



## La ciencia de datos en el baloncesto y el error que cometemos

### Descripción

### Introducción

Vivimos una era en la que el baloncesto está rodeado de números. Rankings, métricas avanzadas, gráficos de impacto, modelos predictivos. Todo parece cuantificable, comparativo y objetivo.

El baloncesto moderno vive una escena curiosa que se repite en vestuarios, despachos y redes sociales. Alguien abre una tabla de estadísticas, señala una cifra con el dedo y pronuncia la sentencia como si estuviera leyendo la presión arterial de un paciente. Aquí está la verdad, dicen, aquí está el rendimiento, aquí está lo que vale este jugador. Luego otro responde con otra tabla, otra web, otro proveedor, otro número distinto con el mismo nombre y la verdad cambia de forma como *slime* deslizándose en la mano.

Lo que debería ser un lenguaje común se convierte en una torre de Babel y lo más inquietante es que, a menudo, ni siquiera estamos discutiendo de interpretación, sino de algo más básico. Estamos discutiendo de si el dato existe como tal.

Si una fuente es débil, el cálculo es frágil. Si el cálculo es frágil, la decisión es un acto de fe con traje de Excel. La ciencia de datos aplicada al baloncesto tiene logros reales y también tiene un pecado original que contamina muchas métricas desde el nacimiento. Se ha llamado datos a lo que con frecuencia son opiniones estandarizadas. Se ha llamado objetividad a lo que en ocasiones es criterio humano disfrazado de oficialidad. Y se ha llamado evaluación individual a lo que, muchas veces, solo mide el acierto gracias a otros

a los que no se mide justamente.

Este artículo no trata de acumular métricas, trata de algo más incómodo, el preguntarnos si estamos midiendo realmente el juego o si estamos midiendo una traducción simplificada del baloncesto y tratándola como si fuera la verdad completa. Al final de este artículo se puede [descargar el informe completo](#).

El baloncesto sucede en continuo. Diez jugadores se desplazan, interactúan, generan ventajas, compensan errores y reaccionan en fracciones de segundo. Sin embargo, nuestras estadísticas tradicionales convierten ese fenómeno dinámico en una lista de eventos discretos: tiros, rebotes, asistencias, pérdidas.

El box score no registra procesos. Registra desenlaces, ni mide cómo se creó la ventaja, solo quién anotó, ni la rotación que evitó una penetración, solo el tapón, si es que lo hubo.

En ese tránsito del juego real al dato estadístico se produce una traducción. Y toda traducción implica pérdida.

El problema no es simplificar, pues toda medición simplifica. El problema aparece cuando olvidamos que estamos simplificando y tratamos el conteo como si fuera equivalente al fenómeno completo.

## **Lo oficial no siempre es científicamente sólido**

Durante años se ha asumido que las estadísticas oficiales son fiables por definición. Pero oficial no significa metodológicamente robusto, sino reglamentado.

Existen categorías con alto componente interpretativo, como la asistencia o el tapón. Diferentes anotadores pueden aplicar criterios ligeramente distintos ante situaciones ambiguas.

La investigación académica ha demostrado que existen variaciones sistemáticas asociadas al pabellón o al anotador, especialmente en estadísticas subjetivas. Si un registro necesita ser ajustado mediante modelos para poder compararse entre contextos, entonces ese registro no es un suelo completamente firme. Es una capa más del sistema, con su propio margen de error.

Las estadísticas oficiales son útiles, pero no son neutras.

Es habitual que dos plataformas ofrezcan la misma métrica con valores diferentes. El problema rara vez es un error simple. Suele estar en los supuestos.

- ¿Se calcula por partido o por posesión?
- ¿Cómo se estiman las posesiones?
- ¿Qué umbral define un tiro contestado?
- ¿Cómo se tratan los datos perdidos en tracking?
- ¿Es una métrica observada o inferida?

Muchas métricas avanzadas no están estandarizadas por ninguna liga, siendo construcciones analíticas legítimas, pero no reguladas. Y cuando el procedimiento no es transparente, la reproducibilidad se resiente. Sin reproducibilidad no hay ciencia, solo hay resultados.

## Decisión y azar, el error clásico

Un triple completamente liberado puede fallarse. Un triple forzado puede entrar.

Si sólo evaluamos el resultado, premiamos la mala decisión y castigamos la buena.

La estadística clásica tiende a confundir proceso con desenlace. Los modelos basados en valor esperado corrigen parcialmente este problema y no preguntan solo qué ocurrió, sino qué era razonable esperar dadas las condiciones del momento.

Separar decisión y azar es fundamental si queremos evaluar coherentemente el rendimiento.

El baloncesto no es una suma lineal de acciones individuales, es un sistema de interacciones. El rendimiento depende del rol.

Un base con muchas asistencias puede estar generando juego o monopolizándolo. Un pívot con pocos puntos puede estar sosteniendo el sistema con bloqueos y sellados. Un alero con números discretos puede estar multiplicando la eficiencia colectiva.

Cuando ignoramos el rol, aplicamos el mismo patrón de evaluación a funciones distintas y eso produce errores estructurales en scouting, contratos y desarrollo.

Las métricas basadas en plus-minus intentan capturar impacto colectivo ajustando por compañeros y rivales. Son un avance frente al conteo simple, pero presentan problemas estructurales como la colinealidad y la dependencia contextual. La regularización

estabiliza estimaciones, pero no convierte el modelo en explicación táctica.

Cuantifican impacto, pero no explican comportamiento. Y sin comportamiento identificable, no hay entrenamiento accionable.

Cuando el número se convierte en arma retórica, deja de ser herramienta de aprendizaje.

Toda métrica moldea incentivos. Si premiamos volumen, incentivamos volumen, si castigamos pérdidas sin contexto, reducimos creatividad y si premiamos tapones, incentivamos riesgo innecesario. Medir no es neutral.



## El salto del evento al proceso

El verdadero cambio de paradigma llega con el tracking óptico. Por primera vez, el baloncesto puede registrarse como fenómeno espaciotemporal.

Ya no solo sabemos que hubo un tiro, sabemos dónde estaban los diez jugadores antes del tiro, cuánto espacio había, qué rotación se produjo y cuánto cambió la expectativa de puntos tras cada decisión.

El tracking no elimina la interpretación, pero cambia la naturaleza del dato. Pasamos de eventos etiquetados por humanos a coordenadas medibles en tiempo real y eso permite modelar procesos.

Y cuando modelamos procesos, podemos empezar a separar decisión y resultado.

## **Qué aporta realmente el data science cuando se hace bien**

Bien aplicado, el data science:

- Detecta patrones en miles de posesiones.
- Distingue señal estructural de fluctuación coyuntural.
- Separa correlación espuria de relación estable.
- Personaliza feedback individual.
- Explicita incertidumbre y estabilidad.

También descubre valor en lo invisible, de defensores que no aparecen en el box score, espaciadores que alteran geometrías o jugadores que sostienen el sistema sin volumen ofensivo.

Pero solo si se respeta el orden metodológico:

1. Medir bien.
2. Definir bien.
3. Modelar con propósito claro.
4. Interpretar con rol y contexto.
5. Validar con vídeo y estabilidad temporal.

Si se hace al revés, el análisis se convierte en justificación.

## **Conclusiones**

El problema no está en las estadísticas, está en olvidar que son una traducción del juego.

Cuando tratamos una medición interpretada como si fuera el fenómeno completo, cometemos un error conceptual. Cuando confundimos resultado con calidad de decisión, mezclamos azar con rendimiento. Cuando ignoramos el rol, confundimos el mapa con el territorio.

---

La ciencia de datos aplicada al baloncesto puede acercarnos al juego real, pero solo si acepta sus límites, declara sus supuestos, muestra su incertidumbre y contextualiza sus métricas. El rigor no está en tener más números, está en saber qué miden, cómo se construyen y cuándo dejan de ser estables.

Cuando no mostramos incertidumbre, lo que vendemos no es ciencia, es fe con gráficos bonitos.

## Descarga el informe completo

Este artículo es un resumen divulgativo del informe técnico: “Análisis de modelos estadísticos en baloncesto. La ciencia de datos en el baloncesto y el error que cometemos”

En el informe completo encontrarás:

- Desarrollo metodológico detallado
- Marco formal de evaluación por rol
- Umbrales mínimos de estabilidad por variable
- Modelos matemáticos completos
- Referencias científicas contrastadas
- Propuesta de guía operativa para staff técnico

Si este artículo te ha resultado útil y quieres profundizar en el marco completo, puedes descargar el informe íntegro desde aquí:



[Descarga el informe](#)

## Referencias

- Cervone, D., D'Amour, A., Bornn, L., & Goldsberry, K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association*, 111(514), 585–599. <https://doi.org/10.1080/01621459.2015.1108174>
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. John Wiley & Sons.
- Deshpande, S. K., & Jensen, S. T. (2016). Estimating an NBA player's impact on his team's chances of winning. *Journal of Quantitative Analysis in Sports*, 12(2), 51–72. <https://doi.org/10.1515/jqas-2015-0027>
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38(1), 105–123. <https://www.erpjournal.net/wp-content/uploads/2020/02/ERPv38-1.-Drost-E.-2011.-Validity-and-Reliability-in-Social-Science-Research.pdf>
- Fédération Internationale de Basketball. (2024). *FIBA statisticians' manual 2024 (Version 1.0)*. <https://assets.fiba.basketball/image/upload/documents-corporate-fiba-statisticians-manual-2024.pdf>
- Franks, A., Miller, A., Bornn, L., & Goldsberry, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics*, 9(1), 94–121. <https://doi.org/10.1214/14-AOAS799>
- Ghimire, S., Ehrlich, J. A., & Sanders, S. D. (2020). Measuring individual worker output in a complementary team setting: Does regularized adjusted plus minus isolate individual NBA player contributions? *PLOS ONE*, 15(8), e0237920. <https://doi.org/10.1371/journal.pone.0237920>
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Hvattum, L. M. (2019). A comprehensive review of plus-minus ratings for evaluating individual players in team sports. *International Journal of Computer Science in Sport*, 18(1), 1–23. <https://doi.org/10.2478/ijcss-2019-0001>

- National Basketball Association. (2013, September 5). NBA expands partnership with STATS LLC to unveil player tracking technology in every NBA arena [Press release]. <https://pr.nba.com/nba-stats-llc-partnership/>
- National Basketball Association. (2016, January 19). STATS LLC and NBA to make STATS SportVU player tracking data available to teams, media and fans [Press release]. <https://pr.nba.com/stats-llc-nba-sportvu-player-tracking-data/>
- Pelechrinis, K., & Winston, W. L. (2020). The plus-minus statistic and the evaluation of NBA players. *PLoS ONE*, 15(8), e0237920. <https://doi.org/10.1371/journal.pone.0237920>
- Second Spectrum. (s.f.). Player tracking and analytics technology in professional basketball. <https://www.secondspectrum.com>
- Sill, J. (2010, March 6). Improved NBA adjusted +/- using regularization and out-of-sample testing. In Proceedings of the MIT Sloan Sports Analytics Conference. <https://www.sloansportsconference.com/research-papers/improved-nba-adjusted-using-regularization-and-out-of-sample-testing>
- van Bommel, M., & Bornn, L. (2016). Adjusting for scorekeeper bias in NBA box scores (arXiv:1602.08754). arXiv. <https://arxiv.org/abs/1602.08754>
- van Bommel, M., & Bornn, L. (2017). Adjusting for scorekeeper bias in NBA box scores. *Journal of Quantitative Analysis in Sports*, 13(3), 119–133. <https://doi.org/10.1515/jqas-2016-0075>