

## Grau en Estadística

---

**Títol:** Anàlisi dels índexs individuals d'impacte dels jugadors de bàsquet segons la seva posició.

**Autor:** Miquel de Jover Boira

**Director:** Josep Lluís Carrasco Jordan

**Departament:** Departament de Fonaments Clínics

**Convocatòria:** Curs 2022-2023, Q1

:



## RESUM:

Dins del bàsquet, un esport amb gran diversitat de jugadors distribuïts en equip i diferents posicions, els índexs d'avaluació de l'impacte dels jugadors intenten mesurar amb la major precisió possible la incidència dels jugadors. Aquest estudi se centra a analitzar una selecció de 6 índexs per determinar si es veuen esbiaixats per la posició de cada jugador. Per assolir aquest objectiu, s'estima un model per a cada índex estudiat, que inclogui la posició del jugador i altres efectes. Es troben evidències per creure que l'efecte de la posició té incidència en els valors de la majoria dels índexs, premiant més a les posicions interiors. Això fa que aquests índexs no resultin útils per comparar jugadors de diferents posicions.

## PARAULES CLAU:

Bàsquet, Estadística esportiva, Estadístiques avançades de bàsquet, Índex d'impacte de jugadors, Comparació entre posicions de joc, factors d'influència en rendiment esportiu.

# Índex

1. INTRODUCCIÓ.....	3
2. MARC TEÒRIC .....	5
2.1. Bàsquet i NBA .....	5
2.2. Posicions de joc.....	6
2.3. Índex d'avaluació d'impacte de jugador .....	7
3. ESTRUCTURA DE LA BASE DE DADES .....	12
3.1. Obtenció de les dades .....	12
3.2. Tractament i selecció de les variables d'interès .....	13
3.3. Estructura final de la base de dades.....	17
4. METODOLOGIA.....	18
4.1. Software Utilitzat .....	18
4.2. Anàlisi descriptiva i anàlisi de comparació de grups.....	18
4.3. Anàlisi estadística: Model Lineal Mixt .....	19
5. ANÀLISI DESCRIPTIVA I DE COMPARACIÓ DE GRUPS .....	21
5.1. Equip (TEAM) .....	21
5.2. Edat (Age_Factor) .....	22
5.3. Minuts (MIN) .....	24
5.4. Formació a Estats Units (FormacioUSA) .....	26
5.5. Posició (POS).....	28
5.6. Índexs.....	30
6. RESULTATS DE LA MODELITZACIÓ DELS ÍNDEXS.....	34
6.1. PlusMinus .....	34
6.2. PER .....	37
6.4. RPM.....	48
6.5. PIE .....	51
6.6. LEBRON .....	54
7. DISCUSSIÓ.....	58
8. BIBLIOGRAFIA.....	61
ANNEX: Codi D'R.....	63

# 1. INTRODUCCIÓ

El món de l'esport està en constant evolució i això es reflecteix en una major sofisticació i una atenció més meticulosa en cada un dels seus àmbits. Un dels àmbits que avança a passos agegantats és el de l'estadística esportiva. Concretament, l'àmbit de l'estadística esportiva cada cop atreu més l'atenció a professionals i aficionats de l'esport. Cada cop s'aposta per una orientació més analítica del joc per tal de reforçar opinions o prendre decisions. Per gaudir d'aquesta visió, en els darrers anys s'ha generat un creixement en tècnica de recollida de dades massiva per a tota mena d'esport, per tal de captar en variables tot el que passa i poder extreure'n la màxima informació possible per la seva posterior anàlisi. L'estadística esportiva aporta una nova dimensió a l'anàlisi de l'esport, ja que té com a principal funció generar una aproximació més objectiva de la realitat del joc. Aquest fet és molt valuós per ajudar a prendre decisions a directives, per ajudar als departaments esportius i tècnics a preveure lesions o aconseguir èxits esportius.

Més particularment, l'esport del basquetbol, per estructura i normativa del mateix esport, sempre han abundat les estadístiques de partit. Els diferents valors dels punts així com les puntuacions altes amb les quals es decideixen els partits, generen moltes situacions i factors diferents per analitzar. En un inici, les estadístiques utilitzades eren les conegudes com a estadístiques tradicionals o senzilles, que van ser introduïdes al principi de la dècada dels 80 amb la creació de la línia de 3 punts. En aquell moment, les estadístiques senzilles eren recollides mitjançant unes plantilles anomenades "box score". Aquestes plantilles recollien informació sobre els punts, analitzaven els tirs intentats, convertits i el seu respectiu percentatge d'encert, els minuts, rebots, assistències, taps, pèrdues, recuperacions i faltes comeses. Avui en dia, encara que el box score es continua usant, ha quedat una mica curt per a recollir informació i a la Nacional Basketball Association (NBA) s'utilitza el sistema SportVU. Aquesta tecnologia es tracta d'un software específic amb tecnologia de seguiment, que recull dades sobre els jugadors i el joc mitjançant sis càmeres a cada pavelló. Així doncs, amb aquesta evolució constant de les estadístiques, neixen les estadístiques avançades, que fan ús de les estadístiques més senzilles per intentar aprofundir i assolir una perspectiva més completa. La quantitat d'estadístiques avançades és immensa i està en continu creixement, però totes busquen avaluar de manera més precisa un aspecte del joc o l'aportació d'un jugador<sup>4,6</sup>. En aquest mercat amb una gran varietat d'estadístiques avançades, hi ha una classe que és molt perseguida i anhelada per tothom. Estem parlant de les estadístiques de mesura d'impacte d'un jugador. Aquest tipus, comunament anomenades *all-in-one metrics*, intenten buscar la

fórmula perfecta per tal d'agrupar la informació ofensiva i defensiva i poder determinar amb una sola mesura l'impacte general d'un jugador al partit.<sup>7</sup>

Hi ha molts factors que influeixen en aquests índexs d'avaluació d'impacte de jugadors, però l'objectiu de tots ells és intentar que siguin el màxim justos possible. Els índexs pretenent estar poc influïts pel context, ja que per molt que evidentment el joc d'un jugador es veu influït per l'equip on està o els companys que té, ha d'intentar minimitzar aquest soroll. Per altra banda, també pretenen avaluar a tothom per igual, donat que han de poder avaluar les habilitats i aportacions de tots els esportistes per molt diferents que siguin.<sup>8</sup>

Aquest treball es centre en aquest últim punt, ja que té com a objectiu veure si els diferents índexs d'avaluació d'impacte es veuen condicionats per les posicions de joc dels jugadors. Per tal de veure si els índexs estan esbiaixats per posició, es realitzarà l'estudi amb els jugadors de la lliga NBA i amb una selecció de 6 índexs d'avaluació d'impacte diferents. En primer lloc, es seleccionaran totes les estadístiques mitjanes de cada jugador durant la temporada 2021-2022 a la lliga NBA. Al mateix temps, per a cada jugador, s'agafaran els valors mitjans dels 6 índexs estudiats. La base de dades obtinguda per tal de fer l'estudi consta en un inici de 605 observacions, referent a cada jugador que hagi participat en la lliga durant la temporada 2021-2022. A partir d'aquí, s'avaluarà l'efecte que pot tenir la posició de joc dels jugadors en les seves respectives puntuacions als diferents índexs. Per assolir aquest objectiu, s'estima un model per a cada índex estudiat, que inclogui la posició i d'altres efectes.

## 2. MARC TEÒRIC

### 2.1. Bàsquet i NBA

El basquetbol va ser creat per James Naismith l'any 1891 a una escola a Springfield, Massachusetts, on James era professor d'educació física. A petició del seu superior, James Naismith va crear l'esport per tal de tenir una activitat que es pogués jugar en interiors durant els hiverns. El bàsquet és un esport on competeixen dos equips, cada un format per cinc jugadors, en una pista rectangular generalment coberta. L'objectiu de cada equip és anotar els màxims punts possibles, aquests punts s'aconsegueixen llançant i introduint la pilota dins la cistella. Aquesta cistella consta d'un anella horitzontal ubicada a 3.05m d'alçada de la qual hi penja una xarxa. En el bàsquet, cada cistella o canastra anotada val dos o tres punts, segons la posició des de la qual es realitzi el tir a cistella. També existeix el cas del tir lliure, causat per una falta d'un jugador de l'equip rival, en aquest cas la puntuació és d'un punt. Les normes d'aquest esport són majoritàriament comunes per totes les diferents competicions que existeixen, només solen canviar en matisos.<sup>19</sup>

El bàsquet actualment s'ha convertit en un dels esports més populars a nivell mundial i això ha fet que existeixin una gran quantitat de lligues professionals arreu del món. D'entre totes aquestes, destaca per sobre la que és considerada la millor lliga del món, l'anomenada NBA (acrònim de National Basketball Association). L'NBA va ser creada l'any 1946 als Estats Units com a fusió de dues grans organitzacions rivals, la National Basketball League (fundada en 1937) y la Basketball Association of America (fundada en 1946). Actualment, gaudeix d'una gran reputació i seguiment gràcies a tenir els millors jugadors del món repartits en els seus equips. La lliga de l'NBA reuneix un total de 30 equips que estan dividits en dos conferències: la conferència Est i la conferència Oest. Durant la temporada regular, els equips de cada conferència juguen un total de 82 partits, 41 com a local i 41 com a visitant. Després d'això, els vuit millors equips de cada conferència passen a jugar els anomenats playoffs, una sèrie eliminatòria per tal d'acabar proclamant un equip com a campió de l'NBA. Els equips s'enfronten a sèries al millor de 7 partits i cada guanyador avança a la següent ronda. L'última ronda enfronta al campió de cada conferència per tal de proclamar el campió de l'NBA. Quant a la normativa, l'NBA té la particularitat que els seus partits tenen una durada total de 48 minuts, ja que els seus 4 quarts es divideixen en 12 minuts, a diferència dels 10 minuts habituals.<sup>18</sup>

## 2.2. Posicions de joc

Des dels seus inicis, el bàsquet ha sigut un esport d'equip, format per 5 jugadors. En un inici aquests 5 jugadors es dividien en 3 classes: 2 defenses, 2 davanters i 1 centre. Actualment, s'estableixen 5 posicions separades, definides com a Base, Escolta, Aler, Ala-pivot i Pivot. A cada posició se li atribueix un número per fer més fàcil la seva identificació, d'aquesta manera el Base és l'1, l'escorta el 2, l'aler el 3, l'ala-pivot el 4 i el pivot el 5. A mesura que s'ha anat professionalitzant el joc, s'ha arribat a la conclusió que no tots els jugadors han de ser iguals en un equip. De fet, resulta més beneficiós tenir jugadors diferents que puguin adoptar diversos rols i posicions dins de la pista. La diversitat de rols i posicions també ajuda a una millor distribució de les responsabilitats en l'equip. Tenir jugadors que s'especialitzin en diferents àrees del joc permet cobrir millor totes les necessitats tàctiques del joc, cosa que fomenta una millor coordinació entre els jugadors i afavoreix a l'èxit col·lectiu. A més a més, també et permet tenir diferents registres com a equip, cosa que ajuda a ser un equip més imprevisible i difícil de guanyar pel rival. En aquestes posicions, normalment també venen adherides una sèries de característiques físiques com a jugadors, així doncs a mesura que creix el número de la posició, també sol créixer l'alçada i pes del jugador.<sup>13</sup>

1. **Base (Point Guard, PG):** És l'encarregat de dirigir i manar en el joc de l'equip. Necessiten tenir una gran habilitat amb la pilota a les mans, ja que habitualment les possessions de l'equip començaran passant per ells. Habitualment es demana que tinguin una gran visió de joc i gestió del partit, per la qual cosa solen acumular moltes assistències als seus companys. En atac, també acostumen a dominar molt el tir de mitjana distància o de triple. Físicament, tendeixen a ser jugadors molt àgils però amb poca alçada i pes.<sup>12</sup>
2. **Escorta (Shooting Guard, SG):** Perfil de jugador tirador, que pot finalitzar de moltes maneres a cistella. Idealment, hauria de controlar el tir de 2 com el de 3 punts, així com també poder finalitzar a prop de la cistella. Requereix també tenir una bona habilitat amb la pilota a les mans per tenir recursos a nivell ofensiu i recolzar el base quan ho necessiti. En l'àmbit físic acostumen a ser dels jugadors més baixets de l'equip, només per davant dels bases, també solen ser jugadors físicament molt atlètics, amb una bona capacitat de salt.
3. **Aler (Small Forward, SF):** Acostuma a ser el jugador més polivalent de l'equip, amb un físic prominent i alt però mantenint-se àgil. En el seu rol versàtil anoten fent penetracions a cistella i dominant el tir. Es caracteritzen per ser habitualment els millors defensors i per poder defensar a tota classe de jugadors. Per altra banda, també

s'identifiquen per tenir una tasca important a l'hora d'ajudar a capturar rebots, tant ofensius com defensius.

4. **Ala-Pivot (Power Forward, PF):** Es caracteritza per ser un tipus de jugador molt físic, que normalment es mou per dins la línia de tres punts. Tot i ser normalment el segon jugador físicament més potent de l'equip, també és important que sigui hàbil. Les seves habilitats ofensives solen ser les de finalitzar a prop de la cistella, encara que normalment també són uns grans dominants del tir de mitjana distància de 4-5 metres. Defensivament, solen defensar a prop de la cistella intentant condicionar els tirs o capturant els rebots que se'n generen.
5. **Pivot (Center, C):** Tradicionalment, és el jugador més alt i físicament més fort de l'equip, encara que amb un físic menys mòbil. Es caracteritzen per jugar a prop de la cistella i anotar tirs des de curta distància. Una de les seves tasques principals és també capturar rebots i condicionar els tirs dels rivals amb la seva gran envergadura.

### *2.3. Índex d'avaluació d'impacte de jugador*

La idea que hi ha darrere del concepte de l'índex d'avaluació d'impacte de jugador està en la base d'aquest treball, així que és necessari aprofundir en la idea que hi ha darrere d'aquest concepte. Primer de tot, és important recordar que una mètrica all-in-one és aquesta que intenta resumir de manera holística la contribució d'un jugador en múltiples àrees del bàsquet, és a dir, busca resumir el rendiment global d'un jugador en una única mesura.

Els innovadors Dean Oliver i John Hollinger van ser els pioners en promoure l'anàlisi dels jugadors en funció dels minuts jugats en lloc de fer-ho pels partits jugats i van desenvolupar les primeres mesures per avaluar l'impacte i el valor general d'un jugador, l'índex d'eficiència de jugador (PER: Player efficiency rating)<sup>5</sup>. Avui en dia, no hi ha consens per saber quina és la millor all-in-one mètric, cosa que comporta que entitats, equips i professionals de l'esport, estiguin en constant recerca per intentar obtenir la fórmula perfecta. En la realització d'aquest treball s'han seleccionat 6 índexs diferents. Dins del gran mercat que hi ha de mètriques que intenten avaluar aquest concepte, s'ha triat aquestes 6 per tal d'intentar representar l'evolució d'aquests propis índexs, recollint els més senzills i tradicionals i els més revolucionaris i innovadors. Cal destacar també que la manera de calcular aquests índexs no sempre és coneguda, ja que sovint la fórmula és confidencial per no ser replicada. Els 6 índexs escollits són aquests:



- **PLUSMINUS: (+/-)** És una mètrica creada i introduïda per l'NBA la temporada 2007 a la lliga. És utilitzada per mesurar la contribució d'un jugador a l'equip d'una manera molt senzilla. El seu nom prové explícitament del que representa: el més (+) s'encarrega de mesurar la quantitat de punts que anota l'equip quan el jugador avaluat és a pista i el menys (-) mesura els punts rebuts quan el jugador avaluat està a pista. Així doncs, la diferència entre el més i el menys acaba resultant l'índex PlusMinus. D'aquesta manera com més és alt sigui el valor del PlusMinus, millor serà l'actuació del jugador i més impacte tindrà en l'equip. Tot i la seva popularitat, l'índex PlusMinus és altament criticat, ja que la seva anàlisi està clarament esbiaixat i és poc fiable. És evident que és un índex massa ampli per determinar l'impacte d'un jugador, donat que té en compte molt pocs factors. Imaginem per exemple la situació hipotètica en què un jugador és tan important en un equip que juga tots els minuts que dura un partit, si el partit s'acaba perdent de 1, aquest jugador tindrà un PlusMinus igual a -1. Al mateix temps, un jugador menys rellevant d'aquest mateix equip, només juga 3 minuts del partit, però just en aquest interval de temps l'equip fa un parcial de 5 punts a favor. En aquesta situació, el jugador acabarà amb un PlusMinus igual a +5, encara que el partit s'hagi acabat perdent per 1 punt.<sup>6</sup>
- **PER (Player efficiency rating):** Aquest índex va ser creat pel columnista d'ESPN John Hollinger. Pioner en aquest tipus d'estadístiques, els seus llibres revelen que la seva base de dades s'estén fins a la temporada 1988-89, encara que el va introduir de manera oficial per ESPN l'any 2007. Gaudeix d'una fórmula molt complexa per la seva antiguitat, que s'ajusta per ritme i per minut. L'índex PER és segurament l'índex més popular i comú, ja que a banda de la seva validesa, la seva interpretació és molt senzilla. Tot i això, també presenta algunes limitacions, cosa que comporta que per molts professionals no sigui escollit com l'índex de referència.<sup>20</sup> Com més alt sigui el PER, més alt és l'impacte d'aquell jugador a l'equip. D'altra banda, també és sabut que el PER mitjà de la lliga és de 15.00, per la qual cosa és fàcil comparar jugadors o saber si un jugador té un PER elevat. La fórmula del PER és la següent:<sup>1</sup>

$$\begin{aligned}
 uPER = (1 / MP) * [ & 3P + (2/3) * AST + (2 - factor * (team\_AST / team\_FG)) * FG + \\
 & (FT * 0.5 * (1 + (1 - (team\_AST / team\_FG)) + (2/3) * (team\_AST / team\_FG))) - VOP \\
 & * TOV - VOP * DRB\% * (FGA - FG) - VOP * 0.44 * (0.44 + (0.56 * DRB\%)) * (FTA - FT) \\
 & + VOP * (1 - DRB\%) * (TRB - ORB) + VOP * DRB\% * ORB + VOP * STL + VOP * DRB\% \\
 & * BLK - PF * ((lg\_FT / lg\_PF) - 0.44 * (lg\_FTA / lg\_PF) * VOP) ]
 \end{aligned}$$

<b>MP</b>	Minuts jugats	<b>TOV</b>	Pèrdues
<b>3P</b>	Triples anotats	<b>DRB%</b>	Percentatge de rebots defensius
<b>AST</b>	Assistències	<b>TRB</b>	Rebots totals (ofensius i defensius)
<b>team_AST</b>	Assistències de l'equip	<b>ORB</b>	Rebots defensius
<b>team_FG</b>	Tirs de camp de l'equip	<b>STL</b>	Recuperacions
<b>FG</b>	Tirs de camp	<b>BLK</b>	Taps
<b>FGA</b>	Tirs de camp intentats	<b>PF</b>	Faltes comeses
<b>FT</b>	Tirs lliures anotats	<b>lg_FT</b>	Tirs de camp de la lliga
<b>FTA</b>	Tirs lliures intentats	<b>lg_FTA</b>	Tirs de camp intentats a la lliga
<b>VOP</b>	Valor de la possessió (en referència a la lliga)		

Taula 1. Llegenda dels components de la fórmula de uPER

$$\begin{aligned}
 \text{factor} &= (2 / 3) - (0.5 * (\lg\_AST / \lg\_FG)) / (2 * (\lg\_FG / \lg\_FT)) \\
 \text{VOP} &= \lg\_PTS / (\lg\_FGA - \lg\_ORB + \lg\_TOV + 0.44 * \lg\_FTA) \\
 \text{DRB\%} &= (\lg\_TRB - \lg\_ORB) / \lg\_TRB \\
 \text{pace adjustment} &= \lg\_Pace / \text{team\_Pace} \\
 \text{estimated pace adjustment} &= 2 * \lg\_PPG / (\text{team\_PPG} + \text{opp\_PPG}) \\
 \text{aPER} &= (\text{pace adjustment}) * \text{uPER} \\
 \boxed{\text{PER} = \text{aPER} * (15 / \lg\_aPER)}
 \end{aligned}$$

- **PIE (Player Impact Estimate):** Creada per la mateixa lliga NBA, la mètrica PIE busca ser una versió de l'índex PER. Essencialment, busca donar-li més valor i pes a la defensa al mateix temps que implica quasi totes les categories estadístiques del box score dins de la seva fórmula. La seva fórmula bastant senzilla i la seva fàcil interpretació n'ha propiciat l'ús i la seva popularització, així com també la seva fàcil comparació amb altres índexs. El PIE busca calcular la diferència entre el rendiment de l'equip mentre el jugador estudiat és a pista respecte al rendiment de l'equip quan no és a pista. El valor resultant d'això, es normalitza perquè vagi de 0 a 100. D'aquesta manera, la suma de tots els valors dels jugadors de l'equip ha d'acabar donant 100. La fórmula del PIE és:

$$\begin{aligned}
 \text{PIE} = & (PTS + FGM + FTM - FGA - FTA + \text{Deff.REB} + \text{Off.REB}/2 + AST + STL + BLK/2 - \\
 & PF - TO) / (\text{Game.PTS} + \text{Game.FGM} + \text{Game.FTM} - \text{Game.FGA} - \text{Game.FTA} + \\
 & \text{Game.Deff.REB} + \text{Game.Off.REB}/2 + \text{Game.AST} + \text{Game.STL} + \text{Game.BLK}/2 - \\
 & \text{Game.PF} - \text{Game.TO})
 \end{aligned}$$

<b>PTS</b>	Punts	<b>Game.PTS</b>	Punts del partit
<b>FGM</b>	Tirs de camp anotats	<b>Game.FGM</b>	Tirs de camp anotats del partit
<b>FTM</b>	Tirs lliures anotats	<b>Game.FTM</b>	Tirs lliures anotats del partit
<b>FGA</b>	Tirs de camp intentats	<b>Game.FGA</b>	Tirs de camp intentats del partit
<b>FTA</b>	Tirs lliures intentats	<b>Game.FTA</b>	Tirs lliures intentats del partit
<b>Deff.REB</b>	Rebots defensius	<b>Game.Deff.REB</b>	Rebots defensius del partit
<b>Off.REB</b>	Rebots ofensius	<b>Game.Off.REB</b>	Rebots ofensius del partit
<b>AST</b>	Assistències	<b>Game.AST</b>	Assistències del partit
<b>STL</b>	Recuperacions	<b>Game.STL</b>	Recuperacions del partit
<b>BLK</b>	Taps	<b>Game.BLK</b>	Taps del partit
<b>PF</b>	Faltes comeses	<b>Game.PF</b>	Faltes del partit
<b>TO</b>	Pèrdues	<b>Game.TO</b>	Pèrdues del partit

*Taula 2. Llegendra dels components de la fórmula de PIE*

- **RPM (Real Plus Minus):** Creada el 2014 per Jeremias Engelmann i popularitzada per l'ESPN, parteix de la base de l'estadística PlusMinus. En aquest cas per això, l'RPM intenta mesurar l'impacte del jugador a partir de la diferència de punts del teu equip per cada 100 possessions ofensives i defensives. Degut a la diferència d'estils de joc entre equips, neix el concepte de possessió. Aquest concepte t'elimina el problema que els equips juguin a diferents ritmes i sigui injust comparar variables usant mitjanes per partit. S'entén que una possessió s'acaba quan el teu rival recupera la pilota, sigui perquè has realitzat un tir o hagis perdut la pilota. Més enllà d'això, el seu mètode de càlcul és una caixa negra i no es coneix de manera pública.
- **RAPTOR:** Creada per l'estadístic Nate Silver per FiveThirtyEight, l'índex RAPTOR va començar a introduir-se de manera oficial l'any 2018 i 2019, encara que la versió completa està disponible des de 2013-2014. Aquesta mètrica que gaudeix d'un gran prestigi degut a la seva gran robustesa, està basat en el mètode del Plusminus. Concretament, intenta veure l'impacte del jugador quan està dins la pista, mesurant el nombre de punts que aporta a l'atac o a la defensa de l'equip per cada 100 possessions en relació a un jugador mitjà de la lliga. Tal com expliquen els seus creadors, l'índex RAPTOR té dos objectius principals: en primer lloc, intenta ser una mesura disponible públicament que aprofiti les dades modernes de l'NBA, especialment el seguiment dels jugadors.<sup>14</sup> Per altra banda, pretén ser una estadística que reflecteix millor com actualment els equips de l'NBA avaluen realment als jugadors. Sota aquestes dos premisses, el càlcul de la mètrica RAPTOR està compost per una fórmula molt complexa, que aprofita milers d'estadístiques descriptives molt específiques com

pèrdues de pilota per aïllament o distància recorreguda per defensor al perímetre. La modelització i creació de la fórmula de l'índex és explicada més detalladament pels seus creadors, FiveThirtyEight en un article. D'aquesta manera, un índex RAPTOR positiu suggereix que un jugador té un impacte general positiu en el rendiment del seu equip, tant en atac com en defensa. Un RAPTOR negatiu indica que un jugador té un impacte general negatiu en el rendiment del seu equip.<sup>9</sup>

- **LEBRON:** L'índex LEBRON (acrònim de Luck-adjusted Player Estimate using a Box prior Regularized ON-OFF) és una de les noves mètriques que s'han fet més rellevants en els últims anys. Creada per BBALL-Index.com, utilitza de manera innovadora resultats del box-score i la metodologia PlusMinus ajustats per rols i pel component de la sort per obtenir una puntuació d'impacte mesurada per cada 100 possessions. Per altra banda, els seus creadors empren un model estadístic complex per intentar contextualitzar les actuacions i rendiment dels jugadors. La seva interpretació és senzilla, un valor de l'índex LEBRON per sobre de zero indica que un jugador té un impacte per sobre de la mitjana de la lliga, mentre que un valor negatiu indica un impacte per sota de la mitjana. BBALL- Index expliquen detalladament l'índex i els seus components en una entrada a la seva pàgina web.<sup>10,11</sup>

### 3. ESTRUCTURA DE LA BASE DE DADES

#### 3.1. *Obtenció de les dades*

Per tal de realitzar l'estudi, es necessita obtenir les dades de la temporada passada, la 2021-2022, de tots els jugadors de la NBA (National Basketball Association). D'aquestes observacions s'ha de disposar sobre un conjunt d'informació biològica dels jugadors, juntament amb les seves dades i estadístiques professionals. Un cop fet això, caldrà buscar les dades sobre tots els índexs que es volen analitzar. És important destacar que, amb la intenció d'obtenir influència i prestigi, entitats, associacions, o professionals de l'esport lluiten per demostrar la superioritat del seu índex respecte als de la competència. Aquest clima de competència, fa impossible trobar les dades unificades, per la qual cosa s'ha d'indagar per trobar i ajuntar cada un dels índexs en una base de dades global.

Les dades descriptives de cada jugador, així com totes les seves estadístiques mitjanes sobre la temporada 2021-2022 s'aconsegueixen de manera senzilla des de la pàgina web oficial de la lliga. L'apartat d'estadístiques d'aquesta pàgina web et permet accedir i utilitzar lliurement tota aquesta informació. S'obtenen unes 605 observacions, corresponent a tots els jugadors que han jugat durant la temporada 2021-2022. S'aconsegueix informació bàsica com l'equip on el que juguen i el nombre total de victòries i derrotes. Més concretament, d'aquí també s'extreu informació d'estadístiques simples que s'inclouen en un box score, és a dir, estadístiques com els minuts jugats, punts, tirs de camps anotats, tirs de camp intentats, percentatge de tirs de camp, triples anotats, triples intentats, percentatge de triples, tirs lliures anotats, tirs lliures intentats, percentatge de tirs lliures, rebots totals, rebots ofensius, rebots defensius, assistències, pèrdues, recuperacions, taps i faltes comeses. També es contemplen dades biogràfiques com l'edat, l'alçada, el pes, la nacionalitat, el centre de formació, l'any que van ser escollits al draft, la ronda que en la que van ser escollits al draft i el número en el qual van ser escollits al draft.

Pel que respecte als índexs, la pàgina web oficial de l'NBA també considera la mètrica del PlusMinus i l'índex PIE. Això fa que també sigui fàcil d'obtenir aquesta informació per qualsevol usuari. A l'hora d'obtenir l'índex PER i RPM, distribuïdes ambdós casos per l'ESPN, s'extreuen els valors directament des de la web oficial. En aquest cas, l'ESPN no facilita l'obtenció d'aquestes dades, davant d'aquesta situació de dificultat per recollir les dades des de la pàgina web oficial, s'ha escollit el webscrapping com a solució viable per recopilar la informació desitjada de manera eficient. Aquesta eina consisteix en l'extracció d'informació de manera automàtica mitjançant la programació. La realització d'aquest procés s'ha fet a través del software d'R. S'ha pres aquesta decisió degut a la gran versatilitat del llenguatge de

programació d'R, juntament amb la gran quantitat de llibreries i paquets específics per fer aquest tipus de tasques que ofereix. Concretament, s'ha decidit utilitzar el paquet "rvest",<sup>3</sup> ja que ofereix funcionalitats avançades per navegar i obtenir el contingut desitjat. Com a dificultat afegida, és important destacar que la taula de la web oficial d'ESPN que conté la informació desitjada es troba segmentada en un sistema de paginació de taula de 12 fulls. Per resoldre-ho, s'ha creat un bucle previ que inclou tot el procés de webscrapping, tenint en compte que l'URL a analitzar va canviant lleugerament. Això va permetent extreure la informació progressivament, pàgina per pàgina, al mateix temps que va sobreescrivint la base de dades que es va extreient per fer-la cada cop més gran.

S'actua de manera igual per tal d'aconseguir la base de dades que conté els valors de l'índex LEBRON. Els valors d'aquest índex es troba dins de la pàgina web de BBall Index, empresa que va desenvolupar l'índex, que tampoc ofereix la possibilitat de descarregar les dades del seu índex. El procediment és el mateix, senzillament canvia l'URL utilitzat a l'hora de fer el procés de webscrapping. Cal afegir, que en aquest cas no existeix la problemàtica de tenir una taula amb paginació, per la qual cosa s'extreu d'un sol cop tota la informació.

En aquests processos de webscrapping, cal remarcar que és important tenir en compte les restriccions legals i polítiques de cada pàgina web. És necessari assegurar-se de no estar incomplint els termes i normatives d'ús. S'ha consultat prèviament en les dues webs on s'ha emprat aquest procediment i no incompleix en cap cas la normativa.

Pel cas de l'índex RAPTOR, s'accedeix a la pàgina oficial de fivethirtyeight, creador i distribuïdor oficial. Dins d'aquesta web oficial, es permet accedir a totes aquestes dades de manera lliure mitjançant Github. De manera simplificada, Github és una plataforma de desenvolupament col·laboratiu de codi, que permet compartir projectes. D'aquesta manera s'aconsegueix descarregar directament la base de dades amb tots els valors de l'índex d'interès.

### *3.2. Tractament i selecció de les variables d'interès*

La base de dades obtinguda inicialment consta de 605 observacions, però després de realitzar-se una neteja de les observacions amb informació faltant, s'obté una base de dades de 557 observacions i 47 variables. Cada observació pertany a un jugador que hagi jugat en algun moment durant la temporada 2021-2022 del qual es té tot el conjunt de variables. Pel que fa a les diferents variables, la base de dades inicial conté informació sobre 3 aspectes generals de cada jugador. En primer lloc, estadístiques esportives de la temporada, que es troben en forma de recompte en els casos de les variables de nombre de partits jugats, de victòries i de derrotes i, per altra banda, també en forma de mitjana de la temporada, com les variables com els minuts jugats, punts, tirs de camps anotats, tirs de camp intentats, percentatge de tirs de

camp, triples anotats, triples intentats, percentatge de triples, tirs lliures anotats, tirs lliures intentats, percentatge de tirs lliures, rebots totals, rebots ofensius, rebots defensius, assistències, pèrdues, recuperacions, taps i faltes comeses. El segon subgrup definiria informació sobre el jugador, de caràcter esportiu, biogràfic i antropomètric. Són variables com l'equip, la posició, la nacionalitat, l'edat, l'alçada, el pes, el centre de formació, l'any que van ser escollits al draft, la ronda que en la que van ser escollits al draft i el número en el qual van ser escollits al draft. Per últim, es troba el grup de variables que recullen la informació relativa a tots els índexs de mesura d'impacte del jugador, que es presenten de manera mitjana per tota la temporada. Aquest conjunt conté les variables PLUSMINUS, PER, RAPTOR, RPM, PIE i LEBRON.

Partint d'aquesta estructura, es decideix tractar la base de dades per tal d'obtenir una base de dades final que reculli la informació important, recopilant les variables que poden ser d'interès. En un inici, es procedeix a examinar les múltiples observacions que hi ha, fet que fa destapar que existeixen jugadors amb molt pocs partits jugats i això pot interferir enormement a l'hora d'analitzar els índexs. Quan un jugador participa en un nombre molt limitat de partits, resulta difícil veure i avaluar amb precisió els seus índexs d'impacte com a jugador, ja que molts altres factors poden influir en aquests resultats. Així doncs, tenint en compte que durant la lliga regular de l'NBA es juguen un total de 82 partits, s'estipula com a mostra prou representativa aquells jugadors que han jugat un mínim d'aproximadament un 25% dels partits durant la temporada 2021-2022. D'aquesta manera es decideix excloure totes les observacions que pertanyin a jugadors que hagin jugat menys de 20 partits, que equival a haver jugat en un 24.39% dels partits de la lliga regular. Un cop fet això, s'acaben separant 114 jugadors, per la qual cosa la base de dades passa a tenir 443 observacions.

De la totalitat de variables de les quals es disposa, hi ha una gran quantitat que, com s'ha apuntat anteriorment, donen informació sobre estadístiques esportives de la temporada. Aquest conjunt de variables no és d'interès per aquest estudi, així que la majoria són descartades. De fet, totes aquest grup d'estadístiques tenen influència en el valor resultant dels índexs, però ja apareixen representades de manera implícita, ja que es fan servir en les seves fórmules de càlcul. Tot i això, es conserva la variable que recull els minuts jugats mitjans per partit durant la temporada. La raó recau en considerar que el factor dels minuts jugats pot ser un factor condicionant i amb una influència significativa en l'explicació dels índexs d'avaluació d'impacte dels jugadors. Estudis anteriors parlen dels punts febles d'aquests índexs, posant èmfasis en la dificultat que tenen per no veure's condicionats per la quantitat de minuts que juguen els jugadors.<sup>20</sup> La inclusió d'aquesta variable permetrà tenir en compte aquest factor i ajustar adequadament les anàlisis posteriors.

Del segon subgrup de variables, es manté la variable que ens indica l'equip de cada jugador, ja que és un factor que intuïtivament porta a pensar que tindrà efecte en l'actuació dels jugadors i al seu impacte. No s'ha de perdre de vista que el basquetbol és un joc col·lectiu d'equip, per la qual cosa incloure el component de l'equip és crucial per tal d'analitzar el possible efecte dels companys d'equip, de l'estil de joc o dels entrenadors en l'impacte que acaba produint un jugador.

Per motius evidents, també es conserva la variable que recull la informació sobre la posició de cada jugador. La variable posició presenta en un inici una classificació en 10 factors, on a part de la diferenciació clàssica en 5 posicions: C, PF, PG, SF,SG, s'inclouen també jugadors amb posicions intermèdies que prenen valors de C-PF, PF-SF, PG-SG, SF-SG i SG-PG. Davant d'aquesta situació, es pren la decisió de recodificar la variable per tal que només existeixin les 5 categories referents a les posicions tradicionals. Aquesta modificació simplificarà l'ús i la interpretació de la variable, al mateix temps que reflectirà millor la realitat actual del bàsquet. A més, les 5 posicions són àmpliament reconegudes i acceptades, fet que farà més senzilla la comprensió i comparació amb altres estudis o treball. La presència de les categories que representen les categories intermèdies és mínima, ja que només representa un total de 17 jugadors. Això permet que la seva recodificació es fa de manera individualitzada perquè sigui més precisa.

Per altra banda, també s'aprofita la variable que recull l'edat del jugador. Aquesta variable de caràcter biogràfic, s'ha decidit incloure-la per la seva possible influència al valor dels índexs d'avaluació d'impacte dels jugadors. L'edat és rellevant per tal de controlar l'afecte de l'experiència en el joc adquirida. A mesura que els jugadors envelleixen, poden experimentar canvis en el seu estat o condició física. De la mateixa manera, els jugadors més joves solent tenir més vitalitat i tenir una motivació més alta per voler demostrar les seves habilitats. A part d'això, l'edat també comporta canvis a nivell personal i això també pot afectar directament al rendiment esportiu. Per tal de simplificar aquest factor, es recodifica la variable numèrica que es té en un primer moment per crear només 3 factors. Aquestes categories busquen agrupar els jugadors en 3 franges d'edat segons el moment vital en què estar el jugador. La primera categoria agrupa tots els jugadors molt joves de la lliga, considerats jugadors novells que no tenen massa experiència. Aquests són tots els jugadors que tenen 23 o menys anys. La segona categoria reuneix el volum més gran de jugadors de la lliga. Es tracta de la categoria que busca agrupar tots els jugadors que estan en l'etapa madura de la seva carrera, que comprèn entre els 24 i els 31 anys. Per últim, es troba la categoria que ajunta els jugadors veterans de la competició i que, per tant, tenen 32 o més anys. Resumidament, s'acaba tenint la variable en 3 categories: 23 o menys, de 24 a 31, 32 o més, corresponents als nivells de maduresa de tots els jugadors.



Sembla important també poder incloure una variable que ens parli de la procedència o origen dels jugadors. Es busca la manera d'incloure d'alguna manera aquest factor d'entre les variables de caràcter biològic que queden. Trobem la variable "País de procedència", que ens parla de la nacionalitat de cada jugador. Tot i això, la procedència d'un jugador no sempre reflecteix el lloc on ha viscut i s'ha format el jugador en concret. És per aquest motiu que s'hauria de trobar una variable que reflectís la manera en què el jugador ha adquirit les seves habilitats i ha desenvolupat el seu potencial. D'aquesta manera es descobreix la variable "College" referent al Centre d'estudis que indica si el jugador ha rebut o no formació a Estats Units i en cas afirmatiu, a quin centre ho ha fet. D'entre totes les variables es comptabilitzen un total de 140 centres, així que l'anàlisi tenia poc sentit tenint en compte que el màxim de jugadors en un centre eren 27 i amb un 95% dels centres amb menys de 10 jugadors. Es considera que el tret diferencial que pot condicionar l'evolució d'un jugador és si ha estat format a Estats Units o no. Així doncs, es crea una nova variable dicotòmica anomenada "FormacioUsa" que recull amb un "SI" els casos dels jugadors que han estat formats a Estats Units independentment de quin centre hagin anat. La variable pren el valor "NO" en cas contrari.

Finalment, també s'inclouen totes les variables referents als valors dels diferents índexs de mesura d'impacte de jugadors. S'inclou la variable PLUSMINUS, la variable PER, la variable PIE, la variable RPM, la variable OALLTOT referent a l'índex RAPTOR i la variable LEBRON.

### 3.3. Estructura final de la base de dades

Variable	Descripció	Nivells
<b>PLAYER</b>	Nom del jugador	
<b>TEAM</b>	Equip del jugador	Atlanta Hawks(ATL), Boston Celtics (BOS), Brooklyn Nets (BKN), Charlotte Hornets (CHA), Chicago Bulls (CHI), Cleveland Cavaliers (CLE), Dallas Mavericks (DAL), Denver Nuggets (DEN), Detroit Pistons (DET), Golden State Warriors (GSW), Houston Rockets (HOU), Indiana Pacers (IND), Los Angeles Clippers (LAC), Los Angeles Lakers (LAL), Memphis Grizzlies (MEM), Miami Heat (MIA), Milwaukee Bucks (MIL), Minnesota Timberwolves (MIN), New Orleans Pelicans (NOP), New York Knicks (NYK), Oklahoma City Thunder (OKC), Orlando Magic (ORL), Philadelphia 76ers (PHI), Phoenix Suns (PHX), Portland Trail Blazers (POR), Sacramento Kings (SAC), San Antonio Spurs (SAS), Toronto Raptors (TOR), Utah Jazz (UTA), Washington Wizards (WAS)
<b>MIN</b>	Minuts jugats de mitjana	
<b>POS</b>	Posició del jugador	Pivot(C), Ala-Pivot(PF), Aler(SF), Escorta(SG), Base(PG)
<b>Age_Factor</b>	Edat del jugador	23 o menys; 24 a 31; 32 o més
<b>FormacioUSA</b>	Informació de si el jugador ha rebut formació als Estats Units	SI, NO
<b>PLUSMINUS</b>	Valors de l'índex PlusMinus	
<b>PER</b>	Valors de l'índex PER	
<b>OALLTOT</b>	Valors de l'índex RAPTOR	
<b>RPM</b>	Valors de l'índex RPM	
<b>PIE</b>	Valors de l'índex PIE	
<b>LEBRON</b>	Valors de l'índex LEBRON	

Taula 3. Estructura final de la base de dades

## 4. METODOLOGIA

### 4.1. Software Utilitzat

L'estudi s'ha realitzat íntegrament amb el llenguatge de programació i software de codi obert d'R, utilitzat per l'anàlisi de dades. S'han usat funcions predeterminades, encara que també molts paquets especialitzats per a diferents àrees de l'estudi. Resumidament, s'ha utilitzat per a la millor visualització de dades el paquet ggplot2, el paquet gridExtra i el paquet ggcorrplot. D'altra banda, per completar i dur a terme l'anàlisi destaquem l'ús de la llibreria compareGroups i la llibreria nlme, entre d'altres.

### 4.2. Anàlisi descriptiva i anàlisis de comparació de grups

Per realitzar una anàlisi exploratòria i comparativa de les dades, s'ha començat utilitzant, per a les variables categòriques, la funció de *compareGroups* que es troba dins de la llibreria que porta el mateix nom<sup>15</sup>. S'empra aquesta llibreria amb l'objectiu de facilitar la comparació de característiques i variables entre grups, al mateix temps que es duen a terme anàlisis descriptives per identificar possibles diferències. Es calculen per a cada un dels grups definits dins de cada variable categòrica les proporcions, la mitjana, la desviació tipus i els intervals de confiança.

Hi ha 3 variables categòriques d'interès a analitzar. En primer lloc, hi ha la variable categòrica Posició (POS). En segon lloc, hi ha la variable "Age\_Factor", que recull en forma de 3 categories l'edat del jugador. En últim lloc hi ha la variable "FormacioUSA". Amb aquestes tres variables categòriques es realitza una anàlisi comparativa amb les variables que representen cada un dels índexs (PLUSMINUS,PER, PIE,LEBRON,RPM,RAPTOR).

L'objectiu és avaluar si hi ha diferències significatives en els valors mitjans d'aquestes variables numèriques entre les diferents categories de cada variable, per exemple, en el cas de FormacioUSA, si hi ha diferències entre FormacioUSA= SI i FormacioUSA=NO. Aquesta anàlisi de comparació de mitjanes es fa mitjançant un test-t. A continuació, s'observa també el resum descriptiu de cada una de les variables numèriques en cada categoria. Per visualitzar aquesta anàlisi descriptiva, s'usa també la funció ggplot per un diagrama de mitjanes per a cada índex.

S'inclou també l'anàlisi descriptiva de totes les variables numèriques que fan referència a cada un dels índexs estudiats. Per aquestes variables, s'utilitza la funció *boxplot* per realitzar un boxplot, alhora que es calculen les següents estadístiques descriptives bàsiques: Mínim,

Primer quartil, Mediana, Mitjana, Tercer quartil i Màxim. S'obtenen aplicant la funció `summary` que et retorna un resum de totes aquestes variables.

#### 4.3. Anàlisi estadística: Model Lineal Mixt

Per tal de determinar l'efecte que té la posició del jugador en la seva puntuació en els diferents índexs estudiats, es procedeix a estimar un model lineal mixt. S'escull aquesta opció ja que et permet combinar efectes fixos i aleatoris en un mateix model estadístic, per la qual cosa és molt útil quan es vol fer una anàlisi de dades amb múltiples nivells.

En aquest estudi, s'analitzen tots els jugadors de la lliga NBA durant una temporada, però aquests jugadors competeixen formant part de diversos equips. És per això que s'identifiquen aquests equips com a diferents clústers, i per tant s'introdueix l'efecte aleatori de l'equip. Tenint en compte això, es té en compte aquesta estructura i es permet que la mitjana dels índexs variï aleatòriament entre els diferents clústers. És a dir, s'introdueix l'efecte aleatori de l'equip per tal de controlar les diferències inherents que hi ha entre els equips i controlar aquest efecte en les variables d'interès. Per tal de programar el model lineal mixt s'escull la funció `lme` dins de la llibreria "nlme". D'aquesta manera es fan 6 models diferents, 1 per cada índex d'interès. La fórmula matemàtica que hi ha darrere dels models lineals mixtes és: <sup>17</sup>

$$Y = X\beta + Zb + \varepsilon$$

On:

- Y: vector de variables resposta
- X: matriu que conté els efectes fixos
- $\beta$ : vector dels coeficients dels efectes fixos
- Z: matriu dels efectes aleatoris
- b: vector dels coeficients aleatoris,  $b \sim N(0, S_b)$
- $\varepsilon$ : vector d'errors,  $\varepsilon \sim N(0, S_e)$

En aquest estudi en particular, s'executa un model lineal mixt on l'efecte fix principal és la posició del jugador (variable "POS") i s'integren també com a possibles variables predictores la variable categòrica que dona informació sobre l'edat del jugador (variable "Age\_Factor"), la que informa sobre si el jugador s'ha format a Estats Units (variable "FormacioUSA") i la variable que dona informació sobre els minuts jugats pel jugador (variable "MIN"). Per altra banda, l'efecte aleatori del model serà l'equip on juga cada jugador. Per últim, com a variables resposta, es tindrà cada un dels 6 índexs que es volen estudiar. D'aquesta manera el model lineal mixt executat tindrà la següent forma:

```
lm1<- lme(PLUSMINUS ~ POS + FormacioUSA + Age_factors+ MIN,  
data= total3, random= ~ 1|TEAM)
```

Cal remarcar per això que aquesta serà la forma específica pel primer model, on s'intentarà tastar l'impacte de l'efecte fix sobre la variable resposta "PLUSMINUS". Existiran 6 models com aquests, així que la part que fa referència a la variable resposta (PLUSMINUS en aquest primer cas) anirà canviant per cada un dels diferents índexs.

Seguidament, per tal de testar els efectes fixos del model de manera genèrica, s'utilitzarà la funció *anova* sobre el model generat, per tal de fer una anàlisi de variància marginal per efecte de cada una de les variables predictores del model. Aquesta funció permet realitzar un test F per avaluar la significació dels efectes fixos en el model lineal mixt. Per últim, també s'utilitzarà per realitzar l'anàlisi post hoc la funció *emmeans* sobre el model i les variables predictores que hagin resultat significatives. Això permetrà obtenir les estimacions de les mitjanes marginals per als diferents nivells i combinacions per a les variables predictores en el model lineal mixt. Aquesta funció realitzarà aquestes estimacions utilitzant l'ajust de Tukey.

## 5. ANÀLISI DESCRIPTIVA I DE COMPARACIÓ DE GRUPS

### 5.1. Equip (TEAM)

La variable equip conté 30 categories diferents, atribuïdes als 30 equips diferents que hi ha actualment a la lliga de l'NBA. La variable categòrica ens recull en quin equip ha jugat cada jugador durant la temporada 2021-2022. La distribució de jugadors per equips és la següent:

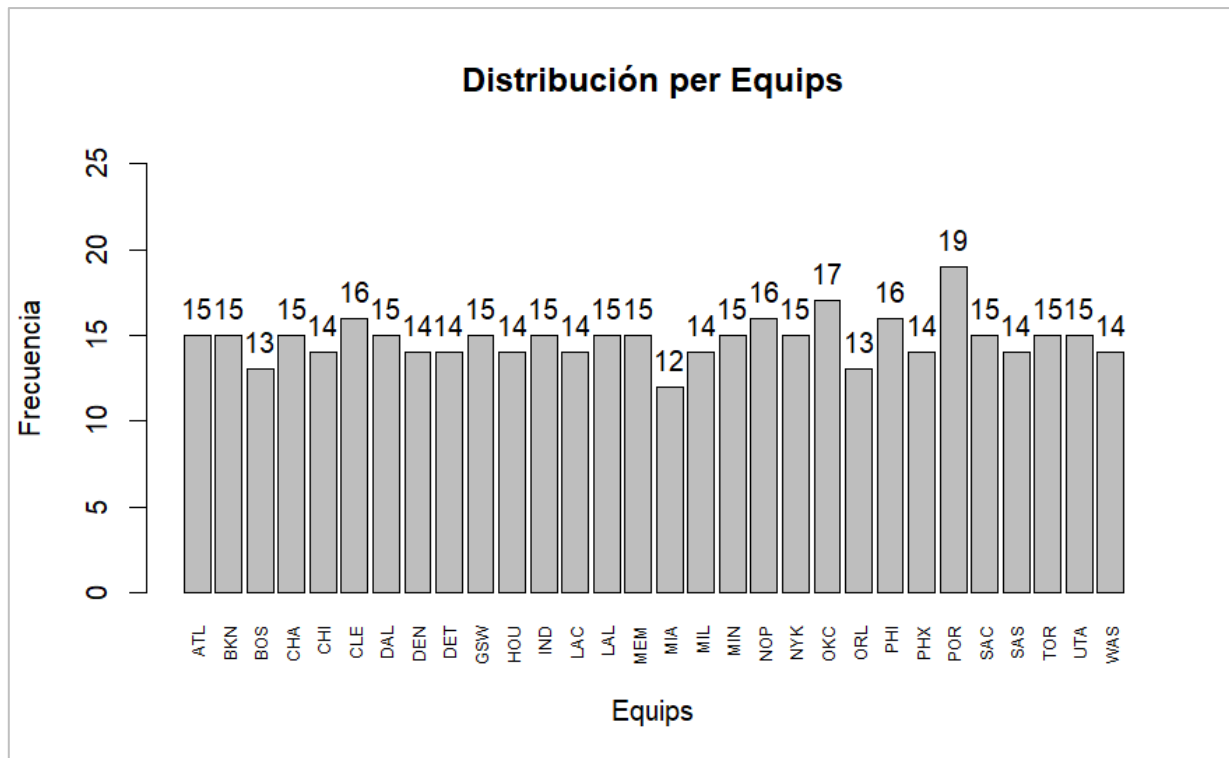


Figura 1. Diagrama de barres sobre l'acumulat del nombre de jugadors per equip

Es pot veure com el recompte dels equips marca com hi ha 13 equips amb 15 jugadors, 9 equips amb 14 jugadors, 3 equips amb 16 jugadors, 2 equips amb 13 jugadors, 1 equip amb 12 jugadors, un altre amb 17 i per últim un altre amb 19.

	Jugadors per equip						
	12	13	14	15	16	17	19
Nombre d'equips	1	2	9	13	3	1	1

Taula 4. Recomppte de la grandària dels equips

D'aquesta informació en podem extreure que el màxim de la variable és l'equip de Portland Trail Blazers (POR) amb 19 jugadors, el mínim està marcat per l'equip de Miami Heat amb només 12 jugadors, mentre que la moda és de 15 jugadors per equip.

	Valor
<b>Mínim</b>	12
<b>Moda</b>	15
<b>Màxim</b>	19

Taula 5. Mínim, moda i màxim de la variable TEAM

## 5.2. Edat (Age\_Factor)

Un cop s'han recodificat els grups d'edat, es pot veure com la distribució dels jugadors segueix un ordre lògic. Per una banda, la categoria "23 o menys" que recull els jugadors en l'etapa de jugadors novells, conté 132 jugadores. Per altra banda, la categoria d'edats entre 24 i 31 conté 243 jugadors. Si tenim en compte que és la categoria que conté un rang d'edats més gran, és lògic que sigui la categoria més poblada. Per últim, hi ha 68 jugadors en la categoria de jugadors més veterans que tenen una edat de "32 o més".

	Categories per edats		
	23 o menys	24 a 31	32 o més
<b>Número de jugadors</b>	132	243	68

Taula 6. Recompte d'observacions en cada categoria de Age\_Factors

S'efectua també un estudi de comparació de grups i una anàlisi de diferència de mitjanes entre grups. En aquesta anàlisi s'intenta veure si existeixen diferències significatives en les variables contínues (els diferents índexs) amb les diferents categories de la variable "Edat". Un cop realitzat, es pot veure com en 5 dels 6 índexs estudiats resulta ser significatiu l'efecte de l'edat. Més concretament, l'índex PlusMinus, RAPTOR i LEBRON es mostren com a més significatius, amb un p-valor associat <0.001. Per altra banda, s'observa amb un p-valor de 0.019 l'índex RPM i per últim amb un p-valor de 0.01 l'índex PIE. Seguidament, s'obté un resum descriptiu de totes les variables referents als diferents índexs agrupades per les tres categories diferents de la variable "Age\_Factor". D'aquesta manera es pot veure de manera clara el valor de les mitjanes de i de les desviacions estàndards associades a cada categoria d'Age\_factors. És destacable com per tots els índexs la mitjana més alta es troba en l'últim rang d'edats, corresponent al grup d'edat més gran que fa referència als jugadors més veterans. Així doncs, per molt que les desviacions estàndard també siguin les més altes, en tots els índexs la mitjana

més alta es troba en els jugadors de 32 o més anys. De fet, és curiós veure com per a tots els índexs la mitjana va creixent a mesura que va creixent també l'edat. Per a tots els índexs, els jugadors més joves en conjunt tenen una mitjana més baixa que el conjunt de jugadors que van de 24 a 31 anys, que al mateix temps tenen com a grup una mitjana més baixa que el conjunt de jugadors que tenen 32 o més anys.

```

----- Summary of results by groups of 'Age_factors'-----

  var      N   p.value  method      selection
1 PLUSMINUS 443 <0.001** continuous normal ALL
2 PER       443 0.120    continuous normal ALL
3 OALLTOT   443 <0.001** continuous normal ALL
4 RPM       443 0.019**  continuous normal ALL
5 PIE       443 0.010**  continuous normal ALL
6 LEBRON    443 <0.001** continuous normal ALL
-----

```

Figura 2. Resultat del test-t d'anàlisi de comparació de mitjanes entre els índexs i Age\_Factors

```

-----Summary descriptives table by 'Age_factors'-----

```

	23 o menys N=132	24 a 31 N=243	32 o més N=68	p.overall
PLUSMINUS	-1.20 (2.97)	0.23 (2.80)	0.69 (2.98)	<0.001
PER	13.5 (4.10)	14.4 (4.58)	14.6 (4.63)	0.120
OALLTOT	-1.83 (3.01)	-0.29 (3.21)	-0.17 (3.21)	<0.001
RPM	-1.71 (2.86)	-0.78 (3.51)	-0.59 (3.64)	0.019
PIE	8.74 (2.69)	9.50 (3.17)	10.1 (3.32)	0.010
LEBRON	-0.86 (1.50)	-0.15 (1.79)	-0.04 (1.89)	<0.001

```

-----

```

Figura 3. Taula descriptiva dels índexs segons les diferents categories de Age\_Factors



El diagrama de mitjanes reafirma i exposa d'una manera més clara el que s'ha vist anteriorment. La línia que dibuixen les diferents mitjanes dels grups és creixent a mesura que van també creixent els grups d'edat. Aquesta pujada no és igual d'exagerada en tots els casos, però és visible en tots els diferents índexs estudiats.

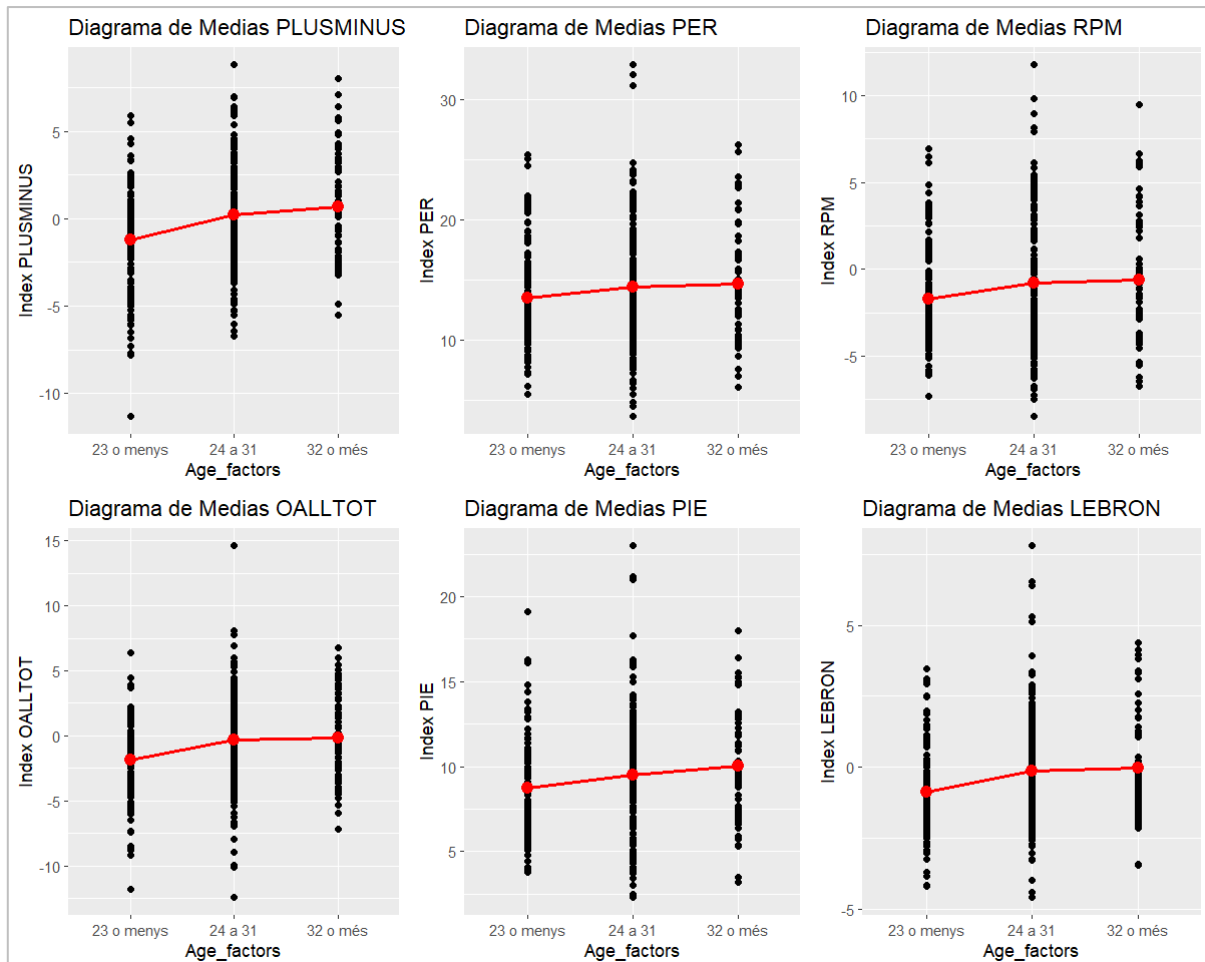


Figura 4. Diagrames de mitjanes entre tots els índexs i Age\_factors

### 5.3. Minuts (MIN)

La variable recull quants minuts de mitjana han jugat els jugadors durant tota la temporada 2021-2022. L'anàlisi descriptiva ens ensenya com la mitjana d'aquesta variable es situa en 22.04, valor molt pròxim als 24 minuts, que reflectiria jugar mig partit. El mínim es situa a 3 minuts, registrat pel jugador Kai Jones dels Charlotte Hornets. A l'altre extrem, el màxim està situat en 37.9 minuts, que el registren Fred VanVleet i Pascal Siakam, ambdós jugadors del mateix equip, els Toronto Raptors.

La forma del boxplot ens mostra molta variabilitat en el conjunt de les dades, fet coherent amb el què la variable reflexa, ja que els minuts jugats es distribueixen en tots els valors possibles.

	Valor
<b>Mínim</b>	3
<b>Quartil 1</b>	15.75
<b>Mitjana</b>	22.0386
<b>Mediana</b>	22
<b>Quartil 3</b>	29.05
<b>Màxim</b>	37.9

Taula 7. Valors del Mínim, Q1, Mitjana, Mediana, Q3 i Màxim de la variable MIN

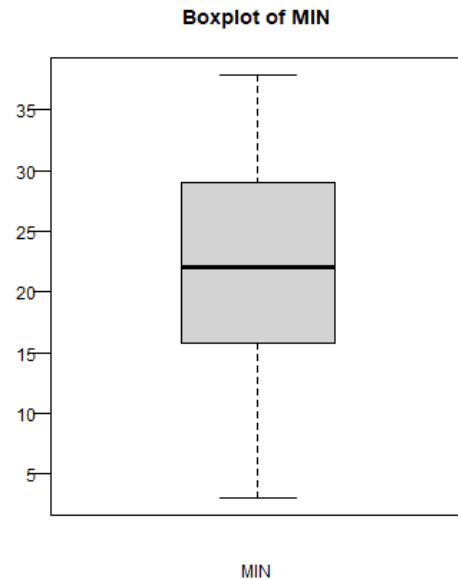


Figura 5. Boxplot de la variable MIN

Els coeficients de correlació no porten a pensar que cap índex estigui molt correlacionat amb els minuts. El valor més alt que s'observa és amb l'índex RPM de 0.76 que no és un valor molt alt. La resta de coeficients es són més o menys semblants, a excepció de l'índex PLUSMINUS que presenta un valor més baix que la resta de 0.25.

Els scatterplots ens donen més informació sobre la relació dels índexs amb la variable que recull els minuts jugats mitjans de cada jugador. Es pot veure com en el cas dels índexs PER, PIE i LEBRON els gràfics presenten una forma molt semblant, amb una forma relació lleugerament lineal i en forma de W. En el cas dels índexs PLUSMINUS i RAPTOR es veu un núvol de punts més dispers i distribuït de manera més horitzontal. Per últim, en el cas de l'índex RPM s'observa com tots els punts es situen més junts, indicatiu que hi ha una relació més forta i s'intueix una relació lineal positiva.

	PLUSMINUS	PER	RAPTOR	RPM	PIE	LEBRON
<b>MIN</b>	0.25	0.45	0.52	0.76	0.51	0.51

Taula 8. Coeficients de correlació dels índexs amb la variable MIN

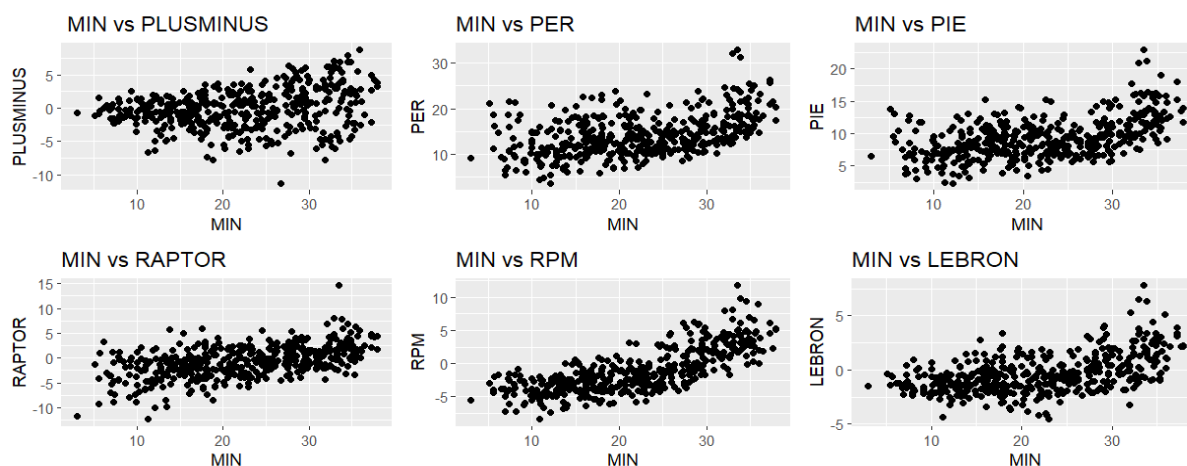


Figura 6. Gràfics de dispersió dels índexs amb la variable MIN.

#### 5.4. Formació a Estats Units (FormacioUSA)

Aquesta variable dicotòmica informa sobre si el jugador ha rebut o no formació als Estats Units. Els jugadors es distribueixen de manera molt desigual entre els 2 grups, on la categoria del SI conté 385 jugadors mentre que la categoria del NO 58 jugadors. És lògic pensar que sigui així, ja que l’NBA és una lliga americana on la majoria dels seus jugadors ho són. A part d’això, és conegut que els Estats Units sempre ha estat un país pioner, que ha deixat empremta en aquest esport a nivell internacional.

Categories	SI	No
Número de jugadors	385	58

Taula 9. Recompte de jugadors per categoria de la variable FormacioUSA

En el cas de l’anàlisi de comparació de grups de variable “FormacioUSA” amb tots els índexs estudiats, no es troben diferències significatives entre les diferents categories. Així doncs, l’efecte d’haver rebut formació als Estats Units no sembla ser significatiu a l’hora d’explicar els valors dels índexs. Tot i així, es pot veure com a l’anàlisi descriptiva, tots els índexs menys el PlusMinus presenten una mitjana major pel grup que recull els jugadors que no han estat formats a Estats Units. D’altra banda, també ressalten els valors alts de les desviacions estàndards, que reflecteixen una alta dispersió de les dades al voltant de les mitjanes. Encara que aquestes desviacions són, en tots els casos, semblants per ambdues categories de FormacioUSA per la qual això dona la visió que la variabilitat no canvia molt entre les dues categories.

```

----- Summary of results by groups of 'FormacioUSA'-----

```

var	N	p.value	method	selection
1 PLUSMINUS	443	0.504	continuous normal	ALL
2 PER	443	0.152	continuous normal	ALL
3 OALLTOT	443	0.865	continuous normal	ALL
4 RPM	443	0.481	continuous normal	ALL
5 PIE	443	0.061*	continuous normal	ALL
6 LEBRON	443	0.445	continuous normal	ALL

Figura 7. Resultat del test-t d'anàlisi de comparació de mitjanes entre els índexs FormacioUSA

```

-----Summary descriptives table by 'FormacioUSA'-----

```

	NO N=58		SI N=385		p.overall
PLUSMINUS	-0.38	(3.12)	-0.09	(2.94)	0.504
PER	15.2	(5.75)	14.0	(4.23)	0.152
OALLTOT	-0.66	(3.77)	-0.74	(3.14)	0.865
RPM	-0.69	(3.98)	-1.08	(3.27)	0.481
PIE	10.2	(3.88)	9.22	(2.93)	0.061
LEBRON	-0.13	(2.28)	-0.37	(1.67)	0.445

Figura 8. Taula descriptiva dels índexs segons les diferents categories de Age\_Factors

Gràficament es pot obtenir una visió més general d'aquesta informació amb el diagrama de mitjanes. Permet veure clarament com en tots els índexs menys el PlusMinus, les mitjanes corresponents al grup dels jugadors que no han sigut formats als Estats Units són més altes que les mitjanes del grup dels jugadors formats a Estats Units. També es veu reflectida la alta variabilitat dels valors dels índexs dins de cada categoria.

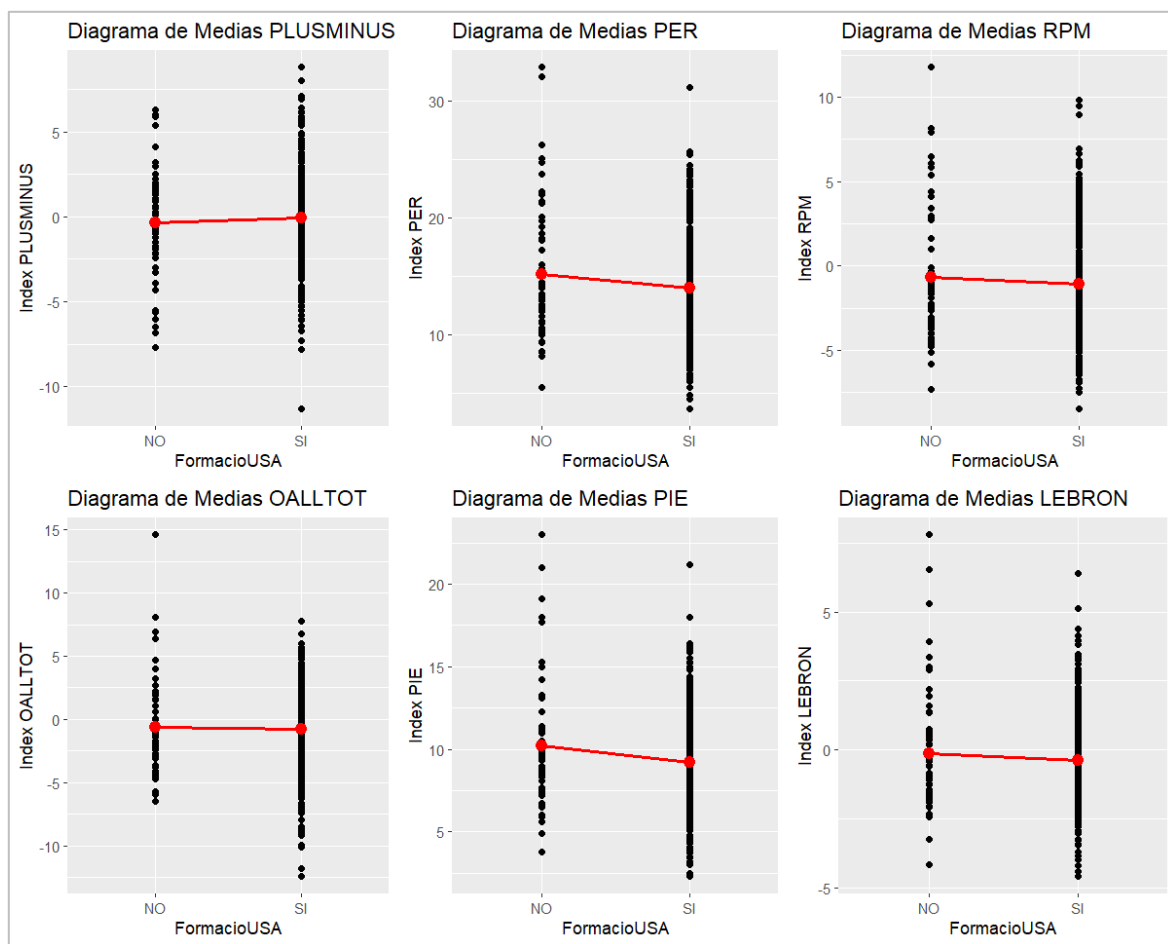


Figura 9. .Diagrames de mitjanes entre tots els índexs i FormacioUSA

### 5.5. Posició (POS)

La distribució per posicions és molt equitativa, trobant una variació molt petita en les diferents categories. Destaca com a màxim la posició d'Escorta (SG) amb un recompte de 103 jugadors, mentre que el mínim ve donat per la posició de Base (PG) amb 83 jugadors. Les tres categories restants es situen molt a prop d'aquests valors, ja que els jugadors que juguen com a Pivots (C) són 84, els jugadors que juguen d'Alers (SF) són 85 i els jugadors que juguen d'Ala-pivots (PF) són 88.

Categories	Pivot (C)	Ala-Pivot (PF)	Aler (SF)	Escorta(SG)	Base (PG)
Número de jugadors	84	88	85	103	83

Taula 10. Recompte de jugadors per cada categoria de la variable POS

En l'anàlisi de comparació de grups, es troben diferències significatives en les mitjanes dels diferents índexs amb les diferents posicions. Dels 6 índexs analitzats en surten 4 de significatius, on destaquen els índexs PER, PIE i LEBRON amb un p-valor associat <0.01 i l'índex RPM amb un p-valor de 0.014. L'anàlisi descriptiva també mostra com en línies generals la posició de Pivot (C) té mitjanes més altes que la resta de posicions, mentre que les mitjanes més baixes les tenen en tots els casos els Escortes (SF). Tot i així, no es veuen grans diferències entre els valors de les mitjanes entre posicions; només en els casos dels índexs PER i PIE existeixen diferències més destacables. També és destacable els baixos valors que presenta la desviació estàndard de l'índex LEBRON respecte als altres índexs, el valor màxim de desviació estàndard que exposa és de 1.95, mentre que fora d'aquest índex el valor mínim de desviació que es troba és de 2.26.

```

----- Summary of results by groups of 'POS' -----

  var      N    p.value  method      selection
1 PLUSMINUS 443 0.368    continuous normal ALL
2 PER       443 <0.001** continuous normal ALL
3 OALLTOT   443 0.113    continuous normal ALL
4 RPM       443 0.014** continuous normal ALL
5 PIE       443 <0.001** continuous normal ALL
6 LEBRON    443 <0.001** continuous normal ALL
-----

```

Figura 10. Resultat del test-t d'anàlisi de comparació de mitjanes entre els índexs i POS

```

-----Summary descriptives table by 'POS'-----

```

	C N=84	PF N=88	PG N=83	SF N=85	SG N=103	p.overall
PLUSMINUS	0.01 (2.79)	0.30 (2.53)	-0.07 (3.44)	-0.23 (3.15)	-0.55 (2.86)	0.368
PER	17.9 (4.38)	14.8 (4.07)	13.8 (4.56)	12.6 (3.83)	12.2 (3.21)	<0.001
OALLTOT	-0.33 (3.77)	-0.96 (3.00)	-0.08 (3.30)	-1.02 (3.20)	-1.15 (2.81)	0.113
RPM	-0.12 (3.61)	-0.83 (3.11)	-0.83 (3.93)	-1.49 (3.18)	-1.71 (2.88)	0.014
PIE	11.5 (3.19)	9.57 (2.83)	9.59 (3.18)	8.32 (2.80)	8.08 (2.26)	<0.001
LEBRON	0.54 (1.88)	-0.12 (1.57)	-0.17 (1.95)	-0.78 (1.56)	-1.03 (1.42)	<0.001

Figura 11. Taula descriptiva dels índexs segons les diferents categories de POS

Amb l'ús del diagrama de mitjanes, s'obté una visió més completa d'aquesta informació. Destaca clarament la diferència de mitjanes entre posicions dels índexs PER i PIE, representada amb una pendent molt més pronunciada de la línia vermella que uneix totes les mitjanes.

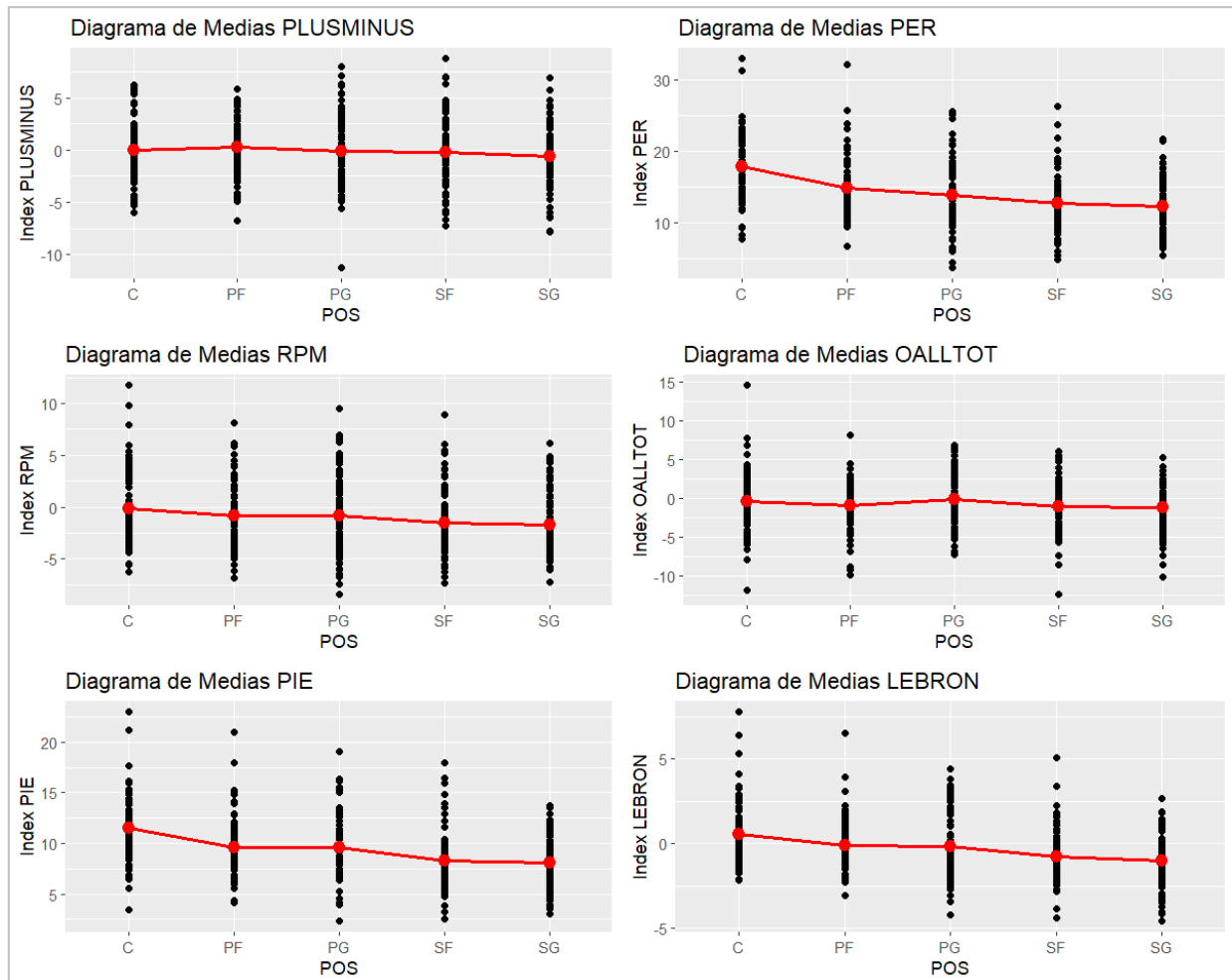


Figura 12. Diagrames de mitjanes entre tots els índexs i la variable POS

## 5.6. Índexs

L'anàlisi descriptiva dels índexs que es volen analitzar porten a treure certes conclusions. En primer lloc, es pot veure com tots els boxplots són semblants, amb un aspecte ample indicador d'alta variabilitat. Per altra banda, es pot apreciar com totes les representacions dels índexs presenten valors extrems, normalment en la part superior del gràfic. Aquest fet és lògic si es té en compte el tipus de variables analitzades. Quan es vol analitzar índex que intenten avaluar l'impacte i el rendiment dels jugadors és normal que apareguin valors extrems, donat que els jugadors excepcionals existeixen i mostren rendiments molt per sobre de la mitjana. Per la qual cosa, és bona senyal que hi hagi valors extrems, ja que reflecteix la veracitat de l'índex.

- **PLUSMINUS:** Destaca un màxim i un mínim molt allunyats, de 8.80 i -11.30 respectivament, ambdós valors extrems. També és important distingir l'estreta caixa del boxplot, on el primer quartil es situa a -2.00 i el tercer quartil a molt poca distància, concretament a 1.650.
- **PER:** El boxplot de la variable PER presenta una forma molt semblant al generat per la variable PIE. La seva mitjana es situa a 14.17, el seu primer quartil a 11.035 i el seu tercer quartil a 16.705. En aquest cas també és destacable la presència de molts valors extrems en la part superior, amb un màxim de 32.94.
- **RAPTOR:** L'anàlisi descriptiva de la variable Raptor ens indica que la variable té un rang de valors d'entre el -12.4 i el 14.6, valors que representen el seu màxim i mínim respectivament. Destaquen el seu màxim 14.6, relatiu al jugador Nikola Jokic, que està molt per sobre de la resta d'observacions. La mitjana de la variable es troba a -0.73 mentre que els seus primer i tercer quartil es troben a -2.65 i 1.5 respectivament.

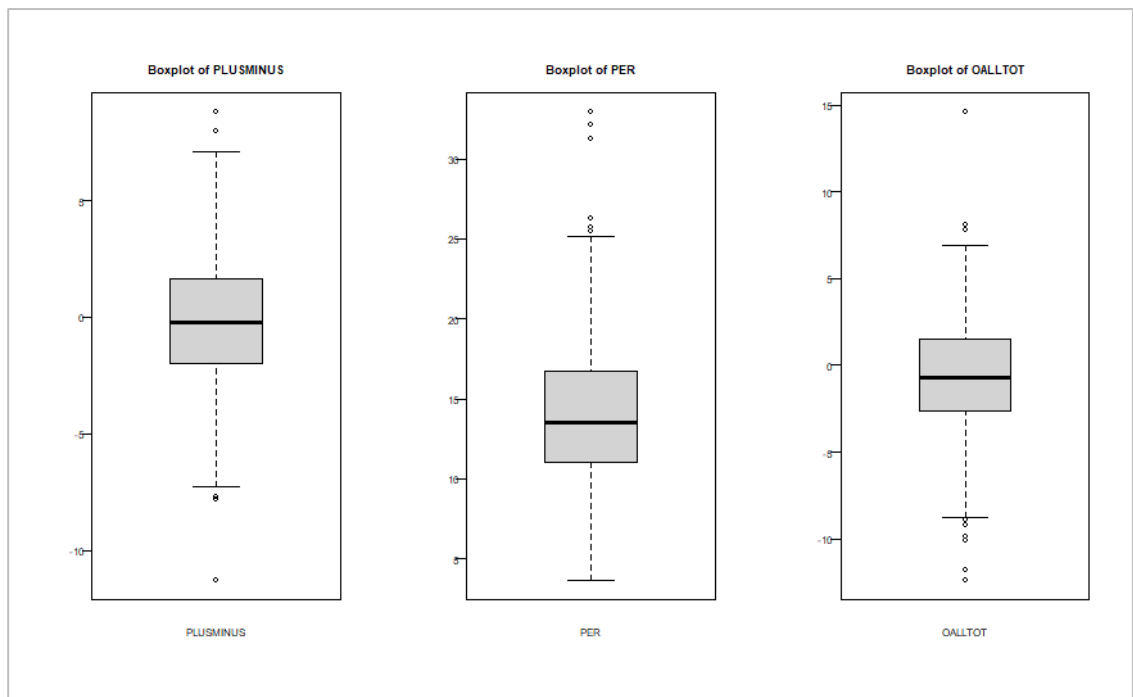


Figura 13. Boxplots dels índexs PLUSMINUS, PER i RAPTOR ordenats d'esquerra a dreta



- **RPM:** Exhibeix una cua més llarga per la part superior que per la part inferior, la qual cosa indica una asimetria de les dades. A més, destaca la presència de molts valors extrems a la part superior del gràfic. El boxplot informa també que la mitjana de l'índex és de -1.02, mentre que el seu màxim i mínim és de -8.46 i 11.78 respectivament.
- **PIE:** El boxplot que presenta la variable PIE ens mostra una caixa estreta, cosa que ens suggereix una concentració en les dades en un rang pròxim al tercer quartil situat a l'11.1. Les llargues cues representen casos excepcionals i poc freqüents. La mitjana de la variable PIE es situa a 9.35 mentre que el màxim es situa a 23 i el mínim a 2.3.
- **LEBRON:** El boxplot que representa la variable LEBRON té una forma allargada que mostra una alta variabilitat. La mitjana està situada a -0.342099 i els seus primer i tercer quartils a -1.520 i 0.575 respectivament. Com passa amb les altres anàlisis, la variable LEBRON presenta també valors extrems per la part superior, amb un màxim de 7.8 i un mínim de -4.59.

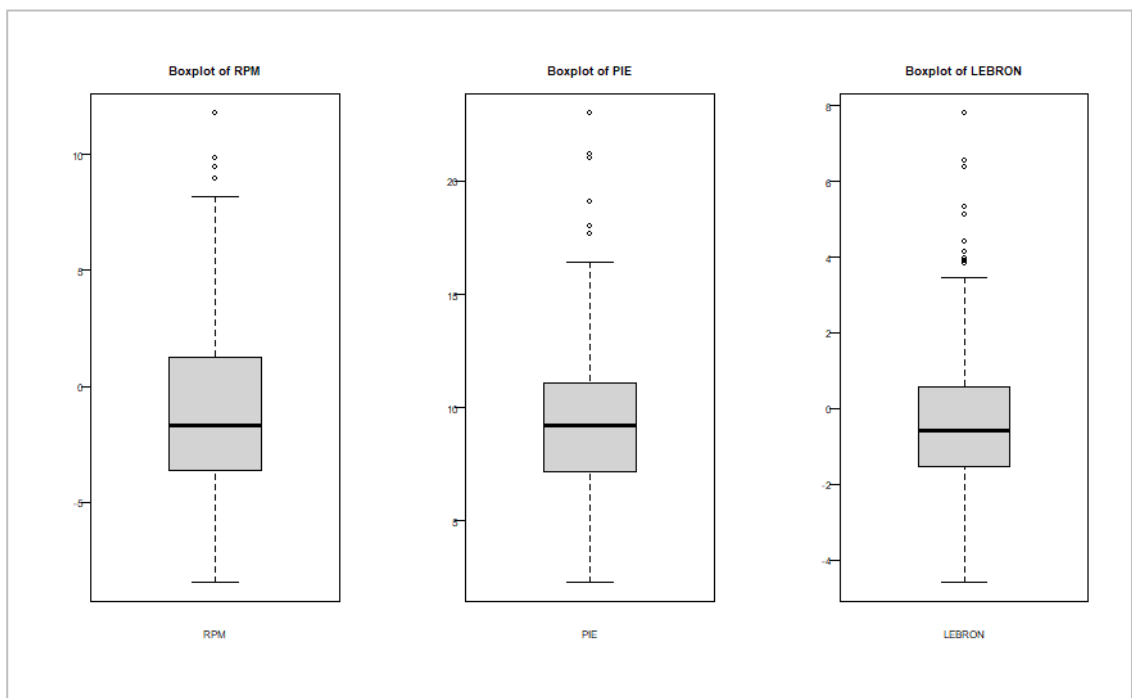


Figura 14. Boxplots dels índexs RPM, PIE i LEBRON, ordenats d'esquerra a dreta

Per altra banda, també s'efectua una matriu de correlacions per saber la correlació que tenen els índexs entre si. S'observa com clarament les correlacions són altres, encara que si tenim en compte el tipus de variables que s'estan analitzant, té sentit que sigui així. Com totes les variables intenten mesurar el mateix, seria motiu d'alerta pensar que no estan correlacionades. La correlació elevada és indicadora que totes assignen als jugadors valors similars o que els avaluen de manera semblant. Tot i així, hi ha un valor que sobresurt de la resta i és la correlació entre els índexs PER i PIE. Aquesta correlació que pren per valor 0.95 és massa elevada i ens indica que aquests dos índexs avaluen pràcticament el mateix.

Donada aquesta situació, s'analitzaran igualment els dos índexs per separat, cosa que no treu que realment els matisos entre els dos siguin ínfims i els resultats obtinguts seran pràcticament iguals.

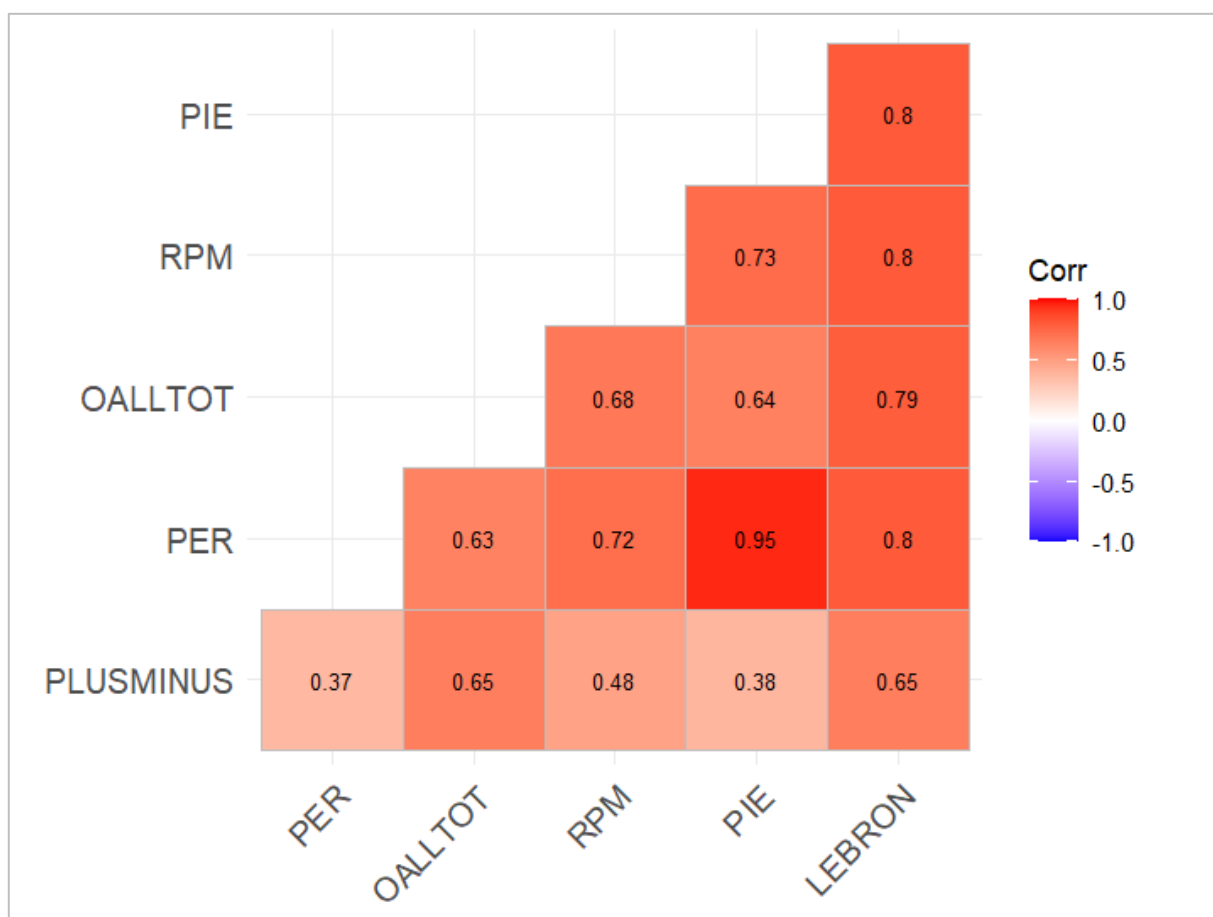


Figura 15. Correlacions entre tots els índexs.

## 6. RESULTATS DE LA MODELITZACIÓ DELS ÍNDEXS

### 6.1. PlusMinus

Un cop realitzat el model lineal mixt es poden treure una sèrie de conclusions. En primer lloc, es destaca clarament com l'efecte aleatori de l'equip és important, ja que es parteix de la situació inicial que hi ha una desviació típica de 1,978 en els valors de PlusMinus entre els equips. Si s'analitzen els efectes fixos, es veu com per l'índex PlusMinus l'efecte principal de posició del jugador no és significatiu, encara que la Posició d'Escorta (SG) sí que presenta un p-valor  $< 0.05$ . Això indica que, quan totes les altres variables es mantinguin en els seus valors de referència (Age\_Factor= 23 o menys, FormacioUSA= NO), el valor de l'índex PlusMinus s'espera que variï en un -0.690 entre els jugadors en la posició de SG i els jugadors en la posició de referència Pívol (C). També s'observa que l'efecte dels grups d'edat té és significatiu en el seu conjunt sobre la variable resposta. Concretament, el grup d'edat de jugadors de 24 a 31 anys, amb un p-valor de 0.0066 i el grup d'edat de jugadors de 32 o més anys. En ambdós casos es preveu un canvi positiu en el valor de l'índex PlusMinus quan totes les altres variables es mantinguin en els seus valors de referència. Així doncs per a jugadors Pívots(C) que no han tingut formació als Estats Units, hi haurà un variació positiva en el valor de PlusMinus quan també siguin jugadors de més de 24 anys. Més concretament, quan siguin jugadors de 24 a 31 anys la variació serà de 0.639 amb una error estàndard de 0.234 i quan siguin jugadors de 32 o més anys la variació serà de 0.730 amb un error estàndard de 0.335. Per últim, es pot veure també que l'efecte dels Minuts és significativa, mostrant un p-valor de  $1.33e-15$ . Això indica que la variació mitjana de la variable resposta "PlusMinus" per cada unitat de canvi en la variable Minut, sempre tenint en compte les altres variables del model, és de 0.0996.

```
## Random effects:
## Formula: ~1 | TEAM
## (Intercept) Residual
## StdDev: 1.978386 2.037253
##
## Fixed effects: PLUSMINUS ~ POS + FormacioUSA + Age_factors + MIN
## Value Std.Error DF t-value p-value
## (Intercept) -2.2517764 0.5655532 405 -3.981547 0.0001
## POSPF -0.2696617 0.3181218 405 -0.847668 0.3971
## POSPG -0.4607118 0.3263631 405 -1.411654 0.1588
## POSSF -0.4475397 0.3217247 405 -1.391064 0.1650
## POSSG -0.6906831 0.3111715 405 -2.219622 0.0270
## FormacioUSASI -0.1173363 0.3059544 405 -0.383509 0.7015
## Age_factors24 a 31 0.6390445 0.2342404 405 2.728157 0.0066
## Age_factors32 o més 0.7305806 0.3357127 405 2.176208 0.0301
## MIN 0.0996620 0.0119668 405 8.328208 0.0000
```

Figura 16. Informació sobre els efectes fixos i aleatoris del model lineal mixt

ANOVA	DF	F-Value	p-value
(Intercept)	405	15.85271	0.0001
POS	405	1.34335	0.2531
FormacioUSA	405	0.14708	0.7015
Age_factors	405	4.18359	0.0159
MIN	405	69.35905	<.0001

Taula 11. Resultats de l'anova del model per testar els efectes fixos

Amb l'ajuda de la funció *emmeans* es fa l'anàlisi post-hoc i s'aconsegueixen les comparacions entre els nivells de la variable significativa "Age\_factors" i també les mitjanes ajustades a la mitjana de la variable numèrica MIN. Per la variable que agrupa els grups d'edat, el mitjana estimat de PlusMinus per la categoria d'edat de 23 o menys és de -0.488, en el cas del segon grup d'edat, de 24 a 31 anys el mitjana estimat és de 0.151 i per l'últim grup "32 o més" la mitjana és de 0.243. També en aquesta variable es pot destacar el contrast significatiu entre els grups "23 o menys" i "24 a 31". Això proporciona informació sobre la diferència estimada, que diu concretament que s'estima una diferència de -0.639 en les mitjanes entre els nivells de "23 o menys" i "24 a 31". El valor negatiu d'aquesta estimació implica que s'espera que el grup de "24 a 31" tingui un valor més baix en PlusMinus en comparació al grup "23 o menys".

```

$emmeans
Age_factors emmean    SE df lower.CL upper.CL
23 o menys  -0.488 0.421 29  -1.348    0.373
24 a 31      0.151 0.403 29  -0.673    0.976
32 o més     0.243 0.460 29  -0.699    1.184

Results are averaged over the levels of: POS, FormacioUSA
Degrees-of-freedom method: containment
Confidence level used: 0.95

$constrasts
contrast          estimate    SE df t.ratio p.value
23 o menys - 24 a 31  -0.6390 0.234 405  -2.728  0.0182
23 o menys - 32 o més  -0.7306 0.336 405  -2.176  0.0765
24 a 31 - 32 o més    -0.0915 0.301 405  -0.304  0.9502

MIN emmean    SE df lower.CL upper.CL
22  -0.0312 0.394 29  -0.837    0.774

```

Figura 17. Mitjanes estimades de les variables Age\_Factors i MIN i contrastos entre les diferents categories de Age\_Factors

Els gràfics de validació mostren en un primer lloc un QQ-plot dels residus. El conjunt de punts es mostra aproximadament seguint la línia diagonal de referència i això indica que els residus

es comporten seguint una distribució normal. Per altra banda, el gràfic que ens compara els valors ajustats i els residus estandarditzats també indica homoscedasticitat i que no hi ha cap patró alarmant en el model. Es veu com el núvol de punts es situa al voltant del zero i estan distribuïts aleatòriament.

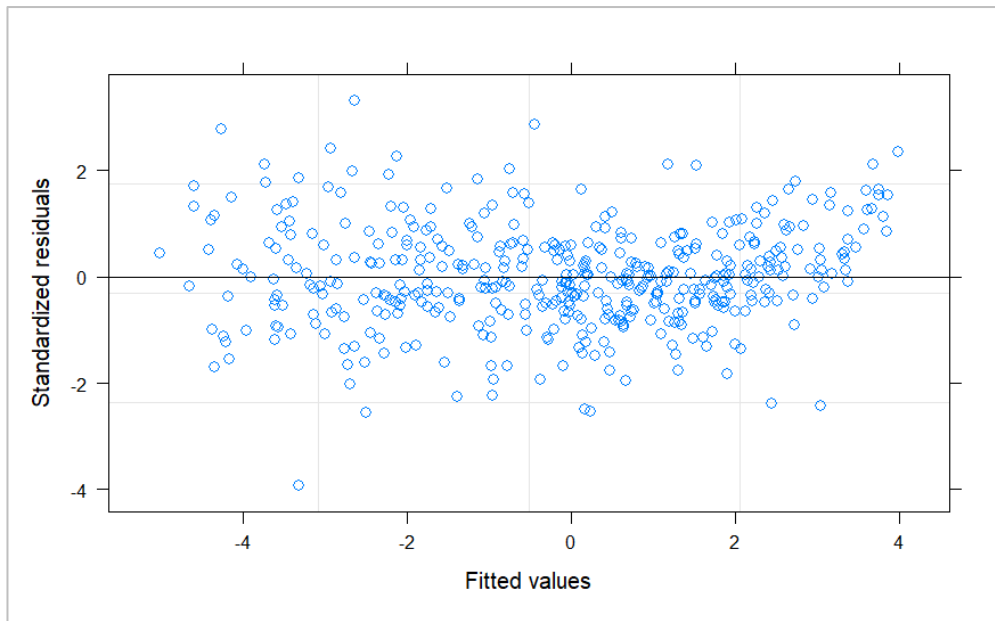


Figura 18. Gràfic dels valors ajustats vs residus estandarditzats

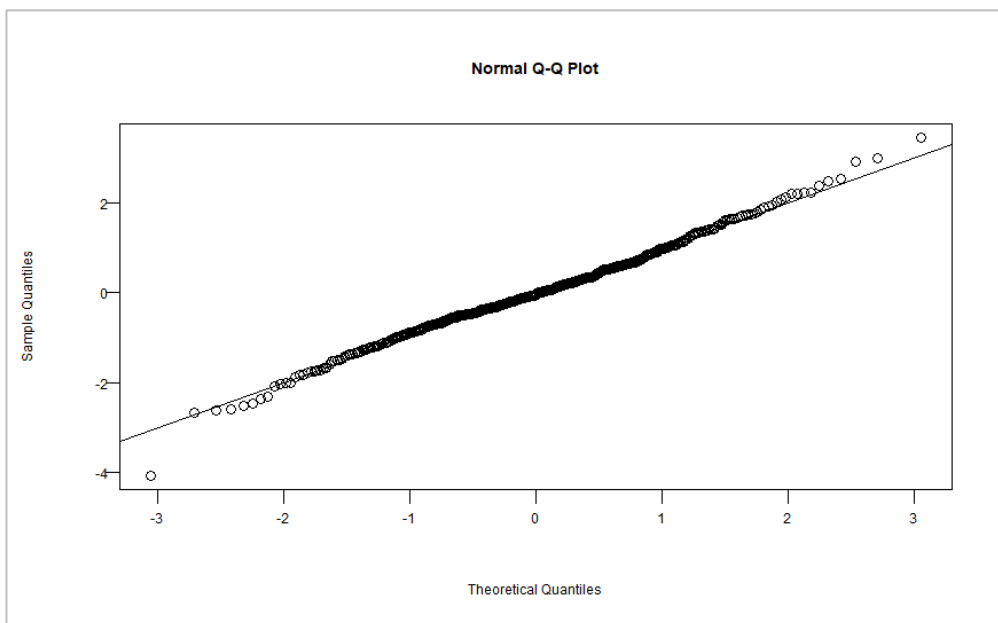


Figura 19. Gràfic QQ-Plot dels residus

## 6.2. PER

En el cas de l'índex PER els resultats ens ensenyen que l'efecte aleatori de l'equip és molt baix, d'un 0,000616. L'efecte principal estudiat que dona informació sobre la posició de jugador és molt significatiu en conjunt. Entrant més en detalls, totes les posicions donen un p-valor significatiu al model lineal. Ressalta el valor de l'estimació significativa de l'índex PER en els seus valors de referència, és a dir, quan el jugador és Pivot (C), de 23 o menys anys i sense formació als Estats Units, que pren un valor de 12.84 amb un error de 0.688. Respecte aquests valors referència, es pot extreure també la informació dels valors estimats en el cas que el jugador jugués a una altra posició. En el cas que jugués d'Ala-Pivot (PF), el valor del seu índex PER s'espera que descendeixi en 3.415 punts. En el cas que jugués de Base (PG) el seu valor s'espera que baixi 5.037 punts. En el cas de ser Aler(SF) el valor esperat baixaria en 5.6319 punts i per últim, si fos un Escorta (SG) el seu valor esperat seria de 6.236 punts menys. No s'observa significació dels efectes d'Edat o de formació als Estats Units. En el cas de l'efecte dels minuts, si que és significatiu. El valor estimat de variació en l'índex PER per cada unitat de canvi en la variable Minut és de 0.2693.

```
Random effects:
Formula: ~1 | TEAM
          (Intercept) Residual
StdDev:  0.0006160731  3.335221

Fixed effects:  PER ~ POS + FormacioUSA + Age_factors + MIN
               Value Std.Error DF   t-value p-value
(Intercept)    12.841262  0.6883345  405   18.655555  0.0000
POSPF          -3.415108  0.5115248  405   -6.676328  0.0000
POSPG          -5.037458  0.5257212  405   -9.581996  0.0000
POSSF          -5.631985  0.5182805  405  -10.866672  0.0000
POSSG          -6.236514  0.4991133  405  -12.495187  0.0000
FormacioUSASI  -0.510332  0.4735276  405   -1.077723  0.2818
Age_factors24 a 31  0.085664  0.3644005  405    0.235082  0.8143
Age_factors32 o més -0.333537  0.5063199  405   -0.658748  0.5104
MIN             0.269386  0.0193623  405   13.912875  0.0000
```

Figura 20. Informació sobre els efectes fixos i aleatoris del model lineal mixt

ANOVA	DF	F-Value	p-value
(Intercept)	405	348.0297	<.0001
POS	405	47.4080	<.0001
FormacioUSA	405	1.1615	0.2818
Age_factors	405	0.4086	0.6648
MIN	405	193.5681	<.0001

*Taula 12. Resultats de l'anova del model per testar els efectes fixos*

El mitjana estimat de l'índex PER per la posició de Pívol (C) és el més alt amb un valor de 18.4, seguit de la mitjana estimat dels Aler-Pívols(PF) amb un 15. En tercera posició es troben els jugadors que juguen a la posició de Bases (PG) amb una mitjana estimada de 13.4. En penúltim lloc, es troba la posició de d'Aler (SF) amb un valor de 12.8, seguit molt de prop per la mitjana estimada dels jugadors que juguen d'Escortes(SG), que és de 12.2. En les comparacions entre els nivells de la variable significativa Posició, és important destacar els contrastos molt significatius, tots amb un p-valor associat <0.001 entre totes les posicions amb la posició de Pívol(C).

El contrast entre la posició de Pívol(C) i Ala-Pívol(PF) estima una diferència entre les mitjanes d'aquestes dues posicions de 3.415 punts. D'aquesta mateixa manera s'analitzaria el contrast entre les posicions Pívol( C) i Base(PG), que és de 5.037, el contrast entre Pívol( C) i Aler(SF) que és de 5.632 i el contrast entre Pívol( C) i Escorta(SG) que és de 6.237. Seguint amb les comparacions entre els nivells de la posició dels jugadors, el contrast entre els Aler-Pívols(PF) i els Bases(PG) és també significativa, amb un valor de 1.622, que informa sobre la diferència estimada de mitjanes entre aquestes dues posicions. Seguint amb la posició d'Ala-pívol (PF), existeixen també més diferències significatives en els contrastos amb les posicions d'Aler (SF) i d'Escorta(SG). Els valors dels contrastos són de 2.217 i 2.821 respectivament.

\$emmeans					
POS	emmean	SE	df	lower.CL	upper.CL
C	18.4	0.400	29	17.6	19.3
PF	15.0	0.403	29	14.2	15.8
PG	13.4	0.412	29	12.6	14.2
SF	12.8	0.417	29	12.0	13.7
SG	12.2	0.394	29	11.4	13.0

Results are averaged over the levels of: FormacioUSA, Age\_factors  
Degrees-of-freedom method: containment  
Confidence level used: 0.95

\$contrasts					
contrast	estimate	SE	df	t.ratio	p.value
C - PF	3.415	0.512	405	6.676	<.0001
C - PG	5.037	0.526	405	9.582	<.0001
C - SF	5.632	0.518	405	10.867	<.0001
C - SG	6.237	0.499	405	12.495	<.0001
PF - PG	1.622	0.514	405	3.159	0.0146
PF - SF	2.217	0.508	405	4.362	0.0002
PF - SG	2.821	0.487	405	5.788	<.0001
PG - SF	0.595	0.519	405	1.146	0.7821
PG - SG	1.199	0.498	405	2.407	0.1154
SF - SG	0.605	0.489	405	1.235	0.7308

MIN					
	emmean	SE	df	lower.CL	upper.CL
22	14.4	0.248	29	13.9	14.9

Figura 21. Mitjanes estimades de les variables POS i MIN i contrastos entre les diferents categories de POS.



En els gràfics de validació, s'observa en el QQ-plot dels residus si segueixen una distribució normal. En aquest cas, els punts segueixen en termes generals la línia diagonal, indicant que els residus tenen una distribució normal. És cert per això que les dues cues s'allunyen una mica de la línia, però no és una preocupant, ja que és només en aquesta zona. A més, el diagrama que contrasta els valors ajustats amb els residus estandarditzats ens dona indicis per validar la presència d'homoscedasticitat, donat que el conjunt de punts es concentra a prop del zero i s'estén de forma aleatòria.

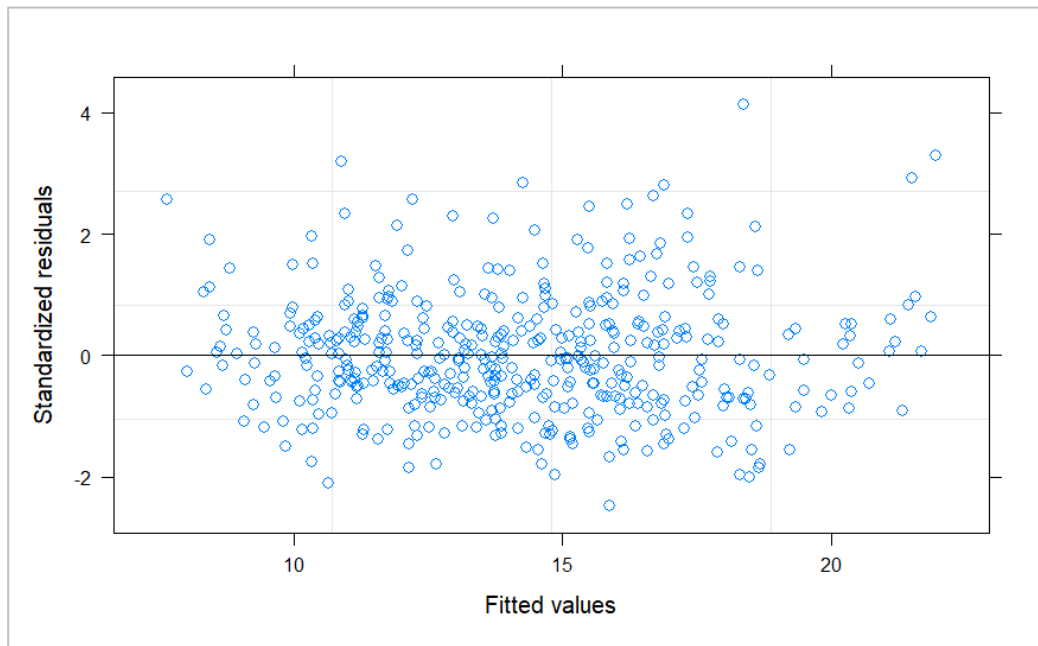


Figura 22. Gràfic dels valors ajustats vs residus estandarditzats

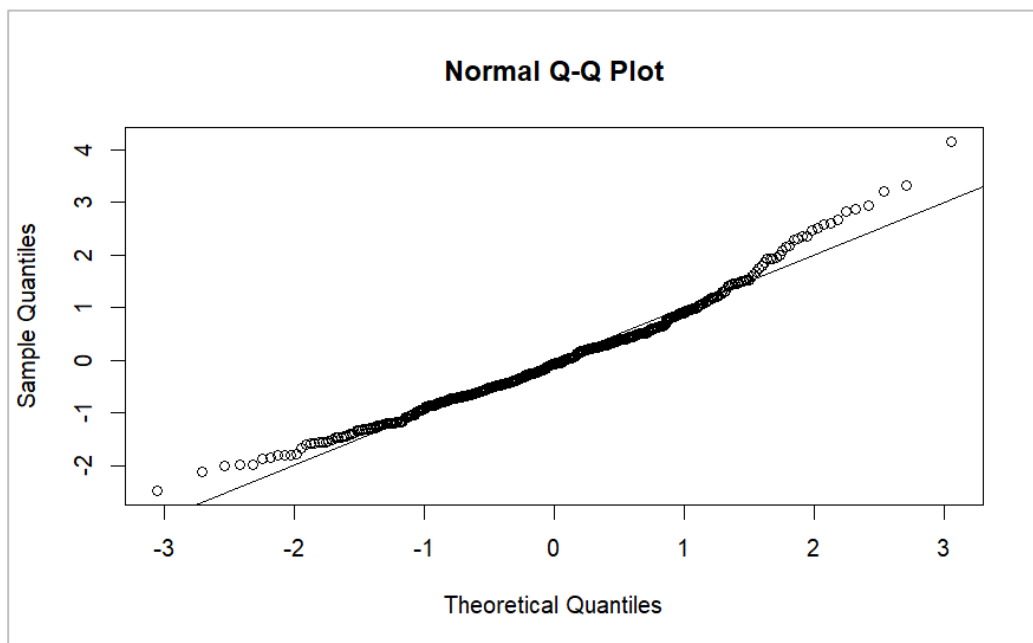


Figura 23. Gràfic QQ-Plot dels residus

### 6.3. RAPTOR

L'efecte de l'equip en el cas de l'índex RAPTOR sembla prou important, ja que té un valor del 0.899. L'anàlisi de la variància marginal per l'efecte general de la variable de Posició en el model té un efecte significatiu, amb un p-valor associat de 0.027. Els resultats del model lineal mixt donen informació sobre que el valor de l'estimació de l'índex RAPTOR quan un jugador està en els seus valors de referència és significatiu. Així doncs, un jugador Pivot(C) de 23 anys o menys i sense formació en Estats Units s'estima que tingui un valor de -4.956 d'índex RAPTOR. Aquest valor empitjora de manera significativa si, en aquestes mateixes condicions, en comptes de jugar de Pivot(C) juga d'Ala-Pivot(PF), que llavors cau -0.9783 punts. Si es segueix variant la posició del jugador i en aquest cas és Aler(SF) o d'Escorta(SG), els valors estimats decreixen de manera significativa 0.856 i 1.139 punts respectivament. És important destacar també l'efecte significatiu de l'Edat, que és significatiu en el seu conjunt. El model ens ensenya que, sota els valors de referència en les altres variables, s'estima que el valor de l'índex RAPTOR millori en 1.049 punts si el jugador té de 24 a 31 anys. De la mateixa manera, també puja en 0.847 punts si el jugador en comptes de tenir 23 anys o menys, té 32 anys o més. Per últim, també es veu com l'efecte dels Minuts és significatiu per explicar l'índex RAPTOR, amb un p-valor proper a zero.

```
Random effects:
  Formula: ~1 | TEAM
          (Intercept) Residual
StdDev:   0.8990733 2.552872

Fixed effects:  OALLTOT ~ POS + FormacioUSA + Age_factors + MIN
               Value Std.Error  DF  t-value p-value
(Intercept)   -4.956860 0.5636325 405 -8.794490 0.0000
POSPF         -0.978331 0.3963578 405 -2.468304 0.0140
POSPG         -0.433571 0.4068450 405 -1.065690 0.2872
POSSF         -0.856986 0.4011187 405 -2.136491 0.0332
POSSG        -1.139559 0.3873413 405 -2.942003 0.0034
FormacioUSASI -0.196794 0.3764783 405 -0.522722 0.6015
Age_factors24 a 31  1.049883 0.2887172 405  3.636373 0.0003
Age_factors32 o més 0.847237 0.4095733 405  2.068584 0.0392
MIN           0.200248 0.0149403 405 13.403210 0.0000
```

Figura 24. Informació sobre els efectes fixos i aleatoris del model lineal mixt

ANOVA	DF	F-Value	p-value
<b>(Intercept)</b>	405	77.34305	<.0001
<b>POS</b>	405	2.75309	0.0278
<b>FormacioUSA</b>	405	0.27324	0.6015
<b>Age_factors</b>	405	6.67341	0.0014
<b>MIN</b>	405	179.64603	<.0001

*Taula 13. Resultats de l'anova del model per testar els efectes fixos*

En l'anàlisi Post-hoc mostra com la mitjana estimada per l'índex RAPTOR per totes les posicions és negatiu. En el cas més proper a zero hi ha la posició de Pivot(C) amb -0.0097 i una desviació estàndard de 0.350, seguit dels jugadors Bases(PG) amb -0.4433 i una desviació estàndard de 0.360. En el cas de les altres tres posicions, la mitjana estimada dels Ala-Pivots(PF) és -0.988 amb una desviació de 0.354, en el cas dels Alers(SF) és de -0.8667 amb una desviació de 0.365 i com a darrer de la llista hi ha els jugadors Escortes(SG) amb una mitjana estimada de -1.149 i una desviació de 0.352.

Pel que fa a les comparacions entre les diferents posicions, és només significatiu el contrast entre la variable Pivot(C) i la variable Escorta(SG) amb una diferència estimada entre les mitjanes d'aquestes dues posicions de 1.140 punts.

Pel que fa referència als altres efectes significatius, s'obté que el contrast entre les categories 23 o menys i 24 a 31 anys és significatiu. Així doncs, la diferència estimada entre les mitjanes de les dues categories és de 1.050, és a dir, s'estima que el grup de "24 a 31 anys" sigui 1.05 punts més alt que el de "23 o menys".

```

$emmeans
POS  emmean    SE df lower.CL upper.CL
C   -0.0097  0.350 29   -0.725    0.706
PF  -0.9880  0.354 29   -1.711   -0.265
PG  -0.4433  0.360 29   -1.180    0.294
SF  -0.8667  0.365 29   -1.613   -0.120
SG  -1.1493  0.352 29   -1.869   -0.430

```

Results are averaged over the levels of: FormacioUSA, Age\_factors  
Degrees-of-freedom method: containment  
Confidence level used: 0.95

```

$contrasts
contrast estimate    SE  df t.ratio p.value
C - PF      0.978 0.396 405   2.468 0.1000
C - PG      0.434 0.407 405   1.066 0.8240
C - SF      0.857 0.401 405   2.136 0.2067
C - SG      1.140 0.387 405   2.942 0.0283
PF - PG     -0.545 0.398 405  -1.370 0.6475
PF - SF     -0.121 0.396 405  -0.306 0.9981
PF - SG      0.161 0.379 405   0.426 0.9931
PG - SF      0.423 0.407 405   1.041 0.8360
PG - SG      0.706 0.388 405   1.820 0.3634
SF - SG      0.283 0.380 405   0.743 0.9462

```

```

$emmeans
Age_factors emmean    SE df lower.CL upper.CL
23 o menys -1.324 0.312 29   -1.963   -0.685
24 a 31     -0.274 0.276 29   -0.838    0.290
32 o més    -0.477 0.387 29   -1.268    0.315

```

Results are averaged over the levels of: POS, FormacioUSA  
Degrees-of-freedom method: containment  
Confidence level used: 0.95

```

$contrasts
contrast                estimate    SE  df t.ratio p.value
23 o menys - 24 a 31    -1.050 0.289 405  -3.636 0.0009
23 o menys - 32 o més  -0.847 0.410 405  -2.069 0.0977
24 a 31 - 32 o més      0.203 0.370 405   0.548 0.8474

```

```

MIN emmean    SE df lower.CL upper.CL
22 -0.691 0.254 29   -1.21   -0.171

```

Figura 25. Mitjanes estimades i contrastos de les variables POS, Age\_Factors i MIN

El gràfic QQPlot dels residus s'ajusta molt a la línia diagonal. Només és preocupant el valor extrem que es situa molt separat de la línia i allunyat de la resta de punts. Seria convenient repetir el model obviant aquest valor extrem. El gràfic que compara els valors ajustats amb els residus estandarditzats presenta una forma normal, encara que es segueix apreciand el mateix valor extrem que s'ha detectat al QQPlot.

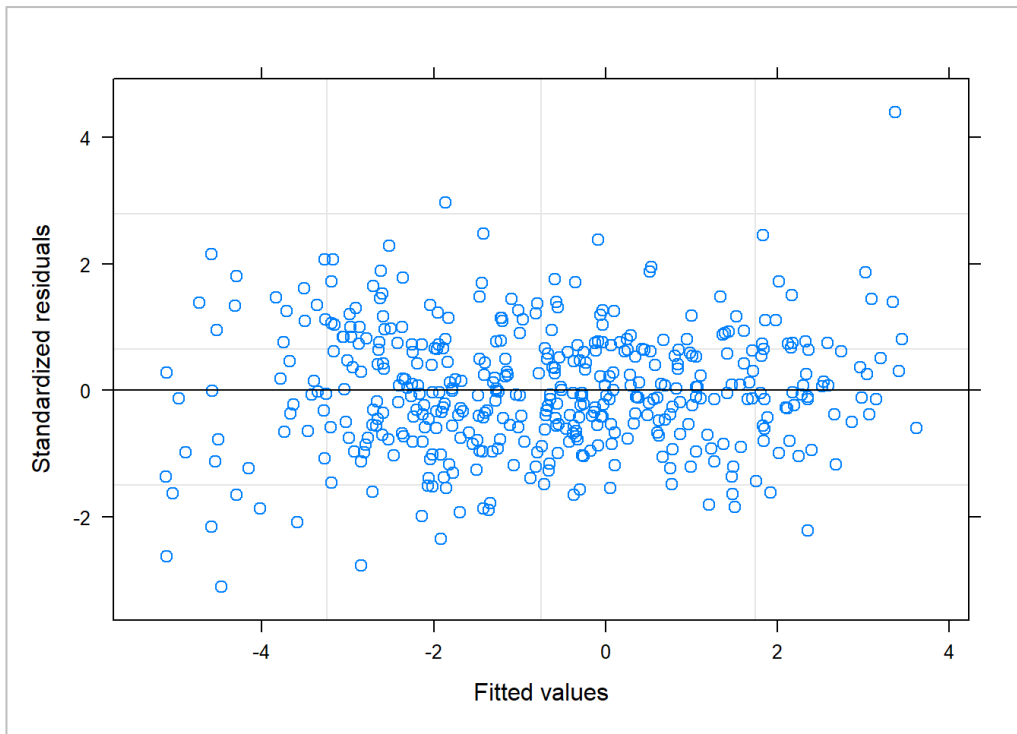


Figura 26. Gràfic dels valors ajustats vs residus estandarditzats

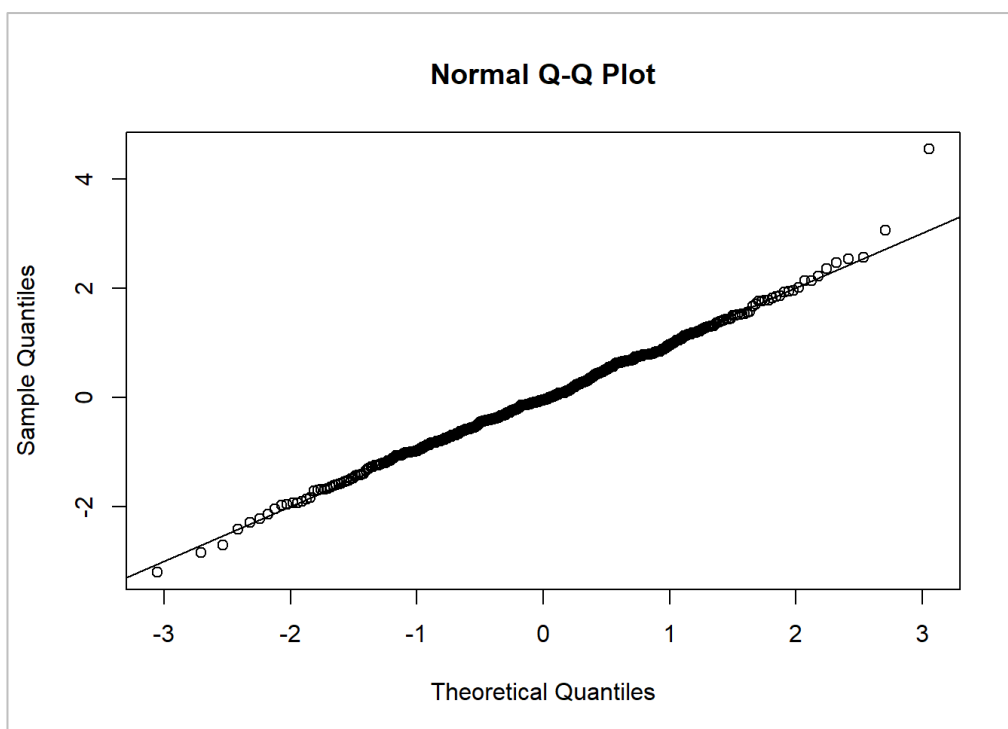


Figura 27. Gràfic QQ-Plot dels residus

Repetició del model obviant el valor extrem detectat anteriorment mitjançant els gràfics de validació. S'observa que quan es repeteix el model sense aquesta observació sorgeixen variacions. Destaca el canvi que fa la variable explicativa Posició, que perd significació globalment com a efecte, ja que ara mostra un p-valor de 0.0574, per la qual cosa no hi ha evidències per tal de rebutjar la hipòtesi nul·la. De la mateixa manera s'observa com ho fa la posició d'Aler(SF), que passa a tenir un p-valor associat de 0,0598 en comptes de 0.0332 . Es veu, per altra banda, que en l'anàlisi post-hoc els mateixos contrastos segueixen sent significatius, destacant el contrast entre les categories d'edat "23 o menys" i "24 a 31".

```

Random effects:
  Formula: ~1 | TEAM
            (Intercept) Residual
StdDev:    0.8855703  2.49231

Fixed effects:  OALLTOT ~ POS + FormacioUSA + Age_factors + MIN
                Value Std.Error  DF   t-value p-value
(Intercept)    -5.142566 0.5521966 404  -9.312925  0.0000
POSPF          -0.842358 0.3880881 404  -2.170532  0.0305
POSPG          -0.287114 0.3984525 404  -0.720574  0.4716
POSSF          -0.740606 0.3924147 404  -1.887305  0.0598
POSSG          -0.991388 0.3795077 404  -2.612299  0.0093
FormacioUSASI   0.011953 0.3703453 404   0.032274  0.9743
Age_factors24 a 31  1.017330 0.2820130 404   3.607388  0.0003
Age_factors32 o més 0.881624 0.4000758 404   2.203644  0.0281
MIN             0.194684 0.0146353 404  13.302380  0.0000
  
```

Figura 28. Informació sobre els efectes fixos i aleatoris del model lineal mixt

ANOVA	DF	F-Value	p-value
(Intercept)	404	86.73058	<.0001
POS	404	2.30893	0.0574
FormacioUSA	404	0.00104	0.9743
Age_factors	404	6.64404	0.0014
MIN	404	176.95331	<.0001

Taula 14. Resultats de l'anova del model per testar els efectes fixos

```

$emmeans
Age_factors emmean      SE df lower.CL upper.CL
23 o menys  -1.423 0.306 29   -2.050   -0.797
24 a 31     -0.406 0.272 29   -0.962    0.149
32 o més    -0.542 0.379 29   -1.316    0.233

Results are averaged over the levels of: POS, FormacioUSA
Degrees-of-freedom method: containment
Confidence level used: 0.95

$constrasts
contrast          estimate      SE  df t.ratio p.value
23 o menys - 24 a 31   -1.017 0.282 404  -3.607  0.0010
23 o menys - 32 o més  -0.882 0.400 404  -2.204  0.0718
24 a 31 - 32 o més     0.136 0.361 404   0.376  0.9252

```

```

MIN emmean  SE df lower.CL upper.CL
22  -0.79 0.25 29   -1.3   -0.279

```

```

Results are averaged over the levels of: POS, FormacioUSA, Age_factors
Degrees-of-freedom method: containment
Confidence level used: 0.95

```

*Figura 29. Mitjanes estimades de les variables Age\_Factors i MIN i contrastos entre les diferents categories de Age\_Factors*

Es tornen a repetir també els gràfics de validació per tal de comprovar que s'ha eliminat el valor extrem. Es pot observar tant en el gràfic QQ-Plot dels residuals com en el gràfic que compara els valors ajustats amb els residus estandarditzats com no existeix l'observació que es volia descartar.

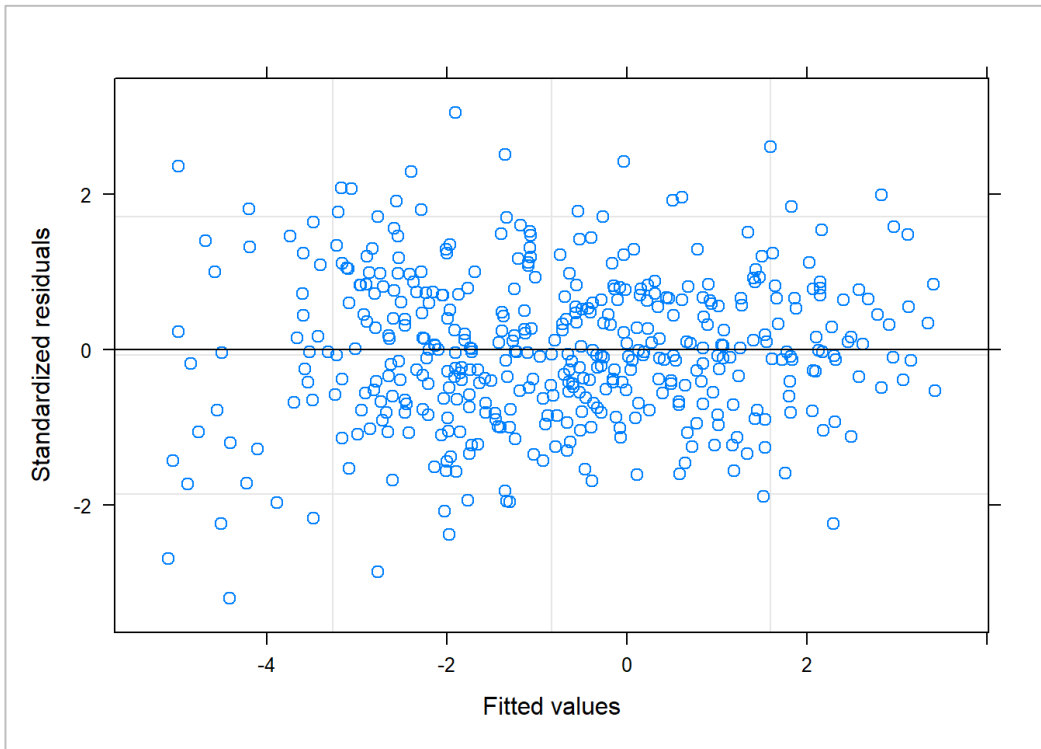


Figura 30. Gràfic dels valors ajustats vs residus estandarditzats

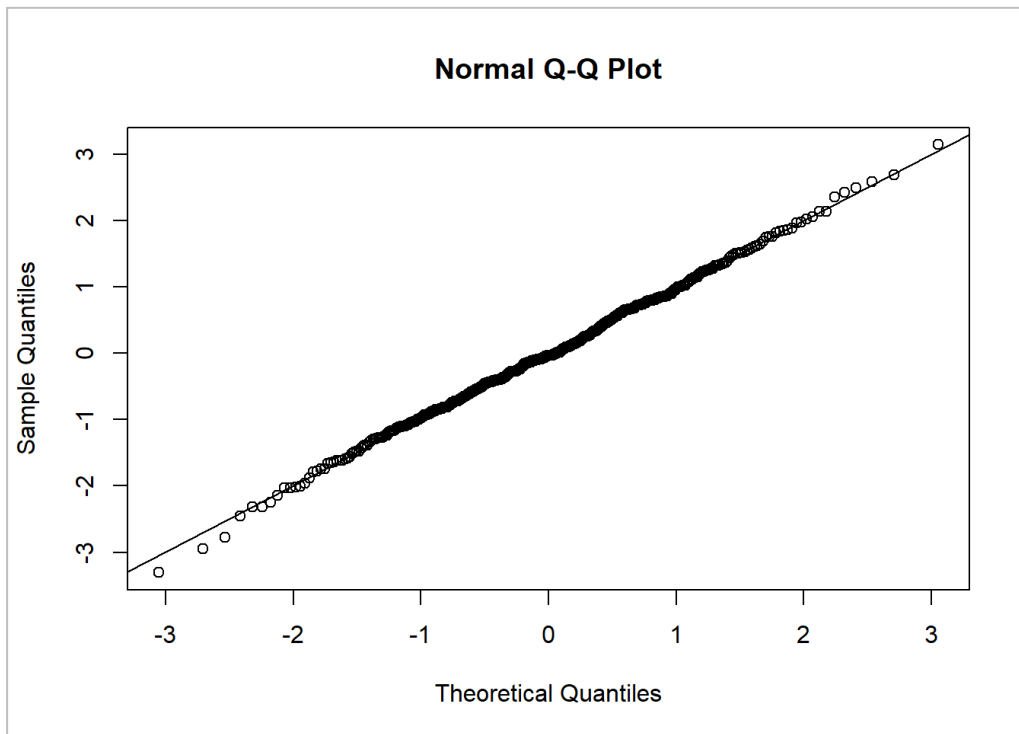


Figura 31. Gràfic QQ-Plot dels residus



#### 6.4. RPM

Al fer l'anàlisi sobre l'índex RPM es pot veure com l'efecte principal de la Posició és clarament significatiu en el seu conjunt. El model lineal mixt mostra com l'efecte posició és significatiu en totes les seves categories, començant per la de Pivot (C) que està definida com a valor de referència en aquesta variable. En aquest primer cas, sota la premissa que el jugador també sigui de 23 anys o menys i no hagi estat format en els Estats Units, el valor esperat serà de -6.642. Aquest valor s'espera que decreixi quan el jugador jugui en una altra posició. En el cas d'Ala-Pivots (PF) hi haurà una variació de 1.173 punts, en el cas dels Bases(PG) de 1.917 punts, en el cas dels Alers(SF) de 1.853 punts i en el cas dels Escortes(SG) de 2.299 punts. Totes aquestes variacions estimades són significatives i presenten uns errors estàndards petits similars. A banda de l'efecte principal de posició, també resulta significatiu l'efecte dels Minuts jugats, que indica una variació mitjana de 0.323 en el valor de l'índex RPM per cada variació unitària en la variable Minut.

```

Random effects:
  Formula: ~1 | TEAM
           (Intercept) Residual
StdDev:    0.6877922 1.924809

Fixed effects:  RPM ~ POS + FormacioUSA + Age_factors + MIN
               Value Std.Error DF   t-value p-value
(Intercept)   -6.642772 0.4255948 405  -15.608207 0.0000
POSPF         -1.173654 0.2988838 405   -3.926791 0.0001
POSPG         -1.917998 0.3067881 405   -6.251866 0.0000
POSSF         -1.853939 0.3024693 405   -6.129344 0.0000
POSSG         -2.299371 0.2920903 405   -7.872123 0.0000
FormacioUSASI -0.156367 0.2839724 405   -0.550640 0.5822
Age_factors24 a 31 0.185118 0.2177672 405    0.850073 0.3958
Age_factors32 o més -0.007239 0.3089943 405   -0.023427 0.9813
MIN           0.323946 0.0112656 405   28.755302 0.0000

```

Figura 32. Informació sobre els efectes fixos i aleatoris del model lineal mixt

ANOVA	DF	F-Value	p-value
(Intercept)	405	243.6161254	0.000000e+00
POS	405	18.3427398	7.305268e-14
FormacioUSA	405	0.3032049	5.821838e-01
Age_factors	405	0.4850513	6.160228e-01
MIN	405	826.8673907	0.000000e+00

Taula 15. Resultats de l'anova del model per testar els efectes fixos

Es veu clarament com les comparacions entre els nivells de la variable significativa Posició tots els contrastos amb la posició Pivot( C) són significatius. El contrast entre la posició de Pivot( C), amb una mitjana estimada de 0.478, i Ala-Pivot(PF), amb una mitjana estimada de -0.696, és significatiu. En el cas de la posició de Base(PG), amb mitjana estimada de -1.440, també s'indica que l'estimació de la variació de 1.918 entre les dues posicions és significativa. En el cas de la variació estimada entre les variables Pivot( C) i les d'Aler(SF) i Escorta(SG) són de 1.853 i de 2.299 respectivament, ambdues parelles significatives amb un p-valor associat <0.001. Per últim, l'únic contrast que destaca com a significatiu és l'establert entre les posicions d'Ala-Pivot(PF) i Escorta(SG), amb una variació estimada entre les seves mitjanes de 1.125 punts.

```

Semmeans
 POS emmean      SE df lower.CL upper.CL
 C      0.478 0.265 29  -0.0635  1.019
 PF     -0.696 0.267 29  -1.2431  -0.149
 PG     -1.440 0.273 29  -1.9977  -0.883
 SF     -1.376 0.276 29  -1.9411  -0.811
 SG     -1.822 0.266 29  -2.3661  -1.277

Results are averaged over the levels of: FormacioUSA, Age_factors
Degrees-of-freedom method: containment
Confidence level used: 0.95

$contrasts
 contrast estimate      SE df t.ratio p.value
 C - PF      1.1737 0.299 405   3.927 0.0010
 C - PG      1.9180 0.307 405   6.252 <.0001
 C - SF      1.8539 0.302 405   6.129 <.0001
 C - SG      2.2994 0.292 405   7.872 <.0001
 PF - PG      0.7443 0.300 405   2.482 0.0968
 PF - SF      0.6803 0.299 405   2.277 0.1545
 PF - SG      1.1257 0.285 405   3.944 0.0009
 PG - SF     -0.0641 0.307 405  -0.209 0.9996
 PG - SG      0.3814 0.293 405   1.304 0.6891
 SF - SG      0.4454 0.287 405   1.553 0.5284

```

```

MIN emmean      SE df lower.CL upper.CL
 22 -0.971 0.193 29  -1.37  -0.576

```

Figura 33. Mitjanes estimades de les variables POS i MIN i contrastos entre les diferents categories de POS.

Els gràfics de validació revelen diverses característiques rellevants. En primer lloc, el QQ-Plot dels residus coincideix amb la línia diagonal, indicant així que els residus segueixen una distribució normal. Es veu com els punts de les cues s'allunyen una mica de la línia, encara que aquesta desviació es considera normal i s'atribueix a alguns valors extrems. Aquest mateix fet també es detecta al gràfic de residus estandarditzats respecte valors ajustats, que tot i tenir una forma on tots els punts es distribueixen aleatòriament, hi ha algun valor extrem.

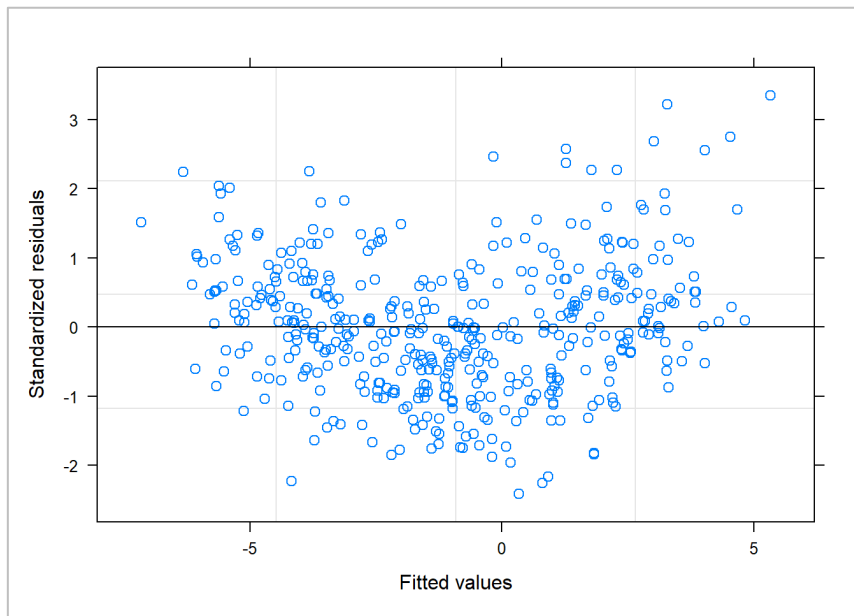


Figura 34. Gràfic dels valors ajustats vs residus estandarditzats

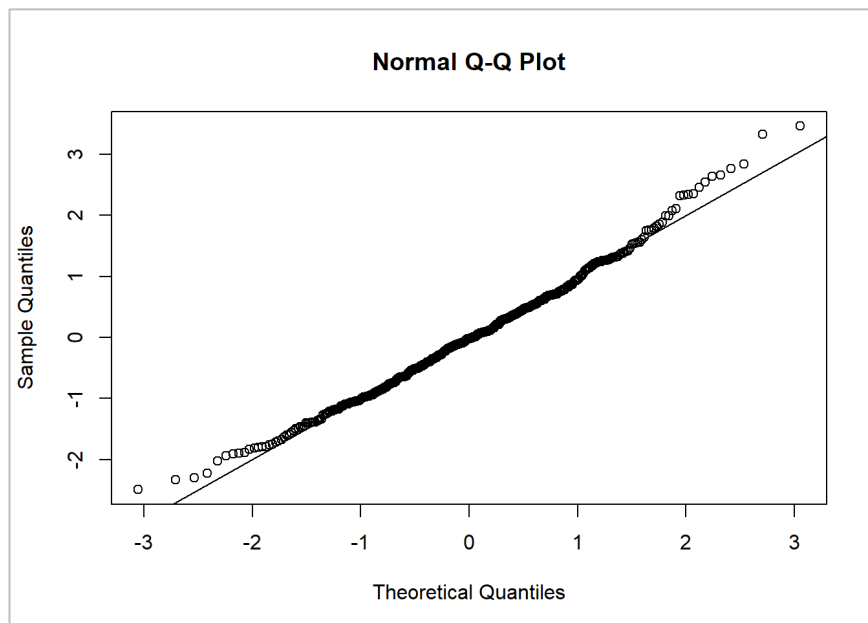


Figura 35. Gràfic QQ-Plot dels residus

## 6.5. PIE

A l'anàlisi de l'índex PIE es pot destacar la significació de l'efecte principal de Posició del jugador, així com també la dels minuts jugats. Cal destacar també que l'anàlisi de variància marginal per l'efecte de la formació als Estats Units en el seu conjunt té un p-valor associat de 0.0508, per la qual cosa no és significatiu però per molt poc. Posant atenció en l'efecte principal a estudiar, el model lineal mixt manifesta que l'estimació de l'índex PER en els seus valors de referència és significativa. Mostra que quan el jugador és Pivot(C), de 23 anys o menys i sense formació als Estats Units, pren un valor de 7.738. Tenint en compte aquest mateix perfil de jugador, aquest valor disminueix de manera significativa si varia només la posició. En el cas que el jugador fos Ala-Pivot(PF) s'estima una disminució de 2.127 punts, en el cas d'un jugador Base(PG) 2.627, en el cas de ser Aler(SF) 3.421 i per últim, la baixada més significativa és si el jugador es Escorta(SG) que és de 3.775 punts.

```

Random effects:
  Formula: ~1 | TEAM
            (Intercept) Residual
StdDev:    0.0001057749 2.292643

Fixed effects:  PIE ~ POS + FormacioUSA + Age_factors + MIN
               Value Std.Error DF   t-value p-value
(Intercept)    7.738260 0.4731637 405   16.354299 0.0000
POSPF          -2.127710 0.3516241 405   -6.051093 0.0000
POSPG          -2.627027 0.3613827 405   -7.269378 0.0000
POSSF          -3.421030 0.3562680 405   -9.602408 0.0000
POSSG          -3.775782 0.3430923 405  -11.005149 0.0000
FormacioUSASI  -0.637432 0.3255046 405   -1.958287 0.0509
Age_factors24 a 31  0.236969 0.2504902 405    0.946022 0.3447
Age_factors32 o més 0.304560 0.3480461 405    0.875058 0.3821
MIN             0.201703 0.0133097 405   15.154509 0.0000
    
```

Figura 36. Informació sobre els efectes fixos i aleatoris del model lineal mixt

ANOVA	DF	F-Value	p-value
(Intercept)	405	267.4630971	0.00000000
POS	405	35.5702531	0.00000000
FormacioUSA	405	3.8348882	0.05088239
Age_factors	405	0.5656978	0.56841180
MIN	405	229.6591422	0.00000000

Taula 16. Resultats de l'anova del model per testar els efectes fixos

L'anàlisi exploratòria posterior dona informació sobre les mitjanes estimades de l'índex PIE per totes les posicions. En el cas dels Pivots (C) la mitjana estimada és de 12.05, pels Ala-Pivots(PF) és de 9.92, pels Bases(PG) és de 9.42, pels Alers(SF) és de 8.62 i pels Escortes(SG) és de 8.27. Totes les desviacions estàndard associades a cada estimació són petites i molt similars. Les comparacions entre posicions, tots els contrastos entre la categoria Pivot(C) són significatius amb p-valors associats < 0.0001. En el cas de les altres posicions, ressalten les diferències estimades entre mitjanes entre la posició d'Ala-Pivot(PF) i les categories Aler(SF) i Escorta(SG). Resultant sempre favorables a la posició d'Ala-Pivot(PF), de la qual s'estima una mitjana d'1.293 punts més alta que a la dels Alers(SF) i 1.648 punts més alta que la dels Escortes(SG). Per últim, també resulta significativa la diferència entre mitjanes estimades entre les categories Base(PG) i Escorta(SG), amb una variació estimada de 1.149 punts favorable a la primera categoria.

```

Semmeans
  POS emmean    SE df lower.CL upper.CL
  C      12.05 0.275 29    11.48    12.61
  PF      9.92 0.277 29     9.35    10.48
  PG      9.42 0.283 29     8.84    10.00
  SF      8.62 0.287 29     8.04     9.21
  SG      8.27 0.271 29     7.71     8.82

Results are averaged over the levels of: FormacioUSA, Age_factors
Degrees-of-freedom method: containment
Confidence level used: 0.95

Scontrasts
  contrast estimate    SE df t.ratio p.value
  C - PF      2.128 0.352 405    6.051 <.0001
  C - PG      2.627 0.361 405    7.269 <.0001
  C - SF      3.421 0.356 405    9.602 <.0001
  C - SG      3.776 0.343 405   11.005 <.0001
  PF - PG      0.499 0.353 405    1.414 0.6187
  PF - SF      1.293 0.349 405    3.702 0.0023
  PF - SG      1.648 0.335 405    4.918 <.0001
  PG - SF      0.794 0.357 405    2.226 0.1723
  PG - SG      1.149 0.342 405    3.355 0.0077
  SF - SG      0.355 0.336 405    1.054 0.8296

MIN emmean    SE df lower.CL upper.CL
  22   9.65 0.17 29     9.31     10

```

Figura 37. Mitjanes estimades de les variables POS i MIN i contrastos entre les diferents categories de POS

Els gràfics de validació resulten bons, es veu gràcies a el gràfic QQ-plot dels residus com aquests segueixen una distribució normal. Tampoc s'observa cap patró en el gràfic de residus estandarditzats i valors ajustats, indicador de homoscedasticitat del model.

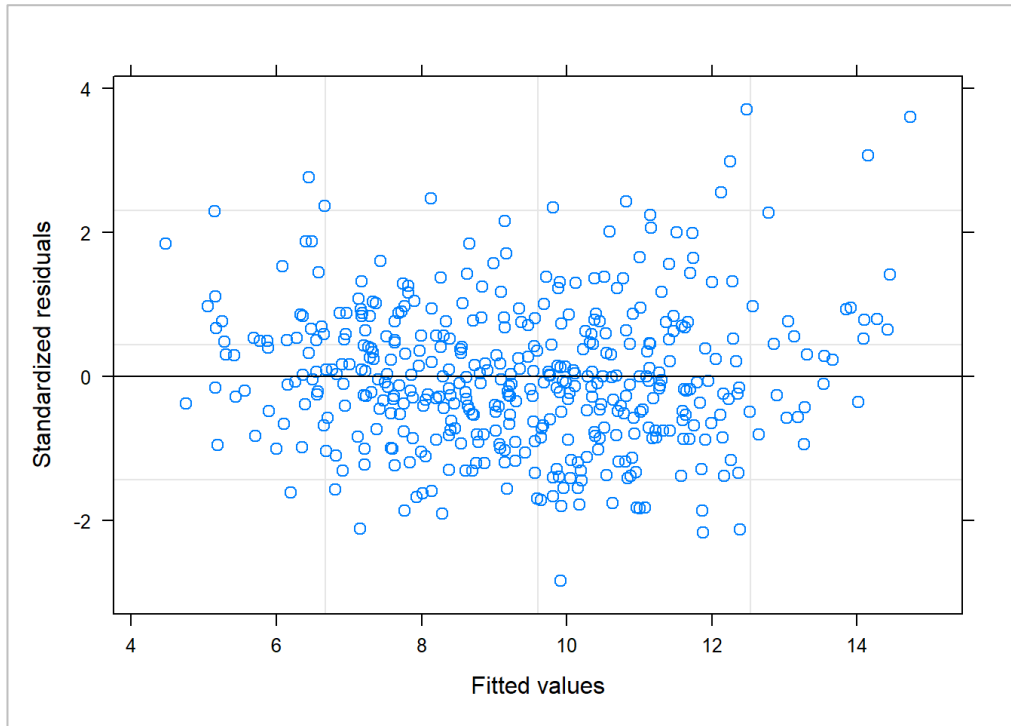


Figura 38. Gràfic dels valors ajustats vs residus estandarditzats

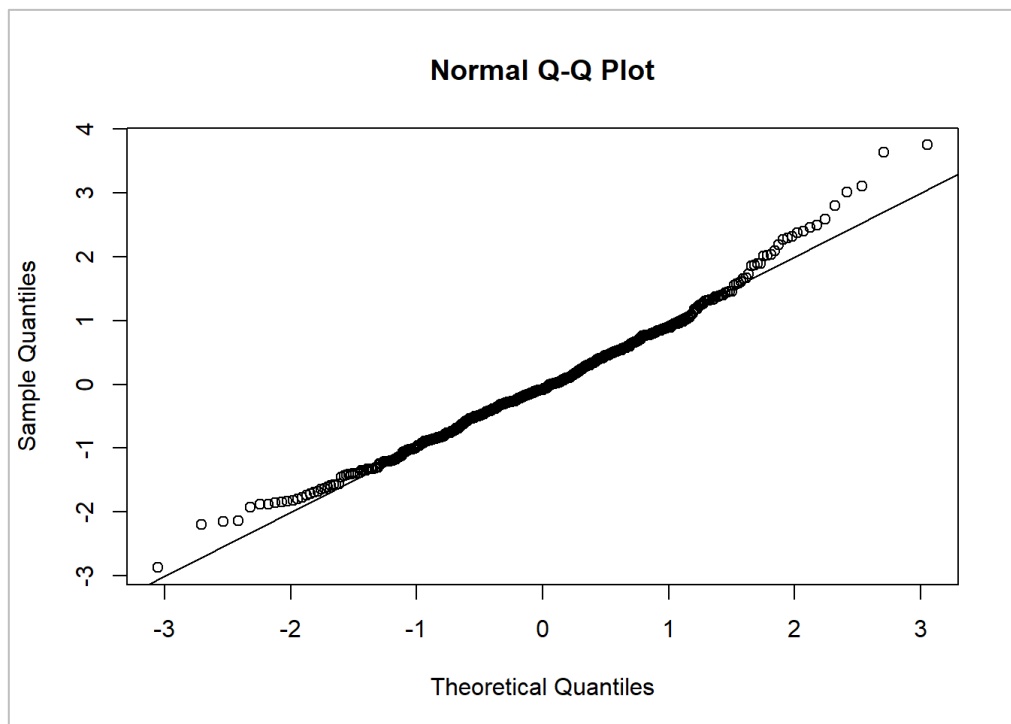


Figura 39. Gràfic QQ-Plot dels residus

## 6.6. LEBRON

En el cas de l'índex LEBRON els resultats ens ensenyen que l'efecte principal testat en el seu conjunt és significatiu en la variable resposta, també ho són l'edat i els minuts jugats. Més detalladament, els efectes fixos assenyalen que hi ha evidències per pensar que totes les posicions tenen un impacte significatiu en la variable resposta LEBRON. Així doncs, el valor predit quan les variables explicatives estan en els seus valors de referència és de -1.843. Aquest valor predit de l'índex LEBRON empitjora si el jugador, en els mateixos valors de referència, és Ala-Pivot(PF), que llavors es preveu que disminuirà 0.902 punts. De la mateixa manera, l'índex disminuirà 1.129 punts si el jugador és Base(PG), 1.503 punts si el jugador és Aler(SF) i 1.821 punts si és Escorta(SG). Sempre tenint en compte que el perfil de jugador no varia en les altres variables, és a dir, que és un jugador de 23 anys o menys i que no té formació als Estats Units. Pel que fa als factors d'edat, es preveu una pujada significativa de 0.332 punts si, tenint en compte els valors de referència, el jugador pertany al grup de 24 a 31 anys en comptes del grup de 23 anys o menys.

```

Random effects:
  Formula: ~1 | TEAM
            (Intercept) Residual
StdDev:    0.4797288 1.295137

Fixed effects:  LEBRON ~ POS + FormacioUSA + Age_factors + MIN
                Value Std.Error DF t-value p-value
(Intercept)    -1.8430121 0.28745629 405 -6.411452 0.0000
POSPF          -0.9022210 0.20117223 405 -4.484819 0.0000
POSPG          -1.1393366 0.20648617 405 -5.517738 0.0000
POSSF          -1.5039842 0.20357830 405 -7.387743 0.0000
POSSG          -1.8215522 0.19660912 405 -9.264841 0.0000
FormacioUSASI  -0.1522775 0.19126621 405 -0.796155 0.4264
Age_factors24 a 31 0.3328355 0.14666095 405 2.269421 0.0238
Age_factors32 o més 0.1690998 0.20821627 405 0.812135 0.4172
MIN            0.1150059 0.00758179 405 15.168693 0.0000
  
```

Figura 40. Informació sobre els efectes fixos i aleatoris del model lineal mixt

ANOVA	DF	F-Value	p-value
(Intercept)	405	41.1067126	4.018429e-10
POS	405	24.1189681	0.000000e+00
FormacioUSA	405	0.6338624	4.264085e-01
Age_factors	405	2.6293346	7.335748e-02
MIN	405	230.0892531	0.000000e+00

Taula 17. Resultats de l'anova del model per testar els efectes fixos

En el cas dels Pivots(C) la mitjana esperada és de 0.783, en el cas dels Ala-Pivots(PF) és de -1.119, en el cas dels Bases(PG) és de -0.357, pel cas dels Alers(SF) és de -0.721 i pel cas dels Escortes(SG) és de -1.039. Totes aquestes estimacions tenen unes desviacions estàndard associades molt petites, que van des de 0.18 la més petita a 0.187 la més gran. Les comparacions entre les posicions són significatives en el cas dels Pivots(C), que es preveu que tingui una mitjana més alta que totes les altres posicions.

En el cas del Ala-Pivots(PF) aquesta diferència s'espera que sigui 0.902, en el cas dels Bases(PG) de 1.139, en el cas dels Alers(SF) de 1.604 i en el cas del Escortes(SG) de 1.822. També resulta significativa la diferència entre mitjanes estimades entre les categoria d'Ala-Pivot(PF) i les categories d'Aler(SF) i d'Escorta(SG), amb una diferència esperada de 0.602 favorable a la categoria d'Ala-Pivot(PF) respecte als Alers(SF) i una diferència esperada de 0.919 positiva en el cas dels Escortes(SG).

Pel que fa a les altres posicions, l'única comparació significativa és la de Base(PG) amb Escorta(SG) que es preveu una diferència positiva de 0.682 favorable als Bases(PG). Per últim, si s'avaluen les mitjanes estimades pels diferents grups d'edat, s'observa que pel grup de 23 o menys anys es preveu que la mitjana sigui de -0.428 amb una desviació estàndard de 0.161. Pel grup d'edat que recull els jugadors de 24 a 31 anys la mitjana s'espera que sigui de -0.125 amb una desviació estàndard de 0.143, mentre que pel darrer grup d'edat, 32 o més, s'espera que sigui de -0.289 amb una desviació estàndard de 0.198. L'única diferència significativa entre les mitjanes estimades ve donada per les categories 23 anys o menys i 24 a 31 anys. S'estima que la diferència sigui significativa de 0.333 punts favorable als jugadors de 24 a 31 anys.



```

$emmeans
  POS emmean    SE df lower.CL upper.CL
C     0.783 0.180 29    0.415  1.1500
PF    -0.119 0.182 29   -0.491  0.2517
PG    -0.357 0.185 29   -0.735  0.0215
SF    -0.721 0.187 29   -1.104 -0.3381
SG    -1.039 0.181 29   -1.408 -0.6693

Results are averaged over the levels of: FormacioUSA, Age_factors
Degrees-of-freedom method: containment
Confidence level used: 0.95

$constrasts
  contrast estimate    SE df t.ratio p.value
C - PF      0.902 0.201 405   4.485 0.0001
C - PG      1.139 0.206 405   5.518 <.0001
C - SF      1.504 0.204 405   7.388 <.0001
C - SG      1.822 0.197 405   9.265 <.0001
PF - PG     0.237 0.202 405   1.175 0.7659
PF - SF     0.602 0.201 405   2.992 0.0244
PF - SG     0.919 0.192 405   4.785 <.0001
PG - SF     0.365 0.206 405   1.766 0.3949
PG - SG     0.682 0.197 405   3.464 0.0053
SF - SG     0.318 0.193 405   1.645 0.4694

```

```

$emmeans
  Age_factors emmean    SE df lower.CL upper.CL
23 o menys  -0.458 0.161 29   -0.787 -0.129
24 a 31     -0.125 0.143 29   -0.417  0.167
32 o més    -0.289 0.198 29   -0.695  0.117

Results are averaged over the levels of: POS, FormacioUSA
Degrees-of-freedom method: containment
Confidence level used: 0.95

$constrasts
  contrast                estimate    SE df t.ratio p.value
23 o menys - 24 a 31    -0.333 0.147 405  -2.269 0.0613
23 o menys - 32 o més  -0.169 0.208 405  -0.812 0.6957
24 a 31 - 32 o més     0.164 0.188 405   0.872 0.6583

```

```

  MIN emmean    SE df lower.CL upper.CL
  22 -0.291 0.132 29   -0.561 -0.0207

```

Figura 41. Mitjanes estimades de les variables POS, Age\_Factors i MIN i contrastos entre les diferents categories de POS i Age\_Factors

Els gràfics de validació donen no donen informació alarmant respecte a les estimacions. Es veu clarament la tendència normal que segueixen els residus, a excepció d'algun valor extrem situat a les cues. El gràfic de residus estandarditzats en relació als valors ajustats es veu un núvol de punts que no es distribueix amb un ordre preestablert.

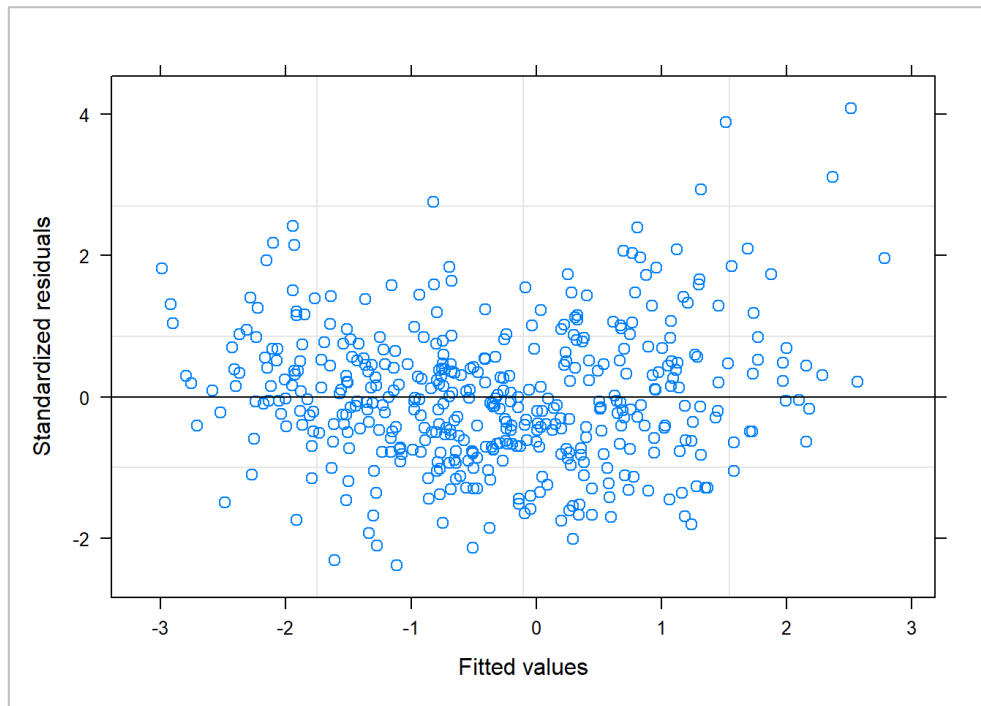


Figura 42. Gràfic dels valors ajustats vs residus estandarditzats

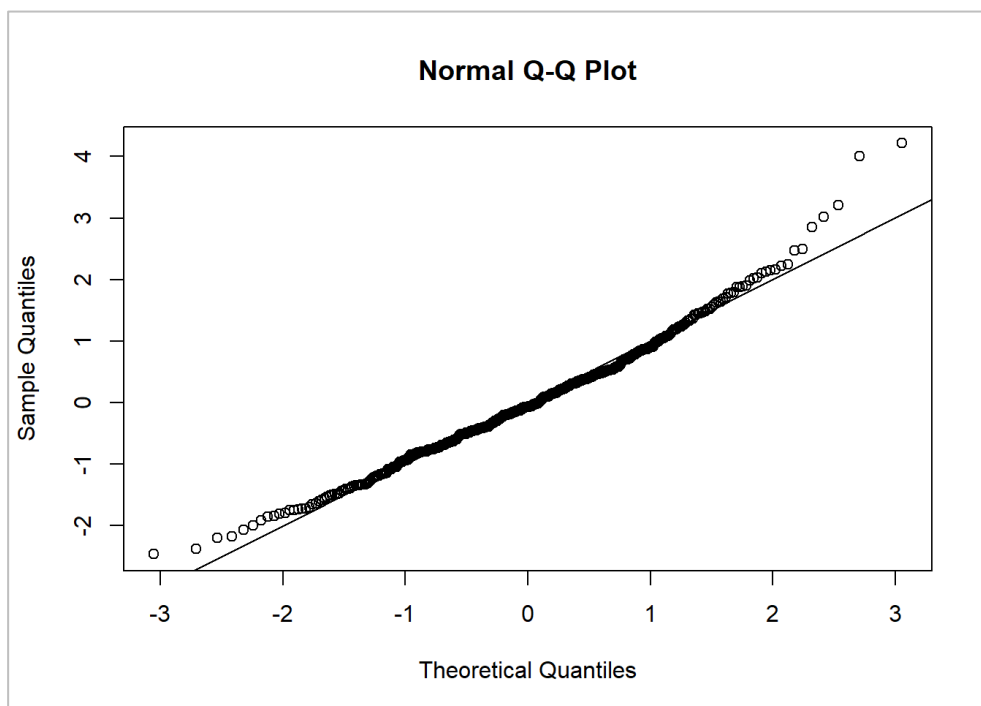


Figura 43. Gràfic QQ-Plot dels residus

## 7.DISCUSSIÓ

L'objectiu d'aquest treball és determinar si els diferents índexs d'avaluació d'impacte es veuen condicionats per la posició de joc dels jugadors d'un equip de bàsquet. Per tal de fer-ho, s'ha estimat un model per a cada índex tenint com a efecte principal la posició dels jugadors. En aquest model també s'ha inclòs l'efecte aleatori de l'equip en el què juga, així com els efectes fixos que donen informació sobre l'edat del jugador, els minuts jugats i la seva formació.

Centrant-nos en l'efecte principal estudiat, la posició del jugador, resulta significatiu en tots els índexs estudiats, exceptuant l'índex PlusMinus. A cada model generat es troben evidències per creure que les posicions dels jugadors tenen incidència en el valor resultant dels índexs. Com a excepció d'això només es troba, a part del cas de l'índex PlusMinus mencionat anteriorment, el cas de l'índex RAPTOR que resulta significatiu en totes les posicions menys en el cas de la posició de Base (PG). És interessant observar que en tots els models executats on la variable Posició és significativa, sempre ha resultat ser la posició de Pivot (C) la que presenta una estimació mitjana més alta. Per altra banda, es descobreix que els jugadors que juguen a la posició de Escorta (SG) presenten les estimacions mitjanes més baixes per a tots els índexs. A partir d'aquí, l'ordre per les altres posicions va canviant. L'índex PER,PIE i LEBRON ordenen de la mateixa manera la resta de posicions, predient que, en segon lloc, la mitjana esperada més alta serà pels jugadors Ala-Pivots (PF), la tercera posició amb la mitjana esperada més alta serà per els Bases (PG) i la quarta serà pels Alers (SF). No és sorprenent que els índexs PER i PIE es comportin de manera molt semblant, ja que s'ha descobert que tenen una molt forta correlació. Pels índexs restants, l'índex RAPTOR col·loca en segona posició els Bases(PG) seguits pels Alers (SF) i pels Ala-Pivots (PF); i l'índex RPM estima que la segona mitjana esperada més alta ve dels Ala-pivots (PF), seguida pels Alers (SF) i després pels Bases (PG).

En el conjunt d'efectes analitzats, no podem oblidar l'efecte aleatori del model, que pren en consideració els equips dels jugadors. En primer lloc, es pot veure com en el conjunt de tots els models l'efecte aleatori de l'equip no recull massa variabilitat en l'estimació dels índexs. Amb l'ajuda del coeficient de determinació intraclasse, es pot veure com en tots els models la variància explicada per l'efecte d'equip és petita. Aquest fet és interessant, ja que contradiu frontalment la creença generalitzada que considera que el rendiment d'un jugador es veu fortament condicionat per l'equip on juga. Com a excepció d'això tenim l'índex PlusMinus, que presenta un percentatge de variabilitat explicada per l'efecte equip d'un 48,534%, valor que es troba molt per sobre dels obtinguts amb els altres models. És lògic que sigui així, ja que l'índex PlusMinus per definició està formulat donant molt pes a la victòria de l'equip, per la qual cosa és evident que l'efecte de l'equip es veurà representat.

Els efectes aleatoris també ens ensenyen que la gran variabilitat ve donada dins dels propis equips. En tots els índexs, ressalten els valors de la desviació estàndard dels residuals, que recullen la variabilitat no explicada per la resta d'efectes del model. La proporció de la variància explicada per l'efecte aleatori de l'equip en relació amb la variància total del model ens permeten observar una sèrie de similituds entre índexs. Els valors que presenten els índexs RAPTOR, RPM i LEBRON són molt semblants amb un 11,035%, un 11,323% i un 12,065% respectivament. Aquests valors ens indiquen que només aquest percentatge de variància total del model és explicada per l'efecte aleatori i que la variació es deu principalment a la variació dins dels equips. Aquest fet s'accentua encara més quan s'observen els valors de PER i PIE, que són molt semblants a un 0%. En aquests dos índexs, una gran proporció de la variació es deu a les diferències dins dels equips i la variació donada per la diferència entre equips és quasi nul·la. Per la qual cosa, s'intueix que la variabilitat pot ser donada més per diferències individuals.

	<b>Desv. típica Equip</b>	<b>Var. Equip</b>	<b>Desv. típica Error</b>	<b>Var Error</b>	<b>% Variabilitat</b>
<b>PLUSMINUS</b>	1,978386	3,914011	2,037253	4,150400	48,534%
<b>PER</b>	0,000616	0,000000	3,335221	11,123699	0,000%
<b>RAPTOR</b>	0,899073	0,808333	2,552872	6,517155	11,035%
<b>RPM</b>	0,687792	0,473058	1,924809	3,704890	11,323%
<b>PIE</b>	0,000106	0,000000	2,292643	5,256212	0,000%
<b>LEBRON</b>	0,479729	0,230140	1,295137	1,677380	12,065%

*Taula 18. Coeficient de determinació intraclasse per a cada índex.*

És important destacar també la forta significació que ha tingut l'efecte dels Minuts en tots els models. Hi ha evidències per pensar que els minuts jugats pels jugadors resulten rellevants per tal d'explicar els valors dels índexs d'avaluació d'impacte de jugadors. De la mateixa manera, cal destacar també que en els índexs PlusMinus, RAPTOR i LEBRON ha resultat significatiu també l'efecte de l'Edat. Així doncs, també es pot concloure que els jugadors en edat madura presenten estimacions mitjanes més altes pels índexs PlusMinus, RAPTOR i LEBRON de manera significativa.

Cal tenir en consideració que l'anàlisi s'ha centrat exclusivament en les dades d'una sola temporada, cosa que pot considerar-se una limitació. Lligat a això, també s'ha de tenir present que les dades utilitzades són resums per temporada i seria més potent poder utilitzar les dades per partit. Aquest estudi s'ha limitat a analitzar l'efecte de 4 variables explicatives per tal

d'intentar explicar els índexs d'avaluació d'impacte dels jugadors. Degut a la complexitat de càlcul d'aquests, és important tenir en compte les limitacions d'aquest estudi. Seria convenient entendre quines altres variables poden tenir una forta influència a l'hora d'explicar aquests índexs. D'aquesta manera, es podrien incloure en el model i fer una predicció més precisa. Una proposta interessant seria la de tenir en compte les característiques físiques de cada jugador. D'alguna manera, poder incloure al model no només l'alçada del jugador, sinó també el pes i l'envergadura. És important per això trobar una estadística que pogués relacionar més d'un aspecte fisiològic, ja que la ràpida evolució del bàsquet fa que els jugadors siguin molt diferents entre ells i cada sigui més difícil trobar un perfil físic que encaixi amb cada posició del joc. Una bona opció seria l'índex de massa corporal ajustat (IMC) que s'obté dividint el pes d'una persona en kg per el producte de la seva alçada elevada al quadrat.

Resumidament, es tenen evidències per pensar que els càlculs dels índexs no són prou complets com per eliminar l'efecte de la posició del jugador. Sembla ser que tendeixen a recompensar més les posicions interiors, oferint puntuacions més altes a jugadors com els Pivots( C) i els Ala-Pivots(PF). Això indica un cert biaix cap a aquestes posicions, avaluant de manera més favorable el seu rendiment. Amb les conclusions extretes, sembla indicar que els índexs d'avaluació d'impacte de jugadors poden resultar una eina molt útil a l'hora de comparar i avaluar jugadors, sempre que es faci tenint en compte la posició del jugador.

## 8.BIBLIOGRAFIA

1. BasketballReference (no date) *Calculating per*, Basketball Reference. Disponible a: <https://www.basketball-reference.com/about/per.html>
2. Berri, David J. "Who Is 'Most Valuable'? Measuring the Player's Production of Wins in the National Basketball Association." *Managerial and Decision Economics*, vol. 20, no. 8, 1999, pp. 411–27. JSTOR, Disponible a: <http://www.jstor.org/stable/3108257>
3. Easily harvest (scrape) web pages (no date) *rvest.tidyverse*. Disponible a: <https://rvest.tidyverse.org/>
4. González,P (2019) *Las estadísticas avanzadas en el baloncesto*, Revistasuma. Disponible a: <https://revistasuma.fespm.es/wp-content/uploads/2021/08/Las-estadisticas-avanzadas.pdf>
5. Hollinger, J. (2007) *What is per?*, ESPN. Disponible a: [https://www.espn.com/nba/columns/story?columnist=hollinger\\_john&id=2850240](https://www.espn.com/nba/columns/story?columnist=hollinger_john&id=2850240)
6. Kubatko, Justin, et al. "A Starting Point for Analyzing Basketball Statistics" *Journal of Quantitative Analysis in Sports*, vol. 3, no. 3, 2007. <https://doi.org/10.2202/1559-0410.1070>
7. Martínez, J. A. (2010). *Una revisión de los sistemas de valoración de jugadores de baloncesto (I). Descripción de los métodos existentes*. Disponible a: *Revista Internacional de Derecho y Gestión del Deporte*, 10, [https://www.upct.es/~beside/Textos/SVI\\_\\_JAM.pdf](https://www.upct.es/~beside/Textos/SVI__JAM.pdf)
8. Martinez, J. (2012) 'Factors determining production (FDP) in basketball', *Economics and Business Letters*, 1(1), p. 21. doi:10.17811/eb1.1.1.2012.21-29.
9. NBAstuffer (2021) *Raptor explained*, NBAstuffer. Disponible a: <https://www.nbastuffer.com/analytics101/raptor/>
10. NBAstuffer (2021) *Lebron metric in basketball explained*, NBAstuffer. Disponible a: <https://www.nbastuffer.com/analytics101/lebron/>

11. Narsu, K. (2022) *Lebron introduction*, Basketball Index. Disponible a: <https://www.bball-index.com/lebron-introduction/>
12. Rai, S. (2015) *Physical and physiological characteristics of basketball player by playing position.*, Academia.edu. Disponible a: [https://www.academia.edu/10032307/Physical\\_and\\_physiological\\_characteristics\\_of\\_Basketball\\_player\\_by\\_playing\\_position](https://www.academia.edu/10032307/Physical_and_physiological_characteristics_of_Basketball_player_by_playing_position)
13. Ransone, J. (2017) *Physiologic profile of basketball athletes*, Gatorade Sports Science Institute. Disponible a: <https://www.gssiweb.org/sports-science-exchange/article/physiologic-profile-of-basketball-athletes>
14. Silver, N. (2019) *How our raptor metric works*, FiveThirtyEight. Disponible a: <https://fivethirtyeight.com/features/how-our-raptor-metric-works/>
15. Subirana, I. et al. (2023) *CompareGroups 4.0: Descriptives by groups*. Disponible a: [https://cran.r-project.org/web/packages/compareGroups/vignettes/compareGroups\\_vignette.html](https://cran.r-project.org/web/packages/compareGroups/vignettes/compareGroups_vignette.html)
16. Terner, Z. and Franks, A. (2021) 'Modeling player and team performance in basketball', *Annual Review of Statistics and Its Application*, 8(1), pp. 1–23. doi:10.1146/annurev-statistics-040720-015536.
17. Torres-Saavedra, P. (2018) *Modelos Estadísticos Avanzados (ESMA6995)*, Pegasus. Disponible a: <https://pegasus.uprm.edu/~pedro.torres/book/modelos-lineales-mixtos-mlm.html>
18. Wikipedia (2023) *National Basketball Association*, Wikipedia. Disponible a: [https://en.wikipedia.org/wiki/National\\_Basketball\\_Association](https://en.wikipedia.org/wiki/National_Basketball_Association)
19. WorldOfBasketball *Basketball history*, WorldofBasketball Disponible a: <https://www.worldofbasketball.org/basketball-history.htm>
20. Zein (2022) *The definitive per criticism*, The Z score. Disponible a: <https://thezscore.com/2016/02/17/the-definitive-per-criticism/>
21. Zuccolotto, P. et al (2021), *Alley-oop! Basketball analytics in R*. Significance, <sup>18</sup>: 26-31. <https://doi.org/10.1111/1740-9713.01507>

## ANNEX: CODI D'R

```
# Carreguem llibreries necessàries
library(tidyr)
library(ggplot2)
library(gridExtra)
library(dplyr)
library(ggcorrplot)
library(compareGroups)
library(tinytex)
library(broom)
library(kableExtra)
library(nlme)
library(emmeans)
library(rvest)
library(xml2)

bbdd <- read.delim("C:/Users/mdejo/Desktop/BBDD_Bo/Completa.txt") bdd <- bdd[, -1]
# Canvia el nom de les columnes
colnames(bdd)[11] <- "FG_pct"
colnames(bdd)[14] <- "X3P_pct"
colnames(bdd)[17] <- "FT_pct"
colnames(bdd)[29] <- "PLUSMINUS"
# BIO
bio <- read.delim("C:/Users/mdejo/Desktop/BBDD_Bo/BioPlayers.txt")
bio <- bio[, 1:10]
setdiff(bdd$PLAYER, bio$PLAYER)
bdd_total <- merge(x=bdd, y=bio, by=c("PLAYER", "TEAM", "AGE"))
# PER
url_inicial <- "https://www.espn.com/nba/hollinger/statistics/_/page/"
url_final <- "/year/2022/qualified/false"
tablas <- list()
for (i in 2:13) {
  url <- paste0(url_inicial, i, url_final)
  pagina <- read_html(url)
  tabla <- html_table(html_nodes(pagina, "table")[[1]])
  tablas[[i - 1]] <- tabla
}
per <- do.call(rbind, tablas)
#URL de la primera pagina
url_primera <- "https://www.espn.com/nba/hollinger/statistics/_/year/2022/qualified/false"
primera_pagina <- read_html(url_primera)
taula1 <- html_table(html_nodes(primera_pagina, "table")[[1]])
per <- rbind(per, taula1)
colnames(per) <- c("RK", "PLAYER", "GP", "MGP", "TS_pct", "AST", "TO", "USG", "ORR", "DRR",
"REBR", "PER", "VA", "EWA")
per <- subset(per, PLAYER != "Hollinger Stats - Player Efficiency Rating - All Players")
per <- subset(per, PLAYER != "PLAYER")
for (i in 1:nrow(per)) {
  per$PLAYER[i] <- strsplit(as.character(per$PLAYER[i]), ",")[[1]][1]
}
per <- per[, c("PLAYER", "PER")]
duplicated_cc <- duplicated(per$PLAYER)
cc_clean <- subset(per, !duplicated_cc)
per <- cc_clean
rm(cc_clean)
setdiff(bdd_total$PLAYER, per$PLAYER)
bdd_total <- merge(x=bdd_total, y=per, by=c("PLAYER"))
# RAPTOR
raptor <- read.delim("C:/Users/mdejo/Desktop/BBDD_Bo/RAPTOR.txt")
for (i in 1:nrow(raptor)) {
  raptor$PLAYER[i] <- strsplit(as.character(raptor$PLAYER[i]), "'2")[[1]][1]
}
columnes <- c(2, 12, 13, 14)
raptor <- raptor[, columnes]
setdiff(raptor$PLAYER, bdd_total$PLAYER)
raptor$PLAYER[which(raptor$PLAYER == "OG Anunoby")] <- "O.G. Anunoby"
raptor$PLAYER[which(raptor$PLAYER == "P.J. Dozier")] <- "PJ Dozier"
raptor$PLAYER[which(raptor$PLAYER == "Jeff Dowtin")] <- "Jeff Dowtin Jr."
raptor <- raptor[-which(raptor$PLAYER == "Luca Vildoza")[1], ]
{
  # l'[1] es perque la funcio which et retorna una llista de tothom que compleix la
  # condició i en aquest cas només eliminem la primera (perq només en tenim una)
}
setdiff(raptor$PLAYER, bdd_total$PLAYER)
bdd_total <- merge(x=bdd_total, y=raptor, by=c("PLAYER"))
# RPM
# URL base
url_inicial <- "http://www.espn.com/nba/statistics/rpm/_/year/2022/page/"
# Taules amb cada pagina
tablas <- list()
# Iterar sobre las pagines
for (i in 1:14) {
  # URL de cada pagina
```



```

url <- paste0(url_inicial, i)
pagina <- read_html(url)
# Extraure la taula
tabla <- html_table(html_nodes(pagina, "table")[[1]])

tablas[[i]] <- tabla
}
rpm_df <- do.call(rbind, tablas)
colnames(rpm_df) <- c("RANK", "PLAYER", "TEAM", "GP", "MIN", "ORPM", "DRPM", "RPM", "WINS")
rpm_df <- subset(rpm_df, PLAYER != "NAME")
var_fora <- c(1,3,4,5,6,7,9)
rpm_df <- rpm_df[, -var_fora]
for (i in 1:nrow(rpm_df)) {
  rpm_df$PLAYER[i] <- strsplit(as.character(rpm_df$PLAYER[i]), ",")[[1]][1]
}
rpm_df$RPM <- as.numeric(rpm_df$RPM)
setdiff(bbdd_total$PLAYER, rpm_df$PLAYER)
setdiff(rpm_df$PLAYER, bbdd_total$PLAYER)
length(setdiff(bbdd_total$PLAYER, rpm_df$PLAYER)) # no es te informació de 48 jugadors
bbdd_total2 <- bbdd_total
bbdd_total2 <- merge(x=bbdd_total, y=rpm_df, by=c("PLAYER"))
#
# PIE
pimpact <- read.delim("C:/Users/mdejo/Desktop/BBDD_Bo/PIE.txt")
columnes <- c(2,23)
pimpact <- pimpact[, columnes]
setdiff(bbdd_total2$PLAYER, pimpact$PLAYER)
setdiff(pimpact$PLAYER, bbdd_total2$PLAYER)
bbdd_total2 <- merge(x=bbdd_total2, y=pimpact, by=c("PLAYER"))
#
# LEBRON
url <- "https://www.bball-index.com/lebron-database/"
webpage <- read_html(url)
table <- webpage %>%
  html_nodes("table") %>%
  .[[1]] %>%
  html_table()
# canvis
lebron_df <- table[table$Season=="2021-22", ]
colnames(lebron_df)[which(colnames(lebron_df) == "Player")] <- "PLAYER"
colnames(lebron_df)[which(colnames(lebron_df) == "O-LEBRON")] <- "OLEBRON"
colnames(lebron_df)[which(colnames(lebron_df) == "D-LEBRON")] <- "DLEBRON"
colnames(lebron_df)[which(colnames(lebron_df) == "Pos")] <- "POS"
colnames(lebron_df)[which(colnames(lebron_df) == "Offensive Archetype")] <-
"OFFENSIVE.ARCHETYPE"
setdiff(bbdd_total2$PLAYER, lebron_df$PLAYER)
setdiff(lebron_df$PLAYER, bbdd_total2$PLAYER)
lebron_df$PLAYER[which(lebron_df$PLAYER == "OG Anunoby")] <- "O.G. Anunoby"
lebron_df$PLAYER[which(lebron_df$PLAYER == "P.J. Dozier")] <- "PJ Dozier"
lebron_df$PLAYER[which(lebron_df$PLAYER == "Jeff Dowtin")] <- "Jeff Dowtin Jr."
columnes <- c(1,3,4,6,7,9,13,14,15,16,17,18)
# columnes <- c(2,5,8,10,11,12)
lebron_df <- lebron_df[, -columnes]
bbdd_total2 <- merge(x=bbdd_total2, y=lebron_df, by=c("PLAYER"))
colnames(lebron_df)
#
# PREPROCESSING
bbdd_total2$TEAM <- factor(bbdd_total2$TEAM) # 30 equips diferents
bbdd_total3 <- bbdd_total2
sort(table(bbdd_total3$COLLEGE), decreasing = T)
bbdd_total3$COLLEGE <- factor(bbdd_total3$COLLEGE) # 141 college diferents
# canviar la variable PER a numérica
bbdd_total3$PER <- as.numeric(bbdd_total3$PER)
# Crear una variable de grupos de GamesPlayed
grups_gp <- cut(bbdd_total2$GP, breaks = seq(0, 90, by = 10), right = F)
# Contar el numero d'observacions de cada grup
taula_gp <- table(grups_gp)
# Graficar els resultats amb diagrama de barras
barp <- barplot(taula_gp, main = "Distribución de Partits Jugats", xlab = "Partits Jugats",
ylab = "Frecuencia", ylim = c(0, 130))
text(barp, taula_gp, labels = taula_gp, pos = 3)
Inferior20 <- bbdd_total2[bbdd_total2$GP < 20, ]
bbdd_total3 <- bbdd_total3[bbdd_total3$GP >= 20, ]
# posar ho per ordre i després fer la neteja de les 2 variables sobrants
# 37 i 38 fora
bbdd_total3 <- subset(bbdd_total3, select = -DRAFT.ROUND)
bbdd_total3 <- subset(bbdd_total3, select = -DRAFT.NUMBER)
table(bbdd_total3$POS)
# canviem posició a factor
bbdd_total3$POS <- as.factor(bbdd_total3$POS)
# Crear un vector con las categorías originales y sus correspondientes nuevas categorías
categ <- c("C-PF", "PF-SF", "PG-SG", "SF-SG", "SG-PG", "SG-SF")
new_categ <- c("C", "PF", "PG", "SF", "SG")
# Reemplazar categorías que contienen un guió per la primera part de la cadena
bbdd_total3$POS <- gsub("-", "", bbdd_total3$POS)
# Creació de Variable Binaria de si van anar a college o no
CollageBinari <- ifelse(bbdd_total3$COLLEGE %in% c("None"), "NO", "SI")
table(CollageBinari)
FormacioUSA <- CollageBinari
bbdd_total3 <- cbind(bbdd_total3, FormacioUSA)

```

```

# variable factorial d'Edat.
table(bbdd_total3$AGE)

Age_factors <- ifelse(bbdd_total3$AGE %in% c("19","20","21","22","23"), "23 o
menys",ifelse(bbdd_total3$AGE %in% c("23","24","25","26","27","28","29","30","31"), "24 a
31","32 o més"))
bbdd_total3<-cbind(bbdd_total3,Age_factors)
table(bbdd_total3$Age_factors)
variables<-c("PLAYER",
"TEAM","MIN","POS","Age_factors","FormacioUSA","PLUSMINUS","PER","OALLTOT","RPM","PIE","LEB
RON")
total3<- bbdd_total3[,variables ]
# Anàlisi descriptiu
##TEAM
taula_teams<-table(total3$TEAM)
barp<-barplot(taula_teams, main = "Distribución per Equipos", xlab = "Equipos", ylab =
"Frecuencia",ylim = c(0, 26), cex.names = 0.6, las = 3)
text(barp, taula_teams,labels= taula_teams, pos = 3)
maxim <- max(taula_teams)
print(maxim)
minim <- min(taula_teams)
print(minim)
taula_freq<-table(taula_teams)
taula_freq
moda <- names(taula_freq)[taula_freq == max(taula_freq)]
moda
## MINUTS
par(mfrow = c(1, 2), cex.main = 0.7, cex.axis = 0.6, cex.lab = 0.6, mgp=c(3,0.4,0), las=1)
k<-3
boxplot(total3[,k], horizontal=F)
title(main=paste("Boxplot of", names(total3)[k] ), xlab = names(total3)[k], line = 1)
#taula summary
summary_taula <- tidy(summary(total3$MIN))
summary_table forma<- summary_taula %>%
  kable() %>%
  kable_styling()
summary_table forma
##MIN GRAFICS
plot1<-ggplot(total3, aes(x = MIN, y = PLUSMINUS)) +
  geom_point() +
  labs(x = "MIN", y = "PLUSMINUS", title = " MIN vs PLUSMINUS")
plot2<-ggplot(total3, aes(x = MIN, y = PER)) +
  geom_point() +
  labs(x = "MIN", y = "PER", title = "MIN vs PER")
plot3<-ggplot(total3, aes(x = MIN, y = OALLTOT)) +
  geom_point() +
  labs(x = "MIN", y = "RAPTOR", title = "MIN vs RAPTOR")
plot4<-ggplot(total3, aes(x = MIN, y = RPM)) +
  geom_point() +
  labs(x = "MIN", y = "RPM", title = "MIN vs RPM")
plot5<-ggplot(total3, aes(x = MIN, y = PIE)) +
  geom_point() +
  labs(x = "MIN", y = "PIE", title = "MIN vs PIE")
plot6<-ggplot(total3, aes(x = MIN, y = LEBRON)) +
  geom_point() +
  labs(x = "MIN", y = "LEBRON", title = "MIN vs LEBRON")
grid.arrange(plot1, plot2, plot3,plot4,plot5,plot6,ncol=2)
# Calcular la matriu de correlacions
total_est <- subset(total3, select = c(3,7,8,9,10,11,12))
cor_matrix <- cor(total_est)
ggcorrplot(cor_matrix, type = "lower", lab = TRUE, lab_size = 2.8,)
## Índexs
#Posicions de les variables corresponents als índexs
#7,8,9,10,11,12
par(mfrow = c(1, 3), cex.main = 0.7, cex.axis = 0.6, cex.lab = 0.6, mgp=c(3,0.4,0), las=1)
k<-7
boxplot(total3[,k], horizontal=F)
title(main=paste("Boxplot of", names(total3)[k] ), xlab = names(total3)[k], line = 1)
k<-8
boxplot(total3[,k], horizontal=F)
title(main=paste("Boxplot of", names(total3)[k] ), xlab = names(total3)[k], line = 1)
k<-9
boxplot(total3[,k], horizontal=F)
title(main=paste("Boxplot of", names(total3)[k] ), xlab = names(total3)[k], line = 1)
k<-10
boxplot(total3[,k], horizontal=F)
title(main=paste("Boxplot of", names(total3)[k] ), xlab = names(total3)[k], line = 1)
k<-11
boxplot(total3[,k], horizontal=F)
title(main=paste("Boxplot of", names(total3)[k] ), xlab = names(total3)[k], line = 1)
k<-12
boxplot(total3[,k], horizontal=F)
title(main=paste("Boxplot of", names(total3)[k] ), xlab = names(total3)[k], line = 1)
total_est <- subset(total3, select = c(7,8,9,10,11,12))
# Calcular la matriu de correlacions
cor_matrix <- cor(total_est)
# Graficar la matriu de correlacions com un heatmap

```

```

ggplot(data = reshape2::melt(cor_matrix), aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", midpoint = 0,
    limits = c(-1, 1), name = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1,
    size = 10, hjust = 1),
    axis.text.y = element_text(size = 10))
ggcorrplot(cor_matrix, type = "lower", lab = TRUE, lab_size = 2.8,)
# Analisis variables categoriques
### FORMACO USA
resu2<-compareGroups(FormacioUSA ~ PLUSMINUS+PER+OALLTOT+RPM+PIE+LEBRON, data=total3,
  Q1=0.025, Q3=0.975)
resu2
createTable(resu2)
summary(resu2[c(1,2,3,4,5,6)])
plot1<-ggplot(total3, aes(x = FormacioUSA, y = PLUSMINUS)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "FormacioUSA", y = "Index PLUSMINUS", title = "Diagrama de Medias PLUSMINUS")
plot2<-ggplot(total3, aes(x = FormacioUSA, y = PER)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "FormacioUSA", y = "Index PER", title = "Diagrama de Medias PER")
plot3<-ggplot(total3, aes(x = FormacioUSA, y = RPM)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "FormacioUSA", y = "Index RPM", title = "Diagrama de Medias RPM")
plot4<-ggplot(total3, aes(x = FormacioUSA, y = OALLTOT)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "FormacioUSA", y = "Index OALLTOT", title = "Diagrama de Medias OALLTOT")
plot5<-ggplot(total3, aes(x = FormacioUSA, y = PIE)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "FormacioUSA", y = "Index PIE", title = "Diagrama de Medias PIE")
plot6<-ggplot(total3, aes(x = FormacioUSA, y = LEBRON)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "FormacioUSA", y = "Index LEBRON", title = "Diagrama de Medias LEBRON")
grid.arrange(plot1, plot2, plot3, plot4, plot5, plot6, ncol = 3)
### POSICIO
resu2<-compareGroups(POS ~ PLUSMINUS+PER+OALLTOT+RPM+PIE+LEBRON, data=total3, Q1=0.025,
  Q3=0.975)
resu2
createTable(resu2)
summary(resu2[c(1,2,3,4,5,6)])
plot1<-ggplot(total3, aes(x = POS, y = PLUSMINUS)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "POS", y = "Index PLUSMINUS", title = "Diagrama de Medias PLUSMINUS")
plot2<-ggplot(total3, aes(x = POS, y = PER)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "POS", y = "Index PER", title = "Diagrama de Medias PER")
plot3<-ggplot(total3, aes(x = POS, y = RPM)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "POS", y = "Index RPM", title = "Diagrama de Medias RPM")
plot4<-ggplot(total3, aes(x = POS, y = OALLTOT)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "POS", y = "Index OALLTOT", title = "Diagrama de Medias OALLTOT")
plot5<-ggplot(total3, aes(x = POS, y = PIE)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "POS", y = "Index PIE", title = "Diagrama de Medias PIE")
plot6<-ggplot(total3, aes(x = POS, y = LEBRON)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "POS", y = "Index LEBRON", title = "Diagrama de Medias LEBRON")
grid.arrange(plot1, plot2, plot3, plot4, plot5, plot6, ncol = 2)
# AGE FACTORS
resu2<-compareGroups(Age_factors ~ PLUSMINUS+PER+OALLTOT+RPM+PIE+LEBRON, data=total3,
  Q1=0.025, Q3=0.975)

```

```

resu2
createTable(resu2)
summary(resu2[c(1,2,3,4,5,6)])
plot1<-ggplot(total3, aes(x = Age_factors, y = PLUSMINUS)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "Age_factors", y = "Index PLUSMINUS", title = "Diagrama de Medias PLUSMINUS")
plot2<-ggplot(total3, aes(x = Age_factors, y = PER)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "Age_factors", y = "Index PER", title = "Diagrama de Medias PER")
plot3<-ggplot(total3, aes(x = Age_factors, y = RPM)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "Age_factors", y = "Index RPM", title = "Diagrama de Medias RPM")
plot4<-ggplot(total3, aes(x = Age_factors, y = OALLTOT)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "Age_factors", y = "Index OALLTOT", title = "Diagrama de Medias OALLTOT")
plot5<-ggplot(total3, aes(x = Age_factors, y = PIE)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "Age_factors", y = "Index PIE", title = "Diagrama de Medias PIE")
plot6<-ggplot(total3, aes(x = Age_factors, y = LEBRON)) +
  geom_point() +
  stat_summary(fun = mean, geom = "point", size = 3, color = "red") +
  stat_summary(fun = mean, geom = "line", size = 1, color = "red", group = 1) +
  labs(x = "Age_factors", y = "Index LEBRON", title = "Diagrama de Medias LEBRON")
grid.arrange(plot1, plot2, plot3, plot4, plot5, plot6, ncol = 3)
# Anàlisi models lineals mixtes
## PLUSMINUS
par(mfrow = c(1, 1), cex.main = 0.7, cex.axis = 0.6, cex.lab = 0.6, mgp=c(3,0.4,0), las=1)
# PLUSMINUS
#Model
fm1<- lme(PLUSMINUS ~ POS + FormacioUSA + Age_factors+ MIN, data= total3,random= ~ 1|TEAM)
summary(fm1)
# testat efectes fixos
anova(fm1,type="marginal")
# Post-hoc
library(emmeans)
emmeans(fm1, pairwise ~ Age_factors, adjust="tukey")
emmeans(fm1, ~ MIN)
# Validació
plot(fm1)
res<-residuals(fm1,type="pearson")
qqnorm(scale(res))
abline(0,1)
## PER
#Model
fm2<- lme(PER ~ POS + FormacioUSA + Age_factors + MIN, data= total3,random= ~ 1|TEAM)
summary(fm2)
# testat efectes fixos
anova(fm2,type="marginal")
# Post-hoc
library(emmeans)
emmeans(fm2, pairwise ~ POS, adjust="tukey")
emmeans(fm2, ~ MIN)
# Validació
plot(fm2)
res<-residuals(fm2,type="pearson")
qqnorm(scale(res))
abline(0,1)
## OALLTÓT
#Model
fm3<- lme(OALLTOT ~ POS + FormacioUSA + Age_factors+ MIN, data= total3,random= ~ 1|TEAM)
summary(fm3)
# testat efectes fixos
anova(fm3,type="marginal")
# Post-hoc
library(emmeans)
emmeans(fm3, pairwise ~ POS, adjust="tukey")
emmeans(fm3, pairwise ~ Age_factors, adjust="tukey")
emmeans(fm3, ~ MIN)
# Validació
plot(fm3)
res<-residuals(fm3,type="pearson")
qqnorm(scale(res))
abline(0,1)
#treiem outlier
outlier <- which(abs(res) == max(abs(res)))
print(outlier)
raptorlm_df<- total3[-outlier,]

```

```

fm3_rectificat <- lme(OALLTOT ~ POS + FormacioUSA + Age_factors + MIN, data = raptorlm_df,
random = ~ 1|TEAM)
summary(fm3_rectificat)
# testat efectes fixos
anova(fm3_rectificat,type="marginal")
# Post-hoc
emmeans(fm3_rectificat, pairwise ~ Age_factors, adjust="tukey")
emmeans(fm3_rectificat, ~ MIN)
# Validacio
plot(fm3_rectificat)
res<-residuals(fm3_rectificat,type="pearson")
qqnorm(scale(res))
abline(0,1)
## RPM
#Model
fm4<- lme(RPM ~ POS + FormacioUSA + Age_factors+ MIN, data= total3,random= ~ 1|TEAM)
summary(fm4)
# testat efectes fixos
anova(fm4,type="marginal")
# Post-hoc
library(emmeans)
emmeans(fm4, pairwise ~ POS, adjust="tukey")
emmeans(fm4, ~ MIN)
# Validacio
plot(fm4)
res<-residuals(fm4,type="pearson")
qqnorm(scale(res))
abline(0,1)
## PIE
#Model
fm5<- lme(PIE ~ POS + FormacioUSA + Age_factors+ MIN, data= total3,random= ~ 1|TEAM)
summary(fm5)
# testat efectes fixos
anova(fm5,type="marginal")
# Post-hoc
library(emmeans)
emmeans(fm5, pairwise ~ POS, adjust="tukey")
emmeans(fm5, ~ MIN)
# Validacio
plot(fm5)
res<-residuals(fm5,type="pearson")
qqnorm(scale(res))
abline(0,1)
## LEBRON
#Model
fm6<- lme(LEBRON ~ POS + FormacioUSA + Age_factors+ MIN, data= total3,random= ~ 1|TEAM)
summary(fm6)
# testat efectes fixos
anova(fm6,type="marginal")
# Post-hoc
library(emmeans)
emmeans(fm6, pairwise ~ POS, adjust="tukey")
emmeans(fm6, pairwise ~ Age_factors, adjust="tukey")
emmeans(fm6, ~ MIN)
# Validacio
plot(fm6)
res<-residuals(fm6,type="pearson")
qqnorm(scale(res))
abline(0,1)

```

