

# Déterminants des performances d'une équipe de basketball et prévision de l'issue des matchs ; cas de la NBA

**Elimane Yassine SEIDOU**



**Université d'Abomey-Calavi**

École Nationale d'Économie Appliquée et  
de Management



Session de Juillet 2022  
Directeur de mémoire :

**Dr M. Nicodème ATCHADE**

Maître Assistant CAMES

# Déterminants des performances d'une équipe de basketball et prévision de l'issue des matchs ; cas de la NBA

**Elimane Yassine SEÏDOU**

**Université d'Abomey-Calavi**

École Nationale d'Économie Appliquée et  
de Management

Session de Juillet 2022

# PLAN

- 1 Problématique**
- 2 Objectifs et hypothèses**
- 3 Méthodologie**
- 4 Résultats et discussions**
- 5 Préconisations opérationnelles**

# Problématique(1/2)

- ♣ L'industrie des sports est l'une des plus répandues dans le monde et en Afrique. Ils sont souvent marqués par des championnats ou des tournois. Aussi divertissant que soient les sports , 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 coachs, 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 ;

## Problématique(2/2)

- ♣ Dans le but d'optimiser les prestations sportives, des prévisions , purement basées sur les chiffres ont été effectuée afin d'ajuster les stratégies des coachs.
- ♣ Dans le cadre du développement du basket ball béninois la NBA et le Bénin se sont engagés dans une collaboration.  
Malheureusement, le basket ball béninois connaît un déficit en terme de simulations chiffrée de matchs, ce qui est un réel retard au vue des techniques utilisées sur la scène internationale

**Quelles sont les performances d'une équipe sportive de basketball et comment prévoir l'issue des matchs de la NBA ?**

# Objectifs

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.

## Objectif n° 1

Déterminer les facteurs qui expliquent le niveau de performance d'une équipe sur 72 matchs.

## Objectif n° 2

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.

# Hypothèses

## Hypothèse n° 1

La performances du coach, l'âge moyen des joueurs et le pourcentage de tirs accordés expliquent la performance d'une équipe sur 72 matchs.

## Hypothèse n° 2

Le pourcentage de 3 points accordés, de lancés francs et le nombre de rebonds défensifs influencent l'issue d'un matchs.

# Méthodologie (1/6)

## Données

Cette étude se fonde sur les données relatives aux statistiques des 30 équipes , collectées pendant les matchs de la saison NBA 2020-2021.

Les données ont été collectées sur le site officiel de la NBA :

<https://www.nba.com>

# Méthodologie (2/6)

## \* Liste des variables

La variable dépendante est "Performance de l'équipe".

Variables	Libellés
Performance de l'équipe (PE)	Nombre de victoire sur nombre total de matchs
PMFP	Points moyens du Franchise Player
Performance à domicile (PAD)	Nbre de victoire sur nbre de matchs joués à domicile
PDC	Performance du coach
SC	Salary Cap
Age	Age moyen des joueurs de l'équipe
FGP	Pourcentage de tirs réussis
3PP	Pourcentage de trois points réussis
FTP	Pourcentage de lancers francs réussis
DRB	Nombre moyen de rebonds défensifs
STL	Nombre moyen de vole balles

## Méthodologie (3/6)

La variable dépendante est "l'issue du match".

$$ISSUE = \begin{cases} 1 & \text{si l'équipe A gagne} \\ 0 & \text{sinon} \end{cases}$$

Variables	Libellés
DFGA	Différence (équipe A - équipe B) des tentatives de tirs
D3PP	Différence du pourcentage de trois points
DFTP	Différence du pourcentage de lancés francs
DFGP	Différence du Pourcentage de tirs réussis
DORB	Différence du nombre de rebonds offensifs
DDRB	Différence du nombre de rebonds défensifs
DAST	Différence du nombre de passes décisives
DSTL	Différence du nombre de vole de balle
DTOV	Différence du nombre de contre attaques

# Méthodologie (4/6)

## Analyse descriptive univariée

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.

## Analyse descriptive bivariée

L'analyse bivariée a pour but d'identifier les liaisons entre la variable performance de l'équipe et les différentes variables explicatives et d'apprécier la significativité statistique de ces liaisons.

# Méthodologie (5/6)

## Le modèle

Le modèle de l'étude relative aux déterminants des performances sportives d'une équipe de basketball, nous avons donc :

$$\widehat{PE} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (1)$$

## Choix du modèle RLM

- ♣ - utiliser les variables liées à la variable d'intérêt à partir de la matrice de corrélation ;
- ♣ - choisir le meilleur modèle en comparant les critères AIC.

## Choix du modèle logit

- ♣ - utiliser les variables liées à la variable d'intérêt à partir du test de Khi-deux ;
- ♣ - choisir le meilleur modèle en comparant les critères AIC.

# Méthodologie (6/6)

## Pertinence du modèle

- ♣ la règle de pouce ;
- ♣ le test de Hosmer-Lemeshow ;
- ♣ le test des résidus de la déviance ;
- ♣ le test des résidus de Pearson.

## Validation du modèle RLM

- ♣ Normalité des erreurs
- ♣ Autocorrélation des erreurs
- ♣ Hétéroscédasticité
- ♣ Facteur d'inflation de la variance

## Qualité du modèle

- ♣ le taux d'erreur ;
- ♣ la courbe ROC ;

# Résultats et discussions (1/9)

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

Figure 1 – Matrice de corrélation

## Résultats et discussions (2/9)

Comme annoncé dans la méthodologie, nous avons choisi les variables à partir d'une matrice de corrélation mais également en automatisant une régression pas à pas nous permettant d'obtenir le modèle RLM avec l'AIC le plus petit. Le modèle suivant est celui retenu après estimations :

$$\text{Modèle A : } \widehat{PE} = -3.536 + 0.013.\text{Age} + 0.071.\text{PDC} + 2.78.\text{FGP} \\ + 1.94.\text{3PP} + 0.85.\text{FTP} + 0.02.\text{DRB} + 0.01.\text{STL}$$

(2)



# Résultats et discussions (3/9)

Table 1 – 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
DRB	0.02626	0.0339
STL	0.01218	0.5307

## Résultats et discussions (4/9)

Table 2 – Validation du modèle de régression linéaire multiple

Tests	Critère de significativité
Normalité des erreurs	$p\text{-value}=0.2038$
Hétéroscédasticité	$p\text{-value}=0.44002$
Autocorrélation	$p\text{-value}=0.4688$
Validation Globale	$p\text{-value}=0.3177$

## Résultats et discussions (5/9)

Nous avons procédé à deux estimations pour donner la meilleure prédition 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 elles. Le modèle 1 estimé est un logit et le modèle 2 un probit. :

Table 3 – Résultats des tests du choix de modèle

	Resid. Dev	AIC
Modèle logit	161.53	173.53
Modèle probit	162.70	174.7

Source : Auteur, 2022.

## Résultats et discussions (6/9)

Variables	Coefficients	Odds Ratio	<i>p-value</i>
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***

Figure 2 – Estimation du logit

## Résultats et discussions (7/9)

Table 4 – Résultat de la pertinence du modèle

Tests	Critère de significativité
Règle de pouce	$D/v = 0.5032021$
Test de Hosmer-Lemeshow	$Pr(Chi^2) = 0.976$
Test de résidus de Pearson	$Pr(Chi^2) = 0.7167697$
Test de résidus de la déviance	$Pr(Chi^2) = 1$

Table 5 – Qualité du modèle

Tests	Valeurs
Taux d'erreur	0,1
Aire sous la courbe ROC	0.9621

## Résultats et discussions (8/9)

Après entraînement du modèle logit, il fut testé sur 100 matchs et obtint les résultats suivants :

$$\widehat{p(ISSUE)} = \text{logit}^{-1}(0.25DFGA + 0.21D3PP + 0.27DFTM + 0.37DBLK + 0.30DDREB)(3)$$

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

Figure 3 – Matrice de confusion

## Résultats et discussions (9/9)

**L'hypothèse 1** selon laquelle "la performance du coach", l'âge moyen des joueurs et le pourcentage de tirs accordés expliquent la performance d'une équipe sur 72 matchs est **confirmée**.

**L'hypothèse 2** selon laquelle "le pourcentage de 3 points accordés, de lancés francs, et le nombre de rebonds défensifs influencent l'issue d'un match est **confirmée**.

# Préconisations opérationnelles

## *A l'endroit des clubs professionnels*

- ♣ Relever dans un premier temps les différentes statistiques personnelles des joueurs ;
- ♣ Tenir des feuilles de statistiques régulièrement et compiler de manière informatisée les données des matchs ;
- ♣ Utiliser le modèle A 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 et ainsi mieux se préparer pour les matchs.



*Merci de votre attention*