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

Axel-Cleris Gailloty

2021-06-28

Table of Contents

En moyenne un joueur de football quelconque 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 séparons calculons les moyennes 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. En apparence nous pouvons avancer la raison que les joeurs étrangers sont discriminés positivement parce que les clubs de football y trouvent un intérêt et que cela est une caractéristique du modèle économique du football. Nous pouvons aussi penser que ce fort écart de salaire est dû aux performances des joueurs de football étrangers qui sont en moyenne supérieures à celles des joueurs ayant la même nationalité que le club et ce un salaire plus élevé permet de compenser le différentiel positif de performances. Dans ce cas nous serions en raison de nous demander quelle est la rémunération marginale d’une unité de performance ?  
Penser ainsi suppose que les salaires des footballeurs sont normalement distribués (avec un salaire moyen et un écart-moyen entre les salaires constants).  
Ou simplement nous pouvons dire que l’écart des salaires contastés est dû à une répartition très inégale des salaires en sorte que seulement quelques joueurs arrivent à eux seuls expliquer une grande partie de la différence. Se poser ces questions énonce inévitablement la question de la discrimination et du modèle économique du football.

Introduction

# 1 Spécification technique

Les questions que nous nous sommes posées en préambule de ce travail nous forcent à réflechir à des outils mathématiques, statistiques et économétriques qui permettent de répondre avec rigueur aux intérrogations. Dans un premier temps, nous estimons nécessaires de nous orienter vers une approche orientée données afin de répondre efficacement aux questions. 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 ici. D’où l’importance de faire un travail de collecte de données sur la question. Les données doivent permettre de : - Réaliser des statistiques descriptives : calculer les moyennes par groupes

Nous aurons besoin d’outils de statistiques inférentielles pour tester les hypothèses que nous avons énoncées. 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 voudrons 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’échantillonage qui nous permettent de comparer les joeurs 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 statitistique. 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.

Ensuite nous allons nous concentrer

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 dimensionalité (ACP). Ces étapes qui ont été très utiles à la formation de la problématique et qui ont pris environ 60% du temps total du travail sont placés en annexes du document. Nous y ferons allusion dans le travail.  
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 varité 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. Le cas typique souvent employé pour illustrer la notion de discrimination en économie est celui de l’écart de salaire entre les hommes et les femmes. Un écart de salaire qui se justifie par un choix d’investissement prioritaire dans la vie familiale pour la femme qui se traduit par un faible investissement dans l’éducation n’est pas une discrimination en économie. On parle de discrimination économique 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 selon le sexe. 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 : le sexe si on veut étudier étudier s’il y a discrimination entre les hommes etles femmes. C’est ce qu’on appelle aussi la condition “*ceteris paribus*”, toute chose égale par ailleurs.

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. Comment donc trouver des groupes d’individus comparables ?

Nous utilisons la méthode d’échantillonage sur le score de propension.

## 3.2 Les méthodes de mesure de discrimination

Plusieurs méthodes économétriques existent pour mesurer la discrimination en économie. Une première méthode que nous pouvons présenter est celle de la régression par quantile. La logique derrière la régression par quantile est simple si on est familier avec la méthode de régression linéaire. La régression par quantile est une méthode de régression pour estimer ces fonctions quantiles conditionnelles. Tout comme la régression linéaire estime la fonction moyenne conditionnelle comme une combinaison linéaire des prédicteurs, la régression quantile estime la fonction quantile conditionnelle comme une combinaison linéaire des prédicteurs.

# 4 Le modèle économique du football

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.

En fait si l’Arrêt Bosman est un tel tremblement de terre dans le monde du football c’est pour une autre raison. Avant 1995 les clubs pouvaient pas compter plus de trois joueurs d’une nationalité différente que celle donc du club en question.

La décision rendue par la Cour de Justice Européenne, le célèbre “Arrêt Bosman” 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.

# 5 Applications

## 5.1 Analyse des données

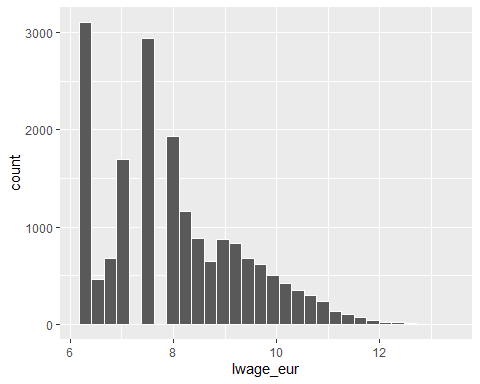
Pour illustrer la discrimination dans le monde de football nous avons choisi d’utiliser les données officielles des joueurs de football publiées par FIFA. Les données ne proviennent pas directement de FIFA mais du site [Sofifa](https://sofifa.com/). La recolte des données s’est faite grâce à un programme informatique d’extraction de contenu web.

Pour répondre à la problématique de ce travail nous aurons besoin d’échantillonner les données. Toutefois pour ce qui concerne les caractéristiques des joueurs, nous nous limitons seulement aux plus pertinentes. Nous avons un large éventail de choix car les caractéristiques des joueurs comprennent les performances techniques des joueurs, leur réputation internationale, la valeur et le salaire des joueurs en euro, le club dans lequel ils jouent, la date où le joueur a rejoint le club, la durée de leur contrat …

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’homogéinété.

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 |
| Valeur du joueur | Le montant auquel le club est prêt à céder le joueur |
| Position du joueur en sélection national | A quel poste le joueur joue dans sa sélection nationale. |
| Position du joueur dans le club | A quel poste le joueur joue dans le club |
| Ancienneté | Ancienneté du joueur dans le club en année |



# 6 Estimations économétriques sur toutes les données

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

Call:  
lm(formula = lwage\_eur ~ age + height\_cm + weight\_kg + overall +   
 potential + international\_reputation + weak\_foot + league1 +   
 european\_club + Dim.1 + Dim.2 + Dim.3 + Dim.4, data = cleaned\_data)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-4.6800 -0.4193 0.0852 0.5604 2.5357   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -2.669147 0.326655 -8.171 3.25e-16 \*\*\*  
age -0.031245 0.002938 -10.636 < 2e-16 \*\*\*  
height\_cm 0.020377 0.001748 11.655 < 2e-16 \*\*\*  
weight\_kg -0.010868 0.001589 -6.841 8.09e-12 \*\*\*  
overall 0.131301 0.002789 47.080 < 2e-16 \*\*\*  
potential -0.007606 0.002486 -3.060 0.002217 \*\*   
international\_reputation 0.416106 0.020453 20.344 < 2e-16 \*\*\*  
weak\_foot 0.019579 0.010572 1.852 0.064056 .   
league1 0.012987 0.015765 0.824 0.410045   
european\_club 0.038013 0.013708 2.773 0.005559 \*\*   
Dim.1 0.034414 0.002371 14.511 < 2e-16 \*\*\*  
Dim.2 0.012027 0.003530 3.407 0.000659 \*\*\*  
Dim.3 0.058189 0.007890 7.375 1.71e-13 \*\*\*  
Dim.4 0.002824 0.006709 0.421 0.673812   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.8824 on 18705 degrees of freedom  
Multiple R-squared: 0.5837, Adjusted R-squared: 0.5834   
F-statistic: 2017 on 13 and 18705 DF, p-value: < 2.2e-16

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.

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.

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 supplémntaires. 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 multicolinéarité 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 joeurs 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 aggressifs 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 sur 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 :

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 donnée 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 transcofidification 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.

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 impressionantes, 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 |  |
| player\_european (mean (SD)) | 0.48 (0.50) | 0.27 (0.44) | <0.001 |  |
| lwage\_eur (mean (SD)) | 7.83 (1.27) | 8.27 (1.43) | <0.001 |  |
| weak\_foot (mean (SD)) | 2.90 (0.64) | 2.98 (0.69) | <0.001 |  |
| international\_reputation (mean (SD)) | 1.06 (0.28) | 1.13 (0.44) | <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.

Call:  
lm(formula = lwage\_eur ~ foreigner + age + height\_cm + weight\_kg +   
 overall + potential + international\_reputation + weak\_foot +   
 league1 + european\_club + Dim.1 + Dim.2 + Dim.3 + Dim.4,   
 data = cleaned\_data)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-4.6845 -0.4204 0.0839 0.5598 2.5455   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -2.638197 0.326918 -8.070 7.46e-16 \*\*\*  
foreigner 0.030155 0.013576 2.221 0.026351 \*   
age -0.030813 0.002944 -10.467 < 2e-16 \*\*\*  
height\_cm 0.020176 0.001751 11.526 < 2e-16 \*\*\*  
weight\_kg -0.010949 0.001589 -6.891 5.72e-12 \*\*\*  
overall 0.130671 0.002803 46.618 < 2e-16 \*\*\*  
potential -0.007147 0.002494 -2.866 0.004167 \*\*   
international\_reputation 0.414936 0.020458 20.282 < 2e-16 \*\*\*  
weak\_foot 0.019249 0.010572 1.821 0.068661 .   
league1 0.008023 0.015921 0.504 0.614320   
european\_club 0.041554 0.013799 3.011 0.002604 \*\*   
Dim.1 0.034069 0.002376 14.337 < 2e-16 \*\*\*  
Dim.2 0.011672 0.003534 3.303 0.000958 \*\*\*  
Dim.3 0.058043 0.007889 7.357 1.95e-13 \*\*\*  
Dim.4 0.002921 0.006708 0.435 0.663226   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.8823 on 18704 degrees of freedom  
Multiple R-squared: 0.5838, Adjusted R-squared: 0.5835   
F-statistic: 1874 on 14 and 18704 DF, p-value: < 2.2e-16

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.

## 7.1 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 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.

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

L’usage que nous allons faire de l’échantillonage 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’exranéité de telle sorte qu’elle prenne la valeur 1 si le joueur est étranger au club et 0 s’il est natif du club.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 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 |  |
| player\_european (mean (SD)) | 0.43 (0.49) | 0.27 (0.44) | <0.001 |  |
| wage\_eur (mean (SD)) | 7270.16 (14833.48) | 11568.70 (24702.35) | <0.001 |  |
| international\_reputation (mean (SD)) | 1.07 (0.30) | 1.13 (0.44) | <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 |  |

Call:  
matchit(formula = comparisonFormula, data = cleaned\_data, method = "nearest")  
  
Summary of Balance for All Data:  
 Means Treated Means Control Std. Mean Diff. Var. Ratio  
distance 0.5085 0.4170 0.6408 0.9832  
age 25.7389 24.7017 0.2334 0.8491  
height\_cm 181.5335 180.8987 0.0919 1.0544  
overall 67.0949 64.3965 0.3768 1.1702  
potential 71.5332 70.7134 0.1272 1.2370  
international\_reputation 1.1328 1.0571 0.1730 2.4770  
weak\_foot 2.9814 2.8971 0.1218 1.1590  
league1 0.8494 0.7010 0.4149 .  
european\_club 0.3771 0.4765 -0.2052 .  
Dim.1 0.5281 -0.4578 0.2362 0.9192  
Dim.2 0.0500 -0.0411 0.0421 1.1333  
Dim.3 0.2810 -0.2545 0.3252 1.0330  
Dim.4 -0.0580 0.0456 -0.0769 1.2188  
 eCDF Mean eCDF Max  
distance 0.1735 0.2480  
age 0.0366 0.1293  
height\_cm 0.0127 0.0473  
overall 0.0578 0.1790  
potential 0.0207 0.0639  
international\_reputation 0.0151 0.0527  
weak\_foot 0.0182 0.0547  
league1 0.1484 0.1484  
european\_club 0.0995 0.0995  
Dim.1 0.0877 0.1398  
Dim.2 0.0203 0.0427  
Dim.3 0.0947 0.1443  
Dim.4 0.0248 0.0470  
  
  
Summary of Balance for Matched Data:  
 Means Treated Means Control Std. Mean Diff. Var. Ratio  
distance 0.5085 0.4524 0.3927 1.2926  
age 25.7389 25.1966 0.1220 0.8998  
height\_cm 181.5335 181.0674 0.0675 1.0554  
overall 67.0949 65.2538 0.2571 1.2440  
potential 71.5332 70.7051 0.1285 1.1799  
international\_reputation 1.1328 1.0666 0.1513 2.1317  
weak\_foot 2.9814 2.9190 0.0902 1.1569  
league1 0.8494 0.7821 0.1881 .  
european\_club 0.3771 0.4291 -0.1073 .  
Dim.1 0.5281 -0.1091 0.1527 0.9784  
Dim.2 0.0500 -0.0417 0.0424 1.1099  
Dim.3 0.2810 -0.0860 0.2229 1.0786  
Dim.4 -0.0580 0.0158 -0.0548 1.1551  
 eCDF Mean eCDF Max Std. Pair Dist.  
distance 0.1110 0.1919 0.3927  
age 0.0194 0.0782 1.0788  
height\_cm 0.0094 0.0402 1.1193  
overall 0.0409 0.1365 0.8735  
potential 0.0204 0.0673 1.0233  
international\_reputation 0.0132 0.0450 0.3492  
weak\_foot 0.0141 0.0448 0.9334  
league1 0.0673 0.0673 0.5799  
european\_club 0.0520 0.0520 0.9008  
Dim.1 0.0597 0.1013 0.9215  
Dim.2 0.0181 0.0369 1.1171  
Dim.3 0.0656 0.1065 0.9729  
Dim.4 0.0201 0.0376 1.0601  
  
Percent Balance Improvement:  
 Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max  
distance 38.7 -1417.5 36.0 22.6  
age 47.7 35.5 46.9 39.5  
height\_cm 26.6 -1.9 26.0 15.0  
overall 31.8 -38.9 29.2 23.7  
potential -1.0 22.2 1.4 -5.3  
international\_reputation 12.5 16.6 12.5 14.6  
weak\_foot 26.0 1.3 22.4 18.0  
league1 54.7 . 54.7 54.7  
european\_club 47.7 . 47.7 47.7  
Dim.1 35.4 74.1 31.9 27.6  
Dim.2 -0.6 16.7 10.8 13.7  
Dim.3 31.5 -133.0 30.7 26.2  
Dim.4 28.7 27.1 18.9 20.0  
  
Sample Sizes:  
 Control Treated  
All 10127 8592  
Matched 8592 8592  
Unmatched 1535 0  
Discarded 0 0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 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 |  |
| player\_european (mean (SD)) | 0.43 (0.49) | 0.27 (0.44) | <0.001 |  |
| wage\_eur (mean (SD)) | 7270.16 (14833.48) | 11568.70 (24702.35) | <0.001 |  |
| international\_reputation (mean (SD)) | 1.07 (0.30) | 1.13 (0.44) | <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 |  |

Stratified by foreigner  
 0 1   
 n 8585 8585   
 age (mean (SD)) 25.20 (4.68) 25.74 (4.45)   
 height\_cm (mean (SD)) 181.06 (6.72) 181.53 (6.91)   
 overall (mean (SD)) 65.26 (6.42) 67.08 (7.15)   
 potential (mean (SD)) 70.70 (5.93) 71.52 (6.44)   
 weight\_kg (mean (SD)) 74.79 (6.92) 75.58 (7.14)   
 player\_european (mean (SD)) 0.43 (0.49) 0.27 (0.44)   
 international\_reputation (mean (SD)) 1.07 (0.30) 1.13 (0.44)   
 weak\_foot (mean (SD)) 2.92 (0.64) 2.98 (0.69)   
 league1 (mean (SD)) 0.78 (0.41) 0.85 (0.36)   
 european\_club (mean (SD)) 0.43 (0.49) 0.38 (0.48)   
 wage\_eur (mean (SD)) 7274.92 (14838.56) 11521.99 (24644.96)  
 skill\_moves (mean (SD)) 2.33 (0.73) 2.46 (0.79)   
 Dim.1 (mean (SD)) -0.11 (4.22) 0.52 (4.17)   
 Dim.2 (mean (SD)) -0.04 (2.05) 0.05 (2.16)   
 Dim.3 (mean (SD)) -0.09 (1.59) 0.28 (1.65)   
 Dim.4 (mean (SD)) 0.02 (1.25) -0.06 (1.35)   
 Stratified by foreigner  
 p test  
 n   
 age (mean (SD)) <0.001   
 height\_cm (mean (SD)) <0.001   
 overall (mean (SD)) <0.001   
 potential (mean (SD)) <0.001   
 weight\_kg (mean (SD)) <0.001   
 player\_european (mean (SD)) <0.001   
 international\_reputation (mean (SD)) <0.001   
 weak\_foot (mean (SD)) <0.001   
 league1 (mean (SD)) <0.001   
 european\_club (mean (SD)) <0.001   
 wage\_eur (mean (SD)) <0.001   
 skill\_moves (mean (SD)) <0.001   
 Dim.1 (mean (SD)) <0.001   
 Dim.2 (mean (SD)) 0.004   
 Dim.3 (mean (SD)) <0.001   
 Dim.4 (mean (SD)) <0.001

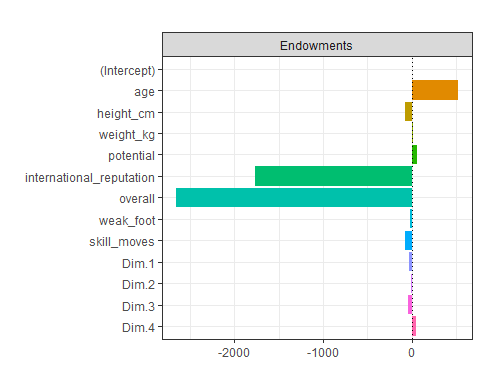
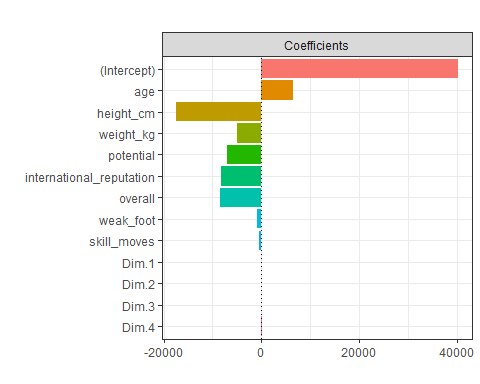
## 7.2 Visualisation et tendance

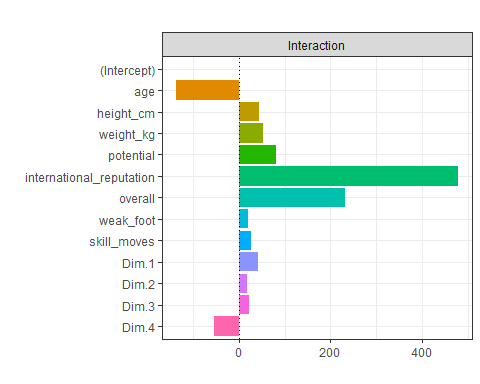
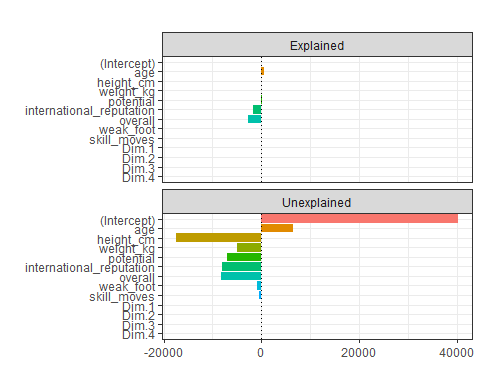
## 7.3 Modélisation économétrique de la discrimination

### 7.3.1 La décomposition de Blinder-Oaxaca

coef(endowments) se(endowments) coef(coefficients) se(coefficients)   
 -4044.2596 NA -1029.8125 NA   
 coef(interaction) se(interaction)   
 827.0016 NA

Les résultats de la décomposition en trois étapes indiquent que 13597 euros de différence entre les salaires des joueurs étrangers et des joueurs nationaux est due aux dotations propres aux joueurs tels que l’âge, le poids, le potentiel ou la réputation internationale. 3655 euros de différence en moyenne est due aux écarts dans les coefficients.

Nous pouvons représenter ces écarts sur  

  # Annexes

## 7.4 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 no 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. Par exemple pour le club Barcelone

|  |  |  |  |
| --- | --- | --- | --- |
| 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 | Atltico Madrid | Spain Primera Division |
| R. Lewandowski | Poland | FC Bayern Mnchen | German 1. Bundesliga |
| Neymar Jr | Brazil | Paris Saint-Germain | French Ligue 1 |
| K. De Bruyne | Belgium | Manchester City | English Premier League |
| K. Mbapp | 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 auquels 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 |
| Atltico Madrid | Spain Primera Division | Spain |
| FC Bayern Mnchen | 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 |

## 7.5 Analyse en composantes principales

### 7.5.1 1. Observation d’individus extrêmes

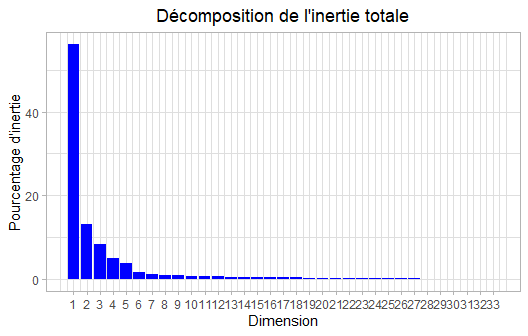
L’analyse des graphes ne révèle aucun individu singulier.

### 7.5.2 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 intertie 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).

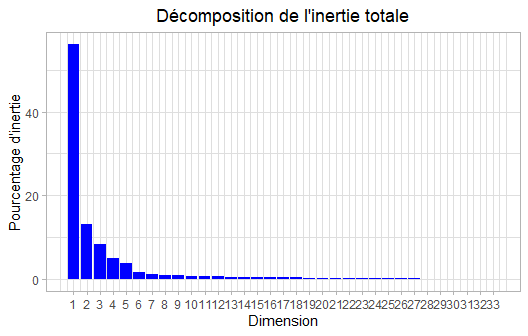
Du fait de ces observations, il serait tout de même probablement préférable de considérer également dans l’analyse les dimensions supérieures ou égales à la troisième.

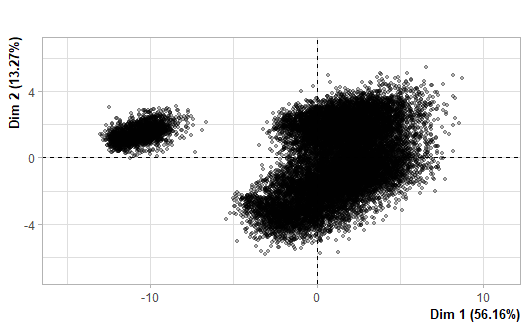


**Figure 2 - Décomposition de l’inertie totale** *Le premier facteur est prépondérant : il explique a lui seul 56.16% de la variabilité totale des données.* *Il convient de noter que dans un tel cas, la variabilité liée aux autres composantes peut être dénuée de sens, en dépit d’un pourcentage élevé.*

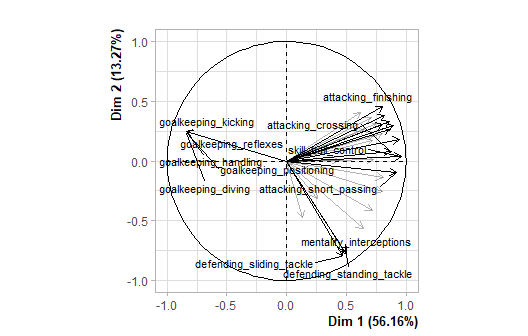
Une estimation du nombre pertinent d’axes à interpréter suggère de restreindre l’analyse à la description des 5 premiers axes. Ces composantes révèlent un taux d’inertie supérieur à celle du quantile 0.95-quantile de distributions aléatoires (86.52% contre 16.17%). Cette observation suggère que seuls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

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





**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).

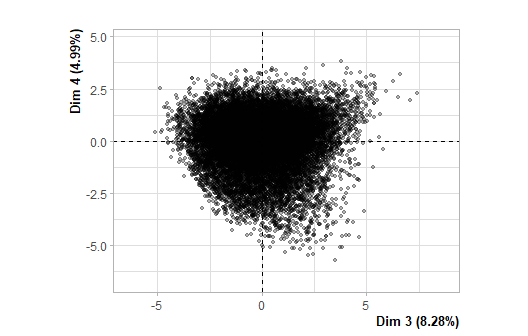
Le groupe 2 (caractérisés par une coordonnée positive sur l’axe) partage :

* 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).

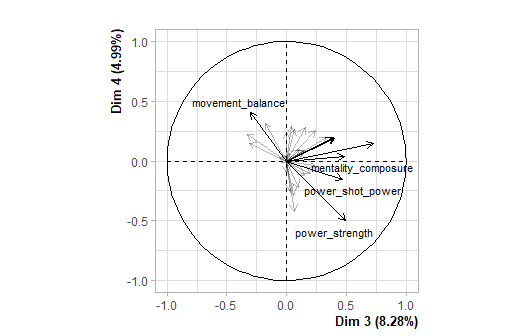
Le groupe 3 (caractérisés par une coordonnées négative sur l’axe) partage :

* 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).

### 7.5.4 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).

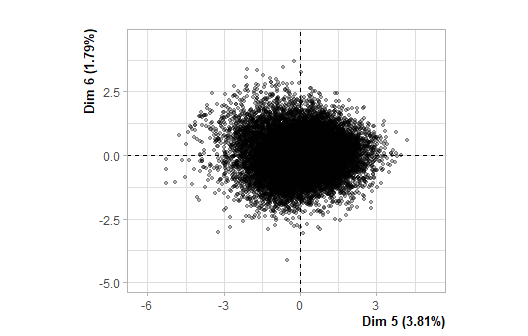
Le groupe 2 (caractérisés par une coordonnée positive sur l’axe) partage :

* 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).

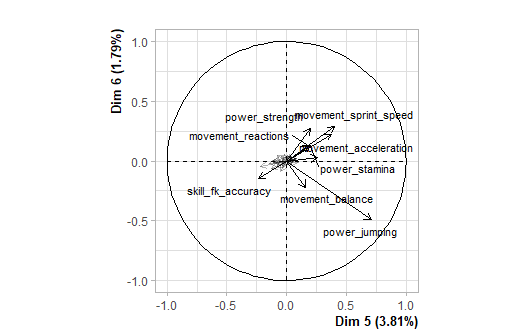
Le groupe 3 (caractérisés par une coordonnées négative sur l’axe) partage :

* 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).

### 7.5.5 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 coordonnées 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 coordonnées 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).

# 8 Bibliographie

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

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

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.

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>