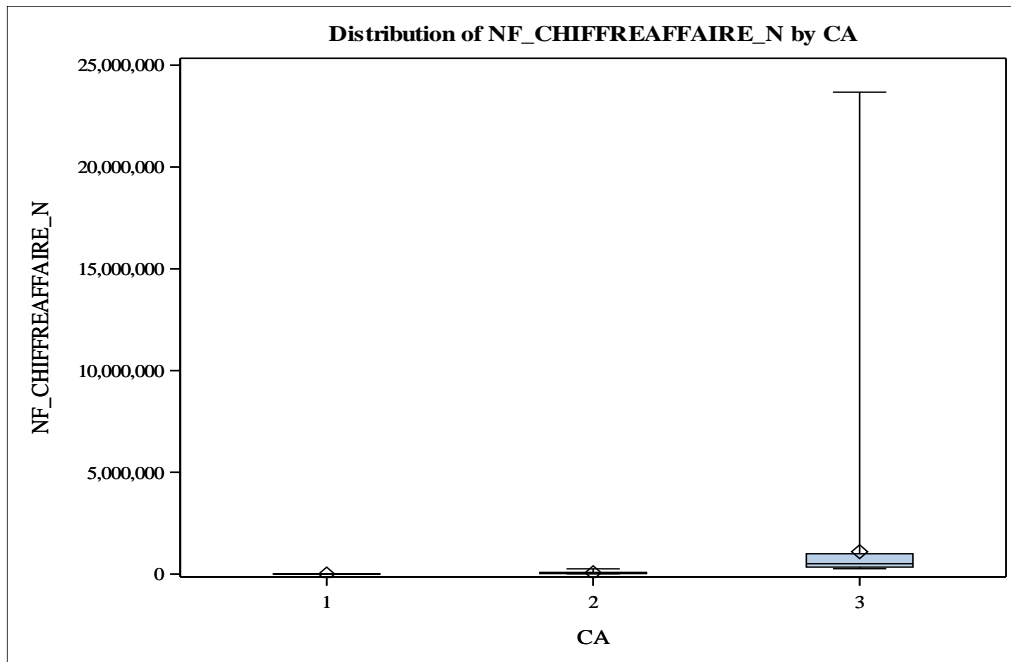


Figure 21: Boxplot de la variable chiffre d'affaire de l'année N par segment

Nous remarquons clairement que le dernier groupe contient probablement des valeurs aberrantes.

- $Q_1 = 344328$ $Q_3 = 1E6$

min (max, $Q_3 + 1.5(Q_3 - Q_1)$) = $1E6 + 1.5 (1E6 - 344328) = 1983643$

max (min, $Q_3 - 1.5(Q_3 - Q_1)$) = $344328 - 1.5 (1E6 - 344328) = -639180$

- nous avons remplacé les valeurs non comprises entre les seuils, du troisième groupe, par la moyenne de ses valeurs non aberrantes, 620415.37016.

1.3 Calcul des ratios financiers

Nous n'utiliserons pas les indicateurs comptables pour la constitution du modèle mais des ratios financiers que nous calculerons à partir des composantes de la base de données. Dans le but d'exploiter les données dont nous disposons de manière optimale, nous avons calculé de nouvelles variables :

Chapitre 3. Modélisation et implémentation

- Dotations aux amortissements= capacité d'autofinancement – résultat net
- Disponibilité= trésorerie nette + dettes à court terme
- Actifs immobilisés= capacité d'autofinancement – fonds de roulement
- Dette totale= trésorerie passive + dettes à long terme + dettes à court terme
- Endettement net= dettes à court terme + dettes à long terme - disponibilités

Voici un tableau récapitulatif des ratios que nous avons calculés :

Ratio	formule	code
Deliquidité générale	$\frac{\textit{Actif circulant}}{\textit{Passif circulant}}$	RLG
De financement de l'actif circulant	$\frac{\textit{Fonds de roulement nets}}{\textit{Actif circulant}}$	RFAC
D'activité	$\frac{\textit{Chiffre d'affaire HT N} - \textit{Chiffre affaire HT N - 1}}{\textit{Chiffre affaire HT N - 1}}$	RA
Taux de rentabilité net N	$\frac{\textit{Résultat net N}}{\textit{Chiffre affaire N}}$	TRN
Taux de rentabilité net N-1	$\frac{\textit{Résultat net N - 1}}{\textit{Chiffre affaire N - 1}}$	TRN1

Chapitre 3. Modélisation et implémentation

Rentabilité des capitaux	$\frac{\text{Résultat net } N}{K_{\text{propres}}}$	RC
Rentabilité de l'activité	$\frac{\text{Capacité d'autofinancement}}{\text{Chiffre affaire } N}$	RA
D'endettement général	$\frac{\text{Dettes totales}}{\text{Capacité d'autofinancement}}$	REG
Marge d'EBE	$\frac{\text{EBE}}{\text{Chiffre affaire } N}$	MEBE
Rentabilité de l'exploitation N	$\frac{\text{Resultat d'exploitation } N}{\text{Chiffre d'affaire HT } N}$	RE
Rentabilité de l'exploitation N-1	$\frac{\text{Resultat d'exploitation } N - 1}{\text{Chiffre d'affaire HT } N - 1}$	RE1
Capacité de remboursement	$\frac{\text{Endettement net}}{\text{Capacité d'autofinancement}}$	CDR
Rotation du BFR	$\frac{\text{BFR} * 360}{\text{Chiffre d'affaire HT } N}$	RBFR
Trésorerie immédiate	$\frac{\text{Disponibilités}}{\text{Dette à court terme}}$	TI

D'immobilisation de l'actif	$\frac{\text{Actifs immobilisés}}{\text{Total bilan}}$	RIDA
Taux de résultat courant N	$\frac{\text{Résultat courant } N}{\text{Chiffre d'affaire HT } N}$	TDRC
Taux de résultat courant N-1	$\frac{\text{Résultat courant } N}{\text{Chiffre d'affaire HT } N}$	TDRC1
Couverture des emplois stables	$\frac{\text{Financement permanent}}{\text{Actifs immobilisés}}$	CDES
Couverture des capitaux investis	$\frac{\text{Financement permanent}}{\text{Actifs immobilisés} + \text{BFR}}$	CDCI
D'indépendance financière	$\frac{K_{\text{propres}}}{\text{Financement permanent}}$	RIF

Tableau 5: Ratios financiers et formules de calcul

Nous retrouvons des valeurs manquantes après le calcul des ratios, pour certaines observations, comme est le cas du ratio d'activité lorsque le chiffre d'affaires N-1 est nul. Nous pouvons supprimer ces valeurs, si elles sont peu nombreuses. Procéder de la sorte, pour chaque ratio, reviendra peut être à éliminer beaucoup d'observations dans l'ensemble. En l'occurrence, il serait préférable de faire abstraction de ces ratios.

Pour les données traitées par le logarithme :

Lorsque nous procédons à la suppression, la base ne contient plus que 1752 observations. Nous avons donc perdu 11.6% des données. Puisque ce pourcentage est non significatif devant la taille de la base de données, nous décidons de garder, pour le moment, l'ensemble des ratios.

Les modèles que nous voulons exploiter, à savoir le Logit et le modèle issu de l'AFD, nécessitent l'indépendance des variables explicatives. Dans le but de repérer les ratios corrélés entre eux, nous utilisons les coefficients de Spearman et de Pearson, présentés précédemment.

La matrice de corrélation des variables, mettant en avant le coefficient de *Pearson*, a révélé des corrélations entre :

- Les variables TDRC1 et RE1 avec $\rho_p=0.91955$
- Les variables TDRC et RE avec $\rho_p=0.79671$

En revanche, la matrice de corrélation utilisant le coefficient de *Spearman*, a affiché des résultats plus poussés :

Variables corrélées	Valeur de ρ_s
CDES et RIDA	-0.74299
RLG et RFAC	0.80910
RE et TRN	0.75706
MEB et RTA	0.83846
MEBE et RE	0.76723
TDRC et TRN	0.81566
TDRC1 et TRN1	0.81363
TDRC1 et RE1	0.89937

Tableau 6: Tableau récapitulatif des coefficients de Spearman les plus élevés

Chapitre 3. Modélisation et implémentation

Lorsque deux grandeurs sont corrélées, nous ne gardons que l'une des deux, afin d'éviter la redondance. Par conséquent, nous retenons les variables : RFAC, RBFR, RTA, TRN, RE, TRN1, RE1 et RIDA.

Nous exploitons la méthode backward, de type pas à pas, pour la sélection des variables. Elle fait intervenir l'ensemble des variables lors de la première étape et ensuite élimine au fur et à mesure la variable la moins significative jusqu'à l'obtention d'un modèle dont toutes les variables explicatives sont significatives. La procédure GENSELECT de SAS, permet de mettre en pratique la méthode et ne retient que la variable RIDA.

Pour les données traitées par segmentation :

La suppression des valeurs manquantes a réduit le nombre d'observation de 14.32%. Nous décidons dans ce cas de garder l'ensemble des ratios. Les matrices de corrélation de Pearson et de Spearman ont détecté les corrélations suivantes :

Spearman		Pearson	
Variabes corrélées	Valeur de ρ_s	Variabes corrélées	Valeur de ρ_p
RIDA/CDES	-0.77105	RTA/TDRC	0.99967
RE/TRN	0.77272	MEBE/TDRC	0.99977
TDRC/TRN	0.82123	TDRC/RE	0.99993
RE1/TRN1	0.76386	RBFR/TDRC1	0.85867
MEBE/RTA	0.84870	TRN1/TDRC1	0.99850
RE/MEBE	0.77348	RTA/RE	0.99962
TDRC1/TRN1	0.84432	MEBE/RE	0.99983
TDRC1/RE1	0.90443	MEBE/RTA	0.99975
		TRN1/RBFR	0.85134

Tableau 7: Tableau récapitulatif des valeurs de rho de Pearson et de rho de Spearman

Nous maintenons les ratios RIDA, RE, RBFR, RTA, RE1, RC, TI, CDR, CDCI, RIF, RLG, RFAC et RA. La procédure SAS a retenu, dans le cas du traitement par segmentation, trois variables : RIDA, TI et RBFR.

Nous décidons d'appliquer les modèles uniquement sur les données traitées par segmentation, suite aux résultats de la sélection backward. En effet, supposé qu'un seul ratio puisse prédire la probabilité de défaut serait absurde.

Nous présentons, dans le tableau ci-dessous, les définitions des ratios que nous avons retenus pour la modélisation.

Ratio	Définition
RIDA	Il renseigne sur le poids de l'actif immobilisé net par rapport au total d'actif, plus il est élevé plus l'entreprise est confronté au risque de non liquidité.
TI	C'est un indicateur de solvabilité à court terme. Plus il est élevé plus la situation financière de l'entreprise est favorable.
RBFR	Il mesure l'évolution du BFR par rapport au chiffre d'affaire en jours. S'il est élevé, cela signifie que le BFR sera important en cas de croissance.

Tableau 8: Définitions des ratios retenus

Malgré les coefficients de corrélation des variables peu élevés, nous avons effectués le test de corrélation de Pearson qui a été négatif pour l'ensemble des variables (tableau 8).

Variables	RIDA	TI	RBFR
RIDA	-	0.2045	0.1566
TI	0.2045	-	0.1134
RBFR	0.1566	0.1134	-

Tableau 9: Tableau des p-value des tests de corrélation de Pearson

Nous avons songé à discrétiser les variables explicatives non seulement pour appliquer la méthode DISQUAL mais encore pour améliorer la performance des modèles. A l'aide de SAS Enterprise Miner, nous avons généré les modalités suivantes :

RIDA	<ul style="list-style-type: none"> Si $RIDA < 0.0636$ alors $r=1$ Si $0.0636 \leq RIDA < 0.632$ alors $r=2$ Si $0.632 \leq RIDA < 0.9515$ alors $r=3$ Sinon $r=4$
TI	<ul style="list-style-type: none"> Si $TI < 1.0016$ alors $t=1$ Si $1.0016 \leq TI < 1.1962$ alors $t=2$ Sinon $t=3$
RBFR	<ul style="list-style-type: none"> Si $RBFR < 175.116$ alors $rb=1$ Sinon $rb=2$

Les derniers tests que nous avons effectués sur les variables explicatives sont les tests de normalité et d'homogénéité des groupes, afin de vérifier les hypothèses de l'analyse factorielle discriminante. Les tests de normalité des variables utilisés, sont ceux de Shapiro-Wilk, de Kolmogorov-Smirnov, de Cramer-vonMises et d'Anderson-Darling. Quant à l'homogénéité des groupes, notre choix s'est porté sur le test de Levene en raison de sa fiabilité.

Normalité

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.791207	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.297116	Pr > D	<0.0100
Cramer-von Mises	W-Sq	26.7125	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	142.1061	Pr > A-Sq	<0.0050

Tableau 10: Tests de normalité de la variable r

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.78815	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.251213	Pr > D	<0.0100
Cramer-von Mises	W-Sq	17.0624	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	121.4681	Pr > A-Sq	<0.0050

Tableau 11: Tests de normalité de la variable t

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.526407	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.474626	Pr > D	<0.0100
Cramer-von Mises	W-Sq	66.7426	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	355.0048	Pr > A-Sq	<0.0050

Tableau 12: Tests de normalité de la variable rb

Les tests de normalité sont tous négatifs pour les trois variables discrétisées.

Homogénéité des groupes

Levene's Test for Homogeneity of Defaut Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
r	3	0.3543	0.1181	2.85	0.0363
Error	1435	59.4503	0.0414		

Tableau 13: test de Levene pour la variable r

Levene's Test for Homogeneity of Defaut Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
t	2	7.5607	3.7804	127.12	<.0001
Error	1436	42.7042	0.0297		

Tableau 14:Test de Levene pour la variable t

Levene's Test for Homogeneity of Defaut Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
rb	1	0.7212	0.7212	17.49	<.0001
Error	1437	59.2564	0.0412		

Tableau 15: Test de Levene pour la variable rb

Les tests d'homogénéité des groupes pour les variables se sont montrés négatifs mis à part pour la variable r.

2. Régression logistique

La procédure LOGISTIC de SAS appliquée aux ratios retenus discrétisés, fourni les sorties suivantes :

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7087	0.2089	11.5123	0.0007
r	1	1	-0.6977	0.2193	10.1204	0.0015
r	2	1	-0.3938	0.2079	3.5879	0.0582
r	3	1	-0.0548	0.2309	0.0564	0.8122
t	1	1	-1.2210	0.1329	84.4363	<.0001
t	2	1	0.3962	0.0982	16.2772	<.0001
rb	1	1	-0.2680	0.0729	13.5155	0.0002

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
r 1 vs 4	0.158	0.034	0.726
r 2 vs 4	0.214	0.047	0.971
r 3 vs 4	0.301	0.065	1.396
t 1 vs 3	0.129	0.086	0.194
t 2 vs 3	0.651	0.496	0.855
rb 1 vs 2	0.585	0.440	0.779

Tableau 16: Estimation des coefficients de la régression

Les paramètres de signe négatif indiquent que les modalités influencent négativement la variable *Défaut* comme est le cas de la modalité r1, r2 ou encore rb1. nous nous intéressons, dans ce sens, aux signes des paramètres et non à leur valeur, puisqu'ils n'ont pas d'interprétation économique.

Si nous fixons les modalités r2, t1, t2, t3 et rb1, l'Odds ratio de la modalité r1 est $\exp(-0.6977)=0.50 < 1$. Nous en concluons que si le Ratio d'immobilité des actifs d'une entreprise est inférieur à 0.0636, les chances que cette dernière fasse défaut sont multipliées par 0.5. De même pour les modalités t1 et t2, dont leur Odds ratio sont respectivement $\exp(-1.221)=0.2949$ et $\exp(0.3962)=1.4861$. Il

est clair que plus le ratio TI augmente, plus les chances que le client soit défaillant augmentent.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1685.511	1520.843
SC	1690.783	1557.745
-2 Log L	1683.511	1506.843

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	176.6674	6	<.0001
Score	156.9834	6	<.0001
Wald	129.3722	6	<.0001

Tableau 17: Test sur le vecteur des paramètres

Les tests de Wald, de vraisemblance et de score, indiquent que le vecteur des beta est différent de zéro, cependant dans l'ensemble, les variables sont significatives. Toutefois, les résultats du test de Student pour chaque modalité indiquent que la modalité r3 et r2 sont non significatives au seuil de 5%.

La méthode pas à pas, backward, effectuée pour le choix des variables montre que le meilleur modèle est celui qui comporte toutes les modalités.

Partition for the Hosmer and Lemeshow Test					
Group	Total	Defaut = 1		Defaut = 0	
		Observed	Expected	Observed	Expected
1	108	3	5.66	105	102.34
2	195	14	13.59	181	181.41
3	114	13	11.50	101	102.50
4	110	27	24.00	83	86.00
5	215	61	58.91	154	156.09
6	114	38	35.13	76	78.87
7	63	21	21.82	42	41.18
8	225	78	82.54	147	142.46
9	158	64	67.30	94	90.70
10	137	72	70.56	65	66.44

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
3.2509	8	0.9176

Tableau 18: Résultat du test de Hosmer-Lemeshow

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.79493	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.320996	Pr > D	<0.0100
Cramer-von Mises	W-Sq	24.55971	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	123.0645	Pr > A-Sq	<0.0050

Tableau 19: Test de normalité sur les résidus de Pearson

Notre modèle s’ajuste parfaitement à nos données, selon le test de Hosmer et Lemeshow avec une p-value supérieure à 5%, mais pas selon le test de normalité sur les résidus de Pearson, avec une p-value inférieure à 5% pour tous les tests.

Le modèle issu de la régression logistique est le suivant :

$$\pi(\mathbf{x}) = \frac{e^{-0.7087 - 0.6977 \cdot 1_{\{r=1\}} - 1.2210 \cdot 1_{\{t=1\}} + 0.3962 \cdot 1_{\{t=2\}} - 0.2680 \cdot 1_{\{rb=1\}}}}{1 + e^{-0.7087 - 0.6977 \cdot 1_{\{r=1\}} - 1.2210 \cdot 1_{\{t=1\}} + 0.3962 \cdot 1_{\{t=2\}} - 0.2680 \cdot 1_{\{rb=1\}}}}$$

3. L'analyse discriminante

La procédure DISCRIM de SAS affiche les résultats ci-dessous :

Class Level Information					
Default	Variable Name	Frequency	Weight	Proportion	Prior Probability
0	_0	1048	1048	0.728284	0.728284
1	_1	391	391.0000	0.271716	0.271716

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
3	-3.15333

Tableau 20: Probabilité à priori estimée par SAS

Les probabilités à priori, d'appartenance aux groupes estimées par le logiciel, sont de 0.728284 et 0.271716, respectivement pour le groupe d'entreprises non défailtantes et défailtantes.

Generalized Squared Distance to Default		
From Default	0	1
0	0.63413	3.19309
1	1.22123	2.60599

Linear Discriminant Function for Default		
Variable	0	1
Constant	-11.72136	-15.92387
r	5.06479	5.40167
t	2.41804	3.27731
rb	7.43648	7.97135

Tableau 21: Fonctions discriminantes

Chapitre 3. Modélisation et implémentation

Nous déduisons, à partir du tableau des fonctions discriminantes linéaires, les fonctions de classement :

- $D(0,X) = -11.72136 + 5.06479*r + 2.41804*t + 7.433648*rb$
- $D(1,X) = -15.92387 + 5.40167*r + 3.27731*t + 7.97135*rb$

Et la fonction de score :

Attribut	Paramètres de la fonction de score
Constant	-4.20251
R	0.33688
T	0.85927
Rb	0.53487

$$D(X) = -4.20251 + 0.33688*r + 0.85927*t + 0.53487*rb$$

Si $D(X_i) > 0$ alors l'entreprise sera affectée au groupe défaillant.

Multivariate Statistics					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.895783	55.65	3	1435	<.0001
Pillai's Trace	0.104217	55.65	3	1435	<.0001
Average Squared Canonical Correlation	0.104217				

The SAS System			
The STEPDISC Procedure			
Backward Elimination: Step 1			
Statistics for Removal, DF = 1, 1435			
Variable	Partial R-Square	F Value	Pr > F
r	0.0082	11.88	0.0006
t	0.0770	119.75	<.0001
rb	0.0085	12.25	0.0005

No variables can be removed.

Tableau 22: Statistiques de la procédure STEPDISC

Le modèle retenu par la procédure STEPDISC comporte l'ensemble des variables introduites. Cependant, le lambda de Wilks est très élevé, ceci est dû aux valeurs de la variable *Defaut* qui sont issues d'une simulation.

	RIDA	TDRRC	TDRCC1	CDES	CDCI	RIF	r	t	rb	_0	_1	Defaut	
1	0.0394620096	0.0559347732	0.0089170936	8.4685212165	0.9267422571	0.9947212368	1	2	2	0.7459906207	0.2540093793	0	
2	0.0937655889	0.0282065762	0.0238437461	1.9654673201	0.4460893458		1	2	3	1	0.6025416692	0.3974583308	0
3	0.1133271502	0.0238437461	0.013895777	1.3410801082	0.553550785		1	2	2	1	0.7816553114	0.2183446886	0
4	0.179237739	0.0943349538	0.0773981934	2.8658332915	0.7302426886		1	2	3	2	0.4703332679	0.5296667321	1
5	0.1946928244	0.0773981934	0.0719730342	2.5247544969	0.7073333147		1	2	3	2	0.4703332679	0.5296667321	1
6	0.0594922406	0.1056762523	0.0770533326	10.666476646	0.8464039889		1	1	3	1	0.6798195925	0.3201804075	0
7	0.0110244525	-0.012697996	0.006582786	43.68652734	1.0873055481		1	1	1	2	0.8739796865	0.1260203135	0
8	0.0018116331	0.006582786	0.0104908309	256.40539354	0.8963560546		1	1	2	2	0.7459906207	0.2540093793	0
9	0.0015323013	0.0104908309	0.0046941758	169.60046145	0.842060445		1	1	2	2	0.7459906207	0.2540093793	0
10	0.0962465108	0.0037273504	0.0179750699	2.6122118288	1.6072132239	0.995992916	2	1	1	1	0.8942219612	0.1057780388	0
11	0.1823322703	-0.033507039	-0.106905451	0.6748057683	0.4000315383	0.3300755214	2	3	1	1	0.6025416692	0.3974583308	0
12	0.0334012521	0.2589238011	-0.023902661	8.5159428822	1.3420878886	0.957947997	1	1	1	1	0.9221186381	0.0778813619	0
13	0.1168887512	0.1144892431	0.0389532088	3.8218084711	2.9813724477	0.9813823671	2	1	1	1	0.8942219612	0.1057780388	0

Les probabilités à postériori ont été calculées par la procédure suivant le critère de Bayes (_0 et _1 dans la table). La nouvelle variable *Defaut* qui apparait dans la base, correspond aux prédictions du modèle pour chaque entreprise.

4. Méthode DISQUAL

Nous allons d'abord apprécier les résultats de l'ACM puis ceux de l'AFD appliquée au nouveau jeu de données généré à partir des axes factoriels.

1.1 L'ACM appliquée au jeu de données

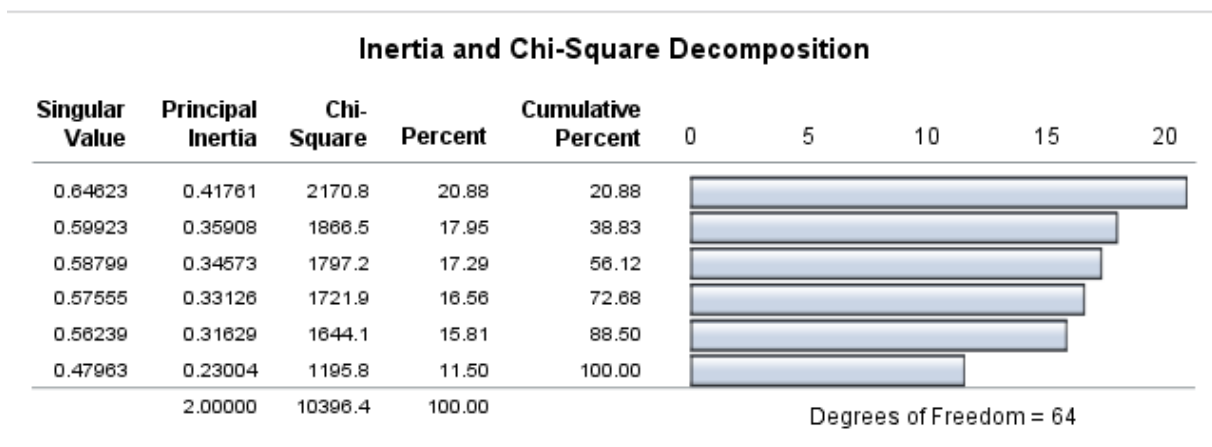


Tableau 23: Tableau de la décomposition de l'inertie

Nous avons retenu les cinq premiers axes factoriels puisqu'ils détiennent 88.50% de l'information.

Les variables qui constituent les axes factoriels sont celles qui lui sont le plus corrélées. Pour cela, nous avons observé le tableau 19, ci-dessous, et nous avons attribué une modalité à l'axe sur lequel ses coordonnées sont les plus grandes en valeur absolue. Exemple : pour la modalité 1 de r, son coefficient le plus élevé est 1.3604 sur l'axe 2. Par conséquent cette dernière est l'une des composantes de l'axe.

Column Coordinates					
	Dim1	Dim2	Dim3	Dim4	Dim5
1	0.3578	1.3604	0.1265	-0.1145	0.2128
2	0.0535	-0.5575	-0.2362	0.4182	-0.4405
3	-0.9667	-0.6280	0.3847	-1.8870	1.1473
4	-1.8378	-0.9890	8.3065	7.8013	6.6083
1	-0.4678	0.3917	-1.1229	0.4413	0.6719
2	-0.5591	0.4185	0.8048	-0.1592	-0.8891
3	0.8243	-0.6483	0.1689	-0.1988	0.2425
1	-0.4496	-0.0812	-0.0461	0.0447	-0.0387
2	1.4531	0.2624	0.1491	-0.1445	0.1251

Tableau 24: Coordonnées des modalités sur les axes factoriels

Par le biais de la même démarche nous parvenons à écrire :

- $A1 = 0.8243 * t3 - 0.4496 * rb1 + 1.4531 * rb2$
- $A2 = 1.3604 * r1 - 0.5575 * r2$
- $A3 = -1.1229 * t1 + 8.3065 * r4$
- $A4 = -1.8870 * r3$
- $A5 = -0.8891 * t2$

1.2 L'analyse discriminante appliquée aux axes factoriels

Forward Selection Summary									
Step	Number In	Entered	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Squared Canonical Correlation	Pr > ASCC
1	1	A3	0.0819	150.12	<.0001	0.91810496	<.0001	0.08189504	<.0001
2	2	A1	0.0139	23.70	<.0001	0.90534746	<.0001	0.09465254	<.0001
3	3	A5	0.0026	4.30	0.0382	0.90303619	<.0001	0.09696381	<.0001
4	4	A2	0.0015	2.47	0.1163	0.90171119	<.0001	0.09828881	<.0001

Tableau 25: Résultat de la méthode forward sur les axes factoriels

La procédure STPEDISC n'a sélectionné que les axes 1, 2, 3 et 4. Nous avons donc lancé l'analyse discriminante pour les quatre axes.

Linear Discriminant Function for Default		
Variable	0	1
Constant	-0.55227	-1.68592
A1	0.31188	0.78106
A2	0.14501	-0.02976
A3	-0.65652	-0.07218
A4	-0.60193	-0.86854

Tableau 26: Fonction discriminante linéaire de l'AFD sur les axes

Les fonctions de classement sont :

- $D(0,X) = -0.55227 + 0.31188 * A1 - 0.14501 * A2 - 0.65652 * A3 - 0.60193 * A4$
- $D(1,X) = -1.68592 + 0.78106 * A1 - 0.02976 * A2 - 0.07218 * A3 - 0.86854 * A4$

La fonction de score est :

$$D(\mathbf{X}) = - 1.13365 + 0.46918 * A1 - 0.17477 * A2 + 0.58434 * A3 - 0.26661 * A4$$

Si $D(\mathbf{X}_i) > 0$ alors l'entreprise sera affectée au groupe défaillant. Les A_i sont considérés comme des variables numériques et sont calculés pour chaque individu. Par exemple, pour un individu j , nous calculerons $A1$ de la sorte : $A1_j = 0.8243 * 1_{t=3} - 0.4496 * 1_{rb=1} + 1.4531 * 1_{rb=2}$

5. Méthode CART

La méthode CART issue de la famille des arbres de décision, est une technique qui ne nécessite pas la vérification d'hypothèses sur les données ou sur les variables. L'arbre que SAS Enterprise Miner a généré, après la validation croisée, est le suivant :



Figure 22: Arbre de décision CART

Les probabilités, à posteriori, calculée par CART (sorties ci-dessous) sont :

Chapitre 3. Modélisation et implémentation

- Si $t=1$ alors $P(\text{Défaut} = 1 / r,t,rb) = 0.08$
- Si $t=2$ ou $t=3$ et $rb=2$ alors $P(\text{Défaut} = 1 / r,t,rb) = 0.45$
- Si $t=3$ et $rb=1$ alors $P(\text{Défaut} = 1 / r,t,rb) = 0.36$
- Si $t=2$ et $rb=1$ alors $P(\text{Défaut} = 1 / r,t,rb) = 0.28$

```

-----*
Node = 3
-----*
if t IS ONE OF: 1
then
Tree Node Identifier = 3
Number of Observations = 360
Valeur prédite : Default=1 = 0.08
Valeur prédite : Default=0 = 0.93
-----*
Node = 5
-----*
if t IS ONE OF: 2, 3 or MISSING
AND rb IS ONE OF: 2
then
Tree Node Identifier = 5
Number of Observations = 227
Valeur prédite : Default=1 = 0.45
Valeur prédite : Default=0 = 0.55
-----*

-----*
Node = 8
-----*
if t IS ONE OF: 3
AND rb IS ONE OF: 1 or MISSING
then
Tree Node Identifier = 8
Number of Observations = 306
Valeur prédite : Default=1 = 0.36
Valeur prédite : Default=0 = 0.64
-----*
Node = 9
-----*
if t IS ONE OF: 2 or MISSING
AND rb IS ONE OF: 1 or MISSING
then
Tree Node Identifier = 9
Number of Observations = 329
Valeur prédite : Default=1 = 0.28
Valeur prédite : Default=0 = 0.72
-----*

```

Règles d'attribution aux groupes (1)

Règles d'attribution aux groupes (2)

Variable à expliquer	Libellé cible	Rôle des données	Evènement	Intervalle de la probabilité a posteriori	Classe	Probabilité a posteriori de la borne inférieure	Probabilité a posteriori de la borne supérieure
Defaut	Defaut	TRAIN	1	0.45-0.50	11	0.45	0.5
Defaut	Defaut	TRAIN	1	0.35-0.40	13	0.35	0.4
Defaut	Defaut	TRAIN	1	0.25-0.30	15	0.25	0.3
Defaut	Defaut	TRAIN	1	0.05-0.10	19	0.05	0.1
Defaut	Defaut	VALIDATE	1	0.45-0.50	11	0.45	0.5
Defaut	Defaut	VALIDATE	1	0.35-0.40	13	0.35	0.4
Defaut	Defaut	VALIDATE	1	0.25-0.30	15	0.25	0.3
Defaut	Defaut	VALIDATE	1	0.05-0.10	19	0.05	0.1

Tableau 27: Tableau de distribution des scores

La variable t est celle qui impacte le plus la variable *Défaut*, comme le montre le tableau d'importance des variables (l'importance de t est égale à 1). Ceci est également observable au niveau de l'arbre puisque t est la première variable qui

a donné lieu à la division. Par contre, la variable r n'est pas considérée par le modèle, son importance est égale à zéro.

Nom de la variable	Libellé	Nombre de règles de découpe	Importance	Nombre de règles dans les arbres CV	Importance relative	Importance de la validation	Ratio de l'importance de la validation par rapport à l'apprentissage
t		2	1.0000	16	1.0000	1.0000	1.0000
rb		1	0.3918	11	0.3736	0.3875	0.9891
r		0	0.0000	0	0.0000	0.0000	

Tableau 28; Tableau d'importance des variables

Un modèle est optimal lorsque les valeurs du Misclassification Rate et de l'AverageSquaredError sont petites. Puisque ces indicateurs renseignent sur le pouvoir de prédiction du modèle, nous nous y sommes intéressés (Tableau 29).

Nous avons constaté que le taux de mauvais classement est légèrement supérieur à 27%, tandis que l'erreur quadratique moyenne est d'environ 0.17, une valeur qui est faible mais satisfaisante.

A expliquer	Libellé cible	Tests d'ajustement	Libellé de la statistique	Apprentissage	Validation
Defaut	Defaut	_NOBS_	Sum of Frequencies	1222	217
Defaut	Defaut	_MISC_	Misclassification Rate	0.270867	0.276498
Defaut	Defaut	_MAX_	Maximum Absolute Error	0.925	0.925
Defaut	Defaut	_SSE_	Sum of Squared Errors	435.3722	80.90274
Defaut	Defaut	_ASE_	Average Squared Error	0.178139	0.186412
Defaut	Defaut	_RASE_	Root Average Squared ...	0.422065	0.431754
Defaut	Defaut	_DIV_	Divisor for ASE	2444	434
Defaut	Defaut	_DFT_	Total Degrees of Freed...	1222	

Tableau 29: Tests d'ajustement sur la méthode CART

6. Comparaison des modèles

Nous avons obtenus les matrices de confusion et les courbes ROC en appliquant les modèles que nous avons relevés sur la base Test.

1.1 Matrice de confusion

L'appréciation des performances des modèles est possible grâce à la matrice de confusion qui récence les VP, FP, VN et les FN.

Modèle de régression logistique

Table of Defaut by _DEFAULT		
Defaut(Default)	_DEFAULT	
	0	Total
0	191 77.64 100.00 77.64	191 77.64
1	55 22.36 100.00 22.36	55 22.36
Total	246 100.00	246 100.00

Tableau 30: Matrice de confusion de la régression logistique

Modèle issu de l'analyse factorielle discriminante

Number of Observations and Percent Classified into Defaut			
From Defaut	0	1	Total
0	117 61.26	74 38.74	191 100.00
1	17 30.91	38 69.09	55 100.00
Total	134 54.47	112 45.53	246 100.00
Priors	0.5	0.5	

Tableau 31: Matrice de confusion de l'analyse discriminante

Modèle DISQUAL

Number of Observations and Percent Classified into Default			
From Default	0	1	Total
0	117 61.26	74 38.74	191 100.00
1	21 38.18	34 61.82	55 100.00
Total	138 56.10	108 43.90	246 100.00
Priors	0.5	0.5	

Tableau 32: Matrice de confusion de la méthode DISQUAL

Arbre CART

Table of Default by _DEFAULT		
Default(Default)	_DEFAULT	
	0	Total
0	191 77.64 100.00 77.64	191 77.64
1	55 22.36 100.00 22.36	55 22.36
Total	246 100.00	246 100.00

Tableau 33: Matrice de confusion de la méthode CART

Grace aux matrices de confusion des méthodes, nous arrivons à établir les taux de bons classements et de mauvais classements.

Méthodes	Taux de bons classements	Taux de mauvais classements
----------	--------------------------	-----------------------------

Régression logistique	0.7764	0.2236
Analyse discriminante	0.6300	0.3700
DISQUAL	0.6138	0.3862
CART	0.7764	0.2236

Tableau 34: Taux de bons et de mauvais classements des méthodes

1.2 Courbe ROC

La courbe ROC est un autre outil graphique de mesure de la qualité des modèles. La méthode la plus performante est celle qui détient l'AUC le plus élevé.

Régression logistique

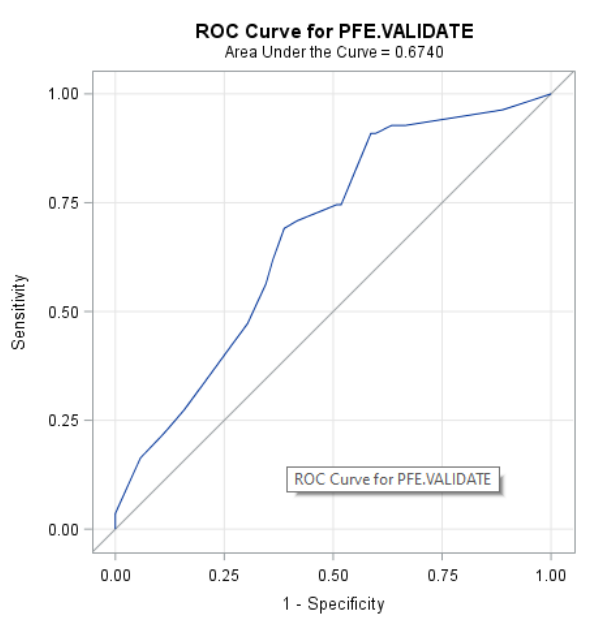


Figure 23: Courbe ROC de la régression logistique

Analyse discriminante

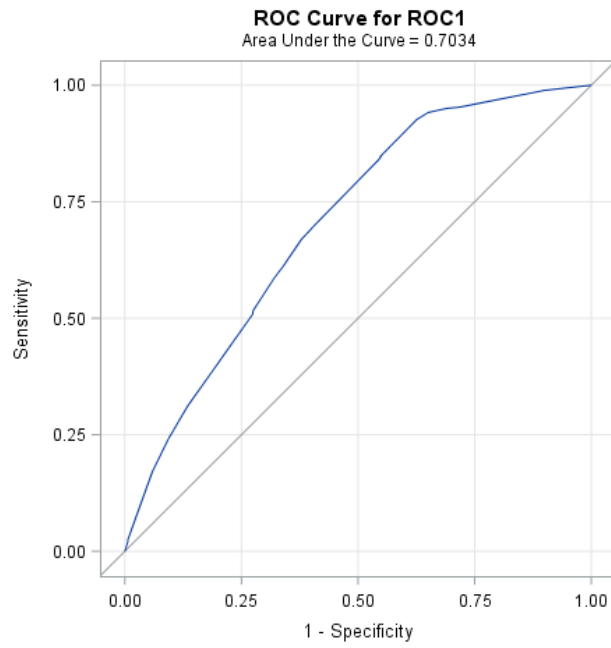


Figure 24: Courbe ROC de l'analyse factorielle discriminante

DISQUAL

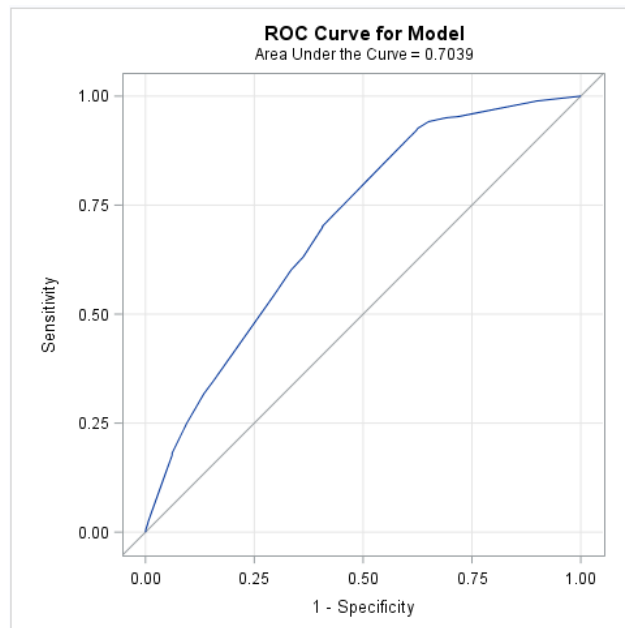


Figure 25: Courbe ROC de la méthode DISQUAL

Arbre CART

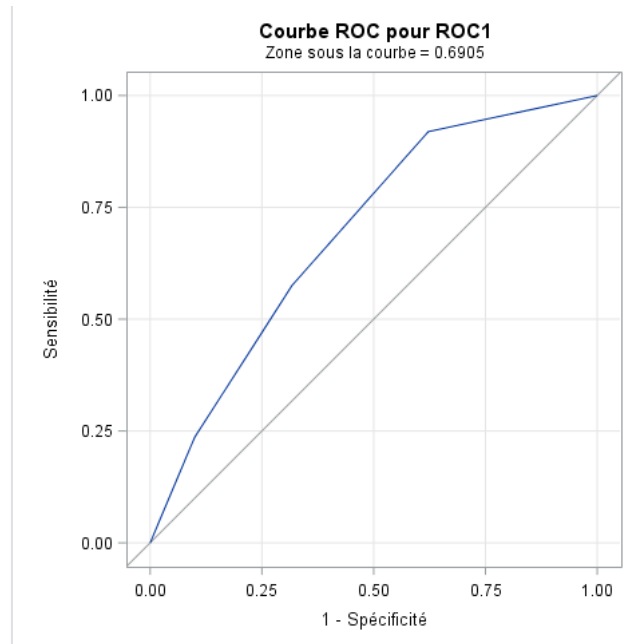


Figure 26: Courbe ROC de l'arbre CART

Le modèle de l'analyse discriminante présente une AUC plus élevée sachant que la surface est presque similaire pour toutes les méthodes.

Commentaire :

Les méthodes d'analyse discriminante, de régression logistique et CART, sont celles qui génèrent les meilleurs taux de classement. Par ailleurs, les AUC de la courbe ROC des trois modèles sont respectivement 0.7034, 0.6740 et 0.6905. Nous avons constaté que le premier modèle détecte mieux les entreprises défaillantes par rapport à celles qui ne le sont pas. A l'opposé, le deuxième et le troisième classent l'intégralité des observations dans la modalité 0. La question que nous nous posons, concerne le choix de la banque face à ces deux alternatives : est-ce qu'elle cherche à minimiser le nombre de mauvaises prédictions pour la modalité défaillant (*Defaut=1*) ou maximiser le nombre de bonnes prédictions pour la modalité non défaillant (*Defaut=0*) ?

Le modèle issu de la méthode DISQUAL détecte aussi bien les entreprises défaillantes que le modèle d'analyse factorielle discriminante.

Pour conclure, après comparaison, le modèle que nous devrions retenir serait celui de l'AFD, si nous prenons en considération uniquement le critère d'AUC.

Conclusion :

L'application des modèles sur la base de données, dont nous disposons, a suscité notre attention et son traitement a nécessité l'exploitation de deux hypothèses : D'une part, nous avons envisagé une transformation logarithmique et une segmentation pour pouvoir détecter les valeurs aberrantes et les imputer. D'autre part, nous avons calculé les ratios financiers mais les tests d'hypothèses, sur les ratios retenus pour la modélisation, ne se sont pas tous montrés positifs.

A l'issue de notre étude, il paraît que les méthodes d'analyse discriminante sont meilleures en termes de prédiction en comparaison avec la régression logistique. Nous avons trouvé les mêmes taux de bons classements pour les modèles CART et de régression logistique, même si généralement le premier se démarque du deuxième.

خاتمة

لقد جذب تطبيق النماذج الموجودة على قاعدة البيانات انتباهنا ، وتطلبت منا معالجتها من خلال استغلال فرضيتين:

- الفرضية الأولى : نظرنا في إمكانية القيام بتحويل لوغاريتمي وتجزئة، لنكون قادرين على اكتشاف القيم المتطرفة وتقييدها
- الفرضية الثانية: قمنا باحتساب النسب المالية ولكن اكتشفنا أن اختبارات الفرضيات على النسب المستخدمة للنمذجة لم تكن كلها إيجابية.

في نهاية بحثنا ، لاحظنا أن طرق التحليل التمييزية أفضل وسيلة من حيث التنبؤ مقارنة بالانحدار

Conclusion

اللوجستي، حيث وجدنا نفس معدلات الترتيب الجيد لنماذج CART مقارنة مع نماذج الانحدار اللوجستي، على الرغم من أن الأول يختلف عمومًا عن النموذج الثاني.

Annexe 1 : Statistiques descriptives sur les variables :

1. Matrice de corrélation de Pearson

Pearson Correlation Coefficients Prob > r under H0: Rho=0 Number of Observations			
	Default	NF_CHIFFREAFFAIRE_N	NF_CHIFFREAFFAIRE_N_1
Default Default	1.00000 1982	0.05812 0.0097 1982	0.07226 0.0013 1982
NF_CHIFFREAFFAIRE_N NF_CHIFFREAFFAIRE_N	0.05812 0.0097 1982	1.00000 1982	0.73664 <.0001 1982
NF_CHIFFREAFFAIRE_N_1 NF_CHIFFREAFFAIRE_N_1	0.07226 0.0013 1982	0.73664 <.0001 1982	1.00000 1982
NF_DETTECOURTTERME NF_DETTECOURTTERME	0.08456 0.0002 1982	0.83848 <.0001 1982	0.61584 <.0001 1982
NF_DETTELONGTERME NF_DETTELONGTERME	0.04216 0.0606 1982	0.16726 <.0001 1982	0.12945 <.0001 1982
NF_ACTIFCIRCULANT NF_ACTIFCIRCULANT	0.07925 0.0004 1982	0.82973 <.0001 1982	0.60050 <.0001 1982
NF_EBE NF_EBE	0.04882 0.0298 1982	0.85519 <.0001 1982	0.61617 <.0001 1982
NF_TRESORERIENETTE NF_TRESORERIENETTE	0.14861 <.0001 1982	0.34603 <.0001 1982	0.28723 <.0001 1982
NF_TOTALBILAN NF_TOTALBILAN	0.05574 0.0131 1982	0.38162 <.0001 1982	0.28704 <.0001 1982
NF_RESULTATNET_N NF_RESULTATNET_N	-0.00059 0.9790 1982	0.72521 <.0001 1982	0.49346 <.0001 1982
NF_RESULTAT_NET_N_1 NF_RESULTAT_NET_N_1	0.02011 0.3710 1982	0.43390 <.0001 1982	0.80774 <.0001 1982
NF_FPNETS NF_FPNETS	0.04573 0.0418 1982	0.30873 <.0001 1982	0.23474 <.0001 1982
NF_KPROPRES NF_KPROPRES	0.04523 0.0441 1982	0.30584 <.0001 1982	0.23262 <.0001 1982
NF_FINANPERMANENT NF_FINANPERMANENT	0.04400 0.0502 1982	0.23363 <.0001 1982	0.17956 <.0001 1982
NF_FDR	-0.00229 0.9190	0.19664 <.0001	0.10764 <.0001

Annexe 1 : Statistiques descriptives sur les variables

NF_FDR	1982	1982	1982
NF_BFR NF_BFR	0.11177 <.0001 1982	0.22576 <.0001 1982	0.16729 <.0001 1982
RESULTAT_COURANT_N RESULTAT_COURANT_N	0.00559 0.8035 1982	0.80267 <.0001 1982	0.55976 <.0001 1982
RESULTAT_COURANT_N_1 RESULTAT_COURANT_N_1	0.02446 0.2764 1982	0.44641 <.0001 1982	0.83988 <.0001 1982
RESULTAT_EXP_N RESULTAT_EXP_N	0.03885 0.0838 1982	0.89534 <.0001 1982	0.62981 <.0001 1982
RESULTAT_EXP_N_1 RESULTAT_EXP_N_1	0.05132 0.0223 1982	0.47947 <.0001 1982	0.87414 <.0001 1982
CAPACITE_AUTO_FIN CAPACITE_AUTO_FIN	0.02044 0.3632 1982	0.88877 <.0001 1982	0.64432 <.0001 1982
TRESORERIE_ACTIF TRESORERIE_ACTIF	0.05905 0.0086 1982	0.74062 <.0001 1982	0.53534 <.0001 1982
PASSIF_CIRCULANT PASSIF_CIRCULANT	0.05057 0.0244 1982	0.86577 <.0001 1982	0.62520 <.0001 1982
DCT DCT	0.20551 <.0001 1972	0.29302 <.0001 1972	0.22476 <.0001 1972

Pearson Correlation Coefficients Prob > r under H0: Rho=0 Number of Observations			
	NF_DETTECOURTTERME	NF_DETTELONGTERME	NF_ACTIFCIRCULANT
Default Default	0.08456 0.0002 1982	0.04216 0.0606 1982	0.07925 0.0004 1982
NF_CHIFFREAFFAIRE_N NF_CHIFFREAFFAIRE_N	0.83848 <.0001 1982	0.16726 <.0001 1982	0.82973 <.0001 1982
NF_CHIFFREAFFAIRE_N_1 NF_CHIFFREAFFAIRE_N_1	0.61584 <.0001 1982	0.12945 <.0001 1982	0.60050 <.0001 1982
NF_DETTECOURTTERME NF_DETTECOURTTERME	1.00000 1982	0.59048 <.0001 1982	0.96210 <.0001 1982
NF_DETTELONGTERME NF_DETTELONGTERME	0.59048 <.0001 1982	1.00000 1982	0.46443 <.0001 1982
NF_ACTIFCIRCULANT NF_ACTIFCIRCULANT	0.96210 <.0001 1982	0.46443 <.0001 1982	1.00000 1982
NF_EBE NF_EBE	0.88687 <.0001 1982	0.50265 <.0001 1982	0.83751 <.0001 1982

Annexe 1 : Statistiques descriptives sur les variables

NF_TRESORERIENETTE NF_TRESORERIENETTE	0.61547 <.0001 1982	0.59848 <.0001 1982	0.52203 <.0001 1982
NF_TOTALBILAN NF_TOTALBILAN	0.75609 <.0001 1982	0.96740 <.0001 1982	0.64159 <.0001 1982
NF_RESULTATNET_N NF_RESULTATNET_N	0.43243 <.0001 1982	-0.26848 <.0001 1982	0.51614 <.0001 1982
NF_RESULTAT_NET_N_1 NF_RESULTAT_NET_N_1	0.21836 <.0001 1982	-0.25936 <.0001 1982	0.26470 <.0001 1982
NF_FPNETS NF_FPNETS	0.67162 <.0001 1982	0.94952 <.0001 1982	0.55361 <.0001 1982
NF_KPROPRES NF_KPROPRES	0.66976 <.0001 1982	0.95159 <.0001 1982	0.55128 <.0001 1982
NF_FINANPERMANENT NF_FINANPERMANENT	0.63416 <.0001 1982	0.98993 <.0001 1982	0.50872 <.0001 1982
NF_FDR NF_FDR	0.09068 <.0001 1982	-0.36664 <.0001 1982	0.34985 <.0001 1982
NF_BFR NF_BFR	0.36527 <.0001 1982	0.09699 <.0001 1982	0.52595 <.0001 1982
RESULTAT_COURANT_N RESULTAT_COURANT_N	0.49738 <.0001 1982	-0.23883 <.0001 1982	0.56740 <.0001 1982
RESULTAT_COURANT_N_1 RESULTAT_COURANT_N_1	0.23595 <.0001 1982	-0.20321 <.0001 1982	0.26206 <.0001 1982
RESULTAT_EXP_N RESULTAT_EXP_N	0.76508 <.0001 1982	0.14857 <.0001 1982	0.79481 <.0001 1982
RESULTAT_EXP_N_1 RESULTAT_EXP_N_1	0.42113 <.0001 1982	0.12089 <.0001 1982	0.40844 <.0001 1982
CAPACITE_AUTO_FIN CAPACITE_AUTO_FIN	0.80987 <.0001 1982	0.32424 <.0001 1982	0.75999 <.0001 1982
TRESORERIE_ACTIF TRESORERIE_ACTIF	0.78126 <.0001 1982	0.37724 <.0001 1982	0.71496 <.0001 1982
PASSIF_CIRCULANT PASSIF_CIRCULANT	0.96703 <.0001 1982	0.49506 <.0001 1982	0.95303 <.0001 1982

Pearson Correlation Coefficients				
Prob > r under H0: Rho=0				
Number of Observations				
	NF_EBE	NF_TRESORERIENETTE	NF_TOTALBILAN	NF_RESULTATNET_N
Default	0.04882	0.14861	0.05574	-0.00059
Default	0.0298 1982	<.0001 1982	0.0131 1982	0.9790 1982
NF_CHIFFREAFFAIRE_N	0.85519 <.0001	0.34603 <.0001	0.38162 <.0001	0.72521 <.0001

Annexe 1 : Statistiques descriptives sur les variables

NF_CHIFFREAFFAIRE_N	1982	1982	1982	1982
NF_CHIFFREAFFAIRE_N_1	0.61617 <.0001	0.28723 <.0001	0.28704 <.0001	0.49346 <.0001
NF_CHIFFREAFFAIRE_N_1	1982	1982	1982	1982
NF_DETTECOURTTERME	0.88687 <.0001	0.61547 <.0001	0.75609 <.0001	0.43243 <.0001
NF_DETTECOURTTERME	1982	1982	1982	1982
NF_DETTELONGTERME	0.50265 <.0001	0.59848 <.0001	0.96740 <.0001	-0.26848 <.0001
NF_DETTELONGTERME	1982	1982	1982	1982
NF_ACTIFCIRCULANT	0.83751 <.0001	0.52203 <.0001	0.64159 <.0001	0.51614 <.0001
NF_ACTIFCIRCULANT	1982	1982	1982	1982
NF_EBE	1.00000	0.41117	0.67036	0.62511
NF_EBE	1982	<.0001 1982	<.0001 1982	<.0001 1982
NF_TRESORERIE NETTE	0.41117 <.0001	1.00000	0.65119 <.0001	-0.15380 <.0001
NF_TRESORERIE NETTE	1982	1982	1982	1982
NF_TOTALBILAN	0.67036 <.0001	0.65119 <.0001	1.00000	-0.07254 0.0012
NF_TOTALBILAN	1982	1982	1982	1982
NF_RESULTATNET_N	0.62511 <.0001	-0.15380 <.0001	-0.07254 0.0012	1.00000
NF_RESULTATNET_N	1982	1982	1982	1982
NF_RESULTAT_NET_N_1	0.32899 <.0001	-0.10574 <.0001	-0.12918 <.0001	0.56317 <.0001
NF_RESULTAT_NET_N_1	1982	1982	1982	1982
NF_FPNETS	0.62205 <.0001	0.60571 <.0001	0.98009 <.0001	-0.09682 <.0001
NF_FPNETS	1982	1982	1982	1982
NF_KPROPRES	0.62033 <.0001	0.60621 <.0001	0.98087 <.0001	-0.10081 <.0001
NF_KPROPRES	1982	1982	1982	1982
NF_FINANPERMANENT	0.56219 <.0001	0.60985 <.0001	0.98551 <.0001	-0.19779 <.0001
NF_FINANPERMANENT	1982	1982	1982	1982
NF_FDR	0.07376	-0.19369	-0.27106	0.49383
NF_FDR	0.0010 1982	<.0001 1982	<.0001 1982	<.0001 1982
NF_BFR	0.15135 <.0001	0.56394 <.0001	0.16990 <.0001	0.09746 <.0001
NF_BFR	1982	1982	1982	1982
RESULTAT_COURANT_N	0.67863 <.0001	-0.10239 <.0001	-0.02955 0.1885	0.98129 <.0001
RESULTAT_COURANT_N	1982	1982	1982	1982
RESULTAT_COURANT_N_1	0.35616 <.0001	-0.05147 0.0219	-0.07987 0.0004	0.51925 <.0001
RESULTAT_COURANT_N_1	1982	1982	1982	1982
RESULTAT_EXP_N	0.89352 <.0001	0.18504 <.0001	0.35015 <.0001	0.87965 <.0001
RESULTAT_EXP_N	1982	1982	1982	1982
RESULTAT_EXP_N_1	0.50531 <.0001	0.19349 <.0001	0.22995 <.0001	0.40069 <.0001
RESULTAT_EXP_N_1	1982	1982	1982	1982
CAPACITE_AUTO_FIN	0.94393 <.0001	0.27751 <.0001	0.51748 <.0001	0.73788 <.0001
CAPACITE_AUTO_FIN	1982	1982	1982	1982
TRESORERIE_ACTIF	0.82185 <.0001	0.51074 <.0001	0.54250 <.0001	0.54105 <.0001
TRESORERIE_ACTIF	1982	1982	1982	1982

Annexe 1 : Statistiques descriptives sur les variables

	1982	1982	1982	1982
PASSIF_CIRCULANT	0.90114	0.39446	0.67111	0.55386
PASSIF_CIRCULANT	<.0001	<.0001	<.0001	<.0001
	1982	1982	1982	1982
DCT	0.24704	0.34517	0.19275	0.10593
DCT	<.0001	<.0001	<.0001	<.0001
	1972	1972	1972	1972

Pearson Correlation Coefficients				
Prob > r under H0: Rho=0				
Number of Observations				
	NF_RESULTAT_NET_N_1	NF_FPNETS	NF_KPROPRES	
Default	0.02011	0.04573	0.04523	
Default	0.3710	0.0418	0.0441	
	1982	1982	1982	
NF_CHIFFREAFFAIRE_N	0.43390	0.30873	0.30584	
NF_CHIFFREAFFAIRE_N	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_CHIFFREAFFAIRE_N_1	0.80774	0.23474	0.23262	
NF_CHIFFREAFFAIRE_N_1	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_DETTECOURTTERME	0.21836	0.67162	0.66976	
NF_DETTECOURTTERME	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_DETTELONGTERME	-0.25936	0.94952	0.95159	
NF_DETTELONGTERME	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_ACTIFCIRCULANT	0.26470	0.55361	0.55128	
NF_ACTIFCIRCULANT	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_EBE	0.32899	0.62205	0.62033	
NF_EBE	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_TRESORERIENETTE	-0.10574	0.60571	0.60621	
NF_TRESORERIENETTE	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_TOTALBILAN	-0.12918	0.98009	0.98087	
NF_TOTALBILAN	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_RESULTATNET_N	0.56317	-0.09682	-0.10081	
NF_RESULTATNET_N	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_RESULTAT_NET_N_1	1.00000	-0.13932	-0.14234	
NF_RESULTAT_NET_N_1	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_FPNETS	-0.13932	1.00000	0.99949	
NF_FPNETS	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_KPROPRES	-0.14234	0.99949	1.00000	
NF_KPROPRES	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_FINANPERMANENT	-0.20922	0.98388	0.98529	
NF_FINANPERMANENT	<.0001	<.0001	<.0001	
	1982	1982	1982	
NF_FDR	0.29488	-0.29169	-0.29452	
NF_FDR	<.0001	<.0001	<.0001	
	1982	1982	1982	

Annexe 1 : Statistiques descriptives sur les variables

NF_BFR NF_BFR	0.03675 0.1019 1982	0.12348 <.0001 1982	0.12257 <.0001 1982
RESULTAT_COURANT_N RESULTAT_COURANT_N	0.58262 <.0001 1982	-0.05983 0.0077 1982	-0.06386 0.0045 1982
RESULTAT_COURANT_N_1 RESULTAT_COURANT_N_1	0.98462 <.0001 1982	-0.09184 <.0001 1982	-0.09438 <.0001 1982
RESULTAT_EXP_N RESULTAT_EXP_N	0.48725 <.0001 1982	0.29933 <.0001 1982	0.29630 <.0001 1982
RESULTAT_EXP_N_1 RESULTAT_EXP_N_1	0.89781 <.0001 1982	0.21000 <.0001 1982	0.20825 <.0001 1982
CAPACITE_AUTO_FIN CAPACITE_AUTO_FIN	0.43443 <.0001 1982	0.48452 <.0001 1982	0.48172 <.0001 1982
TRESORERIE_ACTIF TRESORERIE_ACTIF	0.34843 <.0001 1982	0.50612 <.0001 1982	0.50306 <.0001 1982
PASSIF_CIRCULANT PASSIF_CIRCULANT	0.28875 <.0001 1982	0.58733 <.0001 1982	0.58499 <.0001 1982
DCT DCT	0.08781 <.0001 1972	0.16147 <.0001 1972	0.16109 <.0001 1972

Tableau 35: Matrice de corrélation de Pearson

Annexe 1 : Statistiques descriptives sur les variables

Spearman Correlation Coefficients, N = 1982 Prob > r under H0: Rho=0		
	NF_CHIFFREAFFAIRE_N	NF_CHIFFREAFFAIRE_N_1
NF_CHIFFREAFFAIRE_N NF_CHIFFREAFFAIRE_N	1.00000 <.0001	0.91452 <.0001
NF_CHIFFREAFFAIRE_N_1 NF_CHIFFREAFFAIRE_N_1	0.91452 <.0001	1.00000
NF_DETTECOURTTERME NF_DETTECOURTTERME	0.76952 <.0001	0.73865 <.0001
NF_DETTELONGTERME NF_DETTELONGTERME	0.28683 <.0001	0.26608 <.0001
NF_ACTIFCIRCULANT NF_ACTIFCIRCULANT	0.83141 <.0001	0.81420 <.0001
NF_EBE NF_EBE	0.73433 <.0001	0.68058 <.0001
NF_TRESORERIENETTE NF_TRESORERIENETTE	0.60941 <.0001	0.61771 <.0001
NF_TOTALBILAN NF_TOTALBILAN	0.78464 <.0001	0.75750 <.0001
NF_RESULTATNET_N NF_RESULTATNET_N	0.54253 <.0001	0.50630 <.0001
NF_RESULTAT_NET_N_1 NF_RESULTAT_NET_N_1	0.48918 <.0001	0.54381 <.0001
NF_FPNETS NF_FPNETS	0.70862 <.0001	0.71263 <.0001
NF_KPROPRES NF_KPROPRES	0.71257 <.0001	0.71461 <.0001
NF_FINANPERMANENT NF_FINANPERMANENT	0.69533 <.0001	0.68436 <.0001
NF_FDR NF_FDR	0.37833 <.0001	0.39144 <.0001
NF_BFR NF_BFR	0.54635 <.0001	0.56618 <.0001
RESULTAT_COURANT_N RESULTAT_COURANT_N	0.56670 <.0001	0.51401 <.0001
RESULTAT_EXP_N RESULTAT_EXP_N	0.65807 <.0001	0.60880 <.0001
RESULTAT_COURANT_N_1 RESULTAT_COURANT_N_1	0.50875 <.0001	0.56843 <.0001
TRESORERIE_ACTIF TRESORERIE_ACTIF	0.55745 <.0001	0.53960 <.0001
RESULTAT_EXP_N_1 RESULTAT_EXP_N_1	0.62026 <.0001	0.67617 <.0001

Annexe 1 : Statistiques descriptives sur les variables

CAPACITE_AUTO_FIN	0.62998	0.57761
CAPACITE_AUTO_FIN	<.0001	<.0001
PASSIF_CIRCULANT	0.71907	0.68699
PASSIF_CIRCULANT	<.0001	<.0001

Spearman Correlation Coefficients, N = 1982 Prob > r under H0: Rho=0			
	NF_DETTECOURT TERME	NF_DETTELONGT ERME	NF_ACTIFCIRCULANT
NF_CHIFFREAFFAIRE_N NF_CHIFFREAFFAIRE_N	0.76952 <.0001	0.28683 <.0001	0.83141 <.0001
NF_CHIFFREAFFAIRE_N_1 NF_CHIFFREAFFAIRE_N_1	0.73865 <.0001	0.26608 <.0001	0.81420 <.0001
NF_DETTECOURTTERME NF_DETTECOURTTERME	1.00000	0.39800 <.0001	0.90866 <.0001
NF_DETTELONGTERME NF_DETTELONGTERME	0.39800 <.0001	1.00000	0.34433 <.0001
NF_ACTIFCIRCULANT NF_ACTIFCIRCULANT	0.90866 <.0001	0.34433 <.0001	1.00000
NF_EBE NF_EBE	0.53472 <.0001	0.22865 <.0001	0.61604 <.0001
NF_TRESORERIENETTE NF_TRESORERIENETTE	0.67157 <.0001	0.33265 <.0001	0.65813 <.0001
NF_TOTALBILAN NF_TOTALBILAN	0.95051 <.0001	0.47414 <.0001	0.91779 <.0001
NF_RESULTATNET_N NF_RESULTATNET_N	0.34679 <.0001	0.05585 0.0129	0.46645 <.0001
NF_RESULTAT_NET_N_1 NF_RESULTAT_NET_N_1	0.30206 <.0001	0.05923 0.0084	0.42371 <.0001
NF_FPNETS NF_FPNETS	0.59708 <.0001	0.22183 <.0001	0.73157 <.0001
NF_KPROPRES NF_KPROPRES	0.60707 <.0001	0.23085 <.0001	0.73694 <.0001
NF_FINANPERMANENT NF_FINANPERMANENT	0.69984 <.0001	0.50236 <.0001	0.76749 <.0001
NF_FDR NF_FDR	0.18608 <.0001	-0.00231 0.9181	0.44274 <.0001
NF_BFR NF_BFR	0.41926 <.0001	0.12876 <.0001	0.65406 <.0001
RESULTAT_COURANT_N RESULTAT_COURANT_N	0.33997 <.0001	0.00711 0.7518	0.46391 <.0001
RESULTAT_EXP_N	0.44317	0.09829	0.55918

Annexe 1 : Statistiques descriptives sur les variables

RESULTAT_EXP_N	<.0001	<.0001	<.0001
RESULTAT_COURANT_N_1 RESULTAT_COURANT_N_1	0.31546 <.0001	0.00600 0.7897	0.44453 <.0001
TRESORERIE_ACTIF TRESORERIE_ACTIF	0.49910 <.0001	0.24612 <.0001	0.49607 <.0001
RESULTAT_EXP_N_1 RESULTAT_EXP_N_1	0.43661 <.0001	0.11330 <.0001	0.55634 <.0001
CAPACITE_AUTO_FIN CAPACITE_AUTO_FIN	0.43454 <.0001	0.18537 <.0001	0.52327 <.0001
PASSIF_CIRCULANT PASSIF_CIRCULANT	0.96499 <.0001	0.38185 <.0001	0.86064 <.0001

Spearman Correlation Coefficients, N = 1982 Prob > r under H0: Rho=0				
	NF_EBE	NF_TRESORERIENETTE	NF_TOTALBILAN	NF_RESULTATNET_N
NF_CHIFFREAFFAIRE_N NF_CHIFFREAFFAIRE_N	0.73433 <.0001	0.60941 <.0001	0.78464 <.0001	0.54253 <.0001
NF_CHIFFREAFFAIRE_N_1 NF_CHIFFREAFFAIRE_N_1	0.68058 <.0001	0.61771 <.0001	0.75750 <.0001	0.50630 <.0001
NF_DETTECOURTTERME NF_DETTECOURTTERME	0.53472 <.0001	0.67157 <.0001	0.95051 <.0001	0.34679 <.0001
NF_DETTELONGTERME NF_DETTELONGTERME	0.22865 <.0001	0.33265 <.0001	0.47414 <.0001	0.05585 0.0129
NF_ACTIFCIRCULANT NF_ACTIFCIRCULANT	0.61604 <.0001	0.65813 <.0001	0.91779 <.0001	0.46645 <.0001
NF_EBE NF_EBE	1.00000	0.43562 <.0001	0.58045 <.0001	0.71877 <.0001
NF_TRESORERIENETTE NF_TRESORERIENETTE	0.43562 <.0001	1.00000	0.65356 <.0001	0.25388 <.0001
NF_TOTALBILAN NF_TOTALBILAN	0.58045 <.0001	0.65356 <.0001	1.00000	0.38721 <.0001
NF_RESULTATNET_N NF_RESULTATNET_N	0.71877 <.0001	0.25388 <.0001	0.38721 <.0001	1.00000
NF_RESULTAT_NET_N_1 NF_RESULTAT_NET_N_1	0.52989 <.0001	0.27588 <.0001	0.34861 <.0001	0.61341 <.0001
NF_FPNETS NF_FPNETS	0.66380 <.0001	0.47970 <.0001	0.71390 <.0001	0.58297 <.0001
NF_KPROPRES NF_KPROPRES	0.66570 <.0001	0.48791 <.0001	0.72482 <.0001	0.57721 <.0001
NF_FINANPERMANENT NF_FINANPERMANENT	0.60168 <.0001	0.54736 <.0001	0.84909 <.0001	0.44463 <.0001

Annexe 1 : Statistiques descriptives sur les variables

NF_FDR NF_FDR	0.37336 <.0001	0.15074 <.0001	0.27546 <.0001	0.46086 <.0001
NF_BFR NF_BFR	0.47995 <.0001	0.50667 <.0001	0.49018 <.0001	0.43498 <.0001
RESULTAT_COURANT_N RESULTAT_COURANT_N	0.75923 <.0001	0.24334 <.0001	0.37125 <.0001	0.88463 <.0001
RESULTAT_EXP_N RESULTAT_EXP_N	0.84609 <.0001	0.37997 <.0001	0.47324 <.0001	0.83818 <.0001
RESULTAT_COURANT_N_1 RESULTAT_COURANT_N_1	0.54901 <.0001	0.27746 <.0001	0.35384 <.0001	0.61355 <.0001
TRESORERIE_ACTIF TRESORERIE_ACTIF	0.38742 <.0001	0.23578 <.0001	0.54195 <.0001	0.28736 <.0001
RESULTAT_EXP_N_1 RESULTAT_EXP_N_1	0.64096 <.0001	0.41878 <.0001	0.47332 <.0001	0.60707 <.0001
CAPACITE_AUTO_FIN CAPACITE_AUTO_FIN	0.86337 <.0001	0.31049 <.0001	0.49554 <.0001	0.80215 <.0001
PASSIF_CIRCULANT PASSIF_CIRCULANT	0.48951 <.0001	0.54782 <.0001	0.90627 <.0001	0.33251 <.0001

Spearman Correlation Coefficients, N = 1982 Prob > r under H0: Rho=0				
	NF_RESULTAT_NET_N_1	NF_FPNET S	NF_KPROPRES	
NF_CHIFFREAFFAIRE_N NF_CHIFFREAFFAIRE _N	0.48918 <.0001	0.70862 <.0001	0.71257 <.0001	
NF_CHIFFREAFFAIRE_N_1 NF_CHIFFREAFFAIRE _N_1	0.54381 <.0001	0.71263 <.0001	0.71461 <.0001	
NF_DETTECOURTTERME NF_DETTECOURTTER ME	0.30206 <.0001	0.59708 <.0001	0.60707 <.0001	
NF_DETTELONGTERME NF_DETTELONGTER ME	0.05923 0.0084	0.22183 <.0001	0.23085 <.0001	
NF_ACTIFCIRCULANT NF_ACTIFCIRCULAN T	0.42371 <.0001	0.73157 <.0001	0.73694 <.0001	
NF_EBE NF_EBE	0.52989 <.0001	0.66380 <.0001	0.66570 <.0001	

Annexe 1 : Statistiques descriptives sur les variables

NF_TRESORERIE NETTE NF_TRESORERIE NETTE	0.27588 <.0001	0.47970 <.0001	0.48791 <.0001
NF_TOTALBILAN NF_TOTALBILAN	0.34861 <.0001	0.71390 <.0001	0.72482 <.0001
NF_RESULTAT NET N NF_RESULTAT NET N	0.61341 <.0001	0.58297 <.0001	0.57721 <.0001
NF_RESULTAT NET N_1 NF_RESULTAT NET N_1	1.00000	0.54191 <.0001	0.54046 <.0001
NF_FPNETS NF_FPNETS	0.54191 <.0001	1.00000	0.99541 <.0001
NF_KPROPRES NF_KPROPRES	0.54046 <.0001	0.99541 <.0001	1.00000
NF_FINANPERMANENT NF_FINANPERMANENT	0.42081 <.0001	0.84937 <.0001	0.85822 <.0001
NF_FDR NF_FDR	0.42855 <.0001	0.55184 <.0001	0.54171 <.0001
NF_BFR NF_BFR	0.41920 <.0001	0.65570 <.0001	0.65253 <.0001
RESULTAT_COURANT N RESULTAT_COURANT N	0.60886 <.0001	0.57146 <.0001	0.56533 <.0001
RESULTAT_EXP N RESULTAT_EXP N	0.58983 <.0001	0.62724 <.0001	0.62204 <.0001
RESULTAT_COURANT N_1 RESULTAT_COURANT N_1	0.87501 <.0001	0.55077 <.0001	0.54595 <.0001
TRESORERIE ACTIF TRESORERIE ACTIF	0.27445 <.0001	0.48559 <.0001	0.48802 <.0001
RESULTAT_EXP N_1 RESULTAT_EXP N_1	0.80441 <.0001	0.62185 <.0001	0.62127 <.0001
CAPACITE_AUTO_FIN CAPACITE_AUTO_FIN	0.54362 <.0001	0.64316 <.0001	0.64207 <.0001
PASSIF_CIRCULANT	0.27942	0.54522	0.55479

Annexe 1 : Statistiques descriptives sur les variables

PASSIF_CIRCULANT		<.0001	<.0001	<.0001	
Spearman Correlation Coefficients, N = 1982 Prob > r under H0: Rho=0					
	NF_FINANPERMANENT	NF_FDR	NF_BFR	RESULTAT_COURANT_N	
NF_CHIFFREAFFAIRE_N NF_CHIFFREAFFAIRE_N	0.69533 <.0001	0.37833 <.0001	0.54635 <.0001	0.56670 <.0001	
NF_CHIFFREAFFAIRE_N_1 NF_CHIFFREAFFAIRE_N_1	0.68436 <.0001	0.39144 <.0001	0.56618 <.0001	0.51401 <.0001	
NF_DETTECOURTTERME NF_DETTECOURTTERME	0.69984 <.0001	0.18608 <.0001	0.41926 <.0001	0.33997 <.0001	
NF_DETTELONGTERME NF_DETTELONGTERME	0.50236 <.0001	-0.00231 0.9181	0.12876 <.0001	0.00711 0.7518	
NF_ACTIFCIRCULANT NF_ACTIFCIRCULANT	0.76749 <.0001	0.44274 <.0001	0.65406 <.0001	0.46391 <.0001	
NF_EBE NF_EBE	0.60168 <.0001	0.37336 <.0001	0.47995 <.0001	0.75923 <.0001	
NF_TRESORERIENETTE NF_TRESORERIENETTE	0.54736 <.0001	0.15074 <.0001	0.50667 <.0001	0.24334 <.0001	
NF_TOTALBILAN NF_TOTALBILAN	0.84909 <.0001	0.27546 <.0001	0.49018 <.0001	0.37125 <.0001	
NF_RESULTATNET_N NF_RESULTATNET_N	0.44463 <.0001	0.46086 <.0001	0.43498 <.0001	0.88463 <.0001	
NF_RESULTAT_NET_N_1 NF_RESULTAT_NET_N_1	0.42081 <.0001	0.42855 <.0001	0.41920 <.0001	0.60886 <.0001	
NF_FPNETS NF_FPNETS	0.84937 <.0001	0.55184 <.0001	0.65570 <.0001	0.57146 <.0001	
NF_KPROPRES NF_KPROPRES	0.85822 <.0001	0.54171 <.0001	0.65253 <.0001	0.56533 <.0001	
NF_FINANPERMANENT NF_FINANPERMANENT	1.00000	0.43350 <.0001	0.56683 <.0001	0.42439 <.0001	
NF_FDR NF_FDR	0.43350 <.0001	1.00000	0.74203 <.0001	0.47752 <.0001	
NF_BFR NF_BFR	0.56683 <.0001	0.74203 <.0001	1.00000	0.45002 <.0001	
RESULTAT_COURANT_N RESULTAT_COURANT_N	0.42439 <.0001	0.47752 <.0001	0.45002 <.0001	1.00000	
RESULTAT_EXP_N RESULTAT_EXP_N	0.51279 <.0001	0.46914 <.0001	0.52696 <.0001	0.91712 <.0001	
RESULTAT_COURANT_N_1	0.41132 <.0001	0.45341 <.0001	0.45026 <.0001	0.65687 <.0001	

Annexe 1 : Statistiques descriptives sur les variables

RESULTAT_COURANT_N_1				
TRESORERIE_ACTIF TRESORERIE_ACTIF	0.50712 <.0001	0.25585 <.0001	0.26111 <.0001	0.28831 <.0001
RESULTAT_EXP_N_1 RESULTAT_EXP_N_1	0.52017 <.0001	0.46321 <.0001	0.54176 <.0001	0.62939 <.0001
CAPACITE_AUTO_FIN CAPACITE_AUTO_FIN	0.56184 <.0001	0.36833 <.0001	0.39999 <.0001	0.80576 <.0001
PASSIF_CIRCULANT PASSIF_CIRCULANT	0.64552 <.0001	0.15499 <.0001	0.32182 <.0001	0.32413 <.0001

Spearman Correlation Coefficients, N = 1982 Prob > r under H0: Rho=0			
	RESULTAT_EXP_N	RESULTAT_COURANT_N_1	TRESORERIE_ACTIF
NF_CHIFFREAFFAIRE_N NF_CHIFFREAFFAIRE_N	0.65807 <.0001	0.50875 <.0001	0.55745 <.0001
NF_CHIFFREAFFAIRE_N_1 NF_CHIFFREAFFAIRE_N_1	0.60880 <.0001	0.56843 <.0001	0.53960 <.0001
NF_DETTECOURTTERME NF_DETTECOURTTERME	0.44317 <.0001	0.31546 <.0001	0.49910 <.0001
NF_DETTELONGTERME NF_DETTELONGTERME	0.09829 <.0001	0.00600 0.7897	0.24612 <.0001
NF_ACTIFCIRCULANT NF_ACTIFCIRCULANT	0.55918 <.0001	0.44453 <.0001	0.49607 <.0001
NF_EBE NF_EBE	0.84609 <.0001	0.54901 <.0001	0.38742 <.0001
NF_TRESORERIE NETTE NF_TRESORERIE NETTE	0.37997 <.0001	0.27746 <.0001	0.23578 <.0001
NF_TOTALBILAN NF_TOTALBILAN	0.47324 <.0001	0.35384 <.0001	0.54195 <.0001
NF_RESULTATNET_N NF_RESULTATNET_N	0.83818 <.0001	0.61355 <.0001	0.28736 <.0001

Annexe 1 : Statistiques descriptives sur les variables

NF_RESULTAT_NET_N_1 NF_RESULTAT_NET_N_1	0.58983 <.0001	0.87501 <.0001	0.27445 <.0001
NF_FPNETS NF_FPNETS	0.62724 <.0001	0.55077 <.0001	0.48559 <.0001
NF_KPROPRES NF_KPROPRES	0.62204 <.0001	0.54595 <.0001	0.48802 <.0001
NF_FINANPERMANENT NF_FINANPERMANENT	0.51279 <.0001	0.41132 <.0001	0.50712 <.0001
NF_FDR NF_FDR	0.46914 <.0001	0.45341 <.0001	0.25585 <.0001
NF_BFR NF_BFR	0.52696 <.0001	0.45026 <.0001	0.26111 <.0001
RESULTAT_COURANT_N RESULTAT_COURANT_N	0.91712 <.0001	0.65687 <.0001	0.28831 <.0001
RESULTAT_EXP_N RESULTAT_EXP_N	1.00000	0.62532 <.0001	0.33652 <.0001
RESULTAT_COURANT_N_1 RESULTAT_COURANT_N_1	0.62532 <.0001	1.00000	0.28871 <.0001
TRESORERIE_ACTIF TRESORERIE_ACTIF	0.33652 <.0001	0.28871 <.0001	1.00000
RESULTAT_EXP_N_1 RESULTAT_EXP_N_1	0.68126 <.0001	0.89306 <.0001	0.35898 <.0001
CAPACITE_AUTO_FIN CAPACITE_AUTO_FIN	0.81004 <.0001	0.56755 <.0001	0.33992 <.0001
PASSIF_CIRCULANT PASSIF_CIRCULANT	0.39899 <.0001	0.29185 <.0001	0.47837 <.0001

Tableau 36: Matrice de corrélation de Spearman

2. Statistiques descriptives sur les variables

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Defaut	1982	0.24067	0.42760	477.00000	0	1.00000
NF_CHIFFREAFFAI RE_N	1982	101287	649523	200750220	0	23677899
NF_CHIFFREAFFAI RE_N_1	1982	107313	795351	212693404	0	23677899
NF_DETTECOURT TERME	1982	68533	372515	135832747	-36.08450	10342527
NF_DETTELONGTE RME	1982	35783	794691	70922834	-34.70339	34981414
NF_ACTIFCIRCULA NT	1982	68329	364345	135427199	-129.82583	10628967
NF_EBE	1982	8186	66514	16225271	-68195	1911076
NF_TRESORERIE TTE	1982	19525	103237	38697864	-49.16900	2708607
NF_TOTALBILAN	1982	155148	1698778	307503884	0	71217136
NF_RESULTATNET N	1982	2108	42160	4177599	-585221	1356612
NF_RESULTAT_NET _N_1	1982	2573	40901	5098929	-521289	1356612
NF_FPNETS	1982	46949	625211	93052542	-146047	25910832
NF_KPROPRES	1982	48371	640938	95871891	-133729	26629143
NF_FINANPERMAN ENT	1982	86722	1437942	171882549	-65398	62417080
NF_FDR	1982	6192	97703	12271635	-1866636	1763327
NF_BFR	1982	19320	113780	38292316	-1013923	2095910
RESULTAT_COURA NT_N	1982	3332	59009	6604924	-726127	2021984
RESULTAT_COURA NT_N_1	1982	3753	59069	7438526	-579517	2021984
RESULTAT_EXP_N	1982	4791	50972	9495009	-298252	1765581
RESULTAT_EXP_N 1	1982	4932	51015	9775955	-153708	1765581
CAPACITE_AUTO_F IN	1982	5477	50545	10856212	-103690	1527455
TRESORERIE_ACTI F	1982	6289	39498	12465771	-2768	819911
PASSIF_CIRCULAN T	1982	49009	319510	97134883	0	10182289

Tableau 37: Statistiques descriptives générales

3. Statistiques descriptives des indicateurs les plus importants

Chiffre d'affaire de l'année N:

The UNIVARIATE Procedure
Variable: NF_CHIFFREAFFAIRE_N (NF_CHIFFREAFFAIRE_N)

Moments			
N	1982	Sum Weights	1982
Mean	101286.69	Sum Observations	200750220
Std Deviation	649523.39	Variance	4.21881E11
Skewness	26.2091817	Kurtosis	891.926911
Uncorrected SS	8.56079E14	Corrected SS	8.35746E14
Coeff Variation	641.272205	Std Error Mean	14589.5861

Basic Statistical Measures			
Location		Variability	
Mean	101286.7	Std Deviation	649523
Median	9585.3	Variance	4.21881E11
Mode	0.0	Range	23677899
		Interquartile Range	42112

Tests for Location: Mu0=0			
Test	Statistic	p Value	
Student's t	t 6.942396	Pr > t	<.0001
Sign	M 930	Pr >= M	<.0001
Signed Rank	S 865365	Pr >= S	<.0001

Active Windows

Tests for Location: Mu0=0			
Test	Statistic	p Value	
Student's t	t 6.942396	Pr > t	<.0001
Sign	M 930	Pr >= M	<.0001
Signed Rank	S 865365	Pr >= S	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	23677899.30
99%	1481973.43
95%	361542.27
90%	170628.93
75% Q3	44422.10
50% Median	9585.28
25% Q1	2309.91
10%	491.26
5%	0.00
1%	0.00
0% Min	0.00

EBE :

The UNIVARIATE Procedure
Variable: NF_EBE (NF_EBE)

Moments			
N	1982	Sum Weights	1982
Mean	8186.3122	Sum Observations	16225270.8
Std Deviation	66514.3991	Variance	4424165289
Skewness	19.9473708	Kurtosis	473.271511
Uncorrected SS	8.8971E12	Corrected SS	8.76427E12
Coeff Variation	812.507482	Std Error Mean	1494.04558

Basic Statistical Measures			
Location		Variability	
Mean	8186.312	Std Deviation	66514
Median	693.583	Variance	4424165289
Mode	0.000	Range	1979271
		Interquartile Range	2924

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	5.479292	Pr > t 	<.0001
Sign	M	664.5	Pr >= M 	<.0001
Signed Rank	S	710644.5	Pr >= S 	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	1.91108E+06
99%	1.33972E+05
95%	2.31130E+04
90%	1.10291E+04
75% Q3	3.04492E+03
50% Median	6.93583E+02
25% Q1	1.20746E+02
10%	-2.36172E+02
5%	-1.44138E+03
1%	-1.24564E+04
0% Min	-6.81952E+04

Trésorerie nette :

The UNIVARIATE Procedure
Variable: NF_TRESORERIENETTE (NF_TRESORERIENETTE)

Moments			
N	1982	Sum Weights	1982
Mean	19524.6539	Sum Observations	38697864.1
Std Deviation	103237.382	Variance	1.0658E10
Skewness	16.4995065	Kurtosis	360.921085
Uncorrected SS	2.1869E13	Corrected SS	2.11134E13
Coeff Variation	528.753966	Std Error Mean	2318.91675

Basic Statistical Measures			
Location		Variability	
Mean	19524.65	Std Deviation	103237
Median	381.84	Variance	1.0658E10
Mode	0.00	Range	2708657
		Interquartile Range	5642

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	8.41973	Pr > t	<.0001
Sign	M	690.5	Pr >= M	<.0001
Signed Rank	S	478362	Pr >= S	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	2.70861E+06
99%	3.32816E+05
95%	8.51446E+04
90%	3.22650E+04
75% Q3	5.64196E+03
50% Median	3.81844E+02
25% Q1	0.00000E+00
10%	0.00000E+00
5%	0.00000E+00
1%	0.00000E+00
0% Min	-4.91690E+01

Capacité d'autofinancement:

Annexe 1 : Statistiques descriptives sur les variables

The UNIVARIATE Procedure
Variable: CAPACITE_AUTO_FIN (CAPACITE_AUTO_FIN)

Moments			
N	1685	Sum Weights	1685
Mean	6478.253	Sum Observations	10915856.3
Std Deviation	54736.1965	Variance	2996051209
Skewness	18.646997	Kurtosis	429.188363
Uncorrected SS	5.11607E12	Corrected SS	5.04535E12
Coeff Variation	844.922181	Std Error Mean	1333.44364

Basic Statistical Measures			
Location		Variability	
Mean	6478.25	Std Deviation	54736
Median	626.67	Variance	2996051209
Mode	-1073.71	Range	1631145
		Interquartile Range	2160

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	4.858288	Pr > t	<.0001
Sign	M	609.5	Pr >= M	<.0001
Signed Rank	S	512138.5	Pr >= S	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	1527454.605
99%	96138.709
95%	17370.016
90%	7845.715
75% Q3	2312.137
50% Median	626.666
25% Q1	152.357
10%	-265.297
5%	-2361.052
1%	-16144.569
0% Min	-103690.120

Annexe 1 : Statistiques descriptives sur les variables

The UNIVARIATE Procedure
Variable: NF_BFR (NF_BFR)

Moments			
N	1685	Sum Weights	1685
Mean	19142.7631	Sum Observations	32255555.9
Std Deviation	48758.6603	Variance	2377406952
Skewness	4.23071975	Kurtosis	21.4005527
Uncorrected SS	4.62101E12	Corrected SS	4.00355E12
Coeff Variation	254.71067	Std Error Mean	1187.82322

Basic Statistical Measures			
Location		Variability	
Mean	19142.76	Std Deviation	48759
Median	1771.77	Variance	2377406952
Mode	-807.90	Range	455637
		Interquartile Range	12514

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	16.11584	Pr > t	<.0001
Sign	M	440.5	Pr >= M	<.0001
Signed Rank	S	502735.5	Pr >= S	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	446678.4820
99%	262225.2394
95%	113349.0096
90%	54266.0646
75% Q3	12568.4711
50% Median	1771.7740
25% Q1	54.6747
10%	-807.8979
5%	-2555.0494
1%	-6420.1233
0% Min	-8958.0216

Bibliographie :

- « *Comparaison de la performance relative de l'analyse discriminante prédictive linéaire polytypique et de la régression logistique polytypique nominale en contexte de quasi-multi normalité et d'erreur de mesure* » Luc Briere, 2008.
- « *Datamining classification modeles à base de règles : arbres de décision foret aléatoires* »MohemmedHenySelmi, 2016.
- « *Scoring sur données d'entreprises : instrument de diagnostic individuel et outil d'analyse de portefeuille d'une clientèle* »Mireille Bardos, 2008.
- « *Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature*» Hussein A. Abdou, John Pointon, 2011.
- « *Scoring* »Marie Chavent, 2015.
- « *Les déterminants d'octroi de crédits bancaires aux entreprises : étude comparative entre la banque Conventiionnelle et la banque Islamique* »FaizaBenabdelmoula,, 2018.
- « *la notation statistique des emprunteurs ou « scoring »* »Gilbert Saporta.
- « *Rapport annuel 2017*»Crédit Agricole du Maroc, 2017.
- « *Communication financière Activité et Résultats au 31 Décembre 2019* » Crédit Agricole du Maroc, 2019.
- « *CreditScoring and It's Applications* » Jonathan N.Crook, David B. Edelman et L. C. Thomas, 2002.
- « *Probabilités, analyse des données et statistiques* » Gilbert Saporta, 1990.
- « *Analyse discriminante* » Mireille Bardos, 2001.
- « *Un score sur variables qualitatives pour la détection précoce des défaillances d'entreprises* » L. Lelogeais, 2003.

Bibliographie

- « *Benchmarking state-of-the-art classification algorithms for creditscoring* » Baesens et al, 2003.
- « *Statistique et méthodes neuronales* » S. Thiria, Y. Lechevallier, O. Gascuel et S. Canu, 1997.
- « *Discriminant analysis and statistical pattern recognition* » McLachlan et G. J., 1992.
- « *Weighting and selection of variables for cluster analysis* » Gnanadesikan, J. R. Kettenring et L. Tsao, 1995.

