



UNIVERSITÉ D'ABOMEY-CALAVI



ÉCOLE NATIONALE D'ÉCONOMIE APPLIQUÉE ET DE MANAGEMENT

DÉPARTEMENT DE LA STATISTIQUE

Mémoire de fin de formation en vue de l'obtention du diplôme de Licence

Formation : Statistique

Option : Statistiques Économiques et Sectorielles (SES)

Déterminants des performances d'une
équipe de Basketball et prévision de
l'issue des matchs : cas de la NBA

Rédigé par :

Elimane Yassine SEÏDOU

Directeur de mémoire :

Dr M. Nicodème ATCHADE

Statisticien Economètre

Enseignant Chercheur à l'UNSTIM

Maître Assistant

Directeur de stage :

M. Théophile GUIDIBI

Spécialiste BI

*Manager BI à la Direction Générale
de MTN Bénin*

40ème promotion

L'Ecole Nationale d'Economie Appliquée et de Management (ENEAM) n'entend donner ni approbation ni improbation aux opinions émises dans ce document. Les raisonnements et arguments reviennent donc aux auteurs.

Approbation

*Nous certifions que le présent mémoire a été réalisé par son auteur et exempt de tout plagiat.
Il est arrivé à terme et peut être soutenu devant un jury.*

Cotonou le.....

Signatures

Directeur de mémoire :

Directeur de stage :

Dr M. Nicodème ATCHADE

M. Théophile GUIDIBI

Dédicace

A mon père , ma mère et mes soeurs pour leurs soutiens, leurs encouragements et la force qu'ils m'ont donné.

Elimane Yassine SEÏDOU

Remerciements

Cette étude n'aurait été possible sans la contribution directe ou indirecte de plusieurs personnes à qui nous exprimons nos profondes gratitude et nos sincères remerciements. Nos remerciements vont à l'endroit de :

- Professeur Albert HONLONKOUN, Directeur de l'Ecole Nationale d'Economie Appliquée et de Management (ENEAM) ;
- Professeur Théophile K. DAGBA, Directeur Adjoint de l'Ecole Nationale d'Economie Appliquée et de Management (ENEAM) ;
- Dr Nicodème ATCHADE, Statisticien Economètre, Enseignant chercheur à l'Ecole Nationale Supérieure de Génie Mathématique et Modélisation de l'UNSTIM, notre Directeur de mémoire, pour sa patience, ses observations, ses orientations et ses conseils ;
- Monsieur Théophile GUIDIBI notre Maître de stage qui a encadré ce travail ;
- Dr Barthélémy SENOU Chef du Département de la statistique ;
- Monsieur Olivier AWANGNONGBE, et tout le personnel du Marketing de la Direction Générale de MTN Bénin qui n'ont ménagé aucun effort, malgré leurs multiples occupations , pour assurer un encadrement technique appréciable au cours de notre stage ;
- tout le personnel enseignant de l'ENEAM pour la qualité de la formation transmise ;
- enfin, nos parents et amis, tous ceux qui de près ou de loin ont contribué à l'élaboration de ce mémoire ;
- toutes les personnes qui directement ou indirectement ont contribué à la réussite de ce travail.

Avant propos

L'Ecole Nationale d'Economie et de Management (ENEAM) est une école d'enseignement professionnel de l'Université d'Abomey-Calavi (UAC). Elle offre des formations de trois ans dans plusieurs domaines, au nombre desquelles la nôtre : Statistique Economique et Sectorielle. Il est exigé, au bout des trois ans, à tout étudiant admissible d'effectuer un stage pratique en entreprise au terme duquel un mémoire de fin de cycle est rédigé. Pour répondre à cette exigence et obtenir notre licence en statistique à l'ENEAM, nous avons effectué un stage de trois mois à la Direction Générale de MTN Bénin.

Liste des sigles et abréviations

AIC :	Akaike Information Criterion
ARCERP :	Autorité de Régulation des Communications Electroniques et de la Poste
BENINCELL :	Cellulaire Bénin
BI :	Business Intelligence
ENEAM :	Ecole Nationale d'Economie Appliquée et de Management
FBBB :	Fédération Béninoise de Basket Ball
FIBA :	Fédération International de BasketBall
GSM :	Global System for Mobile Communication
MTN :	Mobile Telecommunication Network
NBA :	National BasketBall Association
TIC :	Technologie de l'Information et de la Communication
SBIN :	Société Béninoise des Infrastructures Numériques
UNSTIM :	Université Nationale des Sciences, Technologies, Ingenierie et Mathématiques

Sommaire

Approbation	ii
Dédicace	iii
Remerciements	iv
Avant propos	v
Liste des sigles et abréviations	vi
Résumé	x
Abstract	xi
Introduction	1
1 Cadre institutionnel de l'étude	2
2 Cadre théorique et méthodologique de l'étude	7
3 Résultats, analyse et interprétations	23
Conclusion	33
Références	33
Annexe	a

Liste des tableaux

3.1	Matrice de corrélation	26
3.2	Résultats de l'estimation du modèle	27
3.3	Résultats des tests du choix de modèle	29
3.4	Résultats de l'estimation du logit	29
3.5	Résultat de la pertinence du modèle	30
3.6	Qualité du modèle	30
3.7	Matrice de confusion	31

Table des figures

3.1 Répartition des équipes suivant le niveau performance	23
3.2 Répartition des équipes par âge moyen des joueurs	24
3.3 Répartition des équipes par performances à domicile	24
3.4 Répartition des équipes par performance des entraîneurs	25
3.5 Distribution normale des erreurs	27
3.6 Hétéroscédasticité	28

Résumé

Cette étude a pour but de déterminer les facteurs qui influencent les performances d'une équipe de basketball et de développer un modèle de prévision de l'issue des matchs. Elle se fonde sur les informations collectées sur le site officiel de la NBA et sur d'autres sites tels que USA today et Basketballreference . L'étude a été réalisée sur une population de 30 Clubs professionnels de basketball en ce qui concerne la performance des équipes, et sur une population de 327 matchs en ce qui concerne la prévision de l'issue des matchs . Après l'analyse descriptive des variables considérées, une modélisation économétrique a été effectuée grâce aux logiciels EXCEL et R v3.5.2.

Des résultats obtenus, il ressort d'une part que le niveau de performance d'une équipe est influencé par l'âge moyen des joueurs, par le pourcentage de trois points accordés, par le pourcentage de lancés francs accordé, par la performance de l'entraîneur, par le pourcentage des points accordés et par les nombre moyen de rebonds défensifs de cette équipe. D'autre part l'issue d'un match est influencée par les écarts du nombre de tentatives de tirs , du pourcentage de tirs à trois points , de lancés francs accordés, de rebonds défensifs et de blocs.

Au regard de ces résultats, nous préconisons donc aux différents clubs professionnels qui souhaitent évaluer leurs performances en fin de saisons et ainsi de réguler leur niveau de jeu, de tenir compte l'âge des joueurs et des différents pourcentages de points totaux, de trois points et de lancés francs. Il sera alors aisé de s'auto évaluer et de déduire sur quels points il faudrait travailler pour remonter le niveau du club. Également, nous préconisons à l'ensemble des clubs professionnels de basketball de réaliser des simulation de matchs en prenant en compte le nombre de tentatives de tirs qu'ils réalisent, leurs pourcentage de tirs à trois points, le nombre de lancés francs que les joueurs de l'équipe accordent en ce qui concerne l'attaque, et le nombre de rebonds défensifs et de blocs en ce qui concerne la défense.

Abstract

The purpose of this study is to determine the factors that influence the performance of a basketball team and to develop a model for predicting the outcome of games. It is based on information collected on the official NBA website and other sites such as USA today and Basketballreference. The study was conducted on a population of 30 professional basketball clubs in terms of team performance, and on a population of 327 games in terms of predicting the outcome of games. After the descriptive analysis of the variables considered, econometric modelling was carried out using EXCEL and R v3.5.2 software.

From the results, it appears that the level of performance of a team is influenced by the average age of the players, the percentage of three points granted, the percentage of free throws granted, the performance of the coach, the percentage of points granted and by the average number of defensive rebounds of this team. In addition, the outcome of a game is influenced by differences in the number of shot attempts, percentage three of shots points, free throws granted, defensive rebounds and blocks.

In view of these results, we therefore recommend to the different professional clubs who wish to evaluate their performance at the end of the season and thus to regulate their level of play, to take into account the age of the players and the different percentages of total points, three points and free throws. It will then be easy to evaluate and deduce on what points it would be necessary to work to raise the level of the club. Furthermore, we recommend to professional basketball clubs to carry out simulations of matches taking into account the number of shot attempts they make their percentage of three-point shots, the number of free throws the team's players allow for offense, and the number of defensive rebounds and blocks for defense.

Introduction

Depuis sa création, le niveau de jeu du basketball compétitif a connu une grande évolution d'année en année, autant dans les ligues professionnelles que dans les compétitions amateurs comme les Jeux Olympiques ou les Championnats mondiaux. Les habiletés personnelles des joueurs n'ont cessé d'augmenter en raison des entraînements personnels et collectifs. On remarque par exemple une augmentation de la hauteur moyenne des sauts, de la précision des lancers, tout comme de la vitesse de jeu et des transitions lorsque les équipes prennent possession du ballon. Tout comme pour la majorité des activités de même genre, les jeux en équipe suivent une stratégie bien définie par les entraîneurs et leurs assistants, dans le but de décrocher une victoire. Ces stratégies sont mises en place suivant une certaine logique et en connaissance des aptitudes personnelles des joueurs. L'objectif étant de remporter le titre de champion, les entraîneurs se sont vus obligés d'optimiser leurs stratégies de jeux en se procurant des vidéos de matchs et en utilisant les chiffres collectés sur les équipes adverses. Ce qui leur permettra d'améliorer les tactiques mises en place et de réaliser de belles performances aux termes des saisons. De plus, de manière beaucoup plus précise, les statistiques collectées servent à des simulation de matchs, en confrontant les données de deux équipes. Le match a ainsi lieu avant d'avoir lieu, et les chiffres prédisent déjà le gagnant (mais pas de manière absolue). Le Basketball fait partie des sports les plus pratiqués au Bénin. Principalement marqué par des championnats nationaux ou des tournois non officiels, le basketball béninois a été longtemps absent de la scène internationale. Mais depuis quelques années, le Gouvernement et les grands acteurs du basketball Béninois mobilisent leurs efforts afin non seulement de faire passer ce sport au niveau supérieur avec des entraînements plus qualitatifs les uns que les autres et la construction de plusieurs terrains de basket-ball, mais également de révéler les talents béninois sur le plan international à travers des partenariats internationaux. Ces réformes étant encore très récentes (début 2021), elles ne prennent en compte que l'aspect pratique du sport et ne s'intéressent pas profondément à l'aspect théorique que représente la mise en place de stratégies gagnantes à travers l'étude des chiffres. Le présent travail vise donc à travers les analyses statistiques qui s'y trouvent, apporter sa pierre à l'édifice en corrigeant cet état de chose. En nous inspirant des techniques stratégiques américaines, il s'agira donc dans un premier temps de déterminer les différents facteurs permettant à une équipe de gagner un maximum de matchs pendant une saison de basketball. Nous déterminerons dans un second temps les facteurs qui influencent le plus l'issue d'un matchs afin de réaliser des simulation et développerons un modèle de prédiction. Ce travail permettra aux entraîneurs béninois et d'ailleurs, d'avoir un aperçu des possibilités qui s'offrent à eux pour évoluer.

1 | Cadre institutionnel de l'étude

1.1 Présentation de MTN-Bénin

MTN Bénin nous a servi d'environnement pour notre stage académique. Il s'agira dans un premier temps de présenter la direction générale d'MTN et dans un second temps les différentes activités menées au cours du stage.

1.1.1 Historique, Mission et Objectifs, Structure

Présentation de la Direction de Mobile Telephone Network Bénin (MTN Bénin)

MTN Bénin, anciennement SPACETEL-BENIN S.A filiale du Groupe INVESTCOM GLOBAL LIMITED est une société fournisseuse de services de communication et de transactions Mobiles. Créée le 07 janvier 1998 au Bénin et agréementée par le décret N°99-298 le 11 juin 1999, elle a officiellement démarré son exploitation dans la télécommunication en Novembre 1999 sous le nom commercial de BéninCell. Depuis le temps, elle a su évoluer et est devenue le leader sur le marché GSM au Bénin en donnant une nouvelle dimension à la communication à travers ses nombreux services. Jour après jour, les services de BéninCell se sont développés en fonction d'une priorité unique et constante : fournir le meilleur service de technologie à ses abonnés.

Evolution : De BéninCell à Areeba puis MTN

De BéninCell à Areeba puis MTN. Le temps passait et le groupe INVESTCOM voulait rendre Spacotel plus grand et plus fort. Ses conquêtes dans le monde des GSM se sont multipliées en Afrique et au Moyen-Orient. Le seul problème qui se posait est que chacune des filiales du groupe voulait garder son propre nom commercial, ce qui ne démontrait pas du tout la puissance du groupe. Il fallait donc trouver une formule pour les harmoniser. C'est ainsi qu'une mutation s'opéra : celle du passage des noms commerciaux des filiales au nom commun d'Areeba. Un réseau GSM international qui met ses clients au cœur de ses activités. Areeba devint donc le Label d'INVESTCOM. Un nouveau réseau avec une vision différente pour :

- maintenir le contact avec ses abonnés ;
- fournir une parfaite communication sans fil (mobile) ;
- être connecté au monde GSM International ;
- offrir de nouveaux services ;
- améliorer la couverture du réseau ;
- avoir un meilleur rapport qualité / prix.

Ainsi, en 2005, Areeba Bénin devint le nouveau nom commercial de la société Spacotel Bénin. Après six années de labeur avec pour seul objectif : la satisfaction totale de la clientèle, en 2006, le Groupe

Sud-Africain M-Cell devenu par la suite MTN International a acquis la licence du réseau de téléphonie mobile Areeba (nom commercial de la société de télécommunication Spacetel appartenant à INVESTCOM) en bourse. Selon le tableau de bord du 31/12/2021 de l'ARCEP Bénin, MTN détient 56,69% de la part du marché des réseaux GSM au Bénin ce qui fait de lui le leader du secteur de la téléphonie mobile au Bénin. La nouvelle Vision de Spacetel Bénin (MTN Bénin) est d'être à l'avant-garde de la fourniture d'un audacieux et nouveau monde numérique pour ses clients.

Mission et Objectifs de MTN Bénin

MTN Bénin hérite de la même mission que son fondateur MTN International sur son territoire avec pour objectif de :

- construire un monde meilleur à travers le financement de projets de développement communautaire ayant un impact durable ;
- contribuer à l'amélioration des conditions de vie des couches vulnérables à travers des programmes d'investissement dans le domaine social, du développement de la capacité des communautés à se prendre en charge grâce au financement de projets qui concourent à leur bien-être tout en participant activement au développement économique et social de la nation béninoise.

1.1.2 Structure organisationnelle et environnementale de MTN Bénin

Cette partie de notre travail a été consacrée à la présentation du fonctionnement et de la structure organisationnelle de MTN BÉNIN.

ORGANE DE DECISIONS

Conseil Administratif CA ; Le Conseil Administratif détermine l'orientation des activités de MTN Bénin et veille à leur mise en œuvre sous réserve des pouvoirs spécialement attribués à l'assemblée des actionnaires et dans les limites de l'objet social. Il se saisit de toutes les questions touchant à la bonne marche de la société et règle par ses délibérations les affaires la concernant. Il est composé d'administrateurs nommés par l'Assemblée Générale des actionnaires et du Directeur Général en fonction qu'elle a préalablement élus. Ce conseil se réunit deux (02) fois par an sauf en cas de nécessité où il est convoqué pour des séances extraordinaires.

La Direction Générale : Le Directeur Général bénéficie d'un large pouvoir de décision lui permettant de mettre en œuvre les dispositions du Conseil d'Administration. Il assure la gestion technique, administrative, financière, commerciale et informatique de MTN Bénin. Il a pour mission de :

- définir des stratégies relatives aux objectifs fixés ;
- coordonner les moyens disponibles ;
- définir des approches logistiques ;
- définir des politiques financières, d'investissement et celle de nomination.

Le Directeur Général est assisté d'une Secrétaire Particulière qui s'occupe des courriers, des relations extérieures et des fonctions habituelles de secrétariat.

ORGANES FONCTIONNELS

Pour mieux accomplir sa mission, MTN-BÉNIN s'est doté d'un organigramme composé de diverses fonctions regroupées autour de la Direction Générale. Ainsi, MTN possède à son actif onze(11) départements dont neuf sont efficacement fonctionnels. Il s'agit du :

- département Technique et Informatique ;
- département Commercial (Marketing et vente) ;
- département Experience Client (Customer Care and Costumer experience) ;
- département Finances ;
- département Gestion des Audits et Risques ;
- département des Ressources Humaines et des Relations ;
- département MTN Business Unit (EBU) ;

1.1.3 Le micro environnement de MTN Bénin

Il est composé des acteurs en contact direct avec l'entreprise, qui influencent son activité et sur lesquels cette dernière peut agir. Il s'agit de :

Les Clients : Les opérateurs GSM desservent plusieurs catégories socioprofessionnelles.

Les Concurrents : MTN Bénin évolue dans un environnement très concurrentiel tant direct qu'indirect.

Les concurrents directs : ETISALAT Bénin : filiale du groupe ETISALAT opérant sous la marque de MOOV Africa

Les concurrents indirects :

- ISOCEL : opérateur qui est un fournisseur d'accès internet et qui s'occupe de mettre sur le marché toutes les offres/forfaits pour les entreprises ou pour les domiciles. C'est un concurrent assez important car il tire sa grande part de marché au niveau des clients data (internet) ;
- Les Opérateurs GSM partenaires Roaming et Interconnect : ce sont les opérateurs qui permettent à MTN Bénin de réaliser des communications ON et OFF net locales et internationales. Ils sont souvent gérés par des carriers des sociétés de liaison pour opérateurs.
- SBIN

1.1.4 Le macro environnement de MTN Bénin

Ce sont les facteurs dits incontournables auxquels MTN Bénin doit s'adapter. Ils sont à la fois des sources d'opportunité et de menace. Les aspects démographiques, économiques, technologiques, politico-légaux et socioculturels obligent MTN à une mutation perpétuelle.

- Environnement démographique : estimé en 2018 à 11.469.140, la population béninoise est dominée par une jeunesse active qui prend de plus en plus conscience de l'utilité de la communication. Le niveau d'éducation croissant qu'atteint notre pays reste aussi une source d'opportunité pour les opérateurs mobiles.

- Environnement économique : la paix sociale qui règne au Bénin, la réhabilitation du réseau routier facilitant les tracs, la mise aux normes internationales du port autonome de Cotonou et les nombreux investissements favorisant un taux de croissance rendent l'environnement économique favorable au développement des activités de la société. MTN Bénin contribue à accroître les recettes budgétaires de l'Etat à travers le paiement de diverses taxes. Elle participe aussi à l'amélioration du bien-être de la population béninoise et à la création de richesse nationale.
- Environnement politico-légal : Ce domaine est soutenu par le MINISTÈRE DE L'ÉCONOMIE NUMÉRIQUE qui s'occupe de la réglementation de l'espace téléphonique au Bénin, puis par l'Autorité de Régulation des Communications Electroniques et de la Poste (ARCEP), créée par décret No 2007-209 du 10 mai 2007 déssant ses attributs, son organisation et son fonctionnement. Elle a pour but de veiller au respect des dispositions des cahiers des charges.
- Environnement technologique : le Bénin, bien qu'il soit un pays en voie de développement essaie de s'armer en matière de technologie. Les opérateurs mobiles exploitent une norme numérique de téléphonie cellulaire fonctionnant sur la bande des 900Mhz. L'essentiel des équipements qui constituent l'ossature des télécommunications au Bénin, hormis les installations des opérateurs GSM eux-mêmes, appartiennent à BENIN TELECOM.

1.2 Déroulement du stage et observation

1.2.1 Déroulement du stage

Au sein de MTN Bénin, notre stage a été effectué dans le Département Commercial (Marketing)

Le département Commercial (Marketing) : a pour mission de s'occuper de l'élaboration des prévisions de vente et est responsable de la mise en oeuvre des plans d'actions. Il consacre son travail surtout sur : l'étude du marché, la définition du produit ou du service adapté aux attentes des clients segmentés, la détermination du prix de vente d'un produit, le choix des canaux de distribution, l'élaboration d'une stratégie de communication à travers la publicité. Il veille au respect par les distributeurs agréés et tout le réseau de distribution des conditions et des règles établies par l'entreprise.

Mes activités de stage ont consisté en une rotation dans les différentes sections du marketing que sont : le consumer segment (section dans laquelle les différentes propositions de valeurs sont pensées, mise en place et soumises au BI pour évaluation de faisabilité et d'implication et de conséquences) , le product and services (section s'occupant de la mise en place effective des propositions de valeurs pensées par le segment en relation avec l'IT), le brand and com (section chargée de tout ce qui a rapport avec la communication sur les réseaux et l'image de la marque) et le Business Intelligence (section chargée de l'évaluation des différentes propositions de valeurs, des performances réalisées au vue des différentes propositions de valeur préalablement mise en place).

- Au consumer segment : j'ai participé à l'exécution du projet de concours de logo de la Fresh Zone
- Au product and services : j'ai joué le rôle de testeur d'offres déjà en place et j'ai rapporté les différentes anomalies rencontrées
- Au Brand and Com : Ma tâche a consisté à répondre à des courriers de sponsorings et prendre

part à l'élaboration du calendrier des événements

- Au Business Intelligence : Il s'agissait à ce niveau d'extraire des informations utiles à l'entreprise en faisant des calculs mathématiques et statistiques à partir de différentes bases de données dans Excel, de produire des rapports comparatifs par période.

Chacune de ces activités don't principalement celles réalisées au Brand and com et au B.I ont participé à la définition et à l'élaboration de mon thème : Analyse des déterminants des performances sportives d'une équipe de basket ball et prévision de l'issu des matchs : cas de la NBA.

1.2.2 Observation

L'entreprise MTN nous a offert un cadre exceptionnel de travail où nous avons eu le plaisir de mettre en application nos connaissances théoriques. Notre stage s'est déroulé dans une ambiance de travail agréable et conviviale. Le personnel de ce département nous a permis de consolider notre logique de raisonnement scientifique, notre démarche , et de définir clairement notre thème. Néanmoins, il faut relever que cette entreprise fait face à certaines faiblesses telles que :

- l'instabilité du réseau ;
- la non couverture totale du territoire national ;
- les prix élevés des abonnements,

Nous avons été confrontés à certaines difficultés au cours de nos recherches notamment à la confidentialité des données de l'entreprises.

2 | Cadre théorique et méthodologique de l'étude

Le présent chapitre expose le cadre théorique et méthodologique de l'étude. Il s'articule autour de trois (3) parties à savoir : l'énoncé du problème et l'intérêt de l'étude, les objectifs et les hypothèses de la recherche et enfin la revue de littérature et la méthodologie.

2.1 Enoncé du problème et intérêt de l'étude

2.1.1 Enoncé du problème

L'industrie des sports est l'une des plus répandues dans le monde et en Afrique. Les activités sportives sont souvent marquées par des championnats ou des tournois. Aussi divertissant qu'elles soient, les professionnels du domaine fournissent d'énormes efforts afin d'optimiser leurs prestations sportives et de remporter la victoire lors des compétitions. Le basketball, l'un des sports les plus pratiqués au monde ne fait pas exception à la règle. Face aux différents enjeux sportifs et économiques, les dirigeants d'associations de basketball comme la NBA ou la FIBA, les détenteurs de franchises et les entraîneurs, ont dû mettre en place des stratégies permettant d'optimiser ce sport, d'améliorer les performances d'une équipe, d'attirer de plus en plus de supporters. C'est donc dans le but d'améliorer les performances des équipes que les entraîneurs de la NBA choisissent des joueurs présentant certaines capacités et pouvant apporter une plus value sur l'un des nombreux points d'une équipe. De plus toujours dans le but d'optimiser les prestations sportives, certaines simulations de matchs, purement basées sur les chiffres ont été effectuées afin de savoir, au vu des statistiques connues, quelle équipe a le plus de chance de gagner et, comment éventuellement battre cette équipe. Les équipes mettant en place des stratégies de simulations de matchs, présenteront de ce fait de meilleurs résultats que celles n'ayant pas fait de simulations.

Au Bénin, les différentes compétitions officielles de basket ball sont régulées par la Fédération Béninoise de Basket Ball actuellement dirigée par Ismaïhil Onifadé. Dans le cadre du développement du basket ball béninois la NBA et le Bénin se sont engagés dans une collaboration. Les deux parties ont de ce fait signé un protocole d'accord (un mémorandum), le mardi 17 mai 2022, au ministère des affaires étrangères et de la coopération. Le ministre a ainsi informé que la NBA est entrain de mettre en place une grosse ligue professionnelle (la basket ball african league) en Afrique depuis 3 saisons et qu'il faut que le Bénin en fasse partie.

Malheureusement, dans son processus de développement, le basket ball béninois connaît un déficit stratégique en terme de simulations chiffrée de matchs, ce qui est un réel retard au regard des techniques utilisées sur la scène internationale. Autrement, il apparaît nécessaire pour une évolution stratégique, de mettre en place des dispositifs permettant non seulement de déterminer les réels besoins des équipes, mais aussi de préparer de meilleures stratégies gagnantes. L'organisation du championnat national de Basket Ball étant similaire à celle de la NBA, et la NBA étant l'association de

basketball la plus organisée à ce jour et mettant déjà en place diverses techniques visant à dynamiser le Basketball, elle apparaît comme le modèle idéale à partir duquel les réflexions faites sur les sujets des performances et de la prévision peuvent servir aux différentes leagues de basketball. C'est dans ce schéma que nous viens les interrogations sur : Quelles sont les déterminants des performances d'une équipe sportive de basketball et comment prévoir l'issue des matchs de la NBA ?

2.1.2 Intérêt de l'Etude

Dans le cadre de la redynamisation du secteur du basketball Béninois, plusieurs objectifs ont été énoncés par la fédération de basket ball béninoise présidée par monsieur Ismahini Onifadé à savoir :

- oeuvrer pour une augmentation de 30% du nombre de pratiquants au niveau national et multiplier par 3 le nombre de licenciés soit passer de 1600 à 500 don't 50% de jeunes garçons et filles ;
- renforcer la capacité des entraîneurs, arbitres et officiels et atteindre la barre des 150 000 pratiquants à l'horizon 2025 ;
- faire en sorte que les équipes béninoises prennent part aux compétitions internationales telles que les éliminatoires afro basketball sénior u20, u16 et u18 masculins et féminins en 2022 et à la coupe d'afrique Can 3x3 2022 au Maroc ;
- elle va continuer l'organisation du MTN Basket league 3x3 et de MTN super ligue junior (supportées toutes les 2 par MTN) ;
- créer et animer des centres de formation régionnaux de la FBBB.

De plus, cette redynamisation est passée par la signature d'un mémorandum entre le ministre des sports Oswald Omeke et le président de la NBA Africa Victor Williams, le sponsoring de la ligue junior et de la league 3x3 par MTN et le réaménagement des terrains de basket ball de l'ancien centre Soweto construit par MTN.

Cette étude permettra d'une part à travers la détermination des facteurs influençant les performances sportives d'une équipe de basketball, de mettre en place de meilleures stratégies de constitution des clubs professionnels et de l'équipe nationale et de ce fait, disposer d'équipes très compétitives sur le plan national et international. Cela permettra également de renforcer l'aspect économique de ce sport et d'attirer les investisseurs et sponsors dans le domaine du Basketball. MTN étant déjà un grand collaborateur en ce qui concerne le sponsoring du basketball Béninois, cela renforcera la confiance de l'entreprise en ce qui concerne la continuité de son appuie pour ce sport au Bénin. D'autre part, l'étude de la prévision sur l'issue des matchs, ce travail permettra d'effectuer des simulation de matchs utiles à la prise de décision en terme de tactique de jeu.

2.2 Objectifs et hypothèses de la recherche

2.2.1 Objectifs de la recherche

L'objectif général de l'étude est d'identifier les déterminants des performances sportives d'une équipe sportive de basket ball et de prévoir l'issue des matchs. De façon spécifique, il s'agira de :

- déterminer les facteurs qui expliquent le niveau de performance d'une équipe sur 72 matchs ;

- déterminer les facteurs qui influencent l'issue d'un match à partir de l'écart des points collectés pendant des matchs entre deux équipes et prévoir l'issue des matchs

2.2.2 Hypothèses de la recherche

Pour atteindre ces objectifs, deux hypothèses ont été formulées :

H_1 : les performances de l'entraîneur, l'âge moyen des joueurs et le pourcentage de tirs accordés expliquent la performance d'une équipe sur 72 matchs.

H_2 : le pourcentage de 3 points accordés, de lancers francs et le nombre de rebonds défensifs influencent l'issue d'un match

2.3 Revue de littérature

Cette partie aborde la revue de la littérature et la méthodologie de la recherche.

2.3.1 Clarification conceptuelle

Le basket-ball a été inventé en décembre 1891 par James Naismith, professeur d'éducation physique canado-américain au Springfield College, dans l'État du Massachusetts (États-Unis). Le basket-ball ou basketball, fréquemment désigné en français par son abréviation basket, est un sport collectif de balle opposant deux équipes de cinq joueurs sur un terrain rectangulaire. L'objectif de chaque équipe est de faire passer un ballon au sein d'un arceau de 45 cm de diamètre, fixé à un panneau et placé à 3,05 m du sol : le panier. Chaque panier inscrit rapporte deux points à son équipe, à l'exception des tirs effectués au-delà de la ligne des trois points (qui rapportent trois points) et des lancers francs accordés à la suite d'une faute (qui rapportent un point). L'équipe avec le nombre de points le plus important remporte la partie. Les matchs de basketball se déroulent en 4 quarts temps de 12 minutes séparés entre eux par des pauses de 2 min sauf entre le deuxième et le 3ème quart temps où la pause est de 15 min. Le basket-ball se pratique exclusivement à la main, et les joueurs peuvent se déplacer avec la balle en main en dribblant sur le sol ou en effectuant deux pas maximum sans dribbler. L'équipe en possession du ballon (l'équipe qui attaque) tente d'inscrire des points en réalisant des tirs, des double-pas ou des dunks, tandis que l'équipe en défense essaie de les en empêcher en réalisant des interceptions de balle ou des contres. Si le tir échoue, les joueurs des deux équipes tentent d'attraper la balle au rebond. La NBA (National Basketball Association, créée en 1949) est l'organe qui coordonne la pratique professionnelle du basketball aux États-Unis. Au Bénin, la professionnalisation des équipes de basketball apparaît encore récente avec une quasi absence sur la scène internationale. Les équipes du Bénin de basket-ball regroupent la sélection des meilleurs joueurs béninois. Elles sont placées sous l'égide de la Fédération béninoise de basket-ball. Classé 145ème au 7 Décembre 2021 avec 49.1 points d'après le classement Mondial FIBA Masculin, le palmarès de l'équipe masculine comprend la participation aux phases finales du championnat d'Afrique de 1974, où l'équipe a terminé 9e du tournoi.

Il s'avère important de définir quelques concepts :

- Conférence : Il y a trente équipes en NBA réparties en deux conférences, Ouest et Est. Pour limiter les longs déplacements, les équipes de conférences différentes ne s'affrontent que deux

fois dans la saison, contre trois fois pour la même conférence et quatre fois pour la même division.

- Division : Il y a six divisions dans la NBA, trois par conférence. Toutes comportent cinq équipes. On retrouve à l'Ouest, la division Nord-Ouest, la division Pacifique et la division Sud-Ouest. À l'Est, il y a la division Atlantique, la division Centrale et enfin la division Sud-Est .

- MVP : Most Valuable Player. Trophée remis au meilleur joueur de la saison, remporté par le Grec Giannis Antetokounmpo l'année dernière. Un peu comme le Ballon d'Or mais en différent.

- Franchise player : leader d'une franchise, titulaire d'un contrat long et autour de qui l'équipe est organisée.

- Playoffs : Les playoffs correspondent au championnat en fin de saison régulière pour décerner le titre de champion. Il y a d'abord un championnat interne entre les conférences pour élire le champion de conférence. Les séries se font au meilleur des sept matches (le premier arrivé à quatre victoires remporte donc la série). Le septième match se jouant automatiquement chez l'équipe la mieux classée en saison régulière. La finale NBA oppose ensuite le champion de l'Est au champion de l'Ouest.

- Saison : La saison de NBA débute fin octobre et s'achève début avril. Chacune des 30 équipes, réparties en 2 conférences de 15, doit disputer 82 matches, 41 à domicile et 41 à l'extérieur. Pour atteindre ce total, les équipes affrontent 4 fois les 4 autres franchises de leur division (16 matches au total), 3 ou 4 fois les autres équipes de leur conférence (36 matches) et 2 fois les équipes des équipes de l'autre conférence (30 matches).

- Assist (assiste) : Une passe décisive.

- Rebond défensif : Une récupération de balle après un tir adverse raté.

- Tir à trois points : Un tir accordé derrière la ligne des 3 points, à 6 mètres environ du panier.

- Faute : Une violation résultant d'un contact illégal avec un joueur adverse.

- Panier accordé : Un tir ayant été concluant et rapportant des points à l'équipe.

- Lancés francs : Des lancés de paniers derrière la ligne réservée aux lancés francs, sans aucune défense, qui surviennent après une faute sur un joueur. Lorsqu'ils sont accordés, ils valent chacun 1 point.

- Rebond offensif : Une récupération de balle par l'équipe tireuse après un tir manqué de cette dernière.

- Contre-attaque (Turnover) : Perte de possession de la balle due à une erreur ou à une faute et attaque de l'équipe adverse.

- bloc : arrêt de balle par l'équipe en défense, lors d'un tir de l'équipe attaquante

- bloc illégal : arrêt de balle par l'équipe en défense au moment où la balle retombe vers le cerceau

- La salary cap : Le salary cap est constitué de l'ensemble des salaires prévus par l'ensemble des contrats de l'effectif

2.3.2 Revue empirique

Dans cette revue, il sera présenté les approches les plus utilisées pour traiter les questions relatives aux facteurs influençant la performance sportive d'une équipe de basketball, aux facteurs influençant l'issue d'un match, et à la prévision de l'issue des matchs.

Performance sportive d'une équipe de basket ball La performance d'une équipe sportive de basket ball est un produit direct des actions de l'équipe mais également des actions individuelles menées par les joueurs. Ainsi la performance des joueurs en sports collectifs ne se limite pas seulement aux points marqués par une équipe (Nadeau, 2001). La performance est davantage la contribution de l'ensemble des actions des joueurs au projet commun d'une équipe qui vise, ultimement, à marquer plus de points que l'adversaire au cours d'un match (Nadeau, 2001). Grâce à une analyse détaillée des détails de chaque portion de jeu, les entraîneurs d'une équipe peuvent soit constater plus clairement les carences dans le jeu d'un joueur en particulier ou de certains joueurs ayant un rôle particulier (comme les joueurs en défensive par exemple), ou encore les carences d'une équipe entière (Collins Hodges, 1978). L'aspect le plus important à considérer pour bien comprendre la performance en sport collectif est que ces composantes sont toutes inter reliées entre elles et en constante interaction. Ainsi, une action motrice réalisée par un ou des athlètes en situation de jeu est généralement liée à plusieurs de ces composantes à la fois. Cela est à tel point vrai qu'il est difficile, voire erroné, de tenter de les dissocier pour expliquer ladite performance (Nadeau, 2001).

Les statistiques de jeu permettent de faire un état des lieux d'une équipe, d'évaluer où sont ses points forts et ses points faibles de manière quantitative. Les statistiques de jeu se basent sur le recueil des actions motrices considérées déterminantes pour la performance d'un joueur ou d'une équipe (Nadeau et al, 2008). Les actions répertoriées sont cumulées d'un match à l'autre de manière de plus ou moins détaillée selon les ressources disponibles (Nadeau et al, 2008). Avec l'avancement de la technologie, les statistiques de jeu sont de plus en plus nombreuses et explicites (Nadeau et al, 2008). Leur gros avantage est qu'elles fournissent des données essentiellement objectives des actions motrices liées à la performance. Les statistiques de jeu renseignent principalement sur les actions du porteur et se focalisent essentiellement sur les points marqués (Nadeau, et al., 2008 ; Renger, 1994 ; Litalien, 2014, Fortier 2014). De plus, elles fournissent une grande quantité d'informations sur des catégories d'actions réalisées par les joueurs selon le jeu effectué telles que les minutes jouées, les rebonds, les balles perdues à l'adversaire, les interceptions, les fautes personnelles, les fautes provoquées (Ahlinvi, 2019), etc.

Mais malgré l'abondance de ces données souvent disponibles, ces informations sont habituellement traitées de manière isolée ou ne renseignent pas sur l'ensemble des joueurs qui ont participé au match (Georget, 2013). Par exemple, il est possible de savoir le nombre de tirs ou le nombre de passes qu'un joueur a effectué dans un match. Par contre, il peut être difficile de décrire plus précisément ce qui a mené à ce tir ou à cette passe et qu'est-ce qui est arrivé ensuite. Or, dans certains cas, certains joueurs ne feront que rarement des tirs au panier même si leur rôle est essentiel à l'équipe afin de permettre aux autres coéquipiers de tirer. (Ahlinvi, 2019). Les équipes professionnelles ont habituellement des ressources presque illimitées ce qui explique la grande quantité

de données recueillies à chaque match (Reid et al., 2010). Enfin, les données recueillies grâce aux principales statistiques de jeu expliquent généralement peu la performance défensive des joueurs ou d'une équipe (Litalien, 2014). Les principales statistiques défensives concernent surtout des actions qui empêchent l'adversaire de marquer, mais ne fournissent que peu d'informations sur le processus, c'est-à-dire sur ce qui a mené au résultat de l'action observée (Godbout, 1988).

Facteur influençant l'issue d'un match de basketball : Des heures d'analyse ont été passées en NBA pour savoir quels sont les facteurs clés pour gagner un match. Dean Oliver, analyste de données et entraîneur de basket-ball qui a travaillé pour Seattle Supersonics et Denver Nuggets entre autres, a développé une théorie entre 2002 et 2004 qui était sur le point de changer la façon dont les chiffres et le basketball sont associés. Selon sa théorie des « quatre facteurs », ces aspects sont basés sur la façon dont une possession pourrait éventuellement prendre fin.

L'efficacité au tir, aux turnovers, aux rebonds et aux lancers francs peut apporter des victoires aux équipes. Ils sont mesurés à l'aide de quatre statistiques d'équipe, avec un poids différent attribué à chacun d'eux. Le field goal efficace ou EFG% (40%), le taux de rotation ou TOV% (25%), le taux de rebond offensif ou ORB% (20%) et le taux de lancers francs ou FTR (15%) ont été définis par Oliver. Ces facteurs s'appliquent également à la défense. Cette théorie semble en fait être assez proche de la réalité, car ils prennent en compte les principes fondamentaux du basketball. Marquez beaucoup, ne retournez pas le ballon, attrapez chaque rebond et tirez des fautes de tir (Kotzias, 2018). Le l'entraîneur Mike Sorrell affirme que 5 facteurs influencent l'issue d'un match de basketball :

- 1) Le pourcentage des points marqués ;
- 2) Le pourcentage des trois points ;
- 3) les rebonds ;
- 4) les contre attaques ;
- 5) les fautes personnelles.

Prévision de l'issue des matchs : Grâce à divers algorithmes capables de prédire l'enregistrement attendu du joueur dans chacune de ces quatre catégories (ou huit comme expliqué) et comment ils contribuent au record total de l'équipe, il est possible de construire des modèles qui fournissent des estimations. Les travaux de Eric Jones en 2016 portant sur la prévision de l'issue des matchs, ont montrer que la différence des statistiques des équipes permet de prédire l'issue des matchs : 7 facteurs ont été retenus : FGS, 3PS, FTS, ORS, ASTS, TOS, FTAS.

Les travaux de Wei-jen et al sur le développement d'un modèle hybride intégrant 5 des méthodes d'explorations des données ont montrés que les statistiques importantes (caractéristiques) basé sur quatre décalages de jeu sont des rebonds défensifs moyens, en moyenne de deux points pourcentage de field goal, pourcentage moyen de lancers francs, moyenne de rebonds offensifs, moyenne passes décisives et moyenne de tentatives de field goal à trois points

2.4 Méthodologie de recherche

2.5 Méthodologie

Sous cette rubrique, nous allons présenter l'approche méthodologique ayant permis d'analyser les données collectées dans le cadre de cette recherche. Il sera question de décrire notre méthodologie d'analyse, de spécifier nos modèles et enfin d'exposer notre procédure d'estimation et d'analyse.

2.5.1 Choix et présentation des variables

L'analyse porte sur les matchs de la saison régulière de la NBA 2020-2021. À partir des informations de la littérature et celles mises à notre disposition sur les statistiques collectées pendant les différents matchs et sur les données externes aux matchs, dans un premier temps une base 30 individus constituant les 30 équipes de la NBA a été sélectionnée et traitée en ce qui concerne les performances sportives, puis un échantillon de 327 matchs a été sélectionné de manière aléatoire. Les variables qui entrent dans la construction du modèle d'analyse sont :

2.5.1.1 Variables relatives aux déterminants des performances sportives d'une équipe de Basketball :

- Performance équipe (PE) : la variable dépendante correspondant aux nombres de matchs gagnés sur les 82 matchs joués au cours de la saison, une équipe est performante lorsqu'elle gagne plus de 50% des matchs joués.
- Points moyens du franchise player (PMFP) : elle correspond aux points moyens accordés par le joueur phare de l'équipe.
- Performance à domicile (PAD) : il s'agit du nombre de matchs gagnés sur le nombre de matchs joués à domicile.
- Salary cap (SC) : Le salary cap est constitué de l'ensemble des salaires prévus par l'ensemble des contrats de l'effectif pour chaque équipe.
- Age Moyen(Age) : désigne l'âge moyen des joueurs par équipe.
- Performance de l'entraîneur (PDC) : désigne le nombre de matchs gagnés par le l'entraîneur sur le nombre total de matchs joués depuis le début de sa carrière.
- FGP : Pourcentage de paniers accordés par l'équipe.
- FGA : Moyenne des tentatives de tirs.
- 3PP : Pourcentage de 3 points accordés par l'équipe.
- 2PP : Pourcentage de 2 points accordés par l'équipe.
- FTP : Pourcentage de lancés francs accordés par l'équipe.
- ORB : Nombre moyen de rebonds offensifs par l'équipe par match.
- DRB : Nombre moyen de rebonds défensifs par l'équipe par match.
- STL : Nombre moyen de vol de balle par l'équipe par match.
- TOV : Nombre moyen de contre-attaque réalisé par l'équipe par match.

- PF : Nombre de moyen de fautes personnelles comptabilisées par les joueurs de l'équipe par match.
- PTS : désigne en moyenne combien de points l'équipe met par matchs.

2.5.1.2 Variables relatives aux déterminants de l'Issue d'un match de Basketball :

En considérant l'équipe A comme l'équipe pour laquelle nous observons l'issue du match ;

- ISSUE : C'est la variable dépendante correspondant à la victoire ou à la défaite d'une équipe lors d'un match. Lorsque l'équipe observée gagne, la variable Issue vaut 1. Elle vaut 0 dans le cas contraire.
- DFGA : Différence (ou écart) entre les points accordés de l'équipe A et de l'équipe B pendant le match joué.
- D3PP : Différence entre les pourcentages de tir à trois points accordés de l'équipe A et ceux de l'équipe B pendant le match joué.
- DFTM : Différence entre les lancés francs accordés de l'équipe A et ceux de l'équipe B pendant le match joué.
- DBLK : Différence des arrêts de balles après un tir de l'équipe adverse entre l'équipe A et l'équipe B pendant le match joué.
- DDREB : Différence des récupérations défensives de balles entre l'équipe A et l'équipe B pendant le match joué.

2.6 Analyse descriptive et économétrique des variables

2.6.1 Analyse descriptive

Les méthodes descriptives regroupent des techniques d'analyses univariée et bivariée. L'analyse univariée permet de dresser un portrait de la population étudiée à partir de sa répartition suivant la variable d'intérêt de l'étude et les principales variables explicatives retenues. L'analyse bivariée a pour but d'identifier les associations entre deux variables et d'apprécier la significativité statistique de cette association à partir d'un seuil donné. Dans le cadre de cette étude, elle permettra non seulement de voir la corrélation existante entre la variable dépendante et les variables explicatives.

2.6.2 Modèle de régression

Modèle de régression linéaire multiple

On considère une population P de 30 individus distinguables par des variables X_1, \dots, X_p . Soit Y la variable quantitative à expliquer. On souhaite expliquer Y à partir de X_1, \dots, X_p . Le principe de la régression linéaire multiple est simple : Déterminer la variable expliquée Y. Déterminer (p - 1) variables explicatives X_1, \dots, X_{p-1} . Il ne reste plus qu'à appliquer un modèle linéaire :

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon \quad (2.1)$$

Lorsqu'on détermine la droite qui passe au plus près de tous les points (on dit qu'on ajuste une droite aux données observées), on estime aussi les paramètres a et b . Ces estimations peuvent également être obtenues à partir de formules dérivées de la méthode du maximum de vraisemblance.

Hypothèses :

- X n'est pas stochastique (X est non aléatoire) et est de plein rang soit $\text{rang}(X) = p$ c'est à dire que la matrice $(X'X)$ est inversible.
- Les erreurs sont des variables aléatoires, les variables explicatives sont non aléatoires ;
- la relation entre les deux variables doit être globalement linéaire, au moins grossièrement. C'est pour cette raison, qu'il faut toujours représenter graphiquement les données avec un scatter plot avant de choisir la méthode d'analyse ;
- les réponses doivent être indépendantes. C'est le plan d'échantillonnage qui renseigne sur cette condition. Si les données proviennent d'individus ou d'unités expérimentales différentes, elles sont généralement indépendantes. En revanche, si la variable indépendante est temporelle, les données ne sont sans doute pas indépendantes ;
- Les covariances des erreurs sont nulles : les erreurs ne sont pas autocorrélées

$$\text{Cov}(X, \epsilon) = 0 \quad (2.2)$$

- les erreurs doivent suivre une loi normale (hypothèse de normalité) et être homogènes (hypothèse d'homoscédasticité). Ces conditions doivent être testées. Des méthodes visuelles ou des tests statistiques peuvent être employés. Pour la normalité, il est possible de réaliser un QQplot ou d'utiliser le test de Shapiro Wilk. Un défaut de normalité est mis en évidence lorsque la p -value du test est ≤ 0.05 , *etsur le QQplot lorsque les points s'écarteraient systématiquement de la droite* ;

$$\epsilon \rightarrow N(0, \sigma_\epsilon^2) \quad (2.3)$$

où σ_ϵ^2 est la variance du terme d'erreur

- Pour l'homoscédasticité, on utilise un "standardized residuals vs fitted plot" ou bien le test de Breusch-Pagan. Le plot des résidus standardisés vs. les fitted values met en évidence un défaut d'homoscédasticité lorsque la dispersion des résidus augmente avec l'augmentation des valeurs prédites (fitted values)

$$E(\epsilon\epsilon') = \sigma^2 \quad (2.4)$$

Modèle de régression logistique

On considère une population P divisée en 2 groupes G_1 et G_2 distinguables par des variables X_1, \dots, X_p . Soit Y la variable qualitative valant 1 si l'individu considéré appartient à G_1 et 0 sinon. On souhaite expliquer Y à partir de X_1, \dots, X_p .

On souhaite estimer la probabilité inconnue qu'un individu w vérifiant $X_1, \dots, X_p = x$ appartienne au groupe G_1 :

$$p(x) = P(Y = 1 | (X_1, \dots, X_p) = x), \quad x = (x_1, \dots, x_p), \quad (2.5)$$

à l'aide des données.

On appelle transformation logit la fonction :

$$\text{logit}(y) = \ln\left(\frac{y}{1-y}\right) \in \mathbb{R}, \quad y \in]0, 1[\quad (2.6)$$

Son inverse est la fonction :

$$\text{logit}^{-1}(y) = \frac{\exp(y)}{1 + \exp(y)} \in]0, 1[, \quad y \in \mathbb{R} \quad (2.7)$$

Le modèle de régression logistique est caractérisé par : pour tout $i \in \{1, \dots, n\}$,

$(x_{1,i}, \dots, x_{p,i})$ est une réalisation du vecteur aléatoire réel (X_1, \dots, X_p) ,

sachant que $(X_1, \dots, X_p) = (x_{1,i}, \dots, x_{p,i}) = x_i, y_i$ est une réalisation de :

$$Y_i \sim B(p(x_i)), \quad p(x_i) = P(\{Y = 1\} | \{(X_1, \dots, X_p) = x_i\}), \quad (2.8)$$

on a :

$$\text{logit}(p(x)) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, \quad x = (x_1, \dots, x_p), \quad (2.9)$$

où β_0, \dots, β_p sont des coefficients réels inconnus.

Ainsi, $p(x)$ et $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ sont liés par la transformation logit ; on parle de lien logit.

On en déduit l'expression de $p(x)$:

$$p(x) = \text{logit}^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + (\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (2.10)$$

On souhaite estimer β_0, \dots, β_p à partir des données, ce qui amènera à une estimation de $p(x)$ par substitution. Pour ce faire, on utilise la méthode du maximum de vraisemblance.

Estimation

1. Vraisemblance

Soit $\beta = (\beta_0, \dots, \beta_p)$. La vraisemblance associée à (Y_1, \dots, Y_n) est :

$$L(\beta, z) = \prod_{i=1}^n p(x_i)^{z_i} (1 - p(x_i))^{1-z_i}, \quad z = (z_1, \dots, z_n) \in \{0, 1\}^n, \quad (2.11)$$

avec $x_i = (x_{1,i}, \dots, x_{p,i})$.

2. Estimation du maximum de vraisemblance

Les estimateurs du maximum de vraisemblance de β_0, \dots, β_p , notés $\hat{\beta}_0, \dots, \hat{\beta}_p$, vérifient les équations :

$$\sum_{i=1}^n (y_i - \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \dots + \hat{\beta}_p x_{p,i})) = 0 \quad (2.12)$$

et, pour tout $j \in 1, \dots, p$,

$$\sum_{i=1}^n x_{j,i} (y_i - \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \dots + \hat{\beta}_p x_{p,i})) = 0 \quad (2.13)$$

Dans le cas général, il n'y a pas d'expression analytique pour $\hat{\beta}_0, \dots, \hat{\beta}_p$; leurs valeurs peuvent être approchées par l'algorithme IWLS (Iteratively Reweighted Least Squares).

3. Variance et écart type

En posant $\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_p)^t$, on a :

$$\hat{V}(\hat{\beta}) = \left(-\frac{\partial^2}{\partial \beta^2} \ln L(\beta, Y) \right)^{-1} \Big|_{\beta=\hat{\beta}} = (X^t W X)^{-1}, \quad (2.14)$$

où X est la matrice :

$$\begin{pmatrix} 1 & x_{1,1} & \cdots & x_{p,1} \\ 1 & x_{1,2} & \cdots & x_{p,2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{p,n} \end{pmatrix}$$

et

$$W = \text{diag}(\hat{p}(x_1)(1 - \hat{p}(x_1)), \dots, \hat{p}(x_n)(1 - \hat{p}(x_n))). \quad (2.15)$$

Dans la suite, on posera $\hat{\sigma}(\hat{\beta}_j) = \sqrt{\hat{V}(\hat{\beta}_j)}$, racine carrée de la $j+1$ -ème composante du vecteur $\hat{V}(\hat{\beta})$.

4. **Critère AIC** Le critère AIC s'applique aux modèles estimés par une méthode de maximum de vraisemblance : les analyses de variance, les régressions linéaires multiples, les régressions logistiques et de Poisson peuvent rentrer dans ce cadre.

Le critère AIC est défini par :

$$AIC = -2 \log \tilde{L} + 2k \quad (2.16)$$

où \tilde{L} est la vraisemblance maximisée et k le nombre de paramètres dans le modèle.

Avec ce critère, la déviance du modèle $-2 \log(\tilde{L})$ est pénalisée par 2 fois le nombre de paramètres.

Après comparaison de deux modèles, le meilleur celui dont **l'AIC est le plus faible**.

Détection des valeurs anormales

La détection de valeurs anormales dans les données est cruciale car elles peuvent avoir une influence négative dans les estimations et, a fortiori, dans les prévisions (effet levier de la fonction de régression). Les méthodes les plus utilisées sont la méthode des résidus normalisés de Pearson et le critère des distances de Cook. Dans le cas de notre étude, nous allons utiliser le critère des distances de Cook.

117 Critères des distances de Cook

Pour tout $i \in 1, \dots, n$, on définit la distance de Cook de la i -ème observation par :

$$d_i = \frac{[W^{1/2}X(X^tWX)^{-1}X^tW^{1/2}]_{i,i}}{(p+1)(1 - [W^{1/2}X(X^tWX)^{-1}X^tW^{1/2}]_{i,i})} (e_i^*)^2. \quad (2.17)$$

Si $d_i > 1$, on envisage l'anormalité de la i -ème observation. Le seuil $4/(n - (p + 1))$ est parfois utilisé.

Pertinence du modèle :

1. La règle du pouce

Si le modèle de régression logistique est pertinent, la déviance D ne doit pas être trop éloignée de $\mathbb{E}(\hat{D})$, laquelle est proche de $\nu = n - (p + 1)$ (voir égale) :

$$\frac{D}{\nu} \simeq 1 \quad (2.18)$$

Si tel est le cas, le modèle est pertinent.

2. Test de Hosmer-Lemeshow

Pour évaluer la pertinence du modèle de régression logistique, on préconise le test de Hosmer-Lemeshow. La p -valeur associée utilise la loi du Khi-deux : si p -valeur > 0.05 , on admet que le modèle est bien adapté aux données. La statistique du test est :

$$\hat{c} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{(n'_k \bar{\pi}_k (1 - \bar{\pi}_k))} \quad (2.19)$$

où g est le nombre de groupes,

o_k est le nombre d'événements parmi les n'_k combinaisons de facteurs/covariables,

n'_k est le nombre d'essais dans le k^e groupe,

$\bar{\pi}_k$ est la probabilité estimée moyenne pour chaque groupe.

3. Test des résidus de la déviance

Le test des résidus de la déviance est similaire à celui des résidus de Pearson, mais avec les déviances résiduelles :

pour tout $i \in 1, \dots, n$, la i -ème déviance résiduelle est la réalisation de :

$$\hat{D}_i = \text{Sign}(Y_i - \hat{p}(x_i)) \sqrt{2(Y_i \ln(\frac{Y_i}{\hat{p}(x_i)}) + (1 - Y_i) \ln(\frac{1 - Y_i}{1 - \hat{p}(x_i)}))}. \quad (2.20)$$

et la déviance est la réalisation de :

$$\hat{D} = \sum_{i=1}^n \hat{D}_i^2. \quad (2.21)$$

Soient χ_{obs}^2 cette réalisation et $K \sim \chi^2(\nu)$.

Alors la p -valeur associée est :

$$p\text{-valeur} = P(K \geq \chi_{obs}^2). \quad (2.22)$$

4. Test des résidus de Pearson

Le test des résidus de Pearson permet d'évaluer la pertinence du modèle.

On considère les hypothèses :

$$\begin{aligned} H_0 : p(x) &= \text{logit}^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) \\ H_1 : p(x) &\neq \text{logit}^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p). \end{aligned}$$

On calcule :

$$\chi_{obs}^2 = \sum_{i=1}^n (e_i^\circ)^2. \quad (2.23)$$

On considère une $\text{var } K \sim \chi^2(\nu)$.

Alors la p – valeur associée est :

$$p\text{ – valeur} = P(K \geq \chi_{obs}^2). \quad (2.24)$$

Si $p\text{ – valeur} > 0.05$, alors on admet que le modèle est bien adapté aux données.

Significativité de la régression

1. Test de Wald

Soit $j \in 0, \dots, p$. Le test de Wald permet d'évaluer l'influence de X_j sur Y .

On considère les hypothèses :

$$H_0 : \beta_j = 0 \quad \text{contre} \quad H_1 : \beta_j \neq 0.$$

On calcule la réalisation z_{obs} de :

$$Z_* = \frac{\hat{\beta}_j}{\hat{\sigma}(\hat{\beta}_j)}.$$

On considère une

$$\text{var } Z \sim N(0, 1), \quad (2.25)$$

Alors la p – valeur associée est

$$p\text{ – valeur} = \mathbb{P}(|Z| \geq |z_{obs}|). \quad (2.26)$$

Ce test repose sur le fait que l'emv a pour loi asymptotique la loi normale.

2. Test de Déviance

Soit $j \in 0, \dots, p$. Le test de la déviance vise à évaluer l'influence (ou la contribution) de X_j sur Y . La p – valeur associée utilise la loi du Khi-deux.

3. Test de lr

On considère les hypothèses :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

H_1 : il y a au moins un coefficient non nul.

Pour ce faire, on utilise le test du rapport de vraisemblance (lr pour likelihood ratio) (asymptotique). La p – valeur associée utilise la loi du Khi-deux.

Rapports des côtes ou odds ratio

On appelle rapport des côtes (ou odds ratio) de 2 valeurs x_* et x_0 de $X = (X_1, \dots, X_p)$ le réel :

$$RC(x_*, x_0) = \frac{\frac{p(x_*)}{1-p(x_*)}}{\frac{p(x_0)}{1-p(x_0)}} = \frac{p(x_*)(1-p(x_0))}{(1-p(x_*))p(x_0)}. \quad (2.27)$$

Soient $j \in 1, \dots, p$ et $e_j = (0, \dots, 0, 1, 0, \dots, 0)$ (le 1 se situe à la j -ème composante).

Si X_j augmente d'une unité, alors le rapport des côtes est :

$$RC_j = RC(x + e_{j,x}) = \exp(\beta_j), \quad x \in \mathbb{R}^p. \quad (2.28)$$

Par conséquent,

- si $RC_j > 1$, l'augmentation d'une unité de X_j entraîne une augmentation des chances que $Y = 1$ se réalise,
- si $RC_j = 1$, l'augmentation d'une unité de X_j n'a pas d'impact sur Y ,
- si $RC_j < 1$, l'augmentation d'une unité de X_j entraîne une augmentation des chances que $Y = 0$ se réalise.

Un estimateur de RC_j est :

$$\hat{RC}_j = \exp(\hat{\beta}_j). \quad (2.29)$$

Avec l'observation de celui-ci, on peut interpréter l'influence de X_j sur $Y = 1$ en la comparant à 1, comme on l'a fait précédemment.

Qualité du modèle

Pour évaluer la qualité du modèle de régression logistique, on préconise :

1. Le taux d'erreur

On appelle le taux d'erreur (de prédiction) le réel :

$$t = \frac{1}{n} \left(\sum_{i=1}^n 1_{y_i=0 \cap \tilde{y}_i=1} \right) + \left(\sum_{i=1}^n 1_{y_i=1 \cap \tilde{y}_i=0} \right). \quad (2.30)$$

Ce taux est la proportion des modalités prédites qui diffèrent des modalités observées (c'est aussi la somme des 2 valeurs non-diagonales de la matrice de confusion divisée par n).

Plus t est proche de 0, meilleure est la qualité prédictive du modèle.

On convient que la qualité prédictive du modèle est mauvaise lorsque $t > 0.5$.

2. La courbe ROC

Soit $\tau \in [0, 1]$. On appelle i -me prédiction du groupe au niveau τ la réalisation $\tilde{y}_i(\tau)$ de :

$$\hat{Y}_i(\tau) = \begin{cases} 1 & \text{si } \hat{p}(x_i) \geq \tau, \\ 0 & \text{sinon.} \end{cases}$$

À partir de ces prédictions, on définit :

- la fréquence de fausse alarme ("1-specificity") :

$$f f a(\tau) = \frac{\sum_{i=1}^n 1_{y_i=0 \cap \tilde{y}_i=1}}{\sum_{i=1}^n 1_{y_i=0}} \quad (2.31)$$

- la fréquence de bonne détection ("sensitivity") :

$$fbd(\tau) = \frac{\sum_{i=1}^n 1_{y_i=1 \cap \tilde{y}_i=0}}{\sum_{i=1}^n 1_{y_i=1}} \quad (2.32)$$

On appelle courbe ROC, la courbe passant par les points :

$$\{(ffa(\tau), fbd(\tau)); \tau \in [0, 1]\}.$$

Plus la courbe longe les axes $x = 0$ et $y = 1$, meilleur est le modèle. (Autrement dit, plus l'aire sous la courbe ROC est proche de 1, meilleur est le modèle).

Choix du modèle

Dans le cadre de la modélisation, il est important que le modèle choisie explique au mieux notre variable d'intérêt. Le choix du modèle s'est fait suivant méthodes. La première consiste à sélectionner les variables liées à la variable d'intérêt à partir du test de Khi-deux. La seconde méthode est le rééchantillonnage.

1. **Sélection des variables par régression pas à pas** La régression pas à pas est un outil automatisé qui permet, dans les phases exploratoires de l'élaboration d'un modèle, d'identifier un sous-ensemble utile de prédicteurs. A chaque étape, le procédé ajoute la variable la plus significative ou supprime la variable la moins significative.

La régression pas à pas standard ajoute et supprime des prédicteurs selon les besoins à chaque étape.

La sélection ascendante démarre avec un modèle vide et Minitab ajoute le terme le plus significatif à chaque étape

L'élimination descendante démarre avec tous les prédicteurs dans le modèle et Minitab supprime la variable la moins significative à chaque étape

2. **La matrice de Corrélation**

Une matrice de corrélation est un tableau indiquant les coefficients de connexion entre les facteurs. Chaque cellule du tableau indique le lien entre les deux facteurs. Une grille de connexion est utilisée pour présenter des informations, comme contribution à une enquête plus approfondie, et comme indication pour des examens de pointe. Les principaux choix à faire lors de l'établissement d'un réseau de relations comprennent la décision de mesurer la connexion, le codage des facteurs, le traitement des informations manquantes et l'introduction Dans le résultat, les variables sont réordonnées en fonction de la force de la corrélation ce qui permet de voir très rapidement les variables les plus associées

2.7 La différence des points (écart des points)

Pour faire de la prévision, le principe de l'écart des points sera utilisé. Cette pratique répandue aux États Unis consiste à effectuer la différence entre les scores de l'équipe A et de l'équipe B (Nombre de tirs accordés, nombre de 3 points accordés, pourcentage des trois points, score final, nombre de passes décisives...) obtenus pendant le match. En cas de victoire de l'équipe A, la différence des scores sera positives. Dans le cas contraire, elle sera négative. Par exemple, en prenant les Lakers comme

l'équipe A et les Bulls comme l'équipe B, si lors de ce match, le score final est de 130 points pour les Lakers et de 125 pour les Bulls, la différence (Score équipe A - Score équipe B) donnera "+5". La valeur positive de cette différence justifie la victoire des Lakers. Si cette différence avait été négative, cela aurait traduit la défaite des Lakers.

2.8 Prévision

Pour effectuer la prévision, une base de données de 327 matchs a été utilisée. Après avoir repéré les variables qui expliquent le mieux l'issue du match, la base de données, nous avons procédé à la sélection des 163 premiers matchs de la base dans le but d'entraîner notre modèle puis des 163 matchs suivants de la base dans le but de tester le modèle entraîné. Après entraînement du modèle, une matrice de confusion a été générée comportant le nombre de matchs prédit correctement et le nombre de matchs mal prédit. Le modèle sera considéré comme bon lorsque le pourcentage de bonnes prédictions sera supérieur ou égal à 55%. Lorsque la probabilité est supérieure à 0.5, une victoire est prédite pour l'équipe A.

2.9 Limites de l'étude

Durant cette étude, nous avons été confrontés à des difficultés de plusieurs ordres. Il était compliqué de choisir quelles statistiques seraient utilisées dans les bases de données, entre les statistiques totales et les statistiques moyennes des 82 matchs de chaque équipe. En raison de la visible multicollinéarité des données collectées par matchs, les analyses ont été plus difficiles que prévues. Le traitement des données était également l'une des tâches les plus longues et difficiles. Nous n'avons pas été en mesure d'explorer plus d'une méthode de prévisions afin de déterminer la meilleure de toutes. En raison de l'absence des données relatives au basketball béninois nous n'avons pas pu aller plus loin dans nos études en terme de tests des modèles sur la performance et sur la prévision.

3 | Résultats, analyse et interprétations

Le présent chapitre expose les résultats obtenus après le traitement des données et leurs interprétations. Il présente également les recommandations qui découlent de la présente recherche.

3.1 Analyse descriptive

3.1.1 Analyse univariée

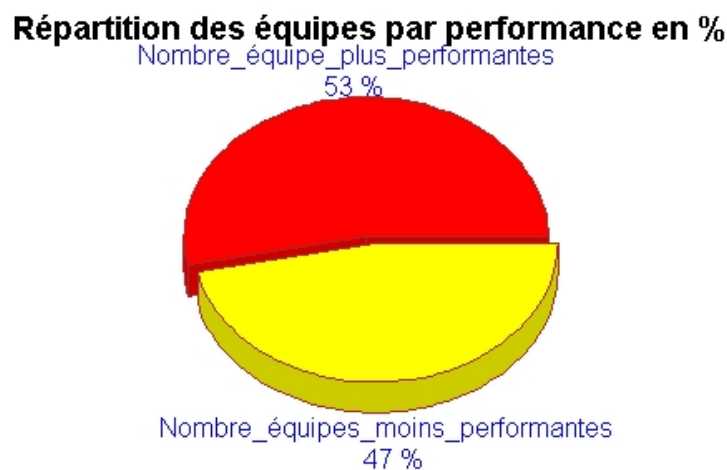


FIGURE 3.1 – Répartition des équipes suivant le niveau performance

Source : Auteur, 2022.

L'analyse de la figure 3.1 révèle que 53% des équipes de notre population sont plus performantes le long de la saison. Autrement dit, 53% des équipes ont gagné 36 matchs.

Répartition des équipes par âge moyen en %

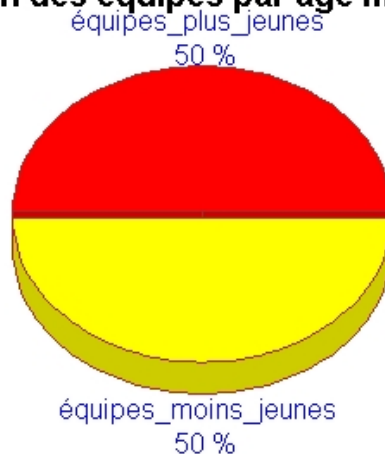


FIGURE 3.2 – Répartition des équipes par âge moyen des joueurs

Source : Auteur, 2022.

La figure 3.2 présente la répartition des équipes suivant l'âge. Les équipes ayant un âge moyen de joueurs supérieurs ou égale à 26 ans sont considérées comme moins jeunes en terme de joueurs. Ainsi, il y a exactement 50% d'équipes constituée de joueurs âgés.

Performante à domicile



FIGURE 3.3 – Répartition des équipes par performances à domicile

Source : Auteur, 2022.

L'analyse de la figure 3.3 révèle que sur l'ensemble des 30 équipes de la NBA, 53% sont plus performantes à domicile. Autrement dit la majorité des équipes gagne plus de matchs en jouant à domicile qu'ailleurs.

FIGURE 3.4 – Répartition des équipes par performance des entraîneurs

Source : Auteur, 2022.

La figure 3.4 révèle que 60% des équipes ont des entraîneurs très performants. Autrement dit dans 60% des équipes, les ont un ratio victoire sur nombre matchs joués supérieur à 0.5

3.1.2 Analyse bivariée

TABLEAU 3.1 – Matrice de corrélation

Variables	Performance équipe	
	<i>r</i>	<i>p-value</i>
PAD	0.17	0.3584
Age	0.62	0.0003
PDC	0.47	0.0095
PMFP	0.33	0.0705
SC	0.41	0.0256
FGP	0.74	0.0229
3PP	0.65	0.0001
2PP	0.44	0.0145
FTP	0.44	0.0146
ORB	-0.12	0.5412
DRB	0.45	0.2659
AST	0.11	0.5607
STL	-0.02	0.9250
TOV	-0.26	0.1643
PF	-0.02	0.8934
PTS	0.48	0.0073

Source : Auteur, 2022.

Les résultats de la matrice de corrélation de la variable expliquée « Performance équipe » avec les variables explicatives indiquent que les variables les plus liées à la variables d'intérêt au seuil des 5% sont les suivantes : Age, PDC , FGP , 3PP et PTS

3.2 Analyse économétrique

3.2.1 Choix des variables explicatives

Après analyse de la matrice de corrélation, une régression pas à pas fut effectuée à partir de l'ensemble des données de la base. à l'issue de cette régression pas à pas, le modèle suivant a été retenu :

$$\text{Modèle A : } \widehat{PE} = -3.536 + 0.013.Age + 0.071.PDC + 2.78.FGP + 1.94.3PP + 0.85.FTP + 0.02.DRB + 0.01.STL$$

(3.1)

3.2.2 Résultat de la régression linéaire multiple et validation du modèle

Une régression linéaire multiple a été réalisée afin de déterminer les facteurs expliquant le niveau de performance des équipes à la fin de la saison NBA. Le tableau suivant présente les résultats finaux de l'estimation après le diagnostic du modèle.

TABLEAU 3.2 – Résultats de l'estimation du modèle

Variables	Coefficients	p-value
	-3.536	8e-05
Age	0.01281	0.2753
PDC	0.07133	0.6920
FGP	2.78508	0.0224
3PP	1.94706	0.1268
FTP	0.85301	0.2473
DRBDRB	0.02626	0.0339
STL	0.01218	0.5307

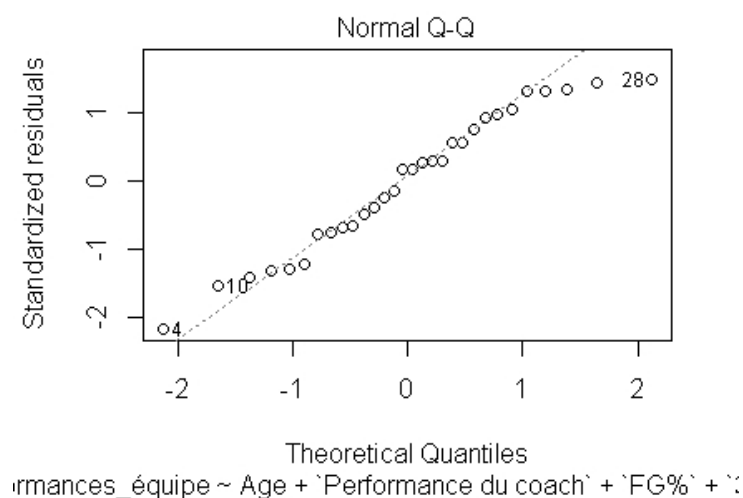
Notes : (ns) Non significatives ; (.) 10% ; (*) 5% ; (**) 1% ; (***) 0,1%

Source : Auteur, 2022.

3.3 Normalité des erreurs

Test de Shapiro-Wilk : $p\text{-value} = 0.6394$; les erreurs suivent une loi normale

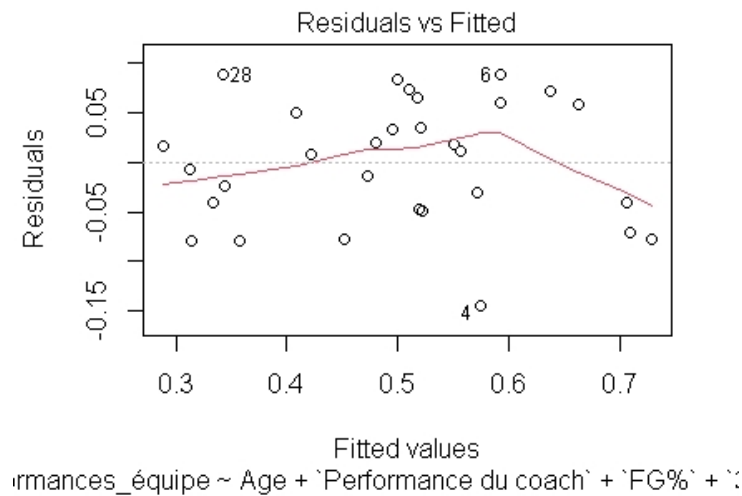
FIGURE 3.5 – Distribution normale des erreurs



3.4 Homoscédasticité

Non Constant Variance Test : $p\text{-value} = 0.96552$; il n'y a pas d'hétéroscédasticité

FIGURE 3.6 – Hétéroscédasticité



3.5 Autocorrélation des erreurs

Durbin Watson Test : = 0.8941 ; il n'y a pas d'autocorrélation des erreurs

3.6 Interprétation des résultats

Selon les résultats de l'estimation du modèle A, pour une augmentation d'un pourcent de "FGP", une équipe aura une performance additionnelle de 2.78508. Autrement dit, pour une augmentation d'un pourcent des tirs, l'équipe aura gagné 2.78508 match de plus sur les matchs totaux de la saison. Cela peut s'expliquer par le fait que le nombre de tirs accordés par une équipe, indépendamment du fait que cela soit des 3 points ou deux points augmente la chance de l'équipe d'avoir un score lui permettant de remporter des matchs; cela traduit la précision des joueurs des équipes. Pour une augmentation d'une unité de "age" une équipe aura une performance additionnelle de 0.01281. Autrement dit , pour une augmentation d'une unité de l'âge moyen des joueurs, l'équipe gagnera 0.01281 matchs de plus sur les matchs totaux de la saison. Cela vient montrer que l'âge influence la performance d'une équipe bien que cela soit faiblement. Également, une augmentation d'une unité de "PDC" entraînerait une augmentation de 0.07 de la performance de l'équipe. Une augmentation d'une unité des performances de l'entraîneur entraînerait un gain de 0.07 matchs sur le nombre total de matchs joués. Cela montre en un sens que les facultés des joueurs ont plus d'impact sur le déroulement d'un match que la performance des entraîneurs. Cela voudrait aussi dire que les tactiques des entraîneurs ne sont pas si différentes les unes des autres et que par conséquent, ce sont aux capacités des joueurs de faire la différence. En outre, une augmentation d'un pourcent des 3 points entraînera un gain additionnel de 1.94706 matchs. Les 3 points étant le plus grand nombre de point que les joueurs puisse mettre en une fois, ils apparaissent comme déterminant dans la victoire des équipes. En examinant de plus près ces résultats, on peut dire que les principaux facteurs du niveau de performance d'une équipe sont l'âge, les pourcentages des lancers francs, 3 points, de tirs, les nombres moyen de vole de balles et des rebonds défensifs et enfin la performance de l'entraîneur.

3.6.1 Résultat de la régression logistique

3.6.2 Choix des variables explicatives

Nous avons procédé à deux estimations pour donner la meilleure prédiction que possible sur la victoire des équipes lors d'un match. Après avoir effectué une régression pas à pas, nous avons réalisé une matrice de corrélation et retiré les variables qui présentait une forte liaison entre elle. Le modèle 1 estimé est un logit et le modèle 2 un probit.

TABLEAU 3.3 – Résultats des tests du choix de modèle

	Resid. Dev	AIC
Modèle 1	161.53	173.53
Modèle 2	162.70	174.7

Source : Auteur, 2022.

En comparant l'AIC de chaque modèle, on remarque que le modèle 1 a l'AIC le plus faible (173.53) que le modèle 2 (174.7) Ainsi, le meilleur modèle retenu pour la prédiction de notre variable d'intérêt est le modèle 1.

3.6.3 Résultat de la régression logistique

Une régression logistique binaire multivariée a été réalisée afin de déterminer l'influence des variables indépendantes sur la variable dépendante "ISSUES". Le tableau suivant présente les résultats finals de l'estimation après le diagnostic du modèle.

TABLEAU 3.4 – Résultats de l'estimation du logit

Variables	Coefficients	Odds Ratio	p-value
DFGA	0.25	1.29032	1.44e-11***
D3PP	0.21	1.24024	1.18e-12***
DFTM	0.27	1.31644	1.03e-09***
DBLK	0.37	1.44861	2.45e-07***
DDREB	0.30	1.34796	3.05e-12***

Notes : (ns) Non significatives ; (.) 10% ; (*) 5% ; (**) 1% ; (***) 0,1%

Source : Auteur, 2022.

A la lecture de ce tableau, on constate que toutes les variables sont significativement différentes de 0 au seuils de 0.1%. Ce qui atteste de la significativité du modèle.

3.7 Pertinence du modèle

On remarque que la valeur D/v n'est pas très proche de 1. Par contre, les probabilités de Khi-deux des tests d'Hosmer-Lemeshow, des résidus de Pearson et des résidus de la déviance sont toutes supérieures au seuil de 5%, donc le modèle est bien adapté aux données. Ainsi, le modèle est pertinent.

TABLEAU 3.5 – Résultat de la pertinence du modèle

Tests	Critère de significativité
Règle de pouce	D/v = 0.5032021
Test de Hosmer-Lemeshow	Pr(Chi2)= 0.976
Test de résidus de Pearson	Pr(Chi2)=0.7167697
Test de résidus de la déviance	Pr(Chi2)= 1

Source : Auteur, 2022.

3.8 Qualité du modèle

TABLEAU 3.6 – Qualité du modèle

Tests	Valeurs
Taux d'erreur	0.0379
Aire sous la courbe ROC	0.9621

Source : Auteur, 2022.

Le taux d'erreur du modèle est inférieur à 0,5 ; donc le modèle est de bonne qualité. De plus, l'aire sous la courbe ROC (0,90) est supérieur à 0,80 ; le modèle est donc d'une bonne qualité prédictive.

3.9 Interprétation des résultats

Selon les résultats de l'estimation du modèle B, lorsque l'équipe A augmentera d'une unité ses tentatives de tirs par rapport à l'équipe B , elle aura 1,29 fois plus de chances de mettre plus de points que l'équipe B .Autrement dit, dans le cas où l'équipe A tente un tir de plus, elle aura plus de chance de voir son score augmenter et par la même occasion réduira l'écart des points avec l'équipe B, s'il était négatif, ou augmentera l'écart des points avec l'équipe B si cette différence était positive. Cela s'explique par le fait qu'à partir de 45% au tir (toutes zones confondues), on peut commencer à parler d'un bon shooteur (source : Trashtalk, 2021) et que le pourcentage de tir réussis de la grande majorité des équipes est supérieur à 45%, d'où à chaque tentatives de tir, les joueurs des équipes ont plus de chance de mettre que de rater. Une augmentation d'un pourcent des tirs à 3 points de l'équipe Lorsque l'équipe A augmentera d'un pourcent ses tirs à 3 points, elle aura 1,24 fois plus de chance de mettre plus de points que l'équipe B. Autrement dit, plus le ratio 3 points accordé - 3 points tentés sera grand , plus les 3 points qui ont effectivement été accordés seront grands. Les 3 points étant le plus grand nombre de points qu'un joueur puisse mettre en une fois, cela explique totalement le fait que l'équipe A puisse mettre plus de points que l'équipe B. Une augmentation d'un point des lancers francs accordés de l'équipe A par rapport à l'équipe B offrira 1,32 fois plus de chances à l'équipe A de mettre plus de points que l'équipe B. Cela s'explique par le fait que les lancers francs étant des tirs libres offerts aux joueurs à la suite d'une faute sur tir, d'une faute technique ou d'une faute après les 5 fautes autorisées par équipes, les joueurs des équipes offrent un ou deux points de plus à leurs équipes. Une augmentation d'une unité des blocs de l'équipe A par rapport à l'équipe B offrira 1,34 fois plus de chances à l'équipe A d'avoir plus de points que l'équipe B. Cela s'explique par le

fait que lorsqu'un joueur bloc une tentative de tir de l'équipe adverse, il réduit la possibilité de cette équipe à mettre plus de points que son équipe. Une augmentation d'un point des rebonds défensifs de l'équipe A par rapport à l'équipe B offrira 1,45 fois plus de chances à l'équipe A de mettre plus de points que l'équipe B. Ainsi, sachant qu'un rebond défensif survient à la suite d'une tentative ratée de tir, cette récupération de balle par l'équipe à la défense empêche non seulement l'équipe tireuse de réessayer un tir mais aussi offre une possession de balle à l'équipe en défense et de ce fait, fait passer cette équipe en attaque pour de nouvelles tentatives de tirs.

En examinant de plus près ces résultats, on peut dire que les principaux facteurs de l'issue d'un match sont l'écart des tentatives de tirs, l'écart du pourcentage des 3 points, l'écart des lancers francs accordés, l'écart des arrêts de balles, l'écart des rebonds défensifs. Malgré l'importance que la littérature accorde aux fautes personnelles et aux contres attaque, ces variables ne se sont pas révélées associées à l'issue du match dans notre étude.

3.10 Prévision

TABLEAU 3.7 – Matrice de confusion

	Défaites	Victoires	Total
Défaites	81	7	88
Victoires	11	64	75
Total	92	71	163

Source : Auteur, 2022.

L'analyse de cette matrice de confusion nous montre que le modèle prédit correctement 88,96 % des matchs. Le modèle est bon et il est donc possible de prédire l'issue des matchs à partir des variables retenues.

$$\text{Modèle B : } \widehat{p(ISSUE)} = \text{logit}^{-1}(0.25DFGA + 0.21D3PP + 0.27DFTM + 0.37DBLK + 0.30DDREB) \quad (3.2)$$

3.11 Vérification des hypothèses et recommandations

3.11.1 Vérification des hypothèses

Au regard toute cette synthèse, il ressort que les deux hypothèses de notre étude sont validées.

3.11.2 Préconisations opérationnelles

Pour la constitution ou l'amélioration des différentes équipes de basketball au Bénin, nous inviterons les clubs professionnels Béninois à :

- relever dans un premier temps les différentes statistiques personnelles des joueurs ;
- tenir des feuilles de statistiques régulièrement et compiler de manière informatisée les différentes données relatives à chaque match ;
- faire utiliser le modèle généré pour mieux observer les performances et se catégoriser de ce fait ;
- organiser des compétitions en vue de détecter de potentiels joueurs qui pourront être utiles ;
- faire des matchs d'entraînement avec plusieurs équipes et récolter les données relatives à ces équipes ;
- utiliser le modèle de prévision généré pour simuler mathématiquement les matchs ;
- se servir des simulations pour mieux préparer les matchs.

Conclusion

La victoire lors d'un championnat et la capacité à mettre en œuvre des stratégies gagnantes sont de réels enjeux dans tous les sports notamment au basketball . L'étude s'est particulièrement intéressée aux statistiques des équipes de la saison 2020-2021. Afin de trouver les déterminants des performances sportives d'une équipe de basketball, nous avons réalisé une régression linéaire multiple à l'aide du logiciel R. A l'issue de ces résultats, il ressort que les performances de l'entraîneur, l'âge moyen des équipes et le pourcentage de tirs expliquent le nombre de victoire d'une équipe sur le nombre total de matchs joués. Afin de trouver les déterminants d'un match de basketball, nous avons réalisé une régression logistique binaire à l'aide du logiciel R. A l'issue de ces résultats, il ressort que le pourcentage des 3 points accordé, de lancés francs et de rebonds défensifs influencent l'issue d'un match. Au terme cette présentation consacrée aux déterminants des performances sportives d'une équipe de basket ball et prévision de l'issue des matchs : cas de la NBA, il ressort qu'il est important de baser la constitution des équipes sur les points stratégiques tels que les capacités des joueurs à lancé des 3 points ou à arrêter des balles.

La prise en compte des résultats de cette recherche devrait permettre à la FBBB et aux clubs professionnels et aux autres associations de basketball de relever le défi et d'améliorer les conditions d'entraînement des équipes de basketball. L'amélioration des conditions d'entraînement des équipes de basketball seront un atout majeur dans l'évolution de cette discipline sportive au Bénin, et renforcera la confiance des investisseurs et sponsors, notamment celle de MTN. Au terme de cette étude, il serait intéressant de connaître de manière plus spécifique les facteurs explicatifs de la faible évolution des jeunes béninois dans le domaine du basketball international.

Bibliographie

- [1] Site Officiel de la NBA. « Box Score ». *www.NBA.com*.
- [2] Site Officiel de Basketreference. « Saison 2020-2021 ». *www.Basketreference.com*.
- [3] Petar Vračar, Erik Štrumbelj, Igor Kononenko. « Modelling Basketball Play-by-Play Data ». *Expert System With Application*
- [4] Nadeau, L. (2001). « La validation d'un outil de mesure de la performance au hockey sur glace en situation réelle de match ». *Thèse de doctorat, Université Laval, Québec, 88 Canada*
- [5] Collins, R.D., Hodges, P.B. (1978). « A comprehensive guide to sports skills tests and measurement ». *Springfield, Il. : C. C. Thomas*.
- [6] Cardinal, C.H., Chouinard, R., Roy, M. (2006). « Développement à long terme du participant et de l'athlète : Matière à réflexion ». . *Le sommet des entraîneurs, Montréal, Québec*.
- [7] Garganta, J., Maia, J., Basto, F. (1997). « Analysis of goal-scoring patterns in European top level soccer teams ». *Science and football III, 246-250*.
- [8] Gréhaigne, J.-F. (2009). « Autour du temps : Apprentissages, espaces, projets dans les sports collectifs ». *Toulouse, France : Presses universitaires de Franche-Comté*.
- [9] Nadeau, L., Richard, J.-F., Godbout, P. (2008). « The validity and reliability of a performance assessment procedure in ice hockey ». *Physical Education and Sport Pedagogy, 13 (1), 65-83*
- [10] Dodjivi Ahlinvi J.-C. A. (2019). « Mesure de la performance sportive au basketball compétitif à l'aide d'une adaptation du "Team Sport Assessment Procedure" » *https://corpus.ulaval.ca*.
- [11] Erik Štrumbelj, Petar Vračar (2012). « Simulating a basketball match with a homogeneous Markov model and forecasting the outcome ». *International Journal of Forecasting*
- [12] Feifang Hu and James V. Zidek (2014). « Forecasting NBA Basketball Playoff Outcomes Using the Weighted Likelihood. » *Institute of Mathematical Statistics*.

- [13] Wei-Jen Chen , Mao-Jhen Jhou, Tian-Shyug Lee and Chi-Jie Lu (2012). « Hybrid Basketball Game Outcome Prediction Model by Integrating Data Mining Methods for the National Basketball Association ». *Entropy* 2021, 23, 477.
- [14] Hans Manner (2015). « Modeling and forecasting the outcomes of NBA Basketball games ». *Institute of Econometrics and Statistics, University of Cologne*.
- [15] Walid Klibi, Mohamed Zied Babai, Yves Ducq, Haytham Omar Abd El Akher (2021).« Basket data-driven approach for omnichannel demand forecasting » . *hal-03195611*.
- [16] Eric Jones. (2016). « PREDICTING OUTCOMES OF NBA BASKETBALL GAMES ». *A Thesis Submitted to the Graduate Faculty of the North Dakota State University of Agriculture and Applied Science*.
- [17] Fadi Thabtah · Li Zhang · Neda Abdelhamid (2019). « NBA Game Result Prediction Using Feature Analysis and Machine Learning ». *Annals of Data Science* (2019) 6(1) :103–116.
- [18] Magel, R. Unruh, S. (2013). « Determining Factors Influencing the Outcome of College Basketball Games ». *Open Journal of Statistics*. 3.4 (August, 2013)

Annexe

Cette section contient les sorties brutes de nos analyses sous le logiciel R ainsi que d'autres résultats de notre travail.

Annexe 1 : Matrice de corrélation et sélection des variables

```
> Data_model_pg <- data.frame( Performances_équipe,
+                               Age, `Performance du coach`, `FG%`
+                               , `3P%` , `FT%` , ORB , STL, DRB)
> cor(Data_model_pg)
```

	Performances_équipe	Age	Performance du coach	FG.	X3P.
Performances_équipe	1.00000000	0.6489032	0.6140700	0.74098177	0.78347147
Age	0.64890322	1.0000000	0.4876396	0.61564374	0.53537746
Performance du coach	0.61407005	0.4876396	1.0000000	0.43025892	0.57681114
FG.	0.74098177	0.6156437	0.4302589	1.00000000	0.62970162
X3P.	0.78347147	0.5353775	0.5768111	0.62970162	1.00000000
FT.	0.57033181	0.3903412	0.5719689	0.36091402	0.67180711
ORB	-0.08343316	-0.4797794	-0.2766341	-0.29965044	-0.08028054
STL	-0.07251072	-0.0789897	-0.1029013	0.05992778	-0.12055011
DRB	0.48517646	0.2601222	0.3868832	0.22531889	0.35105019

	FT.	ORB	STL	DRB
Performances_équipe	0.57033181	-0.08343316	-0.07251072	0.48517646
Age	0.39034116	-0.47977936	-0.07898970	0.26012219
Performance du coach	0.57196895	-0.27663405	-0.10290130	0.38688317
FG.	0.36091402	-0.29965044	0.05992778	0.22531889
X3P.	0.67180711	-0.08028054	-0.12055011	0.35105019
FT.	1.00000000	-0.17484831	-0.20731938	0.09618318
ORB	-0.17484831	1.00000000	0.09539168	0.15581768
STL	-0.20731938	0.09539168	1.00000000	-0.28538030
DRB	0.09618318	0.15581768	-0.28538030	1.00000000

```
> step(Model_pg1 , direction = "both", k=2)
Start: AIC=-144.71
Performances_équipe ~ `Performance à domicile` + `Performance du coach` +
  `FG%` + `3P%` + `FT%` + ORB + STL + TOV + PF + PTS
```

	Df	Sum of Sq	RSS	AIC
- PTS	1	0.0000183	0.11582	-146.71
- `FT%`	1	0.0001609	0.11596	-146.67
- STL	1	0.0008733	0.11668	-146.49
- `Performance à domicile`	1	0.0009705	0.11677	-146.46
- ORB	1	0.0021204	0.11792	-146.17
- TOV	1	0.0028105	0.11861	-145.99
<none>			0.11580	-144.71
- PF	1	0.0100800	0.12588	-144.21
- `Performance du coach`	1	0.0110699	0.12687	-143.97
- `3P%`	1	0.0168464	0.13265	-142.64
- `FG%`	1	0.0303114	0.14611	-139.74

Annexe 2 : Validation globale du modèle

```
> library(gvlma)
> gvlma_mod_pg <- gvlma(Model_pg1)
> summary(gvlma_mod_pg)

Call:
lm(formula = Performances_équipe ~ `Performance à domicile` +
  `Performance du coach` + `FG%` + `3P%` + `FT%` + ORB + STL +
  TOV + PF + PTS, data = stat_per_game)

Residuals:
    Min       1Q   Median       3Q      Max
-0.114482 -0.050770 -0.000821  0.047159  0.110816

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.1977400   0.8764588   -2.508   0.0214 *
`Performance à domicile`  0.1224885   0.3069575    0.399   0.6943
`Performance du coach`    0.2650092   0.1966401    1.348   0.1936
`FG%`           3.8954724   1.7467896    2.230   0.0380 *
`3P%`           2.4655451   1.4830065    1.663   0.1128
`FT%`           0.1472082   0.9060007    0.162   0.8726
ORB             0.0140484   0.0238176    0.590   0.5623
STL             0.0085720   0.0226455    0.379   0.7092
TOV            -0.0129331   0.0190457   -0.679   0.5053
PF             -0.0207094   0.0161035   -1.286   0.2139
PTS            0.0003704   0.0067555    0.055   0.9568
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07807 on 19 degrees of freedom
Multiple R-squared:  0.7936,    Adjusted R-squared:  0.6849
F-statistic: 7.304 on 10 and 19 DF,  p-value: 0.0001152
```

```
ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of significance = 0.05

Call:
gvlma(x = Model_pg1)

      value p-value      Decision
Global Stat    4.49914  0.3426 Assumptions acceptable.
Skewness       0.02309  0.8792 Assumptions acceptable.
Kurtosis       1.08479  0.2976 Assumptions acceptable.
Link Function   2.00428  0.1569 Assumptions acceptable.
Heteroscedasticity 1.38697  0.2389 Assumptions acceptable.
>
```

Annexe 3 : Test d'hypothèses

Normalité des erreurs

```
> res_pg1 <- Model_pg1$residuals
> shapiro.test(res_pg1)

      Shapiro-Wilk normality test

data:  res_pg1
W = 0.97306, p-value = 0.6258
```

Hétéroscédasticité

```
> library(carData)
> library(car)
> ncvTest(Model_pg1)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.8979356, Df = 1, p = 0.34334
```

Autocorrélation

```
> library(lmtest)
> dwtest(Model_pg1)

      Durbin-Watson test

data:  Model_pg1
DW = 2.2326, p-value = 0.7478
alternative hypothesis: true autocorrelation is greater than 0
```

Multicolinéarité

```
> library(carData)
> library(car)
> vif(Model_pg1)
`Performance à domicile`  `Performance du coach`  `FG%`  `3P%`
1.718373                2.003519            4.174965      3.647940
`FT%`                    ORB                    STL      TOV
3.154491                1.677346            1.391208      2.366707
PF                      PTS
1.758645                4.137313
```

Annexe 4 : Régression logistique

Choix du modèle

```
> forecast_probit$aic
[1] 174.7015
> forecast_logit_1$aic
[1] 173.5279
```

Sélection variables logit

```
> step(forecast_logit_1, direction = "both", K=2)
Start: AIC=173.53
ISSUE ~ DFGA + `D3P%` + DFTM + DDREB + DBLK

      Df Deviance   AIC
<none>    161.53 173.53
- DBLK    1   199.13 209.13
- DFTM    1   220.10 230.10
- DFGA    1   239.22 249.22
- DDREB    1   251.77 261.77
- `D3P%`  1   260.96 270.96

Call: glm(formula = ISSUE ~ DFGA + `D3P%` + DFTM + DDREB + DBLK, family = binomial(logit),
  data = Echantillon_match_327)

Coefficients:
(Intercept)          DFGA          `D3P%`          DFTM          DDREB          DBLK
  0.07371      0.25489      0.21531      0.27493      0.29859      0.37060

Degrees of Freedom: 326 Total (i.e. Null); 321 Residual
Null Deviance: 452.4
Residual Deviance: 161.5      AIC: 173.5
```

Odds.ratio

```
> library(questionr)
> odds.ratio(forecast_logit_1)
waiting for profiling to be done...
      OR    2.5 % 97.5 %      p
(Intercept) 1.07649 0.72279 1.6143 0.7174
DFGA        1.29032 1.20445 1.3976 1.443e-11 ***
`D3P%`      1.24024 1.17409 1.3231 1.184e-12 ***
DFTM        1.31644 1.21232 1.4478 1.030e-09 ***
DDREB       1.34796 1.24754 1.4770 3.054e-12 ***
DBLK        1.44861 1.27049 1.6863 2.450e-07 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Annexe 5 : Tests de pertinence

▷ Règle de pouce

```
> library(stats)
> deviance(forecast_logit_1)/df.residual(forecast_logit_1) #0.4287509 0.5032021
[1] 0.5032021
```

▷ Hosmer-Lemeshow

```
> library(performance)

Attaching package: 'performance'

The following object is masked from 'package:ROCR':

    performance

> performance_hosmer (forecast_logit_1) #p_value=0.823 0.976
# Hosmer-Lemeshow Goodness-of-Fit Test

Chi-squared: 2.157
  df: 8
  p-value: 0.976

Summary: model seems to fit well.
```

▷ Tests de Pearson et test de déviance

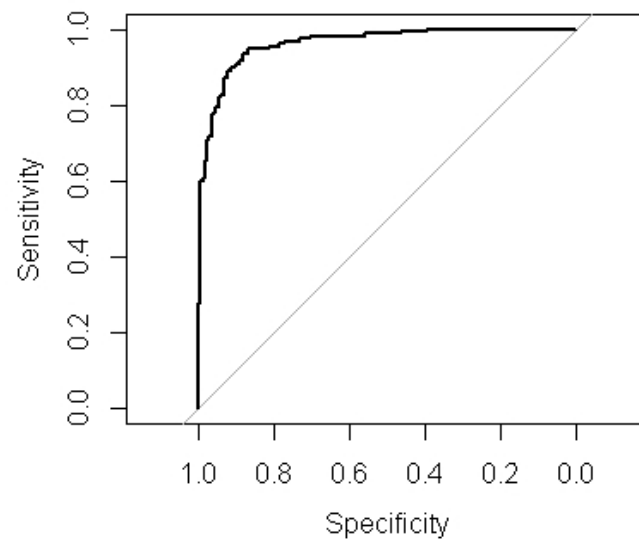
```
> pvaleur1=1-pchisq(deviance(forecast_logit_1),df.residual(forecast_logit_1))
> pvaleur1 #1
[1] 1
```

Annexe 6 : Qualité du modèle

▷ Area Under the Curve

```
> library(pROC)
> library(ROCR)
> prediction <- predict(forecast_logit_1, data = predictions, type="response")
> roc_object <- roc(Echantillon_match_327$ISSUE, prediction)
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> plot(roc_object)
> auc(roc_object) #0.9738 0.9621
Area under the curve: 0.9621
```

▷ Courbe ROC



Annexe 7 : Prévision

Découpage de la base en 2, entraînement du modèle et matrice de confusion

```
> library(dplyr)
> test_ <- slice(Echantillon_match_327, 1:163)
> train_ <- slice (Echantillon_match_327, 163:326)
> library(caret)
> library(InformationValue)
> library(ISLR)
> predicted <- predict(forecast_logit_1, test_ , type="response")
> #find optimal cutoff probability to use to maximize accuracy
> optimal <- optimalCutoff(test_$ISSUE, predicted)[1]
> #create confusion matrix
> confusionMatrix(test_$ISSUE, predicted)
  0  1
0 81  7
1 11 64
```

Specifity, Sensibility

```
> sensitivity(test_$ISSUE, predicted)
[1] 0.9014085
> #calculate specificity
> specificity(test_$ISSUE, predicted)
[1] 0.8804348
> #calculate total misclassification error rate
> misClassError(test_$ISSUE, predicted, threshold=optimal)
[1] 0.0982
```

Table des matières

Approbation	ii
Dédicace	iii
Remerciements	iv
Avant propos	v
Liste des sigles et abréviations	vi
Résumé	x
Abstract	xi
Introduction	1
1 Cadre institutionnel de l'étude	2
1.1 Présentation de MTN-Bénin	2
1.1.1 Historique, Mission et Objectifs, Structure	2
1.1.2 Structure organisationnelle et environnementale de MTN Bénin	3
1.1.3 Le micro environnement de MTN Bénin	4
1.1.4 Le macro environnement de MTN Bénin	4
1.2 Déroulement du stage et observation	5
1.2.1 Déroulement du stage	5
1.2.2 Observation	6
2 Cadre théorique et méthodologique de l'étude	7
2.1 Enoncé du problème et intérêt de l'étude	7
2.1.1 Enoncé du problème	7
2.1.2 Intérêt de l'Etude	8
2.2 Objectifs et hypothèses de la recherche	8
2.2.1 Objectifs de la recherche	8
2.2.2 Hypothèses de la recherche	9
2.3 Revue de littérature	9
2.3.1 Clarification conceptuelle	9
2.3.2 Revue empirique	11
2.4 Méthodologie de recherche	13
2.5 Méthodologie	13
2.5.1 Choix et présentation des variables	13
2.6 Analyse descriptive et économétrique des variables	14

2.6.1	Analyse descriptive	14
2.6.2	Modèle de régression	14
2.7	La différence des points (écart des points)	21
2.8	Prévision	22
2.9	Limites de l'étude	22
3	Résultats, analyse et interprétations	23
3.1	Analyse descriptive	23
3.1.1	Analyse univariée	23
3.1.2	Analyse bivariée	26
3.2	Analyse économétrique	26
3.2.1	Choix des variables explicatives	26
3.2.2	Résultat de la régression linéaire multiple et validation du modèle	27
3.3	Normalité des erreurs	27
3.4	Homoscédasticité	27
3.5	Autocorrélation des erreurs	28
3.6	Interprétation des résultats	28
3.6.1	Résultat de la régression logistique	29
3.6.2	Choix des variables explicatives	29
3.6.3	Résultat de la régression logistique	29
3.7	Pertinence du modèle	29
3.8	Qualité du modèle	30
3.9	Interprétation des résultats	30
3.10	Prévision	31
3.11	Vérification des hypothèses et recommandations	31
3.11.1	Vérification des hypothèses	31
3.11.2	Préconisations opérationnelles	31
	Conclusion	33
	Références	33
	Annexe	a