

El uso de indicadores de desempeño normalizados para la valoración de jugadores: El caso de las estadísticas por minuto en baloncesto

Martínez García, Jose A.
Martínez Caro, Laura

Universidad Politécnica de Cartagena

Jose Antonio Martínez García
Departamento de Economía de la Empresa.
Universidad Politécnica de Cartagena.
Facultad de Ciencias de la Empresa. Calle Real, 3. 30201. Cartagena. España.
Tel.: +34 968 32 57 76; Fax: +34 968 32 70 81.
E-mail: josean.martinez@upct.es

Cómo citar este artículo:

Este artículo es una versión del aceptado para publicación en la revista *Motricidad. European Journal of Human Movement*.

Martínez, J. A. y Martínez, L. (2010). El uso de indicadores de desempeño normalizados para la valoración de jugadores: el caso de las estadísticas por minuto en baloncesto. *Motricidad. European Journal of Human Movement*. Aceptado.

How to cite this paper:

This article is a draft of the accepted paper to be published in the following journal: *Motricidad. European Journal of Human Movement*.

Martínez, J. A. y Martínez, L. (2010). El uso de indicadores de desempeño normalizados para la valoración de jugadores: el caso de las estadísticas por minuto en baloncesto. *Motricidad. European Journal of Human Movement*. Accepted for publication.

El uso de indicadores de desempeño normalizados para la valoración de jugadores: El caso de las estadísticas por minuto en baloncesto

Resumen

Esta investigación somete a prueba la hipótesis subyacente al uso de indicadores normalizados por minuto en baloncesto: No existe asociación entre los índices normalizados y los minutos jugados por cada jugador. Esa asunción es básica para la comparación de jugadores con índices normalizados, ya que si existe algún tipo de relación funcional entre el ratio normalizado y los minutos jugados, existiría un sesgo en las clasificaciones realizadas en cada apartado estadístico, ya que esa normalización quedaría desvirtuada. Para ello, aplicamos varios métodos hasta ahora no utilizados en el estudio de este tema, como el análisis de la variabilidad intra-jugador y los modelos de ecuaciones estructurales. Los resultados de varios análisis de naturaleza confirmatoria y exploratoria muestran que no existe una asociación lineal relevante entre los minutos jugados y las diferentes categorías normalizadas por minuto: puntos, rebotes, asistencias, recuperaciones y tapones. Las consecuencias relacionadas con el rendimiento deportivo y el marketing son discutidas. De este modo, esta investigación contribuye de forma novedosa a la pujante línea de investigación relacionada con la investigación científica en baloncesto

Palabras clave: Normalización, baloncesto, desempeño, estadísticas.

Abstract

This research empirically tests the hypothesis underlying the use of performance indicators in basketball, when they are normalized using the “per-minute” criterion: There is no association among normalized indicators and the minutes played by each player. This assumption is a necessary requirement in order to compare players using normalized indicators, because if this assumption is not met, then there would be a possible bias in the rankings of each statistical category. We apply several under-used methods in this type of research, such as the analysis of the intra-player variability and the structural equation modelling. Results show that there is no lineal association between the minutes played and the different per-minute normalized categories as: points, rebounds, assists, steals, and blocks. We discuss implications for sport performance and marketing. Therefore, this study contributes to the advance in one of the growing areas of sports research; such as basketball metrics.

Key words: Normalization, basketball, performance, statistics.

Introducción

En los últimos años se está produciendo un cambio profundo en el papel que las estadísticas juegan en baloncesto. El fenómeno “Moneyball” (Lewis, 2003), ha propiciado una revolución en Estados Unidos, la cual se está diseminando poco a poco por el resto del mundo. La creación de APBRmetrics (*Association for Professional Basketball Research*), por ejemplo, ha supuesto un vehículo de discusión e innovación en cuanto a las posibilidades que tiene la aplicación de la estadística en baloncesto. Prácticamente todos los investigadores y profesionales analistas más reconocidos mundialmente participan en este foro, donde se comentan multitud de artículos científicos publicados en revistas especializadas y libros, además de nuevas aportaciones de los miembros, no necesariamente filtradas por el proceso de doble revisión académica. En España, los trabajos de Felipo (2005) o Martínez y Martínez (2009, 2010) permiten profundizar en la importancia de esta temática, pudiendo consultarse las principales referencias al respecto.

Básicamente son tres las ventajas que confiere este análisis científico del baloncesto a través de la estadística: La primera de ellas es la evaluación, tanto del desempeño de jugadores como de los equipos, en pos de conseguir una visión más objetiva de la productividad, eficiencia, eficacia y valor de los componentes del juego, tanto a nivel individual como global¹. Prueba de ello, por ejemplo, son los numerosos índices de valoración surgidos al respecto (en www.nbastuffer.com pueden consultarse muchos de ellos). Esa evaluación tiene además una consecuencia capital en la revalorización de mercado de los jugadores, ya que existe un consenso generalizado sobre la importancia de la valoración de jugadores (ya sea a través de índices de valoración o a través de la creación de rankings de liderazgo estadístico), y su valor de mercado². Las siguientes referencias son un reflejo de esta primera utilización de las estadísticas: Berri (1999; 2008), Berri y Eschker (2005), Fernández, Camerino, Anguera y Jonsson (2009), Hoon-Lee y Berri (2008), Esteller-Moré y Eres-García (2002), Mavridis, Tsamourtzis, Karipidis y Laios (2009), Rimler, Song y Yi (en prensa), Piette, Sathyanarayan y Kai (2010). La segunda de ellas (ciertamente relacionada con la primera) es la predicción, es decir, intentar obtener la información necesaria para poder tomar decisiones directivas con menor riesgo (contratación de jugadores, manipulación de factores que inciden en la probabilidad de victoria, etc.). Así, la literatura es abundante en referencia a esta temática, por ejemplo: Alferink, Critchfield, Hitt y Higgins (2009), Berri, Brook y Schmidt (2007), Berri y Schmidt (2002), Berry, Reese y Larkey (1999), Gómez y Lorenzo (2005), Hitt, Alferink, Critchfield y Wagman (2007), Ibáñez, Sampaio, Sáenz-López, Jiménez y Janeira (2003), Romanowich, Bourret y Vollmer (2007), Sánchez, Castellanos y Dopico (2007), Skinner (2010), Tauer, Guenther y Rozek, (2009), Trininic, Dizdar y Luksic (2002), Vollmer, y Bourret, (2000). Existen además, aplicaciones muy complejas de algoritmos basados en la inteligencia artificial, donde a través de procesos de minería de datos se pueden buscar patrones que permitan resumir la información de miles de datos y realizar predicciones. Esto es usado para conseguir diferentes objetivos: por ejemplo, desde el punto de vista de las apuestas deportivas o juegos de simulación (*fantasy games*), existen programas informáticos basados en algoritmos de aprendizaje, como NeuralBet (www.neuralbet.com). Un uso alternativo de este tipo de tecnología es su aplicación al análisis del juego por parte de los entrenadores, como el programa Advanced Scout (www.virtualgold.com, <http://www-03.ibm.com/press/us/en/pressrelease/1190.wss>), cuya aplicación es bastante popular en la NBA. Otras aplicaciones, como los modelos multiagente o las

¹ En el comienzo de la temporada 2009/2010, la NBA ha presentado su herramienta StatsCube. Este software permite tener acceso instantáneo a los datos del “play by play” desde la temporada 1996/1997, y manejar en un entorno amigable múltiples posibilidades de cruce de datos y análisis. Hasta ahora, esos datos eran más difíciles de obtener, ya que requerían del análisis de miles de líneas de código de “play by play”. Con esta nueva herramienta, la NBA pretende facilitar el análisis para los responsables estadísticos y entrenadores de los equipos de la Liga. Además de StatsCube, la NBA tiene a prueba actualmente un sistema revolucionario de análisis de vídeo, con seis cámaras de alta definición que cubren cada aspecto del juego, y que posteriormente debe integrarse con los datos numéricos pertinentes sobre las acciones del partido. No obstante, la empresa Synergy Sports, ofrece ya un servicio basado en esta integración de vídeo y datos de “play by play”, que actualmente utilizan varios equipos de la NBA (NBA.COM, 2009)

² Actualmente estamos realizando una investigación sobre la importancia de las estadísticas en baloncesto. Hemos creado varios paneles de expertos que están evaluando varias cuestiones desde su óptica particular (entrenadores, jugadores, periodistas especializados, catedráticos de marketing, economía y deporte, *bloggers*, aficionados, expertos analistas, etc.) Es un panel mundial, por lo que están participando expertos de todo el mundo. Hasta ahora, en las 175 encuestas que llevamos recibidas (desde Mayo hasta Octubre de 2009), existe un amplio consenso en que los sistemas de valoración estadísticas pueden revalorizar a unos jugadores e infravalorar a otros, desde el punto de vista de su valor de mercado.

redes bayesianas dinámicas, están suscitando el interés de la literatura académica (ej. Perše, Kristan, Kovacic, Vuckovic y Perš, 2009; Zhang, Gatica-Perez, Bengio, y Roy, 2005). Finalmente, la tercera aplicación de la estadística hace referencia a la capacidad de su enfoque para analizar temas controvertidos referentes a la igualdad de las competiciones, la posible manipulación de éstas, la determinación de salarios, o la discriminación de género o racial. Como muestra de ello podemos citar: Berri, Brook, Frick, Fenn y Vicente-Mayoral (2005), Balsdon, Fong y Thayer (2007), Fort, Hoon-Lee y Berri (2008), Fort y Maxcy (2003), Humphreys (2000; 2002), Michaelides (en prensa), Price and Wolfers (2007), Zimmer y Kuethe (2007).

En esta investigación, se aborda un problema relacionado con el primer uso de las estadísticas comentado anteriormente, como es el de la evaluación del desempeño de jugadores. Esos índices de desempeño reflejan la capacidad del jugador para realizar diferentes acciones del juego, permitiendo elaborar clasificaciones o rankings, en función de la comparación entre ellos. Además de las categorías clásicas de valoración en baloncesto (puntos, rebotes, asistencias, recuperaciones, etc.), existen otros índices más elaborados basados en aproximaciones econométricas y ponderaciones lineales (Felipo, 2005; Oliver, 2004; www.nbastuffer.com). Un rasgo distintivo de la inmensa mayoría de éstos índices es su normalización o ponderación por los minutos en pista de cada jugador. Ejemplos del uso de este tipo de normalización en el baloncesto profesional pueden encontrarse en la Liga ACB en España, la NBA, y las ligas italiana, griega o francesa. Existen, asimismo, webs especializadas creadas por reconocidos analistas donde se pueden obtener diferentes índices de valoración y estadísticas avanzadas ponderados por minutos, como www.basketball-reference.com (Justin Kubatko), www.bellotibasketball.com (Bob Belloti), www.dougstats.com (Doug Steele)³. Entre los índices más famosos de valoración de jugadores que se ponderan por minutos destacamos: TENDEX (David Heeren), Player Efficiency Rating (John Hollinger), Wins Produced (David Berri) o Diamond Rating (Kevin Broom). Además, existen análisis comparativos del desempeño de combinaciones distintas de jugadores en cancha, donde se utiliza la normalización por minutos (Winston, 2009). Esa normalización por minuto, ha sido destacada como uno de los criterios básicos para analizar el desempeño de los jugadores en cualquier disciplina deportiva (Hughes, 2004).

Específicamente, esta investigación pone a prueba la hipótesis subyacente del uso de este tipo de normalización: que no existe asociación entre los índices normalizados y los minutos jugados por cada jugador. Esa asunción es básica para la comparación de jugadores con índices normalizados, ya que si existe algún tipo de relación funcional entre el ratio normalizado y los minutos jugados, existiría un sesgo en las clasificaciones realizadas en cada apartado estadístico, ya que obviamente, esa normalización quedaría desvirtuada. Para ello, aplicamos varios métodos hasta ahora no utilizados en el estudio de este tema, como el análisis de la variabilidad intra-jugador⁴ y los modelos de ecuaciones estructurales, metodología ésta que proporciona una serie de ventajas con respecto a las aproximaciones metodológicas hasta ahora utilizadas en investigaciones similares. De este modo, esta investigación contribuye de forma novedosa a la pujante línea de investigación relacionada con la investigación científica en baloncesto.

Antecedentes

No se ha podido encontrar investigaciones publicadas en ninguna revista académica en relación al tema de nuestro estudio. Por tanto, parece evidente que existe la oportunidad de cubrir un hueco evidente en la literatura especializada. Sin embargo, sí existen estudios realizados por diferentes analistas, aunque ninguno de ellos se ha sometido a procesos de revisión ciega. No obstante, merecen ser tenidos en consideración, dado que como hemos especificado anteriormente, es muy común que analistas e investigadores relacionados con el movimiento “Moneyball”, realicen pequeños estudios que someten a discusión en foros especializados⁵.

³ Podemos encontrar, además, otras webs como www.espn.com, <http://basketball-statistics.com/howpacworks.html>, www.draftexpress.com.

⁴ En un artículo coetáneo con esta investigación, Arkes (2010), utiliza también un diseño de investigación intra-jugador para poner a prueba la hipótesis de la “mano caliente” (*hot hand*), es decir, que la probabilidad de anotar un lanzamiento se incrementa si se ha anotado el anterior. Los resultados de Arkes (2010) contradicen los múltiples estudios realizados anteriormente sobre esta temática, indicando que efectivamente, existe esa asociación. No obstante, la investigación de Arkes (2010) se restringe únicamente a los lanzamientos libres, y no a los lanzamientos de campo.

⁵ Esta no es una práctica habitual en la investigación científica en España, al menos en esta área de las ciencias sociales. Es común que revisores y editores de revistas editadas en España tomen con recelo las

La única referencia encontrada proviene de diferentes aportaciones realizadas en APBRmetrics⁶ por distintos analistas, y del estudio realizado por Kurylo (2007), el cual viene motivado precisamente por la discusión generada en ese foro. Kurylo (2007) realiza un interesante análisis relacionando una medida de desempeño normalizada en minutos, el Player Efficiency Rating - PER - (Hollinger, 2005), frente a los minutos jugados por los jugadores de la NBA en varias temporadas (desde 1981 hasta 2006). Las condiciones para formar parte de la muestra de estudio eran: (1) haber jugado más de 40 partidos, (2) promediar entre 10 y 25 minutos por partido, (3) tener como edad un número par. Esta última condición permite a Kurylo (2007) formar una especie de grupo de control para su análisis, con el fin de comparar posteriormente los resultados con un grupo que podíamos llamar experimental, formado por jugadores con edad impar. Un total de 1840 jugadores formaban la muestra total. Los resultados de su modelo de regresión lineal son claros: existe una relación lineal positiva entre el PER y los minutos jugados ($R^2 = 0.109$). Sin embargo, Kurylo (2007) no toma este resultado como evidencia suficiente, y realiza un subsecuente análisis a otro grupo de jugadores, cuyas condiciones son: (1) haber jugado más de 40 partidos, (2) promediar el año anterior entre 10 y 25 minutos por partido, (3) haber incrementado el promedio de minutos en el año corriente en +5, (4) tener edad par. Las conclusiones de su nuevo modelo de regresión lineal muestran otra vez una relación positiva ($R^2 = 0.174$). Sin embargo, el tamaño de efecto encontrado es interpretado como bastante pequeño, ya que el efecto marginal de los minutos sobre el PER es de 0.17, cifra que Kurylo (2007) considera poco importante. De este modo, concluye que el desempeño de los jugadores normalizado por minutos no está asociado linealmente (al menos no de forma relevante desde el punto de vista práctico) con los minutos jugados.

Kurylo (2007) concluye que los jugadores que juegan 10 o más minutos por partido muy probablemente mantendrán sus mismas estadísticas por minutos independientemente del tiempo que estén en cancha. Kurylo (2007) se hace eco de otros estudios, especialmente los de los reconocidos analistas John Hollinger y Justin Kubatko, los cuales, insistimos, no han sido expuestos a revisión académica. El estudio de Kubatko obtiene resultados similares a los de Kurylo (2007), mientras que Hollinger concluye que aquellos jugadores que juegan 15 minutos o más rendían significativamente mejor que aquellos que jugaban por debajo de esa cifra.

Se pueden realizar varias críticas importantes al trabajo de Kurylo (2007). En primer lugar, utiliza como variable dependiente el PER, índice que es una composición de diferentes estadísticas individuales. El PER ha sido ampliamente criticado por carecer de base econométrica y premiar a jugadores con bajos porcentajes de tiro (Berri, 1999; 2008; Winston, 2009), pudiendo además enmascarar los efectos de los minutos jugados sobre las diferentes categorías estadísticas individuales (puntos, rebotes, asistencias, etc.). Son precisamente estas categorías individuales las que predominan en los rankings estadísticos en todas las competiciones a nivel mundial.

En segundo lugar, Kurylo (2007) toma los valores medios por año de cada jugador, por lo que no se estudia la variabilidad individual de cada jugador, es decir, la variabilidad intra-jugador. Por tanto, no se tiene en cuenta la dispersión de la distribución de probabilidad que refleja las actuaciones de cada jugador en los partidos de la temporada.

En tercer lugar, Kurylo (2007) toma varias temporadas en su análisis, por lo que los jugadores que aparecen varias veces en la muestra pueden estar sujetos a cambios sistemáticos debidos a envejecimiento, incremento o decrecimiento de su habilidad, lesiones, etc.

Existen otros factores que merecen ser tenidos en consideración para abordar este problema de investigación. Algunos de ellos también han sido comentados espontáneamente por diferentes analistas en

referencias a este tipo de foros. No ocurre así en otras latitudes. Por ejemplo, SEMNET es un foro de discusión sobre modelos de ecuaciones estructurales (<http://bama.ua.edu/cgi-bin/wa?A0=SEMNET>), donde participan algunos de los más reconocidos metodólogos en ciencias sociales. Discusiones generadas en ese foro han sido debatidas en importantes revistas científicas, como *Structural Equation Modeling: A Multidisciplinary Journal*, o *Personal and Individual Differences*, referenciando explícitamente diferentes mensajes y contribuciones de los miembros del foro. En una investigación muy reciente, Berri y Bradbury (2010) reflexionan sobre este hecho en el campo de la estadística aplicada al deporte.

⁶ <http://sonicscentral.com/apbrmetrics/viewtopic.php?t=85&postdays=0&postorder=asc&start=45>

Kurylo (2007) o en APBRmetrics. Por ejemplo, la Ley de los Rendimientos Decrecientes indica que existiría un punto en el que el incremento marginal de minutos jugados provocaría un descenso del desempeño por minuto del jugador, ya que obviamente, el cansancio podría incidir en su rendimiento. Sin embargo, [se puede](#) objetar que el jugador de baloncesto debe estar preparado y entrenado para realizar su desempeño independientemente de los minutos que jugara en cada partido, siempre teniendo en cuenta que esa preparación iría enfocada a partidos de 40 o 48 minutos dependiendo de la competición. Por tanto, no está tan claro que pueda existir un efecto no lineal relativo a este fenómeno⁷.

Otros aspectos a tener en cuenta que podrían afectar de manera sistemática al desempeño de los jugadores son los comentados por Reed y O'Donoghue (2005): si el partido es en casa o fuera, el tiempo de descanso después del último partido, la posición en la liga en el momento del partido, la oposición del equipo contrario, la distancia recorrida en el viaje hacia el partido, la distancia recorrida por el equipo contrario, o la relevancia en el ranking de la liga.

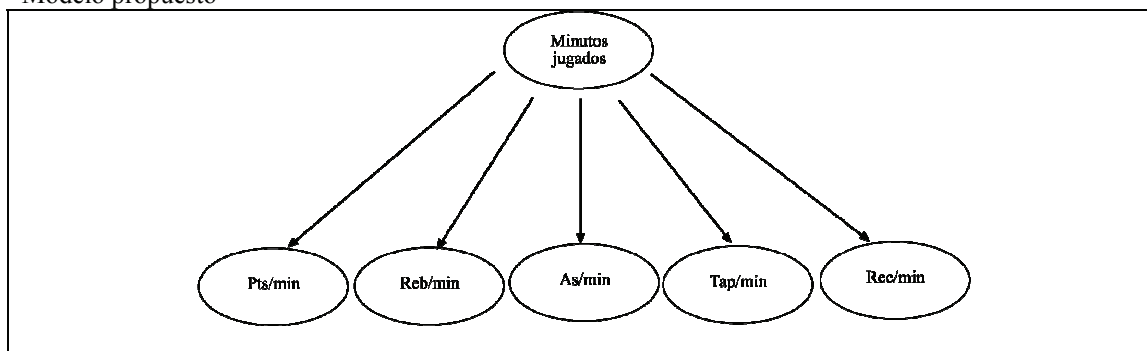
Estos factores son solo una muestra de aquellos que podrían afectar al rendimiento de los jugadores (ciertamente por Reed y O'Donoghue, 2005, reconocen que existen muchos más), pero podrían ser considerados algunos de los más importantes. En cualquier caso, la existencia de estas variables advierte del peligro de muestrear jugadores en distintas condiciones, es decir, de equipos distintos y situaciones deportivas distintas, ya que no se podría controlar por estos posibles factores contaminantes. Por ello, sería más prudente que los estudios realizados sobre esta temática se circunscribieran a unas circunstancias concretas, analizando las especificidades de cada jugador. A continuación, se explica, la aproximación metodológica seguida al respecto en esta investigación.

Metodología

Modelo

Proponemos el siguiente modelo teórico, que refleja esquemáticamente la hipótesis que queremos someter a prueba (Figura 1).

Figura 1
Modelo propuesto



La Figura 1 representa una relación funcional entre diferentes categorías estadísticas normalizadas por minutos y los minutos jugados. La hipótesis de partida es que no existe esa relación funcional, es decir, las diferentes medidas de rendimiento del jugador no dependen de los minutos jugados. Las categorías elegidas para el análisis son: puntos, rebotes, asistencias, taponos, recuperaciones. La elección de esas categorías sigue la línea de razonamiento expuesta en Martínez y Martínez (2009; 2010), ya que son las más utilizadas en diferentes contextos, tanto en la creación de rankings individuales, como para formar parte en la construcción de índices de valoración (ej. TENDEX, Valoración ACB, PER, etc.).

⁷ No obstante, hay un estudio realizado por Dick Mays y Jeff Ganter, en el que muestran como existe un decrecimiento en la productividad del jugador cuando éste juega más del 90% de los minutos de un partido. Para ello, estudian específicamente a aquellos jugadores que permanecen en cancha prácticamente todo el partido, y analizan el efecto producido sobre una medida de productividad creada por ellos: el “Mays Magic Metric”, sistema de ponderaciones lineales derivado del *box-score*, y que mezcla las contribuciones ofensivas y defensivas de los jugadores. Mays y Ganter son dos matemáticos que durante varios años estuvieron realizando análisis sistemáticos de la productividad de los jugadores de la NBA (Mays, comunicación personal).

Este esquema básico puede trasladarse a un modelo de ecuaciones, de la siguiente forma (1):

$$Y_{ij} = f(X_i, Z_{im}, W_{ik}, e_i) \quad (1)$$

representando Y_{ij} el valor de la categoría estadística $j=1\dots5$ para el jugador i , es decir, los puntos, rebotes, asistencias, tapones y recuperaciones por minuto. X_i representa los minutos jugados por cada jugador, mientras que Z_{im} hace referencia al conjunto de m variables que podrían afectar sistemáticamente a Y_{ij} y que están relacionadas con X_i . Por ejemplo, los minutos jugados por un jugador en una liga podrían estar asociados con su edad, con la distancia recorrida en el viaje al partido, o a la participación en partidos de competición entre semana (caso de las ligas europeas). W_{ik} representa el conjunto de k variables que podrían afectar sistemáticamente a Y_{ij} y que no estarían asociadas con X_i . Así, se podrían identificar variables como la posición en la liga en el momento del partido, la oposición del equipo contrario, la distancia recorrida en el viaje hacia el partido, la distancia recorrida por el equipo contrario, o la relevancia en el ranking de la liga. La existencia de estas variables no afectaría a la estimación de los coeficientes del modelo (y por tanto a la hipótesis planteada), únicamente a la capacidad explicativa de éste. Por tanto, el no considerar esas variables en el modelo planteado, no tendría influencia sobre la estimación de los parámetros que relacionan la variable independiente y las dependientes. Por último, e_i representa al conjunto de variables desconocidas y que no tienen un efecto sistemático sobre Y_{ij} , pudiendo considerarse como una perturbación aleatoria con valor esperado nulo. La función f relaciona ambos lados de la ecuación, estableciendo la relación funcional existente. La forma más básica y simple es la relación de identidad, caso de la regresión lineal. Esa linealidad se refiere a la parámetros a estimar que relacionan X_i con Y_{ij} , pudiendo incorporar características no lineales para el conjunto de variables del modelo (caso de tomar logaritmos o valores al cuadrado, por ejemplo), manteniendo la linealidad en los parámetros. Esos parámetros, se interpretan como el efecto del cambio en el valor esperado de la variable dependiente cuando la independiente cambia en una unidad.

El modelo estructural planteado puede expresarse entonces de la siguiente manera (2):

$$Y_{ij} = f(X_i, Z_{im}, W_{ik}, e_i) \quad (2)$$

Muestra

El estudio se circunscribe al análisis de aquellos jugadores que han jugado al menos 70 partidos en la temporada regular 2008/2009 en la NBA. Esos 70 partidos son el criterio de corte estipulado oficialmente por la NBA para que los jugadores formen parte de los rankings estadísticos de liderazgo. Esto supone jugar alrededor del 85% de los partidos de la temporada. Como Martínez y Martínez (2010) indican, es un criterio excesivamente general, pero que se aproxima bastante al que un enfoque probabilístico determinaría como óptimo para cada jugador. Un total de 178 jugadores cumplen con este criterio exigido (según los datos tomados de www.basketball-reference.com), siendo muy heterogéneos sus minutos de media (desde 39.9 de Andre Iguadala, a 12.1 de Royal Ivey). Sin embargo, se busca centrar el análisis en aquellos jugadores que tienen los máximos valores de variabilidad en sus minutos a lo largo de la temporada, es decir, con una varianza mayor en la distribución de sus minutos. Este hecho incorpora valor añadido a al análisis, ya que la dispersión de la variable independiente es mayor.

Para ordenar a los 178 jugadores en función de su varianza, se eliminaron primero aquellos casos que podrían considerarse como atípicos o *outliers*. Obviamente, la varianza es una medida descriptiva muy sensible a casos atípicos. El criterio fue eliminar aquellos casos (partidos) en los que los jugadores habían jugado menos de 5 minutos. El establecimiento de ese punto de corte tiene su base en el concepto de posesión. Según Winston (2009), la media en los últimos años en la NBA es de 92 posesiones por equipo en cada partido, lo que se traduce en prácticamente 2 posesiones por minuto. En 5 minutos, por tanto, alrededor de 10 posesiones son utilizadas por los equipos, siendo este tiempo generalmente suficiente para que un jugador en pista realice acciones de juego relevantes.

Los 178 jugadores generaron 13972 registros (partidos). Tras la eliminación de los casos atípicos la base de datos total se componía de 13807 registros. El siguiente paso fue calcular la varianza intra-jugador, para posteriormente ordenar los jugadores de forma descendente. Pese a la gran disponibilidad de datos sobre los jugadores de la NBA existente en diferentes webs especializadas, no existen datos sobre la varianza de los jugadores en cada apartado del juego y durante una temporada. Esta circunstancia, ha

hecho que se tuviera que calcular uno a uno la varianza en minutos para cada jugador, sin posibilidad de usar los diferentes motores de búsqueda y potentes filtros de las webs especializadas. Los resultados de esta ordenación mostraron como Dominique McGuire ocupaba la primera posición, con una varianza de 110.874, mientras que Jason Kidd, ocupaba la última, con una varianza de 14.957. Obviamente, es mucho más indicado estudiar jugadores con la varianza lo más alta posible, ya que jugadores como Jason Kidd, por ejemplo, juegan un número de minutos bastante similar partido a partido, por lo que no existiría una dispersión suficiente para realizar un análisis fiable de la influencia de los minutos de juego en el rendimiento por minuto.

El hecho de circunscribir los datos a un solo año permite minimizar la posible aparición de Z_{im} , ya que factores como la edad de cada jugador, por ejemplo, no tendrían un efecto notorio. Este es el mismo razonamiento usado por Alferink, Critchfield, Hitt y Higgins (2009),

Análisis de datos

Filosofía de análisis

Las características de las diferentes variables dependientes a considerar hacen que el modo de proceder a nivel analítico sea un poco complejo. Los Ptos/Min y los Reb/Min tienen distribuciones normales o no demasiado alejadas de la normalidad. Sin embargo, las As/Min, Rec/Min y Tap/Min, presentan la particularidad de la presencia de un gran número de ceros en su distribución. Obviamente, esa distribución es distinta si se compara un base con un pivot, ya que los tapones son una característica mucho más prevalente en los hombres interiores que en los exteriores. No obstante, la presencia elevada de ceros, hace que ciertas distribuciones se asemejen a “distribuciones infladas en cero”, con una interpretación a caballo entre las distribuciones de Poisson y continuas censuradas (ver Vermunt y Magidson, 2005), o distribuciones binomiales negativas. Por ello, se han acometido dos estrategias distintas de análisis:

Análisis de los Ptos/Min y Reb/Min

Se han utilizado los modelos de ecuaciones estructurales para testar la hipótesis de partida en esta investigación: los minutos no afectan a los diferentes indicadores normalizados. Para ello, a través de esta potente metodología, se puede plantear una estructura de covarianza con dos variables dependientes (Ptos/Min y Reb/Min) y una independiente (Minutos), obteniendo un índice de ajuste en una única estimación.

Se parte de la premisa de que el modelo planteado es el modelo real que existe en la población. Aunque cada jugador está sujeto a unas condiciones particulares (equipos, ciudad, viajes, compañeros, etc.), se asume que el modelo es el mismo para todos, lo que refleja la hipótesis subyacente a la normalización. De este modo, se puede comparar el ajuste de un modelo restringido (la influencia de los minutos sobre las variables dependiente es “cero”), con un modelo en el que se estime libremente esa influencia. Como en este caso, este segundo modelo es un modelo saturado, esa comparación resulta poco informativa. De este modo, se confía únicamente en el ajuste del modelo restringido para evaluar la hipótesis de partida, siguiendo la filosofía de ajuste exacto de Hayduk (1996) o Hayduk, Cummings, Boadu, Pazderka-Robinson, y Boulianne (2007). No obstante, el modelo saturado sirve para estimar los coeficientes de influencia cuando se relaja la hipótesis de partida. Así, y partiendo del mismo modelo de población, la comparación de diferentes matrices muestrales (una para cada jugador) lleva a una estrategia implícita de replicación (Hitchcock, 2002), reflejada en las diferentes situaciones test (particulares para cada jugador). Así, se podría construir una distribución empírica de efectos, que no es más que la base de un meta-análisis. Para ello, se han utilizado los 50 primeros jugadores ordenados en función de su varianza, tamaño suficiente para construir una distribución de efectos. De este modo, la estrategia a seguir es eminentemente confirmatoria.

Análisis de los As/Min, Rec/Min y Tap/Min

Las características de los datos en estas tres categorías complican enormemente los análisis. Para muchos jugadores, la distribución de sus estadísticas en estos apartados sigue formas muy alejadas de la normalidad. Así, una característica fundamental es la presencia de un gran número de ceros en muchas de esas distribuciones. El problema surge para intentar construir modelos que expliquen la variación en esas variables dependientes (recordemos que la única variable independiente a incluir en los modelos serían

los minutos jugados). Pueden estimarse modelos de regresión censurada o regresión censurada inflada en ceros, así como regresión de Poisson, binomial negativa, Poisson inflado en cero, o binomial negativos inflado en cero⁸. Como puede intuirse, la elección de una u otra especificación cambia los resultados del análisis. Y esa elección no es fácil. Por ejemplo, en el caso de los Taponos/Minuto, existen jugadores como Steve Nash que en 65 de sus 74 partidos no han puesto ningún tapón. La elección de una regresión censurada inflada en ceros podría ser óptima si consideramos que los Taponos/Minuto son una variable continua. Sin embargo, esos 9 valores distintos de cero podrían hacer pensar que la distribución es discreta, en este caso del tipo de Poisson inflada en cero. Precisamente Steve Nash, en el apartado de Recuperaciones/Minuto tiene 35 de sus partidos en cero. Aquí la elección entre una regresión de Poisson inflada en cero y una regresión de Poisson podría realizarse aplicando el test de Vuong (1989). No obstante, desde el punto de vista teórico, si asumimos que la variable dependiente sigue una distribución inflada en cero, podría interpretarse que existe heterogeneidad no observable, con dos tipos de casos en la población: para el primer caso la variable siempre toma cero, y para el segundo se comporta como una variable de Poisson o negativa binomial.

Dadas todas estas dificultades, que ciertamente merecerían una investigación mucho más profunda, se ha optado por adoptar un análisis más simple y de carácter eminentemente exploratorio. Así, vamos a explorar la relación que existe entre los minutos jugados y esas categorías estadísticas en los 10 primeros puestos del ranking para cada una de ellas. Esta aproximación facilita el análisis, ya que esos primeros puestos están ocupados por jugadores con gran desempeño en esas categorías, con lo que es de esperar que las distribuciones se acerquen a la normalidad⁹. Así, el uso de la regresión lineal será la herramienta a utilizar.

Resultados

En primer lugar se describe la distribución de minutos de los 50 jugadores seleccionados en función de su varianza, con el fin de mostrar la alta dispersión de los datos, que como hemos indicado es una circunstancia deseable. La Figura 2 muestra la dispersión jugador a jugador, mientras que la Figura 3 da una visión global de la variabilidad agregada.

⁸ Paquetes estadísticos como Latent Gold o Stata pueden estimar este tipo de especificaciones.

⁹ Es la aproximación que también toman Martínez y Martínez (2010), para la creación de rankings en asistencias, recuperaciones y taponos.

Figura 2. Distribución de minutos de los 50 jugadores con mayor varianza

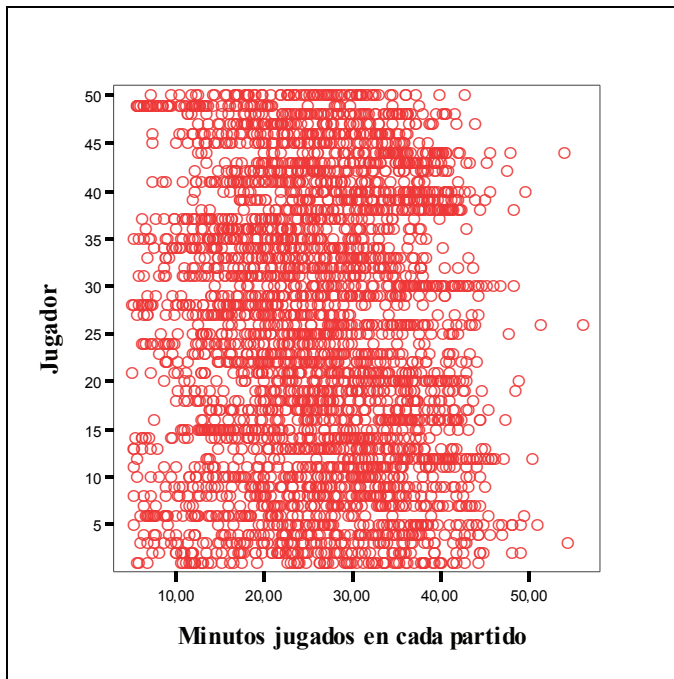
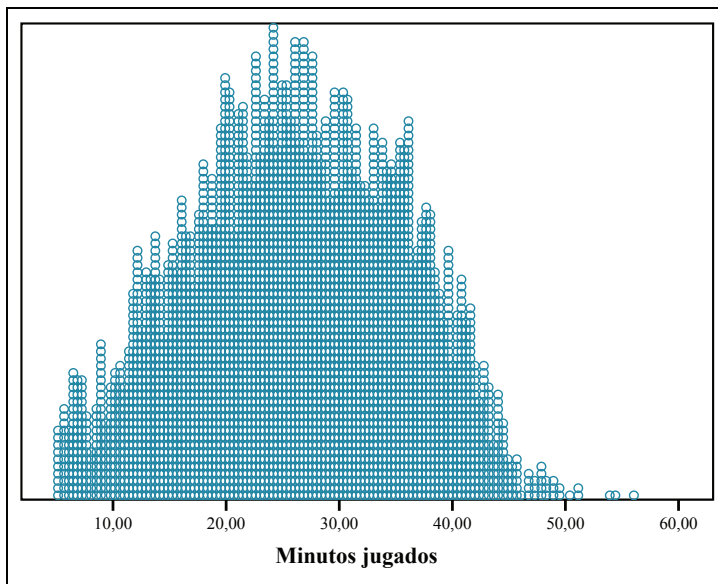


Figura 3. Distribución global de los datos



Análisis confirmatorio

La Tabla 1 muestra los resultados de los análisis para los 50 jugadores seleccionados. En ella se describe el valor del estadístico χ^2 para el modelo restringido, los grados de libertad, y su correspondiente nivel de significación, basados en una estimación por máxima verosimilitud. Hay que recordar que valores por debajo de .05 se consideran significativos, es decir, no favorecen la hipótesis nula, por lo que el modelo planteado sería inconsistente con la matriz de datos muestrales, una vez tenido en cuenta el error muestral. Para ello, se plantea un modelo en LISREL 8.8 (Jöreskog y Sörbom, 2006), especificando una covariación no nula entre las variables dependientes. Esta especificación es consistente con el modelo teórico, especificado en la Ecuación (2), ya que se está considerando implícitamente que los minutos jugados no es la única variable que afecta sistemáticamente a los Pts/Min y Reb/Min, sino que existen

otras variables W_{ik} no correlacionadas con los minutos jugados y que son importantes en la determinación de las variables dependientes, como puede ser, por ejemplo, la habilidad de cada jugador. Dado que existen algunos casos en los que la distribución de los datos se aleja moderadamente de la normalidad multivariante, se ha computado también la corrección por curtosis propuesta por Satorra y Bentler (1999)¹⁰ utilizando el método de máxima verosimilitud robusto. No obstante, en situaciones de no normalidad poco severas, el estadístico χ^2 se comporta de forma confiable. La Tabla 1 también muestra el valor de los parámetros estimados en el modelo no restringido, que son los coeficientes estructurales que reflejan la influencia de los minutos jugados en los Pts/min y Reb/min, con el intervalo de confianza asociado al 95%. Finalmente, incluimos el valor del coeficiente de determinación R^2 , que indica la varianza explicada de cada una de las variables dependientes.

Tabla 1. Influencia de los Min. Sobre los Pts/min y Reb/min

Jugador	χ^2 Modelo restringido	χ^2 Satorra-Bentler	Min → Pts/min	IC-	IC+	R^2	Min → Reb/min	IC-	IC+	R^2
D. McGuire	1.092 (2): .579	.180 (2): .914	.001	-.001	.003	.011	.000	-.002	.002	.002
C. J. Watson	5.559 (2): .0621	4.298 (2): .117	.003	-.001	.007	.028	.002	.000	.004	.050
B. Rush	9.140 (2): .010	3.903 (2): .192	.005	-.001	.011	.059	.002	.000	.004	.049
R. Sessions	6.565 (2): .037	.371 (2): .831	.005	.001	.009	.075	.000	-.002	.002	.004
K. Azubuike	2.429 (2): .297	1.063 (2): .588	.002	-.004	.008	.013	-.001	-.003	.001	.019
A. Afflalo	3.300 (2): .192	.727 (2): .695	.004	.000	.008	.035	-.001	-.003	.001	.014
S. Hawes	3.105 (2): .212	2.844 (2): .241	-.001	-.005	.003	.005	-.002	-.004	.000	.037
C. Lee	2.471 (2): .291	2-045 (2): .360	.001	-.005	.007	.002	-.001	-.003	.001	.026
C. Villanueva	4.853 (2): .088	.895 (2): .639	.006	.000	.012	.052	-.001	-.003	.001	.010
J. Noah	3.265 (2): .195	.450 (2): .799	.003	-.001	.007	.039	.001	-.003	.005	.006
S. Telfair	16.709 (2): .0002*	2.486 (2): .288	.008	.004	.012	1.777	-.001	-.003	.001	.047
E. Gordon	15.608 (2): .0004*	.563 (2): .754	.008	.002	.014	.161	-.001	-.005	.003	.024
T. Thomas	18.124 (2): .0001*	1.064 (2): .587	.008	.004	.012	.196	.002	-.002	.006	.021
D. Cook	10.049 (2): .00657	.150 (2): .928	.009	.005	.013	.124	.000	-.002	.002	.002
G. Davis	4.497 (2): .106	.120 (2): .942	.006	.000	.012	.058	.000	-.002	.002	.001
R. Butler	3.475 (2): .176	1.643 (2): .440	.002	-.002	.006	.018	-.001	-.003	.001	.024
D. J. Agustin	.093 (2): .957	.0253 (2): .987	-.001	-.007	.005	.001	.000	-.002	.002	.000
Q. Richardson	9.747 (2): .00765*	.421 (2): .810	.009	.003	.015	.114	-.001	-.003	.001	.005
L. Ridnour	6.597 (2): .0369	.692 (2): .708	.005	.001	.009	.072	-.001	-.003	.001	.010
A. Bargnani	21.140 (2): .000*	1.044 (2): .593	.011	.007	.015	.224	.001	-.001	.003	.018
J. Moon	5.794 (2): .055	2.564 (2): .277	.004	.000	.008	.051	.002	.000	.004	.029
A. Brooks	3.136 (2): .208	.009 (2): .995	.005	-.001	.011	.038	.000	-.002	.002	.000
C. Bell	15.069 (2): .0005	4.300 (2): .117	.011	.007	.015	.174	.002	.000	.004	.054
S. Novak	6.175 (2): .0456	1.108 (2): .575	.009	.001	.017	.097	-.001	-.005	.003	.018
M. Beasley	2.036 (2): .361	2.011 (2): .836	.003	-.001	.007	.014	-.001	-.003	.001	.004
B. Udrich	6.588 (2): .037	2.415 (2): .299	.005	.001	.009	.08	.001	-.001	.003	.033
W. Szczerbiak	2.799 (2): .247	.137 (2): (.934)	.004	-.002	.010	.035	.001	-.001	.003	.002
R. Nesterovic	2.726 (2): .256	1.323 (2): .516	.003	-.005	.011	.012	.002	-.002	.006	.019
L. Mbah	6.554 (2): .037	.401 (2): .818	.005	.001	.009	.078	.001	-.001	.003	.005

¹⁰ El problema de computar este estadístico es que está basado en la matriz de covarianzas asintótica, la cual puede no producir estimaciones fiables cuando las muestras son pequeñas (Jöreskog y Sörbom, 2001), como ocurre en nuestro caso. Esa corrección en situaciones de no-normalidad también se produce sobre los errores estándar de los parámetros estimados. Se ha comprobado como esos errores producidos por el método robusto son prácticamente idénticos a los producidos por máxima verosimilitud.

B. Diaw	1.399 (2): .497	.296 (2): .862	.002	-.002	.006	.013	-.001	-.003	.001	.003
B. Jackson	8.956 (2): .011*	2.518 (2): .284	.009	.005	.013	.117	.002	.000	.004	.028
A. Nocioni	3.563 (2): .168	3.513 (2): .173	.002	-.004	.008	.011	.002	.000	.004	.037
H. Warrick	4.555 (2): .103	.337 (2): .845	.005	.001	.009	.051	.001	-.003	.005	.005
R. Turiaf	2.123 (2): .346	.567 (2): .753	.003	-.001	.007	.02	-.001	-.005	.003	.007
J. Graham	15.789 (2): .0003*	1.319 (2): .517	.01	.004	.016	.138	-.002	-.006	.002	.016
C. J. Miles	3.130 (2): .209	.228 (2): .892	.005	-.001	.011	.043	.000	-.002	.002	.003
J. Barea	1.810 (2): .405	.181 (2): .914	.004	-.002	.010	.021	.000	-.002	.002	.002
J. Jack	5.451 (2): .065	.963 (2): .618	.005	.001	.009	.054	-.001	-.003	.001	.009
M. Conley	8.009 (2): .018	1.342 (2): .511	.006	.002	.010	.09	-.001	-.003	.001	.013
R. Stuckey	8.498 (2): .014	.033 (2): .984	.007	.001	.013	.101	.000	-.002	.002	.000
M. Evans	24.662 (2): .000*	8.990 (2): .011*	.012	.006	.018	.224	.003	.001	.005	.087
K. Dooling	2.708 (2): .258	.0066 (2): .967	.004	.000	.008	.035	.000	-.002	.002	.001
G. Hill	9.189 (2): .010	.556 (2): .757	.008	.004	.012	.102	.001	-.001	.003	.005
M. Okur	.904 (2): .637	.236 (2): .889	-.003	-.013	.007	.01	-.001	-.003	.001	.004
A. Blatche	.513 (2): .774	.346 (2): .841	.001	-.003	.005	.005	.001	-.003	.005	.003
S. Dalembert	14.195 (2): .0008*	15.024 (2): .0005*	.004	-.002	.010	.042	.007	.003	.011	.149
L. Barbosa	0.845 (2): .655	.005 (2): .997	.004	-.002	.010	.011	.000	.000	.000	.000
J. Thompson	2.069 (2): .355	.653 (2): .721	.003	-.003	.009	.021	.001	-.003	.005	.008
A. Daniels	1.763 (2): .414	1.574 (2): .455	.001	-.005	.007	.002	-.002	-.004	.000	.024
R. Murray	5.566 (2): .061	1.287 (2): .562	.007	-.001	.015	.054	-.001	-.003	.001	.016

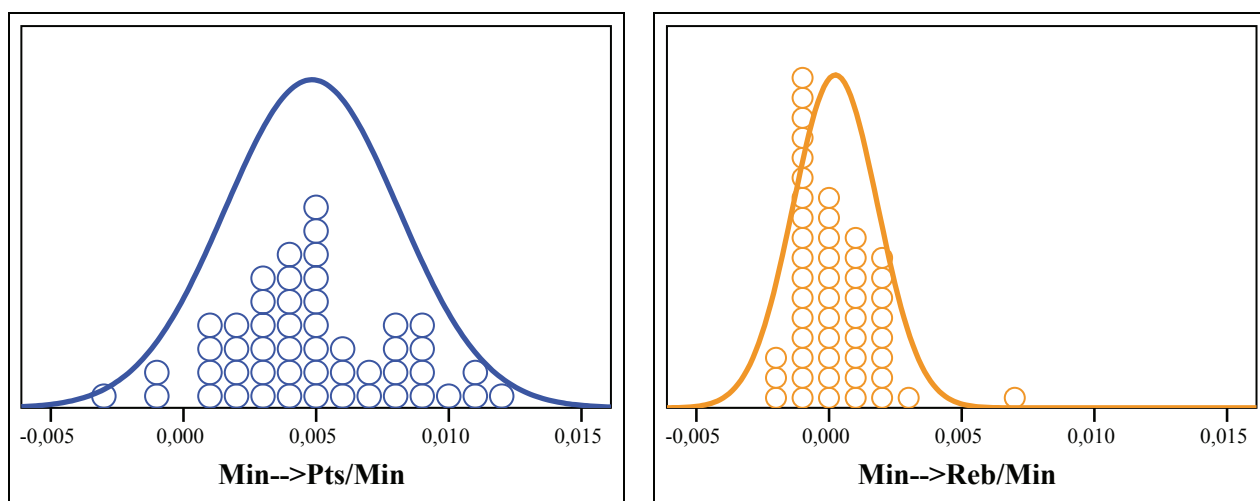
* $p < 0.05$

Los resultados que ilustra la Tabla 1 son muy claros. El modelo que refleja la hipótesis de partida se ajusta en 42 de los 50 casos bajo el criterio de la chi-cuadrado, mientras que en 48 de 50 casos si utilizamos la corrección por curtosis. Por tanto, es evidente que cuando se fijan los coeficientes estructurales de influencia a cero, los modelos se ajustan generalmente bien, por lo que es un claro indicador de que los efectos de esos coeficientes son muy poco importantes. De hecho, al estimar todos los modelos de nuevo liberando esos coeficientes (modelos saturados), los efectos encontrados son muy cercanos a cero, como muestran los intervalos de confianza estimados. Es más, el valor de la R^2 es, como era de esperar, muy bajo también, ya que los minutos jugados explican un porcentaje de varianza prácticamente despreciable de las dos variables dependientes. No obstante, y aún en este contexto de influencia mínima, puede apreciarse una ligera diferencia en cuanto a la importancia del efecto de los minutos en los Pts/Min, con respecto a los Reb/Min. En el primer caso, los efectos son mínimamente mayores que en el segundo.

Esto puede apreciarse mejor en la Figura 4, donde se representan ambas distribuciones de efectos, manteniendo en ambas la misma escala para ilustrar esas diferencias. La distribución de los efectos de los Pts/Min es un poco más dispersa y está centrada sobre el valor de .005. No obstante, hay que insistir en que estos valores son prácticamente despreciables. De hecho un efecto de .005 significa que, si se incrementan los minutos en una unidad sobre la media, entonces se espera que el jugador incremente .005 Pts/Min sobre su media¹¹.

¹¹ Esta interpretación referida a las medias es propia de los modelos de ecuaciones estructurales, en los que existe una reparametrización previa de las ecuaciones de regresión, tomando los valores de las variables como desviaciones sobre su media, con lo que desaparece el tradicional término constante, o *intercept*.

Figura 4. Influencia de los minutos jugados en los Pts/Min y Reb/Min. Distribución de efectos



Análisis exploratorio

En primer lugar, se analizaron los 10 primeros jugadores del ranking de asistencias por minuto, y que han jugado al menos 70 partidos durante la temporada 2008/2009. La Tabla 2 muestra los estadísticos del modelo de regresión lineal por mínimos cuadrados ordinarios. Como puede contemplarse, el efecto sobre la variable dependiente es mínimo, generalmente centrado en cero, y con importancia prácticamente despreciable.

Tabla 2. Influencia de los minutos jugados sobre las asistencias por minuto

Jugador	As/Min	Min → As/min	IC-	IC+	R^2
Steve Nash	.289	-.002	-.005	.002	.012
Chris Paul	.287	-.002	-.006	.003	.007
Rajon Rondo	.249	.000	-.004	.005	.000
Jason Kidd	.243	-.006	-.011	-.001	.057
Sergio Rodriguez	.233	.001	-.004	.006	.002
Ramon Sessions	.208	.002	-.001	.004	.031
Anthony Carter	.204	.000	-.004	.004	.000
Brevin Knight	.203	.002	-.002	.005	.012
Tony Parker	.202	-.001	-.004	.001	.016
Chris Duhon	.195	.003	.001	.006	.075

En segundo lugar, se analizaron los 10 primeros jugadores del ranking de recuperaciones por minuto, y que han jugado al menos 70 partidos durante la temporada 2008/2009. La Tabla 3 muestra los estadísticos del modelo de regresión lineal por mínimos cuadrados ordinarios. Al igual que en el caso de las asistencias por minuto, el efecto es prácticamente cero. No obstante, existe una pequeña diferencia, relativa a que la variable dependiente aquí es más sensible a los cambios. La separación en los puestos del ranking es a veces sensible al tercer decimal, y cambios en un minuto jugado podrían hacer variar ese decimal, y por ende la clasificación.

Tabla 3. Influencia de los minutos jugados sobre las recuperaciones por minuto

Jugador	Rec/Min	Min → Rec/min	IC-	IC+	R ²
Brevin Knight	.072	-.001	-.003	.001	.010
Chris Paul	.072	.001	-.001	.004	.018
Trevor Ariza	.069	.001	-.001	.003	.019
Mario Chalmers	.061	.001	-.001	.002	.012
Sasha Vujacic	.060	-.001	-.004	.002	.004
Dwyane Wade	.057	-.002	-.004	.000	.052
Rajon Rondo	.056	.000	-.002	.001	.002
Jason Kidd	.055	.001	-.001	.004	.020
Anthony Carter	.054	.000	-.001	.002	.002
Ronnie Brewer	.053	.000	-.001	.002	.004

Finalmente, se analizaron los 10 primeros jugadores del ranking de tapones por minuto, y que han jugado al menos 70 partidos durante la temporada 2008/2009. La Tabla 4, nuevamente, muestra los estadísticos del modelo de regresión lineal por mínimos cuadrados ordinarios. Como en los casos anteriores, los efectos son prácticamente despreciables.

Tabla 4. Influencia de los minutos jugados sobre los tapones por minuto

Jugador	Tap/Min	Min → Tap/min	IC-	IC+	R ²
Chris Andersen	.120	.000	-.005	.004	.001
Ronny Turiaf	.099	.000	-.002	.002	.000
Dwight Howard	.082	-.001	-.003	.001	.009
Roy Hibbert	.075	-.005	-.010	-.001	.071
Samuel Dalembert	.072	.001	-.001	.003	.018
Tyrus Thomas	.069	.001	-.001	.002	.010
JaVale McGee	.066	-.001	-.004	.001	.017
Kendrick Perkins	.067	.001	-.001	.003	.008
Louis Amundson	.062	-.001	-.004	.001	.013
Brook Lopez	.060	-.001	-.002	.001	.006

Discusión

Esta investigación ha cubierto un hueco en la literatura académica sobre normalización en baloncesto, mostrando cómo los jugadores de la NBA tienen un rendimiento por minuto que no está asociado linealmente a los minutos en los que están en pista. Aunque esta es la asunción que está detrás de este tipo de normalización, no había sido hasta ahora puesta a prueba, y mucho menos de la manera propuesta en este estudio. Se ha realizado una aproximación estadísticamente formal para estudiar la relación entre los minutos jugados y los puntos por minuto y rebotes por minuto. A través del modelado estadístico, y de una forma confirmatoria, hemos mostrado cómo los efectos son prácticamente despreciables. Bien es cierto que esa distribución de efectos está más centrada en cero en los rebotes por minuto, por lo que en este apartado la conclusión es, si cabe, más aplastante.

En cuanto al resto de categorías: asistencias, recuperaciones y tapones por minuto, la complejidad de las distribuciones para cada jugador ha hecho imposible utilizar la metodología confirmatoria anterior. Ciertamente, sería un error plantear un modelo general para todos los jugadores, dado que el desempeño en esas categorías depende fuertemente de las características de cada jugador, siendo muy prevalente la presencia de ceros dependiendo de la posición que ocupen en el campo. Se ha optado, no obstante, por explorar a los 10 mejores jugadores en cada uno de esos apartados, con el fin de obtener un primer acercamiento a la relación entre las variables. De nuevo, se puede decir que no existe una asociación lineal relevante entre los minutos jugados y el resto de categorías. En el caso de los tapones y

recuperaciones por minuto, al ser acciones de juego con muy poca prevalencia en un partido, esos pequeñísimos efectos encontrados podrían cambiar alguna de las posiciones de los rankings de liderazgo. No obstante, insistimos que esos cambios serían prácticamente inapreciables, y desde el punto de vista de la importancia de los efectos (ver Martínez y Martínez, 2010), no serían relevantes.

Este estudio muestra, por tanto, conclusiones similares a las de Kurylo (2007), aunque el planteamiento metodológico es totalmente diferente. De hecho, las críticas realizadas al trabajo de Kurylo (2007), sobre todo en relación a la no consideración de la varianza intra-jugador, ponen en duda su validez. Por ello, son importantes los resultados que hemos mostrado para este campo de estudio, como una aproximación novedosa y formal al estudio de la normalización de indicadores de rendimiento en baloncesto.

Por tanto, y con las reservas propias en las categorías de asistencias, recuperaciones y taponos por minuto, se pueden extraer dos conclusiones principales de este estudio. En primer lugar, el rendimiento de los jugadores reflejado en estas cinco categorías estadísticas no depende linealmente de los minutos que los jugadores están en pista. Esta conclusión, como ya se ha indicado, es mucho más consistente para los puntos y rebotes por minuto. Así, por ejemplo, los jugadores podrían conseguir sus mismos “números” normalizados jugando 10 o 30 minutos, de forma indistinta. Esta consecuencia es atractiva para los entrenadores, que pueden utilizar este factor para repartir los minutos en cancha con un mejor criterio. En segundo lugar, los rankings realizados utilizando este tipo de normalización ofrecen un dibujo “justo” de la realidad del desempeño de los jugadores, ya que permiten una comparación en igualdad de condiciones para aquellos jugadores que, por ejemplo, juegan 15 minutos de media, frente a los que juegan 35. Las consecuencias de marketing también son evidentes: Los líderes de cada categoría normalizada “por minutos” son “merecedores” de esa distinción, mientras que para aquellos en los que la normalización es “por partido”, esas clasificaciones son más discutibles (Martínez y Martínez, 2010).

Es preceptivo indicar, no obstante, que desde el punto de vista de los analistas de los equipos de baloncesto, que como hemos comentado son una figura de alto aprecio en los *staff* actuales, los análisis deben de hacerse jugador a jugador, para profundizar en las especificidades de cada uno de ellos, y ver si existe algún jugador en el equipo que sistemáticamente realice actuaciones dependientes de los minutos que esté en pista. Del mismo modo, se pueden elaborar predicciones sobre el futuro desempeño de ciertos jugadores en base a un incremento alto de minutos en cancha. Por ejemplo, Drazen Petrovic consiguió prácticamente idénticos puntos por partido cuando jugaba 12.6 minutos de media en su primera temporada en Portland, (1988/1989) que cuando jugaba 38.0 de media en su última temporada en Nueva Jersey (1992/1993). Sin embargo, la importancia que para un equipo tiene un jugador que es capaz de mantener su gran anotación por minuto cuando juega tres veces más minutos es capital, ya que la contribución al global equipo se vuelve muchísimo más importante.

Aún así, futuras líneas de investigación deben profundizar sobre aspectos que este estudio ha dejado un poco en el aire. Además del planteamiento de modelos estadísticos que consideren la aparición de tantos “ceros” en las distribuciones de los jugadores, habría que explorar la posible existencia de relaciones no lineales entre las variables. La existencia de curvas no lineales que reflejen la asociación entre las variables podría ser plausible. De hecho, es una característica de los rendimientos decrecientes. Aún así, en este tipo de curvas existiría una parte lineal que debería haber sido detectada, al menos mínimamente, por los procedimientos lineales que hemos utilizado. No ha sido el caso, lo que hace pensar que tampoco existan relaciones no lineales. No obstante, es indicado sugerir que se explore empíricamente en futuros estudios.

La extensión de las consecuencias de este estudio a otras ligas no parece aventurada. Así, en principio, no tiene porqué haber cambios si se analizan otras competiciones fuera de la NBA. Sin embargo, puede ser otra idea para la investigación futura, en aras de confirmar que, efectivamente, la normalización por minutos no depende de la competición bajo estudio.

Asimismo, también se podrían analizar otras categorías estadísticas de indudable interés, como los balones perdidos, la eficacia en el tiro, o los índices de valoración global de cada jugador, como el PER, TENDEX, etc. Ya se ha comentado que estos índices se construyen a través de las categorías individuales. Una vez que se conocen que cada categoría no está asociada a los minutos jugados, entonces se podría estudiar, utilizando este mismo procedimiento, si una combinación lineal (caso del TENDEX, NBA *Efficiency*, o Valoración ACB) o no lineal de esas categorías (caso del PER) está asociada a los minutos jugados.

Finalmente, sería de sumo interés para futuros estudios sobre normalización, el analizar la relación existente entre el número de posesiones por partido y las categorías estadísticas. La posesión es un concepto fundamental en los procesos de normalización en el baloncesto actual (Kubatko, Oliver, Pelton y Rosenbaum, 2007), por lo que se convierte en una atractiva línea de investigación.

Agradecimientos

Mi más sincero agradecimiento a Dick Mays, por facilitarme información sobre sus aportaciones al análisis de la productividad de los jugadores.

Referencias

- Arkes, J. (2010). Revisiting the Hot Hand Theory with free throw data in a multivariate framework. *Journal of Quantitative Analysis in Sports*, 6 (1), Article 2.
- Alferink, L. A., Critchfield, T. S., Hitt, J. L., y Higgins, W. J. (2009). Generality of the matching law as a descriptor of shot selection in basketball. *Journal of Applied Behavior Analysis*, 42 (3), 595-608.
- Balsdon, E., Fong, L., y Thayer, M. (2007). Corruption in College basketball? Evidence of tanking in postseason conference tournaments. *Journal of Sports Economics*, 8 (1), 19-38
- Berri, D. J. (1999). Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association. *Managerial and Decision Economics*, 20, 411-427.
- Berri, D. J. (2008). A simple measure of worker productivity in the National Basketball Association. in *The Business of Sport*, eds. Brad Humphreys and Dennis Howard, editors, 3 volumes, Westport, Conn.
- Berri, D. J., y Bradbury, J. C. (2010). Working in the land of metricians. *Journal of Sports Economics*. 11 (1), 29-47.
- Berri, D. J., Brook, S. L., Frick, B, Fenn, A. J., y Vicente-Mayoral, R. (2005) The short supply of tall people: competitive imbalance and the National Basketball Association. *Journal of Economic Issues*, 39 (4), 1029-1041
- Berri, D. J., Brook, S. L., y Schmidt, M. B. (2007). Does One Simply Need to Score to Score? *International Journal of Sport Finance*, 2 (4), 190-205
- Berri, D. J., y Eschker, E. (2005). Performance when it counts? The myth of the prime time performer in professional basketball. *Journal of Economic Issues*, 1, 798-807
- Berry, S. M, Reese, C. S., y Larkey, P. L. (1999) Bridging different eras in sports. *Journal of the American Statistical Association*, 94, 661-686.
- Berri, D. J., y Schmidt, M. B. (2002). Instrumental versus bounded rationality: a comparison of Major League Baseball and the National Basketball Association. *Journal of Socio-Economics*, 31, 191-214.
- Berri, D. J., y Schmidt, M. B. (2006),. On the road with the National Basketball Association's superstar externality. *Journal of Sports Economics*, 7 (4), 347-358.
- Bodvarsson, O. B., y Brastow, R. T. (1998). Do employers pay for consistent performance?: Evidence from the NBA. *Economic Inquiry*, 36 (1), 145-160.
- Esteller-Moré, A., y Eres-García, M. (2002). A note on consistent players' valuation. *Journal of Sports Economics*, 3 (4), 354-360.
- Felipo, J. (2005). *Fórmulas para ganar: La revolución estadística del basket*. Barcelona: Zona 131.
- Fernández, J., Camerino, O., Anguera, M. T., y Jonsson, G. K. (2009). Identifying and analyzing the construction and effectiveness of offensive plays in basketball by using systematic observation. *Behavior Research Methods*, 41, 719-730.
- Fort, R., y Maxcy, J. (2003). Competitive balance in sports leagues: An introduction. *Journal of Sports Economics*, 4 (2), 154-160.
- Fort, R., Hoon-Lee, Y., y Berri, D. J. (2008). Race, technical efficiency and retention. The case of NBA coaches. *International Journal of Sport Finance*, 3, 84-97.
- Gómez, M. A. y Lorenzo, A. (2005). Diferencias entre equipos ganadores y perdedores en el rendimiento de competición en baloncesto femenino. *Kronos*, 8, 16-19.
- Hayduk, L. A. (1996). *LISREL Issues, Debates and Strategies*. Ed. Johns Hopkins University Press, Baltimore, MD.
- Hayduk, L. A., Cummings, G., Boadu, K., Pazderka-Robinson, H., y Boulianne, S. (2007). Testing! Testing! one, two, three Testing the theory in structural equation models!. *Personality and Individual Differences*, 42 (5), 841-850.
- Hitchcock, C. (2002). *Probabilistic causation*. Stanford Encyclopedia of Philosophy.
- Hitt, J. L., Alferink, L. A., Critchfield, T. S., & Wagman, J. B. (2007). Choice behavior expressed in elite sport competition: Predicting shot selection and game outcomes in college basketball. In L. A.

- Chiang (Ed.), *Motivation of exercise and physical activity* (pp.79–91). Hauppauge, NY: Nova Science.
- Hollinger, J. (2005). *Pro Basketball Forecast*. Washington, D.C.: Potomac, Inc.
- Hoon-Lee, Y., y Berri, D. J. (2008). A re-examination of production functions and efficiency estimates for the National Basketball Association. *Scottish Journal of Political Economy*, 55 (1), 51-66.
- Hughes, M. (2004). Notational analysis: a mathematical perspective. *International Journal of Performance Analysis in Sport*, 4 (2), 97-139
- Humphreys, B. R. (2000). Equal pay on the hardwood: the earnings gap between male and female NCAA. Division I Basketball coaches. *Journal of Sports Economics*, 1 (3) 299-307.
- Humphreys, B. R. (2002). Alternative measures of competitive balance in sports leagues. *Journal of Sports Economics*, 3 (2), 133-148
- Ibáñez, S. J., Sampaio, J., Sáenz-López, P., Jiménez, J. y Janeira, M. A. (2003). Game statistics discriminating the final outcome of junior world basketball championship matches (Portugal 1999). *Journal of Human Movement Studies*, 74, 1-19
- Jöreskog, K. y Sörbom, D. (2001). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.
- Jöreskog, K. y Sörbom, D. (2006). *LISREL 8.80*. Scientific Software International, Inc
- Kubatko, J.; Oliver, D., Pelton, K, y Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, 3 (3), Article 1.
- Kurylo, M. (2007, 17 Septiembre). One more nail in the anti-per minute argument's coffin? Mensaje publicado en el blog KnickerBlogger.net: <http://www.knickerblogger.net/?p=587>
- Lewis, M. M. (2003) *Moneyball: The art of winning an unfair game*. W.W. Norton & Company Inc.
- Martínez, J. A. y Martínez, L. (2009). Un método para reescribir las estadísticas individuales en la Liga ACB de baloncesto. *Working paper*. Universidad Politécnica de Cartagena.
- Martínez, J. A. y Martínez, L. (2010). Un método probabilístico para las clasificaciones estadísticas de jugadores en baloncesto. *Revista Internacional de Ciencias del Deporte*, 18 (6), 13-36.
- Mavridis, G., Tsamourtzis, E., Karipidis, A., & Laios, A. (2009). The inside game in World Basketball. Comparison between European and NBA teams. *International Journal of Performance Analysis of Sport*, 9, 157-164.
- Michaelides, M. (en prensa). A new test of compensating differences: Evidence on the importance of unobserved heterogeneity. *Journal of Sport Economics*.
- NBA.COM (2009, 23 Octubre). NBA dives headlong into new era of statistical analysis. Descargado el día 24 de Octubre desde http://www.nba.com/2009/news/features/john_schuhmann/10/23/stats.analysis/index.html#
- Oliver, D. (2004). *Basketball on paper. Rules and tools for performance analysis*. Washington, D. C.: Brassey's, INC.
- Perše, M., Kristan, M., Kovacic, S., Vuckovic, G., y Perš, J. (2009). A trajectory-based analysis of coordinated team activity in a basketball game. *Computer Vision and Image Understanding*, 113, 612–621
- Price, J., y Wolfers, J. (2007). Racial discrimination among NBA referees. *Quarterly Journal of Economics* (bajo revisión)
- Reed, D., y O'Donoghue, P.G. (2005), Development and application of computer-based prediction methods. *International Journal of Performance Analysis of Sport (e)*, 5 (3), 12-28.
- Rimler, M. S., Song, S., y Yi, D. T. (en prensa). Estimating production efficiency in men's NCAA college basketball: A bayesian approach. *Journal of Sports Economics*. doi:10.1177/1527002509337803
- Romanowich, P., Bourret, J., y Vollmer, T. R. (2007). Further analysis of the matching law to describe two- and three-point shot selection by professional basketball players. *Journal of Applied Behavior Analysis*, 40, 311–315.
- Piette, J., Sathyanarayan, A., y Kai, Z. (2010). Scoring and shooting abilities of NBA players. *Journal of Quantitative Analysis in Sports*, 6 (1), Article 1.
- Sánchez, J. M., Castellanos, P., y Dopico, J. A. (2007). The winning production function: Empirical evidence from Spanish basketball. *European Sport Management Quarterly*, 7 (3), 283-300.
- Satorra, A., y Bentler, P.M., 1999. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66, 507–514.
- Skinner, B. (2010). The price of anarchy in basketball. *Journal of Quantitative Analysis in Sports*, 6 (1), Article 3.
- Vermunt, J. K., y Magidson, J. (2005) *Latent GOLD 4.0 User's Guide*. Belmont, Massachusetts: Statistical Innovations Inc
- Tauer, J. M., Guenther, C. L., y Rozek, C. (2009). Is there a home choke in decisive playoff basketball games?. *Journal of Applied Sport Psychology*, 21, 148-162.

- Trininic, S., Dizdar, D., y Luksic, E. (2002). Differences between winning and defeated top quality basketball teams in final tournaments of European club championship. *Collegium Antropologicum*, 26 (2), 521-531.
- Vollmer, T. R., y Bourret, J. (2000). An application of the matching law to evaluate the allocation of two and three-point shots by college basketball players. *Journal of Applied Behavior Analysis*, 33, 137-150.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57 (2), 307-333.
- Winston, W. L. (2009). *Mathletics*. New Jersey: Princeton University Press
- Zimmer, T., y Kuethe, T. H. (2007). Testing for bias and manipulation in the National Basketball Association playoffs. *Journal of Quantitative Analysis in Sports*, 5 (3), Article 4.
- Zhang, D., Gatica-Perez, D., Bengio, D., y Roy, D. (2005). Learning Influence among Interacting Markov Chains. *Neural Information Processing Systems*, December 5-8, 2005, Vancouver, British Columbia, Canada.

Webs:

<http://bama.ua.edu/cgi-bin/wa?A0=SEMNET>

<http://basketball-statistics.com/howpacworks.html>

<http://sonicscentral.com/apbrmetrics/viewtopic.php?t=85&postdays=0&postorder=asc&start=45>

www.bellotibasketball.com

www.dougstats.com

www.draftexpress.com

www.espn.com,

www.nbastuffer.com

www.neuralbet.com

www.virtualgold.com

www-03.ibm.com/press/us/en/pressrelease/1190.wss