

Football Statistique

CHIRINA Rania, Gonzalez Emmanuel, AKHLOUFI Khalil, Martin Samuel,
VITOFFODJI Adjimon

Département MIASHS, UFR 6 Informatique, Mathématique et Statistique Université Paul
Valéry, Montpellier 3

2024



- ① Problématique et objectifs de l'étude
- ② Méthodologie
- ③ Présentation des résultats

- ❶ Problématique et objectifs de l'étude

Problème posé

- L'analyse statistique des performances des joueurs de football est devenue un outil incontournable pour les entraîneurs, les analystes et les passionnés du ballon rond.
- Des joueurs emblématiques tels que Lionel Messi et Cristiano Ronaldo ont continué à dominer les classements statistiques, tandis que de nouveaux talents ont émergé pour laisser leur empreinte dans l'histoire du football.
- Cependant, malgré l'abondance de données disponibles, il reste des questions cruciales à explorer pour mieux comprendre les facteurs qui influent sur la performance des joueurs et pour anticiper les tendances à venir.

Problème posé

- Comme l'a souligné Johan Cruyff, légendaire joueur et entraîneur de football, "Les statistiques sont comme un bikini : elles montrent beaucoup de choses, mais elles cachent l'essentiel."
- Cette citation emblématique résume parfaitement la complexité des chiffres dans le contexte du football. Alors que les statistiques peuvent offrir des indications précieuses sur les performances individuelles et collectives des joueurs, elles ne capturent jamais entièrement l'essence du jeu, son flair artistique et son imprévisibilité.

Problématique (3/3)

Problème posé

- Quelles sont les variables qui ont une influence sur la performance des joueurs ?
- Qu'est-ce qui seront les potentiels meilleurs buteurs de chaque championnat durant l'année 2024?

objectifs (1/1)

Objectifs général

L'objectif général de cette étude est de mener une analyse approfondie des performances des joueurs de football de 2016 à 2023, en se concentrant sur les principales ligues européennes

Objectifs spécifique

- O_1 : d'examiner les données statistiques des joueurs pour identifier les variables clés qui influent sur leur performance individuelle et collective.
- O_2 : anticiper les tendances futures en matière de performances des joueurs, en utilisant des techniques de modélisation prédictive basées sur les données
- O_3 : prédire les potentiels meilleurs buteurs de chaque championnat pour l'année 2024, en tenant compte des caractéristiques des joueurs, des stratégies des équipes et des dynamiques du jeu.

② Méthodologie

Ce travail a été réalisé en utilisant les logiciels :

- Excel
- PhpMyAdmin
- Rstudio version 4.2.2
- Rmarkdown pour la rédaction
- Rmarkdown à travers Beamer pour la présentation

Natures et Sources des données

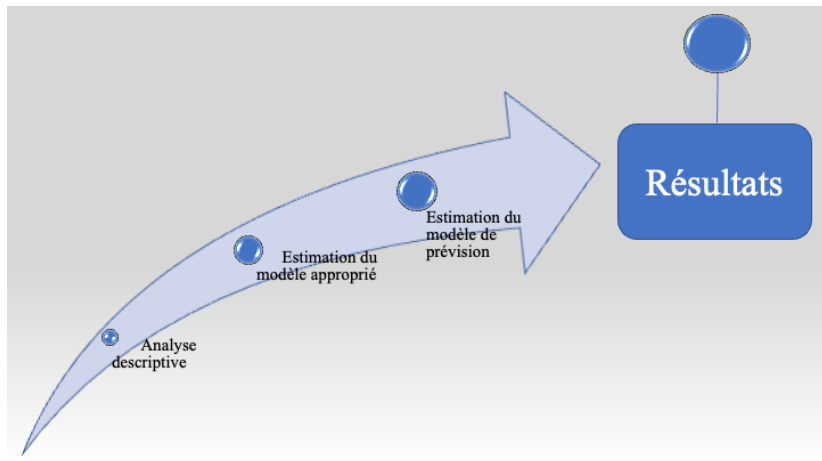
- Les données étutilisé dans la présente études couvrent la période de 2016 à 2023 et concernent les 5 grandes ligues européennes.
- Ces données proviennent de la base de sofascore sur le site de [sofascore.com](https://www.sofascore.com)

Identification des variables de l'étude

- **La variable endogène** : Le nombre de but marqué par chaque joueur au cours d'une année;
- **Les variables exogènes** : Nombre de match jouer, Le nombre de minutes joué pendant la saison, Le nombre de tire tanté pendant la saison, Le nombre de tire cadré pendant une saison, Le ratio de but par match, Grosses occasions créées durant la saison, Identifiant associé à chaque joueur

Methodologie (4/6)

Processus d'analyse



Methodologie (5/6)

Modèle d'analyse

Notre modèle d'analyse repose sur une fonction de type régression linéaire multiple qui se présente comme suit:

Modèle 1

$$But = \beta_0 + \beta_1 Tirs_Total + \beta_2 Gross_occas + \beta_3 tirs_cadrés + \beta_4 matchs + \beta_5 minutes_jouées + \beta_6 Ratio_but_par_match + \beta_7 id_nom + \varepsilon$$

Modèle 2

$$But = \beta_0 + \beta_1 tirs_cadrés + \beta_2 matchs + \beta_3 minutes_jouées + \beta_4 Ratio_but_par_match + \beta_5 id_nom + \varepsilon$$

Modèle 3

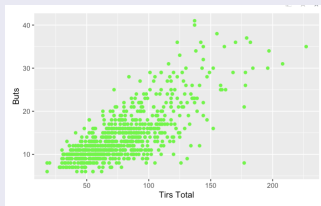
$$\hat{B}ut_t = \hat{\beta}_0 + \hat{\beta}_1 tire_c adr\acute{e}s_t + \hat{\beta}_2 matchs_t + \hat{\beta}_3 minutesjou\acute{e}es_t \\ + \hat{\beta}_4 Ratio_{but_match}_t + \hat{\beta}_5 id_{nom}_t + \epsilon_t$$

③ Présentation des résultats

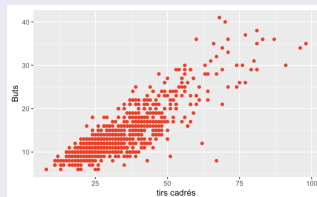
Présentation des résultats (1/8)

Analyse descriptive

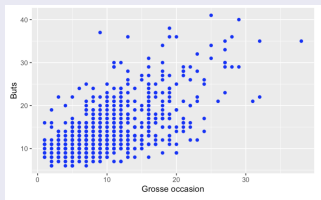
But et Tirs total



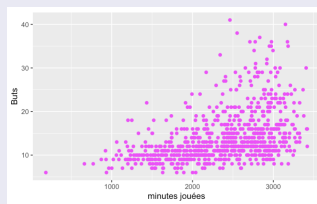
But et tirs cadrés



But et Grosse occasion



But et minutes jouées



Présentation des résultats (2/8)

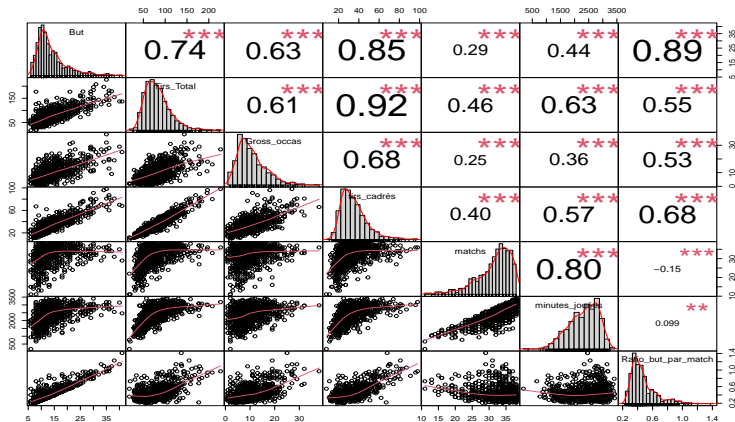
Analyse descriptive

Variables	Statistique W	p-value	Conclusion
But	0,85561	2,2e-16	Absence de normalité
Tirs_Total	0.94029	2.2e-16	Absence de normalité
Gross_occas	0.93052	2.2e-16	Absence de normalité
tirs_cadrés	0.91627	2.2e-16	Absence de normalité
matches	0.89984	2.2e-16	Absence de normalité
minutes_jouées	0.96931	6.374e-12	Absence de normalité
ratio_but_par_match	0.86488	2.2e-16	Absence de normalité

Présentation des résultats (3/8)

Analyse descriptive

Représentation graphique de la liaison entre les variables (Matrice de corrélation de Spearman)



Présentation des résultats (4/8)

Estimation du modèle

Estimation du modèle 1

	But	
	T-Statistics	Coefficients
Tirs_Total	-1,050	0,2939
Gross_occas	-1,050	0,2695
tirs_cadrés	4,974	8,05e-07 ***
matches	43,464	2e-16 ***
minutes_jouées	-2,346	0,0192 *
ratio_but_par_match	99,589	2e-16 ***
ld_Joueur	1,542	0,1235
C	-54,704	2e-16 ***
R²:	0,9803	
F-statistic:	5680	
p-value:	2,2e-16	

*** et * indiquent que les variables sont significatives aux seuils de 1/1000 et 5% respectivement.

Présentation des résultats (5/8)

Estimation du modèle

Estimation du modèle 2

	But	
	T-Statistics	Coefficients
tirs_cadrés	3,116	0,001899 **
matches	-17,277	2e-16 ***
minutes_jouées	-3,497	0,000496 ***
ratio_but_par_match	-34,402	2e-16 ***
Id_Joueur	1,223	0,221614
C	61,631	2e-16 ***
R^2 :	0,8353	
F-statistic:	811,6	
p-value:	2,2e-16	

*** et ** indiquent que les variables sont significatives aux seuils de 1/1000 et 10% respectivement.

Présentation des résultats (6/8)

Estimation du modèle de prévision

Modèle 3

$$\hat{But}_t = 61,631 + 3,116tire_cadrés_t - 17,277matchs_t + -3,497minutesjouées_t \\ + -34,402Ratio_{but_{match}}_t + 1,223id_{nom}_t + \epsilon_t$$

Présentation des résultats (7/8)

Interprétation de prévision

Pévision

Prévisions 2024		
id_nom	nom	But
68	Erling Haaland	35
1	Harry Kane	29
61	Ivan Toney	20
30	Mohamed Salah	19
41	Callum Wilson	18

Présentation des résultats (8/8)

Interprétation de prévision

Conclusion

Merci pour votre attention