

25/04/2023

Rapport

Économétrie et ONRES



BÉGO Liam & LEHMANN Matthieu
LICENCE 3 MIASHS

SOMMAIRE

INTRODUCTION 3

1. Données et modification 4

1.1. Description du jeu de données 4

Présentation générale du jeu de données 4

Nettoyage du dataset 4

Modification du dataset 5

2. Statistiques descriptives. 6

2.1 Variables d'intérêts (Salaire des joueurs) 6

2.2 Informations générales sur les joueurs 8

2.2.1 L'âge des joueurs 8

2.2.2 Latéralité 10

2.2.3 Le championnat des joueurs 10

2.2.4 Le poste des joueurs 12

2.2.5 Nationalité des joueurs 13

2.3 Performances et compétences des joueurs 14

2.3.1 Note général et potentiel 14

2.3.2 Score de dribbles 15

3. Estimation du Salaire 15

3.2 Joueur de référence 16

3.3 Équation à estimer 16

3.4 Démarche empirique 17

a. Tests de significativité des coefficients 17

b. Test de Chow 17

3.5 Présentation et analyse des résultats d'estimation 17

a. Variables exclues du modèle final 18

b. Performance du modèle 18

c. Analyse des coefficients estimés 18

d. Test de Chow sur la latéralité des joueurs 19

INTRODUCTION

1. Données et modification

1.1. Description du jeu de données

Présentation générale du jeu de données

Le jeu de données utilisé pour cette analyse provient du célèbre jeu vidéo de football, FIFA 22, développé par EA Sports. Ce jeu est réputé pour sa simulation réaliste des performances des joueurs de football professionnels et est mis à jour annuellement avec les dernières informations sur les joueurs, les équipes et les compétitions. Les données sont extraites de la base de données du jeu, qui contient des informations détaillées sur des milliers de joueurs évoluant dans des clubs et des équipes nationales à travers le monde.

Le dataset contient des informations sur un grand nombre de joueurs, avec plusieurs variables pour chaque joueur. Parmi les variables disponibles, on trouve notamment le nom, le pays, le club, le poste, l'âge, la note globale, le potentiel, la valeur marchande et le salaire. Ces variables permettent de dresser un portrait précis de chaque joueur et de mieux comprendre les facteurs qui influencent les performances sur le terrain et les rémunérations.

L'objectif de notre analyse est d'étudier les caractéristiques des joueurs et d'expliquer les salaires en fonction de ces variables. Grâce aux statistiques descriptives et aux techniques d'analyse plus avancées, nous pourrions identifier les facteurs qui ont le plus d'impact sur la rémunération des joueurs et éventuellement mettre en évidence des tendances intéressantes ou des anomalies dans les données.

Dans la suite de l'analyse, nous explorerons et nettoierons les données pour nous assurer qu'elles sont prêtes pour l'analyse. Nous examinerons ensuite les statistiques descriptives générales et les distributions des salaires en fonction de différentes catégories, telles que le pays, le club et le poste. Enfin, nous chercherons à établir des relations entre les différentes variables et les salaires des joueurs pour mieux comprendre les facteurs qui influencent la rémunération dans le football professionnel.

Nettoyage du dataset

Lors du nettoyage et de la préparation du dataset, nous avons d'abord travaillé avec un jeu de données comprenant 19 239 lignes et 110 colonnes. Cependant, pour rendre l'analyse plus pertinente et gérable, plusieurs étapes de filtrage et de sélection des variables ont été effectuées.

Tout d'abord, nous avons éliminé un grand nombre de colonnes qui n'étaient pas pertinentes pour notre objectif d'analyse. En supprimant ces colonnes, nous avons pu nous concentrer sur les variables les plus importantes pour expliquer les salaires des joueurs.

Ensuite, nous avons réduit le nombre de joueurs inclus dans notre analyse en ne conservant que les **2 448** joueurs évoluant dans les cinq grands championnats européens. Cette décision a été prise car les données relatives à ces championnats sont généralement plus précises et

représentatives du football professionnel à haut niveau. De plus, les salaires des joueurs évoluant dans ces championnats sont souvent plus élevés et plus comparables entre eux, ce qui facilite l'analyse des facteurs influençant la rémunération.

On a retenue donc le championnat anglais, espagnol, italien, allemand et français.

Enfin, nous avons exclu tous les gardiens de but du dataset, car leurs statistiques sont généralement différentes de celles des joueurs de champ. Les gardiens de but ont des attributs de jeu spécifiques qui ne sont pas directement comparables à ceux des autres joueurs. En ne conservant que les joueurs de champ, nous avons pu nous concentrer sur un ensemble de variables plus homogène.

Grâce à ces étapes de nettoyage et de préparation des données, notre dataset final est composé de **2 448** joueurs et de **26** variables pour les décrire.

Modification du dataset

Ensuite, nous avons apporté des modifications à certaines variables pour les rendre plus exploitables. Par exemple, la date d'arrivée au club, qui était initialement sous forme de date, a été transformée en nombre de mois que le joueur est présent dans le club. De plus, nous avons converti plusieurs variables en facteurs pour faciliter leur utilisation dans notre analyse. Par exemple, le taux de défense, qui était à l'origine en trois niveaux ("Low", "Medium", "High"), a été transformé en valeurs numériques (0, 1, 2) pour simplifier les comparaisons et les analyses ultérieures.

Tableau 1 : Dictionnaire des données

NOM	Description
Nom	Nom du joueur
Poste	Position sur le terrain
Note	Niveau général du joueur
Potentiel	Niveau hypothétique que le joueur pourrait atteindre
Valeur	Valeur du contrat (€)
Salaire	Salaire hebdomadaire (€/semaine)
Age	Age
Taille	Taille en centimètre
Poids	Poids en killogramme
Club	Club actuel
Ligue	Championnat du club
Date_arrivee	Durée que le joueur est dans son club actuel (Mois)
Nationalite	Nationalité du joueur
Pied_prefer	Latéralité du joueur
Pied_faible	Capacité a tiré de son mauvais pied
Mouvements_speciaux	Note sur la possibilité de réalisé des gestes techniques compliqués

Clause_lib	Valeurs en € fixé de la clause libératoire (montant que devras payé un club à un autre pour être assurer d'avoir le joueur)
Vitesse	Rapidité du joueur (Accélération, vitesse maximal...)
Tirs	Capacité de tirs (Coup franc, tir de loin...)
Passe	Capacité à la passe (Passe courte, centre...)
Dribbles	Capacité de dribble (Vitesse, technique...)
Defense	Capacité à défendre (Positionnement,tacle...)
Physique	Note physique du joueur (Endurance, appuies...)
Annee_contrat_restante	Nombre d'années restantes sur son contrat le liant à un club

2. Statistiques descriptives.

L'analyse descriptive est une première étape importante dans la compréhension des données. Elle permet d'obtenir une vue d'ensemble des caractéristiques des données avant de passer à une analyse plus détaillée. Dans le cadre de cette analyse descriptive des données du jeu de FIFA 22, nous allons nous intéresser à trois grands domaines : le salaire des joueurs, les informations générales sur les joueurs, ainsi que leurs performances et compétences.

La variable d'intérêt principale de cette analyse descriptive est le salaire des joueurs, qui représente le montant que les joueurs gagnent chaque semaine pour leur travail. Nous allons explorer la distribution des salaires et identifier les joueurs qui ont les salaires les plus élevés et les plus bas.

En plus des informations sur le salaire, nous allons également examiner les informations générales sur les joueurs, telles que leur âge ou leurs nationalités. Nous allons également étudier les performances et compétences des joueurs, notamment leur note générale, leur potentiel et leur capacité à dribbler.

L'analyse descriptive des données nous permettra de mieux comprendre les caractéristiques des joueurs dans le jeu de FIFA 22, ce qui pourra nous aider à formuler des hypothèses pour des analyses plus avancées et des modèles prédictifs.

2.1 Variables d'intérêts (Salaire des joueurs)

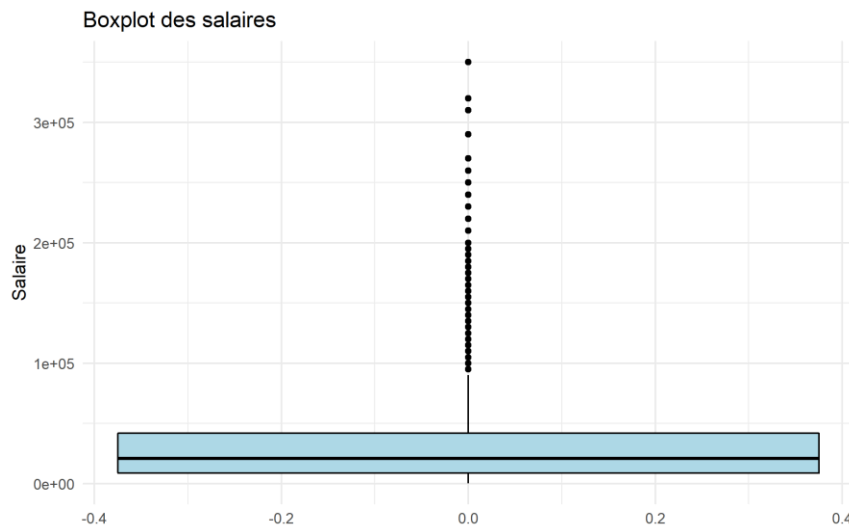
Dans cette section, nous nous concentrons sur l'analyse des statistiques descriptives de notre variable d'intérêt, à savoir le salaire des joueurs. Nous utiliserons divers graphiques pour visualiser et comprendre la distribution des salaires dans notre échantillon de joueurs.

Tableau 2 : Statistique descriptives du salaire

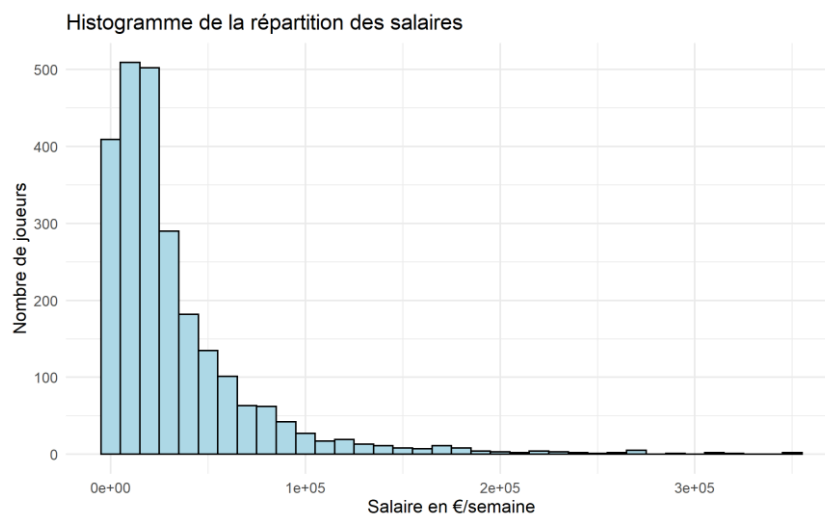
Min	Q1	Median	Moyenne	Q3	Max
500	9000	21000	33731	42000	350000

Nous voyons dans un premier temps que les salaires dans le monde du football sont très dispersés et concentré vers 20 000 € par semaine, même si ce salaire semble élevé une élite est beaucoup plus haute que cela, de nombreuses inégalités subsiste même dans le monde professionnel.

Graphique 1 : Boxplot des salaires des joueurs



Graphique 2 : Histogramme



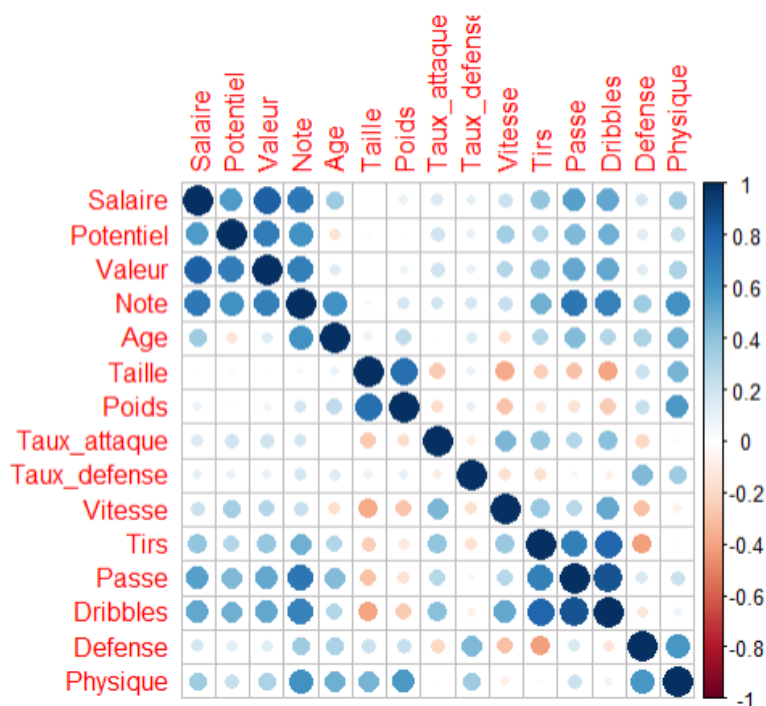
Le boxplot et l'histogramme permet de mieux se rendre compte de la concentration des salaires aux niveau de la médiane et de l'étendu.

2.2 Informations générales sur les joueurs

Nous allons examiner et analyser les données relatives aux caractéristiques de base des joueurs de football présents dans notre jeu de données. Les variables d'intérêt incluent l'âge des joueurs, leur taille, leur nationalité. Ces éléments constituent des informations essentielles pour mieux comprendre la composition et la diversité des joueurs dans les cinq principaux championnats européens. L'analyse de ces données nous permettra de dégager des tendances générales et d'identifier des caractéristiques spécifiques aux différentes populations de joueurs.

Dans un premier temps, on va s'intéresser à la matrice des corrélations de nos variables quantitatives afin d'avoir une idée globale de nos données.

Graphique 3 : Matrice de corrélation des variables quantitatives

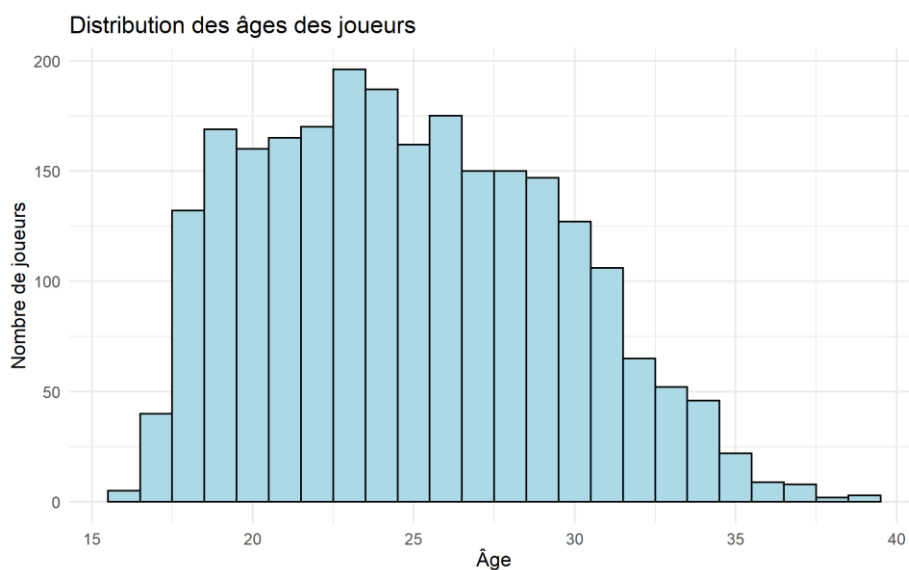


La matrice de corrélation, nous permet de voir une corrélation positive entre a plupart des variables de notre jeux de données, ce qui peut sembler logique pour la plupart étant données la nature des variables qui sont plutôt des notations.

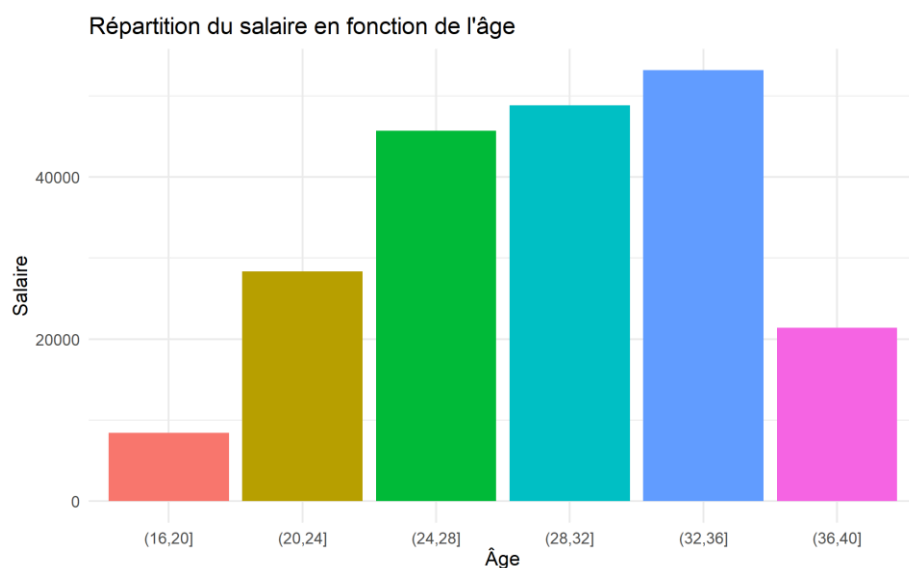
2.2.1 L'âge des joueurs

Dans un premier temps, nous allons explorer la répartition des joueurs en fonction de leur âge afin de comprendre la structure de la population des joueurs. Ensuite, nous pourrons examiner l'impact de l'âge sur le salaire pour mieux comprendre les relations entre ces variables.

Graphique 4 : Histogramme



Graphique 5 : Histogramme



Dans la première partie de l'analyse, nous observons la distribution des âges des joueurs, qui présente une allure plutôt normale. La médiane se situe à 24,5 ans, ce qui semble cohérent pour une carrière professionnelle de footballeur. Lorsqu'on examine la relation entre l'âge et le salaire, on constate une augmentation générale du salaire tout au long de la carrière des joueurs. Les joueurs âgés de 16 à 20 ans gagnent en moyenne 9 000 € par semaine, tandis que ceux entre 32 et 36 ans atteignent un salaire moyen supérieur à 50 000 € par semaine. Cependant, après l'âge de 36 ans, les salaires diminuent drastiquement.

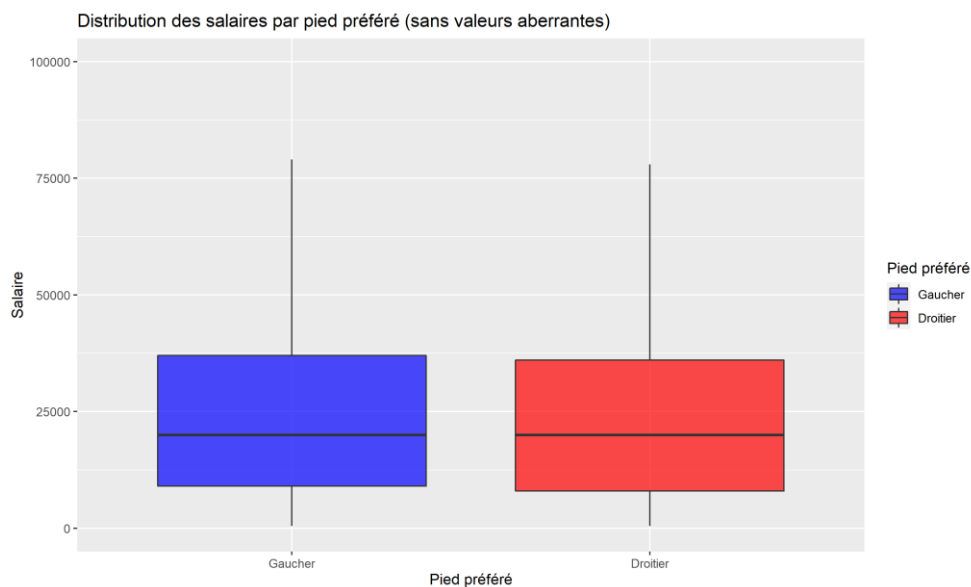
2.2.2 Latéralité

Dans cette section, nous allons analyser l'impact de la préférence de pied des joueurs de football sur leurs salaires. Il est courant de considérer que les joueurs gauchers sont plus rares et peuvent posséder des compétences ou des qualités uniques sur le terrain. Nous explorerons si cette rareté se traduit également par des différences de rémunération entre les joueurs gauchers et droitiers. En examinant les salaires moyens en fonction de la préférence de pied, nous pourrions déterminer si les joueurs gauchers sont effectivement mieux rémunérés que leurs homologues droitiers et mieux comprendre les facteurs qui influencent la rémunération des joueurs de football professionnel.

Proportion de droitier dans l'échantillon : **73%**

Proportion de gaucher dans l'échantillon : **27%**

Graphique 6 : Boxplot



Nous constatons que les valeurs sont quasiment identiques pour les joueurs gauchers et droitiers. Cela suggère que la préférence de pied n'a pas d'impact significatif sur la rémunération des joueurs de football professionnel. Bien que les joueurs gauchers puissent être considérés comme plus rares, cette rareté ne semble pas se traduire par des salaires plus élevés.

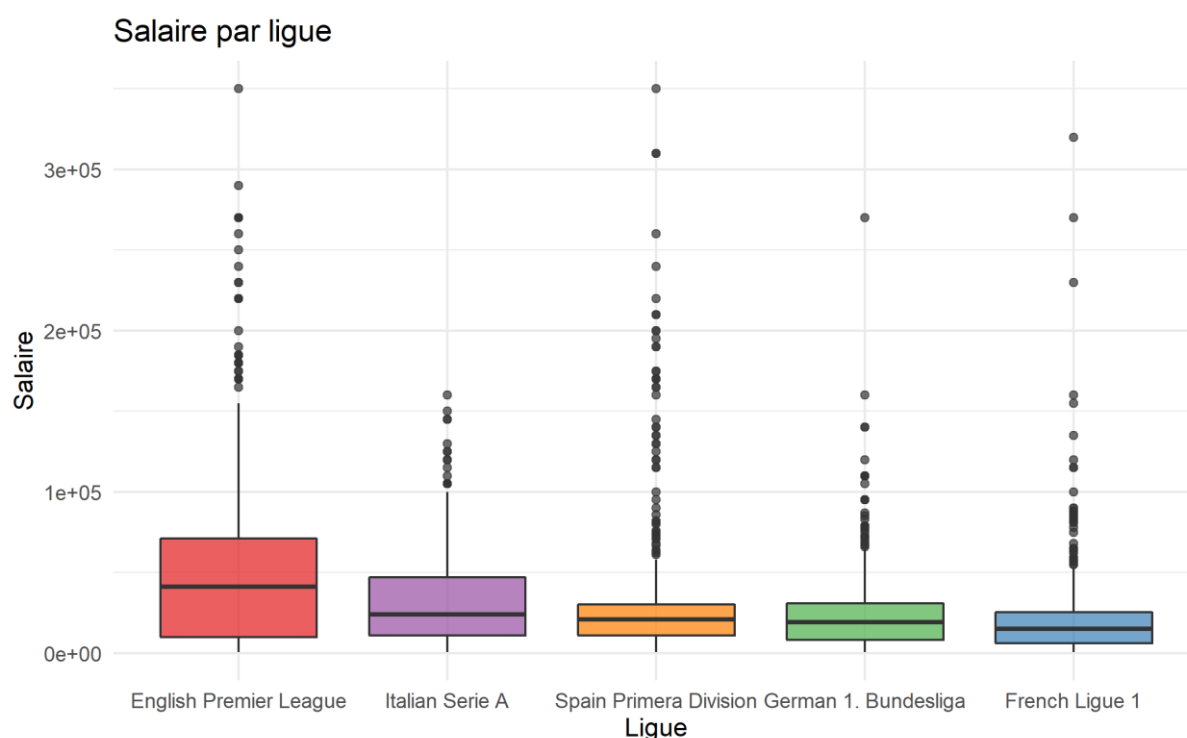
2.2.3 Le championnat des joueurs

Dans cette section, nous examinerons comment les salaires des joueurs varient en fonction des championnats auxquels ils appartiennent. Notre objectif est d'identifier les tendances et les différences salariales entre les cinq grands championnats européens. En analysant ces données, nous pourrions évaluer l'attractivité financière de chaque ligue et mieux comprendre les facteurs qui influencent les salaires des joueurs de football professionnel dans différents pays et compétitions.

Tableau 2 : Nombre de joueurs par championnat

Ligue anglaise	Ligue espagnol	Ligue française	Ligue italienne	Ligue allemande
461	567	524	471	461

Graphique 7 : Boxplot



D'après le graphique des salaires par ligue, nous pouvons constater que la ligue anglaise (Premier League) offre en général les salaires les plus élevés pour les joueurs de football. Cela pourrait s'expliquer par la popularité et les revenus importants générés par cette ligue, ainsi que par le niveau de compétition élevé qui attire des joueurs de talent.

En deuxième position, nous avons la ligue italienne (Serie A), qui présente également des salaires relativement élevés. Cependant, il est intéressant de noter qu'elle possède moins de valeurs très élevées en termes de salaire, comparativement aux quatre autres ligues principales. Cela pourrait signifier que, bien que la Serie A offre des salaires compétitifs, elle ne parvient pas à attirer autant de joueurs superstar que certaines des autres ligues.

Ensuite, nous avons la ligue espagnole (La Liga), suivie de près par la ligue allemande (Bundesliga). Ces deux ligues présentent des salaires moyens légèrement inférieurs à ceux de la Premier League et de la Serie A, mais elles ont néanmoins des salaires élevés pour certaines de leurs stars les plus en vue.

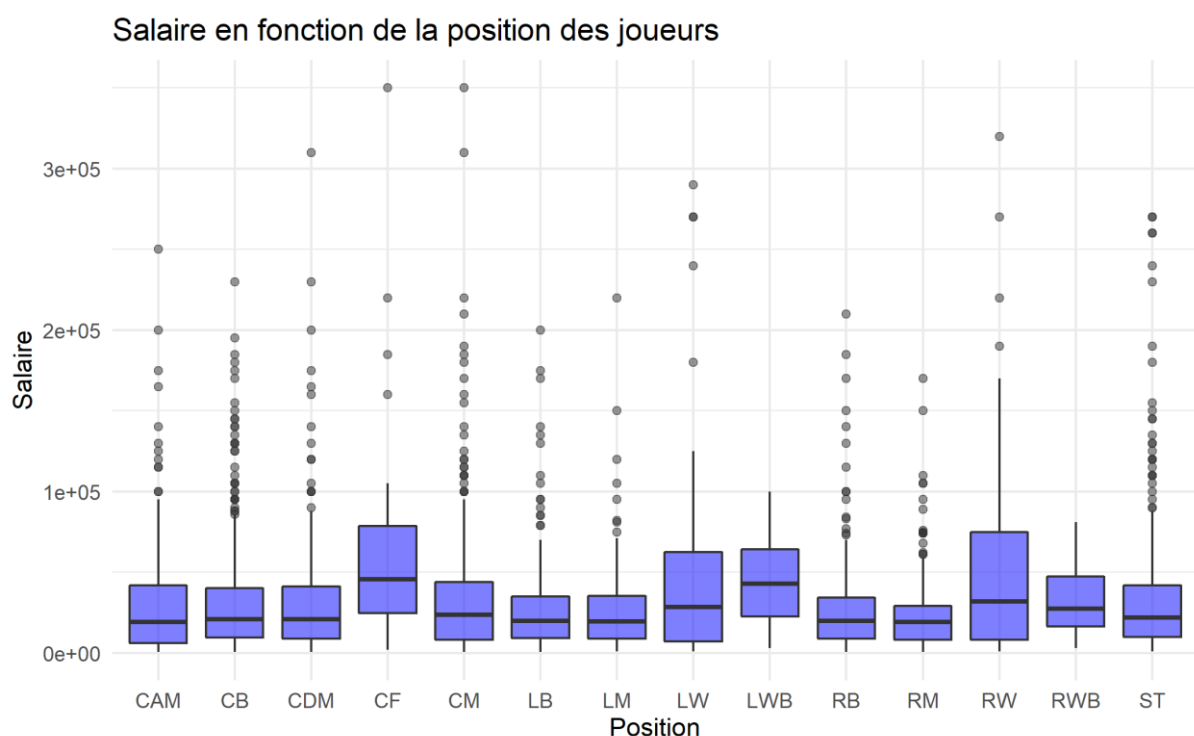
Enfin, la Ligue 1 française se trouve en dernière position parmi les cinq principales ligues européennes en termes de salaire. Bien qu'elle offre des salaires compétitifs pour de nombreux joueurs, elle ne parvient pas à rivaliser avec les autres ligues en termes de rémunération globale. Cela pourrait être dû à divers facteurs, tels que des revenus moins

importants, une compétition moins intense ou une moindre attractivité pour les joueurs internationaux de haut niveau.

2.2.4 Le poste des joueurs

Dans cette partie, nous allons explorer les salaires des joueurs en fonction de leur position sur le terrain. L'objectif est d'analyser les variations de rémunération entre les différents postes et de déterminer si certains rôles, tels que les postes offensifs, sont mieux rémunérés que d'autres.

Graphique 8 : Boxplot



Le graphique des salaires par poste révèle effectivement que les postes offensifs tendent à être mieux rémunérés que les autres postes. Ceci peut s'expliquer par plusieurs raisons :

1. Les attaquants et les milieux offensifs ont souvent un impact direct sur le résultat des matchs en marquant des buts ou en réalisant des passes décisives, ce qui peut leur valoir une plus grande reconnaissance et, par conséquent, des salaires plus élevés.
2. Les joueurs offensifs sont généralement plus médiatisés, en raison de leur rôle clé dans la création et la concrétisation d'occasions de but. Cette visibilité accrue peut se traduire par des contrats de sponsoring plus lucratifs et des salaires plus élevés.
3. Les clubs sont souvent prêts à investir davantage dans des joueurs offensifs de qualité, car ils peuvent faire la différence dans les matchs serrés et contribuer à gagner des titres. Par conséquent, les clubs peuvent offrir des salaires plus élevés pour attirer et retenir les meilleurs talents offensifs.

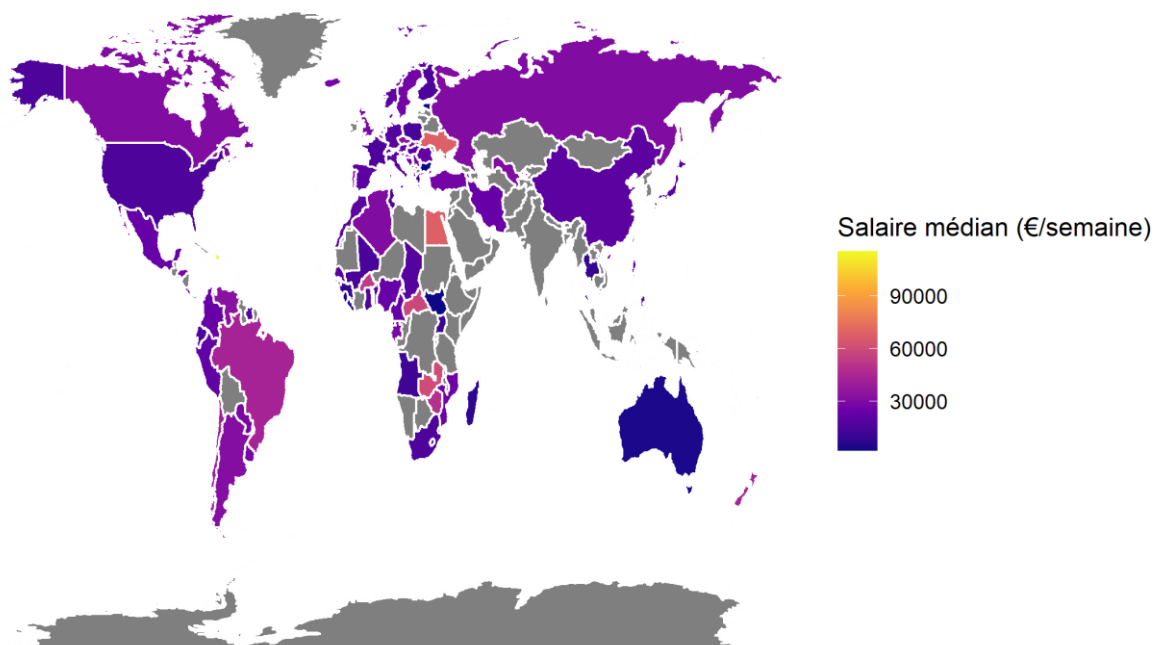
Cependant, il est important de noter que, bien que les postes offensifs soient généralement mieux rémunérés, il existe des exceptions notables, avec certains défenseurs, milieux défensifs et gardiens de but qui peuvent également recevoir des salaires élevés en raison de leur importance et de leur qualité exceptionnelle.

2.2.5 Nationalité des joueurs

Dans cette partie, nous allons étudier la répartition des salaires médians des joueurs de football en fonction de leur nationalité. En utilisant une carte choroplèthe, nous pourrions visualiser les disparités de rémunération entre les différentes nationalités et identifier les pays où les joueurs ont tendance à recevoir des salaires plus élevés ou plus faibles.

Graphique 9 : Choroplète

Salaire médian par nationalité



Le graphique montre effectivement que certains pays présentent un salaire médian élevé, ce qui peut être attribué au faible nombre de joueurs de ces nationalités présents dans notre échantillon. Dans ces cas, quelques joueurs avec des salaires élevés peuvent influencer considérablement la médiane et ne reflètent pas nécessairement une tendance générale pour ces pays.

En outre, il est intéressant de constater que les pays d'Amérique du Sud semblent afficher des salaires médians plutôt élevés pour leurs joueurs. Cela peut s'expliquer par plusieurs facteurs, tels que le niveau élevé de talent et la forte tradition footballistique dans cette région, qui favorisent l'émergence de joueurs de haut niveau. Ces joueurs peuvent attirer des contrats lucratifs et contribuer à l'augmentation du salaire médian pour les joueurs d'Amérique du Sud.

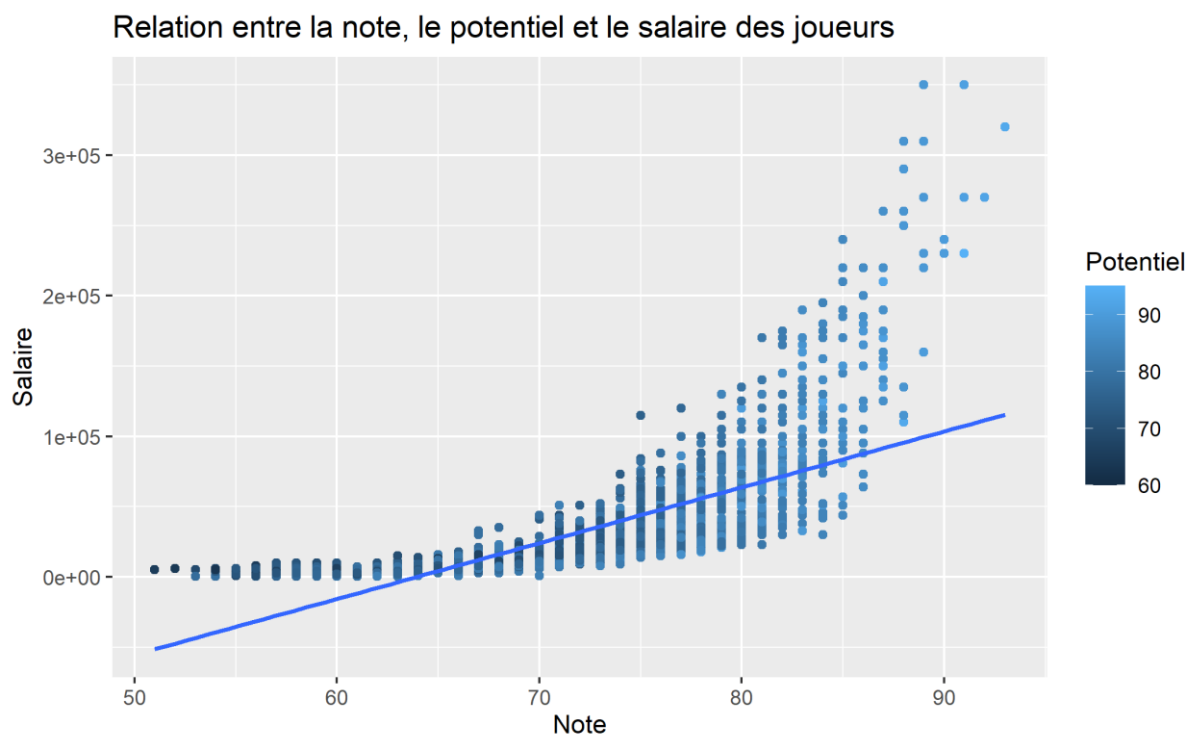
2.3 Performances et compétences des joueurs

Dans cette deuxième partie, intitulée "Performances et compétences des joueurs", nous nous concentrerons sur l'analyse des données relatives aux compétences et aux performances des joueurs de football présents dans notre jeu de données. Les variables d'intérêt incluent la note globale des joueurs, leur potentiel, ainsi que des compétences spécifiques telles que le dribble, la défense, les passes, les tirs et d'autres statistiques de jeu. En examinant ces variables, nous chercherons à comprendre comment ces compétences et performances individuelles varient entre les joueurs et comment elles peuvent être liées à d'autres facteurs tels que l'âge, la position sur le terrain, la nationalité et le club. Cette analyse approfondie des compétences et des performances des joueurs nous permettra également d'explorer les relations potentielles entre ces variables et le salaire des joueurs, offrant ainsi un aperçu des facteurs qui peuvent influencer la rémunération dans le football professionnel.

2.3.1 Note général et potentiel

Dans la section suivante de notre analyse, nous allons étudier la relation entre la note générale, le potentiel des joueurs et leur salaire, afin de mieux comprendre l'importance de ces facteurs. La note générale étant globalement le niveau du joueurs l'impact sur le salaire semble logique

Graphique 10 : Nuage de points



L'analyse graphique révèle que le salaire des joueurs augmente à mesure que leur note générale s'améliore. Il est intéressant de constater qu'il existe un petit groupe de joueurs

d'élite ayant des notes particulièrement élevées, qui se distinguent nettement par des salaires beaucoup plus importants que les autres joueurs.

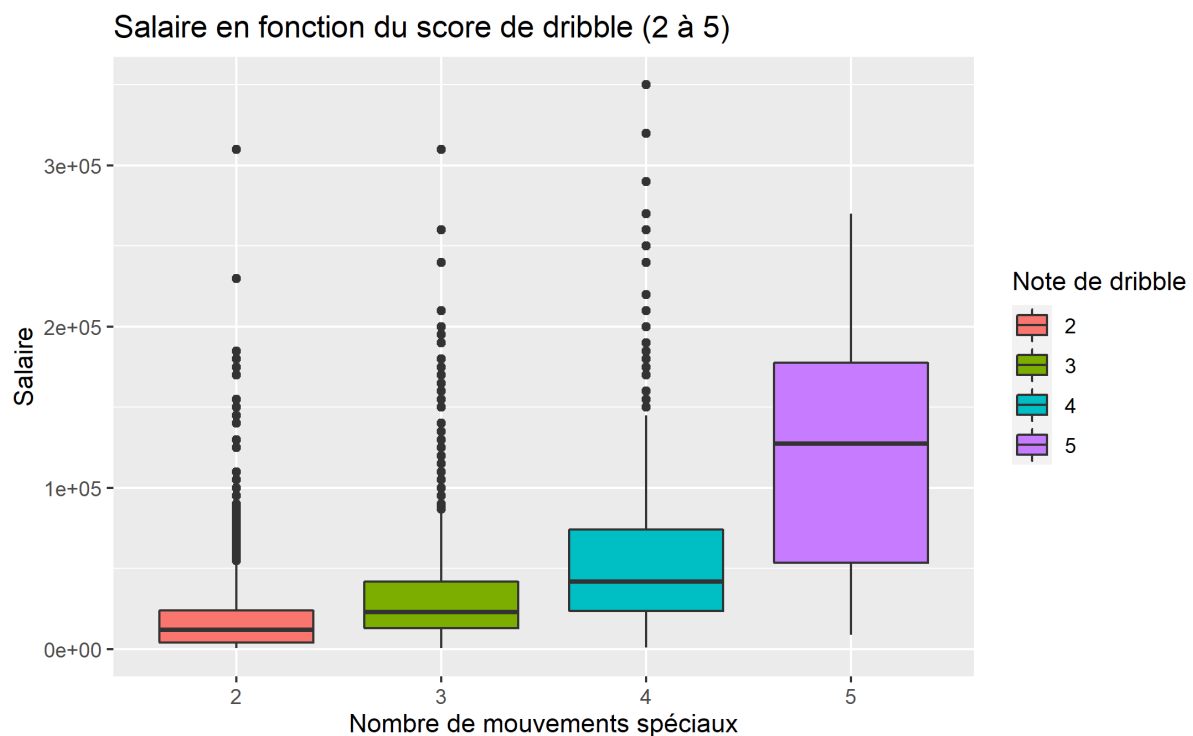
2.3.2 Score de dribbles

Dans cette section de l'analyse, nous nous penchons sur l'impact du score de dribbles sur les salaires des joueurs

Tous joueurs	Défenseurs	Attaquants
2.817	2.349	3.105

On a dans un premier temps observé la moyenne en fonction du poste car généralement les défenseurs étant de mauvais dribbleurs on observe bien une différence entre les moyennes des notes de presque 1 point.

Graphique 11 : Boxplot



Le score de dribbles semble être un facteur significatif dans la détermination du salaire des joueurs de football. Les attaquants, ayant généralement des scores de dribbles plus élevés que les défenseurs, tendent également à avoir des salaires plus élevés. Les joueurs ayant de meilleures compétences en dribble sont donc plus susceptibles de gagner un salaire plus important, indépendamment de leur position sur le terrain.

3. Estimation du Salaire

La variable à expliquer est le salaire des joueurs de football. Les variables explicatives incluent toutes les variables du dataset. Quelques modifications y ont été faites comme le groupement des nationalités en continent. Le club n'est pas inclus dans les variables explicatives car le club implique la ligue ce n'est donc pas une information pertinente.

3.1 Joueur de référence

Lors de l'analyse des facteurs influençant les salaires des joueurs de football, il est essentiel de disposer d'un joueur de référence pour effectuer des comparaisons pertinentes. Un joueur de référence est un joueur dont les caractéristiques servent de point de comparaison pour évaluer l'effet des variables explicatives sur le salaire des autres joueurs. En général, un joueur de référence est choisi en fonction de critères tels que sa position, son niveau de compétence ou son âge.

L'utilisation d'un joueur de référence facilite l'interprétation des coefficients estimés dans le modèle de régression. En effet, les coefficients représentent alors la différence de salaire entre les joueurs ayant une certaine caractéristique et le joueur de référence. Cela permet de mieux comprendre l'ampleur et la direction de l'effet de chaque variable explicative sur le salaire des joueurs.

Enfin, un joueur de référence peut également être utile pour effectuer des tests statistiques, tels que le test de Chow mentionné précédemment. En comparant les coefficients de régression entre les groupes de joueurs (par exemple les droitiers et les gauchers) en utilisant un joueur de référence, on peut détecter s'il existe des différences significatives dans la manière dont les salaires sont influencés par les différentes variables explicatives.

Notre joueur de référence a été choisi pour correspondre à un idéal. C'est-à-dire que dans l'opinion populaire ce sera le mieux payé, il aura les meilleurs statistiques dans les variables qualitatives tel que le taux d'attaque. Il sera également un joueur européen évoluant dans le championnat anglais au poste d'attaquant. Nous avons choisi la latéralité gaucher dû à la rareté.

3.2 Équation à estimer

$$\begin{aligned} \text{Salaire} = & \beta_0 + \beta_1 * \text{Position} + \beta_2 * \text{Note} + \beta_3 * \text{Potentiel} + \beta_4 * \text{Valeur} + \beta_5 * \text{Age} \\ & + \beta_6 * \text{Age}^2 + \beta_7 * \text{Taille} + \beta_8 * \text{Poids} + \beta_9 * \text{Ligue} + \beta_{10} \\ & * \text{Date fin contrat} + \beta_{11} * \text{Latéralité} + \beta_{12} * \text{Pied faible} + \beta_{13} \\ & * \text{Mouvements spéciaux} + \beta_{14} * \text{Taux Attaque} + \beta_{15} * \text{Taux défense} \\ & + \beta_{16} * \text{Clause libératoire} + \beta_{17} * \text{Vitesse} + \beta_{18} * \text{Tir} + \beta_{19} * \text{Passe} \\ & + \beta_{20} * \text{Dribble} + \beta_{21} * \text{Défense} + \beta_{22} * \text{Physique} + \beta_{23} * \text{Continent} \end{aligned}$$

Dans la suite de cette analyse, nous allons estimer les coefficients β_n de l'équation de régression linéaire ci-dessus afin de déterminer les facteurs qui influencent le salaire des joueurs de football. Cette équation prend en compte toutes les variables ont déjà été décrites.

L'objectif principal de cette analyse est de déterminer l'impact de chaque variable explicative sur le salaire des joueurs et d'identifier les facteurs les plus significatifs. Pour ce faire, nous allons estimer les coefficients β_n à l'aide d'une régression linéaire multiple et effectuer des tests statistiques pour évaluer la significativité de chaque coefficient. En fonction des résultats, nous pourrions déterminer les facteurs clés qui influencent les salaires des joueurs de football et fournir des recommandations aux clubs et aux joueurs en matière de négociation salariale.

3.3 Démarche empirique

Dans cette section, nous allons décrire les tests statistiques qui seront réalisés pour évaluer la qualité et la robustesse de notre modèle de régression linéaire. Ces tests comprennent les tests classiques de significativité des coefficients, ainsi que le test de Chow pour vérifier la stabilité des paramètres du modèle entre les différents groupes.

3.3.1 Tests de significativité des coefficients

Ces tests permettent de déterminer si les coefficients estimés des variables explicatives sont significativement différents de zéro. En d'autres termes, ils permettent d'évaluer si les variables explicatives ont un effet significatif sur le salaire des joueurs. Pour chaque coefficient, on effectuera un test t de Student. Si la valeur p associée au test est inférieure à un seuil prédéfini (généralement 0,05), alors le coefficient sera considéré comme significatif.

3.3.2 Test de Chow

Le test de Chow est un test d'hypothèse qui compare les coefficients de régression de deux groupes de données pour déterminer s'ils sont significativement différents. Dans notre cas, nous voulons vérifier si les coefficients de régression pour 2 groupes de joueurs sont différents, ce qui pourrait indiquer que les salaires de ces deux groupes sont influencés par des facteurs distincts.

Ces groupes sont les droitiers et gaucher, puis les joueurs mesurant plus ou moins de 1m82. Cette dernière valeur étant la médiane de l'échantillon.

3.4 Présentation et analyse des résultats d'estimation

Après avoir estimé le modèle de régression linéaire multiple, nous avons obtenu des résultats intéressants que nous allons présenter et analyser.

3.4.1 Variables exclues du modèle final

Le modèle final n'inclut pas les variables suivantes : la note, la taille, le poids, la date de fin de contrat, la latéralité, le pied faible, la vitesse, le tir, le dribble et le taux de défense. Ces variables ont été exclues du modèle en raison de la valeur élevée de leurs p-values (supérieure à 5%). Cela signifie que ces variables n'ont pas d'effet significatif sur le salaire des joueurs de football, au moins dans notre modèle.

3.4.2 Performance du modèle

Le modèle final présente une p-value générale proche de 0, ce qui indique que le modèle est statistiquement significatif et que les variables explicatives incluses dans le modèle ont un effet significatif sur le salaire des joueurs de football. De plus, le R^2 du modèle est de 0.7852, ce qui signifie que 78,52% de la variation du salaire des joueurs peut être expliquée par les variables incluses dans le modèle. Cela montre que notre modèle a une bonne capacité à expliquer les différences de salaire entre les joueurs.

3.4.3 Analyse des coefficients estimés

Les coefficients estimés pour les variables restantes du modèle final (Position, Potentiel, Valeur, Age, Age², Ligue, Mouvements spéciaux, Taux d'attaque, Clause libératoire, Défense, Physique et Continent) montrent l'impact de chaque variable sur le salaire des joueurs de football. Il est important de noter que les coefficients doivent être interprétés en tenant compte de l'échelle des variables et des unités de mesure. Par exemple, nous obtenons un coefficient positif pour l'âge, mais négatif pour l'âge² ce qui montre bien que plus l'âge grandit plus l'âge a un impact négatif sur le salaire. On pourra donc calculer un âge « seuil » auquel les joueurs atteignent leur salaire maximal. De plus, On obtient des valeurs de coefficients négatif pour les positions des joueurs. Or, notre joueur de référence est un attaquant. Donc on montre bien ici que le poste de défenseur ou de milieux de terrain a un impact négatif sur le salaire d'un joueur.

Notre joueur de référence a été choisi car dans l'opinion populaire ce sera le joueur le mieux payé. Nos résultats coïncident avec cette hypothèse. En effet on obtient des coefficients négatifs dans toutes les variables explicatives sous forme de facteur, sauf pour les variables Taux d'attaque et Continent. Ce qui montre un aspect très important du football moderne, un joueur n'a pas besoin d'apporter un travail monstre en attaque pour qu'il soit considéré à sa juste valeur. Pour le continent c'est une surprise, en mettant à part le continent de l'Océanie car trop peu de valeur. Tous les continents ont un impact positif sur le salaire, c'est-à-dire que n'importe quel joueur tant qu'il n'est pas européen aura tendance à être mieux payé qu'un joueur européen.

En conclusion, notre modèle de régression linéaire multiple a permis d'identifier les facteurs clés qui influencent les salaires des joueurs de football. Toutefois, il convient de garder à l'esprit que cette analyse a été réalisée sur un échantillon de joueurs spécifique et dans un contexte particulier, avec des variables qui n'existent pas forcément dans la réalité, de sorte que les résultats pourraient ne pas être généralisables à l'ensemble du monde du football.

3.4.4 Test de Chow sur la latéralité des joueurs

On obtient une p-value inférieure à 5%. Cela indique que l'on rejette l'hypothèse nulle selon laquelle les coefficients de régression sont identiques pour les deux groupes (droitiers et gauchers). En d'autres termes, cela signifie qu'il existe des différences significatives dans la manière dont les variables explicatives influencent le salaire des joueurs droitiers par rapport aux joueurs gauchers.

Cela peut avoir plusieurs implications pour les clubs et les joueurs. Par exemple, il se peut que certaines caractéristiques soient plus valorisées chez les joueurs gauchers que chez les joueurs droitiers, ou vice versa. Les clubs pourraient ainsi adapter leur stratégie de recrutement en fonction de ces différences, tandis que les joueurs pourraient chercher à développer des compétences spécifiques pour augmenter leur valeur sur le marché, en fonction de leur latéralité.

Après une élaboration d'un modèle similaire à l'ensemble des joueurs on obtient quelques particularités propres à ces deux groupes :

Présentation des résultats

	Droitier	Gaucher
Constante	-76 477***	-191 976***
Défenseur	-8 669***	
Milieux	-8 390***	
Potentiel	868***	1 501***
Valeur	0***	0***
Age	2 391***	10 632***
Ligue 1	-18 105***	-16 841***
Bundesliga	-17 800***	-20 525***
Serie A	-17 655***	-20 310***
Primera Division	-17 883***	-18 724***
Mouvements spéciaux 2	-10 343*	-31 149***
Mouvements spéciaux 3	-10 100*	-33 284***
Mouvements spéciaux 4	-6 133	-32 375***
Clause libératoire	-0***	-0***
Défense	188***	
Physique	-231***	-236*
Afrique	2 057	

Amérique	3 290*	
Asie	1 096	
Océanie	-6 945	
Age ²		-151***
Pied faible 2		-13 438*
Pied faible 3		-11 079*
Pied faible 4		-9 594+
<hr/>		
Nombre d'observation	1793	655
R2	0.791	0.762

Analysons d'abord les éventuels changements de signe des coefficients. Il n'en a pas, les variables en commun entre les deux groupes ont des impacts dans le même sens. Ensuite regardons les variables qui différencie ces deux groupes.

- Poste : Le poste n'a de l'importance que pour un droitier. Le salaire des gauchers est donc plus homogène au travers des différents postes.
- L'âge, il n'est pas polynomial pour les droitiers, ils arrivent donc à faire valoir leur latéralité tout au long de leur carrière.
- Le continent d'origine, comme pour le poste, le salaire des gauchers est plus homogène en fonction des origines.
- La variable pied faible est significatif pour les gauchers de façon négative. Ce résultat est surprenant étant donné que l'on aurait tendance à valoriser un joueur qui sait se servir de ses deux pieds.

4. Prédiction du salaire d'un joueur

Nous avons sélectionné l'un des plus gros transferts de l'été 2022. Il s'agit du joueur Brésilien Antony. Il a été transféré de l'Ajax Amsterdam, un club des Pays-Bas vers Manchester United un club d'Angleterre. Le salaire réel du joueur est de 120.000€. Avec nos informations et nos prévisions avec le modèle gaucher nous obtenons : 110.696€. Nous concluons donc que le joueur est survalorisé par rapport à ses caractéristiques.