

SEED - Noticias

<https://www.seed-deporte.es/>

Big data, alquimia, ciencia y buenas prácticas



La revista *Kinesiology Review* tiene en la recámara, a punto de ser publicado en 2020, un artículo de Stefan Szymanski con título “*Sports analytics: Science or Alchemy?*”. La reputación como uno de los mejores economistas del deporte del autor del trabajo, así como el atractivo título del mismo, captaron mi atención, como probablemente captarán la atención de aquellos que trabajamos en este campo. En mi caso, además se da la circunstancia de que el título me retrotrajo a mis tiempos de doctorando, pues tiene muchas similitudes con el de un artículo de David Hendry publicado en *Economica* en 1980, “*Econometrics: Alchemy or Science?*”, en el que se describen los fundamentos de la tradición econométrica de la *London School of Economics and Political Science*, que durante muchos años fue referente para el análisis macroeconómico. A pesar de la similitud de los títulos, el artículo de David Hendry no aparece referenciado en el trabajo de Stefan Szymanski, aunque ambos comparten una visión parecida de cómo debe abordarse el trabajo empírico, aparentemente en dos contextos distintos: *Econometrics* y *Sports Analytics*. Visión que, ya me avanzo, comparto totalmente. Stefan Szymanski entiende *Sports Analytics* como “el uso de datos sobre rendimiento para entender la estructura de la competición deportiva con el objetivo de obtener una ventaja competitiva”. La idea adquiere especial relevancia con el fenómeno del *Big Data*: “la utilización de grandes bases de datos para descubrir patrones de comportamiento hasta hace poco fuera del alcance de la capacidad de cómputo de los ordenadores más avanzados”. A través de un sencillo ejercicio de predicción de los resultados de los 40 primeros partidos jugados en la *Premier League* en esta temporada, Szymanski constata

que los beneficios de utilizar modelos complejos con un uso intensivo de información no son tan grandes como se esperaría, en comparación con modelos menos complejos, basados en menos información, pero con un fundamento teórico detrás, aspecto que no se da en la mayoría de aplicaciones de *Big Data*. A pesar de lo que algunos creen y predicán, tal y como señala Szymanski, “disponer de (muchos) datos no comporta entender los procesos y relaciones que hay detrás de los mismos”, y en muchas ocasiones, “la complejidad no es necesariamente beneficiosa”. La teoría es un ingrediente indispensable para el (buen) análisis científico. Su ausencia, junto a la poca transparencia en ocasiones de lo que entendemos como *Sports Analytics*, acerca más esta disciplina a la alquimia que a la ciencia. Este papel de la teoría en el análisis económico empírico queda también recogido en el trabajo de David Hendry en la parte final del mismo: “cualquier conclusión obtenida, para ser catalogada de científica, se verá beneficiada del hecho de estar basada en modelos rigurosamente contrastados, que describan de manera adecuada la evidencia disponible, que den cuenta de los resultados obtenidos con anterioridad y que se deriven de teorías bien fundamentadas”. Para Hendry las tres reglas de oro del trabajo econométrico que destacan el papel de la teoría son: “contrastar, contrastar y contrastar”. Esta relevancia del papel de la teoría, que no parece darse suficientemente en *Sports Analytics*, sí que parece asumida en nuestra disciplina, pues en las revistas académicas de referencia tanto de economía como de gestión del deporte, se observa un creciente énfasis en exigir que los trabajos empíricos siempre tengan un marco de referencia en el que los resultados puedan ser interpretados, aunque, en ocasiones, parezca haber un cierta desconexión entre el modelo teórico y el análisis empírico. Pero al mismo tiempo, mi impresión es que no se presta la suficiente atención por parte de evaluadores y editores al análisis de los datos y a los aspectos de modelización. Es como si se diera por supuesto que, por defecto, esa parte siempre está bien resuelta. La resuelven los programas estadísticos o econométricos. Nada más lejos de la realidad. En un momento en que la eclosión del *Big Data* ha hecho que aparezcan “nuevas” disciplinas (*Data Science*) y, supuestamente, nuevas profesiones (*data scientist*) hay que reivindicar el valor añadido que aquellos que realizan trabajo empírico deben aportar porque, a fecha de hoy, la tecnología todavía no lo aporta. Es decir, el conocimiento de los datos con los que se trabaja, la capacidad de elegir los instrumentos adecuados o la interpretación de los resultados. Aspectos que, como el “valor”, a veces se dan por supuestos pero desgraciada o afortunadamente no es así. Como indica Stefan Szymanski en un artículo de opinión “*Soccer Analytics: Science or Alchemy?*”^[1], que es precursor del trabajo anteriormente mencionado, “la palabra “*learning*” en “*Machine learning*” no debe hacernos creer que los ordenadores vayan a hacer algo más que optimizar”, al igual que los métodos utilizados anteriormente ya hacían. Es importante que vayamos incorporando en nuestra disciplina los avances que se vayan produciendo tanto en la gestión de la información como en el tratamiento estadístico y econométrico de la misma, pero siendo conscientes del papel a desarrollar por el investigador para hacer un uso competente de dichos avances. En este sentido, hay que destacar propuestas como la del tema del próximo Congreso de Economía del Deporte de Gijón, dedicado a “*Modern Sport Econometrics*” en la medida en que las nuevas herramientas contribuyan a dar mejores respuestas a las cuestiones relevantes en nuestro campo y no sean únicamente meros ejercicios de sofisticación. Bienvenidos todos los avances tecnológicos y metodológicos que ayuden al desarrollo del análisis empírico en economía y gestión del deporte, bienvenido lo que se conoce como *Big Data*, pero no perdamos de vista el papel crucial que el análisis cuidadoso de la información y la adecuada modelización juegan en este ámbito. [1] <https://www.soccernomics-agency.com/?p=1575> Jaume García Villar Universitat Pompeu Fabra. Miembro de UPF

Sports_Lab

Link to Original article: <https://www.seed-deporte.es/noticias/big-data-alquimia-ciencia-y-buenas-practicas?elem=295422>