

Selección de variables aplicada a la identificación de factores de riesgo de lesión en fútbol profesional

María del Carmen Iglesias Pérez¹, Miguel Martínez González², Luis Casáis Martínez²,
Marta Sestelo¹ e Javier Roca Pardiñas¹

¹Departamento de Estadística e I.O. Universidad de Vigo.

²Facultad de Ciencias de Educación y Deporte. Universidad de Vigo.

RESUMEN

En este trabajo se investigan factores de riesgo de lesión en fútbol profesional mediante regresión logística. El elevado número de variables independientes respecto al tamaño muestral disponible y las correlaciones existentes entre las variables llevan a utilizar distintos métodos de selección de variables.

Palabras y frases clave: lesión en fútbol, regresión logística, selección de variables.

1. INTRODUCCIÓN

A pesar de que son numerosas las referencias acerca de la epidemiología lesional y sus factores y mecanismos de producción en el fútbol, se carece de datos suficientes sobre el control del estado neuromuscular en futbolistas de élite y sobre su relación como indicadores de riesgo de lesión en esa misma población.

El objetivo de este trabajo es la identificación de factores de riesgo de lesión en el fútbol profesional, más concretamente de las áreas anatómicas de la articulación de la rodilla y de la musculatura del muslo (cuádriceps e isquiotibial). Entre las variables a investigar como factores de riesgo se han incluido variables que definen el estado y relación funcional de grupos musculares o articulaciones (rodilla) mediante tecnología actual y de vanguardia como la Tensiomiografía, así como variables obtenidas por distintos medios de la relación y simetría entre grupos musculares con una alta incidencia lesional en fútbol. En total se consideraron 57 variables independientes, medidas a 30 jugadores de un equipo profesional de 2ª división durante la temporada 2007/8. Como variable dependiente se consideró la variable dicotómica Lesión (sí/no). La información recogida por Martínez (2012) es de alto valor al proceder de una muestra de élite y difícil acceso.

2. METODOLOGÍA

Para medir las variables independientes se utilizaron una serie de instrumentos suficientemente validados en la literatura y se siguieron cuidadosamente los protocolos establecidos en cada uno de ellos.

Las mediciones se realizaron en 2 momentos puntuales de la temporada, al principio de la primera vuelta y al principio de la segunda vuelta de la competición de liga.

El orden seguido en la secuencia de las pruebas fue: Tensiomiografía (tono muscular en distintos grupos musculares de la pierna dominante y no dominante, 30 variables), Flexibilidad (4 variables), Test de Bosco (salto CMJ=CounterMovementJump) e Isocinético (fuerza, 20 variables). Además se consideraron las variables Lesión previa y Edad, sumando 57 variables independientes.

La muestra de jugadores en la primera vuelta estuvo formada por los 28 jugadores de la plantilla oficial, mientras que en la segunda vuelta se recogieron datos de 24 de ellos y de 2 fichajes nuevos.

Para identificar los factores de riesgo de lesión en cada vuelta se utilizó regresión logística. Previamente, se efectuaron comparaciones de las variables por grupos (lesión/ no lesión) con la

prueba t y la U de Mann-Whitney y también ACP para reducir el número inicial de variables a seleccionar mediante el método clásico de selección de variables por pasos sucesivos. Dado el elevado número de variables independientes en relación al tamaño muestral se pretende comparar los resultados obtenidos por el procedimiento clásico anterior con los obtenidos mediante tres procedimientos de selección de variables más recientes:

- El mejor subconjunto de q variables, implementado en el paquete bestglm de R (McLeod and Xu (2011)).
- El algoritmo propuesto en Sestelo y otros (2013).
- El método de Lasso, implementado en el paquete glmnet de R (Friedman et al.(2013)).

3. RESULTADOS Y CONCLUSIONES

A continuación se presentan los resultados del análisis clásico, que se compararán con los obtenidos mediante los tres métodos de selección de variables citados en la sección anterior.

En la primera vuelta, los factores de riesgo seleccionados en las comparaciones por grupos fueron DMRectoFemoralDominante, TDRectoFemoralDominante y SimetríaRodillaDominante (Variables de Tensiomiografía; DM=Desplazamiento muscular, TD=Tiempo de reacción muscular).

La regresión logística con selección a partir de todas las variables no converge, pero partiendo de las variables significativas en la comparación por grupos se seleccionaron: DMRectoFemoralDominante, SimetríaRodillaDominante y Lesión previa.

La regresión logística a partir de las componentes principales por ACP selecciona aquellas componentes con mayor peso en DMRectoFemoralDominante, TDRectoFemoralDominante y TDVastoMedialDominante.

En la segunda vuelta hubo 12 variables significativas en la comparación por grupos:

- TDRectoFemoralDominante, TDRectoFemoralNoDominante, TCRectoFemoralNoDominante, TCBícepsFemoralDominante, TCBícepsFemoralNoDominante, TDBícepsFemoralNoDominante, TCVastoMedialNoDominante y SimetríaRodillaDominante (p-valor < 0.05 en U Mann-Whitney o en prueba t)

- TDVastoMedialNoDominante, SimetríaRodillaNoDominante, CMJ y RatioCONNoDominante_60 (p-valor < 0.10 en U Mann-Whitney o en prueba t).

La regresión logística seleccionó, partiendo de las variables anteriores, el TCRectoFemoralNoDominante, TCBícepsFemoralNoDominante (TC=tiempo de contracción) y CMJ.

Respecto a la regresión logística por pasos a partir de las componentes principales obtenidas por ACP se seleccionan las componentes con mayor peso en DMRectoFemoralNoDominante, TCRectoFemoralNoDominante, TDVastoMedialNoDominante, TCVastoMedialNoDominante, TCVastoMedialDominante, TCVastoLateralNoDominante, TCVastoLateralDominante, TDVastoLateralNoDominante, TDVastoLateralDominante, TDRectoFemoralDominante y TDRectoFemoralNoDominante.

Como conclusión de este análisis obtenemos que la Tensiomiografía aporta información relevante para el riesgo de lesión, así como el CMJ.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el proyecto MTM2011-23204 (fondos FEDER incluidos) del Ministerio de Ciencia de España y por el proyecto 10PXIB300068PR de la Xunta de Galicia (España).

REFERENCIAS

Friedman, J., Hastie, T. and Tibshirani, R. (2013). *glmnet: Lasso and elastic-net regularized generalized linear models*. R package version 1.8-5.

- Martínez, M. (2012). *Predictores de lesión artromuscular en futbolistas profesionales*. Tesis doctoral, Universidad de Vigo.
- McLeod, A.I. and Xu, C. (2011). *bestglm: Best Subset GLM*. R package version 0.33.
- Sestelo, M., Villanueva, N.M., Roca-Pardiñas, J. (2013). **FWDselect**: An R package for selecting variables in regression models. *Discussion Papers in Statistics and Operation Research*, 13/02.