Métodos estadísticos aplicados al baloncesto

Paula Moreno Blazquez

Enero 2022

Agradecimientos

Este trabajo me ha permitido profundizar más en un mundo que pensaba que conocía bastante bien, y darme cuenta de que no todo lo que ocurre en un partido o en una temporada se ve bien reflejado en los números que hasta ahora conocía. Me ha permitido conocer nuevos métodos de análisis que hasta ahora desconocía y aprender cómo se haría para aplicarlos.

Resumen

Hoy en día, el deporte es un hobby muy popular por todo el mundo. Desde pequeños, los niños practican algún tipo de deporte, especialmente aquellos que son de equipo. Eso nos lleva a querer saber más del deporte, más detalles, más información. Nos entra la curiosidad de "¿quién es el mejor jugador?", "Qué equipo es mejor?", o incluso intentar prevenir qué equipo ganará según sus resultados anteriores. Y gracias a los avances tecnológicos e informáticos, cada vez se nos facilita más poder seguir un deporte desde casa, ver la estadística de los deportistas e incluso hay plataformas o juegos que nos permiten ser, de manera virtual, managers de los clubs y, por lo tanto, nos facilitan mucha información que antes era más difícil de saber.

Eso hace que, de manera progresiva, también mejore el estudio y el análisis de cada deporte, y cada vez sea más específica para cada deporte, implementando nuevos recursos para mejorar los resultados. Pero, ¿son lo suficientemente eficaces los análisis que se realizan actualmente en Europa? ¿O dichos análisis están anticuados y requieren de una actualización?

Índice

1.	Introducción	5
2.	El baloncesto	5
	2.1. Historia y reglas básicas	5
	2.2. Conceptos y definiciones básicas del baloncesto:	6
3.	Posibles análisis realizables en el baloncesto	6
	3.1. Boxscore	6
	3.2. Play-by-play	7
	3.3. Shot-charts	8
	3.4. Graphic Stats	8
4.	Estadístico Más/Menos ($Plus/Minus$)	9
	4.1. Plus/Minus, PM	9
	4.2. Adjusted Plus/Minus, APM	9
	4.3. Regularized Adjusted Plus/Minus, RAPM	9
	4.4. Real Plus/Minus, RPM	10
	4.5. Otros tipos de métricas basados en el Más/Menos	10
5.	Análisis y métodos	11
	5.1. Descripción de la base de datos	11
Re	ferencias	20
6.	Annexo	21
	6.1. Glosario	21
	6.2. ¿Qué es Git Hub?	22
	6.2.1. Ventajas	22
	6.2.2. ¿Qué es el control de versiones?	22
	6.2.3. ¿Qué es Git?	22
	6.3. Descripción de las variables	23
	6.4. Code	23

1. Introducción

En este trabajo estudiaremos más a fondo el Baloncesto, el segundo deporte más popular de Europa (solo superado por el futbol), y el cual tengo interés personal, ya que lo practico desde los 4 años.

El análisis estadístico siempre ha estado presente en el baloncesto profesional. Como parte del $staff^1$, te permite conocer las fortalezas (en qué puedes confiar) y las debilidades (en qué debes trabajar) de tu equipo y, por otro lado, te permite conocer mejor a tus rivales, para poder determinar sus puntos fuertes (de lo que debes ser consciente) y también sus debilidades (de lo que sacar provecho). Esto significa que una parte importante del partido se juega antes de empezar, cuando el staff analizan el modo de jugar de sus rivales en los partidos anteriores y a lo largo de la temporada.

Tradicionalmente, hay dos estadísticos que resumen un partido: el box-score y el play-by-play. Estos dos documentos, junto con la visualización del video del partido, han sido típicamente la fuente de datos para la evaluación de los equipos.

Pero, como aficionada del baloncesto profesional y como jugadora y entrenadora de baloncesto amater, con el paso de los años me he dado cuenta de que, la mayoria de las veces, estos documentos o estadísticos que comentamos no son del todo fiables. Pierden mucha información relevante o incluso se pueden malinterpretar dependiendo de la situación de la acción.

Por eso surgió este trabajo, por el constante pensamiento de que los análisis actuales que se hacen en este deporte en Europa son bastante pobres a nivel informativo, puesto que se basan en conceptos muy básicos, y principalmente ofensivos (que vendría a ser el $50\,\%$ de un partido). Para que nos hagamos una idea, el estadístico por preferencia es el llamado Valoraci'on y que se origin\'o en 1991 (hace 30 años) y desde entonces nunca se ha modificado.

Es por eso que, considero que actualmente los análisis que se hacen de este deporte necesitan una actualización significativa para llegar a informar de todos aquellos datos que hoy en día si se pueden recoger gracias a los avances tecnológicos, y de los cuales no se analizan por falta de dinero o porque se consideran poco relevantes.

El objetivo principal de este estudio es mejorar los análisis que se elaboran de cada partido, para poder encontrar una variable respuesta que nos diga que aportación al equipo tiene cada jugador personalmente, disminuyendo la diferencia de pesos que hay actualmente entre las aportaciones ofensivas y las aportaciones defensivas.

Este documento se estructura de la siguiente manera: a continuación, se explicará brevemente los conceptos de baloncesto que son necesarios para poner en contexto a los lectores en caso de no conocer el mundo del baloncesto y para entender los tecnicismos del trabajo y se presentaran posibles análisis que se realizan. Finalmente, se describirá la base de datos con la que se ha trabajado y sus variables, y también se explicará en profundidad el análisis que se desarrollará en este trabajo, el Más/Menos Ajustado (Adjusted Plus/Minus, APM). En la sección de resultados presentaremos la resolución del análisis y finalmente discutiremos, en la sección de conclusiones, los resultados obtenidos.

2. El baloncesto

2.1. Historia y reglas básicas

El baloncesto es un deporte de equipo que se originó en 1891, por James Naismith, profesor de educación física en la escuela YMCA de Springfield, Massachusetts, Estados Unidos.

James buscaba idear un deporte que sus alumnos pudieran practicar bajo techo, pues los duros inviernos en Massachusetts dificultaban la realización de ejercicio al aire libre, por lo que inventó el baloncesto utilizando unas cajas de melocotones y unos balones.

Con el paso de los años, este deporte, que empezó como actividad de colegio, ha ido evolucionando mucho, añadiendo más reglas, conceptos nuevos, límites de números de jugadores, se ha determinado tiempos de juego, las canastas tienen un valor distinto según la distancia, etc.



¹conjunto de personas que se encargan de gestionar y dirigir el equipo. Normalmente está formado por un entrenador, uno o dos entrenadores ayudantes, y lueg pueden haber delegados, fisioterapeuta, preparador físico, *scouter*...

Actualmente, las normas más básicas de este deporte son:

- En las ligas superiores, hay un total de 4 cuartos de 10 minutos y pueden estar en pista 5 jugadores por equipo.
- No te puedes desplazar con la pelota en las manos, es obligatorio botar con una mano (si no será una infracción y conllevará la perdida de pelota y saque de banda del equipo rival).
- Cada jugador puede realizar hasta un total de 5 faltas, que será penalizado con un saque de banda o con un tiro libre (dependerá de la situación). El jugador que ejecuta 5 faltas será expulsado del partido.
- El objetivo es encestar el máximo de puntos posibles, teniendo en cuenta que pueden sumar 1, 2 o 3 puntos, según la distancia.

2.2. Conceptos y definiciones básicas del baloncesto:

Para que podamos entender a que nos referimos en este trabajo, es necesario comprender unos conceptos básicos de vocabulario. Tendremos en cuenta los conceptos que se necesitan para realizar la valoración del jugador y/o del equipo que se utilizan en las estadísticas federadas.

- Puntos: Acumulación de tiros encestados multiplicados por su valor, que cada jugador y/o equipo realiza durante el partido
- Minutos: Número de minutos que el jugador está en pista
- Falta: Acción en la que un defensor bloquea el avance de su rival sin tener control de balón o de manera no reglamentaria (empujar, agarrar...)
- Pérdidas de balón: cuando un equipo pierde el control del balón y pasa a ser del equipo rival.
- Rebotes: Recuperación de pelota después de que el tiro sea ejecutado, pero no haya encestado.
- Recuperación de balón: Cuando un equipo consigue robar el balón al equipo rival.
- Asistencia: Es un pase a un jugador que se encuentra en una posición de ventaja o que le ayuda a conseguir una canasta sin hacer ningún bote.
- Tapón: Bloqueo de un tiro en el aire.

3. Posibles análisis realizables en el baloncesto

Viendo la gran cantidad de datos que se pueden extraer de cada partido (y de cada equipo), se han ido creando análisis que recogen estos datos y los analizan para ayudarnos a identificar y desarrollar hipótesis sobre cada jugador y/o equipo.

3.1. Boxscore

El primer análisis que se hizo fue un Box Score (Caja de puntuación) donde se recopilaba únicamente los puntos de cada jugador según el valor de esta y las faltas realizadas. Posteriormente, se fue mejorando añadiendo conceptos como rebotes, tapones, pérdidas de balón, recuperaciones de balón... Y se añadió el estadístico (que actualmente es por defecto) que se hace a partir de todos estos datos: Valoración (en inglés PIR, Performance Index Rating) que engloba todo lo básico que pasa en el partido de manera individual y que, cuanto más positivo, mejor. Este estadístico se calcula utilizando la siguiente fórmula:

```
PIR = (Puntos + Rebotes + Asistencias + Robos + Tapones + FaltasRecibidas) - \\ (Tiros de Campo Fallados + Tiros Libres Fallados + Tapones Recibidos + Prdidas + Faltas Realizadas)  (1)
```

(en el Anexo 1 encontraréis la descripción de cada variable)

Posteriormente, se añadió la variable "Más/Menos" (P/M, Plus/Minus) que tiene que ver con la diferencia de puntos en el marcador durante el tiempo que el jugador esté en pista. Esta variable sirve para ver la contribución de los jugadores cuando están en pista. Todos los jugadores parten inicialmente con un 0, y según van entrando y saliendo de la pista, esta variable se va actualizando. Por ejemplo, los jugadores que son del quinteto inicial, empiezan con el marcador 0 - 0, y un P/M = 0. Si en el minuto 5, se substituye un jugador en cancha del equipo local (J1) por otro que está descansando (J2), y el marcador va 12 - 7, el P/M del J1 pasará a ser +5. Y si al cabo de 3 minutos, se sustituye el J2 por otro (J3) y el marcador ahora va 20 - 9, el P/M del J2 será +6 ((20 - 12) - (9 - 7) = 8 - 2 = +6).

Buckner	16	1-1	4	33
Totals	57	33-41	32	147
PHILD	ELPHI	A (16	9)	
Arizin	F	FT.	F.	Pts.
Meschery	4	2-2	0	16
Meschery Chamberlain	36	98.30	2	100
Rodgers	1	9-12	2545	100
Attles	8	1-1	4	17
Larese	4	1-1	5	9
Conlin	0	0-0	1	900
Ruklick Luckenbill	0	0-2	2	0
Luckenbill	0	0-0	2	0
Totals	63	43-52	28	100
New York	26 4	2 38	41-	-147

Figura 1: Boxscore del partido de la NBA de Philadelphia Warriors contra New York Knicks, del 2 de Marzo de 1962



Figura 2: Boxscore del partido de la Euroliga de Real Madrid contra FC Barcelona, del 11 de Febrero del 2022

Aunque un Box Score es muy útil para realizar análisis básicos, ya que es muy visual y cualquier persona sin la necesidad de muchos recursos puede analizar y predecir ciertos valores, pero estadísticamente perdemos una parte importante de la información de los datos, puesto que no nos los muestra progresivamente, sino que nos da los valores acumulados al final del tiempo establecido, y muchas veces contiene información engañosa, especialmente en las estadísticas defensivas.

Por lo que, para el desarrollo temporal del partido y para conocer cierta información de equipo que piden entrenadores y clubs, no nos sirve (como por ejemplo la eficacia de los quintetos, el desarrollo del marcador o de cualquier otra variable del equipo entero durante un tiempo determinado del partido, etc.)

3.2. Play-by-play

Este tipo de recogida de información se creó para solucionar el problema que teníamos con el *Box Score*. Los datos de *Play-by-Play (PBP)* han sido la fuente principal de muchas estadísticas avanzadas, como el más-menos ajustado, que se desarrollará en este trabajo.

Play-by-Play proporciona una transcripción del juego en un formato de eventos individuales. Los datos típicos de jugada por jugada deben tener la siguiente información:

- El tiempo de la posesión
- El jugador que inició la posesión (en caso de robo o rebote defensivo)
- El jugador contrario que inició la posesión (en caso de un tiro fallado o pérdida de balón), incluida la ubicación en el piso desde donde se realizó el tiro y algunos otros identificadores únicos que usamos para clasificar el tipo de posesión.

Este tipo de análisis se inicializaron en 2007 para la Euroliga (principal competición europea de baloncesto) y en 2012 para la Eurocopa (competición internacional de segundo nivel).



Figura 3: Play by play del partido de la Euroliga de Real Madrid contra LDLC Asvel, del 17 de Marzo del 2022. Se lee de abajo hacia arriba.

3.3. Shot-charts

Este tipo de análisis es de los más visuales, ya que se realiza de una manera muy sencilla: se tiene como plantilla el dibujo de una pista de baloncesto de manera vectorial, y se va colocando cada tiro realizado en la posición del tiro, el jugador y si se encesta o no. De forma general se hace escribiendo el número del jugador en la posición desde donde se ejecuta el tiro, y si encesta, se hace un círculo alrededor del número.

El *Shot-Charts*, proporciona un output visual muy fácil de interpretar, ya que es se parece a un mapa de calor y, por lo tanto, podemos observar de una manera muy rápida desde que zonas de la pista es más efectivo el equipo y/o el jugador.

Una vez realizado, podemos obtener con facilidad el porcentaje de acierto del equipo y/o el jugador, o incluso, determinar el porcentaje de acierto por zonas.

De este análisis, es frecuente encontrar variantes: mapa de calor del equipo, mapa de porcentajes de aciertos por zonas de la pista...

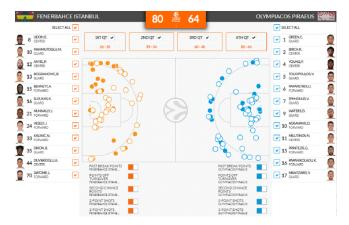


Figura 4: Shot-Chart del partido de la Euroliga del Fenerbahce Istanbul contra Olympiacos Piraeus, del 21 de Mayo del 2021.

3.4. Graphic Stats

Este tipo de análisis es más moderno y se centra en un único equipo o (más frecuentemente) en un jugador, y se visualiza sus números de manera muy visual y con mucho carácter para marketing y fáciles de compartir en redes sociales.

Básicamente, consiste en, de manera creativa y con un diseño moderno, enseñar los números que ha realizado dicho jugador. Normalmente, se visualiza puntos, rebotes y asistencias o si no también se suele mostrar el porcentaje de acierto en los diferentes tiros, conjuntamente con una o varias fotografías del jugador.

La información de este tipo de análisis viene obtenida de alguna de los anteriores tipos, por lo tanto, aunque se considera un análisis deportivo, es más bien una representación gráfica y visual de los datos destacados del jugador o del equipo.



Figura 5: Graphic Stats de Ante Zizic, jugador de Maccabi Tel Aviv en el partido del contra el FC Barcelona del 30 de Marzo de 2021.

4. Estadístico Más/Menos (Plus/Minus)

Como ya hemos comentado, el Más/Menos $(+/-,\pm)$ es una estadística deportiva utilizada para medir el impacto de un jugador en el juego. Conforme se ha ido conociendo y estudiando más este estaístico, se ha revelado que tiene fallas o que no es tan preciso, por lo que confiar únicamente en el Más/Menos es un gran error, por eso se han desarrollado y han tratado de reducir estas fallas para intentar obtener resultados más precisos con respecto al impacto de un jugador. Actualmente, se están publicando diferentes tipos de métricas sobre el Más/Menos:

4.1. Plus/Minus, PM

Este es el estadístico original que está representado por la diferencia entre la puntuación total de su equipo y la de su oponente cuando el jugador está en el juego. Se basa únicamente en la información del marcador de baloncesto tradicional; no se incluyen datos play-by-play ni datos de cuadro de puntuación no tradicionales (como mates o robos).

Fueron Los Montreal Canadiens de la NHL (hockey sobre hielo) el primer equipo en medir el más-menos de sus jugadores, comenzando en algún momento de la década de 1950. Aunque es estadístico fue pionero en el hockey, se ha utilizado en otros deportes y áreas de la vida, llegando al mundo del baloncesto por los Houston Rockets de la NBA, que lo utilizaron por primera vez una versión modificada de la estadística, que ayudó a revelar la efectividad no anunciada de Shane Battier, quien anotó muy poco. Ahora es de uso regular en toda la NBA.

Se calcula de la diguiente manera:

 $PM_i = PointsTeam - PointsOpponent$

4.2. Adjusted Plus/Minus, APM

Plus/Minus ajustado (a menudo abreviado APM) es un análisis de baloncesto que intenta predecir el impacto de un jugador individual en el margen de puntuación de un juego mediante el control del resto de los jugadores en la cancha en un momento dado. La métrica se deriva utilizando datos jugada por jugada para realizar un seguimiento de todas las acciones de finalización de sustitución y posesión. Los Dallas Mavericks lo implementaron por primera vez a principios de la década de 2000 después de que el propietario Mark Cuban encargara a los científicos de datos Jeff Sagarin, Wayne Winston y Dan Rosenbaum, quienes desarrollaron la métrica junto con su conversión WINVAL para ayudar en la determinación del salario de los jugadores.[1] En combinación con otras innovaciones, esto le dio a los Mavericks una de las oficinas centrales más progresistas de la liga en ese momento. Desde la creación de APM, se han creado varias métricas derivadas que intentan mejorar el esqueleto.

4.3. Regularized Adjusted Plus/Minus, RAPM

El Adjusted Plus/Minus convencional, aunque mejora respecto el Plus/Minus clásico, todavia hace un mal trabajo al predecir el resultado de los juegos futuros, particularmente cuando se tiene menos de una temporada de datos. Por eso, para mejorarlo se añaden regularizaciones para compensar estas deficiencias, y mejora en gran medida la precisión, y las calificaciones más-menos de algunos jugadores cambian drásticamente.

Con Regularized Adjusted Plus-Minus" (RAPM), el objetivo es proporcionar resultados más precisos mediante el empleo de una técnica especial llamada regresión de cresta" (también conocida como regularización). Reduce significativamente los errores estándar en el más-menos ajustado (APM).

La mejora con RAPM es una técnica bayesiana en la que los datos se combinan con creencias teóricas sobre rangos de datos amplios y razonables para los parámetros a fin de producir modelos más precisos. Eso es lo que hace la regresión de cresta.

RAPM es aproximadamente el doble de preciso que un APM usando regresión estándar y usando 3 años de datos, donde la ponderación de los últimos años de datos y el límite de minutos del jugador de referencia también se han optimizado cuidadosamente.

4.4. Real Plus/Minus, RPM

4.5. Otros tipos de métricas basados en el Más/Menos

- Defensive Plus/Minus aka DPM
- Net Plus/Minus aka Roland Rating
- Adjusted Plus/Minus aka APM por Wayne Winston and Jeff Sagarin
- Statistical Plus/Minus aka SPM por Dan Rosenbaum
- Regularized Adjusted Plus/Minus aka RAPM por Joe Sill
- Real Plus/Minus aka RPM por ESPN (basado en el tranajo de Steve Ilardi y Jeremias Engellmann)
- Box Plus/Minus aka BPM por Daniel Myers
- Player Tracking Plus/Minus aka PT-PM
- Estimated Plus/Minus aka EPM por Taylor Snarr
- Player Impact Plus/Minus aka PIPM por Jacob Goldstein
- CARMELO, por FiveThirtyEight
- RAPTOR, por FiveThirtyEight
- LEBRON, por Tim Cranjis, Krishna Narsu
- Daily Plus/Minus aka DPM o DARKO por Kostya Medvedovsky
- Individual Player Value aka IPV por Talking Practice Blog
- Augmented Box Plus/Minus aka AuPM por Ben Taylor
- Daily Updated Ranking of Individual Performance aka DRIP por Nathan Walker

(explicar cada concepto y su manera de calcularlo o en base a que se calcula)

Tabla de diferencias ?

5. Análisis y métodos

5.1. Descripción de la base de datos

```
> source(here("R", "fix-lineups.R"))
> ## Function fix_lineups() only takes data from a single game,
> ## so I split the data and apply the function to each splitted data frame.
> pbp_2008 <- read_csv("pbp2008.csv")</pre>
> pbp_2008_fixed <- split(pbp_2008, factor(pbp_2008$game_code)) %>%
  map_df(fix_lineups)
> pbp_2009 <- read_csv("pbp2009.csv")</pre>
> pbp_2009_fixed <- split(pbp_2009, factor(pbp_2009$game_code)) %>%
   map_df(fix_lineups)
> pbp_2010 <- read_csv("pbp2010.csv")</pre>
> pbp_2010_fixed <- split(pbp_2010, factor(pbp_2010$game_code)) %>%
   map_df(fix_lineups)
> pbp_2011 <- read_csv("pbp2011.csv")</pre>
> pbp_2011_fixed <- split(pbp_2011, factor(pbp_2011$game_code)) %>%
+ map_df(fix_lineups)
> pbp_2012 <- read_csv("pbp2012.csv")</pre>
> pbp_2012_fixed <- split(pbp_2012, factor(pbp_2012$game_code)) %>%
+ map_df(fix_lineups)
> pbp_2013 \leftarrow read_csv("pbp2013.csv")
> pbp_2013_fixed <- split(pbp_2013, factor(pbp_2013$game_code)) %>%
  map_df(fix_lineups)
> pbp_2014 <- read_csv("pbp2014.csv")</pre>
> pbp_2014_fixed <- split(pbp_2014, factor(pbp_2014$game_code)) %>%
   map_df(fix_lineups)
> pbp_2015 <- read_csv("pbp2015.csv")
> pbp_2015_fixed <- split(pbp_2015, factor(pbp_2015$game_code)) %>%
+ map_df(fix_lineups)
> pbp_2017 <- read_csv("pbp2017.csv")</pre>
> pbp_2017_fixed <- split(pbp_2017, factor(pbp_2017$game_code)) %>%
  map_df(fix_lineups)
> pbp_2018 <- read_csv("pbp2018.csv")
> pbp_2018_fixed <- split(pbp_2018, factor(pbp_2018$game_code)) %>%
  map_df(fix_lineups)
> pbp <- rbind(pbp_2011_fixed, pbp_2012_fixed, pbp_2013_fixed, pbp_2014_fixed, pbp_2015_fixed, pbp_2017
> data_fixed <- pbp_2011_fixed</pre>
> data_subset <- as.data.frame(data_fixed) %>%
   select(contains(c("home_player", "away_player"))) %>%
   arrange(home_player1)
> WRONG <- c("ABRINES, ALEJANDRO", "ACIK, TUTKU", "AGRAVANIS, DIMITRIS", "ALFRED", "AMAGOU, PAPE", "AND
> RIGHT <- c("ABRINES, ALEX", "AÇIK, TUTKU", "AGRAVANIS, DIMITRIOS", "LUCAS, JAMON ALFRED", "AMAGOU, PA
> players_with_errors <- data.frame("wrong"=WRONG,
                                     "right"=RIGHT)
> data_subset_head <- head(data_subset)</pre>
> #Numero de errores
> source(here("R", "errors_fun.R"))
> ERRORS_totals <- errors_fun(data_subset, players_with_errors)
> #Limpieza errores
>
> #errors_fun(data_clean, players_with_errors) > 0
```

```
> source(here("R", "dummys_fun.R"))
> df_dummys_2011 <- dummys_fun(pbp_2011_fixed)</pre>
> df_dummys_2012 <- dummys_fun(pbp_2012_fixed)
> df_dummys_2013 <- dummys_fun(pbp_2013_fixed)
> df_dummys_2014 <- dummys_fun(pbp_2014_fixed)</pre>
> df_dummys_2015 \leftarrow dummys_fun(pbp_2015_fixed)
> df_dummys_2017 <- dummys_fun(pbp_2017_fixed)</pre>
> df_dummys_2018 <- dummys_fun(pbp_2018_fixed)</pre>
> saveRDS(df_dummys_2011, "C:\\Users\\paula\\Documents\\GitHub\\TFG-Baloncesto2\\Dummys\\dummys_2011.rds
> saveRDS(df_dummys_2012, "C:\\Users\\paula\\Documents\\GitHub\\TFG-Baloncesto2\\Dummys\\dummys_2012.rds
> saveRDS(df_dummys_2013, "C:\Vsers\\rho unla\Documents\GitHub\TFG-Baloncesto2\Dummys\dummys_2013.rds) \\
> saveRDS(df_dummys_2014, "C:\\Users\\paula\\Documents\\GitHub\\TFG-Baloncesto2\\Dummys\\dummys_2014.rds
> saveRDS(df_dummys_2015, "C:\\Users\\paula\\Documents\\GitHub\\TFG-Baloncesto2\\Dummys\\dummys_2015.rds
> saveRDS(df_dummys_2017, "C:\Vsers\paula\Documents\GitHub\TFG-Baloncesto2\Dummys\dummys_2017.rds) \\
> saveRDS(df_dummys_2018, "C:\\Users\\paula\\Documents\\GitHub\\TFG-Baloncesto2\\Dummys\\dummys_2018.rds
> saveRDS(pbp, "C:\\Users\\paula\\Documents\\GitHub\\TFG-Baloncesto2\\Dummys\\pbp.rds")
> saveRDS(pbp_2011_fixed, "C:\\Users\\paula\\Documents\\GitHub\\TFG-Baloncesto2\\Dummys\\pbp_2011_fixed.
> dummys_2011 <- readRDS("Dummys/dummys_2011.rds")</pre>
> dummys_2012 <- readRDS("Dummys/dummys_2012.rds")</pre>
> dummys_2013 <- readRDS("Dummys/dummys_2013.rds")</pre>
> dummys_2014 <- readRDS("Dummys/dummys_2014.rds")</pre>
> dummys_2015 <- readRDS("Dummys/dummys_2015.rds")</pre>
> dummys_2017 <- readRDS("Dummys/dummys_2017.rds")</pre>
> dummys_2018 <- readRDS("Dummys/dummys_2018.rds")</pre>
> pbp <- readRDS("Dummys/pbp.rds")</pre>
> pbp_2011_fixed <- readRDS("Dummys/pbp_2011_fixed.rds")</pre>
> N_rows <- nrow(dummys_2011)+nrow(dummys_2012)+nrow(dummys_2013)+nrow(dummys_2014)+nrow(dummys_2015)+n
> players_2011 <- colnames(dummys_2011)[-c(1:5)]</pre>
> players_2012 <- colnames(dummys_2012)[-c(1:5)]</pre>
> players_2013 <- colnames(dummys_2013)[-c(1:5)]</pre>
> players_2014 <- colnames(dummys_2014)[-c(1:5)]</pre>
> players_2015 <- colnames(dummys_2015)[-c(1:5)]
> players_2017 <- colnames(dummys_2017)[-c(1:5)]</pre>
> players_2018 <- colnames(dummys_2018)[-c(1:5)]
> players <- data.frame("player" = c(players_2011, players_2012, players_2013, players_2014, players_20
    arrange(player) %>%
    unique()
> players_vec <- players$player
> num_players <- length(players_vec)</pre>
```

En este estudio vamos a analizar los datos de todas las jugadas registradas en partidos oficiales de varias temporadas, empezando en 2011 hasta 2018 (con la excepción de la temporada de 2016 que no hemos podido obtener los datos) de la Liga Nacional de ACB para buscar el Más/Menos Ajustado. Tenemos un total de N=79872 jugadas, y 1200 jugadores en 7 temporadas.

De las temporadas se ha recogido el play-by-play de cada partido de cada temporada, una forma de recoger los datos de manera bastante completa: tenemos la temporada del partido, el código de identificación del partido, el id de la acción, el equipo y el jugador que la realiza, el tipo de acción, el momento exacto del partido en que ocurre (en segundos), el cuarto del partido, el marcador, los jugadores que estan en pista en ese momento... entre otras variables que para este estudio no nos son de interés.

En la siguiente página teneis una muestra de como visualizamos los datos originales.

```
> pbp_subset <- pbp %>%
+ select(season, game_code, play_number, team_code, player_name, quarter, points_home, points_away, s
> summary(pbp_subset)

season game_code play_number team_code
Min. :2011 Min. : 1.0 Min. : 2.0 Length:873698
```

1st Qu.:2012 1st Qu.: 61.0 1st Qu.:143.0 Class :character Median: 2014 Median: 123.0 Median: 282.0 Mode: character Mean :2014 Mean :123.9 Mean :283.4 3rd Qu.:420.0 3rd Qu.:2017 3rd Qu.:184.0 Max. :2018 Max. :260.0 Max. :768.0 player_name quarter points_home points_away Length:873698 Min. :1.000 Min. : 0.00 Min. : 0.00 Class :character 1st Qu.:2.000 1st Qu.: 15.00 1st Qu.: 15.00 Mode :character Median : 3.000 Median : 30.00 Median : 29.00 Mean :2.565 Mean : 33.16 Mean : 31.35 3rd Qu.:4.000 3rd Qu.: 48.00 3rd Qu.: 45.00 Max. :6.000 Max. :116.00 Max. :111.00 home_player1 Length:873698 home_player3 home_player2 seconds Min. : 0 Length:873698 Length:873698 Length:873698 1st Qu.: 662 Class :character Median :1240 Mode :character Mode :character Mode :character Mean :1254 3rd Qu.:1848 Max. :3000 home_player4 home_player5 away_player1 away_player2 Length:873698 Length:873698 Length:873698 Length:873698 Class :character Class : character Class : character Class : character Mode :character Mode :character Mode :character Mode :character

away_player3 away_player4 away_player5
Length:873698 Length:873698 Length:873698
Class:character Class:character Class:character
Mode:character Mode:character Mode:character

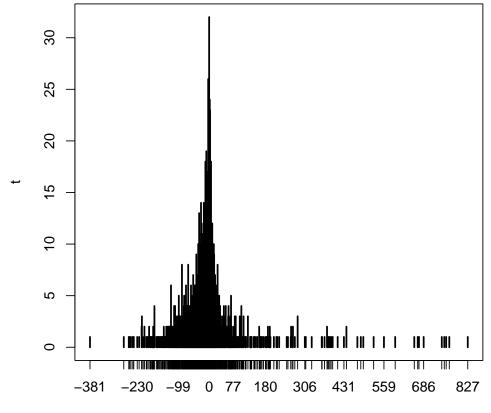
> library(xtable)
> head_pbp <- head(pbp_2011_fixed, n=20)[c(1:5, 8:10, 12:14, 19:28)]
> print(xtable(head_pbp))

	season	game_code	play_number	team_code	player_name	quarter	points_home	points_away	$_{ m seconds}$	home_team	$away_team$	home_player
П	2011.00	1.00	2.00	NAN	AKINGBALA, AKIN	1.00	0.00	2.00	22.00	Bennet Cantu	SLUC Nancy	MICOV, V
2	2011.00	1.00	3.00	NAN	BATUM, NICOLAS	1.00	0.00	2.00	31.00	Bennet Cantu	SLUC Nancy	MICOV, VI
က	2011.00	1.00	4.00	NAN	AKINGBALA, AKIN	1.00	0.00	2.00	37.00	Bennet Cantu	SLUC Nancy	MICOV, VI
4	2011.00	1.00	5.00	$\operatorname{CL}\Omega$	MARCONATO, DENIS	1.00	0.00	2.00	37.00	Bennet Cantu	SLUC Nancy	MICOV, VI
ಬ	2011.00	1.00	0.00	CLO		1.00	0.00	2.00	52.00	Bennet Cantu	SLUC Nancy	MICOV, VI
9	2011.00	1.00	8.00	NAN	AKINGBALA, AKIN	1.00	0.00	2.00	00.99	Bennet Cantu	SLUC Nancy	MICOV, VI
7	2011.00	1.00	9.00	NAN	BATUM, NICOLAS	1.00	0.00	2.00	69.00	Bennet Cantu	SLUC Nancy	MICOV, VI
∞	2011.00	1.00	10.00	CLO	MICOV, VLADIMIR	1.00	0.00	2.00	82.00	Bennet Cantu	SLUC Nancy	MICOV, VI
6	2011.00	1.00	11.00	CLO	LEUNEN, MAARTEN	1.00	0.00	2.00	84.00	Bennet Cantu	SLUC Nancy	MICOV, VI
10	2011.00	1.00	12.00	CLO	MARCONATO, DENIS	1.00	0.00	2.00	96.00	Bennet Cantu	SLUC Nancy	MICOV, VI
11	2011.00	1.00	13.00	NAN	AKINGBALA, AKIN	1.00	0.00	2.00	00.96	Bennet Cantu	SLUC Nancy	MICOV, VI
12	2011.00	1.00	14.00	NAN	SHULER, JAMAL	1.00	0.00	4.00	111.00	Bennet Cantu	SLUC Nancy	MICOV, VI
13	2011.00	1.00	15.00	$\operatorname{CL}\Omega$	MICOV, VLADIMIR	1.00	0.00	4.00	117.00	Bennet Cantu	SLUC Nancy	MICOV, VI
14	2011.00	1.00	16.00	NAN	LINEHAN, JOHN	1.00	0.00	4.00	117.00	Bennet Cantu	SLUC Nancy	MICOV, VI
15	2011.00	1.00	17.00	$\operatorname{CL}\Omega$	MICOV, VLADIMIR	1.00	0.00	4.00	117.00	Bennet Cantu	SLUC Nancy	MICOV, VI
16	2011.00	1.00	18.00	NAN	LINEHAN, JOHN	1.00	0.00	4.00	117.00	Bennet Cantu	SLUC Nancy	MICOV, VI
17	2011.00	1.00	20.00	NAN	SHULER, JAMAL	1.00	0.00	4.00	132.00	Bennet Cantu	SLUC Nancy	MICOV, VI
18	2011.00	1.00	21.00	NAN	SAMNICK, VICTOR	1.00	0.00	4.00	133.00	Bennet Cantu	SLUC Nancy	MICOV, VI
19	2011.00	1.00	22.00	CLO	MARCONATO, DENIS	1.00	0.00	4.00	135.00	Bennet Cantu	SLUC Nancy	MICOV, VI
20	2011.00	1.00	23.00	NAN	SHULER, JAMAL	1.00	0.00	4.00	135.00	Bennet Cantu	SLUC Nancy	MICOV, VI

```
> df_dummys <- dplyr::bind_rows(dummys_2011, dummys_2012, dummys_2013, dummys_2014, dummys_2015, dummys
+ mutate_all(~replace(., is.na(.), 0))
> df_dummys<- df_dummys[df_dummys$stint_seconds != 0, ]</pre>
> df_dummys_PlusMinus <- df_dummys %>%
   group_by(SeasonGame, quarter) %>%
   mutate(across(matches(players_vec), function(x) x*PlusMinus)) %>%
   select(-c(PlusMinus)) %>%
    ungroup()
> ### MasMenos de los mismos lineups (sin tener en cuenta SeasonGame o Quarter):
> df_dummys_PlusMinus_2 <- df_dummys_PlusMinus %>% select(-c(SeasonGame, quarter))
> source(here("R", "Print_MS_fun.R"))
> #combinar mismos stints
> PlusMinus_Lineups <- df_dummys_PlusMinus_2 %>%
                       mutate(stint_seconds = Print_MS(stint_seconds))
> PlusMinus_Classic <- colSums(PlusMinus_Lineups[5:ncol(PlusMinus_Lineups)])</pre>
> PlusMinus_Classic_df <- as.data.frame(PlusMinus_Classic) %>%
    tibble::rownames_to_column(var="Players") %>%
    arrange(desc(PlusMinus_Classic)) %>%
    select(Players, PlusMinus_Classic)
```

> print(xtable(head(PlusMinus_Classic_df, n=10)))

	Players	PlusMinus_Classic
1	RODRIGUEZ, SERGIO	827.00
2	TEODOSIC, MILOS	768.00
3	HINES, KYLE	758.00
4	DE COLO, NANDO	752.00
5	FERNANDEZ, RUDY	744.00
6	REYES, FELIPE	686.00
7	VORONTSEVICH, ANDREY	671.00
8	LLULL, SERGIO	667.00
9	CARROLL, JAYCEE	656.00
10	SLOUKAS, KOSTAS	595.00



PlusMinus_Classic

> head(coef1, n=50)

	Estimate	Std. Error	t value	Pr(> t)
CASELLA, ANDREA	-9.762730	2.6857070	-3.6350690	0.0002781593
UTA, CHISTIAN	-6.670958	2.1248160	-3.1395462	0.0016929917
`MOURATOS, VASSILIS`	-5.272638	2.7279701	-1.9328064	0.0532651733
ZAMANSKIY, IGOR	-4.649898	1.9758771	-2.3533334	0.0186094435
CHARALAMPOPOULOS, V.	-4.376151	1.5488686	-2.8253859	0.0047240691
`PISKIN, SUACAN`	-4.124322	1.7189459	-2.3993319	0.0164283212
FORTE, JOSEPH	-3.765323	2.0736742	-1.8157735	0.0694105582
BAYRAK, ERCAN	-3.695550	1.7638861	-2.0951184	0.0361650269
MARTINEZ, CARLOS	-3.438398	1.9146417	-1.7958441	0.0725247973
YURTSEVEN, OMER FARUK	-3.035248	1.8956287	-1.6011827	0.1093421960
`DAMBRAUSKAS, PAULIUS`	-2.984252	1.9854848	-1.5030345	0.1328358760
SAMB, MAMADOU	-2.878735	2.6635074	-1.0808062	0.2797880602
`SIRIN, OSMAN`	-2.785542	1.2305492	-2.2636574	0.0235990729
MORAY, YANNICK	-2.589865	2.1012160	-1.2325551	0.2177470528
`CONDE, LUIS`	-2.558610	1.4093735	-1.8154237	0.0694642661
`KANYGIN, IGOR`	-2.528429	1.2689061	-1.9926053	0.0463096192
`BALASHOV, NIKITA`	-2.190066	0.9240524	-2.3700668	0.0177882990
`HALILOVIC, MIRALEM`	-2.110729	1.2809626	-1.6477679	0.0994060204
`ADJAGBA, ULYSSE`	-2.045948	1.4339014	-1.4268404	0.1536315768
`DELFINO, CARLOS`	-1.945011	0.8170069	-2.3806539	0.0172853208
SARAKAUSKAS, VYTAUTAS	-1.909972	2.9093356	-0.6564976	0.5115067803
MARTINEZ, IVAN	-1.843183	1.2876231	-1.4314617	0.1523036215
`HARARI, ARAD`	-1.797106			0.2890828067
MAGRO, DANIELE	-1.758020	0.9787938	-1.7961092	0.0724826358

`PEREZ, PABLO` -1.637536 3.7522446 -0.4364149 0.6625374214 LAZAREV, IVAN -1.612350 0.5119991 -3.1491270 0.0016384657 TAROLIS, DONATAS -1.587162 0.9665390 -1.6421085 0.1005732321 -1.569293 1.4036679 -1.1179945 0.2635742416 `KANTE, LAMINE` -1.537271 1.6803381 -0.9148582 0.3602700803 `SIKORA, SLAWOMIR` KURUCS, RODIONS `KHOMENKO, ALEXANDER` -1.365021 4.6220388 -0.2953287 0.7677438845 PEREZ, JOSEP -1.353116 1.0822469 -1.2502838 0.2112012001 -1.333621 1.5468302 -0.8621638 0.3886011368 GANKEVICH, ALEXANDER 'HAKANSON, LUDDE' -1.313231 0.8388095 -1.5655887 0.1174506926 `DRAGIC, GORAN` -1.289523 0.5114805 -2.5211574 0.0116997519 CRAFT, AARON NGUIRANE, MAODO -1.219234 0.6550520 -1.8612779 0.0627102666 VANUZZO, MANUEL `HERNANGOMEZ, GUILLERMO` -1.203979 0.7556713 -1.5932572 0.1111082570 DIOP, MAMADOU -1.196628 1.2440991 -0.9618430 0.3361326960 HERNANDEZ `SANLI, SERTAC` -1.143645 0.5051299 -2.2640615 0.0235742085 -1.137392 0.7122038 -1.5970038 0.1102705958 NTILIKINA, FRANK PACHULIA, ZAZA -1.129146 0.4523522 -2.4961650 0.0125573195 BENSON, KEITH -1.096169 0.7805206 -1.4044071 0.1602032944 -1.078846 0.4870537 -2.2150449 0.0267610598 `VARDA, RATKO` -1.075344 0.4996637 -2.1521364 0.0313909469 GREGORY, KENNY MAYR, DANIEL -1.053958 0.9187801 -1.1471271 0.2513340697 SCHEYER, JONATHAN -1.052088 0.4725759 -2.2262835 0.0259992292

> head(coef2, n=50)

Estimate Std. Error t value Pr(>|t|) MORAY, YANNICK -0.26016551 0.17273779 -1.5061297 0.132039620 ANIKIIENKO, KOSTIANTYN -0.19722743 0.10527833 -1.8733906 0.061019767 -0.17407088 0.10530600 -1.6530006 0.098336435 `HALILOVIC, MIRALEM` `VIGGIANO, JEFF` -0.16610527 0.05995422 -2.7705351 0.005598303 -0.14214304 0.22078791 -0.6437990 0.519708439 CASELLA, ANDREA -0.13743241 0.13935567 -0.9861989 0.324039852 `HARARI, ARAD` -0.13526883 0.04124628 -3.2795398 0.001040411 RANNIKKO, TEEMU `MBODJ, CHEIKH` -0.11906338 0.06543395 -1.8195964 0.068825926 -0.11854491 0.05192838 -2.2828540 0.022442767 RUPNIK, LUKA LUCAS, JAMON ALFRED -0.11728248 0.05446652 -2.1532949 0.031299840 `THALIEN, KEVIN` -0.11027472 0.22370516 -0.4929467 0.622052207 `PINDA, KINGSLEY` -0.10792399 0.13486992 -0.8002079 0.423593779 `DIA, PAPA` -0.10485486 0.06323266 -1.6582388 0.097274968 $\hbox{-0.10340410 0.05041962 -2.0508700 0.040284332}$ `AKPINAR, ISMET` `MOCNIK, JAN` -0.08736160 0.03477096 -2.5124878 0.011991161 AMATO, ANDREA -0.08425539 0.15506695 -0.5433485 0.586892078 GIRDZIUNAS, MINDAUGAS -0.08341743 0.06088274 -1.3701327 0.170651069 IVLEV, VLADIMIR -0.07979035 0.03163966 -2.5218462 0.011676872 `SAMB, MAMADOU` -0.07920208 0.21896291 -0.3617146 0.717566718 `KILICLI, DENIZ` -0.07794003 0.02860856 -2.7243602 0.006444654'UTA, CHISTIAN' -0.07729615 0.17467791 -0.4425067 0.658124287 ZOZULIN, ALEKSEY -0.07502694 0.02859940 -2.6233744 0.008708755 -0.07499093 0.04942372 -1.5173065 0.129195059 WATERS, DOMINIC BOONE, JASON -0.06947053 0.04527629 -1.5343690 0.124944653 FISCHER, LUCAS -0.06810279 0.06202402 -1.0980067 0.272206378 KASTRATI, ERJON -0.06745255 0.04503958 -1.4976285 0.134235531PAVLOVIC, ALEKSANDAR -0.06710476 0.04977352 -1.3482021 0.177598947 PISKIN, SUACAN -0.06702275 0.14131194 -0.4742894 0.635295441 `DIMSA, EGIDIJUS` -0.06684544 0.05997676 -1.1145224 0.265060009 SCHEYER, JONATHAN -0.06387537 0.03884975 -1.6441642 0.100148007 NELSON, SPENCER -0.06366748 0.03603614 -1.7667674 0.077272740

ROSZYK, KRZYSZTOF -0.06333301 0.04607497 -1.3745645 0.169272095 VANUZZO, MANUEL -0.06264341 0.05385083 -1.1632765 0.244722385 `ANDUSIC, DANILO` -0.06262365 0.02833797 -2.2098844 0.027117285 -0.06191744 0.06416542 -0.9649659 0.334566132 `BENSON, KEITH` `MOURATOS, VASSILIS` -0.05828806 0.22426230 -0.2599102 0.794934018 -0.05758832 0.03107211 -1.8533765 0.063833755 `IBAKA, SERGE` CHARALAMPOPOULOS, V. -0.05549014 0.12733014 -0.4357974 0.662985436 `DIASSE, PAPA` -0.05222136 0.04626591 -1.1287222 0.259019979 -0.05180449 0.13813805 -0.3750197 0.707647256 SIKORA, SLAWOMIR DAMBRAUSKAS, PAULIUS -0.04984464 0.16322370 -0.3053762 0.760080765 -0.04930238 0.06143113 -0.8025634 0.422230571 SHAKUR, MUSTAFA `HEYTVELT, JOSH` -0.04927676 0.03330363 -1.4796214 0.138980046 `TAROLIS, DONATAS` -0.04853031 0.07945771 -0.6107691 0.541354995 -0.04852379 0.04930330 -0.9841897 0.325026588`CAPELA, CLINT` `SARAKAUSKAS, VYTAUTAS` -0.04702149 0.23917207 -0.1966011 0.844140434 EZE, BENJAMIN -0.04605558 0.02594380 -1.7752053 0.075869516 `KASUN, MARIO` -0.04453040 0.02734671 -1.6283641 0.103453420 `NIELSEN, MATT` -0.04448529 0.02791176 -1.5937829 0.110990419 `FORTE, JOSEPH` -0.04358043 0.17047362 -0.2556432 0.798227291

> head(coef3, n=50)

Estimate Std. Error t value MORAY, YANNICK -0.26454960 0.17265132 -1.5322767 0.125459906 ANIKIIENKO, KOSTIANTYN -0.19185220 0.10522742 -1.8232149 0.068276277 `HALILOVIC, MIRALEM` -0.17306476 0.10525278 -1.6442774 0.100124625 `VIGGIANO, JEFF` -0.16607244 0.05992414 -2.7713779 0.005583834CASELLA, ANDREA -0.14354780 0.22067623 -0.6504905 0.515378121 RANNIKKO, TEEMU -0.13558069 0.04122543 -3.2887636 0.001006916 HARARI, ARAD -0.12956062 0.13928899 -0.9301570 0.352293889 RUPNIK, LUKA -0.11948019 0.05190226 -2.3020227 0.021337609 LUCAS, JAMON ALFRED -0.11648664 0.05443905 -2.1397624 0.032378359 `THALIEN, KEVIN` -0.11626096 0.22359356 -0.5199656 0.603089633 `MBODJ, CHEIKH` -0.11459077 0.06540348 -1.7520592 0.079769199 -0.10573552 0.13480636 -0.7843511 0.432837485 PINDA, KINGSLEY AKPINAR, ISMET -0.10422103 0.05039422 -2.0681147 0.038633895 -0.10316193 0.06320107 -1.6322813 0.102625965 `DIA, PAPA` MOCNIK, JAN -0.08927213 0.03475427 -2.5686667 0.010211672 -0.08907068 0.21885596 -0.4069831 0.684021966 SAMB, MAMADOU -0.08218647 0.06085214 -1.3505930 0.176831394 GIRDZIUNAS, MINDAUGAS `IVLEV, VLADIMIR` -0.07957679 0.03162366 -2.5163691 0.011859914 `AMATO, ANDREA` -0.07944342 0.15498976 -0.5125721 0.608252745 `UTA, CHISTIAN` $\hbox{-0.07804971 0.17458952 -0.4470469 0.654842948}$ -0.07777896 0.02859409 -2.7201060 0.006528147 `KILICLI, DENIZ` ZOZULIN, ALEKSEY -0.07618580 0.02858533 -2.6652060 0.007696363 WATERS, DOMINIC -0.07579166 0.04939884 -1.5342803 0.124966471 PISKIN, SUACAN -0.07201962 0.14124195 -0.5099025 0.610121816 `BOONE, JASON` -0.07035325 0.04525352 -1.5546469 0.120035932 DIMSA, EGIDIJUS -0.06877234 0.05994694 -1.1472201 0.251295658 KASTRATI, ERJON -0.06800060 0.04501684 -1.5105593 0.130906501 PAVLOVIC, ALEKSANDAR -0.06762704 0.04974837 -1.3593820 0.174031137 `FISCHER, LUCAS` -0.06748104 0.06199268 -1.0885323 0.276364953 -0.06510603 0.05382455 -1.2095973 0.226438668 VANUZZO, MANUEL SCHEYER, JONATHAN -0.06507700 0.03883045 -1.6759269 0.093758183 NELSON, SPENCER -0.06491582 0.03601923 -1.8022543 0.071510873 `BENSON, KEITH` -0.06445479 0.06413381 -1.0050049 0.314898892 `ANDUSIC, DANILO` $\hbox{-0.06362995 0.02832394 -2.2465076 0.024675458}$ MOURATOS, VASSILIS -0.06275394 0.22414955 -0.2799646 0.779505717 ROSZYK, KRZYSZTOF -0.06243654 0.04605184 -1.3557883 0.175172088DAMBRAUSKAS, PAULIUS -0.05951901 0.16314605 -0.3648204 0.715246851CHARALAMPOPOULOS, V. -0.05950423 0.12726679 -0.4675551 0.640104664

```
`SIKORA, SLAWOMIR` -0.05837906 0.13807095 -0.4228193 0.672428753
                          -0.05822308 0.03105650 -1.8747468 0.060832853
-0.05481745 0.04624377 -1.1854019 0.235863696
`IBAKA, SERGE`
`DIASSE, PAPA`
                          -0.05242285 0.10112258 -0.5184090 0.604175029
-0.05237541 0.17039127 -0.3073832 0.758552892
`SIRIN, OSMAN`
`FORTE, JOSEPH`
`TAROLIS, DONATAS`
`HEYTVELT JOSH`
                            -0.05068797 0.07941800 -0.6382428 0.523318279
HEYTVELT, JOSH
                            -0.05028246 0.03328707 -1.5105702 0.130903732
`SARAKAUSKAS, VYTAUTAS` -0.04890028 0.23905114 -0.2045599 0.837916753
SHAKUR, MUSTAFA -0.04876591 0.06140007 -0.7942321 0.427063750
                          -0.04690956 0.04927880 -0.9519218 0.341140810
-0.04616326 0.02593220 -1.7801522 0.075056579
CAPELA, CLINT
`EZE, BENJAMIN`
`KASUN, MARIO`
                           -0.04597409 0.02733375 -1.6819533 0.092583571
```

(añadir texto del script "modelosdatos")

RIDGE

```
> match_firstplayer <- 6
> n <- dim(df_dummys)[2]
> x <- model.matrix(~.-1, data=df_dummys_PlusMinus[,match_firstplayer:n])
> y <- df_dummys$PlusMinus

> library(glmnet)
> grid <- 10^seq(10, -2, length=100)
> ridge.mod <- glmnet(x, y, alpha=0, lambda=grid)
> dim(coef(ridge.mod))
> sqrt(sum(coef(ridge.mod)[-1, 60]^2))
> predict(ridge.mod, s=50, type="coefficients")[1:20, ]
```

Referencias

- [1] Joseph Sill, Improved NBA Adjusted +/- Using Regularization and Out-of-Sample Testing, PDF, 6 Marzo 2010.
- $[2] \ \ Happy \ Git, \ \textit{Let's Git started}, \ url: \ \texttt{https://happygitwithr.com/index.html}.$
- [3] NBASTUFFER.COM, Regularized Adjusted Plus Minus (RAPM), url: https://www.nbastuffer.com/analytics101/regularized-adjusted-plus-minus-rapm/.

6. Annexo

6.1. Glosario

Estos son los conceptos técnicos que se mencionan en este informe:

- Box score: Resumen estructurado de la información individual de cada jugador que participó en el juego, en formato de tabla.
- Play-by-Play: Relato detallado de las acciones en un partido de baloncesto. Solo describe las acciones valiosas, por lo que no se describen los pases ni los *dribblings*.
- Tiempo extra o prórroga: los partidos de baloncesto deben tener un ganador. Si, después del tiempo reglamentario (4 cuartos de 10 minutos cada uno), los dos equipos están empatados, se otorga un tiempo extra (cuarto adicional de 5 minutos). Si es necesario, se otorgan horas extra adicionales hasta que haya un ganador.

Tipos de jugadas descritas en el play-by-play:

- Puntos: Sistema de puntuación en baloncesto. El equipo que obtiene la mayor cantidad de puntos gana el juego.
- Tiro: Intento de canasta. En caso de que entre un intento, el equipo puede obtener uno, dos o tres puntos.
- Tiro libre: intento de 1 punto. El jugador dispara sin oposición desde detrás de una línea que está a 4,6m de la canasta. Este tiro se otorga tras determinadas situaciones ilegales.
- Tiro de campo: Tiro realizado (encestado) durante una jugada.

Intento de 2 puntos: cualquier tiro realizado dentro de la línia de triple.

Intento de 3 puntos: También llamado triple. El jugador tira detrás de un arco de 6,75 metros lejos de la canasta. Esta distancia es la máxima, ya que el arco se convierte en una línea paralela a las líneas laterales cuando se acerca a ellas. Esto deja algo de espacio entre la línea de 3 puntos y la línea lateral. Esta modificación hace que la distancia se reduzca 6,6 metros.

• Rebote: Recogida del balón tras un tiro fallado.

Rebote defensivo: El jugador que consigue el rebote es del equipo que estaba defendiendo.

Rebote ofensivo: El jugador que consigue el rebote es del equipo que intentó anotar.

- Asistencia: Pase que provoca un tiro de campo anotado. Si el jugador X pasa el balón a Y y este tarda poco tiempo en anotar, X recibe una asistencia.
- Robo: Cuando un jugador que defiende consigue quitar la pelota a un jugador que está atacando.
- Pérdida de balón: Pérdida de la posesión por parte de un equipo antes de ejecutar un tiro. Puede deberse a una infracción, al robo del balón, a la comisión de una falta ofensiva, etc.
- Tapón: Desvío por parte de un jugador de un tiro del otro equipo con la intención de impedir que anoté
- Falta: Infracción de las reglas más grave que una violación. La acumulación de faltas por parte de un jugador puede implicar la descalificación.

Falta personal: Falta por contacto personal ilegal con un adversario. Si la falta es antideportiva (excesiva o injustificada), se sanciona con falta flagrante.

Falta técnica: Falta por actitud antideportiva.

 Tiempo muerto: Pausa a la mitad de un cuarto para permitir que los entrenadores den indicaciones a los jugadores.

Métricas en el Box-Score:

■ +/- (más-menos): fórmula estadística para medir el impacto del jugador en un juego. Da cuenta de los puntos que se hicieron mientras el jugador estaba en la cancha.

- Performance Index Rating (PIR): Fórmula estadística utilizada para medir la eficiencia de un jugador en un juego.
- Eficacia (EFF): Fórmula estadística utilizada para evaluar el rendimiento de un jugador a lo largo de una temporada.

6.2. ¿Qué es GitHub?

GitHub es una plataforma de alojamiento, propiedad de Microsoft, que ofrece a los desarrolladores la posibilidad de crear repositorios de código y guardarlos en la nube de forma segura, usando un sistema de control de versiones, llamado Git.

Como he comentado, facilita la organización de proyectos y permite la colaboración de varios desarrolladores en tiempo real. Es decir, nos permite centralizar el contenido del repositorio para poder colaborar con los otros miembros de nuestro grupo des de varios dispositivos.



GitHub está basada en el sistema de control de versiones distribuidas de Git, por lo que se puede contar con sus funciones y herramientas, aunque GitHub ofrece varias opciones adicionales y su interfaz es mucho más fácil de manejar, por lo que no es absolutamente necesario que las personas que lo utilizan tengan un gran conocimiento técnico.

6.2.1. Ventajas

Hay un gran número de razones por las que GitHub es una gran opción para el control y gestión de proyectos de código. Como por ejemplo:

- GitHub permite que alojemos proyectos en repositorios de forma gratuita
- Los repositorios son públicos por defecto. Sin embargo, GitHub te permite también alojar tus proyectos de manera privada
- Puedes crear y compartir páginas web estáticas con GitHub Pages
- Facilita compartir tus proyectos de una forma mucho más fácil y crear un portafolio
- Te permite colaborar para mejorar los proyectos de otros y a otros mejorar o aportar a los tuyos
- Ayuda reducir significativamente los errores humanos y escribir tu código más rápido con GitHub Copilot
- Te da control de versiones, una herramienta muy útil.

6.2.2. ¿Qué es el control de versiones?

Se le llama control de versiones a la administración de los cambios que se realizan sobre los elementos o la configuración de algún proyecto. En otras palabras, el control de versiones sirve para conocer y autorizar los cambios que hagan los colaboradores en tu proyecto, guardando información extra de qué están, incluyendo los cambios y cuándo se hicieron. Este control comienza con una versión básica del documento y luego va guardando los cambios que se hagan a lo largo del proyecto.

El control de versiones es una herramienta muy valiosa, pues con ella puedes tener acceso a las versiones anteriores de tu proyecto si es que en algún momento no llega a funcionar de forma correcta.

6.2.3. ¿Qué es Git?

Git es un software de control de versiones diseñado por Linus Torvalds, pensando en la eficiencia, la confiabilidad y compatibilidad del mantenimiento de versiones de aplicaciones cuando estas tienen un gran número de archivos de código fuente.

Diferencias Git vs GitHub

Entonces, ¿qué diferencia a Git de GitHub? La principal diferencia es que Git es un sistema que permite establecer un control de versiones, mientras que GitHub es una plataforma que ofrece un grupo de funciones que facilitan el uso de Git y la colaboración en tiempo real, así como el almacenamiento en la nube.

6.3. Descripción de las variables

- \bullet MIN (Minutes): Minutos totales jugados
- PTS (*Points*): Puntos totales realizados
- 2FGA (2-point Field Goals Attempted): Número de tiros de 2 puntos intentadas
- 2FGM (2-point Field Goals Made): Número de tiros de 2 puntos anotadas
- 3FGA (3-point Field Goals Attempted): Número de tiros de 3 puntos ("triples") intentadas
- 3FGM (3-point Field Goals Made): Número de tiros de 3 puntos ("triples") anotadas
- FTA (Free Throws Attempted): Número de tiros libres intentados
- FTM (Free Throws Made): Número de tiros libres anotados

6.4. Code

URL GITHUB