



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

# ¿CÓMO EVALUAR EL RENDIMIENTO DE LOS ARQUEROS EN EL FÚTBOL?

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN ECONOMÍA  
APLICADA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL  
INDUSTRIAL

BENJAMÍN IGNACIO MONTESINO URRUTIA

PROFESOR GUÍA:  
ALEJANDRO CORVALÁN AGUILAR

MIEMBROS DE LA COMISIÓN:  
PABLO GALAZ CARES  
ANDRÉS MUSALEM SAID

SANTIAGO DE CHILE

2023

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE MA-  
GÍSTER EN ECONOMÍA APLICADA Y MEMORIA PARA OP-  
TAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL  
POR: **BENJAMÍN IGNACIO MONTESINO URRUTIA**  
FECHA: 2023  
PROFESOR GUÍA: ALEJANDRO CORVALÁN AGUILAR

## ¿CÓMO EVALUAR EL RENDIMIENTO DE LOS ARQUEROS EN EL FÚTBOL?

Cada año los clubes de fútbol profesional buscan contratar a los mejores jugadores según su presupuesto para reforzar el equipo de cara a una nueva temporada. Para esto, es relevante conocer cuales son los atributos buscados en los jugadores prospectos para cada posición. Son varios los equipos de fútbol que han ido incorporando la ciencia de datos y estadística a los análisis realizados por las secretarías técnicas de cara a los procesos de selección. Esto hace relevante mejorar las métricas utilizadas para evaluar jugadores según la posición dentro del campo de juego.

El objetivo principal de la presente tesis es discutir una variable de desempeño y medir el rendimiento de los arqueros de las ligas chilena y argentina en el año 2021. En particular, se busca validar la variable propuesta y la metodología utilizada para la generación de un ranking de arqueros, con la intención de ser extrapolada para el análisis de las posiciones restantes dentro de la cancha.

Actualmente, para evaluar el rendimiento de los arqueros se utilizan mayoritariamente, los goles concedidos, las atajadas/salvadas realizadas y la cantidad de partidos sin goles. Estas variables pueden traer ciertos problemas al momento de medir la habilidad del portero, debido a que son susceptibles a sobre estimaciones y a sesgos. Para eliminar en lo posible estos problemas se propone una nueva variable de medición: Salvadas esperadas controladas por los tiros al arco que el arquero enfrentó,  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ . Esta variable incorpora la dificultad de los tiros a portería y a su vez, ofrece un intervalo continuo en la medición de rendimiento.

Para medir el poder explicativo de  $\frac{(xG - Goles)}{Tiros\ al\ arco}$  se realizan distintas regresiones lineales OLS para evaluar el ajuste con respecto al valor de mercado de los arqueros. Adicionalmente se hace competir contra Goles Concedidos con la intención de comparar el efecto de estas sobre el valor de mercado. Se concluye que  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ , a pesar de tener un rendimiento levemente menor a Goles Concedidos, permite identificar y ordenar a los arqueros en un ranking que se relaciona de forma significativa, positiva e intuitiva con el valor de mercado. Del ranking, se obtiene como resultado a Ignacio González arquero del Club Deportes Antofagasta en la temporada 2021, como el arquero con mejor rendimiento de la liga chilena y argentina. También se realiza una estimación del valor de mercado de los arqueros según la métrica propuesta, concluyendo que Lucas Hoyos, portero de Vélez Sarsfield, es el arquero más sub valorado por el mercado. Esto, a su vez, nos entrega oportunidades de mercado, en el sentido de que su precio es apenas el 30% de su valor estimado por el modelo utilizando  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ .

*A Gladys.*

*Saludos*

# Agradecimientos

Ha sido un camino largo donde he recibido un apoyo incondicional durante todos estos años. Primero que todo, quiero recordar a quienes hoy no están presentes físicamente, a Gladys, la razón y el motivo de esfuerzo de todos estos años. A la Juany y Bernardita, que siempre vivirán en mi memoria.

Quiero agradecerle a mi familia, a Francisco, Claudia, Joaquín y Cristóbal, por invadir la casa y tener que soportarme. De forma especial a Catalina, por ser mi motor de estos últimos años. También a la Maggi, a la Toti y al Pato, por la preocupación constante.

A mis amigos del colegio que fueron fundamentales los primeros años, en especial a Nacho Gálvez, Vale Domínguez y Matute, por los maratones de estudio y las risas que nunca faltaron. Quiero agradecerles también a Nico, Bubu, Mati, Palma, Edu y Félix, por ser los mejores amigos que me podría haber imaginado y por todo lo que vivimos en industrias, gracias por todo.

No quiero dejar de saludar a Pablo, por darme la oportunidad de hacer la tesis en un tema que me gusta tanto, y a la Coni por la buena onda y el apañe en Futbolytics. A la 601 por todo el compañerismo y el ánimo en todo el año pasado.

Finalmente agradecer a Alejandro Corvalán por ser mi profesor guía y apoyarme en cada duda y en cada idea que tuve a lo largo de la construcción de esta tesis.

Gracias totales.

# Tabla de contenido

Capítulo 1 : Introducción .....	1
1.1. Motivación .....	1
Capítulo 2 : Marco Teórico .....	5
2.1. ¿Cómo se mide el rendimiento en el mercado laboral? .....	5
2.2. ¿Cómo se mide el rendimiento en los deportes? .....	7
2.3. ¿Cómo se mide el rendimiento en el Fútbol? .....	9
2.4. ¿Cómo se mide el rendimiento de los Arqueros? .....	10
Capítulo 3 : Midiendo el rendimiento .....	13
Capítulo 4 : Datos.....	15
4.1. InStat.....	15
4.2. Transfermarkt .....	18
4.3. Sports Reference.....	19
4.4. Correlaciones.....	19
4.5. Ranking .....	24
Capítulo 5 : Metodología.....	25
Capítulo 6 : Resultados.....	27
Capítulo 7 : Discusión.....	33
Bibliografía .....	36

# Índice de Tablas

TABLA 1: RESUMEN VARIABLES INSTAT .....	16
TABLA 2: RESUMEN VALOR DE MERCADO .....	18
TABLA 3: RANKING ARQUEROS TEMPORADA 2021 .....	24
TABLA 4: ESTIMACIÓN VALOR DE MERCADO ARQUEROS. VARIABLES REVI- SIÓN BIBLIOGRÁFICA .....	27
TABLA 5: ESTIMACIÓN VALOR DE MERCADO ARQUEROS. VARIABLES LASSO. .....	29
TABLA 6: ESTIMACIÓN VALOR DE MERCADO ARQUEROS. VARIABLES LASSO Y PASES CORRECTOS .....	31
TABLA 7: EVALUACIÓN AUTOREGRESIVIDAD .....	32
TABLA 8: RANKING ARQUEROS MÁS SUB VALORADOS .....	34
TABLA 9: RANKING ARQUEROS MÁS SOBRE VALORADOS .....	34

# Índice de Figuras

FIGURA 1: DISTRIBUCIÓN GOLES CONCEDIDOS.....	17
FIGURA 2A: DISTRIBUCIÓN SALVADAS ESPERADAS.....	17
FIGURA 2B: DISTRIBUCIÓN SALVADAS ESPERADAS VS TIROS AL ARCO .....	17
FIGURA 3A: DISTRIBUCIÓN VALOR DE MERCADO.....	19
FIGURA 3B: DISTRIBUCIÓN LOGARITMO VALOR DE MERCADO .....	19
FIGURA 4: CORRELACIÓN LOGARITMO VALOR DE MERCADO Y GOLES CONCEDIDOS.....	20
FIGURA 5: CORRELACIÓN LOGARITMO VALOR DE MERCADO Y GOLES CONCEDIDOS POR ARQUERO Y LIGA.....	21
FIGURA 6: CORRELACIÓN LOGARITMO VALOR DE MERCADO Y SALVADAS ESPERADAS POR ARQUERO Y LIGA .....	22
FIGURA 7: CORRELACIÓN LOGARITMO VALOR DE MERCADO Y SALVADAS ESPERADAS VS TIROS AL ARCO POR ARQUERO Y LIGA.....	22
FIGURA 8: CORRELACIÓN DISPERSIÓN VS VALOR DE MERCADO .....	23

# Capítulo 1: Introducción

## 1.1. Motivación

Los clubes de fútbol profesional buscan contratar jugadores para reforzar el equipo de cara a una nueva temporada. Para esto, el problema a resolver es traer a los mejores jugadores posibles sujeto a una restricción presupuestaria. En particular, se busca detectar talento que el mercado aún no es capaz de ver. Para esto es relevante detectar cuáles son las variables por posición en la cancha que determinan si un jugador es talentoso o no. Usando estas variables, se podría contratar refuerzos de alta calidad que el mercado aún no valora y adquiriéndolos a un menor precio.

Actualmente, los clubes de fútbol están incorporando cada vez más las decisiones guiadas por datos a sus procesos de selección de jugadores, con la intención de quitar cualquier sesgo que exista al momento de decidirse por un fichaje nuevo. Un ejemplo de esto, es la hackatón organizada por el Sevilla FC a principios de junio del 2022, la cual tuvo como desafío mejorar el algoritmo de búsqueda de jugadores, para que identifique que variables definen a un buen jugador por posición. El mismo equipo, ha sido pionero en contratar jugadores basado en los datos y necesidades del club. En el verano de 2019, el jugador, Diego Carlos fue fichado en un proceso que consideró alrededor de 4.000 defensas centrales. El club, mediante al uso de grandes bases de datos, fue acotando jugadores hasta dejar a aquellos que tenían los mejores promedios de La Liga en duelos aéreos ganados (62%) y pases largos completados (43%), entre otras variables. Finalmente, considerando opciones por edad, personalidad, coste y otras restricciones, él brasileño fue elegido. Costó 15 millones €, rindió tres temporadas a un gran nivel y fue traspasado al Aston Villa, por aproximadamente 33 millones de euros<sup>1</sup> (Lacort, 2022).

No solo el Sevilla FC ha incorporado el uso de grandes datos a la toma de decisiones estratégicas, equipos de la Premier League también lo han incorporado como herramienta. En Soccernomics (Kuper, 2018), se relata que el departamento de estadística del Manchester City realizó un estudio sobre los tiros de esquinas, estudiando más de 400 tiros de diferentes

---

<sup>1</sup> Lacort, J. (2022, 6 de junio). Monchi y el hackaton del Sevilla FC: "Queremos que el algoritmo responda a qué es un buen jugador posición por posición". Xataka. <https://www.xataka.com/empresas-y-economia/monchi-hackaton-sevilla-fc-queremos-que-algoritmo-responda-a-que-buen-jugador-posicion-posicion>



ligas y temporadas, llegando a la conclusión que los tiros de esquinas lanzados al primer palo eran los más peligrosos para el rival. El departamento llevó el resultado a Roberto Mancini, entrenador del Manchester City en ese momento, quien en primera instancia no se convenció de la veracidad del estudio. De igual forma, tomó la recomendación y en la temporada 2011/2012 el Manchester City anotó 15 goles de tiros de esquina más que el resto de los equipos de la Premier League.

En Chile, el uso de datos en el deporte se ha vuelto cada vez más relevante en varias disciplinas, como el fútbol, el tenis y el atletismo. En el fútbol, se han implementado diversas herramientas tecnológicas para recopilar y analizar datos de los jugadores, tales como sistemas de seguimiento GPS, análisis de video y plataformas de software especializado en el análisis de datos deportivos. El Club de Fútbol Universidad de Chile tiene una alianza con Catapult Sports, la cual es una empresa líder a nivel mundial en el desarrollo de tecnología para el rendimiento deportivo, la cual mediante el uso de petos con sensores genera datos de seguimiento de GPS, estado físico, entre otros de los futbolistas. Los cuales son utilizados por el cuerpo técnico para analizar el rendimiento de los jugadores y equipos, identificar fortalezas y debilidades, ajustar tácticas y estrategias, planear el trabajo físico y dosificar las cargas de volumen e intensidad, teniendo un óptimo control del entrenamiento. Esta tecnología también es utilizada para tomar decisiones informadas en tiempo real durante los partidos<sup>2</sup>.

A pesar de entregar información valiosa e importante a los clubes y sus cuerpos técnicos, estos datos y sus análisis, se pueden considerar como descriptivos, dejando espacio para la extrapolación a nuevos y mejores usos de estos. Existen también empresas como Futbolytics, start-up de análisis de datos que provee servicios a clubes profesionales de fútbol, que buscan agregar valor a la información descriptiva recopilada por proveedores de datos como Catapult. Futbolytics ([www.futbolytics.cl](http://www.futbolytics.cl)) ofrece una plataforma con diferentes funciones, entre estas, la comparativa de jugadores, en donde se puede realizar una comparación entre futbolistas que tienen la misma posición dentro de la cancha. Por lo tanto, son capaces de generar valor utilizando datos descriptivos para entregar información valiosa a clubes. Como, por ejemplo, si un club busca maximizar los ataques aéreos, la plataforma entrega una comparativa de delanteros, analizando cabezazos, tiros al arco, movimientos sin balón, entre otros. Esto representa una diferenciación con respecto a la oferta de recopilación de datos,

---

<sup>2</sup> Nos unimos a Catapult Sports para optimizar el trabajo físico del plantel. (2019, 22 de febrero). Club Universidad de Chile. <https://www.udechile.cl/nos-unimos-a-catapult-sports-para-optimizar-el-trabajo-fisico-del-plantel/>

debido a que, le entrega sentido del negocio a los análisis. Las recomendaciones entregadas por la empresa permiten ahorrar tiempo valioso, que actualmente se utilizan para analizar horas de videos de jugadores prospectos, mejorando la calidad de la información y haciendo más eficiente la toma de decisiones.

Siguiendo esta línea de generar valor, se pretende estudiar cuáles son las variables de cada posición que estadísticamente se relacione con el rendimiento de los equipos de fútbol. A partir de variables descriptivas, como lo son balones perdidos, pases correctos o tiros al arco, se quiere evaluar qué variables afectan el rendimiento del equipo. Teniendo en consideración que las posiciones tienen distintos roles. Hace sentido pensar que, para un delantero es más importante el porcentaje de tiros al arco que para un lateral, de la misma forma, para un lateral el porcentaje de balones recuperados es más valioso que para un delantero.

Es común hablar que un equipo jugó bien o mal, pero muchas veces estas afirmaciones son percepciones subjetivas que carecen de análisis sustentados en datos. Dado que, perder un partido puede ser debido a que el equipo rival fue superior o porque el equipo no convirtió goles. Esto puede ser consecuencia de que el delantero no estuvo preciso al momento de patear al arco, o porque el partido se jugó mayoritariamente en el campo propio, donde los mediocampistas y los defensas tuvieron actuaciones más defensivas, lo cual no quiere decir que el equipo como grupo haya tenido un mal rendimiento. Entonces, afirmar que un equipo jugó bien o mal por el resultado, es una mirada miope, debido a que el resultado pudo ser una mala actuación individual como grupal.

Como primer acercamiento se estudiarán a los arqueros, dado que evaluar al resto de los jugadores y aislar el rendimiento de cada posición es complicado, ya que, en general el rendimiento de los jugadores se ve afectado por sus mismos compañeros o el equipo contrario. Por ejemplo, para un delantero su rendimiento se puede ver afectado por si mismo o por las interacciones con sus compañeros de equipos, como recibir pocos centros o pases. Es por esto, que se pretende estudiar a los arqueros, dado que a pesar de depender de cómo sea el rendimiento de sus defensas y cuanto lo protejan de ataques del contrario, al momento de recibir un tiro al arco, este se encuentra solo contra el balón y se puede evaluar su respuesta ante el ataque rival. En resumen, se puede aislar o controlar más fácilmente el rendimiento que tuvo el arquero durante el partido y el campeonato.

Esta tesis busca un primer acercamiento al estudio del rendimiento de arqueros de fútbol. Se analizan las variables que actualmente se utilizan para evaluar el desempeño de estos y se estudiarán alternativas a estas variables. El propósito final es encontrar una

metodología que permita evaluar el rendimiento de los porteros dejando afuera, en lo posible, los sesgos de resultados o cualquier otro sesgo que pueda intervenir en la evaluación. Se propone incorporar una nueva variable para medir el rendimiento del arquero, en particular proponemos incorporar la dificultad de los tiros a los que se enfrenta este. Dado que, el grado de dificultad varia entre disparos al arco, por ejemplo, es diferente recibir un tiro de una distancia de 40 metros a uno de 5 metros de la portería. Así se busca evaluar la habilidad de un arquero de atajar tiros de distintos puntos del campo de juego y de dificultades, para que entrenadores y dirigentes de fútbol puedan tomar decisiones objetivas, dejando de lado sus percepciones u opiniones personales.

# Capítulo 2: Marco Teórico

## 2.1. ¿Cómo se mide el rendimiento en el mercado laboral?

Esta es una clásica pregunta que se estudia en economía, como medir el esfuerzo, el talento y el rendimiento de los trabajadores. Podríamos definir el rendimiento laboral como el conjunto acciones y señales realizadas por un trabajador las cuales impactan en el resultado de una empresa, impacto positivo se asocia como un buen rendimiento, caso contrario, mal rendimiento. El esfuerzo y talento se relacionan con estas acciones y señales, en teoría un trabajador talentoso debiera dedicar menos esfuerzo a las acciones de impacto positivo, en comparación a un trabajador menos talentoso que tiene un rendimiento similar.

Pero, ¿por qué es importante evaluar el rendimiento de un trabajador? En primer lugar, para determinar si cumple o no con los estándares de productividad esperables. Por otro lado, para comprender las áreas de mejora e implementar contramedidas para mejorar el rendimiento. Uno de los puntos más importantes en la teoría económica, relacionado con el conjunto de acciones, sus señales y el rendimiento de los trabajadores, es el principio de informatividad (Holmstrom, 1979, 1982) para la teoría de contratos (Dewatripont, 2005). El principio de informatividad sugiere que un contrato debe estar diseñado para inducir a las partes a revelar información sobre sus habilidades, esfuerzo y otras señales relevantes que pueden afectar el resultado del contrato. Al hacerlo, el contrato puede ayudar a alinear los intereses de las partes y promover la eficiencia, por lo tanto, es fundamental tener bien definidas estas señales.

Existen varios métodos para evaluar el rendimiento, pero dentro de los más destacados se encuentran los siguientes. Evaluación 360, la cual consiste en tener evaluaciones de pares, jefes y/o subordinados junto con los clientes de un trabajador. Evaluación de los incidentes críticos, el cual consiste en que cada supervisor investiga, observa y registra los hechos positivos o negativos más destacados del desempeño de cada subordinado, este método también se conoce como de acontecimientos notables (Montejo, 2001). El último a destacar es el método de administración por objetivos, el cual consiste en establecer objetivos o metas por parte del evaluador y el evaluado.

Entonces, la siguiente pregunta que surge es ¿cuáles son las acciones, señales, objetivos o metas que determinan un buen rendimiento? Por ejemplo, en algunas universidades se mide el rendimiento de los profesores según la cantidad de publicaciones que estos realizan en un periodo de tiempo establecido, o en el fútbol se premia al goleador del torneo según goles convertidos en la temporada. Cada profesión tiene sus propias medidas de rendimiento, algunas más desarrolladas que otras. Pero no todas las industrias pueden evaluar tan directamente el rendimiento, de aquí la importancia de que las señales sean bien identificadas.

Una señal es valiosa, si es informativa del desempeño de un trabajador y todas deberían ser recompensadas (Dewatripont, 2005). Lamentablemente en las organizaciones, las señales de desempeño precisas y objetivas generalmente no están disponible. Los tomadores de decisiones tienen que confiar en las evaluaciones subjetivas del desempeño, las cuales desafortunadamente son propensas a sesgos. Los incentivos de los evaluadores pueden limitar la precisión de la evaluación subjetiva del desempeño en las organizaciones, perjudicando el objetivo principal de evaluar el desempeño, identificar si se están haciendo bien las cosas y donde existen puntos de mejora (Prendergast, 1993, 1996; Bol, 2011).

Realizar evaluaciones objetivas del desempeño de un trabajador no es trivial, el éxito o fracaso de un trabajador a menudo se debe a sus acciones y en parte a un factor de suerte. Como, por ejemplo, recibir un gol al último minuto de partido luego de haber tenido un muy buen rendimiento. En tales situaciones, el evaluador tiene la difícil tarea de evaluar el desempeño de los agentes en función de sus méritos previos, al evento de suerte que puede haber influido en su resultado. Esta dificultad se incrementa con la presencia del sesgo de resultado, donde el evaluador basa su decisión en el resultado final y no en las acciones que llevaron a este resultado (Baron, 1988).

El gran uso de datos que actualmente se ve en las empresas son una gran fuente de información y una oportunidad para incorporar analítica a los procesos de evaluación e ir dejando afuera de la evaluación los juicios subjetivos. Hoy en día el número de publicaciones académicas ligadas al estudio de la relación entre la analítica de grandes datos con el desempeño en las organizaciones aumenta rápidamente. Las aplicaciones de estos estudios ya están siendo incorporadas, agregando valor en distintos rubros, como en la gestión de clientes, información, innovación, tecnología, cadenas de suministro y en el análisis deportivo (Battistič, 2019).

## 2.2. ¿Cómo se mide el rendimiento en los deportes?

Medir el desempeño de los jugadores es clave para cualquier equipo y también está siendo ampliamente utilizada para la estimación del salario de los deportistas. La teoría detrás de los modelos de pago y desempeño tiene dos componentes, cuando no hay señales claras de este último. En primer lugar, los deportistas con escasa habilidad forman emparejamientos laborales con equipos bajo información imperfecta (Farber, 1999). La segunda línea teórica muestra cómo los deportistas, a través de agentes, y los equipos negocian los resultados salariales (Simmons, 2022). De aquí la importancia de identificar las señales que permitan evaluar de buena forma el rendimiento de los deportistas.

En los deportes existen diversas experiencias midiendo el rendimiento, el caso icónico del análisis deportivo es el libro *Moneyball: The art of winning an unfair game* (Lewis, 2004). Donde se da a conocer la historia del equipo de béisbol de Estados Unidos, Oakland Athletics y de su gerente general Billy Bean, junto al economista de la Universidad de Yale Paul DePodesta, quienes utilizaron una revolucionaria estrategia analítica para reforzar el equipo en la temporada 2002. Hasta ese momento los fichajes eran supervisados y decididos por ojeadores expertos, quienes al criterio de los clubes eran los que más sabían de béisbol y eran capaces de seleccionar jugadores bajo la intuición y olfato de estos “expertos”. Bean y DePodesta incorporaron estadística y métricas nuevas, entre las que destacan, por ejemplo, para los bateadores el porcentaje de embasamiento o para los lanzadores, priorizaron brazos finos por encima de la velocidad de lanzamiento. Esta nueva estrategia y filtros analíticos les permitió reforzar el equipo y poder competir con las grandes franquicias de la liga estadounidense de béisbol (MLB), logrando 103 partidos ganados y una racha histórica de 20 victorias consecutivas. Esta estrategia no solo revolucionó el mundo del béisbol, muchos deportes adoptaron la iniciativa y han ido incorporado la analítica, las estadísticas y las decisiones guiadas por datos a sus procesos internos.

Uno de los métodos más antiguos en la analítica deportiva es la Evaluación del rendimiento en deportes de equipo o TSAP por sus siglas en inglés (Team Sport Assessment Procedure). Este método busca evaluar el rendimiento individual de los jugadores con respecto a la contribución en el equipo. Evalúa las intervenciones que tuvo el jugador, evaluando la decisión que este tuvo en cada ocasión que participó y la ejecución de esta (Grehaigne, 1997). Por ejemplo, en un deporte de balón donde se tenga que “atacar” al contrario (fútbol, básquetbol, hockey, etc.) se puede evaluar que hizo el jugador con el balón

y como fue la ejecución de esta decisión. Si el jugador decide lanzar al arco o al aro, y este lanzamiento es efectivo, se cuenta como una decisión y ejecución correcta, finalmente luego de caracterizar todas las intervenciones del jugador, se calculan los porcentajes de decisiones correctas, ejecuciones correctas e intervenciones totales con respecto al equipo, para luego ser utilizadas como métricas de rendimiento.

A pesar de ser un método antiguo sigue teniendo vigencia en el mundo de la analítica deportiva básica y pedagógica. Un ejemplo es el caso de estudio realizado por Catarino (2017) para el caso del básquetbol, deporte en el cual no existe un solo instrumento válido y fiable para evaluar de forma objetiva el rendimiento de los jugadores y su conocimiento técnico. Se estudiaron dos jugadores profesionales mediante vídeo análisis utilizando el TSAP por tres observadores, con la finalidad de validar este método. Se obtuvo como resultado que los observadores tuvieron un acuerdo casi perfecto en el registro e interpretación de las acciones de los jugadores, se obtuvo un coeficiente  $k$  de Cohen de 0,88; métrica estadística que ajusta el efecto del azar en la proporción de la concordancia observada. Resultado que implica que el TSAP es una herramienta válida y fiable para evaluar elementos ofensivos dentro un partido de básquetbol. Si bien este método ofrece una estandarización en el juicio de los observadores, lamentablemente los deportes no solo se componen por fases ofensivas y defensivas. Hay un contexto y elementos de suerte en cada acción realizada por el deportista, que puede afectar en la decisión y en la ejecución de esta. Como se mencionó anteriormente la ciencia de datos ha tenido un gran avance en temáticas como esta, análisis deportivo, demostrando incluso que modelos estadísticos simples pueden coincidir con los más complejos en términos de predicción (Szymanski, 2022).

Dentro de la experiencia de otros deportes, se tienen varios artículos que modelan los salarios de la liga estadounidense de básquetbol (NBA) en los cuales se han utilizado victorias producidas como una medida superior de la productividad del jugador sobre la propia medida de eficiencia de la NBA (Simmons, 2011). En la liga estadounidense de fútbol (NFL), las yardas ganadas parecen ser una buena métrica para los jugadores (Berri, 2009). Esto demuestra la importancia de capturar las buenas señales para poder caracterizar y calcular el rendimiento de los deportistas a lo largo de los deportes.

## 2.3. ¿Cómo se mide el rendimiento en el Fútbol?

El fútbol es uno de estos deportes en donde el equipo interactúa entre sí y el contexto del juego afecta estas interacciones entre jugadores. Con lo cual nace la siguiente pregunta, ¿Cómo se puede evaluar el rendimiento de los jugadores? El fútbol carece de una variable que pueda ser atribuida completamente al rendimiento de los jugadores. Existen estudios que utilizan variables cualitativas y cuantitativas para medir rendimiento de jugadores, desde jugadores sub 12 hasta futbolistas profesionales (Tom, 2021; Di Salvo, 2007).

Un ejemplo de estos estudios es el de Hughes (2012), los cuales realizaron una serie de mesas de trabajo (focus group) con expertos del fútbol, para definir los indicadores clave (KPI) en 5 diferentes categorías (fisiológicas, tácticas, técnicas defensivas, técnicas ofensivas y psicológicas) para medir el rendimiento de los futbolistas en 7 posiciones del campo de juego (arquero, defensa central, defensa lateral, mediocampista central, mediocampista ofensivo, mediocampista exterior y delantero). En general, en las categorías técnicas se mencionaban características similares, pero en distinto orden de importancia, por ejemplo, la precisión en los pases en los mediocampistas era fundamental, pero en los defensas centrales no era primordial. Otro ejemplo de estudios relacionados a las variables de interés por posición dentro de la cancha es el de Tom L. (2021), quienes entrevistaron a 125 ojeadores de jóvenes talentos de distintos clubes, los cuales indicaron una serie de variables en las que se fijan al momento de evaluar un joven talento. El estudio se enfocó en evaluar si realmente los ojeadores se fijan en todas las variables que definieron y si eran evaluadas de forma consistente a lo largo de las observaciones de jugadores. Finalmente, el estudio concluyó que los ojeadores no eran consistentes en la evaluación de la totalidad de variables seleccionadas y tampoco en la importancia de estas al momento de evaluar distintos jugadores. Es por esto que se abre una oportunidad para definir una variable de rendimiento que pueda ser utilizada para discernir entre jugadores y sus actuaciones.

Existe también el caso de PlayeRank (Pappalardo, 2019), que utilizando estadísticas por partido de los jugadores, asignó una clasificación (ranking) al rendimiento de los futbolistas. Utilizaron datos de eventos (nivel 3), contaban con la información de cada evento del partido, sabían si una falta terminaba en gol del equipo rival o si el pase del medio campista terminó en gol. Utilizando como tesis que el rendimiento de los equipos era el resultado del rendimiento individual de los jugadores de este. Así asignaron un puntaje a cada jugada realizada por el futbolista durante el partido, según las implicancias de estas en el encuentro.



El modelamiento fue mediante redes neuronales, el cual permitió validar la intuición, como es el caso del efecto negativo sobre el equipo cuando un jugador le sacan tarjeta roja y debe abandonar el campo. El modelo le entregó un ponderador negativo a aquellos jugadores que recibieron tarjetas rojas, lo cual es intuitivo ya que dejan al equipo con un jugador menos.

El estudio tuvo un resultado importante dado que los datos de eventos, describen cada acción que tuvo un jugador en la cancha y, en consecuencia, cuál fue el aporte de cada jugada, y así aislar el rendimiento de los jugadores. Pero esto a su vez es una gran limitante; los datos necesarios son muy específicos y profundos, dificultando la aplicación de este modelo para otras ligas, como por ejemplo la chilena o argentina. Los datos menos desagregados dificultan un estudio similar, dado que ya no se tiene acción por acción, sino que se tiene, por ejemplo, el total de los pases correctos durante un partido, pero no se sabe cuál de los pases terminó en gol, o cual de los balones perdidos terminó en gol en contra.

## 2.4. ¿Cómo se mide el rendimiento de los Arqueros?

En un intento de aislar el rendimiento de jugadores, se puede tomar la posición de arquero debido a que, si bien, tiene relación con el rendimiento del equipo, al momento de enfrentarse a un tiro al arco, está solo contra el balón y se puede medir su respuesta.

En la literatura existen análisis a porteros, centrados en su rendimiento dentro de la cancha, tanto físico como técnicos. Del análisis físico, destacan variables como kilómetros recorridos, la intensidad de los movimientos, la duración y distancia de pique, etc. Por ejemplo, White (2018), indica que al observar a través de un sistema de seguimiento multi-cámara, los porteros de la Premier League inglesa (EPL) registraron 5,611 metros por partido, mientras que los porteros internacionales en otros lugares promediaron 4,183 metros por partido y obtuvieron índices de 46 metros por minuto. Adicionalmente, un portero profesional holandés el día del partido (incluido el calentamiento) acumuló 5,985 metros.

Con respecto a los análisis de rendimiento, los estudios plantean varias variables de interés, pero las más destacadas son los Goles concedidos, las Atajadas o Salvadas, y el porcentaje de estos. El primer criterio, *Goles concedidos*, contabilizan los goles en contra por minuto jugado en la temporada, por construcción este criterio cuanto menor es la medida, mejor es el resultado en comparación con otros porteros (Clemente, 2018; Longo, 2019; Sainz

de Baranda, 2019; Yam, 2019). Una segunda métrica utilizada son las *Salvadas*, calculadas como los tiros al arco menos los goles recibidos. La tercera métrica es *Porcentaje de Salvadas* la cual representa el porcentaje de goles evitados por el portero obtenido al dividir la diferencia entre tiros a puerta y goles marcados entre el total de tiros a puerta. Este criterio tiene un impacto positivo en el resultado porque cuanto mayor sea el índice, mejor para el portero (Dicks, 2017; Liu, 2015; Montesan, 2016; Sainz de Baranda, 2019, 2008; White, 2018). Estas variables actualmente el mercado les entrega mucha importancia y utiliza principalmente como métrica de rendimiento de los arqueros, siendo la principal, los goles que recibió durante la temporada, en menor medida se utilizan tiros atajados y cantidad de partidos con la portería invicta.

Gauriot (2019) utiliza los tiros al arco que terminan golpeando en palo como un cuasi experimento, estudiando las consecuencias de estas acciones en los jugadores. Obteniendo que el cuerpo técnico premia a los delanteros que marcan un gol luego que su tiro haya golpeado el poste del arco. Los delanteros que marcaron el gol (palo gol), jugaron significativamente más en los partidos posteriores que los delanteros que no marcaron (palo fuera), demostrando que los clubes y los entrenadores premian la suerte de los delanteros. Este estudio deja entrever que los goles y en particular, los eventos aleatorios en el fútbol son premiados, se destaca también que el fútbol es un deporte orientado al gol, por lo que los delanteros son los más famosos y los arqueros que menos goles reciben, se pueden considerar como de los mejores. Por lo tanto, podemos denotar los goles como una señal que tiene un componente de suerte y a su vez, es afectada por el sesgo de resultado dada la naturaleza del fútbol. Por lo tanto, existe un espacio para estudiar y proponer una variable que incorpore el rendimiento de los arqueros distinto a los goles que recibió, dado que estos tienen un componente de suerte, y no es deseable premiar o castigar a un arquero dado un evento aleatorio.

Más allá de la suerte asociada a los goles, Berri (2023) estudia cómo se relaciona el rendimiento de los arqueros y el salario que reciben. Utiliza variables como edad, las ligas en las que juegan, presencia en campeonatos internacionales, tiros recibidos, atajadas, goles recibidos, entre otras variables. Concluyendo que los arqueros que más goles coinciden, reciben una penalización en su sueldo, sugiriendo que los equipos recompensan y penalizan a los porteros según los resultados en lugar de la carga de trabajo. También se obtuvo un resultado contra intuitivo, no se logra probar que los arqueros con más atajadas tienen un mejor sueldo. Es más, para una determinada proporción de atajadas, el aumento de la dificultad del tiro se penaliza con un salario más bajo.

Este último punto se relaciona con el estudio realizado por Gelade (2014), donde observa que los arqueros de equipos más débiles, tienden a enfrentar tiros al arco más difíciles de atajar. Estos equipos a su vez son más pequeños con menor poder económico, lo cual podría explicar el menor sueldo percibido por los arqueros. No solo existen diferencias en los tiros que reciben los equipos más pequeños, Liu (2015) realizó un estudio sobre cómo cambiaban ciertas variables técnicas de los arqueros según el nivel de los equipos donde juega y el de su rival. Encontrando que los porteros de equipos de alto nivel hicieron más atajadas cuando se enfrentaron a un oponente de bajo nivel que cuando se enfrentaron a equipos intermedios u otros de alto nivel. Estas diferencias en las dificultades que enfrentan los arqueros a lo largo del campeonato se pueden atribuir a los distintos clubes, sus plantillas y jerarquías.

Dado que existen diferencias en la cantidad de tiros al arco y la dificultad de estos, en la literatura se utiliza ampliamente la esperanza de gol ( $xG$ ), para normalizar estas diferencias. De Rathke (2016) se tiene que la esperanza de gol toma los tiros y los pondera por la probabilidad de ser gol para cada una de las situaciones generadas, es un modelo que se calibra a nivel mundial, calculando la probabilidad de gol por cada posición dentro de la cancha, como goles partido en tiros desde esa posición. A modo de ejemplo, a un disparo desde mitad de cancha le asignará un  $xG$  de 0,05 y a uno desde el área chica 0,7; es decir en un 70% de las ocasiones generadas idénticas a esa, termina en gol. Esta normalización nos permite normalizar la variable tiro al arco incorporando la dificultad de estos.

# Capítulo 3: Midiendo el rendimiento

Los criterios presentados en la literatura pueden sobre estimar el rendimiento de los arqueros. Por ejemplo, si se miran los goles en contra, es distinto recibir un gol de mitad de cancha, en comparación a recibir uno de penal. Por lo tanto, un arquero que recibe goles de penal o desde el área chica, tendrá el mismo rendimiento que un arquero que solo recibió un gol de media cancha, los cuales en teoría debieran ser más fáciles de atajar, por lo cual el rendimiento de este último estaría siendo sobre estimando. A su vez, Salvadas por construcción puede sobre estimar el rendimiento de los arqueros, es distinto atajar uno de dos tiros al arco que tapar uno de tres tiros a portería, a pesar de que en ambos casos las salvadas son las mismas,  $Salvadas_{Caso 1} = 2 - 1 = 1$  vis a vis  $Salvadas_{Caso 2} = 3 - 2 = 1$ . La tercera métrica, Portejates de Salvadas, viene a hacerse cargo de lo recién mencionado. Pero también puede representar dificultades al momento de discernir si uno tuvo un rendimiento más alto que otro, dado que si un arquero de 4 tiros al arco, le marcan 2 goles, su métrica sería  $\frac{(4-2)}{4} = 0,5$ . Pero otro arquero que de 2 tiros al arco solo atajó 1, su métrica es  $\frac{(2-1)}{2} = 0,5$ ; mismo resultado que el primer arquero, a pesar de tener menos salvadas o balones atajados. Cabe destacar que por construcción si un partido no hubo tiros al arco, esta métrica se indetermina. Lo cual es consistente con la teoría, debido a que, si un arquero no fue puesto a prueba con un tiro al arco, no hay forma de identificar y evaluar su rendimiento.

Al igual que goles concedidos, tanto Salvadas como Porcentaje de Salvadas pueden sobre estimar el rendimiento del arquero si no se toma en cuenta la dificultad de los tiros a los cuales se enfrentan los arqueros. Por lo tanto, se controlan por esperanza de gol (xG) con la finalidad de incorporar la dificultad de los tiros a estas métricas.

Por ejemplo, un arquero que recibe dos tiros desde el área chica y solo atajó uno, sus métricas Salvadas y Porcentaje de Salvadas son  $2 - 1 = 1$  y  $\frac{(2-1)}{2} = 0,5$  respectivamente. Controladas por xG estas métricas cambian a  $1,4 - 1 = 0,4$  y  $\frac{(1,4-1)}{1,4} = 0,29$  respectivamente, dado que los tiros desde el área chica tienen una esperanza de gol de 0,7. Este resultado se puede interpretar como que el arquero del 1,4 goles que le deberían haber marcado, salvó 0,4 goles, traduciéndose en que salvó casi el 30% de las ocasiones de gol. Ahora supongamos que el mismo arquero recibió dos tiros desde la media cancha y le marcaron un gol, entonces sus métricas Salvadas y Porcentaje de Salvadas son  $2 - 1 = 1$  y  $\frac{(2-1)}{2} = 0,5$  respectivamente,

mismas que en el caso anterior. Controlando por xG y asumiendo que el xG de los tiros de mitad de cancha son 0,05, sus métricas cambiarían a  $0,1 - 1 = -0,9$  y  $\frac{(0,1-1)}{0,1} = -9$  respectivamente. Estos resultados nos indican que al arquero recibió 0,9 goles más de los que debería haber recibido, es decir, le convirtieron un gol con un tiro con muy baja probabilidad de ser gol. El resultado de Porcentaje de Salvadas cuando el rendimiento del arquero es negativo, ie, al arquero le marcaron más goles de los que deberían, puede generar confusiones debido a que el porcentaje queda mayor a 100% y negativo. Por lo que se utilizará para el estudio  $\frac{(xG - Goles)}{Tiros\ al\ arco}$  el cual siempre mantendrá los parámetros de un intervalo de [-1, 1].

Controlando por xG, se logra una continuidad en las métricas que permite comparar fácilmente entre rendimientos de arqueros, dado que un arquero que obtuvo salvadas esperadas igual a 0, significa que le marcaron la misma cantidad de goles que la dificultad de los tiros sugiere. Lo mismo para rendimientos positivos y negativos, donde se traduce que el arquero atajó más tiros de lo que el xG sugiere, o le convirtieron más goles de los que deberían haberle hecho, respectivamente. Finalmente, se utilizarán como variables a estudiar las Salvadas esperadas y las mismas Salvadas esperadas, pero serán controladas por la cantidad de tiros al arco recibidos. Definidos de la siguiente forma;

$$Salvadas\ esperadas = (xG - Goles) \text{ y } \frac{Salvadas\ esperadas}{Tiros\ al\ arco} = \frac{(xG - Goles)}{Tiros\ al\ arco}.$$

# Capítulo 4: Datos

Los datos utilizados son provistos por InStat, empresa dedicada a la captura y almacenamiento de datos de fútbol. Se cuenta también con la base de datos de Transfermarkt de los valores de mercado a finales del 2021 de los arqueros estudiados. Por último, se utilizan también los datos de Sports Reference, con las plantillas de los equipos en 2021.

## 4.1. InStat

InStat es una empresa líder mundial en análisis de rendimiento deportivo para ligas, clubes, jugadores y medios profesionales de fútbol, fútbol sala, básquetbol y hockey sobre hielo. Entre sus clientes se encuentran más de 1900 clubes deportivos, incluidos los mejores equipos del mundo como Real Madrid, Manchester City, Bayern Munich, entre otros. InStat ([www.instatsport.com](http://www.instatsport.com)) es una plataforma de pago que permite descargar los datos capturados durante los partidos de fútbol de las distintas ligas. Por ejemplo, los datos capturados durante partidos de la Champions League, están disponibles 4 horas después de la finalización del partido, en comparación a un partido de la liga profesional chilena, que están disponibles en 4 a 5 días hábiles.

En este estudio se cuenta con datos de InStat provistos por Futbolytics para su realización. Se tienen datos agregados a nivel de partido y por jugador. En particular de los arqueros, se tienen variables como tiros al arco recibidos, tiros tapados, goles concedidos, entre otras. La base cuenta con 72 arqueros, con estadísticas de los partidos disputados en el torneo nacional chileno y argentino en 2021. La principal dificultad que se presentó en la base proporcionada por InStat, fue que los equipos de los arqueros estaban actualizados, es decir, los arqueros en la base aparecían con el equipo actual y no con su equipo 2021, a pesar de sí contar con las estadísticas de esa temporada. Por lo que fue necesario buscar las plantillas de los equipos en la temporada 2021, detallada más adelante.

Con la finalidad de tener suficientes partidos para evaluar el rendimiento de los arqueros se filtraron los arqueros con menos partidos jugados. De los 72 porteros presentes en la temporada 2021, se eliminan los arqueros que se encontraran en el primer cuartil de partidos

jugados, aquellos que hayan jugado menos de 7 partidos para la liga chilena y 8 partidos en la liga argentina, dejando un total de 56 arqueros a estudiar.

La Tabla 1 proporciona información de las variables relacionadas a los arqueros estudiados, como edad, altura, peso, número de partidos jugados y cantidad de puntos obtenidos por el equipo. Otras variables son por partido, número de tiros al arco, Salvadas, Porcentaje de Salvadas, Goles Concedidos, Esperanza de gol ( $xG$ ) y Salvadas esperadas ( $xG - Goles$ ). Los resultados muestran que la edad promedio de los jugadores es de 30 años y la altura promedio es de 1,86 m, mientras que el peso promedio es de 80,8 kg. Además, el promedio de Goles Concedidos por partido es de 1,23; lo que indica que en promedio los arqueros reciben más de un gol por partido. A su vez, las atajadas (Salvadas) por partido es de 9,97; lo que demuestra que a pesar de atajar en promedio casi 10 tiros al arco, los arqueros reciben más de un gol. Si controlamos por esperanza de gol de los tiros recibidos, tenemos que en promedio de  $xG$  de los tiros es de 0,1; y restando los goles, indica que en promedio los arqueros obtienen un saldo positivo en el ( $xG - Goles$ ) de 0,04. Este valor se interpreta como que en promedio, los arqueros salvan más goles de los que les deberían haber marcado dada la dificultad de los tiros. Recordar que la esperanza de gol se calcula con todo el universo de datos que tiene InStat, dado que, si se calibrara la esperanza de gol con los datos del torneo chileno y argentino 2021, la resta entre  $xG$  y Goles sería cero.

<b>Variable</b>	<b>Mean</b>	<b>St. Dev</b>	<b>Min</b>	<b>Max</b>
<i>Age</i>	30	4,96	20	42
<i>Height</i>	186	3,68	178	196
<i>Weight</i>	80,8	5,91	70	100
<i>xG-Goles</i>	0,04	0,28	-0,54	0,74
<i>xG per shot saved</i>	0,10	0,02	0,05	0,15
<i>Goles conceded</i>	1,23	0,33	0,67	2
<i>Salvadas</i>	9,97	1,53	7,41	14
<i>Salvadas %</i>	0,89	0,03	0,78	0,93
<i>Matches played</i>	20,4	7,79	7	32
<i>Points</i>	37,7	10,4	21	68
<i>Opponent's shots on target</i>	4,33	0,77	3,17	6,2
<i>Clean sheets</i>	6,39	3,47	1	14
<i>(xG-Goles)/Tiros al arco</i>	0,01	0,06	-0,12	0,16

Tabla 1: Resumen variables InStat

A continuación, se presentan las distribuciones de los goles y del precio de mercado. Los goles concedidos, graficados en la Figura 1, distribuye Poisson con  $\lambda = 1,229$ . Se realizó el test Kolmogorov-Smirnov (KS Test) para comprobar si la distribución de los goles convertidos corresponde a una distribución de Poisson de parámetro  $\lambda$  (Berger, 2014), obteniendo como resultados el estadístico  $D = 0,342$  y el p-valor  $< 2,2e-16$ . Es decir, dado que el p-valor es menor al nivel de significancia  $\alpha = 0,01$ ; no se puede rechazar la hipótesis nula de que los goles distribuyen Poisson de parámetro  $\lambda = 1,229$ ; el cual corresponde al promedio de goles por partido.

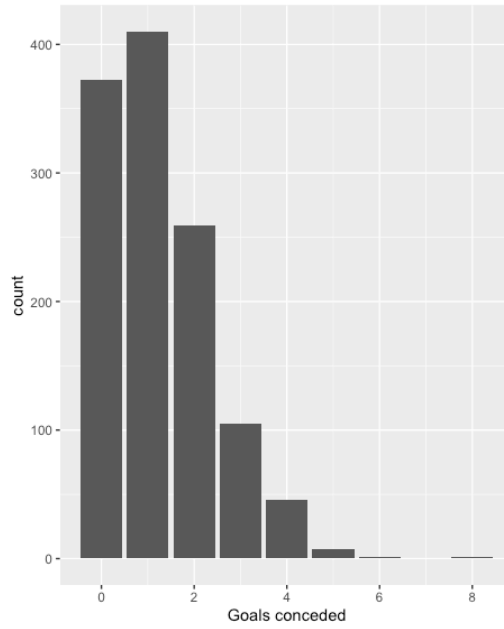


Figura 1: Distribución goles concedidos.

Teniendo la base indexada por arquero, se procede a calcular las variables de interés mencionadas en previamente para cada arquero,  $(xG - Goles)$  y  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ , las cuales distribuyen de forma unimodal, como se aprecia en las Figuras 2a y 2b, respectivamente.

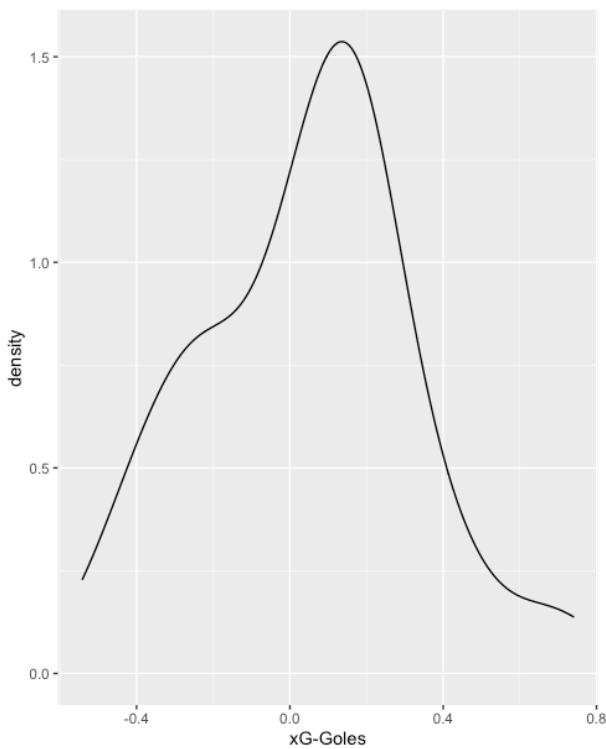


Figura 2a: Distribución Salvadas esperadas

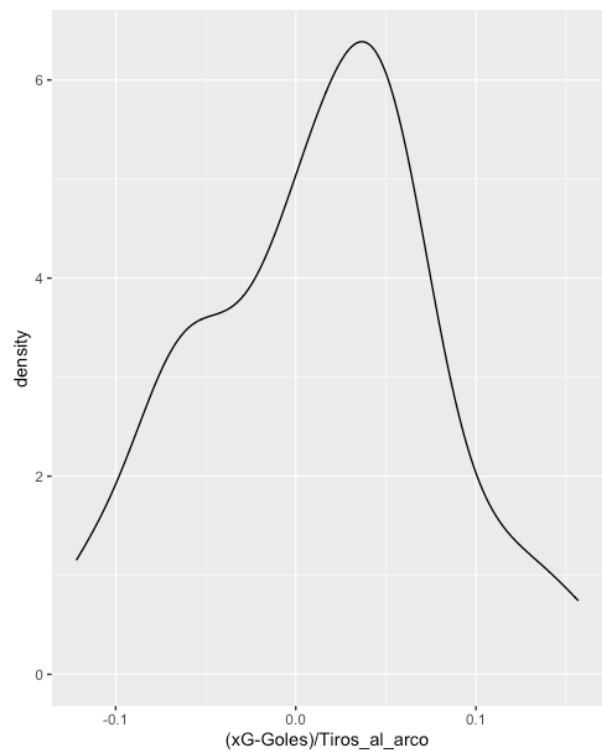


Figura 2b: Distribución Salvadas esperadas vs Tiros al arco



## 4.2. Transfermarkt

Transfermarkt ([www.transfermarkt.com](http://www.transfermarkt.com)) es la plataforma de fútbol digital líder en datos, estadísticas y como elemento diferenciador, presenta una estimación del valor de mercado de los futbolistas a nivel mundial, siendo una de las principales referencias para valorizar jugadores. Opera en varios idiomas, llegando a 30 millones de aficionados al fútbol al mes en todo el mundo. Ofrecen la base de datos de fútbol más grande del mundo con mucha la información sobre jugadores, clubes y competiciones, así como una de las comunidades de fútbol más grandes, sumado al suministro de noticias, estadísticas, videos, opiniones de fanáticos e información de agentes de jugadores. El principal aporte de esta base de datos, es identificar el valor de mercado de los arqueros estudiados, con la finalidad de estudiar la relación que existe entre el valor de mercado con el rendimiento.

La tabla 2 contiene estadísticas sobre el valor de mercado de los futbolistas en euros (€). El valor de mercado promedio es de €958.393 con una desviación estándar de €1.165.500, lo que sugiere que los valores de mercado de los jugadores varían mucho. Los valores del percentil 10 y 25 son relativamente bajos, por debajo de €137.500 y €287.500 respectivamente. La mediana del valor de mercado es de €460.000. Los valores del percentil 75 y 90 son muy altos, lo que sugiere que solo el 25% y el 10% de los jugadores tienen un valor de mercado superior a €1.000.000 y €2.750.000 respectivamente. Además, se proporciona información sobre el número de observaciones y el número total de jugadores incluidos en el análisis de 56 arqueros. En resumen, esta tabla proporciona información útil sobre la distribución de los valores de mercado de los futbolistas en euros (€) y es estándar a lo que se maneja en el mercado laboral, donde la distribución es asimétrica con jugadores muy caros y la gran mayoría no tanto.

<b>Statistics</b>	<b>Valor mercado (Euros)</b>	<b>Ln(Valor mercado)</b>
<i>Mean</i>	958.393	13,23
<i>Standard deviation</i>	1.165.500	1,029
<i>10th percentile</i>	137.500	11,83
<i>25th percentile</i>	287.500	12,57
<i>Median</i>	460.000	13,04
<i>75th percentile</i>	1.000.000	13,82
<i>90th percentile</i>	2.750.000	14,82
<i>N Observations</i>	56	56
<i>N players</i>	56	56

Tabla 2: Resumen Valor de Mercado

Con respecto al valor de mercado de los arqueros, este corresponde al de finales de la temporada 2021, con la intención de estudiar el efecto del rendimiento durante toda la temporada en el precio de mercado del portero. Para este estudio se utilizará el logaritmo del valor de mercado, de modo que la distribución de la variable sea más cercana a una de valor normal, como se aprecia en las Figuras 3a y 3b respectivamente.

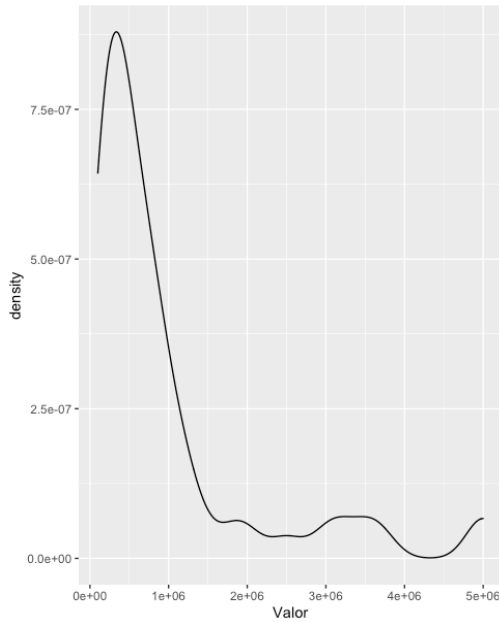


Figura 3a: Distribución Valor de Mercado

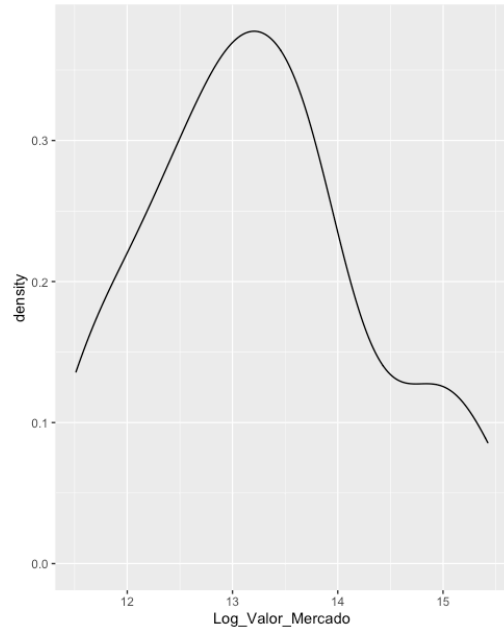


Figura 3b: Distribución Logaritmo Valor de Mercado

### 4.3. Sports Reference

Sports Reference es una empresa estadounidense que opera varios sitios web relacionados con deportes, en particular para datos de fútbol utilizan la plataforma de FBref ([www.FBref.com](http://www.FBref.com)), la cual es gratuita e incluye diversas estadísticas. También operan un servicio basado en suscripción para estadísticas, llamado Stathead. Se utiliza la información de las plantillas de los equipos de las ligas chilena y argentina 2021, para solucionar el problema de la actualización de los equipos de los arqueros. Se utilizó como identificador, el nombre del jugador, y se agregó el equipo donde jugó en la temporada 2021.

## 4.4. Correlaciones

Teniendo las bases de datos consolidados, se grafican en la Figura 4 las relaciones entre las variables de interés, donde cada punto es un arquero y la línea azul es un ajuste lineal. Se puede apreciar una relación inversa entre goles y el logaritmo del valor de mercado, es decir, los arqueros con menos goles son los más caros. Lo cual es consistente con la intuición y con la literatura.

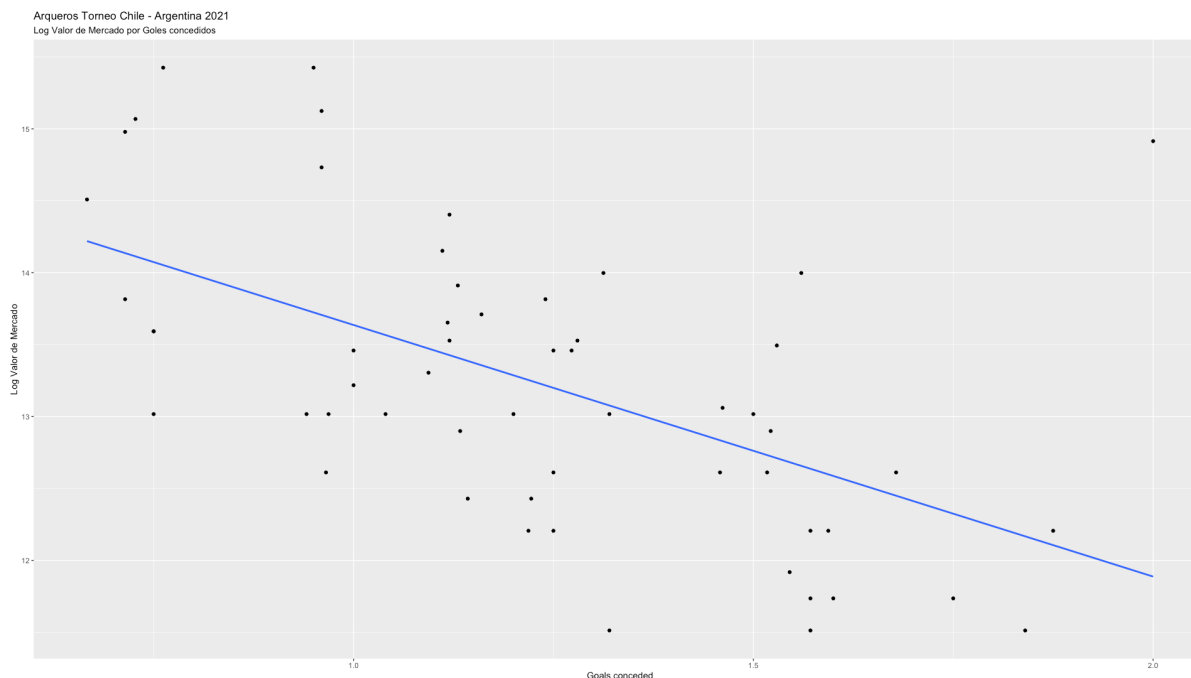


Figura 4: Correlación Logaritmo Valor de Mercado y Goles Concedidos

Se realiza el mismo gráfico anterior de Goles Concedidos contra Logaritmo de mercado, pero identificando cada uno de los arqueros, separándolos por liga y precisando la cantidad de partidos jugados, Figura 5. Los ajustes lineales no consideran los pesos asociados a la cantidad de partidos jugados. Se mantiene la relación negativa entre ambas variables para ambas ligas, chilena y argentina. Se aprecia que los arqueros de la liga argentina son más caros y tienen mayor varianza que los de la liga chilena. Bryan Cortés, arquero de Colo Colo, a pesar de ser el arquero con menos goles concedidos en promedio por partido, no es el arquero más caro, es superado por arqueros de la liga argentina con rendimiento similar.

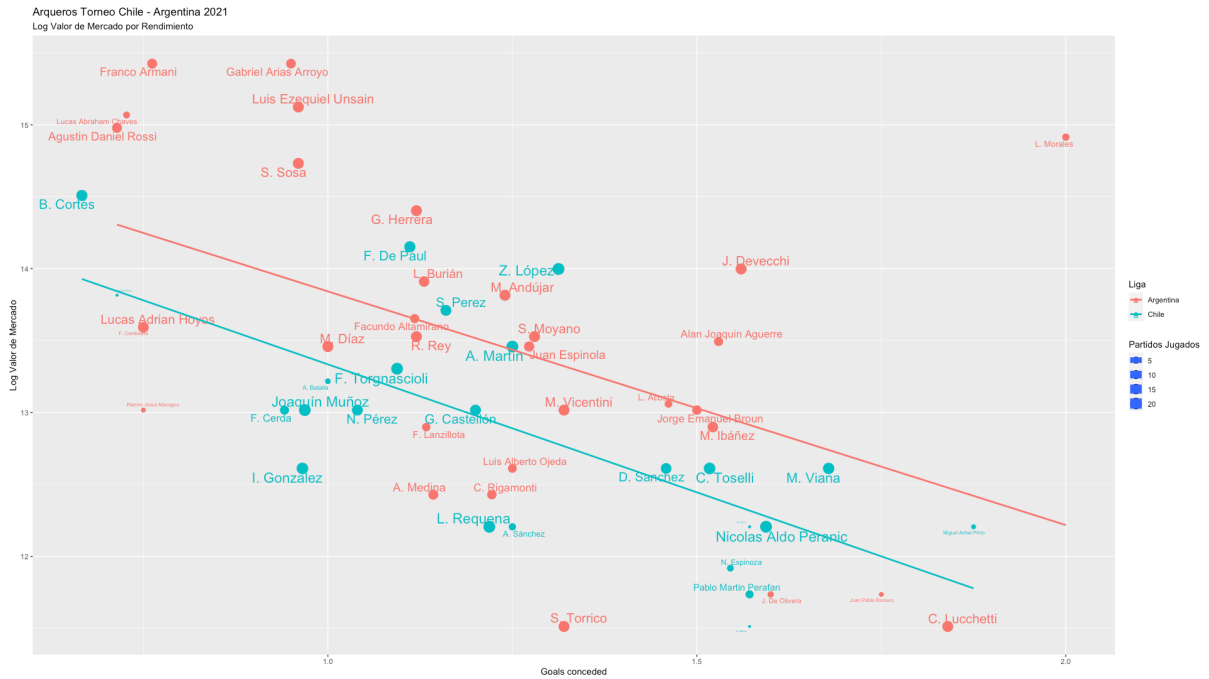


Figura 5: Correlación Logaritmo Valor de Mercado y Goles Concedidos por Arquero y Liga

A continuación, Figura 6 y 7, se grafican también las relaciones de  $(xG - \text{Goles})$  y  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  con el logaritmo del valor de mercado, al igual que el gráfico anterior, indicando cada arquero, identificando en qué liga juegan y cuántos partidos disputaron, al igual que las otras Figuras, los ajustes lineales no consideran los pesos asociados a la cantidad de partidos jugados. La intención es ver cómo se comportan con respecto a goles concedidos, la cual, como se comentó previamente, es una variable ampliamente utilizada para evaluar a los arqueros, donde la relación con el valor de mercado es inversamente proporcional. Podemos apreciar que tanto  $(xG - \text{Goles})$  y  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  tienen una relación positiva con el logaritmo del valor de mercado. Esto es intuitivo, dado que implica que arqueros con mayores salvadas, ajustadas por esperanza de gol y por tiros al arco, tienen un mayor valor de mercado.

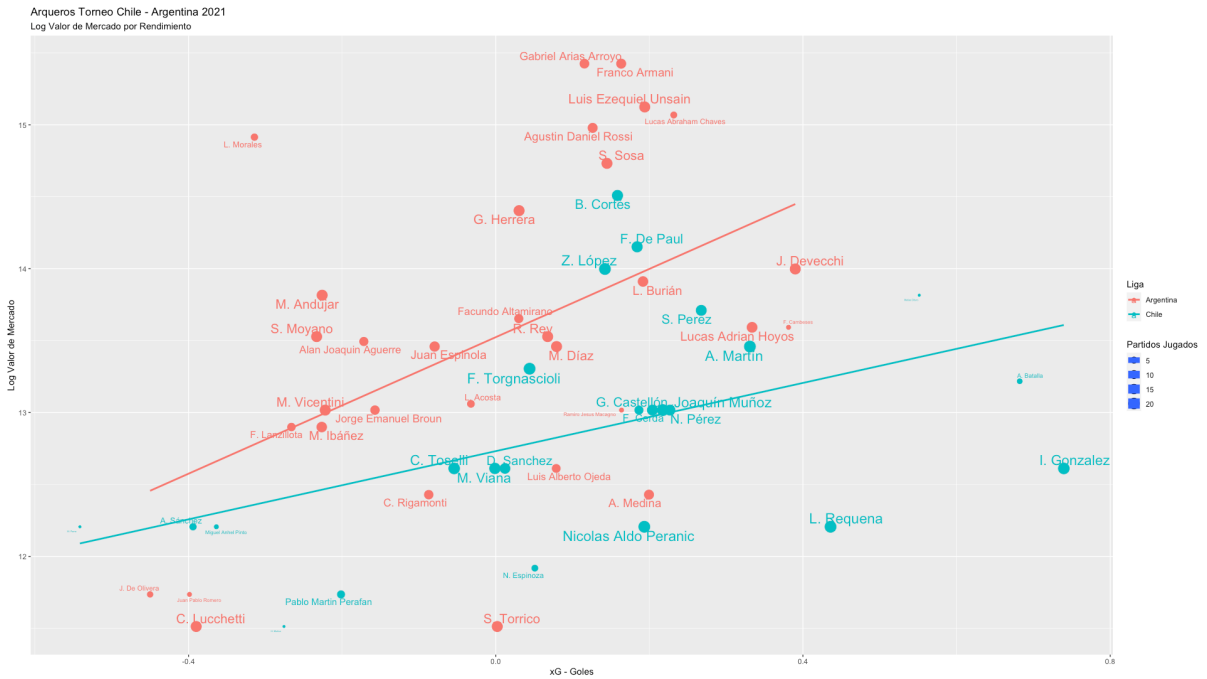


Figura 6: Correlación Logaritmo Valor de Mercado y Salvadas esperadas por Arquero y Liga

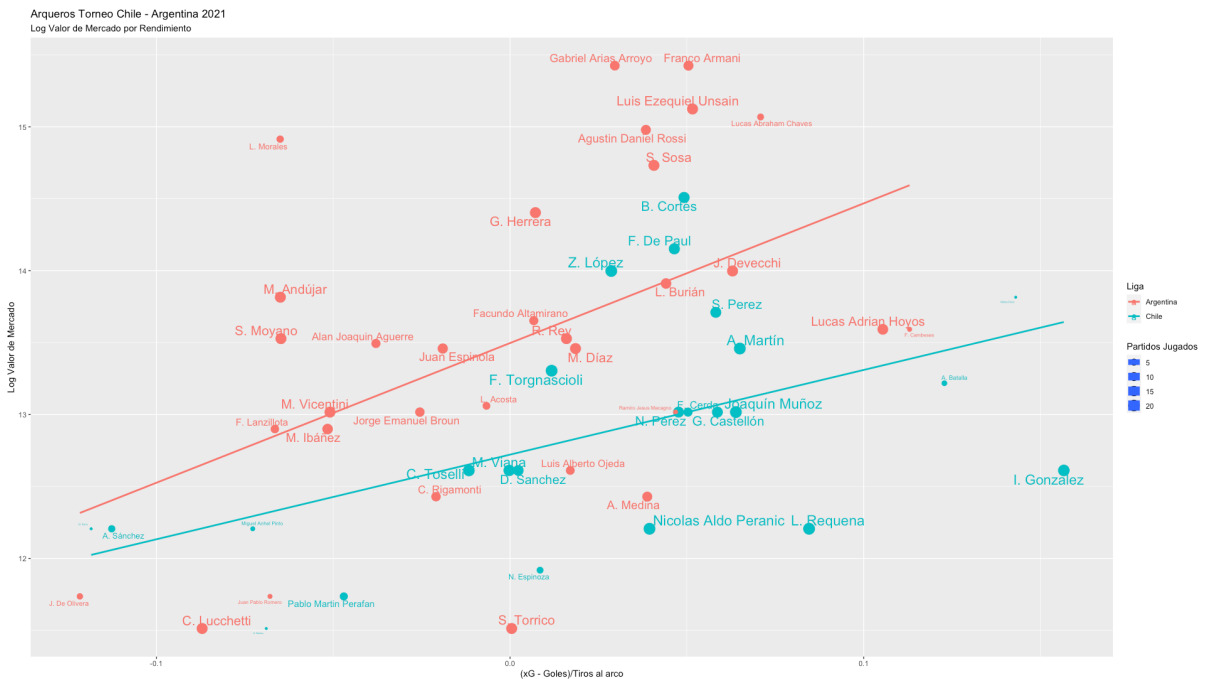


Figura 7: Correlación Logaritmo Valor de Mercado y Salvadas esperadas vs Tiros al arco por Arquero y Liga

Se aprecia como jugadores como Franco Armani, arquero de River Plate, es el de mayor valor de mercado, junto con ser de los arqueros con menos goles concedidos en promedio por partido, a su vez es de los arqueros con  $(xG - Goles)$  y  $\frac{(xG - Goles)}{Tiros\ al\ arco}$  positivo. Esto se puede interpretar como que los arqueros, y en este caso Franco Armani, que tienen  $(xG - Goles)$

y  $\frac{(xG - Goles)}{Tiros\ al\ arco}$  positivo, reciben menos goles de los que deberían haber recibido, en otras palabras, Armani salva 0,5 goles en promedio por partido.

Por último, se estudia la relación entre el valor de mercado y varianza de  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ , la cual se puede interpretar como la consistencia del desempeño partido a partido de los arqueros. En la Figura 8, se aprecia que no existe una relación directa entre el entre estas dos variables. Por ejemplo, Franco Armani, el arquero más caro de la muestra, tiene una dispersión mayor a Guido Herrera o Fernando de Paul con Bryan Cortez.

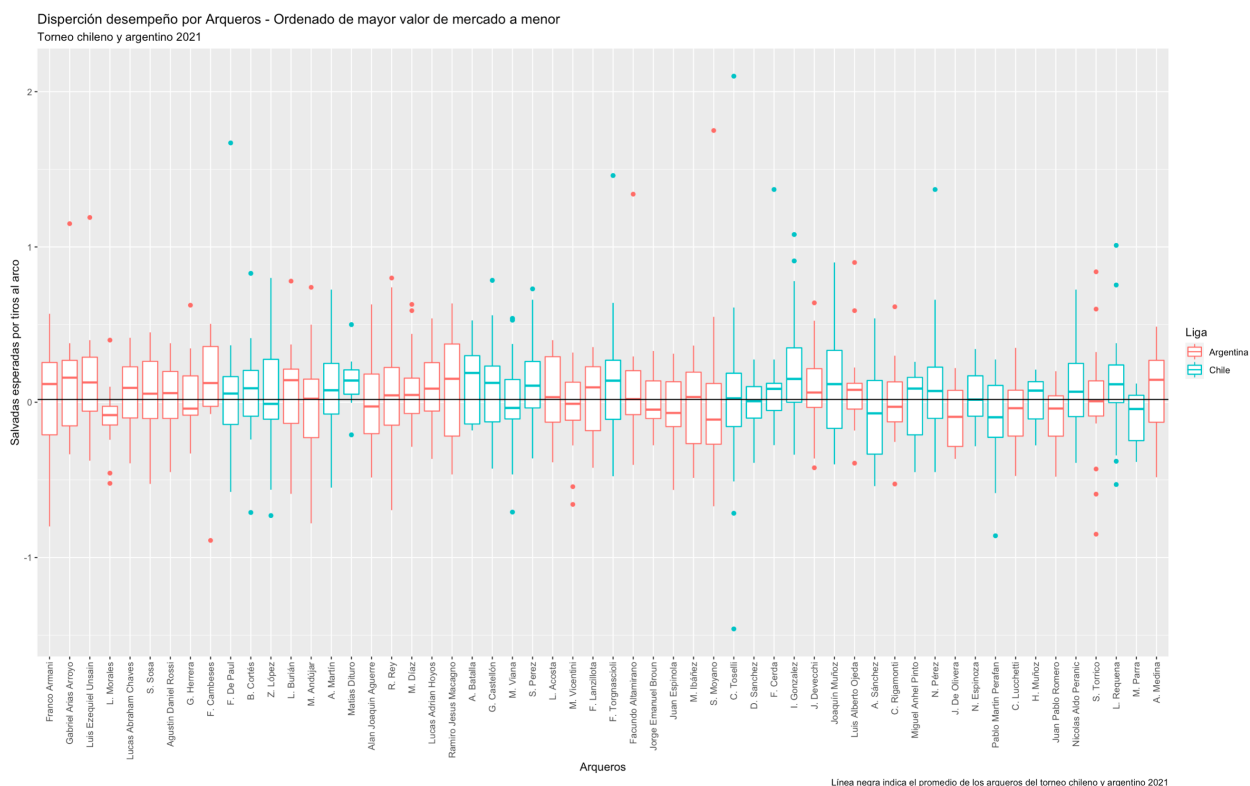


Figura 8: Correlación dispersión de rendimiento vs Valor de Mercado

## 4.5. Ranking

Ordenado los arqueros por su  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  promedio se puede visualizar el siguiente ranking. Donde se aprecia como Ignacio González es el mejor arquero del campeonato chileno/argentino 2021, a pesar de tener más goles concedidos en promedio por partido que Matías Dituero. Esto se debe a que, en promedio, González atajó tiros al arco de una mayor dificultad que los tiros atajados por Dituero.

Ranking	Arquero	(xG - Goles)/ Tiros al arco	Goles Concedidos Totales	Goles Concedidos Promedio por partido	Partidos Jugados	Ranking	Arquero	(xG - Goles)/ Tiros al arco	Goles Concedidos Totales	Goles Concedidos Promedio por partido	Partidos Jugados
1	I. Gonzalez	0,157	28	0,97	29	29	R. Rey	0,016	28	1,12	25
2	Matias Dituero	0,143	5	0,71	7	30	F. Torgnascioli	0,012	35	1,09	32
3	A. Batalla	0,123	9	1,00	9	31	N. Espinoza	0,008	17	1,55	11
4	F. Cambeses	0,113	6	0,75	8	32	G. Herrera	0,007	28	1,12	25
5	Lucas Adrian Hoyos	0,105	18	0,75	24	33	Facundo Altamirano	0,007	19	1,12	17
6	L. Requena	0,085	39	1,22	32	34	D. Sanchez	0,002	35	1,46	24
7	Lucas Abraham Chaves	0,071	8	0,73	11	35	S. Torrico	0,000	33	1,32	25
8	A. Martín	0,065	40	1,25	32	36	M. Viana	0,000	47	1,68	28
9	Joaquín Muñoz	0,064	31	0,97	32	37	L. Acosta	-0,007	19	1,46	13
10	J. Devecchi	0,063	39	1,56	25	38	C. Toselli	-0,012	44	1,52	29
11	G. Castellón	0,059	30	1,20	25	39	Juan Espinoia	-0,019	28	1,27	22
12	S. Perez	0,058	29	1,16	25	40	C. Rigamonti	-0,021	22	1,22	18
13	Luis Ezequiel Unsain	0,052	24	0,96	25	41	Jorge Emanuel Broun	-0,025	27	1,50	18
14	Franco Armani	0,050	16	0,76	21	42	Alan Joaquin Aguerre	-0,038	26	1,53	17
15	F. Cerda	0,050	16	0,94	17	43	Pablo Martin Perafan	-0,047	22	1,57	14
16	B. Cortés	0,049	18	0,67	27	44	M. Vicentini	-0,051	33	1,32	25
17	N. Pérez	0,048	26	1,04	25	45	M. Ibáñez	-0,052	35	1,52	23
18	Ramiro Jesus Macagno	0,047	6	0,75	8	46	S. Moyano	-0,065	32	1,28	25
19	F. De Paul	0,046	30	1,11	27	47	M. Andújar	-0,065	31	1,24	25
20	L. Burián	0,044	26	1,13	23	48	L. Morales	-0,065	24	2,00	12
21	S. Sosa	0,041	24	0,96	25	49	F. Lanzillota	-0,067	17	1,13	15
22	Nicolas Aldo Peranic	0,039	51	1,59	32	50	Juan Pablo Romero	-0,068	14	1,75	8
23	A. Medina	0,039	24	1,14	21	51	H. Muñoz	-0,069	11	1,57	7
24	Agustin Daniel Rossi	0,038	15	0,71	21	52	Miguel Anhel Pinto	-0,073	15	1,88	8
25	Gabriel Arias Arroyo	0,030	19	0,95	20	53	C. Lucchetti	-0,087	46	1,84	25
26	Z. López	0,029	42	1,31	32	54	A. Sánchez	-0,113	15	1,25	12
27	M. Diaz	0,019	25	1,00	25	55	M. Parra	-0,118	11	1,57	7
28	Luis Alberto Ojeda	0,017	20	1,25	16	56	J. De Olivera	-0,122	16	1,60	10

Tabla 3: Ranking Arqueros temporada 2021

# Capítulo 5: Metodología

Para evaluar si las variables de rendimiento tienen poder explicativo para el valor de mercado y en consecuencia ser una opción viable para evaluar el rendimiento de arqueros, se utilizarán regresiones lineales. La regresión lineal es una técnica estadística utilizada para modelar la relación entre dos variables y dar cuenta de la variable dependiente, en este caso, el valor de mercado. Esta técnica se basa en ajustar una línea recta a los datos observados, determinada por análisis de covarianza, para minimizar el error residual. Los parámetros de línea se estiman utilizando el método de mínimos cuadrados. Se basa en optimizar una función objetivo que mide la distancia entre el valor observado y el valor predicho por la línea. La calidad del ajuste se evalúa mediante el coeficiente de determinación, definido como la fracción de la varianza explicada por la variable independiente (en este caso, el conjunto de variables) de la variable dependiente. La regresión lineal es una técnica estadística ampliamente utilizada en diversos campos como la ciencia, la ingeniería y la economía. (Hastie, 2009).

En el fútbol hay muchos factores que pueden afectar el desempeño o el valor de un jugador, como la técnica, la condición física, la edad o la estrategia de juego. Al utilizar una regresión lineal, podemos examinar cómo estas variables se relacionan con el rendimiento del jugador, en este caso el valor de mercado, y determinar cuáles son las más importantes. Además, podemos cuantificar la fuerza de la relación entre estas variables y el valor de mercado del jugador, lo que puede ayudar a los entrenadores y analistas a tomar decisiones informadas sobre cómo mejorar el proceso de búsqueda de jugadores del equipo. En resumen, la regresión lineal es una herramienta apropiada para identificar las variables más relevantes en el valor y/o rendimiento de los jugadores de fútbol y en particular de los arqueros.

Se utilizarán las variables Edad,  $Edad^2$ , Liga y partidos jugados, como variables de control, las cuales fueron seleccionadas de la revisión bibliográfica (Berri, 2023). Las variables de desempeño son, Salvadas, goles recibidos (Goals conceded), (xG – Goles) y  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$ . Dado su alta correlación, se estiman separadamente para evaluar el valor de mercado. La ecuación que estimaremos es la siguiente.

$$\ln(\text{Valor de Mercado}_i) = \beta_0 + Y\beta_1 + X\beta_2 + \varepsilon_i$$



Donde  $Y$  es un vector con las variables de control Edad,  $Edad^2$ , Liga y partidos jugados,  $X$  es el vector con variables de rendimiento mencionadas anteriormente. También se realiza una segunda regresión, utilizando como variables de control, aquellas seleccionadas mediante el método LASSO, con la finalidad de probar la robustez de las variables de rendimiento. Este método se puede utilizar debido a que la base de datos cuenta con variables numéricas y nominales. LASSO es una técnica de regresión muy conocida para estos casos. Es capaz de realizar una selección de variables en el modelo lineal y puede tener una mayor precisión que la regresión lineal en una variedad de escenarios, dependiendo de la elección de lambda  $\lambda$ . A medida que  $\lambda$  aumenta, más coeficientes serán cero, lo que significa que se seleccionan menos variables y se emplea más reducción entre los coeficientes distintos de cero. Con un límite en la suma de los valores absolutos de los coeficientes, minimiza la suma habitual de errores al cuadrado (He, 2015). Las variables seleccionadas por el método son, Edad,  $Edad^2$ , posición en la que finalizó el campeonato el equipo del arquero, Liga, partidos jugados y Tiros al arco (Opponents' shots on target).

Cada regresión será evaluada por su coeficiente de determinación o  $R^2$ , el cual es una medida estadística que indica qué porcentaje de la variabilidad de la variable dependiente en un modelo de regresión lineal puede ser explicado por la variable independiente o variables independientes incluidas en el modelo. En otras palabras, el  $R^2$  mide qué tan bien el modelo se ajusta a los datos. El  $R^2$  ajustado toma en cuenta el número de variables independientes incluidas en el modelo y, por lo tanto, proporciona una medida más precisa de la calidad del ajuste del modelo (Kutner, 2004; Montgomery, 2021).

# Capítulo 6: Resultados

A continuación, se presentan los resultados obtenidos luego de realizar las regresiones lineales. La Tabla 4 muestra los resultados, utilizando las variables extraídas de la regresión con variables seleccionadas del estudio bibliográfico.

Tabla 4: Estimación valor de mercado arqueros. Variables revisión bibliográfica.

	Dependent variable:				
	Log_Valor_Mercado				
	(1)	(2)	(3)	(4)	(5)
Edad	0.350 (0.243)	0.024 (0.236)	0.197 (0.248)	0.186 (0.244)	0.023 (0.238)
Edad_2	-0.007* (0.004)	-0.002 (0.004)	-0.004 (0.004)	-0.004 (0.004)	-0.002 (0.004)
LigaChile	-0.854*** (0.232)	-0.676*** (0.215)	-0.905*** (0.231)	-0.892*** (0.227)	-0.696*** (0.227)
`Matches played`	0.043*** (0.016)	0.033** (0.014)	0.027* (0.016)	0.027* (0.016)	0.032** (0.015)
Salvadas	-0.153** (0.076)				
`Goals conceded`		-1.357*** (0.350)			-1.260** (0.476)
`xG-Goles`			1.076** (0.463)		
`(xG-Goles)/Tiros_al_arco`				5.136** (1.943)	0.753 (2.472)
Constant	10.561*** (3.603)	15.227*** (3.586)	11.231*** (3.604)	11.338*** (3.539)	15.119*** (3.636)
Observations	56	56	56	56	56
R <sup>2</sup>	0.402	0.503	0.417	0.433	0.504
Adjusted R <sup>2</sup>	0.342	0.453	0.358	0.376	0.443
Residual Std. Error	0.835 (df = 50)	0.761 (df = 50)	0.825 (df = 50)	0.813 (df = 50)	0.768 (df = 49)
F Statistic	6.722*** (df = 5; 50)	10.115*** (df = 5; 50)	7.144*** (df = 5; 50)	7.632*** (df = 5; 50)	8.292*** (df = 6; 49)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Con respecto a las variables de rendimiento de los arqueros. Como se mencionó en la metodología, estas van entrando de forma secuencial desde la columna 1 hasta la 4 y en la columna 5 entran Goles Concedidos y  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ . En la primera columna podemos apreciar como los Salvadas es significativo y tiene signo negativo, a mayor cantidad de tiros atajados, menor es el precio de mercado. Se interpreta como por cada ataja más que tiene un arquero,

su valor de mercado varía negativamente en un 15,3%. Esta variable puede tener incluido el efecto del equipo, dado que equipos de menor categoría reciben más tiros al arco.

En la columna 2, se tiene Goles Concedidos, promedio por partido, de los arqueros durante el campeonato, este es significativo y con signo negativo. El coeficiente asociado se interpreta como el efecto de un gol más recibido por el arquero, afecta negativamente su valor de mercado en 136%. Esto, valida lo expuesto en la sección de correlaciones, donde se presenta el gráfico de la relación entre el logaritmo del precio de mercado y el promedio de Goles Concedidos por partido. Esta relación se podría explicar debido a que el precio de mercado de los arqueros, tiene por construcción muy presente los goles recibidos por los arqueros, entonces el precio de mercado es fijado en función de los goles recibidos en una temporada, y de otras variables.

Columna 3, La variable de rendimiento ( $xG - Goles$ ) tiene un efecto positivo y significativo en la explicación del valor de mercado de los arqueros, lo que es intuitivo, dado que a mayor ( $xG - Goles$ ) mayor es la dificultad y cantidad de los tiros atajados, tienen un efecto positivo en el precio. Columna 4,  $\frac{(xG - Goles)}{Tiros\ al\ arco}$  tiene signo positivo y significativo. A mejores atajadas, mayor es el precio de mercado, lo cual es intuitivo y es lo esperable en un arquero. Los coeficientes asociados a estas variables influyen de forma positiva en el valor de mercado, en donde una unidad más de estas variables, tienen un efecto de 108% y 514% respectivamente.

En la columna 5, se hacen competir Goles Concedidos con  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ ; pero en esta ocasión solo goles concedidos es significativo, al igual que en la columna 2, es negativo y un gol más tiene un efecto sobre el valor de mercado de una disminución de 126%. Esto puede ser debido a que el mercado ve los goles recibidos por los arqueros y ajusta el precio.

Las variables Edad y  $Edad^2$  (Edad\_2) se comportan de forma esperable, positiva y negativa respectivamente, pero no son significativas. Los arqueros que juegan en la liga chilena, tienen un valor de mercado significativamente menor a los de la liga argentina y el efecto de jugar más partidos es positivo y significativo. Los arqueros que juegan más partidos (Matches played) son significativamente más caros que los arqueros con menos partidos.

Las regresiones presentadas en las columnas 2 y 5 son las con un mayor  $R^2$ , cual indica que estas son las con mayor poder explicativo sobre el logaritmo de valor de mercado. En ambas regresiones están presente los Goles Concedidos promedio por partido de los arqueros, por lo cual, se puede extraer que esta explica de mejor manera el logaritmo del valor de

mercado. Evaluando el  $R^2$  ajustado, la regresión de la columna 2 es la con mejor ajuste de las 5 expuestas.

A continuación, se presentan en la Tabla 5 los resultados de las regresiones con las variables de control seleccionadas por el método LASSO.

Tabla 5: Estimación valor de mercado arqueros. Variables Lasso.

	Dependent variable:				
	Log_valor_Mercado				
	(1)	(2)	(3)	(4)	(5)
Edad	0.071 (0.203)	-0.018 (0.202)	-0.010 (0.196)	0.002 (0.195)	-0.007 (0.202)
Edad_2	-0.003 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)
LigaChile	-1.016*** (0.207)	-0.975*** (0.197)	-1.063*** (0.192)	-1.051*** (0.192)	-1.039*** (0.204)
`Matches played`	0.028** (0.013)	0.028** (0.012)	0.023* (0.013)	0.023* (0.012)	0.024* (0.013)
`Opponents' shots on target`	-0.374* (0.187)	-0.172 (0.147)	-0.333*** (0.118)	-0.314** (0.118)	-0.288 (0.179)
Pos.	-0.072*** (0.015)	-0.067*** (0.015)	-0.067*** (0.015)	-0.066*** (0.015)	-0.066*** (0.015)
Salvadas	0.035 (0.093)				
`Goals conceded`		-0.668* (0.387)			-0.119 (0.617)
`xG-Goles`			0.773** (0.365)		
`(xG-Goles)/Tiros_al_arco`				3.271** (1.569)	2.889 (2.535)
Constant	15.400*** (3.047)	16.807*** (3.076)	16.668*** (2.979)	16.379*** (2.958)	16.521*** (3.077)
Observations	56	56	56	56	56
R <sup>2</sup>	0.633	0.654	0.663	0.663	0.663
Adjusted R <sup>2</sup>	0.580	0.603	0.614	0.613	0.605
Residual Std. Error	0.667 (df = 48)	0.649 (df = 48)	0.639 (df = 48)	0.640 (df = 48)	0.647 (df = 47)
F Statistic	11.831*** (df = 7; 48)	12.933*** (df = 7; 48)	13.512*** (df = 7; 48)	13.464*** (df = 7; 48)	11.549*** (df = 8; 47)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Las variables de rendimiento, al igual que en la tabla anterior, van entrando de forma secuencial desde la columna 1 hasta la 4 y Goles Concedidos con  $\frac{(xG - Goles)}{Tiros\ al\ arco}$  compiten en la última. De la misma forma que en la tabla anterior, los coeficientes se interpretan como el aumento o disminución porcentual con respecto al valor de mercado, multiplicando el coeficiente por cien. En la primera columna podemos apreciar como las Salvadas es significativo y tiene signo positivo, con un efecto de aumentar en 3,5% el valor de mercado por cada

Salvada adicional efectuada por el arquero, lo cual representa un cambio con respecto a la Tabla 3. Este cambio puede deberse a lo expuesto por Gelade (2014), donde el efecto del equipo esta siendo controlado por tiros al arco, donde Salvadas representaría el efecto que tiene el atajar tiros al arco sobre el valor que el mercado, independiente si es un arquero de equipo grande o no. En la columna 2, se tiene Goles Concedidos promedio de los arqueros durante el campeonato, este es significativo y con signo negativo, lo cual es intuitivo. En la columna 3 y 4, entran respectivamente las variables de rendimiento  $(xG - Goles)$  y  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ , las cuales tienen un efecto positivo y significativo en la relación con el valor de mercado de los arqueros. En la columna 5, nuevamente se hacen competir goles concedidos con  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ , pero esta vez teniendo un resultado distinto, goles concedidos pierde significancia, pero los signos se mantienen igual.

Con respecto a las variables de control seleccionadas por el método LASSO, se aprecia que Edad y  $Edad^2$  (Edad\_2) tienen coeficientes no significativos, positivo y negativo, respectivamente, para el valor de mercado, lo cual es intuitivo. Respecto a la liga en la cual juega el arquero, toma como caso base un arquero de la liga argentina e indica que jugar en Chile influye de forma significativa y negativa en el valor de mercado. Partidos jugados (Matches played) es significativa con coeficientes positivos, coherentes con la intuición. LASSO también selecciona la posición en la que finalizó el campeonato el equipo (Pos.) la cual es significativa y negativa, a medida que el equipo finalizó en una peor posición, menor es el valor de mercado del arquero. Este resultado se podría atribuir que los arqueros de equipos menos competitivos son de menor valor de mercado, debido a que no pueden costear a un mejor arquero o uno de mayor valor. Tiros al arco (Opponents' shots on target) pierde significancia en algunas columnas, pero en las 5 columnas tiene signo negativo, lo cual es consistente a lo expuesto por Gelade (2014), los equipos de menor categoría reciben más tiros al arco que los equipos más grandes. Por lo cual, este efecto negativo puede estar asociado a que los arqueros de equipos más chicos, son en promedio de menor valor de mercado, debido a que los equipos tienen un menor presupuesto.

Enfocándose en los  $R^2$  de las regresiones, las columnas 3, 4 y 5 son las con mayor poder explicativo. Lo cual, representa un cambio con la Tabla 4, donde los goles concedidos dejan de ser la variable que mejora el coeficiente de determinación de las regresiones. Mirando el  $R^2$  ajustado, las regresiones de las columnas 3 y 4,  $(xG - Goles)$  y  $\frac{(xG - Goles)}{Tiros\ al\ arco}$  respectivamente, son las con mejor ajuste. Lo cual indica que al momento de controlar por tiros al arco y la posición en la tabla en la cual terminó el club del arquero, Goles Concedidos pierde

poder explicativo, dando espacio a  $(xG - \text{Goles})$  y  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  para ser utilizadas como variables explicativas del rendimiento.

Adicionalmente, se vuelve a realizar las regresiones ya expuestas, pero agregando la variable de efectividad de los pases de los arqueros. Con el fin, de evaluar si esta habilidad de los arqueros, incentiva a los defensas a realizar pases hacia portería al momento de detener al equipo rival presionando alto y bloqueando posibles pases. Si el defensa no confía en las habilidades con los pies del arquero, esto podría causar que el defensa tome más riesgos en los pases defensivos. De equivocarse en el pase, puede dejar al equipo rival con una mayor oportunidad de gol, aumentando el xG recibido por el arquero. A continuación, en la Tabla 6 se presentan los resultados de estas regresiones.

Tabla 6: Estimación valor de mercado arqueros. Variables Lasso y pases correctos.

	Dependent variable:				
	Log_valor_Mercado				
	(1)	(2)	(3)	(4)	(5)
Edad	0.071 (0.195)	-0.005 (0.196)	0.010 (0.191)	0.020 (0.190)	0.001 (0.197)
Edad_2	-0.002 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.002 (0.003)	-0.001 (0.003)
LigaChile	-0.978*** (0.200)	-0.952*** (0.191)	-1.034*** (0.187)	-1.024*** (0.187)	-0.995*** (0.200)
`Matches played`	0.031** (0.013)	0.031** (0.012)	0.027** (0.012)	0.027** (0.012)	0.028** (0.013)
`Opponents' shots on target`	-0.371** (0.180)	-0.155 (0.142)	-0.307** (0.116)	-0.292** (0.116)	-0.234 (0.176)
`Accurate passes, %`	0.058** (0.027)	0.055** (0.026)	0.050* (0.026)	0.049* (0.026)	0.051* (0.027)
Pos.	-0.064*** (0.015)	-0.059*** (0.015)	-0.060*** (0.015)	-0.060*** (0.015)	-0.059*** (0.015)
Salvadas	0.052 (0.090)				
`Goals conceded`		-0.626 (0.374)			-0.264 (0.605)
`xG-Goles`			0.667* (0.360)		
`(xG-Goles)/Tiros_al_arco`				2.783* (1.550)	1.920 (2.519)
Constant	10.005** (3.830)	11.606*** (3.857)	11.854*** (3.844)	11.629*** (3.829)	11.804*** (3.883)
Observations	56	56	56	56	56
R <sup>2</sup>	0.667	0.684	0.688	0.686	0.688
Adjusted R <sup>2</sup>	0.610	0.630	0.634	0.633	0.626
Residual Std. Error	0.643 (df = 47)	0.626 (df = 47)	0.622 (df = 47)	0.624 (df = 47)	0.629 (df = 46)
F Statistic	11.772*** (df = 8; 47)	12.694*** (df = 8; 47)	12.929*** (df = 8; 47)	12.851*** (df = 8; 47)	11.247*** (df = 9; 46)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

A diferencia de las dos tablas presentadas anteriormente, solo las regresiones 3 y 4, presentan significancia para las variables de rendimiento  $(xG - Goles)$  y  $\frac{(xG - Goles)}{Tiros\ al\ arco}$  respectivamente, al igual que un mayor  $R^2$  ajustado. Estas variables tienen la misma interpretación, el aumento porcentual por cada unidad adicional. Mientras que, las otras variables de rendimiento, estas pierden su poder explicativo, a pesar de que las regresiones tienen un mayor  $R^2$  ajustado que todas las regresiones antes presentadas.

Con respecto a las variables de control, tienen comportamientos similares a los expuestos en las Tablas 4 y 5. En particular, la habilidad con los pases de los arqueros (Accurate passes, %) tiene un comportamiento esperable, su coeficiente es positivo y significativo. Este se puede interpretar como el aumento en el valor de mercado del arquero, por cada aumento en un punto porcentual de sus pases correctos, es de 4,9% del valor, para la regresión 4.

Finalmente, se evalúa si existe correlación entre la consistencia de las salvadas esperadas por tiro al arco,  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ , de los arqueros partido a partido. También se estudia si el rendimiento del partido anterior afecta al partido siguiente, evaluando la presencia de componentes de autoregresion de la variable  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ . Se presenta la Tabla 7, donde se realiza la evaluación de la autoregresion.

Tabla 7: Evaluación autoregresividad

	<i>Dependent variable:</i>
	`(xG-Goles)/Tiros_al_arco.x`
performance_anterior	0.001 (0.032)
Constant	0.051*** (0.010)
Observations	1,043
R <sup>2</sup>	0.00000
Adjusted R <sup>2</sup>	-0.001
Residual Std. Error	0.310 (df = 1041)
F Statistic	0.0004 (df = 1; 1041)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Se aprecia que las salvadas esperadas por tiros al arco, no son afectadas y no se explican por las salvadas esperadas por tiros al arco del partido anterior (performance\_anterior), dado que el coeficiente asociado no tiene poder explicativo. Por lo tanto, no existe autoregresividad en  $\frac{(xG - Goles)}{Tiros\ al\ arco}$ .

# Capítulo 7: Discusión

El principal hallazgo es la validación de Goles Concedidos como variable de medición del rendimiento de los arqueros. Uno de sus mayores atributos es la facilidad de utilizar y de fácil acceso, ya que, no es necesario tener un modelo matemático detrás, como es el caso de la esperanza de gol. Cabe destacar también, que al momento de hacerla competir con  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$ , esta última perdió poder explicativo. Con respecto, a la diferencia entre los  $R^2$  las regresiones son muy similares, lo cual indica, que no hay un modelo preponderante y que tanto Goles Concedidos, como  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  pueden ser utilizados como variable de rendimiento para arqueros.

Por lo expuesto anteriormente,  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  se puede catalogar como una buena señal, la cual es confiable y entrega información valiosa, adicional a los goles, dado que por construcción incluye la dificultad de los tiros recibidos por el arquero y la cantidad de estos. Esta se relaciona positivamente y significativamente con el valor de mercado de los jugadores, escalando de forma intuitiva el rendimiento de los arqueros.

También existe un valor económico en la variable propuesta, dado que se puede utilizar para comparar el valor de mercado actual del arquero con su valor estimado y evaluar si el portero está siendo valorado por sobre su rendimiento o no. Esto contribuye a la toma de decisiones informada por parte de los clubes de Fútbol y sus cuerpos técnicos. Suponiendo que el valor estimado por la regresión LASSO es el verdadero valor que debería tener un arquero dado su rendimiento y sus características demográficas, se procede a calcular el porcentaje de valoración, como el valor de mercado actual sobre el valor estimado. Este indicador nos entrega información de cuanto más o menos valor el mercado le está asignando al portero en cuestión. Los valores menores a 100%, indican que el arquero en cuestión está siendo sub estimado por el mercado, dado que se le está asignando un precio menor del que debería dado su rendimiento. Mientras que los valores por sobre el 100% indica que el mercado está sobre valorando al arquero y que le asignó un precio mayor al que debería.

En la Tabla 6 se exponen los 10 arqueros más sub valorados de la base estudiada, siendo Lucas Hoyos, arquero de Vélez Sarsfield, el portero más sub valorado por el mercado. Esto lo convertiría, y cualquiera de los 10 arqueros listados, en una gran oportunidad de mercado, para algún club que esté en búsqueda de un portero con un rendimiento que el



mercado no haya descubierto o interiorizado aún. A diferencia de Gabriel Arias, el portero arquero más sobre valorado de la muestra, Tabla 7, quien se dispone como una de las peores oportunidades de mercado, dado que su precio estimado según su rendimiento es un 19% de su valor de mercado, presentado una sobre valorización del 538%. Cabe destacar que esta afirmación se hace estrictamente en base al valor estimado en comparación al de mercado.

Sub valorado	Arquero	Liga	Equipos_abr	Valor de mercado	Valor estimado	Porcentaje valoración
1	Lucas Hoyos	Argentina	Vélez Sarsfield	€ 800.000	€ 2.815.356	28%
2	Juan Pablo Romero	Argentina	Rosario Central	€ 125.000	€ 398.331	31%
3	Joaquín Muñoz	Chile	Audax Italiano	€ 450.000	€ 1.176.829	38%
4	Ignacio González	Chile	D. Antofagasta	€ 300.000	€ 750.073	40%
5	Hernán Muñoz	Chile	Ñublense	€ 100.000	€ 234.546	43%
6	Sebastián Torrico	Argentina	San Lorenzo	€ 100.000	€ 229.497	44%
7	Ramiro Macagno	Argentina	Newell's OB	€ 450.000	€ 951.289	47%
8	Lucas Acosta	Argentina	Lanús	€ 470.000	€ 951.474	49%
9	Leandro Requena	Chile	Cobresal	€ 200.000	€ 391.314	51%
10	Federico Lanzillota	Argentina	Argentinos Jr	€ 400.000	€ 681.163	59%

Tabla 8: Ranking arqueros más sub valorados.

Sobre valorado	Arquero	Liga	Equipos_abr	Valor de mercado	Valor estimado	Porcentaje valoración
1	Gabriel Arias	Argentina	Racing	€ 3.800.000	€ 706.532	538%
2	Lautaro Morales	Argentina	Lanús	€ 3.000.000	€ 956.402	314%
3	Fernando De Paul	Chile	U. de Chile	€ 1.400.000	€ 525.263	267%
4	Lucas Chaves	Argentina	Argentinos Jr	€ 3.500.000	€ 1.473.289	238%
5	Franco Armani	Argentina	River Plate	€ 5.000.000	€ 2.218.442	225%
6	Sebastián Sosa	Argentina	Independiente	€ 2.500.000	€ 1.257.077	199%
7	Jorge Broun	Argentina	Rosario Central	€ 450.000	€ 238.971	188%
8	Zacarías López	Chile	D. La Serena	€ 1.200.000	€ 659.126	182%
9	Alan Aguerre	Argentina	Newell's OB	€ 725.000	€ 416.094	174%
10	Augusto Batalla	Chile	O'Higgins	€ 550.000	€ 321.780	171%

Tabla 9: Ranking arqueros más sobre valorados.

Entendiendo que este estudio no cambiará el modo en que el mercado valoriza a los arqueros, es un aporte que entrega información del rendimiento de estos. Lo cual, permite discernir si un arquero está con un rendimiento por debajo de su valor de mercado (Sobre valorado) o si este se encuentra por arriba de su rendimiento (Sub valorado). Ejemplo de lo anterior, es el caso de Ignacio González, arquero de Antofagasta en la temporada 2021, quien a pesar de ser el arquero con mejor rendimiento, utilizando  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  como medición, es de los arqueros con menor valor de mercado en comparación a porteros con rendimientos similares. Por lo tanto, la variable propuesta es una buena señal que aporta información adicional a los Goles Recibidos, por lo que entrega una oportunidad de mercado y en ocasiones permitirá adelantarse al mercado y aprovechar de fichar arqueros que tengan un rendimiento superior al valor que el mercado le asigna.

Con respecto a las limitaciones,  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  no tiene poder de predicción sobre el rendimiento futuro de los arqueros estudiados con la misma, dado que es una variable calculada en base a lo mostrado partido a partido y los datos no son lo suficientemente desagregados para entender cómo se puede predecir esta variable mirando las acciones individuales minuto a minuto del arquero y del equipo.

Finalmente, en miras de futuros pasos, se abre la oportunidad para estudiar cómo se relacionan  $\frac{(xG - \text{Goles})}{\text{Tiros al arco}}$  con otras estadísticas tanto del arquero como del mismo equipo, con la finalidad de explicar esta variable y ser capaces de predecir el rendimiento futuro de distintos arqueros. Utilizando la metodología propuesta en la presente tesis, se busca validar la búsqueda de distintas métricas cuantitativas para ir medir el rendimiento de las distintas posiciones que existen el fútbol e ir cada vez agregando más información a la toma de decisiones.

# Bibliografía

1. Baron, J., and J.C. Hershey. (1988). "Outcome bias in decision evaluation." *Journal of Personality and Social Psychology*, 54(4): 569.
2. Batistič, S., & van der Laken, P. (2019). History, evolution and future of big data and analytics: a bibliometric analysis of its relationship to performance in organizations. *British Journal of Management*, 30(2), 229-251.
3. Berger, V. W., & Zhou, Y. (2014). Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online*.
4. Berri, D., Butler, D., Rossi, G., Simmons, R., & Tordoff, C. (2023). Salary determination in professional football: empirical evidence from goalkeepers. *European Sport Management Quarterly*, 1-17.
5. Berri, D., & Simmons, R. (2009). Race and the evaluation of signal callers in the national football league. *Journal of Sports Economics*, 10(1), 23-43.
6. Bol, Jasmijn C. 2011. "The determinants and performance effects of managers' performance evaluation biases." *The Accounting Review*, 86(5): 1549-1575.
7. Catarino, L. M., Carvalho, H. M., & Gonçalves, C. E. (2017). Analysing tactical knowledge through team sport assessment procedure/TSAP: a case study in basketball.
8. Clemente, F. M. (2018). Performance outcomes and their associations with network measures during FIFA World Cup 2018. *International Journal of Performance Analysis in Sport*, 18(6), 1010-1023.
9. Dewatripont, M., & Bolton, P. (2005). Contract theory (No. 2013/9543). ULB--Universite Libre de Bruxelles.
10. Dicks, M., Pocock, C., Thelwell, R., & van der Kamp, J. (2017). A Novel on-Field Training Intervention Improves Novice Goalkeeper Penalty. *The Sport Psychologist*.

11. Di Salvo, V., Baron, R., Tschan, H., Montero, F. C., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International journal of sports medicine*, 28(03), 222-227.
12. Farber, H. (1999). Mobility and stability: The dynamics of job change in labor markets. In O. Ashenfelter, & D. Card (Eds.), *Handbook of labor economics*, Vol 3, Part B (pp. 2439–3483). Elsevier.
13. Gauriot, R., & Page, L. (2019). Fooled by performance randomness: Overrewarding luck. *Review of Economics and Statistics*, 101(4), 658-666.
14. Gelade, G. (2014). Evaluating the ability of goalkeepers in English Premier League football. *Journal of quantitative analysis in sports*, 10(2), 279-286.
15. Gavião, L. O., Gavião, E. V., Sant’Anna, A. P., Lima, G. B. A., & Garcia, P. A. D. A. (2021). Performance analysis of professional soccer goalkeepers by Composition of Probabilistic Preferences. *Revista Brasileira de Ciências do Esporte*, 43.
16. Grehaigne, J. F., Godbout, P., & Bouthier, D. (1997). Performance assessment in team sports. *Journal of teaching in Physical Education*, 16(4), 500-516.
17. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.
18. He, M., Cachucho, R., & Knobbe, A. J. (2015, September). Football Player's Performance and Market Value. In *Mlsa@ pkdd/ecml* (pp. 87-95).
19. Holmström, B. (1982). Moral hazard in teams. *The Bell journal of economics*, 324-340.
20. Holmström, B. (1979). Moral hazard and observability. *The Bell journal of economics*, 74-91.
21. Hughes, M., Caudrelier, T., James, N., Redwood-Brown, A., Donnelly, I., Kirkbride, A., & Duschesne, C. (2012). Moneyball and soccer-an analysis of the key performance indicators of elite male soccer players by position. *Journal of Human Sport and Exercise*, 7(2), 402-412.
22. Kenley, R., & Seppänen, O. (2010). *Location-Based Management for Construction: Planning, Scheduling and Control*. Abingdon, Inglaterra: Spon Press.

23. Kuper, S., & Szymanski, S. (2018). *Soccernomics: Why England loses, why Germany and Brazil win, and why the US, Japan, Australia, Turkey--and even Iraq--are destined to become the kings of the world's most popular sport*. Hachette UK.
24. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Wasserman, W. (2004). *Applied linear regression models* (Vol. 4, pp. 563-568). New York: McGraw-Hill/Irwin.
25. Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.
26. Liu, H., Gómez, M. A., & Lago-Peñas, C. (2015). Match performance profiles of goalkeepers of elite football teams. *International Journal of Sports Science & Coaching*, 10(4), 669-682.
27. Longo, U. G., Sofi, F., Dinu, M., Candela, V., Salvatore, G., Cimmino, M., ... & Denaro, V. (2018). Functional performance, anthropometric parameters and contribution to team success among Italian "Serie A" elite goalkeepers during season 2016-2017. *The Journal of Sports Medicine and Physical Fitness*, 59(6), 969-974.
28. Montesano, P. (2016). Goalkeeper in soccer: performance and explosive strength. *Journal of Physical Education and Sport*, 16(1), 230.
29. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
30. Montejo, A. P. (2001). Evaluación del desempeño laboral. *Gestión*, 2(9).
31. Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(5), 1-27.
32. Prendergast, Canice, and Robert H Topel. (1996). "Favoritism in Organizations." *The Journal of Political Economy*, 104(5): 958-978.
33. Prendergast, Canice, and Robert Topel. (1993). "Discretion and bias in performance evaluation." *European Economic Review*, 37(2): 355-365.

34. Rathke, A. (2017). An examination of expected goals and shot efficiency in soccer. *Journal of Human Sport and Exercise*, 12(2proc), S514-S529.
35. Sainz de Baranda, P., Adán, L., García-Angulo, A., Gómez-López, M., Nikolic, B., & Ortega-Toro, E. (2019). Differences in the offensive and defensive actions of the goalkeepers at women's FIFA World Cup 2011. *Frontiers in psychology*, 10, 223.
36. Sainz De Baranda, P., Ortega, E., & Palao, J. M. (2008). Analysis of goalkeepers' defence in the World Cup in Korea and Japan in 2002. *European Journal of Sport Science*, 8(3), 127-134.
37. Simmons, R. (2022). Professional labor markets in the Journal of Sports Economics. *Journal of Sports Economics*, 23(6), 728-748.
38. Simmons, R., & Berri, D. (2011). Mixing the princes and the paupers: Pay and performance in the National Basketball Association. *Labour Economics*, 18(3), 381-388.
39. Szymanski, S. (2020). Sport analytics: Science or alchemy?. *Kinesiology Review*, 9(1), 57-63.
40. Tom L. G. Bergkamp, Wouter G. P. Frencken, A. Susan M. Niessen, Rob R. Meijer & Ruud. J. R. den Hartigh (2021): How soccer scouts identify talented players, *European Journal of Sport Science*.
41. White, A., Hills, S. P., Cooke, C. B., Batten, T., Kilduff, L. P., Cook, C. J., ... & Russell, M. (2018). Match-play and performance test responses of soccer goalkeepers: A review of current literature. *Sports Medicine*, 48, 2497-2516.
42. Yam, D. (2019). A data driven goalkeeper evaluation framework. In 13th MIT Sloan Sports Analytics Conference (pp. 1-18). Boston: MIT Sloan.