

Simulacion TemporadaFutbol

Resumen ejecutivo

El objetivo de este trabajo fue simular, mediante métodos estadísticos, un conjunto de temporadas de fútbol correspondientes a una liga cualesquiera. A través de este enfoque se buscó analizar cómo se distribuyen los títulos a lo largo del tiempo y, a partir de ello, obtener una aproximación a la “grandeza” relativa de los distintos clubes. También se estudia la relación éxito-presupuesto, enfocando el rendimiento de la inversión respecto del puntaje obtenido en un torneo. Se analiza entonces la racionalidad en torno a una inversión futbolística, y como esta varía según la etapa en la que la institución se encuentre.

Para construir este modelo se incorporaron diversas variables que afectan directamente el rendimiento deportivo de los equipos. Entre ellas se incluyeron la probabilidad de anotar un gol, la probabilidad de sufrir una lesión y la variación del presupuesto disponible, el cual se actualiza en función de los resultados obtenidos en cada temporada. La combinación de estos elementos genera un entorno dinámico en el que las decisiones y desempeños acumulados influyen en el desempeño futuro. De esta forma, la simulación ofrece una herramienta útil para comprender la evolución competitiva de los clubes en horizontes temporales prolongados.

Presentación del problema

El problema que se aborda en el proyecto consiste en modelar, mediante técnicas de estadística computacional, la evolución de una liga de fútbol profesional a lo largo de un horizonte de varias temporadas. En particular, se busca responder preguntas del tipo:

- ¿Cómo se distribuyen los campeonatos entre clubes cuando se consideran simultáneamente factores futbolísticos (capacidad anotadora, lesiones), económicos (presupuesto) y aleatoriedad inherente al juego?
- ¿Hasta qué punto el presupuesto condiciona la probabilidad de convertirse en un “equipo grande” en términos de títulos acumulados?
- ¿Qué diferencias surgen entre un escenario donde todos los equipos comienzan cada temporada “desde cero” y otro donde el éxito deportivo se traduce en más recursos económicos en el futuro?
- ¿Cuánto cuesta ganar un punto más en la liga según el presupuesto?
- ¿Cómo se distribuye el costo por punto?

Se sabe que en ligas reales como La Liga española, la Uruguaya o la Serie A italiana existe una concentración importante de títulos en pocos clubes. Sin embargo, separar el efecto del presupuesto, del azar y de otros factores es complejo a partir de los datos observados. La simulación computacional permite construir un entorno controlado donde se fijan reglas explícitas de juego (estructura del torneo, premios económicos, procesos de lesión, generación de goles) y se evalúa cómo estas reglas se traducen en resultados agregados tras un gran número de temporadas simuladas.

El uso de simulación resulta especialmente adecuado en este contexto por dos motivos. En primer lugar, el sistema completo (partidos individuales con goles discretos, lesiones dependientes del tiempo, actualización de presupuestos, etc.) es demasiado complejo como para derivar resultados analíticos cerrados sobre la

distribución de títulos. En segundo lugar, la simulación permite incorporar de forma flexible supuestos alternativos —por ejemplo, cambios en la estructura de premios o en la forma en que el presupuesto influye sobre la probabilidad de anotar— y comparar sus efectos de manera directa. De esta forma, el modelo actúa como un “laboratorio virtual” para estudiar la grandeza relativa de los clubes en distintos escenarios.

Metodología

Para generar los distintos resultados, se requiere en un inicio de una base de datos a simular. Este fue fundamentalmente el primer desafío a resolver: obtener una fuente de datos. La misma se compone de esencialmente tres factores aleatorios que interactúan en el transcurso de esta liga: probabilidad de gol, probabilidad de lesión y presupuesto por equipo. Se observa entonces dos variables a nivel de jugadores y una a nivel del equipo.

Al obtener la base de datos inicial, esta funge como nuestro punto de partida para la posterior realización de simulaciones a partir de las cuales se ergirá nuestro análisis.

A continuación se procede a explicar en detalle los mecanismos utilizados para asignar los parámetros de interés a cada observación de jugador-equipo.

Estructura general de la liga

Se considera una liga ficticia compuesta por 20 equipos, cada uno de ellos con un plantel de 22 jugadores. Los dorsales del 1 al 11 corresponden a titulares y los dorsales del 12 al 22 a sus suplentes directos. Las posiciones en el campo se resumen en cuatro categorías: arquero (GK), defensor (DF), mediocampista (MF) y delantero (FW). La liga se juega bajo un formato todos contra todos a doble rueda: cada equipo disputa 38 partidos por temporada (19 como local y 19 como visitante).

La asignación de puntos por partido se da según el sistema clásico (3 puntos por victoria, 1 por empate, 0 por derrota) y la clasificación final se obtiene ordenando a los equipos por puntos y, en caso de empate, por diferencia de goles. Al finalizar la temporada se otorgan premios económicos decrecientes a los diez primeros equipos de la tabla, que impactan en el presupuesto disponible para la próxima temporada.

Módulo de generación de planteles y presupuestos

La instancia base de la liga se construye de forma simulada pero respetando ciertas regularidades observadas en ligas reales:

- Cada equipo recibe un presupuesto inicial medido en millones de dólares. Se consideran dos rangos diferenciados:
 - clubes “grandes”, con presupuestos uniformes entre 800 y 1200,
 - clubes “pequeños”, con presupuestos uniformes entre 20 y 400.
- A nivel de jugadores, se generan 22 futbolistas por equipo, asignándoles posición (GK, DF, MF, FW) según su dorsal y probabilidades individuales de anotar gol coherentes con su rol en el campo.
- Se incorporan probabilidades de lesión dependientes de la posición, más bajas para arqueros y más altas para mediocampistas y delanteros. Las probabilidades se asignan de manera arbitraria. Posteriormente, estas probabilidades se ajustan en función del presupuesto del equipo, de modo que los clubes más ricos (que se supone disputan más competiciones) enfrentan una mayor exposición a lesiones.

Calibración de probabilidades de gol

Un aspecto central de la metodología es que las probabilidades de gol por posición se calibran para reproducir el promedio de goles observado en la temporada 2024/2025 de La Liga (torneo español), en la que se registraron 995 goles, con un promedio de 2.61 goles por partido. El procedimiento sigue dos etapas:

1. Obtención de semillas a partir de datos reales.

A partir de archivos de estadísticas de jugadores de la temporada de referencia, se calcula para cada posición (GK, DF, MF, FW) el promedio de goles por partido. Dichos promedios se utilizan como probabilidades “semilla” por posición, que reflejan la intensidad ofensiva relativa de cada rol. Para los suplentes (dorsales > 11) se utiliza la mitad de la probabilidad de su correspondiente titular, capturando su menor impacto esperado en el marcador.

2. Ajuste mediante simulación Monte Carlo.

Sobre esas semillas se aplica un algoritmo de calibración iterativo que escala un factor común multiplicativo hasta que el promedio de goles generado por el modelo converge al objetivo de 2.61 goles por partido. En cada iteración se simula un número fijo de partidos utilizando el motor de partidos completo (incluyendo efecto presupuestal, suplentes y lesiones) y se calcula el promedio de goles simulados. El factor se actualiza en función de la discrepancia entre el promedio simulado y el objetivo, manteniendo siempre la jerarquía entre posiciones (es decir, si inicialmente $P(FW) > P(MF)$ y $P(MF) > P(DF)$, esa relación se preserva).

El resultado de este módulo es un conjunto de probabilidades de gol calibradas por posición, que garantizan que el modelo reproduzca un nivel de anotaciones similar al observado en la liga real.

Motor de lesiones

Las lesiones se modelan de manera dinámica y dependiente del tiempo. El proceso funciona de la siguiente forma:

- Cada jugador tiene una probabilidad base de lesionarse en un partido, asignada arbitrariamente, que depende de su posición (menor en GK, mayor en MF/FW).
- Esta probabilidad se multiplica por un factor que crece linealmente con el presupuesto del equipo, acotado para que el efecto total quede entre 1 y 2. La idea es que los equipos ricos, al participar en más competiciones, están más expuestos a lesiones.
- Para cada partido, se actualiza la duración remanente de las lesiones ya existentes (disminuye en 1 partido a partido hasta llegar a 0) y se simulan nuevas lesiones solo para los jugadores que actualmente están sanos. La duración de una nueva lesión se asigna al azar entre 1 y 8 partidos.

La variable booleana que indica si un jugador está lesionado en un momento dado se utiliza posteriormente en la selección de titulares y suplentes para cada encuentro.

Motor de partidos

El motor de partidos integra el efecto del presupuesto, las lesiones y las probabilidades individuales de gol para generar el resultado de cada encuentro. El esquema general es el siguiente:

1. Ajuste por presupuesto.

Dado un enfrentamiento entre dos equipos, se identifica cuál de ellos tiene menor presupuesto y se calcula la razón de presupuestos $\text{pairRatio} = \frac{\text{presupuesto_grande}}{\text{presupuesto_chico}}$. A partir de este valor se construye un factor de corrección budgetDif que aumenta las probabilidades de victoria del equipo más débil, aumentando ligeramente sus probabilidades de gol. Este factor se define de manera que siempre haya posibilidad de sorpresa, incluso ante diferencias económicas grandes.

2. Selección de jugadores disponibles.

Para cada equipo se parte de los 11 titulares teóricos. Los jugadores lesionados son descartados y, en su lugar, se activa a su suplente directo (dorsal + 11).

3. Generación de goles.

Una vez definidos los jugadores que efectivamente disputan el encuentro, se utilizan sus probabilidades ajustadas de gol para simular conteos de goles individuales. El modelo se basa en distribuciones de conteo (geométrica) que permiten capturar la sobredispersión observada en goles reales, donde la varianza excede a la media y existen partidos con actuaciones excepcionalmente destacadas.

4. Cálculo del marcador y registro de “batacazos”.

Los goles individuales se agregan a nivel de equipo para obtener el marcador final. Cuando el equipo presupuestalmente más débil gana, el modelo clasifica la magnitud del “milagro” en función de pairRatio y emite una marca cualitativa (una escala de emojis).

El motor de partidos devuelve los goles por jugador y por equipo, así como la información necesaria para actualizar los puntos y las estadísticas acumuladas de la liga.

Motor de liga y actualización económica

El motor de liga se construye sobre el calendario generado por el algoritmo round-robin. Para cada jornada:

- Se recorren todos los partidos de la fecha y se llama al motor de partidos, actualizando goles, puntos y partidos jugados.
- Opcionalmente, se puede visualizar la tabla intermedia tras cada jornada, permitiendo un seguimiento paso a paso de la evolución de la competición.

Al finalizar la temporada:

- Se ordena la tabla final de posiciones por puntos y goles.
- Se asignan premios económicos decrecientes a los 10 primeros equipos (por ejemplo, 50, 30, 20, 15, 10, 8, 6, 5, 4 y 2 millones de dólares).
- Estos premios se suman al presupuesto inicial de cada club, de modo que el rendimiento deportivo tiene un efecto directo sobre los recursos disponibles para la temporada siguiente.

Este mecanismo permite estudiar tanto ligas estacionarias, donde los presupuestos se reinician cada año, como ligas acumulativas en las que el éxito pasado refuerza la ventaja futura.

Simulación de múltiples temporadas

Para analizar la distribución de títulos en el largo plazo se implementa un módulo de simulación de múltiples ligas que repite el siguiente esquema n veces:

1. Inicialización del estado de la liga.

Existen dos modos de trabajo:

- **Monte Carlo puro** (`mantener_evolucion = FALSE`): antes de cada temporada se restaura el estado inicial de equipos y jugadores a partir del momento 0 previamente capturado. Todos los campeonatos se simulan bajo las mismas condiciones iniciales, lo que permite estimar la probabilidad intrínseca de que cada equipo sea campeón.

- **Modo Dinastía** (`mantener_evolucion = TRUE`): se mantienen los presupuestos acumulados al final de la temporada anterior, pero se reinician puntos, goles, partidos jugados y lesiones. En este escenario, los equipos exitosos se vuelven progresivamente más ricos, generando una dinámica de acumulación de la riqueza.

2. Simulación de la temporada.

Se llama al motor de liga para simular las 38 jornadas de la competencia en modo silencioso (sin impresión intermedia).

3. Registro del campeón.

Una vez finalizada la temporada, se identifica al campeón y se registra su nombre en un vector de historial de campeones.

Al concluir las n simulaciones, se construye una tabla de frecuencias de títulos por equipo y se vincula esta información con los presupuestos iniciales, lo que permite explorar visualmente la relación entre riqueza y éxito deportivo. Gráficos de barras coloreados por presupuesto facilitan la interpretación de cuántas ligas gana cada club en función de sus recursos económicos.

Datos utilizados

Datos reales de calibración

Para que el modelo reproduzca un entorno futbolístico plausible, se utilizan datos reales de la temporada 2024/2025 de La Liga (primera división del fútbol español). Concretamente, se emplean archivos de estadísticas de jugadores (carpeta `PlayerStats`) que contienen, entre otras variables:

- posición del jugador (`Pos`),
- partidos disputados (`MP`),
- goles anotados (`Gls`).

A partir de estos datos se construye, para cada equipo real, una tabla de promedio de goles por partido por posición:

- se calcula $goles_{Esp} = Gls/MP$ (con valor 0 si $MP = 0$),
- se agrupa por posición y se obtiene la media de $goles_{Esp}$, resultando en un parámetro λ , es decir la probabilidad de convertir gol, por posición.

Luego se construye una tabla global que resume los valores de λ para todas las posiciones y equipos, y se los ordena según presupuesto real. Esta información se utiliza para:

- derivar las probabilidades semilla de gol por posición, que alimentan el módulo de calibración,
- establecer una correspondencia entre equipos reales y equipos simulados, de modo que los clubes ficticios con mayor presupuesto reciban parámetros consistentes con los equipos grandes de La Liga.

Datos simulados de equipos y jugadores: Estructura inicial

La estructura principal del modelo se basa en datos simulados. Se define previamente la cantidad de clubes grandes y pequeños que compondrán la liga, ya que este dato será definitivo en el comportamiento del resto de variables: presupuesto, probabilidad de gol y probabilidad de lesión.

- **Equipos:**

Se genera un data frame con 20 equipos ficticios a quienes se asignan:

- un presupuesto inicial (en millones de dólares) según el esquema de clubes grandes y pequeños
- variables para registrar: puntos de liga, goles a favor, partidos jugados y cantidad de títulos obtenidos. Todos arrancan en 0.

- **Jugadores:**

Se construye un data frame con 22 jugadores por equipo, incluyendo:

- el equipo al que pertenecen
- el número de camiseta, que determina su condición de titular o suplente (dorsal mayor a 11)
- la posición: golero, defensa, mediocampo o delantero (igualmente derivada del dorsal)
- goles acumulados en el torneo, inicializado en 0
- probabilidad de gol inicialmente asignadas, varían de acuerdo a la posición y titularidad (los suplentes tienen la mitad de gol de los titulares)
- probabilidad de lesión asignada arbitrariamente, según titularidad y posición (los suplentes no se pueden lesionar)
- indicador de lesión (binario) y duración de la lesión, ambos inicializados en 0
- partidos jugados por el equipo, igualmente arranca en 0

Implementación utilizada

Se procede a describir brevemente la estructura funcional del código utilizado para generar el simulador. Se decide añadir esto en función de la complejidad del sistema generado.

Posterior a la inicialización previamente especificada se procede a ajustar los últimos factores de los datos pre-simulación: las probabilidades de gol y lesión.

- La probabilidad de gol se procede a calibrarla de tal forma que se aproxime a los valores de promedio de gol por partido de la liga de referencia. Para eso se generaron dos caminos distintos: - la función ‘calibrar_montecarlo_exacto’ que utiliza un método iterativo de tipo montecarlo para ajustar los valores.
- Las funciones que componen el cuerpo de la sección “Metodo de Bisección” descrito más abajo. Ambos métodos arriban a las mismas conclusiones, con valores similares, por lo que son sustitutivos uno de otro.
- La probabilidad de lesión se modifica para considerar probabilidades mayores de lesión para aquellos equipos con mayor presupuesto considerados grandes. Esto dado que dichos equipos en la realidad suelen jugar mayor cantidad de competiciones externas por lo que sus planteles se ven más expuestos al riesgo de lesionarse. El factor que se busca generar está acotado entre 1 y 2 de tal manera de aumentar las probabilidades pero no de manera grosera.

Pasando a la simulación del torneo, tenemos que considerar la conformación del mismo. Tenemos 20 equipos por lo que se van a jugar 19 partidos de ida y 19 partidos de vuelta cada uno. En total se dan 38 enfrentamientos.

En un partido ocurren cuatro eventos que tenemos que funcionalizar:

- Generación de un ratio de presupuesto que computa las interacciones entre los equipos vía presupuesto, de tal forma de favorecer al equipo más rico (que en nuestro modelo se correlaciona con la grandeza). Se toma el inverso de ese ratio entre presupuestos, que resulta en un número menor a 1 que multiplicara las probabilidades de gol del equipo más débil. Contenido en la función ‘budget_math()’.

- Se inicia simulando lesiones para saber que jugadores estarán sanos a la hora de arrancar el partido. Dicha lesión se simula con una bernoulli que toma como probabilidad de éxito la anteriormente asignada. Si se dan lesiones entre los titulares, los suplentes necesarios (que recordemos tienen la mitad de probabilidad de gol). Al jugador lesionado se le simula un valor aleatorio entre uno a ocho partidos que permanecerá lesionado y que irá disminuyendo partido a partido hasta llegar a 0 (recuperación). Contenido en la función ‘*injury_sim()*’.
- Durante el partido ocurren eventos de gol que se simulan para todos los jugadores titulares de cada equipo. La simulación se hace con una geométrica. La suma de todos los goles de los jugadores de un equipo se comparan con la del contrario, esto determina el resultado. Contenido en la función ‘*sim_goals()*’.
- Se actualiza la tabla de la liga para considerar el resultado del partido recién acontecido. Contenido en la función ‘*update_results()*’.

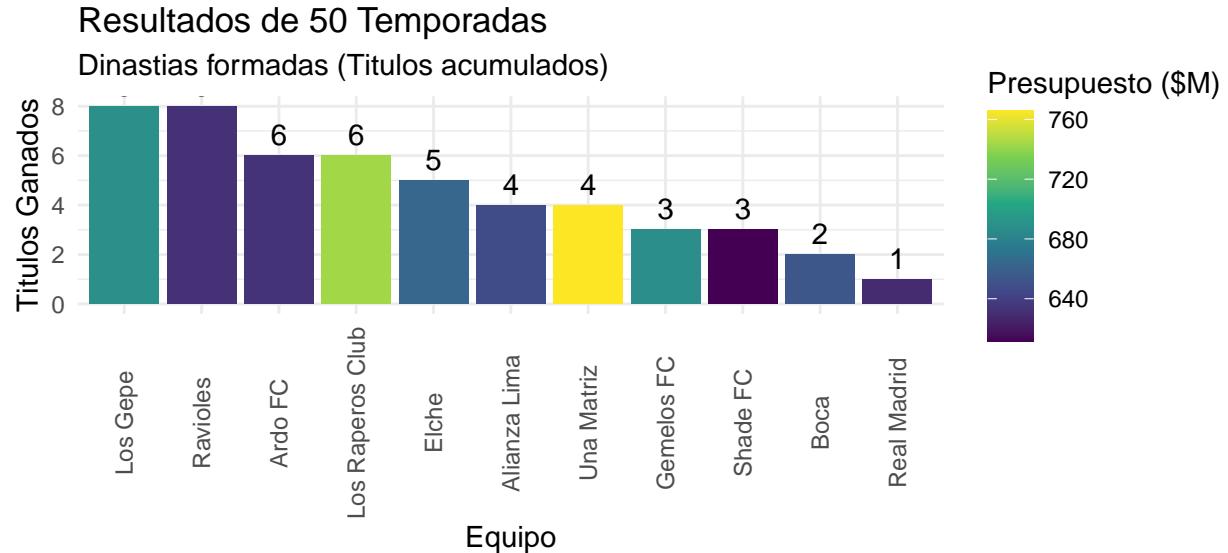
Todas estas funcionalidades están incorporadas en la función ‘*sim_match()*’.

Continuando con el pasaje desde un partido a la liga completa necesitamos generar un fixture para cada equipo durante las 38 fechas. En cada fecha se simularán 10 partidos entre pares de equipos distintos. El fixture se genera con la función ‘*gen_fixture()*’ y se itera a lo largo de todos los encuentros que el mismo genera con la función ‘*sim_liga()*’.

Finalmente con la función ‘*sim_n_ligas()*’ iteramos generando distintas ligas consecutivas. Añadimos la posibilidad de tener en cuenta o no los resultados anteriores de las ligas previas a la hora de simular las contiguas. Como se puede considerar la opción de simularlas de manera independiente, agregamos la función ‘*reset_liga()*’ para volver al estado de condiciones inicialmente establecido.

Simulamos

```
## |
```



Herramientas de análisis numérico aplicadas en el modelo

En esta sección se buscó aportar desde otro método un resultado similar al obtenido con la calibración con Montecarlo. El objetivo es aproximar el promedio de gol por partido de una liga real de referencia (la

española en este caso) a la ficticia que nosotros simulamos.

Metodo de Bisección

```
## [1] 0.60 0.65
```

```
## [1] 0.6398438
```

```
## [1] -0.002105263
```

```
## [1] 7
```

Calibración del parámetro θ y análisis de su efecto en la dinámica goleadora del modelo

Una vez consolidada la estructura del motor de simulaciones, incluyendo la composición de los planteles, la dinámica de lesiones, la diferenciación entre titulares y suplentes y la influencia presupuestal que introduce asimetrías competitivas entre equipos, se avanzó hacia la calibración del parámetro θ . Este parámetro funciona como un factor de escala que modula la probabilidad efectiva de que un jugador convierta un gol, integrando simultáneamente múltiples condicionantes: posición en la cancha, condición física, disponibilidad de reemplazos y fortaleza económica del club.

En términos conceptuales, θ opera como un coeficiente global que amplifica o atenúa la intensidad goleadora de toda la liga, traduciendo los comportamientos individuales en patrones agregados. El objetivo de la calibración consiste en encontrar un valor de θ que reproduzca el promedio empírico de 2.61 goles por partido. Formalmente, esto equivale a resolver la ecuación no lineal:

$$g(\theta) = \mathbb{E}[G(\theta)] - 2.61 = 0,$$

donde $\mathbb{E}[G(\theta)]$ representa el promedio de goles por partido observado en la simulación.

Identificación de un intervalo y resolución numérica

El estudio inicial del comportamiento de $g(\theta)$ permitió detectar un intervalo en el cual la función cambia de signo. Desde el punto de vista del análisis numérico, esto garantiza la existencia de una raíz dentro de dicho intervalo y habilita la aplicación de métodos de resolución que requieren esta propiedad.

Sobre este intervalo se aplicó un proceso iterativo que, en apenas siete pasos, condujo al valor aproximado:

$$\theta^* \approx 0.653,$$

para el cual la diferencia residual es muy cercana a cero:

$$g(\theta^*) \approx 0.003.$$

Dado el carácter estocástico del simulador, esta discrepancia es prácticamente despreciable. El resultado confirma que el modelo es sensible al ajuste de θ y que las múltiples dinámicas internas —lesiones, sustituciones, diferencias posicionales, ventajas presupuestales y distribución de minutos— responden coherentemente a variaciones en la intensidad goleadora.

Herramientas de análisis numérico aplicadas en el modelo

Aunque el motor de simulaciones incorpora módulos específicos (distribución de probabilidades, sistema de lesiones, ajustes económicos, Monte Carlo), su estructura conceptual descansa en herramientas clásicas de análisis numérico. La calibración de θ constituye, en esencia, un problema de encontrar la raíz de una ecuación no lineal del tipo:

$$f(\theta) = \text{goles simulados} - \text{goles objetivo} = 0,$$

equivalente a la formulación más general:

$$g(\theta) = \mathbb{E}[G(\theta)] - 2.61.$$

Para resolver ecuaciones de este tipo suelen emplearse métodos como:

- **bisección**, cuando se conoce un intervalo con cambio de signo;
- **iteración sobre un punto fijo**, cuando se buscan aproximaciones sucesivas sin derivadas;
- **Newton–Raphson**, cuando es posible estimar de manera fiable la derivada de $g(\theta)$.

Estos enfoques comparten la idea de ajustar iterativamente el valor de θ para que la diferencia entre lo producido por el modelo y el valor objetivo desaparezca.

Por qué se eligió el método de bisección

Aunque existen alternativas más rápidas en teoría, como Newton–Raphson o el método de iteración sobre un punto fijo, el método de bisección ofrece ventajas críticas dadas las características del modelo:

1. No requiere derivadas

Newton–Raphson depende de estimar $g'(\theta)$.

Sin embargo, debido al carácter estocástico del simulador —donde cada evaluación de $G(\theta)$ implica una liga completa con lesiones, sustituciones y efectos económicos— la derivada numérica resulta extremadamente ruidosa y poco estable.

2. No requiere que el proceso sea contractivo

El método de punto fijo exige que la transformación sea contractiva en el entorno de la solución, condición que no está garantizada en sistemas con retroalimentaciones complejas como este.

3. Garantiza convergencia si existe cambio de signo

Una vez detectado un intervalo $[a, b]$ con $g(a)g(b) < 0$, la convergencia del método de bisección es segura, independiente del ruido estocástico o de la forma de $g(\theta)$.

Esta robustez es especialmente valiosa en modelos donde una sola evaluación implica una temporada de fútbol completa.

4. Costo computacional razonable

Aunque Newton–Raphson es más rápido teóricamente, la necesidad de reevaluar $g(\theta)$ varias veces para estimar derivadas hace que, en la práctica, la bisección resulte más eficiente cuando cada evaluación es costosa.

Por estas razones, la bisección se presenta como el método natural: estable, simple, robusto frente al ruido y perfectamente adecuado para una función $g(\theta)$ cuyo comportamiento global se conoce gracias a la exploración previa.

Breve comparación con el enfoque Monte Carlo previo

Antes de este enfoque numérico, la calibración se realizaba mediante un procedimiento exclusivamente Monte Carlo: simular miles de partidos, observar el promedio obtenido y modificar θ multiplicándolo por un factor correctivo. Aunque operativamente válido, este método dependía del azar, requería muchas iteraciones y no garantizaba estabilidad ni unicidad en la solución.

El enfoque actual, en cambio:

- estudia explícitamente la estructura matemática del problema,
- identifica un intervalo donde existe raíz,
- aplica un método determinístico y estable,
- converge con rapidez,
- y permite interpretar el resultado desde una perspectiva analítica.

De este modo, la calibración de θ se vuelve más precisa, más transparente y mejor fundamentada dentro del contexto del análisis numérico.

Optimización numérica

El módulo de calibración implementa un proceso iterativo de minimización:

$$\min_{\theta} (\text{goles simulados}(\theta) - 2.61)^2,$$

lo cual corresponde a un problema de optimización unidimensional.

Asimismo, el ajuste económico de presupuestos puede interpretarse como un problema dinámico de optimización, donde los equipos *maximizan* su probabilidad futura de ganar títulos al aumentar su presupuesto a través del rendimiento deportivo.

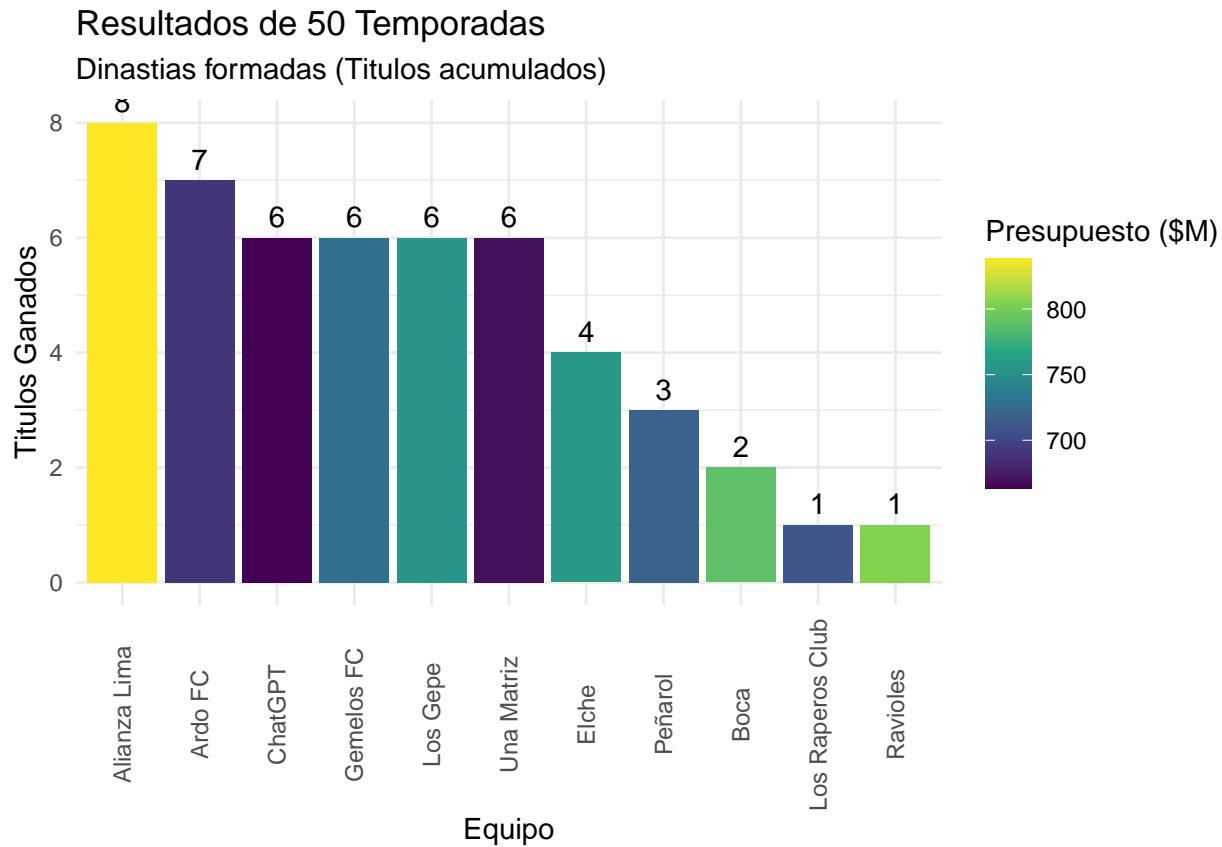
Discusión

Una vez finalizado el motor de simulaciones, tanto para cada liga individual como para el conjunto de n ligas simuladas, se procedió a indagar en las preguntas iniciales.

Concentración de títulos

Se parte estudiando la grandeza relativa de los equipos y su evolución a lo largo del tiempo. El objetivo fue analizar cómo se distribuían los títulos entre los clubes y observar el grado de concentración que emergía bajo distintos escenarios iniciales: ligas con dos equipos grandes o sin clubes grandes en la base inicial. Esta comparación permitió evaluar cómo la configuración de partida influye en la dinámica competitiva y en la acumulación histórica de campeonatos.

Los resultados se muestran a continuación. Si arrancamos con ningun equipo grande dentro de la liga obtenemos la siguiente distribución en los títulos:



Para ligas con tres equipos grandes resulta:

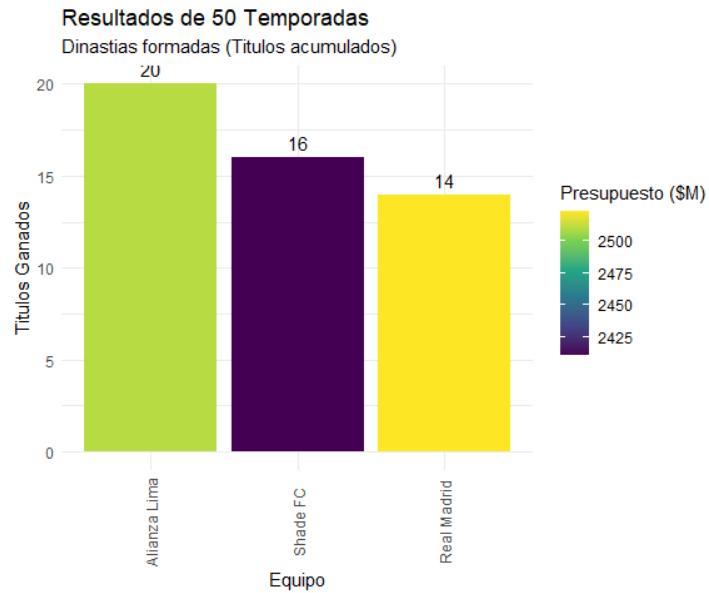


Figure 1: Concentración de titulos con Equipos Grandes

En principio, se destaca la concentración de títulos manifiesta cuando tratamos con la existencia previa de equipos grandes en la liga. En contraste, Cuando la liga parte sin equipos grandes, la distribución inicial de presupuestos es relativamente homogénea y ningún club cuenta con una ventaja marcada. Sin embargo, a

medida que avanzan las temporadas, algunos equipos comienzan a destacarse casi de forma orgánica. Esto responde a que el sistema incorpora un premio económico para el campeón de cada liga, que incrementa su presupuesto y, en consecuencia, mejora sus probabilidades de obtener mejores resultados deportivos en la siguiente temporada. Este ciclo de éxito deportivo seguido de mayor capacidad económica actúa como un mecanismo acumulativo que, con el tiempo, genera una brecha creciente entre los clubes.

En el escenario donde la liga sí comienza con 3 equipos grandes, el proceso es aún más marcado. Estos clubes ya parten con un presupuesto superior al del resto, lo que les otorga un rendimiento inicial más alto y, por lo tanto, una mayor probabilidad de ganar campeonatos. Cuando estos equipos obtienen victorias, el premio económico adicional no solo consolida su ventaja sino que la amplifica, profundizando la desigualdad competitiva. De esta forma, la estructura inicial condiciona fuertemente la evolución de la liga: en ausencia de equipos grandes, la concentración surge de manera natural a lo largo del tiempo; cuando ya existen clubes dominantes, su supremacía se intensifica y la concentración se vuelve todavía más acentuada.

Frontera del Gasto Racional

La incorporación de técnicas de optimización numérica al modelo de simulación nos permite trascender la simple descripción de resultados deportivos para auditar la estructura económica de la competición. Al aislar las variables y someter a un equipo testigo a diferentes niveles de inversión, hemos logrado trazar la “Frontera de Eficiencia”, una curva que revela cuánto cuesta realmente transformar dinero en puntos. Este enfoque es fundamental porque elimina el ruido estadístico de una temporada aislada y nos muestra el comportamiento estructural del mercado: no se trata de quién tuvo suerte un año, sino de cuáles son las reglas financieras inmutables que rigen nuestra liga virtual.

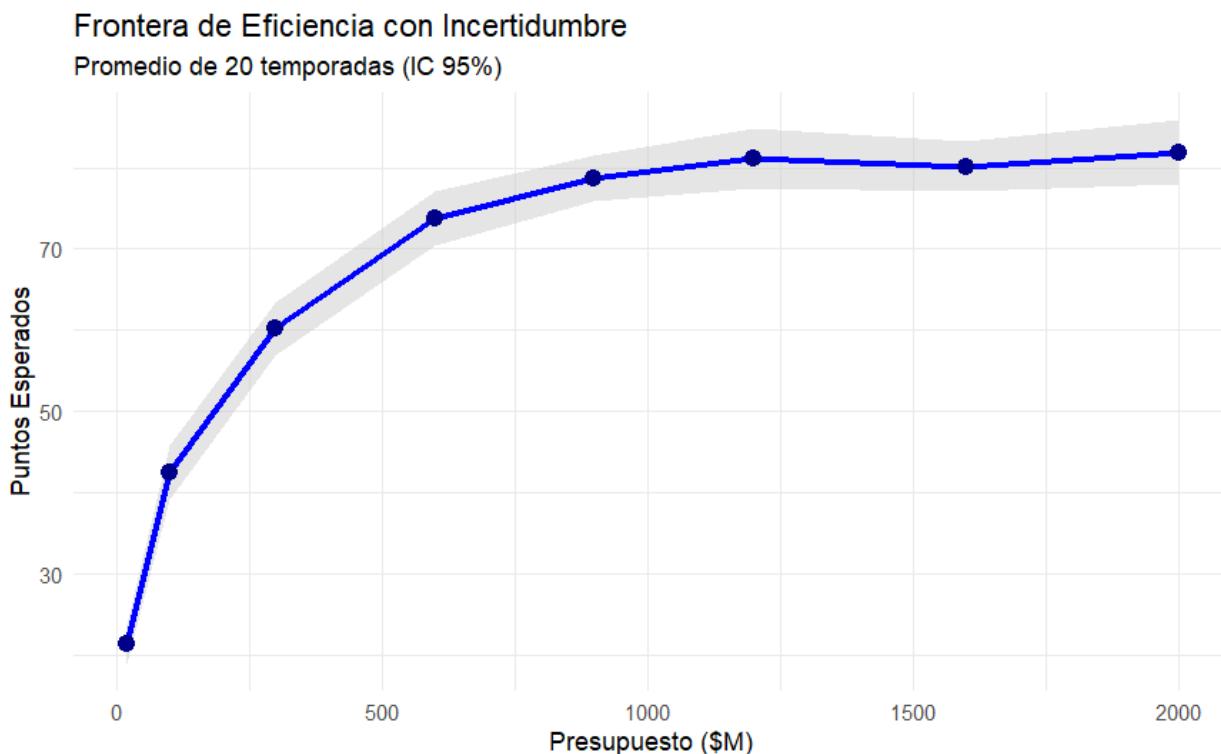


Figure 2: Frontera de eficiencia

Los resultados actualizados confirman, en primer lugar, una fase de expansión acelerada para los presupuestos bajos. El retorno de inversión inicial es extraordinario: un club que quintuplica su presupuesto de 20 a 100 millones logra, de media, duplicar su cosecha de puntos (saltando de 21 a 44,5). En esta etapa temprana, el

“precio del éxito” es sumamente accesible, con un costo marginal de apenas 3,4 millones por punto adicional. Esto sugiere que, para los equipos chicos, la prioridad estratégica debe ser salir de la zona de pobreza extrema, ya que cualquier inyección mínima de capital se traduce inmediatamente en una mejora competitiva drástica.

Sin embargo, al adentrarnos en la zona media-alta, el modelo revela un fenómeno de estancamiento prematuro. Al pasar de 600 a 900 millones de presupuesto, el rendimiento deportivo se congela casi por completo, subiendo imperceptiblemente de 73,7 a 74,2 puntos. Esto indica una zona de inefficiencia severa: gastar 300 millones adicionales apenas mueve la aguja deportiva, elevando el costo marginal a la absurda cifra de 600 millones por punto. Este hallazgo es contraintuitivo y valioso: nos advierte que existe una “trampa de inversión” en la que un club puede gastar como un gigante pero seguir obteniendo resultados de equipo mediano si no logra dar el salto cualitativo definitivo hacia la élite absoluta.

Curiosamente, si el club logra romper esa barrera y escalar hasta los 1200 millones, experimenta un segundo aire de rendimiento, alcanzando el pico máximo observado de 81,6 puntos. Pero este pico es un “falso positivo” financiero. A partir de aquí, la optimización nos muestra que el sistema colapsa bajo su propio peso. Al forzar la inversión hasta los 1600 millones, el rendimiento promedio cae a 80,7 puntos, generando un costo marginal negativo (-444 millones por punto). Esto significa que el club está pagando activamente para empeorar. La optimización numérica demuestra así que el presupuesto de 1200 millones representa el óptimo global: es el punto exacto donde se maximiza la probabilidad de campeonato antes de que la saturación de estrellas y la mecánica de lesiones comiencen a destruir valor deportivo, castigando la ambición desmedida con rendimientos decrecientes.

Costo promedio por punto

Ahora veamos como distribuye el Costo Promedio por Punto (CPP)

$$\text{CPP} = \frac{\text{Presupuesto Total}}{\text{Puntos Totales}}$$

1. **Equipos Chicos (<200M):** La Eficiencia de Supervivencia “Para los equipos chicos, el Costo por Punto sigue una distribución Weibull estrecha y asimétrica. El costo se concentra en un rango de máxima eficiencia (~3M/punto), lo que indica que estos equipos operan cerca de un límite físico de rentabilidad. Al estar protegidos de la saturación (lesiones) y beneficiados por mecánicas de equilibrio (buffs), su desempeño refleja una maximización forzada de recursos donde el despilfarro es sistémicamente imposible.”
2. **Equipos Medianos (200M - 800M):** La Fase de Transición y Costos Crecientes “En el segmento medio, la distribución muta hacia una Lognormal (o Gamma) y se desplaza hacia la derecha. El incremento en el costo medio (~5.5M/punto) y la aparición de colas hacia costos más altos evidencian el impacto de los costos marginales crecientes. Esta fase marca la pérdida de ventajas competitivas iniciales, donde ganar puntos adicionales requiere un esfuerzo financiero desproporcionalmente mayor al de la etapa anterior.”
3. **Equipos Grandes (>800M):** Saturación y Gestión de Riesgo “Para los equipos grandes, la eficiencia deja de ser una variable de optimización y se convierte en una variable aleatoria (Normal). El costo medio se estabiliza en un nivel alto estructural (~13.5M/punto) debido a los rendimientos decrecientes, y la alta varianza observada refleja que, a estos niveles de inversión, el éxito financiero depende más de la mitigación de riesgos (azar/lesiones) que de la estructura de costos base.”

Costo Marginal del Punto
Cuánto cuesta (aprox) subir al siguiente nivel competitivo

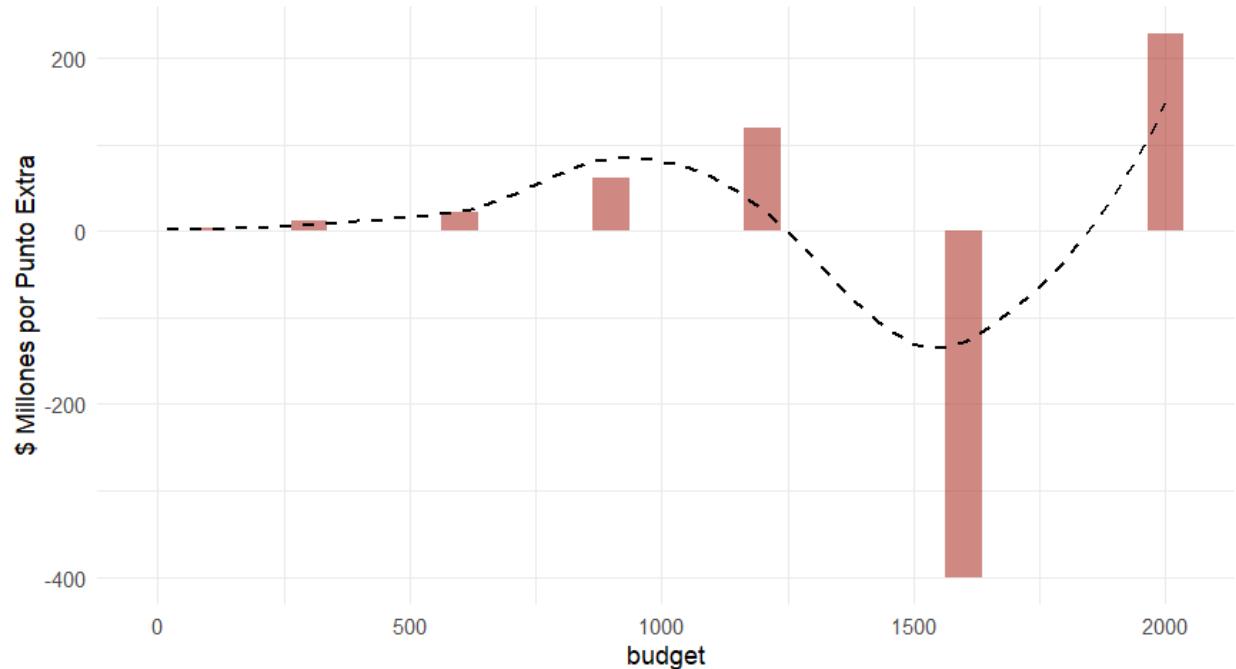


Figure 3: Costo marginal por punto

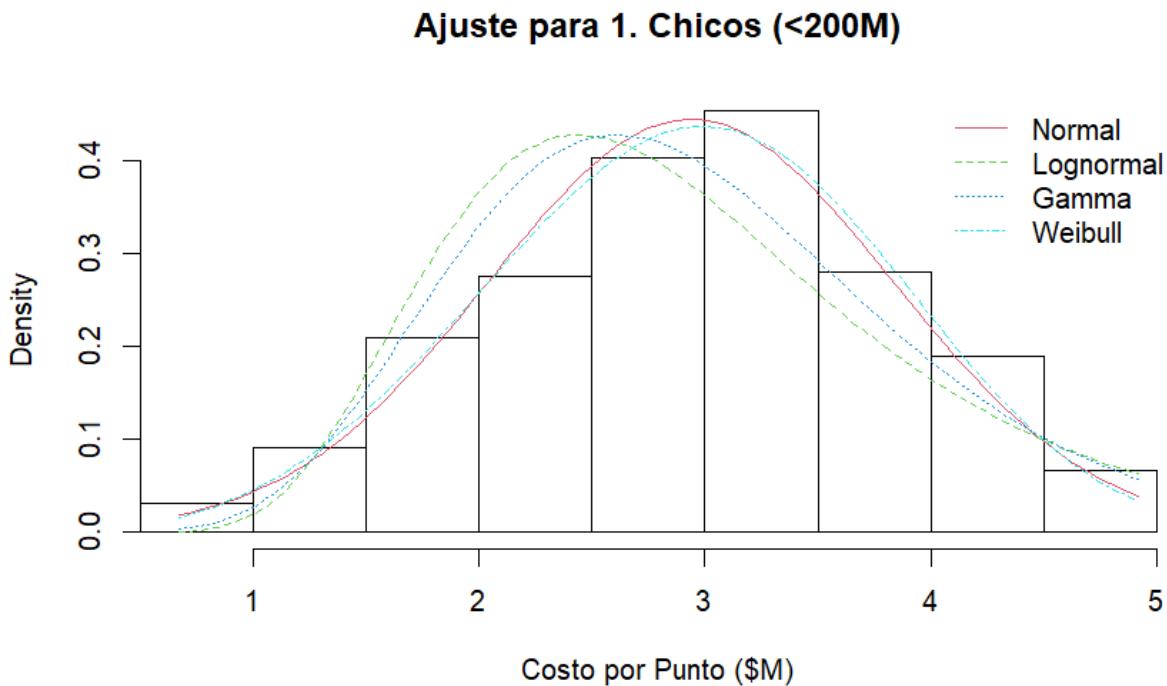


Figure 4: Disrtribucion del costo por punto para equipos chicos

Ajuste para 2. Medianos (200-400M)

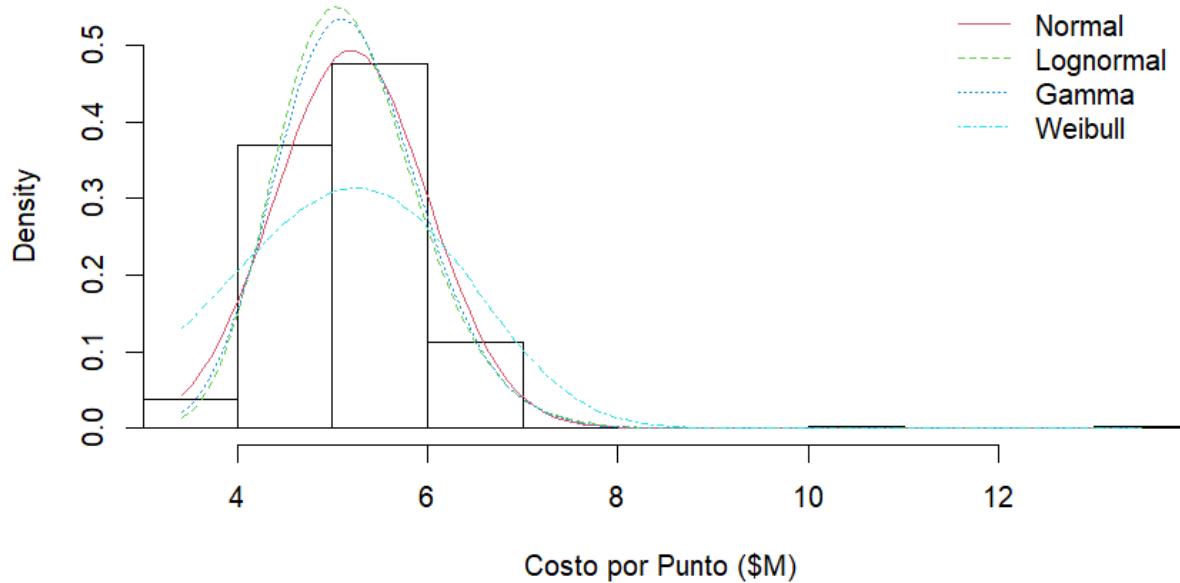


Figure 5: Disrribucion del costo por punto para equipos medianos

Ajuste para 3. Grandes (>800M)

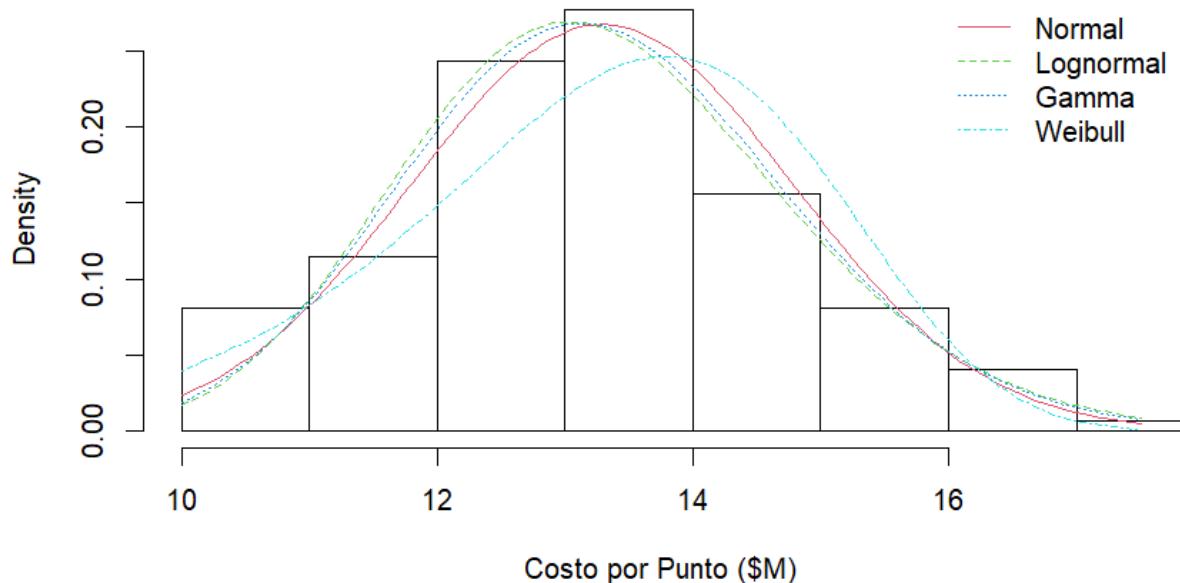


Figure 6: Disrribucion del costo por punto para equipos grandes

Bibliografía de las librerías utilizadas

Wickham H, Hester J, Francois R (2025).

dplyr: A Grammar of Data Manipulation.

R package version 1.1.4,

<https://CRAN.R-project.org/package=dplyr>.

Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C,

Woo K, Yutani H, Dunnington D (2025).

ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.

R package version 3.5.1,

<https://CRAN.R-project.org/package=ggplot2>.

Wickham H, Miller E, Bache S M, Müller K (2025).

readr: Read Rectangular Text Data.

R package version 2.1.5,

<https://CRAN.R-project.org/package=readr>.

Wickham H, Girlich M (2025).

tidyverse: Tidy Messy Data.

R package version 1.3.1,

<https://CRAN.R-project.org/package=tidyr>.

Xie Y (2025).

knitr: A General-Purpose Package for Dynamic Report Generation in R.

R package version 1.50,

<https://yihui.org/knitr/>.

Wickham H, Vaughan D, Girlich M (2025).

purrr: Functional Programming Tools.

R package version 1.0.2,

<https://CRAN.R-project.org/package=purrr>.

Müller K, Wickham H (2025).

tibble: Simple Data Frames.

R package version 3.2.1,

<https://CRAN.R-project.org/package=tibble>.

Pedersen TL (2025).

patchwork: The Composer of Plots.

R package version 1.2.0,

<https://CRAN.R-project.org/package=patchwork>.

Garnier S (2025).

viridis: Colorblind-Friendly Color Maps for R.

R package version 0.6.5,

<https://CRAN.R-project.org/package=viridis>.

Delignette-Muller ML, Dutang C (2025).

fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data.

R package version 1.1.10,

<https://CRAN.R-project.org/package=fitdistrplus>.