

# Ecart de salaires entre les joueurs de football étrangers et nationaux, une approche économique de la discrimination

Axel-Cleris Gailloty

2021-07-05

## Table des matières

<b>1</b>	<b>Spécification technique</b>	<b>4</b>
<b>2</b>	<b>Plan du travail</b>	<b>4</b>
<b>3</b>	<b>Revue de littérature</b>	<b>5</b>
3.1	La discrimination en économie . . . . .	5
<b>4</b>	<b>Le modèle économique du football</b>	<b>6</b>
4.1	L'arrêt Bosman et ses conséquences . . . . .	6
4.2	Les sources de revenus des clubs de football . . . . .	6
4.2.1	La billetterie . . . . .	6
4.2.2	Droits télévisuels . . . . .	6
4.2.3	Sponsoring . . . . .	7
<b>5</b>	<b>Analyse exploratoire des données</b>	<b>8</b>
5.1	Source des données . . . . .	8
5.2	Caractéristiques du jeu de données . . . . .	8
5.3	Statistiques descriptives . . . . .	9
5.4	Visualisation de la distribution des revenus . . . . .	10
<b>6</b>	<b>Estimations économétriques</b>	<b>12</b>
6.1	Première estimation avec les OLS . . . . .	12
6.1.1	Diagnostics du modèle . . . . .	12
6.1.2	Analyse des coefficients . . . . .	12
6.1.2.1	Les coefficients des variables . . . . .	13
6.1.2.2	Les coefficients des variables synthétiques . . . . .	13
6.1.2.3	L'équation de la première estimation . . . . .	14
<b>7</b>	<b>Etudier l'écart des salaires selon l'extranéité</b>	<b>14</b>

7.1	Caractéristiques moyennes des joueurs natifs et étrangers avant appariement	15
7.2	Seconde estimation avec les OLS . . . . .	16
7.3	L'échantillonnage sur score de propension . . . . .	16
7.3.1	Comment faire un essai randomisé sur les joueurs de football . . . . .	17
7.4	Mesure de la qualité de l'appariement . . . . .	18
7.4.1	Distribution des scores de propension . . . . .	18
7.5	Caractéristiques moyennes des joueurs natifs et étrangers après appariement	20
<b>8</b>	<b>Modélisation économétrique de la discrimination</b>	<b>22</b>
8.1	La décomposition de Blinder-Oaxaca . . . . .	22
8.1.1	Fonctionnement de la méthode . . . . .	22
8.1.2	Application de la méthode aux écarts de salaires des footballeurs . . .	22
8.1.2.1	Pourcentage des écarts de salaire . . . . .	22
8.1.2.2	Les écarts imputables aux dotations individuelles . . . . .	22
8.1.2.3	Les écarts imputables aux coefficients . . . . .	23
8.1.2.4	Les écarts imputables à l'interaction des dotations individuelles et des coefficients . . . . .	24
8.1.3	La décomposition en deux facteurs . . . . .	25
<b>9</b>	<b>Conclusion</b>	<b>26</b>
9.1	Limites de l'étude . . . . .	26
9.1.1	Absence dans la base de données du revenu des clubs . . . . .	26
9.1.2	Absence de la composition du salaire . . . . .	26
<b>10</b>	<b>Bibliographie</b>	<b>27</b>
<b>11</b>	<b>Annexes</b>	<b>28</b>
11.1	Code source du projet . . . . .	28
11.2	Transformations des données . . . . .	28
11.2.1	Script R utilisé pour nettoyer les données . . . . .	28
11.2.2	Script utilisé pour factoriser les variables caractéristiques techniques des joueurs en 4 composantes principales . . . . .	29
11.3	Analyse en composantes principales . . . . .	31
11.3.1	1. Observation d'individus extrêmes . . . . .	31
11.3.2	2. Distribution de l'inertie . . . . .	31
11.3.3	3. Description du plan 1:2 . . . . .	32
11.3.4	4. Description du plan 3:4 . . . . .	35
11.3.5	5. Description de la dimension 5 . . . . .	38

---

*En moyenne un joueur de football en club dans le monde gagne 8780 € par mois. Ce montant est supérieur au salaire moyen des pays de l'OCDE. La différence est encore plus forte lorsque nous calculons le salaire moyen conditionnellement à l'extranéité du joueur. En moyenne un joueur étranger à la nationalité du club qui l'engage gagne 11569 € contre 6414 € si le joueur est de la même nationalité que le club qui l'engage. Ce différentiel qui du moins est énorme peut susciter des questions importantes quant à la répartition des salaires dans les clubs de foot. Face à une telle différence dans les rémunérations, un esprit rationnel ne peut s'empêcher d'énoncer des hypothèses pour rendre compte de cet écart. Dans cette étude nous proposons d'étudier facteurs qui entrent dans la composition du salaire d'un footballeur. Nous montrerons notamment que les clubs de football rémunèrent les joueurs de football en fonction de leurs caractéristiques individuelles. Nous montrerons que plus de 30% des écarts de salaire entre un joueur étranger et natif s'expliquent par les dotations personnelles des joueurs étrangers qui sont significativement plus élevées que celles des joueurs natifs.*

**Mots clés : football, régression linéaire, échantillonnage sur score de propension, décomposition de Blinder-Oaxaca**

---

# 1 Spécification technique

Les questions que nous nous sommes posées en préambule de ce travail nous forcent à réfléchir à des outils mathématiques, statistiques et économétriques qui permettent de répondre avec rigueur aux interrogations. Dans un premier temps, nous estimons nécessaires de nous orienter vers une approche par les données afin de quantifier les effets causaux qui entrent dans la détermination des salaires des footballeurs.

Bien que de nombreux travaux théoriques d’une rigueur académique aient été réalisés sur la problématique de la discrimination salariale en économie, nous estimons que les résultats théoriques exposés par ces travaux ne sauraient s’appliquer aux genres de questions que nous nous sommes posées dans le cadre de ce travail.

Nous ferons usages des outils de statistiques inférentielles pour tester nos hypothèses. Il s’agira notamment de tester la significativité des différences observées entre les groupes. Il nous sera donc possible de dire si oui ou non les différences observées dans les salaires sont dues au hasard.

Nous voudrions aussi évacuer les biais statistiques dans le traitement des données afin de poser pouvoir généraliser les résultats que nous obtenons. Pour cela nous utiliserons des méthodes d’échantillonnage qui nous permettent de comparer les joueurs de football en tenant compte de leurs caractéristiques.

Les outils économétriques nous serviront à étudier la causalité entre les variables que nous utilisons.

## 2 Plan du travail

Le plan suivra la trame que nous avons définies dans la spécification technique. Nous commencerons à présenter les données que nous avons collectées pour cette étude. Nous décrivons brièvement le processus mis en place pour obtenir les données. Puis nous montrons le contenu des données : le nombre d’observation, les variables.

Après cette brève présentation des données nous faisons une étude statistique. Nous calculons les tendances centrales : moyenne, médiane, écart-type. Nous visualiserons les données à l’aide des capacités graphiques du langage R.

Nous allons ensuite estimer plusieurs modèles économétriques avec des méthodes d’estimations connues comme les moindres carrés, la régression logistique afin de vérifier certaines hypothèses et poser le diagnostic des résultats intermédiaires. Nous accompagnerons chaque estimation économétrique par une série de tests statistiques.

Pour ne pas allourdir la lecture de ce document nous avons fait le choix de ne pas nous attarder sur le nettoyage des données, la création des variables par combinaison d’autres variables, les méthodes de réductions de dimensionnalité (ACP). Ces étapes qui ont été très utiles à la formation de la problématique et nous avons choisis de les inclure en annexes du document. Nous y ferons allusion dans le développement.

Ce travail a été réalisé avec le souci de la reproductibilité des résultats obtenus. Pour cette raison nous avons écrit entièrement ce document avec Rmarkdown qui est une solution qui permet d’écrire du texte puis du code R simultanément. Le document Rmarkdown peut être exporté dans une variété de format : latex, html, pdf, Microsoft Word etc. . .

Le document Rmarkdown qui contient à la fois le code et le texte peut être consulté librement sur mon site Internet.

## 3 Revue de littérature

Derrière chaque méthode d’estimation et des pratiques se trouve une histoire et un contexte. Nous souhaitons dans ce travail ne pas nous focaliser sur l’implémentation technique d’une approche économétrique sans en connaître les tenants et les aboutissants.

Nous allons dans un travail faire une brève revue de littérature portant sur la discrimination en économie et nous présentons les méthodes de mesure de discrimination. Nous choisirons parmi ces méthodes la méthode de décomposition de Blinder-Oaxaca pour expliquer l’écart de salaire entre les joueurs de football étrangers.

### 3.1 La discrimination en économie

Le mot discrimination ne revêt pas la même signification en économie que dans d’autres sciences sociales ou encore dans le langage courant. Dans le langage courant toute différence de traitement peut être perçue comme de la discrimination. L’exemple typique employé pour illustrer la notion de discrimination en économie est celui de l’écart de salaire sur le marché de travail entre différentes catégories de la population. En France, en moyenne une femme gagne moins de 27% qu’un homme au cours de sa carrière. Un écart de cet ampleur sait bien alimenter les articles de journaux parlant de la discrimination dans les salaires. Les études économétriques réalisées sur la question des écarts de salaire entre hommes et femmes affirment qu’il existe bel et bien une discrimination dans les salaires mais pas à hauteur de 27%. Ce qui nous force à définir la discrimination salariale dans une perspective économique et non seulement comme un écart quelconque.

Pour l’exemple que nous venons de prendre, nous ne pouvons parler de discrimination dans les salaires que lorsque des hommes et des femmes ayant la même productivité dans le même emploi ou dans deux emplois de même type et de même durée touchent des salaires différents.

Mesurer donc la discrimination en économie suppose que les deux catégories qu’on compare possèdent des caractéristiques observables identiques sauf sur la variable d’intérêt sur laquelle on veut tester la discrimination. Pour le reformuler autrement on ne peut parler de discrimination que si la condition “*ceteris paribus*” (toute chose égale par ailleurs) est vérifiée.

Or il est impossible de trouver des individus identiques sur tous les points. Pour mesurer la discrimination en gardant la condition toute chose égale par ailleurs, la science économique compare deux groupes d’individus qui ont les mêmes caractéristiques sauf sur un seul point qui sert à mesurer la discrimination.

## 4 Le modèle économique du football

### 4.1 L'arrêt Bosman et ses conséquences

Pour devenir le sport le plus médiatisé, le football a dû subir des bouleversements profonds. L'année 1995 marque le début de changements profonds dans le monde du football international. L'expression "Arrêt Bosman" est populaire auprès des économistes du sport pour désigner ce changement de visage du football.

En 1990 un joueur de foot le milieu offensif belge du Royal Football Club (RFC) de Liège, Jean Marc Bosman veut quitter son club pour signer en France à Dunkerque car à Liège il était arrivé en fin de contrat. Mais à l'époque un club acheteur doit toujours verser une indemnité de transfert au club détenteur même lorsque le joueur est en fin de contrat. Dans le cas de Bosman, Dunkerque refuse de payer une indemnité au club de Liège qui à son tour refuse de laisser partir son joueur sans indemnité de transfert. Jean Marc Bosman est pris au piège et décide d'attaquer en justice le club de Liège. Cinq ans plus tard sa plainte aboutit : les joueurs en fin de contrat seront désormais libres de s'engager où bon leur semble. Ce dénouement est finalement d'une importance mineure. Cette décision de justice qui sera connue plus tard sous le nom de **l'Arrêt Bosman** aura un retentissement profond dans le monde du football pour une autre raison. En effet, avant 1995 les clubs pouvaient pas compter plus de trois joueurs d'une nationalité différente que celle du club en question.

La décision rendue par la Cour de Justice Européenne, juge que ce principe de quota de joueurs étrangers est en fait tout aussi injuste que les indemnités de transfert qui avaient initialement motivé Bosman à aller en justice. La cour de justice juge que les quotas remettent en question la liberté de circulation des joueurs de football qui sont considérés comme des travailleurs. Le football européen change profondément à partir de cette décision. Si aujourd'hui on observe un grand nombre de joueurs étrangers dans des clubs européens c'est grâce **l'Arrêt Bosman**.

### 4.2 Les sources de revenus des clubs de football

Au cours des cinquante dernières années, les clubs de football sont devenus des entreprises. Ils réalisent aujourd'hui un chiffre d'affaires qui provient de trois sources principales que sont la vente des billets, la vente des droits de retransmissions des matchs et le sponsoring.

#### 4.2.1 La billetterie

Il s'agit de la vente des billets pour assister aux matchs dans les stades. C'est la plus ancienne source de revenus des clubs de football. En France les premiers matchs payant remontent au début des années 1890. On observe toutefois aujourd'hui que la plupart des stades ont réduit leur capacité d'accueil depuis les années 1950. En 2015 le Real Madrid a réalisé 130 million d'euros de chiffre d'affaires uniquement via la vente des billets.

#### 4.2.2 Droits télévisuels

Les clubs de football vendent à des chaînes de télévision le droit de retransmettre les matchs à la télévision. La retransmission des matchs de football en direct à la télévision a débuté

dans les années 1950 mais ce n'est seulement dans les années 1980 que les dirigeants des clubs de football ont décidé de faire de la retransmission des matchs de football un levier important de revenu. Aujourd'hui les droits télévisuels sont la première source de revenu des clubs de football.

#### **4.2.3 Sponsoring**

Les clubs de football cèdent à une entreprise, le droit de communiquer en utilisant l'image du club et de faire du merchandising (vente de maillots essentiellement). Par exemple le FC Barcelone est aujourd'hui associé à des machines à laver et le Paris Saint-Germain à des crèmes pour le visage. Le sponsoring représente une source importante de revenu pour les grands clubs populaires. Les plus petits clubs ont du mal à trouver des sponsors. Le PSG a réalisé en 2015 un revenu de 297 millions d'euros grâce au sponsoring.

## 5 Analyse exploratoire des données

### 5.1 Source des données

Nous allons réaliser notre étude sur un jeu de données provenant de FIFA qui contient 18944 joueurs de football. En fait ce jeu de données ne contient pas la totalité des joueurs de football professionnels, mais les plus connus. Il s'agit des joueurs qu'on retrouve dans les jeux vidéos FIFA. Chaque année une version du jeu est commercialisé en prenant en compte les dernières modifications qui se sont effectuées au sein des clubs et dans les performances des joueurs. Le jeu de données que nous utilisons dans ce travail est celui de FIFA 21.

Les données ne proviennent pas directement de FIFA mais du site [Sofifa](#). La recolte des données dans un format tabulaire exploitable pour des études économétriques s'est faite grâce à un programme informatique d'extraction de contenu web. Le jeu de données ainsi extrait du site Sofifa est rendu disponible gratuitement sur [Kaggle](#) qui est une plateforme de compétition pour les data scientists. Les données disponibles sur Kaggle sont publiées sous la licence *CC0: Public Domain*. Le jeu de données appartient donc au domaine public, toute personne qui dispose d'un compte Kaggle peut le télécharger gratuitement à l'adresse suivante : <https://www.kaggle.com/stefanoleone992/fifa-21-complete-player-dataset>. L'auteur du jeu de données s'appelle Stefano Leone.

### 5.2 Caractéristiques du jeu de données

Dans son format brut, le jeu de données contient 18944 observations (lignes) et 108 variables (colonnes). Une observation correspond à un joueur de football et une variable contient les caractéristiques personnelles du joueur et du club dans lequel il joue. Le tableau suivant affiche un échantillon de quelques joueurs et de quelques variables du jeu de données :

sofifa_id	short_name	age	nationality	club_name
158023	L. Messi	33	Argentina	FC Barcelona
20801	Cristiano Ronaldo	35	Portugal	Juventus
200389	J. Oblak	27	Slovenia	Atlético Madrid
188545	R. Lewandowski	31	Poland	FC Bayern München
190871	Neymar Jr	28	Brazil	Paris Saint-Germain
192985	K. De Bruyne	29	Belgium	Manchester City

Il existe de nombreuses observations et de variables dans le jeu de données. Nous n'allons pas tous les reprendre dans notre analyse. Ce que nous ferons plutôt c'est de ne choisir que les caractéristiques qui sont pertinentes pour l'analyse. Pour alléger ce travail nous choisissons de détailler notre heuristique dans la section 1 des annexes de ce document.

Le tableau suivant résume les caractéristiques que nous avons choisies. Les critères de ce choix sont simples et intuitifs : la caractéristique doit avoir le moins de valeurs manquantes possibles. Elle doit contribuer à l'hétérogénéité des joueurs tout en gardant un certain degré d'homogénéité.



Table 2: Caractéristiques des joueurs retenues dans l’analyse

Caractéristique des joueur	Description
Age	L’âge du joueur en année
Poids	Le poids mesuré en kg
Taille	La taille mesurée en cm
Nationalité	La nationalité du joueur
Club	Le nom du club dans lequel il joue
Potentiel du joueur	Valeur estimée par FIFA allant de 0 à 100. Si la valeur est de 100 le joueur excelle dans toutes les classifications de FIFA.
Réputation internationale	Estimation de la réputation internationale du joueur. Valeur allant de 1 à 5
Caractéristiques techniques du joueur	Variables synthétiques estimés par une ACP sur un ensemble de 60 variables
Salaire du joeur	Le montant du salaire du joueur en euro
League 1	Variable binaire indiquant si le club est en première division

### 5.3 Statistiques descriptives

Nous allons brièvement résumer les données à l’aide des statistiques descriptives et des graphiques. Nous souhaitons connaître la distribution de la variable salaire (que nous avons converti en logarithme) et des autres variables explicatives.

Table 3: Les nationalités les plus présentes dans le jeu de données.

nationality	n
England	1685
Germany	1189
Spain	1072
France	984
Argentina	936
Brazil	885
Japan	489
Netherlands	432
Italy	421
United States	378

Les joueurs anglais sont les plus représentés dans le jeu de données. Cela confirme bien la réputation de l’Angleterre qui est le pays berceau du soccer qui est l’ancêtre du football. Les 4 premières nationalités sont européennes. Les Européens sont bien représentés dans le

football. L'Argentine et le Brésil occupent les 5e et 6e positions.

Le footballeur moyen est âgé de 25 ans, il mesure 1,81m et pèse 75 kg. Il gagne 8780 euros Sa performance globale est de 65%. 38% des joueurs de football sont de nationalité européenne (UE).

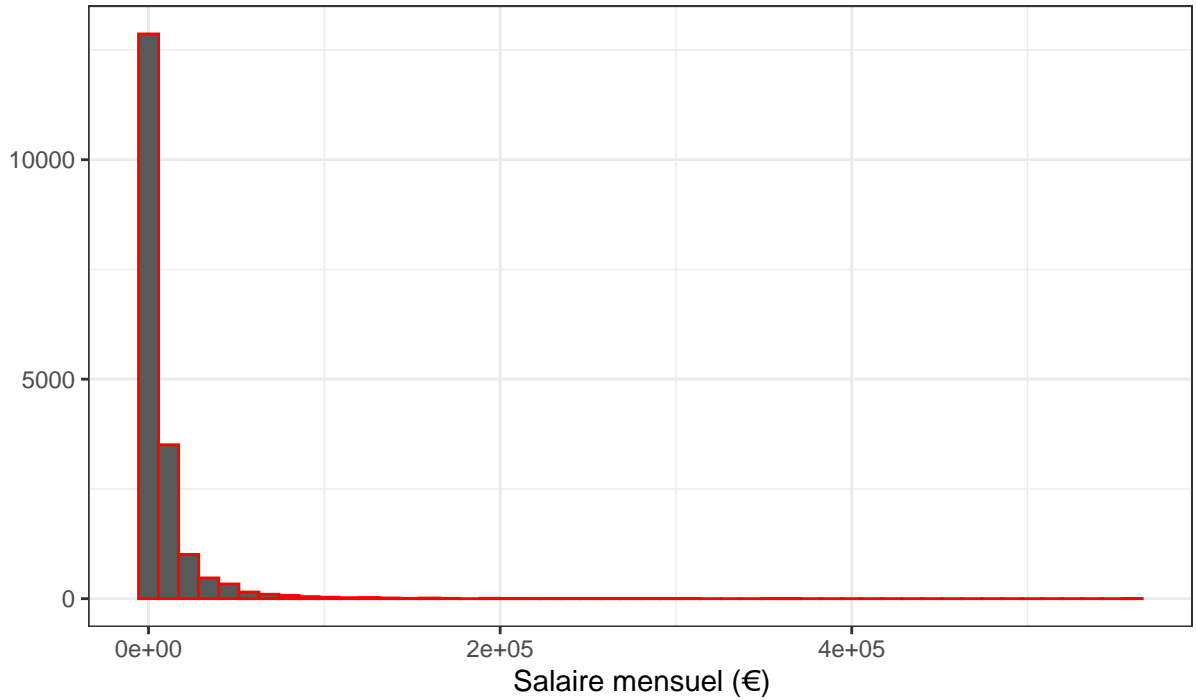
Nous affichons plus en détails ces statistiques dans le tableau suivant.

	Overall
n	18719
age (mean (SD))	25.18 (4.68)
height_cm (mean (SD))	181.19 (6.82)
overall (mean (SD))	65.64 (7.00)
potential (mean (SD))	71.09 (6.11)
weight_kg (mean (SD))	75.01 (7.05)
player_european (mean (SD))	0.38 (0.49)
international_reputation (mean (SD))	1.09 (0.36)
weak_foot (mean (SD))	2.94 (0.67)
league1 (mean (SD))	0.77 (0.42)
european_club (mean (SD))	0.43 (0.50)
wage_eur (mean (SD))	8780.14 (19749.38)
skill_moves (mean (SD))	2.36 (0.77)
Dim.1 (mean (SD))	-0.01 (4.30)
Dim.2 (mean (SD))	0.00 (2.09)
Dim.3 (mean (SD))	-0.01 (1.65)
Dim.4 (mean (SD))	0.00 (1.28)

## 5.4 Visualisation de la distribution des revenus

Il convient d'analyser la distribution des salaires. Pour cela nous représentons un histogramme des salaires.

Distribution du salaire mensuel des footballeurs en euro.



La distribution des salaires est fortement concentrée vers la gauche. Ce qui décrit de fortes inégalités dans les revenus. Nous pouvons calculer les différents moments de cette distribution.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
500	1000	3000	8780	8000	560000

Le plus faible salaire est de 500 euros par mois tandis que le plus haut salaire s'élève à 560000 euros par mois. Nous pouvons retrouver les joueurs 5 joueurs qui gagnent le plus et les 5 joueurs qui gagnent le moins.

Table 5: Top 5 des joueurs qui gagnent le plus.

short_name	wage_eur
L. Messi	560000
K. De Bruyne	370000
K. Benzema	350000
E. Hazard	350000
Casemiro	310000

Table 6: Top 5 des joueurs qui gagnent le moins.

short_name	wage_eur
Mãjrio Fernandes	500
Marlos	500

short_name	wage_eur
Taison	500
Ismaily	500
Adryan Zonta	500

Les statistiques que nous venons de calculer sur le salaire montrent qu’il existe des écarts significatifs dans le salaire des joueurs de football. Toutefois les statistiques que nous venons de calculer ne sont qu’informatives, elles ne nous donnent pas d’explications sur ces écarts de revenus. Pour comprendre l’écart de revenus entre les joueurs il nous faut étudier la causalité. En d’autres termes nous voulons quantifier les facteurs qui entrent dans le salaire des footballeurs qui expliquent ces écarts.

## 6 Estimations économétriques

### 6.1 Première estimation avec les OLS

Nous commençons les estimations par une régression linéaire multiple pour expliquer globalement les facteurs qui entrent dans la détermination des salaires des footballeurs.

term	estimate	std.error	statistic	p.value	signif.
(Intercept)	-2.669147	0.326655	-8.171146	0.000000	***
age	-0.031245	0.002938	-10.636130	0.000000	***
height_cm	0.020377	0.001748	11.654701	0.000000	***
weight_kg	-0.010868	0.001589	-6.841346	0.000000	***
overall	0.131301	0.002789	47.079768	0.000000	***
potential	-0.007606	0.002486	-3.059905	0.002217	**
international_reputation	0.416106	0.020453	20.344247	0.000000	***
weak_foot	0.019579	0.010572	1.851900	0.064056	.
league1	0.012987	0.015765	0.823833	0.410045	
european_club	0.038013	0.013708	2.773014	0.005559	**
Dim.1	0.034414	0.002371	14.511402	0.000000	***
Dim.2	0.012027	0.003530	3.406538	0.000659	**
Dim.3	0.058189	0.007890	7.375233	0.000000	***
Dim.4	0.002824	0.006709	0.420929	0.673812	

#### 6.1.1 Diagnostics du modèle

Les tests statistiques réalisés indiquent une significativité globale du modèle au seuil de 1%. Ce modèle peut être candidat à expliquer la variation du salaire entre les footballeurs.

#### 6.1.2 Analyse des coefficients

**6.1.2.1 Les coefficients des variables** Le coefficient associé à la variable âge est significatif et négatif. En moyenne et toute chose égale par ailleurs un an de plus fait baisser le salaire moyen de -3.1 %. Cela donne une première indication permettant d'expliquer les variables qui rendent compte du salaire des footballeurs : être jeune est un atout.

La taille des joueurs représente également un atout. Plus un joueur est grand de taille plus son salaire est élevé.

Le coefficient de la variable `overall` qui mesure la performance globale du joueur est l'un des plus élevés et le plus significatif. Le coefficient associé à cette variable indique qu'une l'augmentation d'une unité de performance globale entraîne une hausse du salaire de 13.1 % toute chose égale par ailleurs.

La réputation internationale a un très grand impact sur le salaire. L'augmentation d'une unité de réputation augmente le salaire moyen de 41.6 %. (les valeurs de la variables sont des entiers compris entre 1 et 5. Messi et Ronaldo ont une réputation de 5).

Nous remarquons aussi que le fait de jouer dans un club européen augmente le salaire de 3.8 % toute chose égale par ailleurs.

**6.1.2.2 Les coefficients des variables synthétiques** Le jeu de données initial contenait une trentaine de caractéristiques physiques des joueurs. Les inclure toutes dans un modèle économétrique risque d'introduire des biais dans l'analyse. Les caractéristiques individuelles des joueurs sont fortement corrélées entre elles, ce qui n'apporte pas d'informations supplémentaires. Nous avons donc fait le choix de réduire ces caractéristiques individuelles en 4 variables synthétiques qui résument environ 80% de la variance totale. Les variables Dim.1 - Dim.4 sont les résultats de l'analyse en composantes principales. Les composantes principales que nous avons retenues pour résumer les caractéristiques des joueurs sont les composantes 1 à 4. Plus concrètement chaque valeur dans chacune des composante représente la coordonnée d'un joueur dans un espace à n dimensions. Nous prenons avantage sur le fait que les coordonnées produites par l'ACP sont orthogonales d'une dimension à l'autre pour évacuer tout biais de multicollinéarité entre les joueurs. Nous détaillons plus en détails ce que renferme chaque dimension dans les annexes. Mais pour comprendre le résultat de l'estimation que nous venons de faire voici brièvement ce que représente chaque axe :

Sur cette axe se trouvent les joueurs qui ont un profil d'attaquant. Ce sont des joueurs qui ont des tirs cadrés même sur longue distance. Ces joueurs ont une bonne vision du jeu, ils savent très bien dribbler et mener des attaques offensives contre l'équipe adverse. Les joueurs de cet axe ont des performances très médiocres en défenses.

Les joueurs de l'axe 2 ont un profil de gardien de but. Ils savent se positionner pour repousser les attaques offensives. Ces joueurs ont souvent une faible maîtrise de la balle et une endurance limitée.

Les joueurs de l'axe 3 ont un profil de défenseur. Ils savent lire le jeu et sont forts en interceptions de ballon. Ils exécutent bien les tacles défensives et ils possèdent une endurance accrue. Ils peuvent se montrer agressifs sur le terrain et sont plein d'énergies. Ils ont un tir médiocre, ils ne savent pas exécuter une attaque défensive.

Les joueurs de l'axe 4 sont des milieu de terrain : ils sont très agiles et très rapides . Ce sont des joueurs qui accélèrent très rapidement pour intercepter une action offensive et récupérer la balle et tenter une contre-offensive. Ils ont une bonne lecture du jeu et savent bien se positionner sur le terrain.

Les trois premiers axes que nous avons inclus dans le modèle sont significatifs au seuil de 1% et ont tous des coefficients positifs. Ces coefficients confirment bien que les caractéristiques physiques et techniques des joueurs expliquent une partie de la variance des salaires de joueurs de football. Nous voyons que toute chose égale par ailleurs qu'avoir des performances élevées sur les postes d'attaquant (représenté par la variable Dim.1) fait augmenter le salaire de 3.4 %, des performances positives sur les postes de gardien (Dim.2) font augmenter le salaire de 1.2 % . Avoir de bonnes performances sur les postes de défenseur fait augmenter le salaire de 5.8 %. Le coefficient associé aux joueurs ayant de bonnes performances sur les postes de milieu offensif ou milieu défensif n'est pas significatif au seuil de 5%.

La première estimation que nous venons de réaliser a l'intérêt de décomposer la variation du revenu à l'aide des variables explicatives. Nous pouvons résumer la fonction du salaire des footballeurs à l'aide de cette équation :

### 6.1.2.3 L'équation de la première estimation

$$\widehat{\text{lwage\_eur}} = -2.67 - 0.03(\text{age}) + 0.02(\text{height\_cm}) - 0.01(\text{weight\_kg}) + \\ 0.13(\text{overall}) - 0.01(\text{potential}) + 0.42(\text{international\_reputation}) + 0.02(\text{weak\_foot}) + \\ 0.01(\text{league1}) + 0.04(\text{european\_club}) + 0.03(\text{Dim. 1}) + 0.01(\text{Dim. 2}) + \\ 0.06(\text{Dim. 3}) + 0(\text{Dim. 4})$$

Nous remarquons la significativité de la constante ainsi que son coefficient très élevé en valeur absolue. En général une constante ayant un coefficient non nul est une indication du biais de la variable omise. Cela signifie que nous manquons de variables observables pour rendre compte de certains écarts dans le salaire.

Nous tenterons de créer de nouvelles variables pour rendre mieux compte de cet écart dans les salaires.

## 7 Etudier l'écart des salaires selon l'extranéité

L'extranéité est une variable que nous avons créée à partir des autres variables du jeu de données initial. Nous détaillons plus en détails la création de cette variable dans les annexes. En bref nous avons créé cette variable à partir de deux variables qui se trouvent dans le jeu de données initial : la variable représentant la nationalité du joueur et la ligue à laquelle appartient le club. Nous avons remarqué que très souvent le nom du pays apparaît dans le nom de la league. Par exemple : *Spain Primera Division*, *Italian Serie A*, *German 1. Bundesliga*, *French Ligue 1*, *English Premier League*, *Portuguese Liga ZON SAGRES*, *Korean K League Classic*. Voyant cette régularité nous avons écrit une suite d'instruction qui extrait le nationalité de la league puis à l'aide d'une table de transcodification que nous avons

recherchée sur Internet nous avons pu convertir de manière automatique la nationalité en nom du pays par exemple French devient France. La variable qui représente le pays d'origine des joueurs est présente dans le jeu de données. Nous avons donc créé une variable binaire qui prend la valeur 1 si le pays de nationalité du joueur est égal au pays de nationalité du club, 0 sinon. Cette variable est donc l'instrument que nous utilisons pour déterminer l'extranéité d'un joueur.

Pour avoir une idée de la pertinence de cette nouvelle variable, nous choisissons de réaliser une série de tests de Student sur l'ensemble des variables. Nous comparons la moyenne et les écarts-types des variables et nous calculons la significativité des différences entre les moyennes.

## 7.1 Caractéristiques moyennes des joueurs natifs et étrangers avant appariement

Le tableau suivant nous montre que la variable `foreigner` que nous venons de créer est très significative. Cette variable montre qu'il existe de réelles différences entre les caractéristiques des joueurs étrangers et celles des joueurs natifs.

En moyenne les joueurs étrangers sont plus âgé d'une année, ils sont plus grand en taille, ils ont des performances globales plus impressionnantes, ils sont plus réputés à l'international et ont des profils d'attaquant.

	Natif	Etranger	p.value
n	10127	8592	
age (mean (SD))	24.70 (4.82)	25.74 (4.44)	<0.001
height_cm (mean (SD))	180.90 (6.73)	181.53 (6.91)	<0.001
overall (mean (SD))	64.40 (6.62)	67.09 (7.16)	<0.001
potential (mean (SD))	70.71 (5.79)	71.53 (6.44)	<0.001
weight_kg (mean (SD))	74.51 (6.94)	75.59 (7.15)	<0.001
player_european (mean (SD))	0.48 (0.50)	0.27 (0.44)	<0.001
international_reputation (mean (SD))	1.06 (0.28)	1.13 (0.44)	<0.001
weak_foot (mean (SD))	2.90 (0.64)	2.98 (0.69)	<0.001
league1 (mean (SD))	0.70 (0.46)	0.85 (0.36)	<0.001
european_club (mean (SD))	0.48 (0.50)	0.38 (0.48)	<0.001
wage_eur (mean (SD))	6414.25 (13823.45)	11568.70 (24702.35)	<0.001
skill_moves (mean (SD))	2.28 (0.73)	2.46 (0.79)	<0.001
Dim.1 (mean (SD))	-0.46 (4.35)	0.53 (4.17)	<0.001
Dim.2 (mean (SD))	-0.04 (2.03)	0.05 (2.16)	0.003
Dim.3 (mean (SD))	-0.25 (1.62)	0.28 (1.65)	<0.001
Dim.4 (mean (SD))	0.05 (1.22)	-0.06 (1.35)	<0.001

Nous pouvons réestimer la régression linéaire en ajoutant cette variable en plus pour voir l'impact de l'extranéité sur le salaire des joueurs de football.

## 7.2 Seconde estimation avec les OLS

Dans la seconde estimation nous incluons la variable `foreigner` qui traduit ici l'extranéité des joueurs.

term	estimate	std.error	statistic	p.value	signif.
(Intercept)	-2.638197	0.326918	-8.069906	0.000000	***
foreigner	0.030155	0.013576	2.221172	0.026351	*
age	-0.030813	0.002944	-10.466928	0.000000	***
height_cm	0.020176	0.001751	11.525665	0.000000	***
weight_kg	-0.010949	0.001589	-6.891012	0.000000	***
overall	0.130671	0.002803	46.618011	0.000000	***
potential	-0.007147	0.002494	-2.865574	0.004167	**
international_reputation	0.414936	0.020458	20.282424	0.000000	***
weak_foot	0.019249	0.010572	1.820750	0.068661	.
league1	0.008023	0.015921	0.503925	0.614320	
european_club	0.041554	0.013799	3.011382	0.002604	**
Dim.1	0.034069	0.002376	14.337094	0.000000	***
Dim.2	0.011672	0.003534	3.302968	0.000958	**
Dim.3	0.058043	0.007889	7.357274	0.000000	***
Dim.4	0.002921	0.006708	0.435471	0.663226	

Toute chose égale par ailleurs, un joueur étranger à la nationalité du club pour qui il joue gagnent 3 % de plus qu'un joueur qui n'est pas étranger.

Cette différence est intéressante à analyser mais nous ne pouvons pas affirmer hâtivement que le fait d'être étranger procure une hausse du salaire, d'autant plus que les séries de tests de Student indiquent qu'il y a un biais de sélection. Les joueurs étrangers sont sélectionnés en moyenne pour des caractéristiques physiques et techniques supérieures.

Nous ne pouvons donc pas affirmer que l'extranéité soit la cause du salaire plus élevé des joueurs étrangers. Il nous faut comparer les joueurs sur une base de **toutes choses égales par ailleurs**, autrement dit nous devons mesurer la différence de salaire entre les joueurs étrangers et les joueurs natifs qui ont des caractéristiques très similaires.

Nous allons introduire la notion de l'échantillonnage sur score de propension.

## 7.3 L'échantillonnage sur score de propension

Cette méthode permet de définir les groupes de traitement et de contrôle en corrigeant l'éventuel biais de sélection. L'appariement sur score de propension se base sur le calcul d'un «score» qui résume l'influence des variables explicatives. Rajouter ce score permet de contrôler les influences des X et corriger le biais de sélection. Le calcul de ce score s'effectue souvent par un logit (probit) sur la variable de traitement.

La méthode du score de propension est souvent présentée comme une alternative peu coûteuse



aux essais randomisés. Un essai randomisé contrôlé est un protocole expérimental ayant pour but d'évaluer l'efficacité d'une thérapie, d'une action de prévention ou d'un médicament. Il compare un groupe expérimental dit groupe d'intervention à qui l'on administre le traitement et un groupe dit groupe de contrôle ou groupe témoin suivant un traitement standard ou prenant un placebo.

### 7.3.1 Comment faire un essai randomisé sur les joueurs de football

L'usage que nous allons faire de l'échantillonnage sur score de propension est quelque peu inédit. En effet, en temps normal nous utilisons le score de propension pour diviser une population en deux groupes : un groupe témoin et un groupe test. Dans notre cas nous savons au préalable quel joueur est étranger au club dans lequel et quel joueur est natif. Nous voulons donc utiliser le score de propension non pour faire un essai randomisé mais pour répondre à la question suivante : les joueurs étrangers ont-ils des caractéristiques comparables aux joueurs natifs ? Nous allons calculer le score de propension pour chacun des joueurs en prenant l'extranéité comme la variable dichotomique servant à identifier si le joueur est dans le groupe de traitement ou dans le groupe test. Nous avons codé la variable **foreigner** qui représente l'extranéité de telle sorte qu'elle prenne la valeur 1 si le joueur est étranger au club et 0 s'il est natif du club. Nous voulons en quelque sorte reproduire le contrefactuel. Etant donné le salaire actuel d'un joueur étranger, que gagnerait-il s'il était de la même nationalité que le club qui l'engage ?

Pour calculer le score de propension nous n'incluons que les variables qui décrivent les caractéristiques des joueurs et des clubs de football. La variable salaire est par exemple omise car c'est le salaire que nous aimerons expliquer en fonction des caractéristiques des joueurs. Il existe plusieurs manières d'estimer le score de propension. Selon les méthodes et les paramètres que l'algorithme utilise la qualité du score de propension peut varier. Nous avons observé que calculer le score de propension avec ma méthode **nearest** est plus rapide à exécuter. Mais cette méthode n'utilisant pas de l'optimisation ne donne pas de résultats satisfaisant : il n'y a pas de contrainte sur l'algorithme des k-voisins proches, de surcroît pour garder une taille d'échantillon assez conséquent il ne faut pas utiliser un paramètre ratio élevé. Le ratio est un nombre entier qui indique le nombre de voisins plus proches à prendre en compte lors du matching. Un ratio de 2 signifie qu'un joueur étranger sera apparié à deux joueurs natifs. Nous essayons la méthode optimale. Calculer le score de propension en utilisant cette méthode nous a pris plus de 20 minutes sur un ordinateur i5 et 8GB de RAM mais a abouti à une erreur car il n'y a pas assez de mémoire pour allouer les calculs intermédiaires.

La méthode d'estimation utilisée pour calculer le score de propension est la régression logistique. L'appariement que nous faisons sur les scores de propension est un appariement 1:1 sans remplacement, un joueur étranger est apparié à un joueur natif. Les variables que nous utilisons sont l'âge, la taille, la performance globale, le potentiel du joueur, sa réputation internationale, la capacité du joueur à jouer avec son pied faible (pied gauche si droitier et pied droit si gaucher), l'indicatrice qui montre que le club est en première ligne, le club est-il européen et les caractéristiques techniques individuelles des joueurs (Dim.1 à Dim.4). C'est ce que nous résumons dans le tableau suivant.

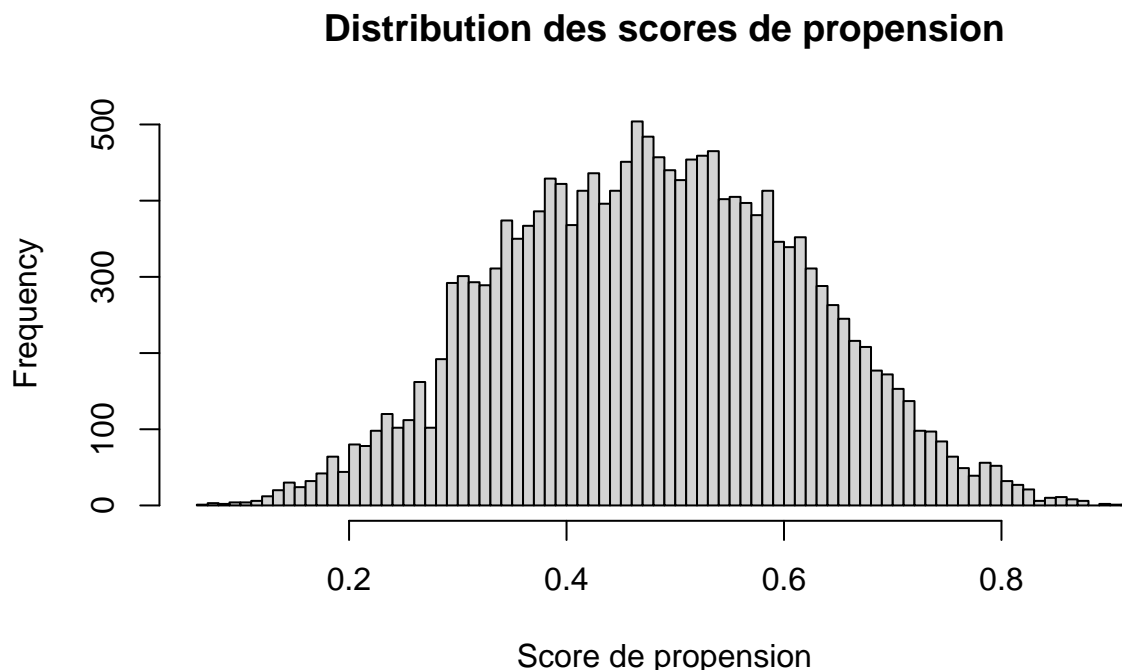
A `matchit` object

- method: 1:1 nearest neighbor matching without replacement
- distance: Propensity score
  - estimated with logistic regression
- number of obs.: 18719 (original), 17184 (matched)
- target estimand: ATT
- covariates: age, height\_cm, overall, potential, international\_reputation, weak\_foot,

## 7.4 Mesure de la qualité de l'appariement

### 7.4.1 Distribution des scores de propension

Nous représentons la distribution des scores de propension à l'aide d'un histogramme. Plus le score est élevé meilleure est la qualité de l'appariement. Sur le graphique suivant nous voyons que l'allure des scores de propension suit une distribution normale. Il y a des scores de propension très élevés et d'autre très faibles. Nous verrons un peu plus tard ce que cela implique que le score de propension soit faible ou élevé (compris entre 0 et 1).



Nous avons construit un tableau qui affiche un échantillon des résultats du matching. La fonction `match.data()` du package `MatchIt` ajoute les colonnes `subclass` et `distance` au jeu de données initial. La colonne `subclass` précise les identifiants des sous-classes uniques qui ont été créées. Une sous-classe contient une paire d'individu apparié. Il y a donc en total  $N/2$  sous-classes avec  $N$  le nombre total d'observations dans le jeu de données. La colonne `distance` de son côté contient le score de propension estimé pour le joueur. Voyons un aperçu de la qualité de l'appariement en affichant quelques entrées du jeu de données.

Nous affichons les scores de propension par ordre décroissant. D'abord nous affichons dans le tableau suivant les 10 premiers scores de propension les plus élevés.

Table 10: Les 10 premiers scores de propension les plus élevés.

subclass	Etranger / Natif	Country	Distance	Club	Nationalité club	Salaire
40	P. Pogba / H. Kane	France / England	0.912578	Manchester United / Tottenham Hotspur	England / England	190000 / 220000
6	K. De Bruyne / D. Alli	Belgium / England	0.883243	Manchester City / Tottenham Hotspur	England / England	370000 / 105000
37	H. Son / M. Rashford	Korea Republic / England	0.876011	Tottenham Hotspur / Manchester United	England / England	165000 / 150000
99	R. Jiménez / T. Alexander-Arnold	Mexico / England	0.852286	Wolverhampton Wanderers / Liverpool	England / England	140000 / 110000
8	V. van Dijk / T. Abraham	Netherlands / England	0.849843	Liverpool / Chelsea	England / England	210000 / 77000
180	M. Arnautovic / Morata	Austria / Spain	0.848694	Shanghai SIPG FC / Atlético Madrid	China / Spain	24000 / 73000
34	P. Aubameyang / T. Müller	Gabon / Germany	0.846870	Arsenal / FC Bayern München	England / Germany	170000 / 130000
45	K. Havertz / A. Sobolev	Germany / Russia	0.845075	Chelsea / Spartak Moscow	England / Russia	105000 / 500
132	G. Bale / K. Mbappé	Wales / France	0.843990	Tottenham Hotspur / Paris Saint-Germain	England / France	130000 / 160000
127	G. Higuaín / R. Barkley	Argentina / England	0.840522	Inter Miami / Chelsea	NF_usa / England	17000 / 80000

Par exemple dans la première ligne du tableau nous observons que la moyenne des scores de propension de Paul Pogba et de Harry Kane est de 0.9125. Un score de propension proche de 1 traduit une bonne qualité d'appariement entre les deux joueurs.

Nous pouvons interpréter la première ligne du tableau comme suivant : Paul Pogba de nationalité française est apparié à Harry Kane de nationalité anglaise. Paul Pogba est de ce fait étranger à la nationalité du club dans lequel il joue tandis que Harry Kane est natif du club dans lequel il joue. Les deux joueurs ne jouent toutefois pas dans le même club.

Dans ce premier exemple nous voyons que Harry Kane qui est natif du club dans lequel il joue gagne plus que Paul Pogba qui est étranger. La qualité de l'appariement est très bonne pour ces deux joueurs. Bien que nous n'ayons pas inclus le salaire dans le calcul du score de propension nous voyons tout de même qu'à compétences presque similaires les deux joueurs gagnent des revenus très élevés.

Ensuite nous affichons les scores de propension les moins élevés.

subclass	Etranger / Natif	Distance	Club	Country	Salaire
8474	A. O'Reilly / S. Zehnder	0.118067	Preston North End / SV Darmstadt 98	Republic of Ireland / Germany	2000 / 2000
8207	D. OntuÅans / L. Ravenhill	0.116115	Bayern München II / Doncaster Rovers	Latvia / England	1000 / 500
7147	A. Lewis / J. Hill	0.111844	Amiens SC / Fleetwood Town	England / England	750 / 500
7148	L. Klassen / O. Nkambadio	0.108404	TSV 1860 München / Paris FC	Russia / France	500 / 500
7792	G. Azzinnari / FortuÅo	0.107308	SC Paderborn 07 / RCD Espanyol	Italy / Spain	2000 / 500
8358	N. Curcija / A. Assal	0.099610	Le Havre AC / AFC Wimbledon	Canada / England	500 / 500
8265	J. Hawkins / M. Cox	0.095535	FC Ingolstadt 04 / AFC Wimbledon	United States / England	950 / 500
8266	T. Booth / A. Boutoutaou	0.082678	Bayern München II / Valenciennes FC	United States / France	800 / 600
7401	C. Riad / D. Waidner	0.078045	CE Sabadell FC / Bayern München II	Morocco / Germany	500 / 650
8262	T. Rushesha / I. DoukourÅ	0.070157	Swansea City / Valenciennes FC	Wales / France	500 / 500

A mesure que le score de propension décroît, la qualité de l'appariement devient médiocre. Les exemples que nous venons d'afficher nous montrent que l'appariement sur le score de propension dépend fortement de la valeur du score.

Voyons maintenant les caractéristiques des joueurs après l'appariement.

## 7.5 Caractéristiques moyennes des joueurs natifs et étrangers après appariement

La taille de l'échantillon a été réduite de 18719 observations à 17184 car 1535 joueurs natifs ont été retirés du jeu de données car ne pouvant pas être apparié à un joueur étranger. Il y a donc maintenant 8592 observations dans chacun des groupes. C'est ce que nous résumons dans le tableau suivant.

	Natif	Etranger	p.value
n	8592	8592	
age (mean (SD))	25.20 (4.69)	25.74 (4.44)	<0.001
height_cm (mean (SD))	181.07 (6.73)	181.53 (6.91)	<0.001
overall (mean (SD))	65.25 (6.42)	67.09 (7.16)	<0.001
potential (mean (SD))	70.71 (5.93)	71.53 (6.44)	<0.001
weight_kg (mean (SD))	74.79 (6.92)	75.59 (7.15)	<0.001
player_european (mean (SD))	0.43 (0.49)	0.27 (0.44)	<0.001
international_reputation (mean (SD))	1.07 (0.30)	1.13 (0.44)	<0.001
weak_foot (mean (SD))	2.92 (0.64)	2.98 (0.69)	<0.001
league1 (mean (SD))	0.78 (0.41)	0.85 (0.36)	<0.001
european_club (mean (SD))	0.43 (0.49)	0.38 (0.48)	<0.001
wage_eur (mean (SD))	7270.16 (14833.48)	11568.70 (24702.35)	<0.001
skill_moves (mean (SD))	2.33 (0.73)	2.46 (0.79)	<0.001
Dim.1 (mean (SD))	-0.11 (4.22)	0.53 (4.17)	<0.001
Dim.2 (mean (SD))	-0.04 (2.05)	0.05 (2.16)	0.004
Dim.3 (mean (SD))	-0.09 (1.59)	0.28 (1.65)	<0.001
Dim.4 (mean (SD))	0.02 (1.25)	-0.06 (1.35)	<0.001

Nous voyons que la différence de salaire entre les deux groupes est passée de 5154 à 4299. Toutefois nous pouvons imputer une partie de la différence de salaire entre les deux groupes à la monnaie dans laquelle est exprimée le salaire des joueurs. En effet, les salaires des joueurs sont exprimés en euros sans prendre en compte la parité du pouvoir d'achat. Il serait donc pertinent de soit calculer les salaires par parité du pouvoir d'achat soit calculer la différence des salaires entre les étrangers et les natifs conditionnellement à la situation géographique des clubs qui engagent les joueurs.

Nous optons pour la seconde option qui est simple à implémenter. Nous allons appliquer un filtre sur la nationalité des clubs pour ne nous limiter qu'aux clubs des pays européens que sont la France, l'Angleterre, l'Allemagne, l'Italie, l'Espagne, la Belgique et le Portugal. Bien que l'Angleterre n'utilise pas l'euro nous estimons que le niveau de salaire des Anglais est comparable à celui des autres pays filtres.

En appliquant le filtre sur les données, quelques propriétés intéressantes ressortent. L'intuition que nous venons d'avoir quant à l'écart des salaires qui devrait être plus faible compte tenu de la situation géographique s'est avérée fausse car nous observons le contraire. Les joueurs étrangers jouant dans les clubs européens gagnent beaucoup plus que leurs coéquipiers

européens natifs des clubs dans lesquels ils jouent. Cette différence de salaire s'élève cette fois à 10109.

Les joueurs recrutés dans les clubs européens ont sensiblement le même âge qu'ils soient étrangers ou natifs du club. Il n'y a non plus de différences en termes de taille. Pour ce qui concerne les caractéristiques techniques nous observons que les clubs européens recrutent des joueurs étrangers qui ont un profil d'attaquant pour combler le manque des joueurs ayant la même nationalité que le club (la variable synthétique Dim.1). Toutefois nous voyons qu'en proportion nous observons un nombre sensiblement élevé de joueurs ayant le profil d'attaquant en Europe.

Nous résumons les différences de caractéristiques dans le tableau suivant.

	Natif	Etranger	p.value
n	3493	3122	
age (mean (SD))	25.25 (4.45)	25.19 (4.27)	0.580
height_cm (mean (SD))	182.54 (6.37)	182.55 (6.84)	0.986
overall (mean (SD))	67.29 (6.70)	69.52 (7.50)	<0.001
potential (mean (SD))	72.53 (6.09)	74.65 (6.28)	<0.001
weight_kg (mean (SD))	75.67 (6.74)	76.28 (7.16)	<0.001
player_european (mean (SD))	0.65 (0.48)	0.37 (0.48)	<0.001
international_reputation (mean (SD))	1.12 (0.40)	1.26 (0.60)	<0.001
weak_foot (mean (SD))	3.02 (0.64)	3.03 (0.69)	0.473
league1 (mean (SD))	0.47 (0.50)	0.59 (0.49)	<0.001
european_club (mean (SD))	0.65 (0.48)	0.66 (0.47)	0.337
wage_eur (mean (SD))	12612.80 (21143.94)	22721.80 (37110.87)	<0.001
skill_moves (mean (SD))	2.43 (0.76)	2.57 (0.84)	<0.001
Dim.1 (mean (SD))	0.44 (4.20)	1.13 (4.20)	<0.001
Dim.2 (mean (SD))	-0.01 (2.14)	0.01 (2.27)	0.734
Dim.3 (mean (SD))	0.31 (1.60)	0.65 (1.72)	<0.001
Dim.4 (mean (SD))	0.01 (1.32)	0.01 (1.38)	0.986

Bien que nous ayons apparié les joueurs à l'aide du score de propension nous observons qu'il existe toujours des différences significatives entre les joueurs étrangers et les joueurs natifs. Le salaire est toujours plus élevé pour les joueurs étranger. Nous aimerons maintenant étudier modéliser les composantes qui entrent dans l'écart du salaire. Nous avons vu qu'en économie, pour affirmer qu'il existe une discrimination nous devons être en mesure de comparer les joueurs sur une base de *toutes choses égales par ailleurs*. Or nous voyons à l'aide de l'appariement qu'il existe des différences significatives entre les joueurs étrangers et les joueurs natifs qui expliquent l'écart du salaire.

Nous voulons maintenant décomposer cette différence de salaire en plusieurs facteurs : quelle est la part du salaire qui est dûe exclusivement aux caractéristiques individuelles des joueurs. Quelle est aussi la part de l'écart du salaire que nous pouvons expliquer grâce aux variables observables telles que l'âge, la taille etc... et aussi quelle est la part de l'écart du salaire que l'interaction entre les dotations individuelles et les coefficients expliquent.

Nous allons pouvoir répondre à cette série de question à l'aide de la décomposition dite de Blinder-Oaxaca.

## 8 Modélisation économétrique de la discrimination

### 8.1 La décomposition de Blinder-Oaxaca

#### 8.1.1 Fonctionnement de la méthode

La décomposition de Blinder-Oaxaca du nom de ces deux auteurs Alan Blinder et Ronald Oaxaca propose une décomposition des écarts constatés entre deux populations (hommes et femmes par exemple) en une part expliquée par les caractéristiques observables de ces deux groupes, c'est-à-dire un effet de composition, et en une part inexpliquée. Isoler cet écart inexpliquée permet de mettre en avant d'éventuels phénomènes discriminatoires, sous certaines hypothèses.

Nous souhaitons utiliser cette méthode pour décomposer l'écart de salaire que nous observons entre les joueurs étrangers et natifs du club dans lequel ils jouent.

#### 8.1.2 Application de la méthode aux écarts de salaires des footballeurs

A l'aide du package `oaxaca` nous estimons décomposons les écarts du salaire en trois étapes : la part de l'écart imputable aux caractéristiques individuelles des joueurs (les dotations propres, aussi appelées *endowments*), l'apart du salaire imputable aux coefficients puis la part de l'écart du salaire imputable aux interactions entre les dotations propres et les coefficients.

**8.1.2.1 Pourcentage des écarts de salaire** Les résultats de cette décomposition indiquent que 31 % des différences de salaires entre les joueurs étrangers et les joueurs natifs s'expliquent par les caractéristiques individuelles des joueurs. NA % de la différence dans les salaires s'expliquent par les coefficients des variables explicatives. Cela signifie que NA % du salaire s'explique par le poids des variables dans la composition des salaires. L'interaction entre les dotations individuelles et le poids des variable explique quant à elle 3 % de l'écart de salaire entre les deux groupes.

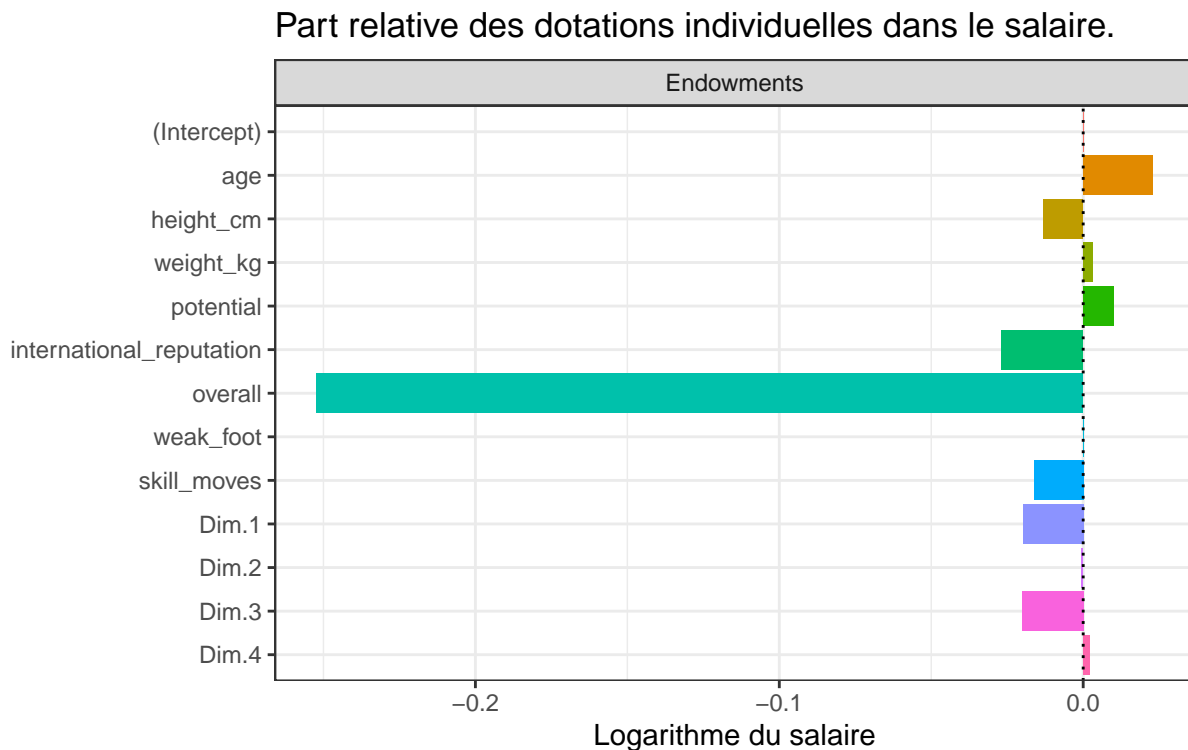
C'est ce que nous résumons dans le tableau suivant en gardant le signe : étant donné que la variable `étranger` prend la valeur 1 si le joueur est effectivement étranger alors le signe négatif des coefficients explique le gain en moins des joueurs natifs.

	% de différence
<code>coef(endowments)</code>	-31.05653
<code>coef(coefficients)</code>	-2.77417
<code>coef(interaction)</code>	2.78946

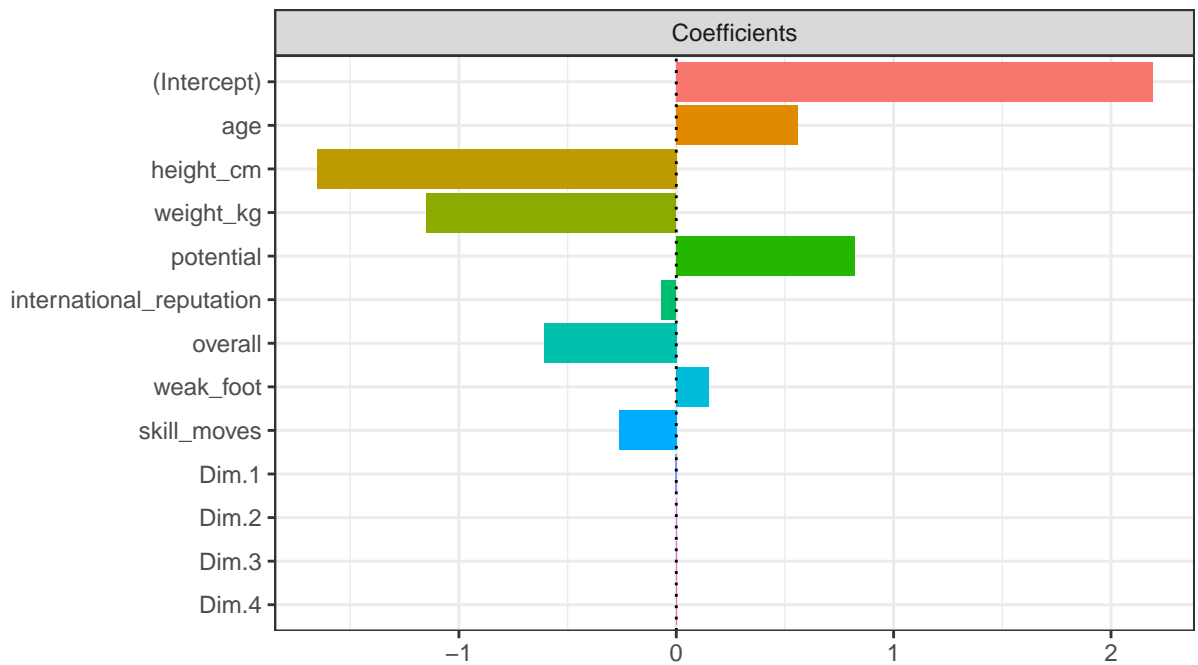
**8.1.2.2 Les écarts imputables aux dotations individuelles** Nous quantifions les écarts de salaires entre les joueurs étrangers et les natifs dus aux dotations individuelles. Les

dotations individuelles représentent les caractéristiques uniques à un groupe. Si les joueurs étrangers sont rémunérés à causes de leurs dotations individuelles, c'est que les joueurs natifs ne possèdent pas ces caractéristiques.

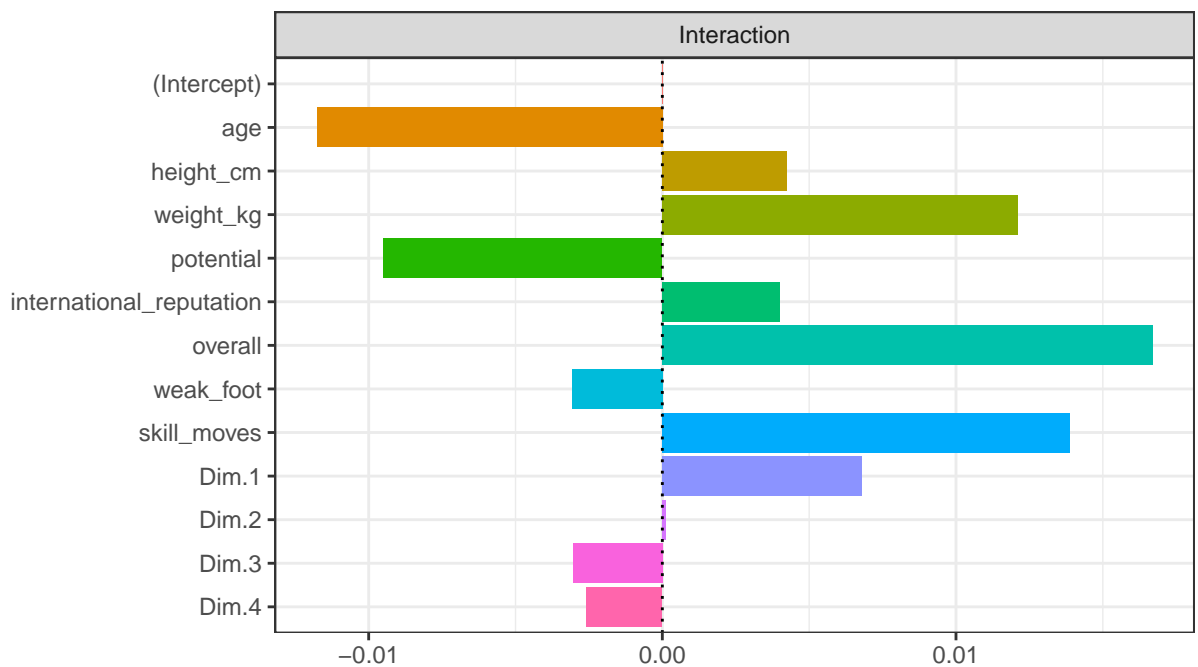
A l'aide du package `oaxaca` nous pouvons faire une représentation des dotations individuelles des joueurs. Le graphique suivant représente la part relative des dotations individuelles dans la composition des salaires. L'axe des abscisses représente la perte ou le gain relatif de salaire des natifs. Nous observons qu'en moyenne 25% des écarts de salaires est expliqué par les performances globales (overall) des joueurs. Cette dotation joue en défaveur des joueurs natifs qui perçoivent l'absence de cette compétence comme une pénalité de 25% sur leur salaire. Près de 2,5% des écarts s'expliquent par la réputation internationale.



**8.1.2.3 Les écarts imputables aux coefficients** Les écarts dus aux coefficients expriment la différence dans le revenu due aux coefficients, c'est-à-dire l'importance de chaque variable dans la composition du salaire. Dans le graphique suivant, le coefficient de la constante explique une part importante des écarts du salaire. En général lorsque la constante prend autant d'ampleur on suspecte un biais de la variable omise car la constante doit être prise telle quelle dans l'équation du salaire. C'est la part de l'écart non explicable par les variables explicatives.



**8.1.2.4 Les écarts imputables à l'interaction des dotations individuelles et des coefficients** Il s'agit ici des écarts de revenus dus à l'interaction entre les dotations individuelles et les coefficients.



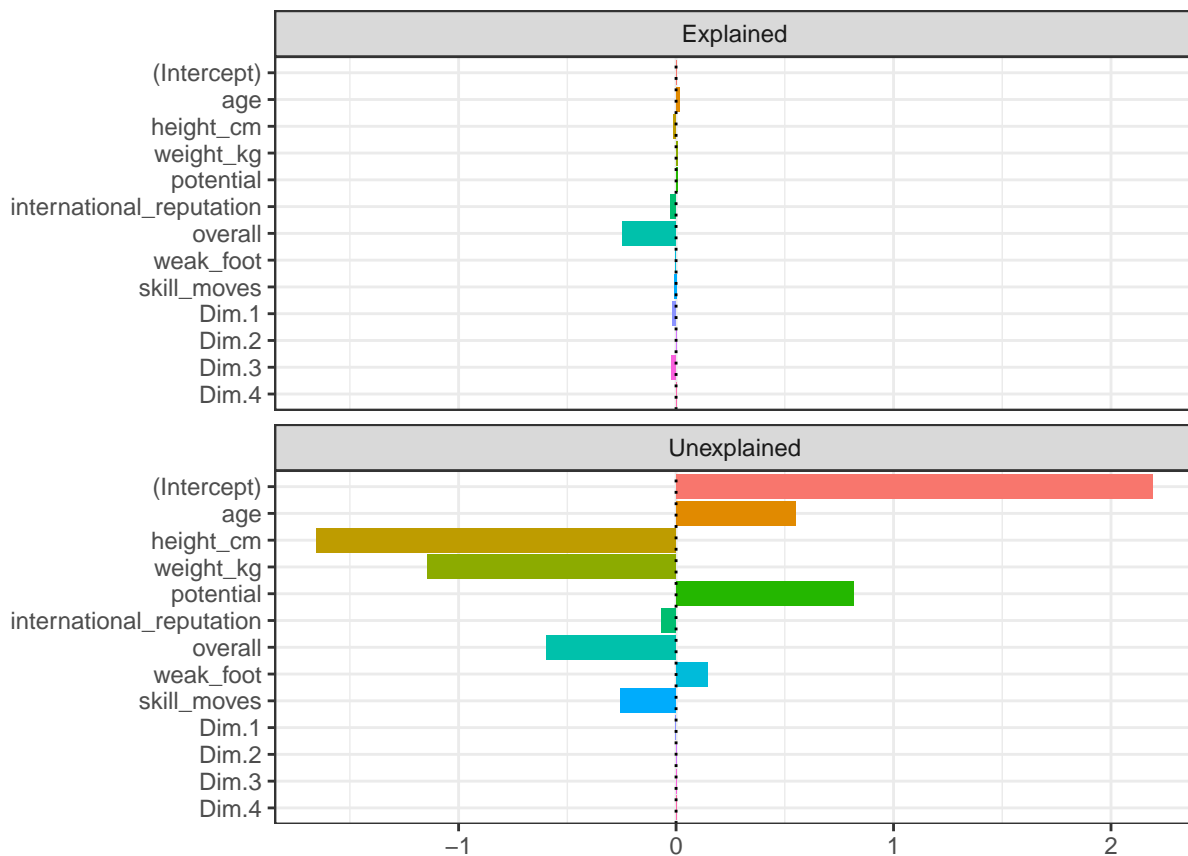


### 8.1.3 La décomposition en deux facteurs

Nous décomposons maintenant l'écart dans les salaires en deux catégories : les écarts expliqués et les écarts non expliqués.

Dans les deux graphiques suivants, nous affichons les variables qui expliquent positivement et négativement l'écart du revenu entre les étrangers et les natifs. Nous voyons par exemple la performance globale (**overall**) explique le fait que les joueurs natifs gagnent moins que les joueurs étrangers. Toutes les autres variables ont une contribution moindre dans l'explication de l'écart des salaires.

Pour ce qui en est des écarts non expliqués, nous voyons sans doute la constante avoir le coefficient le plus élevé. La longueur du diagramme exprime le manque d'intérêt de la variable dans l'explication des salaires.



## 9 Conclusion

Le football demeure aujourd'hui le sport le plus médiatisé de la planète. Les clubs de football génèrent beaucoup d'argent même quand la conjoncture économique n'est pas bonne. La rémunération des joueurs en club est fonction de plusieurs paramètres tels que la réputation du joueur concerné, sa performance globale et ses aptitudes techniques. Cette étude nous a permis de rendre compte de l'écart significatif des salaires entre les joueurs étrangers et les joueurs natifs d'un club de football. En moyenne un joueur étranger gagne 200 % plus qu'un joueur natif. Nous avons vu que cet écart de salaire n'est pas dû au hasard mais le niveau de salaire plus élevé des joueurs étrangers est la rémunération des compétences plus élevées de ces joueurs. Il nous est donc difficile de conclure à une discrimination dans le monde du football professionnel car l'écart des salaires s'explique bien par les compétences. A niveau de compétence égale les joueurs de football gagnent un salaire équivalent. Il existe un écart de salaire très élevé au sein de chaque groupe. La différence de salaire entre les joueurs natif est très élevé et beaucoup plus élevé encore au sein des joueurs étrangers. Cette situation montre que la rémunération dans le football ne tient pas à la nationalité mais à l'image même du joueur.

### 9.1 Limites de l'étude

#### 9.1.1 Absence dans la base de données du revenu des clubs

Le revenu des club nous aurait permis de relativiser l'écart des salaires. Nous pensons que le revenu du joueur est une fonction du revenu du club. Avoir cette variable dans la base de données nous aurait permis de calculer le revenu marginal du joueur par rapport au revenu du club.

#### 9.1.2 Absence de la composition du salaire

La variable salaire qui a servi de variable endogène dans notre étude est une donnée dont nous ignorons la composition. Nous ne savons pas si ce salaire inclut également les droits à l'image du joueur. Nous avons vu brièvement dans le modèle économique des clubs de football qu'une partie importante du revenu des club provient du sponsoring, c'est-à-dire du fait que le club vend à des entreprises le droit d'utiliser son image. Or il se trouve que très souvent les joueurs du club font partie de la communication commerciale. Si le salaire mensuel des joueurs inclut les versements du droit à l'image alors les estimations que nous avons faites devraient être ajustées en prenant en compte le poids des revenus issus du sponsoring dans le salaire, ce qui diminuerait le coefficient de certaines variables comme la réputation internationale qui contribue beaucoup dans l'équation estimée du salaire.

## 10 Bibliographie

Andreff, Wladimir. « Le modèle économique du football européen », Pôle Sud, vol.47, no. 2, 2017, pp.41-59.

Daniel E. Ho, Kosuke Imai, Gary King, Elizabeth A. Stuart (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software, Vol. 42, No. 8, pp. 1-28. URL <https://www.jstatsoft.org/v42/i08/>

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.5. <https://CRAN.R-project.org/package=dplyr>

Havet, Nathalie, et Catherine Sofer. « Les nouvelles théories économiques de la discrimination », Travail, genre et sociétés, vol.7, no. 1, 2002, pp. 83-115.

Hlavac, Marek (2018). oaxaca: Blinder-Oaxaca Decomposition in R. R package version 0.1.4. <https://CRAN.R-project.org/package=oaxaca>

Kazuki Yoshida and Alexander Bartel (2020). tableone: Create ‘Table 1’ to Describe Baseline Characteristics with or without Propensity Score Weights. R package version 0.12.0. <https://CRAN.R-project.org/package=tableone>

Llorca, Matthieu, et Thierry Teste. « Dépenses salariales et performance dans l’industrie du football », Revue française d’économie, vol.volume xxxi, no. 2, 2016, pp. 125-145.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Vie publique.fr.(2013, 7 août).Rapport d’information déposé (. . .) par la commission des affaires culturelles et de l’éducation sur le fair-play financier européen et son application au modèle économique des clubs de football professionnel français. <https://www.vie-publique.fr/rapport/33372-modeles-economiques-des-clubs-de-football-professionnel-francais>

# 11 Annexes

## 11.1 Code source du projet

Le projet a été entièrement rédigé en Rmarkdown, ce qui permet de reproduire tous les résultats obtenus. Une copie de la rédaction, de la base de données et des scripts se trouve sur ma page GitHub : <https://github.com/agailoty/football-discrimination>

## 11.2 Transformations des données

Il existe beaucoup de colonnes dans le jeu de données qui peuvent être intéressant mais leur nombre est trop nombreux. Il est préférable donc de résumer ces colonnes en des variables synthétiques.

Le traitement dans ce projet est le fait que le joueur a la même nationalité que le club dans lequel il joue. L'idée serait de créer une variable qui indique si oui ou non le joueur a la nationalité du club. La difficulté ici réside dans le fait que la nationalité des clubs n'est pas clairement indiquée dans la base de données. Toutefois il existe une colonne qui montre le nom complet du league.

### 11.2.1 Script R utilisé pour nettoyer les données

```
library(readr)
library(dplyr)

source("nationality.R")

output_path <- "/final"

# Importing the data and merging
fifa21 <- read_csv("players_21.csv")
league_nationality <- strsplit(fifa21$league_name, split = " ")

league_nat <- tolower(league_nationality %>% sapply(function(x) x[[1]][1]))
country_adj <- read_csv("demonyms.csv")
colnames(country_adj) <- c("adjective", "country")
country_adj$adjective <- tolower(country_adj$adjective)
country_adj$country <- tolower(country_adj$country)

find_country <- function(adj) {
  ladj <- tolower(adj)
  if (! ladj %in% country_adj$adjective) {
    if (ladj %in% country_adj$country) {
      tools::toTitleCase(country_adj$country[which(country_adj$country == ladj)][1])
    }
  }
}
```

```

    } else {
      paste0("NF_", adj)
    }
  } else {
    tools::toTitleCase(country_adj$country[which(adj == country_adj$adjective)][1])
  }
}

league_nat_country <- sapply(league_nat, find_country)
fifa21$league_nationality <- league_nat_country

# Player nationality == club nationality ?

fifa21$foreigner <- ifelse(fifa21$nationality == fifa21$league_nationality, 0, 1)

write.csv(x = fifa21, file = paste0(output_path, "/fifa21.csv"), row.names = FALSE)

```

### 11.2.2 Script utilisé pour factoriser les variables caractéristiques techniques des joueurs en 4 composantes principales

```

source("helper.R")

library(dplyr)
library(FactoMineR)

fifa21 <- read.csv("final/fifa21.csv")
pca_data <- fifa21[, columns_for_pca]
pca_data$short_name <- make.unique(fifa21$short_name) # Make sure all names are unique'
row.names(pca_data) <- pca_data$short_name
player_names <- pca_data$short_name

pca_data <- subset(pca_data, select = - c(short_name, defending_marking))

acp <- PCA(pca_data, scale.unit = TRUE, ncp = 10, graph = FALSE)

players <- data.frame(acp$ind$coord)

# Combine the PCA coordinates with the main dataset and remove the columns used in the

players <- cbind(fifa21[, columns_to_keep], players[, 1:4])

# Keep only clubs from these countries : ["Spain", "Italy", "Germany", "France", "Engl
selected_countries <- c("Spain", "Italy", "Germany", "France", "England", "Portugal")

```

```

cleaned_data <- filter(players, league_nationality %in% selected_countries)

# Calculate contract length : date joining and date end contract
library(lubridate)

cleaned_data <- cleaned_data %>%
  mutate(contract_length = difftime(ymd(contract_valid_until, truncated = 2L),
                                     as_date(joined), units = "days")/(365.25/12))

write.csv(players, file = "final/players.csv", row.names = FALSE)
write.csv(cleaned_data, file = "final/cleaned.csv", row.names = FALSE)

```

short_name	nationality	club_name	league_name
L. Messi	Argentina	FC Barcelona	Spain Primera Division
Cristiano Ronaldo	Portugal	Juventus	Italian Serie A
J. Oblak	Slovenia	Atlético Madrid	Spain Primera Division
R. Lewandowski	Poland	FC Bayern München	German 1. Bundesliga
Neymar Jr	Brazil	Paris Saint-Germain	French Ligue 1
K. De Bruyne	Belgium	Manchester City	English Premier League
K. Mbappé	France	Paris Saint-Germain	French Ligue 1
M. ter Stegen	Germany	FC Barcelona	Spain Primera Division
V. van Dijk	Netherlands	Liverpool	English Premier League
Alisson	Brazil	Liverpool	English Premier League

L'idée serait donc de parcourir tous les noms des leagues auxquels appartiennent les clubs, extraire la nationalité des clubs puis trouver le nom du pays à partir de la nationalité du club.

club_name	league_name	league_nationality
FC Barcelona	Spain Primera Division	Spain
Juventus	Italian Serie A	Italy
Atlético Madrid	Spain Primera Division	Spain
FC Bayern München	German 1. Bundesliga	Germany
Paris Saint-Germain	French Ligue 1	France
Manchester City	English Premier League	England
Paris Saint-Germain	French Ligue 1	France
FC Barcelona	Spain Primera Division	Spain
Liverpool	English Premier League	England
Liverpool	English Premier League	England

## 11.3 Analyse en composantes principales

### 11.3.1 1. Observation d'individus extrêmes

L'analyse des graphes ne révèle aucun individu singulier.

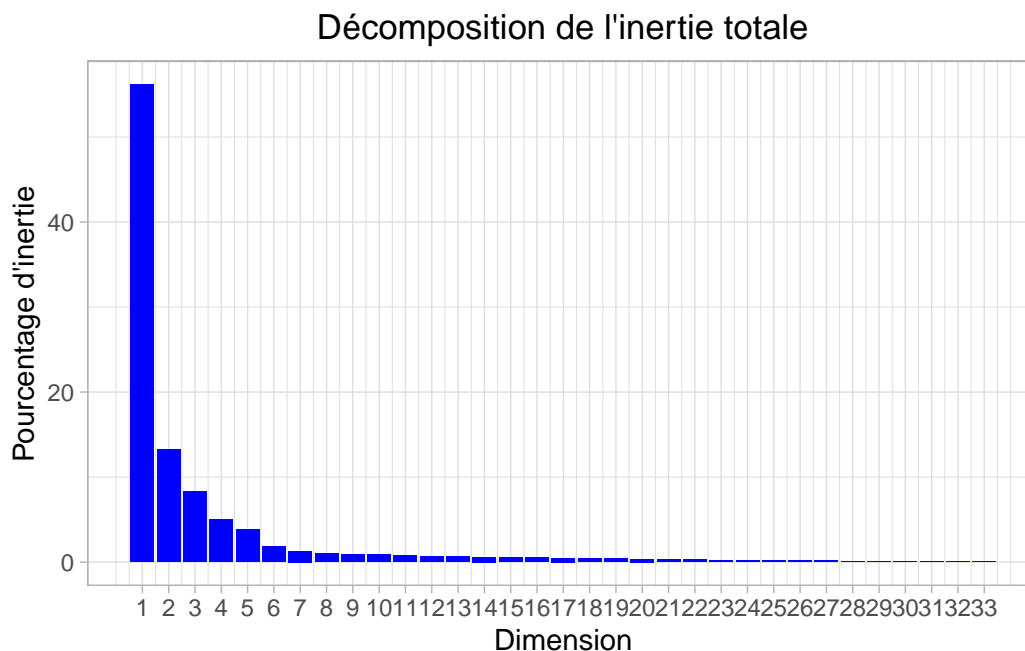
---

### 11.3.2 2. Distribution de l'inertie

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'analyse expriment **69.43%** de l'inertie totale du jeu de données ; cela signifie que 69.43% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage assez important, et le premier plan représente donc convenablement la variabilité contenue dans une grande part du jeu de données actif. Cette valeur est nettement supérieure à la valeur référence de **6.54%**, la variabilité expliquée par ce plan est donc hautement significative (cette inertie de référence est le quantile 0.95-quantile de la distribution des pourcentages d'inertie obtenue en simulant 101 jeux de données aléatoires de dimensions comparables sur la base d'une distribution normale).

Du fait de ces observations, il serait tout de même probablement préférable de considérer également dans l'analyse les dimensions supérieures ou égales à la troisième.



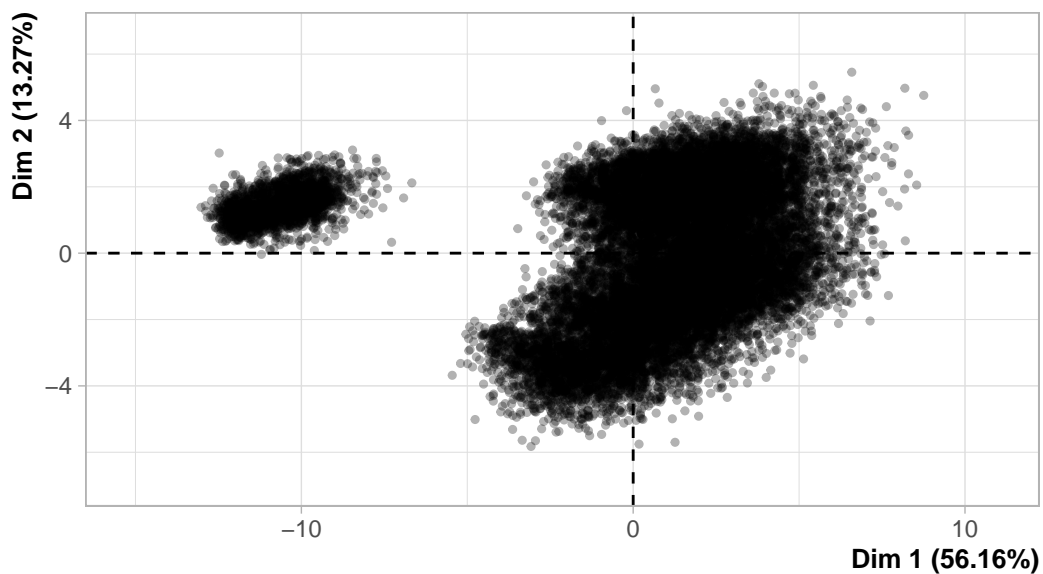
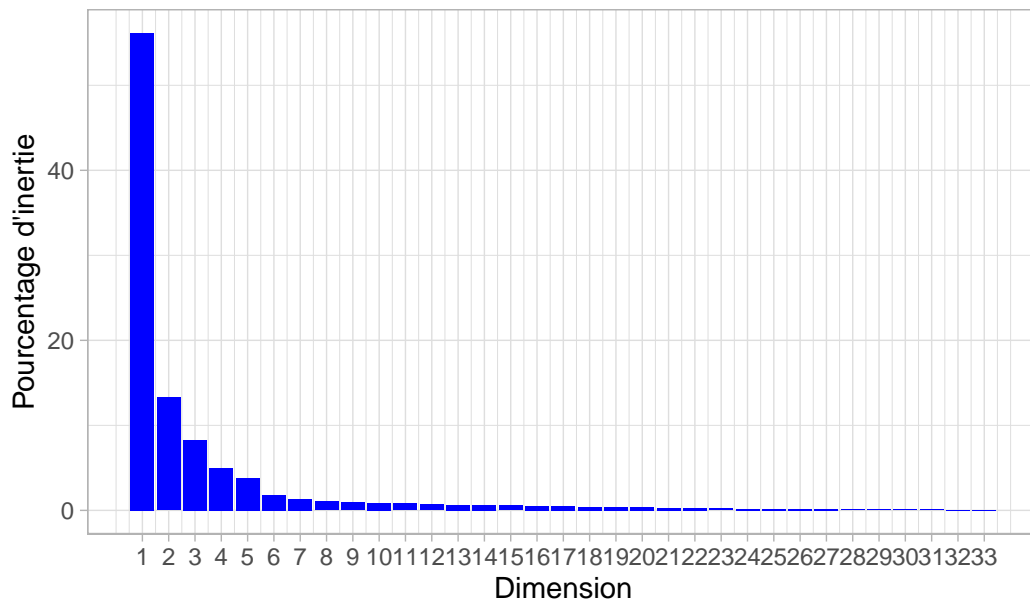
**Figure 2 - Décomposition de l'inertie totale** *Le premier facteur est prépondérant : il explique à lui seul 56.16% de la variabilité totale des données. Il convient de noter que dans un tel cas, la variabilité liée aux autres composantes peut être dénuée de sens, en dépit d'un pourcentage élevé.*

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 5 premiers axes. Ces composantes révèlent un taux d'inertie supérieur à celle du quantile 0.95-quantile de distributions aléatoires (86.52% contre 16.17%). Cette observation suggère que seuls ces axes sont porteurs d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

---

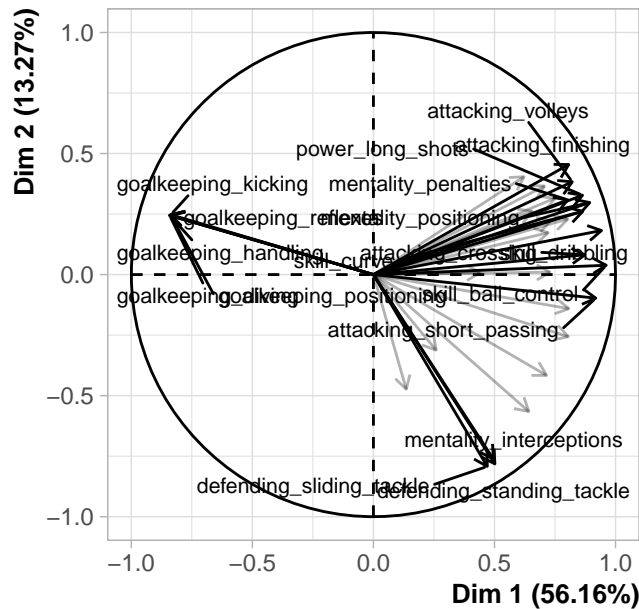
### 11.3.3 3. Description du plan 1:2

Décomposition de l'inertie totale





**Figure 3.1 - Graphe des individus (ACP)** Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.



**Figure 3.2 - Graphe des variables (ACP)** Les variables libellées sont celles les mieux représentées sur le plan.

La **dimension 1** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour des variables telles que *power\_long\_shots*, *attacking\_finishing*, *mentality\_positioning*, *skill\_curve*, *attacking\_volleys*, *skill\_dribbling*, *skill\_fk\_accuracy*, *attacking\_crossing*, *power\_shot\_power* et *mentality\_vision* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *goalkeeping\_diving*, *goalkeeping\_reflexes*, *goalkeeping\_handling*, *goalkeeping\_positioning*, *goalkeeping\_kicking* et *power\_strength* (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour les variables *goalkeeping\_diving*, *goalkeeping\_handling*, *goalkeeping\_reflexes*, *goalkeeping\_kicking* et *goalkeeping\_positioning* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *attacking\_heading\_accuracy*, *skill\_ball\_control*, *power\_stamina*, *attacking\_short\_passing*, *skill\_dribbling*, *mentality\_aggression*, *attacking\_crossing*, *skill\_long\_passing*, *mentality\_positioning* et

*movement\_sprint\_speed* (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour les variables *defending\_sliding\_tackle*, *defending\_standing\_tackle*, *power\_strength*, *mentality\_interceptions*, *attacking\_heading\_accuracy*, *mentality\_aggression*, *power\_jumping* et *power\_stamina* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *power\_shot\_power*, *mentality\_vision*, *attacking\_finishing*, *power\_long\_shots*, *skill\_curve*, *attacking\_volleys*, *skill\_fk\_accuracy*, *movement\_agility*, *mentality\_positioning* et *goalkeeping\_kicking* (de la plus extrême à la moins extrême).

Notons que la variable *skill\_ball\_control* est extrêmement corrélée à cette dimension (corrélation de 0.92). Cette variable pourrait donc résumer à elle seule la dimension 1.

---

La **dimension 2** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (en haut du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (en bas du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour les variables *goalkeeping\_diving*, *goalkeeping\_handling*, *goalkeeping\_reflexes*, *goalkeeping\_kicking* et *goalkeeping\_positioning* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *attacking\_heading\_accuracy*, *skill\_ball\_control*, *power\_stamina*, *attacking\_short\_passing*, *skill\_dribbling*, *mentality\_aggression*, *attacking\_crossing*, *skill\_long\_passing*, *mentality\_positioning* et *movement\_sprint\_speed* (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour des variables telles que *power\_long\_shots*, *attacking\_finishing*, *mentality\_positioning*, *skill\_curve*, *attacking\_volleys*, *skill\_dribbling*, *skill\_fk\_accuracy*, *attacking\_crossing*, *power\_shot\_power* et *mentality\_vision* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *goalkeeping\_diving*, *goalkeeping\_reflexes*, *goalkeeping\_handling*, *goalkeeping\_positioning*, *goalkeeping\_kicking* et *power\_strength* (de la plus extrême à la moins extrême).

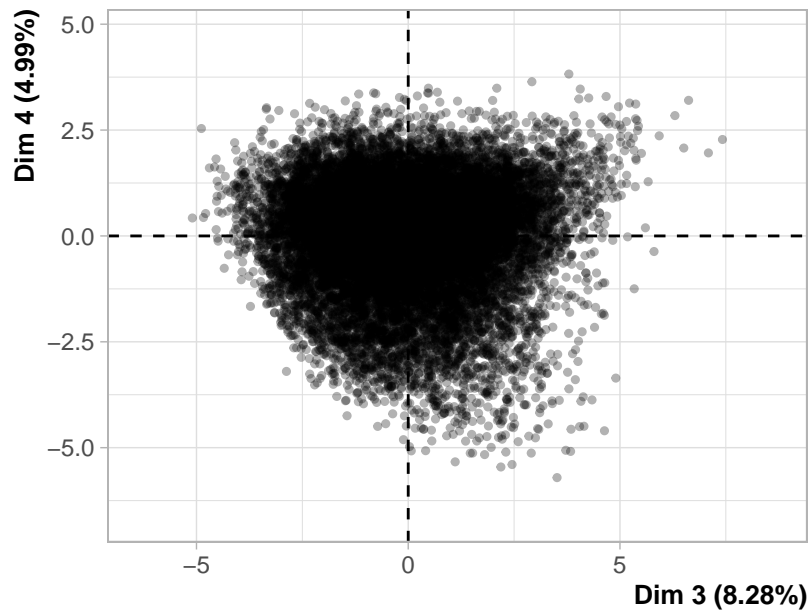
Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour les variables *defending\_sliding\_tackle*, *defending\_standing\_tackle*, *power\_strength*, *mentality\_interceptions*, *attacking\_heading\_accuracy*, *mentality\_aggression*, *power\_jumping* et *power\_stamina* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *power\_shot\_power*, *mentality\_vision*, *attacking\_finishing*, *power\_long\_shots*, *skill\_curve*, *attacking\_volleys*, *skill\_fk\_accuracy*,

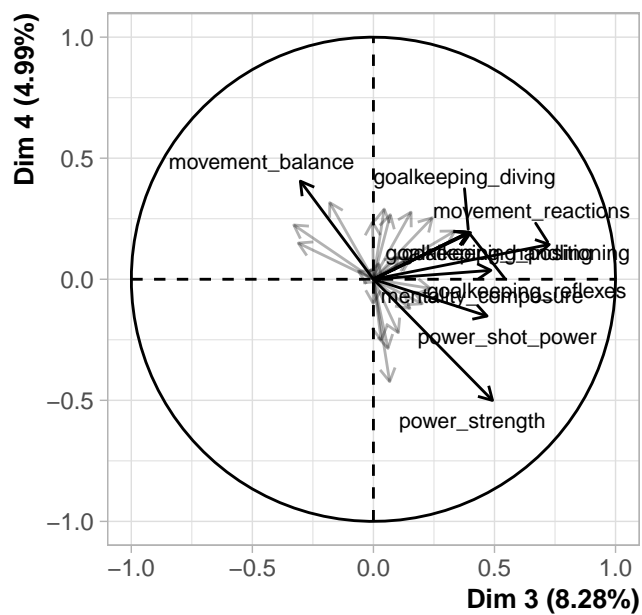
*movement\_agility*, *mentality\_positioning* et *goalkeeping\_kicking* (de la plus extrême à la moins extrême).

---

#### 11.3.4 4. Description du plan 3:4



**Figure 4.1 - Graphe des individus (ACP)** *Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.*



**Figure 4.2 - Graphe des variables (ACP)** *Les variables libellées sont celles les mieux représentées sur le plan.*

---

La **dimension 3** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour des variables telles que *attacking\_finishing*, *mentality\_penalties*, *attacking\_heading\_accuracy*, *power\_strength*, *attacking\_volleys*, *power\_shot\_power*, *power\_long\_shots*, *mentality\_positioning*, *power\_jumping* et *mentality\_composure* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *defending\_sliding\_tackle*, *defending\_standing\_tackle*, *mentality\_interceptions*, *movement\_balance*, *attacking\_crossing*, *skill\_long\_passing*, *movement\_agility*, *movement\_acceleration*, *movement\_sprint\_speed* et *attacking\_short\_passing* (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour des variables telles que *movement\_reactions*, *mentality\_composure*, *skill\_long\_passing*, *mentality\_vision*, *mentality\_interceptions*, *attacking\_short\_passing*, *mentality\_aggression*, *defending\_standing\_tackle*, *defending\_sliding\_tackle* et *power\_shot\_power* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *movement\_sprint\_speed*, *movement\_acceleration*, *movement\_balance* et *movement\_agility* (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour les variables *movement\_balance*, *movement\_acceleration*, *movement\_sprint\_speed*, *movement\_agility*, *defending\_sliding\_tackle*, *defending\_standing\_tackle*, *attacking\_crossing* et *mentality\_interceptions* (de la plus extrême à la moins extrême).
  - de faibles valeurs pour des variables telles que *power\_strength*, *attacking\_heading\_accuracy*, *power\_shot\_power*, *mentality\_penalties*, *power\_long\_shots*, *attacking\_volleys*, *attacking\_finishing*, *mentality\_composure*, *movement\_reactions* et *mentality\_positioning* (de la plus extrême à la moins extrême).
- 

La **dimension 4** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (en haut du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (en bas du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour les variables *movement\_balance*, *movement\_acceleration*, *movement\_sprint\_speed*, *movement\_agility*, *defending\_sliding\_tackle*, *defending\_standing\_tackle*, *attacking\_crossing* et *mentality\_interceptions* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *power\_strength*, *attacking\_heading\_accuracy*, *power\_shot\_power*, *mentality\_penalties*, *power\_long\_shots*, *attacking\_volleys*, *attacking\_finishing*, *mentality\_composure*, *movement\_reactions* et *mentality\_positioning* (de la plus extrême à la moins extrême).

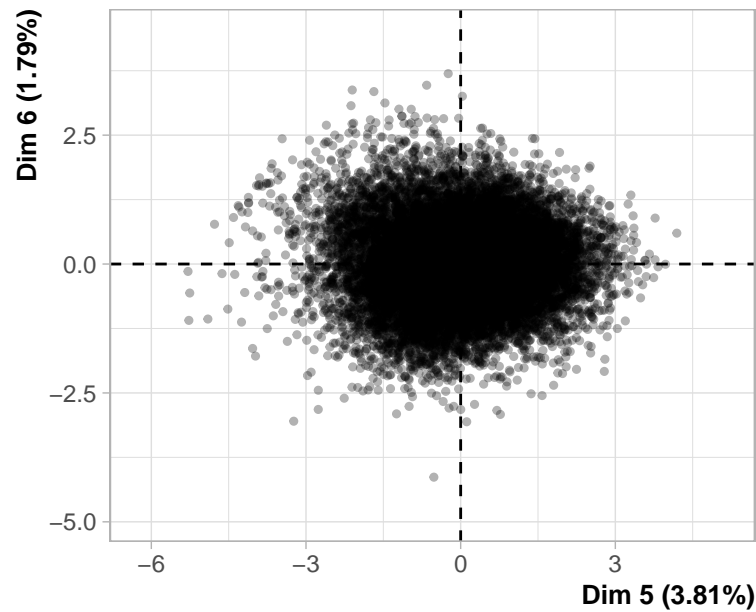
Le groupe 2 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour des variables telles que *movement\_reactions*, *mentality\_composure*, *skill\_long\_passing*, *mentality\_vision*, *mentality\_interceptions*, *attacking\_short\_passing*, *mentality\_aggression*, *defending\_standing\_tackle*, *defending\_sliding\_tackle* et *power\_shot\_power* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *movement\_sprint\_speed*, *movement\_acceleration*, *movement\_balance* et *movement\_agility* (de la plus extrême à la moins extrême).

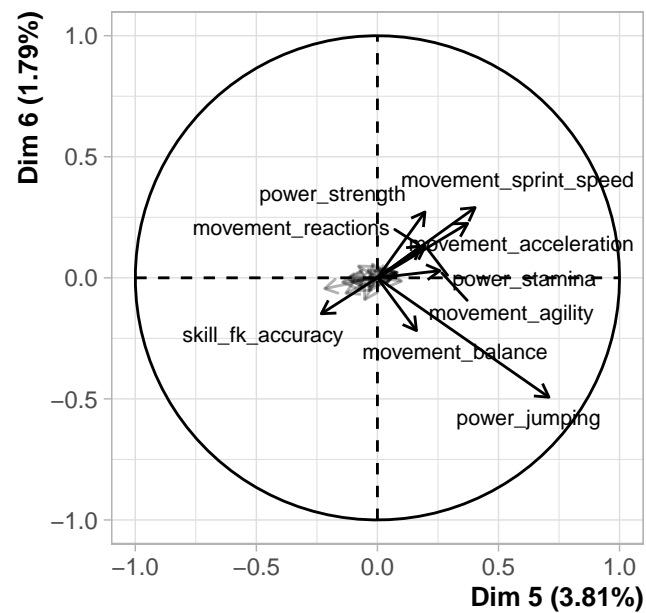
Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour des variables telles que *attacking\_finishing*, *mentality\_penalties*, *attacking\_heading\_accuracy*, *power\_strength*, *attacking\_volleys*, *power\_shot\_power*, *power\_long\_shots*, *mentality\_positioning*, *power\_jumping* et *mentality\_composure* (de la plus extrême à la moins extrême).
  - de faibles valeurs pour des variables telles que *defending\_sliding\_tackle*, *defending\_standing\_tackle*, *mentality\_interceptions*, *movement\_balance*, *attacking\_crossing*, *skill\_long\_passing*, *movement\_agility*, *movement\_acceleration*, *movement\_sprint\_speed* et *attacking\_short\_passing* (de la plus extrême à la moins extrême).
-

### 11.3.5 5. Description de la dimension 5



**Figure 5.1 - Graphe des individus (ACP)** *Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.*



**Figure 5.2 - Graphe des variables (ACP)** *Les variables libellées sont celles les mieux représentées sur le plan.*

La **dimension 5** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour des variables telles que *power\_jumping*, *movement\_sprint\_speed*, *movement\_acceleration*, *movement\_agility*, *power\_stamina*, *movement\_balance*, *attacking\_heading\_accuracy*, *power\_strength*, *mentality\_aggression* et *movement\_reactions* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *skill\_long\_passing*, *skill\_fk\_accuracy*, *attacking\_short\_passing*, *mentality\_vision*, *skill\_curve*, *defending\_standing\_tackle*, *defending\_sliding\_tackle*, *mentality\_interceptions*, *attacking\_crossing* et *power\_long\_shots* (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnée négative sur l'axe) partage :

- de fortes valeurs pour des variables telles que *skill\_long\_passing*, *skill\_fk\_accuracy*, *attacking\_short\_passing*, *mentality\_vision*, *skill\_curve*, *defending\_standing\_tackle*, *defending\_sliding\_tackle*, *power\_long\_shots*, *mentality\_penalties* et *mentality\_interceptions* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *movement\_sprint\_speed*, *movement\_acceleration*, *power\_jumping*, *power\_stamina*, *movement\_agility*, *power\_strength*, *attacking\_heading\_accuracy*, *movement\_reactions*, *movement\_balance* et *mentality\_aggression* (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisés par une coordonnée négative sur l'axe) partage :

- de fortes valeurs pour les variables *movement\_sprint\_speed*, *movement\_acceleration*, *power\_strength*, *power\_stamina*, *attacking\_crossing*, *skill\_dribbling* et *movement\_reactions* (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que *power\_jumping*, *movement\_balance*, *mentality\_penalties*, *skill\_fk\_accuracy*, *attacking\_heading\_accuracy*, *power\_long\_shots*, *attacking\_volleys*, *power\_shot\_power*, *skill\_curve* et *mentality\_aggression* (de la plus extrême à la moins extrême).