REVISTA INTERNACIONAL DE CIENCIAS DEL DEPORTE
International Journal of Sport Science

Rev. int. cienc. deporte

International Journal of Sport Science
VOLUMEN VI - AÑO VI
Páginas:13-36 ISSN:1885-3137
Nº 18 - Enero - 2010

doi:10.5232/ricyde2010.01802

Un método probabilístico para las clasificaciones estadísticas de jugadores en baloncesto.

A probabilistic method to statistically classify players in basketball.

José Antonio Martínez García
Laura Martínez Caro
Facultad de Ciencias de la Empresa. Universidad Politécnica de Cartagena. España

Resumen

En esta investigación presentamos una nueva forma de interpretar las estadísticas individuales de la Liga ACB de baloncesto. Para ello, proponemos un enfoque probabilístico de los números individuales obtenidos por cada jugador al final de la temporada regular. Esto convierte a cada valor conseguido en un estimador del valor real teórico, por lo que tiene un error asociado que, en función de su magnitud, influye en los rankings de líderes estadísticos de la ACB. Asimismo, realizamos una aproximación paramétrica para cuantificar un tamaño de error máximo admisible, que debe servir como criterio para considerar si un jugador debe ser incluido en los rankings de cada apartado estadístico. Dada la importancia creciente que la utilización de la estadística está teniendo en el baloncesto profesional, este método presenta una contribución novedosa al análisis del desempeño de los jugadores, análisis que repercute en su valor económico y mediático, es decir, en el valor de mercado de éstos.

Palabras clave: estadísticas: baloncesto; liga ACB; valor de mercado.

Abstract

We introduce a method to re-elaborate the rankings of individual stats in the ACB League. The method is based on a probabilistic approach to interpret the individual performance achieved by each basketball player at the end of the regular season. Therefore, each individual record is an estimate of the real value of the parameter, with the corresponding associated error. The size of this error influences the final elaborated ranking. Under a parametric approach, we quantify the size of a maximum admissible error. Because of the growing interest of statistics in basketball, our proposal is a valuable contribution to analyse the performance of players, which is highly related to their market value.

Key words: statistics; basketball; ACB league; market value.

Correspondencia/correspondence: José Antonio Martínez García
Universidad Politécnica de Cartagena.Facultad de Ciencias de la Empresa.
Paseo Alfonso XIII, 50. 30203. Cartagena. España.
E-mail: josean.martinez@upct.es

Recibido el 29 de julio de 2009; Aceptado el 19 de octubre de 2009

Introducción

La importancia de las estadísticas en los deportes profesionales es cada vez más prominente, sobre todo desde la aparición de "Moneyball" (Lewis, 2003), un famoso libro que narra la historia de un equipo de béisbol que utiliza enfoques estadísticos novedosos para ayudar en la toma de decisiones directivas y conseguir el éxito deportivo.

En baloncesto también se está adoptando una orientación más analítica del juego, basada en diferentes tratamientos estadísticos de los datos recogidos en cada partido, con el fin de tener una aproximación más objetiva a la realidad del mismo. No cabe duda que "Moneyball" ha tenido una influencia destacada, multiplicándose en los últimos años la aparición de webs especializadas y foros de discusión fundamentados en las estadísticas, como por ejemplo: www.winsproduced.com, www.apbrmetrics.com, www.82games.com, www.basketball-reference.com, etc. En España, destacamos el trabajo de Felipo (2005), donde se recoge buena parte de las nuevas tendencias de análisis provenientes de Estados Unidos.

Además de "Moneyball" (Lewis, 2003), o el reciente "Mathletics" (Winston, 2009), existen influyentes referencias en el ámbito exclusivo del baloncesto, donde investigadores y analistas han contribuido publicando libros como "Basketball on Paper" (Oliver, 2004), "Pro Basketball Forecast" (Hollinger, 2005), o "The Wages of Wins" (Berri, Schmidt, y Brook, 2006). En cuanto a revistas de índole científico, la creación del Journal of Quantitative Analysis in Sports en 2005 ha supuesto un vehículo de comunicación importante para los investigadores, destacando trabajos como los de Page, Fellingham y Reese (2007), Kubatko, Oliver, Pelton y Rosenbaum (2007), o West (2006). Además, otras publicaciones prestigiosas han recogido trabajos similares, como los de Berri (1999; 2008), Cooper, Ruiz y Sirvent (2009), o McGoldrick y Voeks (2005).

Todo este "boom estadístico" influye de manera importante en el valor de mercado de los jugadores. Muchos de los sistemas estadísticos creados están enfocados a la valoración del desempeño del jugador a través de diferentes índices (en www.nbastuffer.com pueden consultarse muchos de ellos). No existe homogeneidad en el uso de los índices de valoración de jugadores entre las diferentes ligas, quienes de forma oficial, proveen sistemas de valoración que difieren en la fórmula de cálculo. Así, por ejemplo, los sistemas de las ligas española, griega, o francesa, difieren entre sí.

Independientemente del criterio seguido para la formación de ese índice de valoración, en muchos casos ese índice es usado para conceder premios individuales. Actualmente, en la Liga ACB, el "Jugador de la Jornada" es un honor que recae sobre el jugador mejor valorado estadísticamente. Hasta hace unos pocos años, concretamente hasta la temporada 2003/2004, el "Jugador más Valioso" o MVP de la temporada era otorgado también por criterios puramente estadísticos, aunque a partir de ese año, esa concesión es responsabilidad de un jurado de expertos. No obstante, las valoraciones estadísticas son un factor clave en la elección de esos expertos. Lo mismo ocurre con el MVP del mes, donde el criterio estadístico fue sustituido en la temporada 2008/2009.

Esos galardones influyen en la proyección mediática del jugador, ya que el eco de esa noticia se difunde a través de todos los medios especializados. Esto, y usando la terminología acuñada por los investigadores Francesc Pujol y Pedro García del Barrio (www.unav.es/econom/sport), incrementa su valor mediático, lo que puede redundar en una revalorización económica del jugador, entendiendo ésta como un incremento del caché del mismo.

Pero no sólo las estadísticas que apuntan a la valoración del jugador son importantes. Las estadísticas individuales (puntos, rebotes, asistencias, etc.) que son las que construyen esos índices de valoración, tienen un protagonismo particular. Así, en todas las competiciones aparecen las clasificaciones de los principales apartados estadísticos individuales. En la ACB, además, se denominan "Triunfadores estadísticos" a quienes vencen en cada categoría, teniendo también su relevancia mediática, la cual se disemina entre los medios especializados, que su vez potencian la noticia. En la Euroliga, por ejemplo, se otorga el "Trofeo Alphonso Ford" al máximo anotador de la competición, teniendo el premio una cobertura mediática similar al de MVP. En el portal global de baloncesto más importante, www.eurobasket.com, se muestran los primeros puestos del ranking de puntos, rebotes, asistencias, recuperaciones y pérdidas de todas las ligas del mundo donde existen datos disponibles. Dado que uno de los objetivos de ese portal es que sirva como referencia a agentes, entrenadores y directivos para contratar jugadores, las clasificaciones estadísticas se convierten en un aspecto importante en la revalorización de los mismos. Estadísticas similares son también reportadas por www.keyhoops.com, aunque sólo referidas a las principales ligas europeas y norteamericanas.

De este modo, puede existir una diferencia notable entre quedar el primero o no en cualquier clasificación estadística, ya que la trascendencia mediática-profesional-social para el jugador puede variar. A pesar de la importancia de este hecho, y de la relevancia comentada de las estadísticas en el baloncesto actual, creemos que el método actual de clasificación de estadísticas individuales es mejorable. Para ello, proponemos un método basado precisamente en los principios de inferencia estadística, donde el valor de la puntuación del jugador en cada apartado estadístico se considera un estimador del valor real del parámetro en cuestión. Así, para un jugador que dispute todos los partidos de la liga regular, el estadístico derivado de la muestra coincidirá con el parámetro poblacional, sin ningún error asociado. Sin embargo, para jugadores que disputen menos partidos de los que corresponderían a la competición en cuestión, la estimación del parámetro poblacional tendría asociado un error. Del tamaño de ese error va a depender que la clasificación final varíe o no con respecto a la ofrecida por las estadísticas oficiales.

El objetivo de esta investigación es proponer un método general de estimación de las estadísticas individuales de cada jugador al final de la temporada, con el fin de buscar un criterio probabilístico para construir la clasificación en cada apartado estadístico. Como veremos, la aplicación de este método en las clasificaciones de la ACB para la temporada 2008/2009 produce variaciones importantes en algunas de ellas, lo que indica que algunos jugadores podrían haberse beneficiado desde el punto de vista profesional (valor económico), haciendo los mismos o menores méritos que otros jugadores que no habrían obtenido esa recompensa.

Descripción del método

Sea $\{X_1...X_j...X_k\}$ el conjunto de las k características de una población N a estudiar, por ejemplo los puntos o las asistencias que un jugador promedia por temporada. El tamaño de la población es igual al número de partidos de los que se compone la competición sobre la que pretende realizar la clasificación estadística. En el caso de la

ACB, en la última temporada (2008/2009) el tamaño de $N = 32$. Para cada jugador, la característica X_j sigue una distribución F_j de probabilidad caracterizada por los parámetros media (μ_j) y desviación típica (σ_j). De este modo, $X_j \quad F_j(\mu_j, \sigma_j)$ es una variable aleatoria continua, cuya forma de distribución, media y desviación típica son en principio desconocidos, y característicos de cada jugador.

Los partidos realmente disputados por el jugador constituyen la muestra n, siendo necesariamente $n \leq N$. Cada valor x_{ij} de la muestra $(x_{1j} \ldots x_{ij} \ldots x_{nj})$ se distribuye idénticamente a X_j, siendo $f_j(x_{ij})$ el valor de su función de densidad, y $F_j(x_{ij})$ el valor de su función de distribución. La función de densidad conjunta es $f_j(x_{1j} \ldots x_{ij} \ldots x_{nj}) = \prod_{i=1}^{n} f_j(x_{ij})$ y la función de distribución acumulada $F_j(x_{ij}) = \int_{-\infty}^{x_{ij}} f_j(x_j)dx_j$ indica la probabilidad de que la característica poblacional sea menor o igual que el valor x_{ij}.

De este modo, podemos definir el conjunto q de jugadores $\{P_1 \ldots P_p \ldots P_q\}$ de la Liga ACB, donde cada jugador P_p tiene asociado el conjunto de características a estudiar, las cuales tienen una distribución desconocida y caracterizada por una media y una varianza. Además, cada jugador compite en la misma liga, que tiene un número finito de partidos, de los que disputa una muestra menor o igual a ese número, y de la que se obtienen las realizaciones muestrales. Formalmente (1):

$$P_p(X_k; F_j; \mu_j; \sigma_j; N; n; x_{ij}) \forall_{\substack{p=1\ldots q \\ j=1\ldots k \\ i=1\ldots n}} \tag{1}$$

Una vez definido el conjunto de las k características a estudiar, el valor de N es evidentemente conocido e idéntico para todos los jugadores, siendo n_p también conocido (el número de partidos que disputa cada jugador). Los valores de x_{ijp} son los datos individuales que reporta la Liga ACB para cada jugador, y por tanto también están determinados.

Llegados a este punto, estamos interesados en conocer μ_{jp}, ya que las clasificaciones de cada apartado estadístico son un ranking descendente de μ_{jp}. La mejor forma de estimar μ_{jp} es a través de $\overline{x}_{jp}$, es decir, a través del valor esperado de la distribución muestral. Cuando $n = N$, $\overline{x}_{jp} = \mu_{jp}$, es decir, la distribución de la población y la muestra es idéntica, por lo que la media muestral coincide perfectamente con la media poblacional. Sin embargo, cuando $n \leq N$, $\overline{x}_{jp}$ es el estimador máximo verosímil e insesgado de μ_{jp}, pero existe un error E_{jp} asociado a la estimación. Ese error depende de la confianza C de la estimación, de la desviación típica poblacional σ_{jp}, y de los tamaños de la muestra y de la población. Así, se puede construir un intervalo de confianza (L_{jp}, U_{jp}) aleatorio sobre μ_{jp} de tal forma (2):

$$\Pr(L_{jp} \leq \mu_{jp} \leq U_{jp}) = C \tag{2}$$

siendo $L_{jp} = \overline{x}_{jp} - E_{jp}$ y $U_{jp} = \overline{x}_{jp} + E_{jp}$. Estableciendo un valor para C, usualmente 95%, el intervalo de confianza indica que contendrá al parámetro poblacional 95 de cada 100 muestras aleatorias. El par (L_{jp}, U_{jp}) es también una variable aleatoria, por lo que la delimitación de los intervalos de confianza depende de la muestra que se analice. Una consecuencia de la interpretación anterior es que cualquier valor constante poblacional θ que se quiera comparar con μ_{jp}, será considerado estadísticamente igual si $(L_{jp} \leq \theta \leq U_{jp})$.

Sin embargo, si la comparación se realiza con otro estadístico θ_{jp} que es un estimador de un parámetro poblacional con $E_{jp} > 0$, se requiere un test estadístico para determinar si existen diferencias entre ambos.

De este modo, si se quieren comparar las estadísticas de dos jugadores $\{P_1, P_2\}$ para una categoría k determinada del juego se pueden dar dos opciones: (a) comparación directa de las medias poblacionales si $n_1 = n_2 = N$; (b) comparación estadística entre las medias cuando $n_1 \vee n_2 \neq N$.

El problema de este enfoque estadístico es que E_{jp} se incrementa a medida que la muestra es más pequeña y la desviación típica más grande. Como el nivel de confianza conviene que se establezca igual para todas las estimaciones, y N es invariante, el tamaño del error en el caso de nuestro estudio depende únicamente de n_p y σ_{jp}. Errores grandes no son permisibles, porque hacen que el intervalo de confianza de la estimación sea más ancho, lo que perjudicaría la claridad de las comparaciones entre jugadores. Este hecho haría que, desde el punto de vista estadístico, no se tuviese potencia suficiente para detectar diferencias entre jugadores.

Por tanto, lo ideal sería calcular un tamaño de muestra mínimo para acotar el error cometido en la estimación en un máximo admisible. Como la desviación típica es específica para cada jugador y característica a analizar, sería arriesgado tomar como referencia una desviación poblacional promedio. De este modo, se necesita estimar σ_{jp}.

Estimación del error

Como puede adivinarse, el problema que se plantea en esta investigación es simplemente un problema de inferencia estadística y test de diferencia de medias. Esto quiere decir que las opciones para resolver este problema son las mismas que cualquier problema similar de inferencia. Básicamente, se pueden distinguir dos tipos de escenarios: (1) la población es normal, es decir, $F_{jp} \equiv Normal$; (2) la distribución de la población no es normal.

Para el primer caso, la opción más adecuada es estimar σ_{jp} a través de la desviación típica muestral S_{jp}, y usar la distribución t de Student para computar el error cometido y calcular los intervalos de confianza (Levy y Lemeshow, 2003). De este modo, el tamaño del error sería (3):

$$E_{jp} = t_c(SE_{jp}) = t_c \frac{S_{jp}}{\sqrt{n_{jp}}} \sqrt{\frac{N - n_{jp}}{N}} \quad (3)$$

siendo SE_{jp} el error estándar de la media $\overline{x}_{jp}$. Cuando la distribución se aleja de la normalidad (segundo caso), usualmente se recomienda la estimación por remuestreo de $\overline{x}_{jp}$ y S_{jp} (Hesterberg, Moore, Monaghan, Clipson y Epstein, 2005), pudiendo computar un intervalo de confianza utilizando varios procedimientos, como los percentiles de la distribución empírica, por ejemplo. De este modo, el valor de E_{jp} diferiría para la cota inferior y superior del intervalo de confianza, que podría no ser simétrico respecto a la media. No obstante, los métodos de remuestreo tradicionalmente implementados en el marco de población infinita, como el *bootstrapping* o el *jacknife,* producirían estimaciones poco precisas, por lo que debería de realizarse ajustes procedimentales en el marco de población finita y con muestreo sin reemplazamiento (Lombardía, González-Manteiga y Prada-Sánchez, 2004).

Cuando $n_1 \vee n_2 \neq N$ se requiere un contraste de hipótesis sobre las medias de los dos jugadores. En el caso más simple, cuando ambas distribuciones poblacionales son normales, se aplica el test de diferencia de medias, y su variante en el caso de heterocedasticidad (Lumley, Diehr, Emerson y Chen, 2002; Sawilowsky, 2002). Cuando las poblaciones no son normales, entonces el test de Wilcoxon-Mann-Whitney es una opción interesante, el cual mide la diferencia entre las medianas de las dos distribuciones de datos.

En todas estas situaciones descritas, el tamaño de la muestra también es importante, ya que, gracias a la aplicación del Teorema Central del Límite, se puede obtener la estimación de E_{jp}, y por ende de los intervalos de confianza, de forma aproximada para muestras mayores de 30. Esta es una recomendación muy general, ya que los intervalos de confianza pierden robustez a medida que la muestra disminuye en su acercamiento a 30. Aunque existen investigaciones que recomiendan el uso del *boostrapping* no paramétrico y del test de Wilcoxon-Mann-Whitney cuando se dan situaciones de no normalidad (Sawilowsky, 2002; 2005), existen también bastantes evidencias sobre la robustez de la aproximación paramétrica, a través de la utilización de la distribución t-Student en condiciones no extremas y con tamaños de muestra por debajo de 30 (Lumley, et al. 2002; Mora y Solomon, 2002), por lo que es una opción atractiva y fácilmente interpretable si los datos no se desvían exageradamente de la normalidad.

Dado que esta situación de no extrema normalidad es la más previsible (como veremos posteriormente en la exploración de datos), la tomaremos como marco de referencia para la estimación del tamaño de muestra mínimo para cometer un error máximo prefijado. Básicamente, el problema se reduce a computar la diferencia máxima admisible $A_{t=0}$ entre el valor de $\overline{x}_{jp}$ y μ_{jp} (4):

$$\Pr(|\,\overline{x}_{jp} - \mu_{jp}\,| \leq A_{j,t=0}) = C \quad (4)$$

Utilizando un poco de álgebra (Casas, 1997) podemos llegar a la siguiente expresión (5):

$$\Pr\left[-\frac{A_{j,t=0}}{S_{jp}/\sqrt{n_{jp}}}\left[\sqrt{\frac{N-n_{jp}}{N}}\right]^{-1} \le t_{\alpha,\upsilon} \le \frac{A_{j,t=0}}{S_{jp}/\sqrt{n_{jp}}}\left[\sqrt{\frac{N-n_{jp}}{N}}\right]^{-1}\right] = C \tag{5}$$

donde $t_{\alpha,\upsilon}$ es el valor de la distribución *t*-Student para una confianza $C = 1-\alpha$ y los grados de libertad υ. Como el valor de S_{jp} es específico para cada jugador y categoría del juego, se hace muy complicado establecer un criterio estándar, por lo que la cuantificación de n_{jp} va a depender de cada situación específica. Lo que sí se puede prefijar es el valor de la diferencia máxima admisible, $A_{j,t=0}$, y para ello proponemos el siguiente criterio:

Creemos oportuno cuantificar como un valor tolerable de error el 5% del valor numérico correspondiente a los líderes de cada categoría estadística, siempre y cuando éstos hubieran disputado todos los partidos de la competición. Sobre un horizonte temporal móvil, que podríamos establecer en 4 años, se calcularía el valor medio de $\mu_{j,t=0}$, de la siguiente forma (6):

$$\mu_{j,t=0} = \frac{\mu_{j,t=-1} + \mu_{j,t=-2} + \mu_{j,t=-3} + \mu_{j,t=-4}}{4} \tag{5}$$

Por ejemplo, en la categoría de "máximos anotadores", para calcular el valor de $\mu_{j,t=0}$, de la temporada 2008/2009, promediaríamos las anotaciones de los jugadores: Lou Roe (20.9 puntos, 3º en 2004/2005), Pete Mickeal (19 puntos, 2º en 2005/2006), Louis Bullock (15.4 puntos, 4º en 2006/2007), y Joseph Gomis (18 puntos, 2º en 2007/2008). Así, el valor de $\mu_{k,t=0} = 18.32$. De este modo (6):

$$A_{j,t=0|2008/2009} = G\mu_{j,t=0|2008/20009} \tag{6}$$

siendo G el valor tolerable de error. Así, para $G = 5\%$ $A_{j,t=0|2008/2009} = 5\%\mu_{j,t=0|2008/20009} = 0.92$. Si fuéramos más restrictivos con la precisión de las estimaciones, por ejemplo, al 1%, entonces $A_{j,t=0|2008/2009} = 1\%\mu_{j,t=0|2008/2009} = 0.18$

Una vez calculado el valor del error máximo admisible para cada categoría estadística, se podría calcular el tamaño de muestra mínimo para cada jugador, en función de la estimación de su varianza para cada categoría. Para ello, habría que computar n_{jp} de la siguiente forma (7):

$$n_{jp} = \frac{t^2_{\alpha,\upsilon} S^2_{jp} N}{(A^2_{j,t=0} N) + (t^2_{\alpha,\upsilon} S^2_{jp})} \tag{7}$$

La resolución de (7) puede dar un valor decimal para n_{jp}, por lo que habría que escoger el valor íntegro más cercano. Un aspecto a destacar es que el valor del estadístico t-Student depende de los grados de libertad υ, que a su vez dependen de n_{jp}, ya que

$\upsilon = n_{jp} - 1$. Esta circularidad queda resuelta al tomar un valor fijo de $t_{\alpha,\upsilon} = 2.04$, que es el que corresponde al valor $n_{jp} = N - 1$ cuando $N = 32$, caso de la temporada 2008/2009. Como el valor $t_{\alpha,\upsilon}$ es bastante estable para un nivel de confianza del 95% y sobre una variación moderada de n_{jp} (sólo varía en 5 centésimas como máximo cuando n_{jp} cae hasta 20), podemos tomarlo como fijo, ya que es previsible que el tamaño de muestra mínimo para que el error cometido sea admisible sea siempre superior a 20.

Distancia entre jugadores para establecer el ranking

Una vez realizada las estimaciones de la media poblacional para cada jugador y en cada categoría estadística, se ha de proceder a la clasificación de los jugadores, es decir, la creación de un ranking descendente. El primer paso sería la eliminación de aquellos jugadores cuyas estimaciones produzcan un error por encima del máximo admisible, como hemos explicado anteriormente. Tras realizar esos descartes, la clasificación debe someterse también a criterios de inferencia estadística, es decir, por ejemplo, a la hora de comparar los puntos obtenidos por dos jugadores, habría que realizar un test estadístico que ayude a dilucidar si ambas puntuaciones pueden considerarse estadísticamente iguales o no.

La comparación entre los diferentes jugadores para cada apartado estadístico obedecería a un problema de comparaciones múltiples. Esta clase de problemas se suele resolver a través de tests ómnibus, como el test F de Snedecor en el Análisis de la Varianza. Este tipo de tests han sido criticados por su poca capacidad para detectar los efectos de interés y por proporcionar información demasiado general y poco útil sobre la existencia de esos efectos (ej. Cohen, 1990; Rosenthal, Rosnow y Rubin, 2000; Rosnow y Rosenthal, 1996). Así, se recomienda utilizar contrastes enfocados, basados en la distribución t de Student, analizando las relaciones de interés a través de la utilización de ponderaciones lineales (ej. Olegnik y Algina, 2000; Rosnow y Rosenthal, 1996), o los análisis *post-hoc* para controlar el error Tipo I, como los ajustes de Bonferroni, Scheffé, Newman-Keuls, etc. (ej. Casas, 1997; Curran-Everett, 2000). Esta última opción ha sido criticada por cierto grupo de investigadores (ej. Anderson, Burnham y Thompson, 2000; Cohen, 1990), debido el conservadurismo de alguno de estos métodos y la "distracción" que producen sobre el cómputo de la magnitud de interés, es decir, del tamaño del efecto o diferencia sustantiva entre los tratamientos.

De cualquier modo, el caso del estudio que nos ocupa tiene unas particularidades que difieren del uso más tradicional de los métodos estadísticos para el análisis de experimentos. En este caso, las diferencias entre los distintos jugadores no proceden del efecto de ningún "tratamiento". Así, los contrastes enfocados del tipo de análisis de tendencia por ordenación de intensidad en la aplicación de los tratamientos (ej. Rosenthal, Rosnow y Rubin, 2000) no son adecuados. Lo que realmente interesa es enfocar las comparaciones entre los jugadores con puntuaciones similares, pero sin utilizar el clásico contraste de hipótesis múltiple, donde la hipótesis nula es la igualdad de las puntuaciones de todos los jugadores. Esta premisa es inadecuada porque es obvio que van a existir diferencias entre los mejores y los peores jugadores en cada categoría. Como bien explica Rothman (1990), la paradoja de penalizar estadísticamente la claridad de los contrastes "dos a dos" por el simple hecho de manejar más información no debe aceptarse. En este caso, no debería penalizarse la comparación entre dos jugadores (a través de ajustes tipo Bonferroni, Holms, Simes-Hochberg, Hommel, etc.) por considerar que están dentro del conjunto de todos los jugadores de la liga, al menos, sin un modelo teórico detrás. Las reflexiones de Rothman (1990), también apoyadas por Savitz y Olsan (1998), Mayo y Cox (2006), o Rothman, Greenland y Lash (2008), están relacionadas con la disyuntiva sobre el uso conveniente de los contrastes de hipótesis en la perspectiva frecuentista sobre probabilidad, que ha llevado a un gran número de autores a recomendar la utilización de los índices de tamaño de efecto en lugar del clásico p-valor asociado a un contraste de hipótesis (ver la revisión realizada por Nickerson, 2000).

La forma de proceder que proponemos en esta investigación para construir el ranking en cada categoría es la simple ordenación del tamaño de efecto resultante de la comparación por pares de jugadores, realizando una previa ordenación basada en la estimación puntual de sus valores medios. Así, los jugadores se ordenan de modo descendente tal y como la ACB lo hace actualmente, y después se calcula el tamaño de efecto de forma jerarquizada para analizar las posibles diferencias entre pares de jugadores, y entre cada uno de ellos con respecto al líder de la categoría. Esta forma de proceder, da una visión "práctica" de distancia para los jugadores de cada apartado estadístico.

Como índice de tamaño de efecto podemos utilizar la propia diferencia de medias Δ (Cohen, 1990), ya que no sería estrictamente necesario estandarizar esa diferencia (caso de la *d* de Cohen (1988)), porque las puntuaciones de cada categoría tienen un significado propio no arbitrario. De este modo (8):

$$\mu_{j1} - \mu_{j2} = \Delta \tag{8}$$

El siguiente paso es establecer qué tamaño de efecto consideramos como relevante, es decir, cuál es el valor de Δ que discrimine a dos jugadores. Una solución obvia es establecer $\Delta > 0$, lo que significaría que cualquier valor decimal diferente entre las puntuaciones de dos jugadores haría que ambos fuesen considerados en diferente lugar en el ranking. Sin embargo, esta solución no nos parece atinada. La ACB clasifica a los jugadores usando sólo un decimal, por lo que se podría tomar ese decimal como valor de Δ. Esta nueva solución parece en principio más acertada. Así, por ejemplo, si un jugador anotase 707 puntos en 32 partidos y otro 704 en esos mismos partidos, $\Delta = 0.093 < 0.1$, por lo que ambos jugadores compartirían la misma posición en el ranking. Es una tolerancia que resulta bastante lógica, y estaría en consonancia con la filosofía del análisis de la importancia práctica de los efectos (Meehl, 1990). Por tanto, si a dos jugadores en toda una temporada sólo les separa una canasta, es lógico pensar que ambos son igualmente óptimos para ocupar la misma posición en el ranking.

Para discernir si dos jugadores ocupan la misma posición en el ranking se puede construir un contraste de la forma siguiente (9):

$$\begin{aligned} H_0 &: \mu_{j1} - \mu_{j2} < \Delta = 0.1 \\ H_1 &: \mu_{j1} - \mu_{j2} \geq \Delta = 0.1 \end{aligned} \tag{9}$$

En este caso se plantea la hipótesis nula de igualdad de medias, algo plausible en contrastes dos a dos con jugadores previamente ordenados en función de sus estimaciones puntuales. Así, si se quieren comparar las estadísticas de dos jugadores $\{P_1, P_2\}$ para una categoría k determinada del juego, la decisión a tomar dependería de la comparación directa de ambas puntuaciones en el caso de que $n_1 = n_2 = N$ o la comparación estadística cuando $n_1 \vee n_2 \neq N$ (10):

$$
\begin{aligned}
& n_1 = n_2 = N \\
& \mu_{j1} = \mu_{j2} \Leftrightarrow \Delta < 0.1 \\
& \\
& n_1 \vee n_2 \neq N \\
& \mu_{j1} = \mu_{j2} \Leftrightarrow \left\{ (\overline{x}_{j1} - \overline{x}_{j2}) < \Delta + t_{\alpha,v}(SE_{jp_{1\_2}}) \right\} \equiv t_{prueba} = \frac{\overline{x}_{j1} - \overline{x}_{j2} - \Delta}{SE_{jp_{1\_2}}} < t_{\alpha,v}
\end{aligned}
\tag{10}
$$

El valor del error estándar de la diferencia de medias $SE_{jp_{1\_2}}$, puede ser estimado utilizando la ponderación de varianzas en el caso de homogeneidad (Casas, 1997), o métodos aproximados (ej. Welch) en situaciones de heterocedasticidad (Grissom y Kim, 2005). Como mostrará más adelante la exploración de datos, la situación realista es la heterocedástica, por lo que los diferentes métodos propuestos (Vegas, 1997; Belloni y Didier, 2008) podrían ser implementados. No obstante, esos métodos (algunos de ellos de cálculo muy complejo), no contemplan en su desarrollo la situación en las que la muestra es muy similar a la población, es decir, cuando el error típico de la media se ve corregido por el factor de finitud $F = \sqrt{\frac{N - n_{jp}}{N}}$. La aproximación de Welch (la más utilizada), requiere una corrección del estadístico t de Student, el cual se recomienda en el caso de desconocer las varianzas poblacionales. Así, la computación del error estándar y de los grados de libertad sería (11)

$$
\begin{aligned}
& SE_{jp_{1\_2}} = \sqrt{\frac{S_1^2}{n_1}F_1 + \frac{S_2^2}{n_2}F_2} \\
& \upsilon = \frac{\left(\frac{S_1^2}{n_1}F_1 + \frac{S_2^2}{n_2}F_2\right)^2}{\frac{\left(\frac{S_1^2}{n_1}F_1\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}F_2\right)^2}{n_2 - 1}}
\end{aligned}
\tag{11}
$$

La comparación de jugadores "dos a dos" podría a primera vista parecer coherente para establecer las posiciones en el ranking descendente. Sin embargo, surge un problema derivado del concepto de distancia y la problemática de las comparaciones estadísticas múltiples. Sean $d(a,b)$, $d(b,c)$, $d(a,c)$ las distancias entre tres posiciones consecutivas (y por ende distintas) en un ranking. Es decir, si estadísticamente $a = b$, entonces $d(a,b) = 0$, por lo que ambos jugadores ocuparían la misma posición del ranking. Sin embargo, si $a \neq b$, $d(a,b) = 1$, por lo que las posiciones serían diferentes. El problema surge cuando se quieran comparar tres jugadores con valores medios muy parecidos, lo que obligaría a varios test "dos a dos". Así, si las comparaciones estadísticas nos dan el siguiente patrón: $a = b$, $b = c$, y $a \neq c$, entonces ocurre una paradoja, y es que $d(a,b) = 0$, $d(b,c) = 0$ y $d(a,c) = 1$, lo que desde el punto de vista cartesiano es incongruente, ya que si $d(a,b) = 0$ y $d(b,c) = 0$, entonces $d(a,c)$ debería ser también cero. Esta paradoja que surge de las particularidades de la inferencia estadística no debe tampoco aceptarse, porque las posiciones relativas en el ranking de dos jugadores no deben depender de la comparación con un tercero, es decir, $d(b,c)$ debería ser independiente de $d(a,b)$.

Para evitar este problema, proponemos no realizar un ranking numérico de jugadores, con las posiciones delimitadas, como se hace actualmente, sino realizar una comparación entre el líder de cada categoría y el resto, con el fin de crear un ranking descendente pero sin delimitación definida de posiciones. Realmente, desde el punto de vista mediático, la primera posición es la que más interesa, por lo que el hecho de que un jugador ocupe, por ejemplo, el quinto o sexto puesto, queda en un segundo plano. Esto no quiere decir que no se ordene descendentemente a los jugadores, ni que se puedan comparar entre ellos. Así, todos los jugadores tendrán una medida de distancia con respecto al primero, por lo que las distancias entre éstos se podrían comparar. Asimismo, siempre se pueden realizar comparaciones “dos a dos” para analizar si las puntuaciones de dos jugadores concretos son iguales o diferentes desde el punto de vista estadístico. De cualquier manera, el ranking resultante clasificará a los jugadores de forma descendente, y los intervalos de confianza de la diferencia de medias con respecto al primero del ranking, servirá como criterio para evaluar su posición en la clasificación.

Por tanto, tenemos ya todos los instrumentos necesarios para realizar la creación de rankings para cada apartado estadístico, proceso que detallaremos en el apartado siguiente.

Muestra

En primer lugar, hemos establecido el siguiente conjunto de categorías estadísticas: Valoración o Ranking, Puntos, Asistencias, Rebotes, Tapones, y Recuperaciones. Como hemos explicado anteriormente, el sistema de clasificación oficial seguido por la ACB ordena a los jugadores en cada categoría en función del valor bruto de los números conseguidos, independientemente del número de partidos jugado por temporada.

En el momento de realización de esta investigación, las estadísticas oficiales que provee la ACB en su página web, www.liga-acb.es, están disponibles a partir de la temporada 2004/2005, en cuanto a las clasificaciones individuales por categoría estadística. Por ello, circunscribimos nuestro análisis únicamente al horizonte temporal donde hay datos oficiales disponibles.

Resultados

El primer paso a realizar es el análisis de las distribuciones de probabilidad empíricas de aquellos jugadores que han ocupado las primeras posiciones en cada categoría, y que han jugado todos los partidos de la temporada. De este modo, no necesitaríamos ningún test estadístico (ej. Shapiro-Wilks, Anderson-Darling, etc.) para contrastar la hipótesis de normalidad, y así evitar el problema de elegir entre la divergencia de resultados proporcionados por los diferentes métodos. La Tabla 1 muestra los 10 primeros clasificados de cada categoría estadística en la temporada 2008/2009, estando sombreados aquellos jugadores que han disputado todos los partidos.

Si la distribución es perfectamente normal, la media y la mediana coinciden, y la asimetría y el exceso de curtosis es cero. Hemos calculado esos valores para los jugadores seleccionados, y como puede contemplarse en la Tabla 2, la desviación de la normalidad no es en ningún momento excesiva. Asimismo, hemos comparado dos cuantiles de la distribución empírica (en ambas colas de la distribución) y los de una distribución normal teórica con la misma media y desviación típica que la distribución empírica. De nuevo las diferencias son muy pequeñas. Como puede apreciarse, las desviaciones típicas difieren entre jugadores, por lo que la situación de heterocedasticidad es la más realista.

Tabla 1. Clasificación en las categorías estadísticas consideradas en la temporada 2008/2009

	Val.	Ptos.	As.	Reb.	Tap.	Rec.
1	Reyes, F.	*Rakocevic, I.	Rubio, R.	Borchardt, C.	Vázquez, F.	Rubio, R.
2	Splitter, T.	*Oleson, B.	Rodríguez, J.	Reyes, F.	*Barnes, L.	Prigioni, P.
3	Borchardt, C.	Quinteros, P.	Prigioni, P.	*Ilyasova, E.	Borchardt, C.	Popovic, B.
4	*Rakocevic, I.	Dean, T.	Thomas, C.	Panko, A.	*Asselin, J.	*Lescano, M.
5	Vázquez, F.	Haislip, Ma.	*Valters, K.	Vázquez, F.	*Gutiérrez, J. P.	Cook, O.
6	*Oleson, B.	*Jeter, P.	Cook, O.	Splitter, T.	Splitter, T.	Jasen, P.
7	Haislip, M.	Reyes, F.	*Salgado, J.	*Barnes, La.	Haislip, M.	*Jeter, P.
8	*Banic, M.	*English, C.	Williams, S.	*Augustine, J.	Hdez-Son., E.	Thomas, C.
9	Mickeal, P.	Navarro, J.C.	Lakovic, J.	*Banic, M.	*Savané, S.	Rodríguez, J.
10	Panko, A.	Bullock, L.	Navarro, J.C.	*Suárez, C.	Ndong, B.	Quinteros, P.

Nota: Los jugadores que han disputado todos los partidos aparecen con un asterisco.

Tabla 2. Propiedades de la distribución de datos

Jugador	Categoría.	Media	Mediana	Asimetría	Desv. típica	Exceso curtosis	Cuantil empírico 1*	Cuantil teórico 1*	Cuantil empírico 2**	Cuantil teórico 2**
Rakocevic, I.	Val.	19.25	19.50	-0.03	8.28	0.31	4.00	6.54	32.00	31.96
Oleson, B.	Val.	18.28	18.00	0.05	9.67	-0.71	3.00	3.45	32.00	33.12
Banic, M.	Val.	15.63	16.00	-0.22	6.39	0.07	7.00	5.83	26.00	25.42
Rakocevic, I.	Ptos.	19.81	20.50	-0.23	6.53	1.31	7.00	9.80	27.00	29.82
Oleson, B.	Ptos.	18.06	19.50	-0.36	7.11	-0.49	6.00	7.16	29.00	28.96
Jeter, P.	Ptos.	16.34	17.00	-0.50	6.42	0.67	3.00	6.49	25.00	26.19
English, C.	Ptos.	15.97	16.00	0.15	7.48	0.13	2.00	4.50	30.00	27.44
Valters, K.	As.	4.81	5.00	0.13	2.43	-0.64	1.00	1.09	8.00	8.54
Salgado, J.	As.	4.16	4.00	1.03	2.09	2.27	1.00	0.94	7.00	7.37
Ilyasova, E.	Reb.	7.63	7.00	0.20	3.75	-0.59	2.00	1.88	14.00	13.37
Barnes, L.	Reb.	6.25	5.50	0.67	4.18	-0.06	0.00	-0.17	13.00	12.67
Augustine, J.	Reb.	6.06	6.00	0.31	2.46	-0.50	2.00	2.29	10.00	9.84
Banic, M.	Reb.	5.94	6.00	-0.15	2.28	-0.52	2.00	2.44	9.00	9.43
Suárez, C.	Reb.	5.94	6.00	0.53	2.77	-0.02	2.00	1.69	10.00	10.19
Barnes, L.	Tap.	1.56	1.50	1.11	1.34	2.18	0.00	-0.50	3.00	3.63
Asselin, J.	Tap.	1.22	1.00	1.06	1.39	0.45	0.00	-0.91	4.00	3.34
Gutiérrez, J. P.	Tap.	1.13	1.00	0.94	0.99	0.89	0.00	-0.40	3.00	2.65
Savané, S.	Tap.	1.03	1.00	0.73	1.07	-0.74	0.00	-0.62	3.00	2.68
Lescano, M.	Rec.	1.81	1.00	1.09	1.89	0.24	0.00	-1.09	6.00	4.72
Jeter, P.	Rec.	1.66	1.00	0.32	1.29	-1.08	0.00	-0.32	4.00	3.63

* $p \leq 0.0625$

** $p \leq 0.935$

Por tanto, podemos afirmar que las distribuciones empíricas de los jugadores que han ocupado los primeros puestos de cada apartado estadístico y han jugado todos los partidos de la temporada son aproximadamente normales, o al menos no se desvían en demasía de la normalidad. Este hecho facilita la computación estadística del error máximo admisible $A_{j,t=0|2008/2009}$, tal y como hemos explicado en epígrafes anteriores. Así, la Tabla 3 muestra los valores del error máximo admisible para cada apartado estadístico, computado a partir de (5) y (6).

Tabla 3. Error máximo admisible

	Val.	Ptos.	As.	Reb.	Tap.	Rec.
$\mu_{j,t=0|2008/2009}$	22.75	18.32	5.01	8.38	1.63	2.03
$A_{j,t=0|2008/2009}$ $G=10\%$	2.27	1.83	0.50	0.84	0.16	0.20
$A_{j,t=0|2008/2009}$ $G=5\%$	1.14	0.92	0.25	0.42	0.08	0.10
$A_{j,t=0|2008/2009}$ $G=1\%$	0.23	0.18	0.05	0.08	0.02	0.02

Llegados a este punto, podemos establecer el tamaño de muestra mínimo para cada jugador en cada apartado estadístico, en función de la estimación de su varianza, como se muestra en (7). Así, varios jugadores que aparecen en los primeros puestos de las clasificaciones deberían quedar eliminados, ya que no han jugado los partidos necesarios para obtener estimaciones con la precisión requerida.

Asimismo, podemos calcular el intervalo de confianza para la diferencia de medias, utilizando la aproximación de Welch. De este modo, podemos conocer la medida de la magnitud de los efectos, es decir, de la distancia práctica entre ambos jugadores. Al margen de ese intervalo de confianza, proveemos un valor crítico que nos acepta o rechaza la hipótesis nula de igualdad, tal y como describimos en (10), para el valor de Δ determinado.

Las Tablas 4, 5, 6, 7, 8 y 9, muestran los resultados del re-análisis para los apartados estadísticos considerados. Las comparaciones y los test estadísticos se han realizado tomando como referencia el jugador subrayado, el cual debe ser el líder de cada categoría según el criterio que hemos adoptado.

En cuanto a la categoría "Valoración", Felipe Reyes sigue siendo claramente el triunfador estadístico, ya que aventaja significativamente al resto de jugadores. Es de destacar que Tiago Splitter y Curtis Borchardt, que originalmente ocupan las posiciones segunda y tercera, respectivamente, deberían ser eliminados del ranking, ya que el error de sus estimaciones está por encima del umbral que arbitrariamente hemos considerado. Y es que sus intervalos de confianza para la media son más amplios de lo que sería deseable, lo que les confiere una ventaja añadida a la hora de "pujar" estadísticamente para la primera posición del ranking. Por tanto, *ceteris paribus*, ambos jugadores deberían de haber jugado algún partido más para ser considerados en el ranking (1 partido en el caso de Splitter y 3 partidos en el caso de Borchardt).

Tabla 4. Valoración 2008/2009

Ranking ACB	Jugador	n_{jp}	$\overline{x}_{jp}$	S_{jp}	$\hat{A}_{j,t=0\|}$	IC95%	n_{jp} mínimo	Dif. medias (Δ)	IC95% (Δ)	t_{prueba}
1	Reyes, F. (RMA)	30	22.73	9.30	0.87	(21.87 , 23.60)	28.69			
2	Splitter, T. (TAU)	26	20.62	7.55	1.31*	(19.31 , 21.92)	27.23	2.11	(0.33 , 3.90)	2.62**
3	Borchardt, C. (GRA)	26	19.92	10.23	1.77*	(18.15, 21.70)	29.21	2.81	(0.55 , 5.07)	2.80**
4	Rakocevic, I. (TAU)	32	19.25	8.42	0	19.25	28.04	3.48	(2.48 , 4.49)	7.97**
5	Vázquez, F. (FCB)	31	18.74	9.81	0.64	(18.11 , 19.38)	28.99	3.99	(2.78, 5.21)	7.39**

*Error por encima del máximo admisible (1.14) cuando G=5%

** $t_{prueba} \geq t_{\alpha=0.05,\upsilon}$

Tabla 5. Puntos 2008/2009

Ranking ACB	Jugador	n_{jp}	$\overline{x}_{jp}$	S_{jp}	$\hat{A}_{j,t=0\|}$	IC95%	n_{jp} mínimo	Dif. medias (Δ)	IC95% (Δ)	t_{prueba}
1	Rakocevic, I. (TAU)	32	19.81	6.63	0	19.81	27.87			
2	Oleson, B. (AGF)	32	18.06	7.22	0	18.06	28.45	1.75		
3	Quinteros, P. (CAI)	30	17.43	6.71	0.63	(16.81, 18.06)	27.96	2.38	(1.65 , 3.10)	7.44**
4	Dean, T. (MUR)	31	16.68	7.82	0.51	(16.17 , 17.18)	28.92	3.14	(2.55 , 3.72)	12.22**
5	Haislip, M. (UNI)	31	16.45	6.65	0.43	(16.02 , 16.88)	27.89	3.36	(2.86 , 3.86)	15.45**

*Error por encima del máximo admisible (0.92) cuando G=5%

** $t_{prueba} \geq t_{\alpha=0.05,\upsilon}$

Tabla 6. Asistencias 2008/2009

Ranking ACB	Jugador	n_{jp}	$\overline{x}_{jp}$	S_{jp}	$\hat{A}_{j,t=0\vert}$	IC95%	n_{jp} mínimo	Dif. medias (Δ)	IC95% (Δ)	t_{prueba}
1	Rubio, R. (DKV)	22	6,14	3,12	0.76*	(5.38 , 6.90)	30.49			
2	Rodríguez, J. (MAN)	28	6,11	2,50	0.34*	(5.77 , 6.45)	29.71	0.03	(-0.93 , 0.99)	-0.17
3	Prigioni, P. (TAU)	31	5,03	2.60	0.17	(4.86 , 5.20)	29.88	1.10	(0.19 , 2,02)	2.64**
4	Thomas, C. (MUR)	31	4,97	2.46	0.16	(4.81 , 5.13)	29.64	0.06	(-0.20 , 0.33)	-0.31
5	Valters, K. (AGF)	32	4,81	2.47	0	4.81	29.66	0.22	(0.02 , 0.41)	1.45

*Error por encima del máximo admisible (0.25) cuando G=5%

** $t_{prueba} \geq t_{\alpha=0.05,\upsilon}$

Tabla 7. Rebotes 2008/2009

Ranking ACB	Jugador	n_{jp}	$\overline{x}_{jp}$	S_{jp}	$\hat{A}_{j,t=0\vert}$	IC95%	n_{jp} mínimo	Dif. medias (Δ)	IC95% (Δ)	t_{prueba}
1	Borchardt, C. (GRA)	26	9,88	3.91	0.68*	(9.21 , 10.56)	29.39			
2	Reyes, F. (RMA)	30	9,40	3.90	0.36	(9.04 , 9.76)	29.38	0.48	(-0.39 , 1.36)	1.02
3	Ilyasova, E. (FCB)	32	7,63	3.81	0.00	7.63	29.26	1.78	(1.35 , 2.20)	9.41**
4	Panko, A. (GBC)	32	6,75	2.87	0.00	6.75	27.48	2.65	(2.23 , 3.07)	14,32**
5	Vázquez, F. (FCB)	31	6,65	3.88	0.25	(6.39 , 6.90)	29.35	2.75	(2.26 ,3.25)	12.27**

*Error por encima del máximo admisible (0.42) cuando G=5%

** $t_{prueba} \geq t_{\alpha=0.05,\upsilon}$

Tabla 8. Tapones 2008/2009

Ranking ACB	Jugador	n_{jp}	$\overline{x}_{jp}$	S_{jp}	$\hat{A}_{j,t=0\|}$	IC95%	n_{jp} mínimo	Dif. medias (Δ)	IC95% (Δ)	t_{prueba}
1	Vázquez, F. (FCB)	31	1,65	1.28	0.08	(1.56 , 1.73)	31.07			
2	Barnes, L. (MUR)	32	1,56	1.37	0	1.56	31.18	0.08	(-0.01 , 0.18)	-0.43
3	Borchardt, C. (GRA)	26	1,38	1.36	0.24*	(1.15 , 1.62)	31.17	0.26	(-0.03 , 0.55)	1.31
4	Asselin, J. (MAN)	32	1,22	1.41	0	1.22	31.22	0.43	(0.33 , 0.52)	8.04**
5	Gutiérrez, J. P. (GRA)	32	1,13	1.01	0	1.13	30.52	0.52	(0.42, 0.62)	10.34**

*Error por encima del máximo admisible (0.08) cuando G=5%

** $t_{prueba} \geq t_{\alpha=0.05,\upsilon}$

Tabla 9. Recuperaciones 2008/2009

Ranking ACB	Jugador	n_{jp}	$\overline{x}_{jp}$	S_{jp}	$\hat{A}_{j,t=0\|}$	IC95%	n_{jp} mínimo	Dif. medias (Δ)	IC95% (Δ)	t_{prueba}
1	Rubio, R. (DKV)	22	2,18	1.79	0.44*	(1.75 , 2.62)	31.25			
2	Prigioni, P. (TAU)	31	2,10	1.42	0.09	(2.00 , 2.19)	30.83	0.09	(-0.44 , 0.61)	-0.07
3	Popovic, B. (GBC)	30	1,83	1.53	0.14*	(1.69 , 1.98)	30.99	0.26	(0.07 , 0.46)	1.96
4	Lescano, M. (CAI)	32	1,81	1.93	0	1.81	31.35	0.28	(0.18 , 0.39)	4.08**
5	Cook, O. (UNI)	32	1,69	1.40	0	1.69	30.79	0.41	(0.30 , 0.52)	6.85**

*Error por encima del máximo admisible (0.10) cuando G=5%

** $t_{prueba} \geq t_{\alpha=0.05,\upsilon}$

Tampoco hay diferencias importantes en la categoría "Puntos" (Tabla 5), ya que Igor Rakocevic mantiene la primera posición, de forma destacada además. Sin embargo, sí que hay cambios destacables en cuanto a las "Asistencias" (Tabla 6), ya que los dos primeros jugadores del ranking ACB, Ricky Rubio y Javi Rodríguez, no han jugado los partidos suficientes para obtener un error admisible. Esto supondría que fuesen eliminados del ranking, con lo que Pablo Prigioni ocuparía la primera plaza. Pero esa primera plaza debería ser compartida con Chris Thomas y Kristaps Valteers, ya que no existen diferencias estadísticas con respecto a Prigioni. Por tanto, esos tres jugadores serían los triunfadores estadísticos en la categoría de "Asistencias". Ocurre, además, un resultado curioso, referido a que a pesar de que Rubio y Rodríguez poseen estimaciones con un error por encima del exigible, ambos siguen siendo superior estadísticamente al resto. El número de asistencias de ambos jugadores es tan alto con respecto a los demás, que, incluso en una situación de alto error de estimación, aún mantienen su hegemonía. Sin embargo, de obviarse el criterio del error admisible, ambos deberían de compartir el honor de la primera plaza del ranking. Este hecho, obligaría a replantearse ciertos ajustes en cuanto al método descrito, que serán discutidos posteriormente.

También encontramos novedades en cuanto a los "Rebotes" (Tabla 7), ya que Curtis Borchardt no alcanza el número de partidos mínimo exigible. Felipe Reyes, por tanto, se convertiría en el triunfador estadístico de esta categoría, con diferencias ostensibles con respecto a sus seguidores. De nuevo, si obviamos el criterio del error admisible, Borchardt y Reyes deberían compartir el galardón, ya que no existen diferencias estadísticas entre ambos. Por tanto, podemos decir que Felipe Reyes se ha visto perjudicado por el sistema actual de clasificación.

En cuanto a los "Tapones" (Tabla 8), Fran Vázquez debería compartir el primer puesto del ranking con Lamont Barnes. Borchardt de nuevo no alcanza el mínimo exigible, aunque si se obviara esta circunstancia se uniría a los dos jugadores mencionados para compartir el galardón.

Finalmente, en cuanto a las "Recuperaciones" (Tabla 9), de nuevo Ricky Rubio debería quedar fuera del ranking, quedando la primera plaza para Prigioni y Popovic, ya que no existen diferencias estadísticas entre ambos jugadores.

Análisis de sensibilidad

Dado que uno de los criterios fundamentales sobre el que hemos basado nuestro razonamiento es la fijación de un error máximo admisible para cada categoría estadística, conviene analizar la sensibilidad de los resultados a esa premisa. Los análisis de sensibilidad son altamente deseables en estas situaciones, siendo además, una de las herramientas metodológicas recomendadas (aunque poco utilizadas, ciertamente) a nivel multidisciplinar en ciencia, por ejemplo: Hayduk (1996) en el ámbito de los modelos de estructura de covarianza, Sterman (2002) en el contexto de la simulación de sistemas dinámicos, o Rothman, Greenland y Lash (2008) en epidemiología.

En primer lugar, hemos de estudiar la sensibilidad de $\mu_{j,t=0}$ a los cambios en el criterio prefijado para su cómputo. Como hemos indicado, hemos escogido un horizonte móvil de 4 años. Esto nos servirá para computar el error máximo admisible, como un porcentaje de ese valor (hemos usado el 5%). Hemos utilizado los datos proveídos por www.keyhoops.com, ya que la web oficial de la ACB no tiene a la vista esos datos en el momento de realización de este trabajo. Por ello, ha sido imposible calcular la categoría "Valoración", porque www.keyhoops.com no la considera (considera alternativamente la "Eficiencia", cuyo cómputo difiere de la "Valoración ACB"). Hemos utilizado el mismo procedimiento que describimos anteriormente, es decir, cuantificar los primeros puestos de cada categoría sólo en el caso de que hubieran jugado todos los partidos de esa temporada, para así evitar escoger un caso donde no se alcanzara un tamaño de muestra mínimo para pertenecer al ranking.

La Tabla 10 muestra que no existen variaciones ostensibles en los resultados obtenidos, si el horizonte temporal para $\mu_{j,t=0}$ se incrementa 5, 6, 7 u 8 años. Esas pequeñas variaciones no afectarían en ningún caso los resultados de nuestro estudio.

Tabla 10. Sensibilidad en el Error máximo admisible

	Horizonte temporal (años)	Val.	Ptos.	As.	Reb.	Tap.	Rec.
$\mu_{j,t=0\|2008/2009}$	4	22.75	18.32	5.01	8.38	1.63	2.03
	5		18.84	4.90	8.60	1.62	2.00
	6		18.72	5.02	8.90	1.60	2.07
	7		18.80	5.00	9.13	1.64	2.06
	8		18.94	4.94	9.38	1.74	2.15
$A_{j,t=0\|2008/2009}$ $G=5\%$	4	1.14	0.92	0.25	0.42	0.08	0.10
	5		0.94	0.25	0.43	0.08	0.10
	6		0.94	0.25	0.45	0.08	0.10
	7		0.94	0.25	0.46	0.08	0.10
	8		0.95	0.25	0.47	0.09	0.11

El siguiente análisis se refiere a la sensibilidad de los resultados al hecho de considerar arbitrariamente $G=5\%$ en lugar de otro porcentaje. Aunque el 5% es una especie de "número mágico" en ciencia, como umbral de error admisible (véase por ejemplo las convenciones sobre el p-valor en los contrastes de hipótesis), vamos estudiar cómo variarían los resultados ante un valor más elevado de error: $G=10\%$

Tras el reanálisis de los datos, podemos afirmar que los resultados son bastante sensibles a la elección de *G*, ya que muchos de los jugadores eliminados del ranking, no lo estarían bajo un criterio de error menos exigente. Sólo Ricky Rubio, en los apartados de "Asistencias" y "Recuperaciones" y Curtis Bordchardt en "Tapones" seguirían quedándose fuera del ranking. Los triunfadores estadísticos bajo este criterio menos exigente serían los que se muestran en la Tabla 11, junto con la comparativa con la situación real en la ACB.

Tabla 11. Triunfadores estadísticos bajo dos valores de error máximo admisible.

	Valoración	Puntos	Asistencias	Rebotes	Tapones	Recup.
ACB	Reyes, F.	Rakocevic, I.	Rubio, R.	Bordchart, C.	Vázquez, F.	Rubio, R.
$G=5\%$	Reyes, F.	Rakocevic, I.	Prigioni, P., Thomas, C., y Valters, K.	Reyes, F.	Vázquez, F., y Barnes, L.	Prigioni, P.
$G=10\%$	Reyes, F.	Rakocevic, I.	Rodríguez, J.	Reyes, F.	Vázquez, F., y Barnes, L.	Prigioni, P., y Popovic, B.

Discusión

En esta investigación hemos propuesto una nueva forma de elaborar los rankings para cada categoría estadística de la Liga ACB, basándonos en una aproximación inferencial al parámetro de interés, es decir, al valor esperado de cada distribución de datos. Para ello, hemos asumido que las realizaciones muestrales son valores independientes y aleatorios, tomados de una población delimitada por el número de partidos que forma una temporada.

Estas premisas, a nuestro entender, permiten elaborar rankings de una forma más justa, ya que consideran la imprecisión de las estimaciones, y por ende, el error asociado a cada media muestral. Por tanto, se premia a los jugadores que han disputado un número de partidos cercano al total de la temporada, es decir, a aquellos cuyas estimaciones son mucho más precisas. Sin embargo, los jugadores con un número de partidos no tan elevado se caracterizan por estimaciones con un margen de error no deseable, que les favorecería a la hora de pujar por los primeros puestos del ranking.

Además, actualmente, los rankings de la ACB no discriminan entre diferencias grandes y pequeñas entre jugadores, a la hora de asignarles una posición en el ranking. A través de nuestro enfoque, diferencias pequeñas serán no significativas en la mayoría de los casos, por lo que no habría diferencia entre dos posiciones del ranking en esas circunstancias. Es más, proponemos una tolerancia de una décima, margen razonable para discernir si al cabo de una temporada dos jugadores son merecedores de la misma posición del ranking.

Los resultados de nuestro estudio aplicado a la última temporada en la ACB (2008/2009) indican que puede haber cambios sustanciales en alguna de las categorías estadísticas, tras aplicar nuestro método. Este hecho es ciertamente importante, dada la trascendencia que tienen las estadísticas en el baloncesto actual, no sólo a nivel mediático, con el incremento de notoriedad de los jugadores que ocupan la primera posición, sino a nivel económico, ya que las estadísticas son un vehículo de comunicación/negociación entre jugadores, clubes, entrenadores y agentes. No obstante, y considerando el papel que juegan los rankings estadísticos en el baloncesto actual, no hemos encontrado ninguna referencia que haya tenido en cuenta esta aproximación inferencial a la creación de rankings,

Nuestra propuesta no pretende criticar los sistemas actuales de creación de rankings, ya que entendemos que las diferentes ligas tienen que vender su "producto", y se dirigen a grupos de interés dispares, donde los aficionados son parte fundamental. Es obvio que para el aficionado es mucho más fácil entender e interpretar el sistema actual de creación de rankings que, como hemos dicho, es común en todas las ligas del mundo. Sin embargo, esto no es razón suficiente para dejar de buscar nuevos sistemas, que aporten nuevas ideas, y que tal vez den una visión más ajustada a la realidad de la temporada. Insistimos de nuevo en la importancia de los rankings estadísticos en el baloncesto, y recordamos una de las máximas de Ries y Trout (1999): "Es mejor ser el primero que ser el mejor". Esta "Ley del liderazgo" es uno de los principios básicos de la teoría del marketing que defienden estos autores, argumentando que aquellas marcas que son las primeras en un mercado, ocupan una posición de privilegio en la mente del consumidor, independientemente de que luego existan otras con una mayor calidad. Desde el punto de vista de marketing, es innegable el efecto positivo que produce quedar primero en una categoría estadística en baloncesto con respecto al segundo. Por ello, creemos firmemente que la propuesta que realizamos en esta investigación debe ser tenida en consideración, en aras de buscar una mayor equidad en los rankings de las competiciones.

Parece también evidente que nuestra propuesta es extensible a otras competiciones con un número definido de partidos, como podría ser la fase regular de la Euroliga o de la NBA. En el caso de la NBA, el tamaño poblacional sería ostensiblemente mayor (82 partidos), por lo que existiría mucha más heterogeneidad en el número de partidos mínimo necesario para no superar el máximo error admisible. No obstante, en la NBA existe un criterio oficial para la elaboración de rankings en base a unos mínimos por cada categoría (www.nba.com/leader_requirements). En el caso de que se jugasen únicamente 32 partidos (en analogía con el análisis realizado en la ACB), el mínimo de partidos necesario para entrar en el ranking sería de 28 (una vez cumplido otros mínimos en cuanto a puntuaciones en cada categoría). Como puede deducirse, la recomendación de la NBA es muy general, y no tiene en cuenta la heterogeneidad en los desempeños de cada jugador para cada categoría estadística.

También es necesario destacar que nuestro procedimiento ha de aplicarse una vez la temporada haya concluido, ya que hasta entonces no se conocen el total de realizaciones muestrales. Durante la temporada, los rankings actuales deben servir como aproximación al ranking final.

Desde el punto de vista metodológico, existen procedimientos que podrían ofrecer resultados ligeramente diferentes. Alentamos a otros autores a profundizar sobre este hecho, considerando las diferentes aproximaciones a la resolución del problema de Behrens-Fisher, o del enfoque no paramétrico. Las referencias que hemos aportado en este estudio pueden ser un buen punto de partida para ello.

El análisis de sensibilidad de las asunciones realizadas ha mostrado que el valor de G es clave a la hora de su influencia en los resultados. Sería conveniente que, si nuestra propuesta se usara de forma "oficial", se discutiera sobre la idoneidad de considerar un valor de G entre el 5 y el 10%. Creemos sinceramente que más de un 10% sería inadecuado, ya que el error se incrementaría en demasía. No obstante, existe también la posibilidad de que un panel de expertos analice esta cuestión, y que fije valores de error máximo admisible para cada categoría estadística. Así, por ejemplo, podría decidirse que 1 punto, o 0.5 tapones son el valor del error máximo admisible. De especial relevancia es el problema que ocurre con las asistencias, donde existe cierta disparidad de criterios sobre qué es una asistencia o no.

Aunque, las diferentes ligas dan directrices sobre cómo registrar las acciones de juego (ej. www.galanissportsdata.com/mainsite/basketr.htm), es indudable que los errores existen. Esos errores podrían considerarse aleatorios, de forma que no afectaran al valor esperado, sino a la varianza observable, de forma análoga a como lo hace el error de medida en la Teoría Clásica de los Test. El capítulo de asistencias es el más polémico, ya que tiene a veces un componente subjetivo. Una forma de considerar este hecho sería corrigiendo las varianzas observables por un error asumido a priori (Hayduk, 1996; Ree y Carreta, 2006). Sin embargo, si el error ocurriera de forma sistemática estaríamos ante una situación más difícil de resolver, ya que se necesitarían estudios empíricos para cuantificar ese sesgo. Este hecho no es baladí, ya que existen ciertas sospechas sobre errores sistemáticos de codificación en la NBA, en pos de beneficiar a ciertos jugadores. Desconocemos si en la ACB existen circunstancias de este tipo, pero sería atractivo profundizar al respecto.

En el cálculo de la media móvil, si nuestro método es adoptado, no haría falta considerar sólo los jugadores que hayan jugado todos los partidos de la temporada, sino los que ocuparan los primeros puestos (por ejemplo, los 5 primeros puestos de cada año), ya que estaríamos asegurando que esos 5 puestos tienen estimaciones con error por debajo del máximo admisible.

Otro hecho merecedor de discusión es el que ha ocurrido en el apartado de "Asistencias", donde Ricky Rubio, supera a los jugadores a partir del segundo puesto del ranking, teniendo un intervalo de confianza muy amplio, debido a los pocos partidos que ha jugado esta temporada. Esto podría obligar a reflexionar sobre qué hacer en casos tan especiales como éste, donde un jugador tiene asociado un error bastante grande en su estimación, pero ha tenido una actuación tal que incluso así sus números son superiores al resto. Nuestra propuesta con respecto a esta cuestión es seguir confiando en el criterio del error máximo admisible para la decisión de excluir a los jugadores del ranking, ya que se corre el peligro de que jugadores hagan grandes números jugando pocos partidos en relación al resto, y que esos pocos partidos fueran fuente de sesgo sistemático. Un ejemplo sería un jugador que disputase sólo un 50% de partidos de la competición y sólo ante los rivales más débiles. Por tanto, insistimos en establecer el criterio de error máximo admisible, aunque quizá siendo más permisivos para algunos apartados estadísticos (incrementando, por ejemplo, el valor de G).

Hace ya más de 40 años que Pedro Ferrándiz (el único español en el Hall of Fame), adaptó las tendencias que venían de Estados Unidos, sobre la necesidad de contar con un registro de estadísticas, más allá de los simples anotadores. Han pasado muchos años, y la tecnología para codificar esos datos y la importancia de éstos se ha multiplicado exponencialmente. Sin embargo, es cierto que en algunos sectores del baloncesto aún se mira con cierta reticencia lo que "los números dicen". En su famosa frase, Aaron Levenstein, profesor de Administración de Empresas en el Baruch College de Nueva York, decía que: *"las estadísticas son como los bikinis, lo que revelan es sugerente, pero lo que esconden es vital"*. Muchos toman esta cita como razón para otorgar una importancia menor a los números, de lo que quizá tienen en estos momentos en el mundo del baloncesto. Por tanto, "algo está pasando" con la revolución estadística en el basket, y se ha de intentar innovar y mejorar los sistemas establecidos.

Conclusión

En esta investigación hemos presentado una nueva forma de plantear los rankings estadísticos de la Liga ACB. Reconocemos que nuestra propuesta se basa en técnicas clásicas de inferencia estadística ya existentes, por lo que los procedimientos de cálculo en sí no son novedosos. Sin embargo, sí que lo es la aplicación que hemos descrito a la creación de rankings, lo que produce cambios ostensibles en las clasificaciones de los jugadores en cada apartado estadístico. Este hecho tiene una importancia creciente en el baloncesto actual, ya que algunos jugadores podrían verse beneficiados o perjudicados a nivel mediático y de mercado. Los resultados de esta investigación concernientes a la temporada 2008/2009 así lo atestiguan. Por ello, alentamos a otros investigadores a que discutan las particularidades de nuestro enfoque, o que profundicen en aspectos más metodológicos del mismo. Además, si nuestra propuesta es tenida en consideración, un comité de expertos de la competición debería discutir sobre la fijación del error mínimo admisible.

Finalmente, la extensión de este método a otras categorías estadísticas no comentadas en esta investigación debería realizarse. En cuanto a los apartados estadísticos "Tiros de 1 convertidos", "Tiros de 2 convertidos", y "Tiros de 3 convertidos", la aplicación de nuestro método es totalmente válida. Sin embargo, se necesitaría estudiar más detenidamente los apartados referidos a porcentajes de tiro, ya que los criterios de clasificación deberían tal vez considerar la posible existencia de una función no lineal de ajuste, que tuviera en cuenta la dificultad de mantener un mismo porcentaje de acierto a medida que el número de lanzamientos se incrementa. Esta es otra atractiva línea de investigación para el futuro.

Referencias

Anderson, D. R.; Burnham, K. P., y Thompson, W. L. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *The Journal of Wildlife Management, 64*, 912-923.

Belloni, A., y Didier, G. (2008). On the Behrens--Fisher problem: A globally convergent algorithm and a finite-sample study of the Wald, LR and LM Tests. *Annals of Statistics*, *36* (5), 2377-2408.

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* (8), 411-427.

Berri, D. J. (2008). A simple measure of worker productivity in the National Basketball Association. En Brad Humphreys and Dennis Howard (Eds). *The Business of Sport* (pp. 1-40*);* Westport, Conn: Praeger.

Berri, D. J.; Schmidt, M. B., y Brook, S. L. (2006). *The wages of wins: Taking measure of the many myths in modern sport*. Palo Alto,CA: Stanford University Press.

Casas, J. M. (1997). *Inferencia estadística.* Segunda Edición. Madrid: Editorial Centro de Estudios Ramón Areces, S. A.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist 45* (12), 1304–1312.

Cooper, W. W.; Ruiz, J. L., y Sirvent, I. (2009). Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. *European Journal of Operational Research, 195* (2), 563–574.

Curran-Everett D. (2000). Multiple comparisons: Philosophies and illustrations. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology, 279*, R1-R8.

Felipo, J. (2005). *Fórmulas para ganar: La revolución estadística del basket.* Barcelona: Zona 131.

Grissom, R. J., y Kim, J.J. (2005). *Effect sizes for research. A broad practical approach* Mahwah, NJ: LEA.

Hayduk, L. A. (1996). *LISREL Issues, Debates and Strategies*. Baltimore and London: John Hopkins University Press.

Hesterberg, T.; Moore, D. S., Monaghan, S., Clipson, A., y Epstein, R. (2005). *Bootstrap Methods and Permutation Tests*, 2nd edition. N.Y.: W. H. Freeman.

Hollinger, J. (2005). *Pro Basketball Forecast*. Washington, D.C.: Potomac, 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.

Levy, P.S., y Lemeshow, S. (1999). *Sampling of populations: Methods and applications* (3ª ed). Wiley series in probability and statistics. Survey Methodology Section.

Lewis, M. M. (2003). *Moneyball: The art of winning an unfair game*. W.W. Norton & Company Inc.

Lombardía, M. J.; González-Manteiga, W., y Prada-Sánchez, J. M. (2004). Una revisión de la estimación de la función de distribución y métodos de remuestreo bootstrap en poblaciones finitas. *Estadística española, 46* (155), 149-175.

Lumley, T.; Diehr, P., Emerson, S., y Chen L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review Public Health, 23*, 151-169.

Mayo, D. G., y Cox, D. R. (2006). Frequentists statistics as a theory of inductive inference. En Rojo, J. (ed.), *2nd Lehmann Symposium- Optimally. IMS Lecture Notes-Monographs Series*: 1-28.

McGoldrick, K., y Voeks, L. (2005). We got game! An analysis of win/loss probability and efficiency differences between the NBA and WNBA. *Journal of Sports Economics, 6* (1), 5-23.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66* (Monograph Suppl. 1-V66), 195-244.

Mora, J. L., y Solomon, P. J. (2002). Worrying about normality. *Critical Care and Resuscitation*, *4* (4), 316-319

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241-301.

Olejnik, S., y Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241-286.

Oliver, D. (2004). *Basketball on paper*. *Rules and tools for performance analysis.* Washington, D. C.: Brassey's, INC.

Page, G. L.; Fellingham, G. W., y Reese, C. S. (2007). Using box-scores to determine a position's contribution to winning basketball games. *Journal of Quantitative Analysis in Sports, 3* (4), Article 1.

Ries, A., y Trout, J. (1999). Las 22 leyes inmutables del marketing. McGraww-Hill.

Ree, M. J., y Carreta, T. H. (2006). The role of measurement error in familiar statistics. *Organizational Research Methods, 9* (1), 99-112.

Rosenthal, R., Rosnow, R. L., y Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge University Press.

Rosnow, R. L., y Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for researchconsumers. *Psychological Methods, 1*, 331-340.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology, 1* (1), 43-46.

Rothman, K. J.; Greenland, S., y Lash, T. L. (2008). *Modern Epidemiology*. Third edition. Philadelphia: Luppincott, Williams & Wilkins.

Savitz, D. A., y Olshan, A. F. (1998). Decribing data requires o adjustment for multiple comparisons: a reply form Savitz and Olshan. *American Journal of Epidemiology, 147*, 813-814.

Sawilowsky, S. S. (2002). Fermat, Schubert, Einstein, and Behrens-Fisher: The probable difference between two means when $\sigma1 \neq \sigma2$. *Journal of Modern Applied Statistical Methods, 1* (2), 461-472.

Sawilowsky, S. S. (2005). Misconceptions leading to choosing the t Test over the Wilcoxon Mann-Whitney test for shift in location parameter. *Journal of Modern Applied Statistical Methods, 4* (2), 598-600.

Sterman, J. (2002). All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review, 18* (4), 501–531.

Vegas, E. (1997). El problema de Behrens-Fisher en la investigación biomédica. Análisis crítico de un estudio clínico mediante simulación. *Qüestiió, 21* (1-2), 293-316.

West, B. T. (2006). A simple and flexible rating method for predicting success in the NCAA basketball tournament. *Journal of Quantitative Analysis in Sports, 2* (3), Article 3.

Winston, W. L. (2009). *Mathletics.* New Yersey: Princeton University Press.