

---

# UN ENFOQUE ANALÍTICO PARA LA CONFECCIÓN DEL RANKING FIFA Y LA REALIZACIÓN DEL SORTEO DEL MUNDIAL DE FÚTBOL

---

SEBASTIÁN CEA<sup>\*</sup>  
GUILLERMO DURÁN<sup>\* \*\*</sup>  
MARIO GUAJARDO<sup>\*\*\*</sup>  
DENIS SAURÉ<sup>\*</sup>  
GONZALO ZAMORANO<sup>\*</sup>

## Resumen

Este trabajo analiza los procedimientos de ranking y sorteo de grupos de la Copa del Mundo utilizados por la FIFA. Calibrando un modelo predictivo, establecemos un ranking ideal de referencia sobre el cual juzgamos el desempeño de simples propuestas de cambio al procedimiento actual de ranking, las que a su vez están guiadas por un análisis cualitativo y estadístico del procedimiento actual. Luego analizamos el uso del ranking de equipos para el sorteo de los grupos en la fase final de la Copa del Mundo. Proponemos un modelo de programación lineal entera para balancear la *dificultad* de los grupos en dicha fase final. Ilustramos este enfoque utilizando datos de la Copa del Mundo 2014.

**Palabras Clave:** Investigación de Operaciones en deportes, Ranking, Copa del Mundo FIFA, Fútbol

---

<sup>\*</sup>Departamento de Ingeniería Industrial, Universidad de Chile, Santiago, Chile

<sup>\*\*</sup>Departamento de Matemática e Instituto de Cálculo, FCEN, Universidad de Buenos Aires, Buenos Aires, Argentina

<sup>\*\*\*</sup>Department of Business and Management Science, NHH Norwegian School of Economics, Bergen, Norway

---

## 1. Introducción

---

La Copa del Mundo FIFA es el evento futbolístico más popular del mundo. Organizado por la FIFA en ciclos de cuatro años, el evento congrega a las selecciones de más de 200 naciones asociadas a la FIFA. La Copa del Mundo se juega en dos fases: la fase clasificatoria y la fase final (comúnmente conocida como *Mundial*). En la fase clasificatoria, los equipos compiten dentro de sus confederaciones continentales por una cierta cantidad de cupos para la fase final. Actualmente, existen seis confederaciones: Confederación Sudamericana de Fútbol (CONMEBOL), Unión de Asociaciones Europeas de Fútbol (UEFA), Confederación Asiática de Fútbol (AFC), Confederación Africana de Fútbol (CAF), Confederación de Fútbol de Norte, Centroamérica y el Caribe (CONCACAF), y Confederación de Fútbol de Oceanía (OFC). En su edición más reciente, la fase final se disputó en Brasil en 2014 y congregó a 32 equipos, 3,4 millones de espectadores en los estadios y 3,2 billones de televidentes alrededor del mundo [8, 9].

En el formato actual de torneo, los 32 equipos clasificados al Mundial son agrupados en 8 grupos de 4 equipos cada uno. La formación de estos grupos ocurre de acuerdo a un sorteo. Una vez que los grupos han sido definidos, los equipos de un mismo grupo juegan una fase de tres partidos, todos contra todos. Los dos de mejor desempeño en cada grupo clasifican a una fase de *playoffs* que define al campeón.

En la metodología actualmente utilizada por la FIFA, el sorteo de los grupos del Mundial ocurre en base a criterios de desempeño y a criterios geográficos. El desempeño de los equipos se mide al momento del sorteo (seis meses antes del Mundial) por su posición en el *ranking FIFA* [10]. Este ranking se construye de acuerdo a una metodología que otorga puntos a cada equipo por los partidos disputados en los últimos cuatro años. Los siete equipos mejor rankeados, además del país anfitrión, son designados como sembrados (o cabezas de serie) de los grupos. Los restantes equipos son ubicados en distintos agrupamientos (llamados “pots”), según la confederación continental a la que pertenecen, y luego son sorteados a alguno de los grupos que encabezan los sembrados. Este procedimiento asegura que los sembrados no jueguen entre sí durante la fase de grupos y, además, promueve la diversidad geográfica entre los equipos que componen un mismo grupo. Si bien estos criterios parecen razonables, tanto el ranking como el sorteo han sido objetos de crítica. Por ejemplo, para el Mundial 2014, circularon varias notas de prensa deportiva y

opiniones de aficionados criticando que Colombia, Bélgica y Suiza consiguieran una posición tan alta en el ranking FIFA, a punto tal que les permitió ser cabezas de serie. Otras críticas apuntaron a la aparentemente amplia diferencia en el grado de dificultad de los distintos grupos sorteados para el Mundial. Más allá de esta evidencia anecdótica, la comunidad académica se ha gradualmente involucrado en el debate mundialístico mediante trabajos científicos. La mayor parte de los trabajos se ha enfocado en la predicción de resultados (e.g. [14], [4], [17], [7]). Algunos se han enfocado específicamente en el ranking FIFA (e.g. [15], [19], [13]) con el fin principal de estudiar o comparar su poder predictivo más que proponer modificaciones a su metodología de cálculo. Recientemente, en [6] reportan técnicas en investigación de operaciones aplicadas a la calendarización de las clasificatorias mundialistas sudamericanas. Por otro lado, la literatura sobre la formación de grupos del Mundial es muy escasa. A nuestro conocimiento, el único esfuerzo académico en esta línea es [12], que discute deficiencias del sistema de formación de grupos actualmente utilizado por la FIFA y elabora algunas propuestas de mejora.

En este artículo analizamos de forma crítica los procedimientos de construcción del ranking y de sorteo de los grupos del Mundial de la FIFA. En primer lugar analizamos la construcción del ranking FIFA. Para esto, tras demostrar una serie de falencias del procedimiento actual, establecimos un ranking *ideal* de referencia, que nos permite calificar el desempeño de posibles modificaciones sencillas a dicho procedimiento. El ranking aludido representa el resultado esperado de un torneo doble round-robin entre los equipos integrantes de la FIFA, el cual es obtenido mediante simulación de Monte Carlo. La componente principal de este procedimiento es la calibración de una variante del modelo predictivo propuesto por [14], y que es la base de la mayor parte de los modelos predictivos propuestos en la literatura. Para esto, usamos datos de partidos entre selecciones jugados entre 2009 a 2013.

Respecto a las propuestas de modificación al procedimiento de ranking, en un primer paso construimos modelos de regresión logística multinomial que nos permiten identificar variables que explican los resultados de partidos entre selecciones de fútbol. En base a nuestro análisis y los resultados de estos modelos, proponemos modificaciones simples a la metodología del ranking FIFA actual, las que son evaluadas utilizando el ranking ideal de referencia.

En segundo lugar, el trabajo estudia el uso de un modelo de programación lineal entera mixta para la formación de grupos del Mundial. Dicho modelo toma en cuenta criterios geográficos y de balance de dificultad de los grupos. Nuestros resultados muestran que es posible mejorar considerablemente la metodología actual de la FIFA para el ranking de equipos y la formación de grupos del Mundial.

El resto del artículo está organizado en cinco secciones. La Sección detalla la construcción del ranking ideal de referencia en base al modelo predictivo de [14]. En la Sección 2.1 analizamos el ranking FIFA actual e identificamos variables importantes para la predicción de resultados. En base al análisis y al resultado de esos modelos, la Sección 3.2 elabora nuevas metodologías de ranking y testea su desempeño mediante simulaciones. La Sección 4.2 formula el modelo de programación lineal entera mixta para la formación de grupos del Mundial y presenta sus resultados. El artículo concluye con una discusión final en la Sección 5.3.

---

## 2. Ranking ideal de referencia

---

Una parte significativa de las críticas contra el procedimiento actual se debe a la inclusión en el top 10 de equipos que tradicionalmente no han obtenido buen desempeño en competencias importantes (ver [15]). Dichas críticas son en esencia subjetivas, por lo que se requiere un procedimiento objetivo para comparar dos rankings. Esto es particularmente valioso en nuestro caso: nuestras propuestas buscan modificar de manera simple las reglas de puntuación en el procedimiento del ranking, de forma que estas tengan potencial de ser implementadas; cada modificación a las reglas impacta en el ranking, y es necesario entonces evaluar la bondad de dicha modificación.

Cualquier criterio de comparación de rankings lleva a la necesidad de un ranking *ideal* o *de referencia*, por lo que podemos pensar en el problema de construir el mismo. En este sentido, un ranking de referencia debiese reflejar la calidad de las selecciones, de forma que un equipo con mejor ranking debiese tener mayor probabilidad de derrotar a un equipo con menor ranking en terreno neutral. Este ranking debiese ser obtenido a partir de datos históricos, sin embargo encuentros entre selecciones normalmente ocurren en el marco de competencias internacionales de forma esporádica, e influenciados por la estructura de los torneos. El problema de inferir datos históricos a partir de comparaciones pareadas se remonta a un antiguo trabajo de [2]. Por otra parte, [13] presentan un estudio de sistemas de ranking para fútbol. En dicho trabajo, la calidad de un ranking está asociada a su capacidad predictiva, la cual se infiere a partir de un modelo de regresión logística que utiliza como variable explicativa el ranking de los equipos y la condición de localía. Tomando en cuenta esta idea, considerando que el resultado de un torneo varía considerablemente dependiendo del formato del mismo, y que el formato *round-robin* (o *todos contra todos*) es el que maximiza la correlación entre el mejor equipo y resultado final ([18]), en este trabajo proponemos utilizar como ranking de

referencia el resultado esperado de un torneo doble round-robin que considera el factor de localía. Para obtener un buen poder predictivo, calibramos una variación del modelo de [14] para simular el resultado de partidos entre dos equipos. Con esta herramienta a nuestra disposición procedemos a calcular el ranking esperado de la competencia round-robin a través de simulación de Monte Carlo. A continuación presentamos nuestra adaptación del modelo de [14].

## 2.1. Modelo Predictivo

Sean  $X_{A,B}$  e  $Y_{B,A}$  el número de goles convertidos por el equipo  $A$  y el equipo  $B$ , respectivamente, en un partido donde  $A$  es local. Siguiendo a [14] suponemos que  $X_{A,B}$  e  $Y_{B,A}$  son variables aleatorias independientes con distribución de Poisson con tasas  $\lambda_{A,B}$  y  $\gamma_{A,B}$ , respectivamente. Además, suponemos que en un partido dado la tasa con la cual un equipo anota depende de su capacidad ofensiva, la capacidad defensiva del equipo rival y la localía. En particular, asumimos que:

$$\ln(\lambda_{A,B}) = a_A - d_B + \rho_h, \quad (1a)$$

$$\ln(\gamma_{A,B}) = a_B - d_A + \rho_a. \quad (1b)$$

En (1a) y (1b), los parámetros  $a_A$  y  $d_A$  representan las capacidades ofensiva y defensiva del equipo  $A$ , respectivamente. Las variables  $\rho_h$  y  $\rho_a$  representan correcciones a las tasas que toman en cuenta la localía (en un partido en terreno neutral ambas variables  $\rho_h$  y  $\rho_a$  son omitidas). Con esto, la probabilidad de obtener el resultado  $\{X_{A,B} = m, Y_{B,A} = n\}$ , con  $m$  los goles de  $A$  y  $n$  los goles de  $B$ , cuando  $A$  y  $B$  se enfrentan, está dada por:

$$P(X_{A,B} = m, Y_{B,A} = n) = \frac{e^{-\lambda_{A,B}} (\lambda_{A,B})^m}{m!} \cdot \frac{e^{-\gamma_{A,B}} (\gamma_{A,B})^n}{n!}. \quad (2)$$

El modelo descrito arriba cuenta con dos parámetros por cada equipo, más aquellos asociados a los factores de localía y visita. Los parámetros del modelo son calibrados usando *maximum likelihood estimation* (MLE), utilizando datos de todos los partidos oficiales FIFA entre los años 2009 y 2013.

Para obtener nuestro ranking de referencia simulamos  $10^5$  torneos doble *round-robin* en terreno neutral considerando sólo a los equipos participantes en la Copa Mundial 2014, y promediamos el número de puntos obtenidos por cada equipo cuando una victoria, empate y derrota otorgan 3, 1 y 0 puntos respectivamente. La Tabla 1 muestra el ranking de referencia y una estimación de la probabilidad de ser campeón de cada país obtenida a través de las simulaciones.

Tabla 1: Simulación de torneos round-robin entre equipos participantes en Copa Mundial 2014.

Orden	País	Prob. Campeón (%)	Orden	País	Prob. Campeón (%)
1	Brasil	24,45	17	Ghana	0,17
2	España	20,49	18	Costa de Marfil	0,15
3	Argentina	15,33	19	Corea del Sur	0,076
4	Holanda	12,49	20	Suiza	0,072
5	Alemania	5,3	21	Japón	0,068
6	Inglaterra	5,26	22	Bélgica	0,042
7	Francia	3,27	23	Australia	0,03
8	Uruguay	3,0	24	Nigeria	0,031
9	Ecuador	2,94	25	USA	0,027
10	Colombia	2,0	26	Portugal	0,026
11	Croacia	1,18	27	Honduras	0,012
12	Rusia	1,07	28	Camerún	0,003
13	Italia	1,04	29	Costa Rica	0,003
14	México	0,95	30	Argelia	0,002
15	Chile	0,19	31	Grecia	0,001
16	Bosnia	0,18	32	Irán	0

---

### 3. Análisis del ranking FIFA

---

Con un ranking referencial disponible, pasamos a estudiar el procedimiento actual del ranking FIFA. Introducida en 2006, la metodología actual del ranking FIFA se basa en el puntaje (conocido como “rating”) de las 211 selecciones adultas asociadas a la FIFA. El rating de un equipo se basa en el puntaje asignado a cada uno de los partidos disputados por dicho equipo, de acuerdo a la metodología que describimos a continuación.

El puntaje  $P$  asignado a un equipo por disputar un partido cualquiera está dado por la multiplicación de 4 factores:

$$P = M \cdot I \cdot T \cdot C. \quad (3)$$

El factor  $M$  denota los puntos obtenidos en el encuentro: 3 si el equipo gana el partido; 1 en caso de empate; 0 en caso de derrota. El factor  $I$  denota la importancia del encuentro: 4 si es Copa del Mundo; 3 si es fase final de torneo de Confederación (Copa América, Copa de Oro, Eurocopa, etc.) o Copa Confederaciones; 2.5 si es eliminatoria mundialista o en el ámbito de confederación; y 1 si es un partido amistoso o torneo menor de confederación. El factor  $T$

denota la fortaleza del rival la que se calcula como 201 menos el ranking FIFA del oponente (e.g. de tal manera que jugar contra el top 1 del ranking tiene asociado un factor  $F$  igual a 200); en caso que el oponente tenga un ranking inferior al 150, la fuerza del contrincante se fija en un mínimo de 50. Finalmente, el factor  $C$  denota el promedio de la fuerza de las confederaciones a la cuales pertenecen los dos equipos que participan del partido.<sup>1</sup> Por ejemplo, la Tabla 2 muestra los puntos obtenidos por Honduras y Chile en el partido que disputaron en el Mundial Sudáfrica 2010, cuyo resultado fue un triunfo de Chile por 1-0.

Tabla 2: Ejemplo Puntaje FIFA.

	Honduras	v/s	Chile
Posición en ranking	38		22
$M$ : Resultado del partido	0		3
$I$ : Importancia del partido	4		4
$T$ : Fuerza del competidor	178		162
Confederación	CONCACAF		CONMEBOL
Fuerza de la confederación	0.94		0.94
$C$ : Fuerza de confederaciones		$(0.88+1)/2$	
$P = M \cdot I \cdot T \cdot C$	0		1227

Para un equipo  $i$ , definimos  $G_i := \{(j_k, t_k) : k = 1 \dots\}$  el conjunto de partidos disputados por el equipo  $i$ , donde  $j_k$  y  $t_k$  denotan el rival y la fecha (medida en años) del encuentro  $k$ . Adicionalmente, para  $s = 1, \dots, 4$ , definimos  $G_{i,s}(t)$  el conjunto de índices correspondientes a los encuentros disputados por el equipo  $i$  entre los años  $t - s$  y  $t - s + 1$ . Sea  $R_i(t)$  el rating del equipo  $i$  en la fecha  $t$ . De acuerdo al procedimiento actual, este rating se calcula como una suma ponderada de los promedios de puntos obtenidos durante los últimos cuatro años, según la siguiente fórmula:

$$R_i(t) := \sum_{s \in \mathbb{N}} \alpha_s \left( \frac{1}{\max\{|G_{i,s}(t)|, 5\}} \sum_{h \in G_{i,s}(t)} P_{i,h} \right), \tag{4}$$

donde  $P_{i,h}$  denota el puntaje obtenido por el equipo  $i$  en el encuentro  $h$ . El procedimiento actual deprecia los puntajes obtenidos en el pasado de acuerdo al siguiente esquema:  $\alpha_1 = 1; \alpha_2 = 0,5; \alpha_3 = 0,3; \alpha_4 = 0,2$ ; y  $\alpha_s = 0$  para  $s > 4$ .

<sup>1</sup>Este factor de fuerza se calcula de acuerdo al número de victorias que han obtenido los equipos de cada Confederación en las últimas tres ediciones de la World Cup. Los valores de fuerza antes del mundial Brasil 2014 fueron 1 para CONMEBOL y UEFA, 0.88 para CONCACAF, 0.86 para AFC y CAF, y 0.85 para OFC. Ver [10] para más detalles.

El ranking FIFA ordena mensualmente a los 211 países asociados, otorgando el ranking número 1 al equipo que posea el mayor rating, y así sucesivamente a los posteriores. El principal uso del ranking es determinar a los sembrados para el sorteo de la fase de grupos con la cual comienza la fase final de la Copa del Mundo: los 7 mejores rankeados más el país anfitrión son sembrados en los 8 grupos con los que comienza la copa. Este hecho les otorga cierta ventaja, dado que las chances de pasar la primera fase o fase de grupos se ven incrementadas dado que evitan partidos difíciles en la ronda inicial.

### 3.1. Deficiencias del ranking actual

Si bien la metodología actual permite rankear a los equipos de acuerdo a un cálculo relativamente sencillo, esta posee serias deficiencias que discutimos a continuación.

**Partidos Amistosos.** El ranking de un equipo, dado en (4), es una suma ponderada de los promedios de los puntos obtenidos en los últimos 4 años por un equipo. Considerando la baja importancia asociada a los partidos amistosos ( $I = 1$ ), es fácil notar que a los países con buen rating no les conviene disputar este tipo de partidos. Ilustramos este punto con el caso de Chile, que en Agosto de 2014 tenía un promedio en el último año de 673,61 puntos (ver Tabla 3). Jugar un partido amistoso contra el equipo rankeado top 1 y vencerlo otorga

Tabla 3: Rating de Chile en Agosto de 2014.

Año	Promedio anual	$\alpha_t$	Contribución
2014	673,61	1	673,61
2013	391,47	0,5	195,73
2012	469,7	0,3	140,91
2011	447,07	0,2	89,41
$R_{aug,2014}$			1099,66

el máximo puntaje posible para estos encuentros, equivalente a 600 puntos ( $3 \cdot 1 \cdot 200 \cdot 1 = 600$ ), que es inferior al promedio de Chile a dicha fecha. Por lo tanto, incluso un triunfo en un partido amistoso contra el mejor equipo en ese momento disminuiría su rating y, a su vez, lo haría probablemente retroceder en el ranking mundial. Este punto quedó en evidencia en la selección de sembrados para el Mundial 2014: La Tabla 4 muestra los diez equipos que encabezaban el ranking FIFA en Octubre 2013, cuando se seleccionaron los sembrados.

Tres de los sembrados (Colombia, Bélgica y Suiza) lograron su ranking, en parte, por haber jugado menos amistosos durante el año 2013 en comparación a los otros equipos que estaban disputando esas plazas. A Octubre de 2013,

Tabla 4: Ranking FIFA Octubre 2013.

Posición	País	Puntos	Posición	País	Puntos
1	España	1513	6	Uruguay	1164
2	Alemania	1311	7	Suiza	1138
3	Argentina	1266	8	Holanda	1136
4	Colombia	1178	9	Italia	1136
5	Bélgica	1175	10	Inglaterra	1080

Colombia y Suiza habían jugado sólo 3 partidos amistosos en el último año y Bélgica lo había hecho en 4 ocasiones, mientras que Holanda e Inglaterra habían disputado 5 partidos de esta categoría e Italia lo había hecho en 6 oportunidades. Un análisis via simulación revela que equipos con menor rating hubiesen logrado ser sembrados en la copa del mundo de no haber jugado partidos amistosos el año previo al sorteo (por ejemplo, Chile, que ocupaba el puesto 12 al momento del sorteo, hubiese sido segundo en el ranking).

**Depreciación de puntaje y calendarización.** Un segundo problema con la metodología actual es que la depreciación anual de puntajes favorece a algunas confederaciones, debido a la calendarización de sus torneos de confederación. De lo explicado anteriormente, es claro que los equipos reciben mayor puntaje desde las copas de confederación (Copa América, Copa de Oro, Eurocopa, Copa de Asia, Copa Africana, Copa de las Naciones de la OFC), sin embargo, estas no se realizan en las mismas fechas, por lo tanto se deprecian de manera distinta: los países pertenecientes a confederaciones que juegan su torneo más cerca del mundial siguiente se ven beneficiados. Por ejemplo, equipos europeos son beneficiados respecto a países sudamericanos dado que su torneo continental se juega sólo 2 años antes del mundial siguiente, mientras que el de los sudamericanos se realiza 3 años antes (en la antesala al Mundial 2014, la Copa América se jugó en 2011 y la Eurocopa en 2012).

**Puntos por resultado.** En primer lugar, si bien el cálculo de  $M$  es acorde al número habitual de puntos obtenidos por partido en un torneo de fútbol (3 en caso de triunfo, 1 en caso de empate, y 0 en caso de derrota), postulamos que esto impone una diferencia desmedida en puntaje obtenido en un partido si se empata con los mejores del mundo en comparación a si se obtiene un triunfo ante equipos de medianía de tabla, en igualdad de formato de partido (ambos amistosos, o eliminatoria, o mundial). Por ejemplo empatar con el top 1 del ranking en un partido amistoso entrega como máximo 200 puntos, mientras que ganarle a un equipo ubicado en la posición 100 reporta como mínimo 255 puntos (estos puntajes dependen de las confederaciones que disputen el partido). Más allá de lo discutible de la diferencia entre los méritos de ambos

casos, lo que parece contradictorio entonces es que sea eventualmente más conveniente para un equipo disputar un partido contra un equipo mediocre en vez del honor de enfrentar al mejor del ranking.

**Localía y Confederación.** Además de las deficiencias sobre los factores actualmente considerados por el ranking FIFA, nos parece importante analizar la relevancia de otros factores que no son actualmente considerados. Un factor relevante en este sentido es el asociado a la condición de localía. De acuerdo a la fórmula utilizada actualmente, el puntaje asignado a un equipo en un partido es independiente de si el equipo jugó de local o visita. Sin embargo, existe consenso respecto a que los equipos mejoran su desempeño cuando juegan de local. Este *efecto de localía* ha sido reconocido en la literatura como un factor clave para explicar resultados deportivos. Particularmente en el fútbol, [5] relacionan esta ventaja de localía a la presión ejercida sobre los árbitros y el aliento de los aficionados, mientras [16] atribuye esta ventaja al mayor conocimiento del campo de juego por parte del equipo local. Otro factor a considerar es la influencia de la confederación en el cálculo de puntos para el ranking. En nuestra opinión, su relevancia no es clara a priori. Adicionalmente nos parece importante extender el concepto de localía mas allá del país de origen de un equipo y llevarlo al ámbito regional. Esto es motivado por los resultados en los torneos más importantes a nivel de selecciones adultas: en 13 de los 15 mundiales disputados en territorio americano o europeo, el campeón fue un equipo del mismo continente que el país anfitrión.

### 3.2. La condición de localía

Para estimar la relevancia de este factor relativo al ranking FIFA, y para sacar prescripciones que puedan transformarse en recomendaciones de mejoras concretas, en esta sección analizamos el poder predictivo de la condición de localía. Considerando los factores que se toman en cuenta en el procedimiento de ranking, procedemos a estimar directamente la probabilidad de un resultado en función de la diferencia en el ranking y el factor de localía. A diferencia de la sección anterior, en esta sección utilizamos una regresión logística multinomial (ver [7] para un enfoque alternativo), remitimos al lector a [1] para más detalles e interpretación de la regresión logística.

Sea  $Y_{i,k} \in \{\text{ganar, empatar, perder}\}$  el resultado del encuentro  $k$  para el equipo  $i$ . Nuestro modelo base, al cual nos referimos como MRLn, plantea que

$$P(Y_{i,k} = x) = \begin{cases} \frac{e^{V_{i,k}}}{1+e^{V_{i,k}}+e^{U_{i,k}}} & x = \text{victoria} \\ \frac{1}{1+e^{V_{i,k}}+e^{U_{i,k}}} & x = \text{empate} \\ \frac{e^{U_{i,k}}}{1+e^{V_{i,k}}+e^{U_{i,k}}} & x = \text{derrota,} \end{cases} \quad (5)$$

donde definimos

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{local} \cdot 1_{\{i \text{ es local}\}} + \beta_{visita} \cdot 1_{\{i \text{ es visita}\}} \quad (6)$$

$$U_{i,k} = \beta_0 + \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{local} \cdot 1_{\{j_k \text{ es local}\}} + \beta_{visita} \cdot 1_{\{j_k \text{ es visita}\}} \quad (7)$$

Adicionalmente, consideramos las siguientes especificaciones para las variables  $V_{i,k}$  (la modificación respectiva a  $U_{i,k}$  sigue la misma lógica).

- Modelo sin factor de localía (MR).

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)).$$

- Modelo con localía de confederación (MRLnLc):

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{local} \cdot 1_{\{i \text{ es local}\}} + \beta_{visita} \cdot 1_{\{i \text{ es visita}\}} + \beta_c \cdot C_{i,k},$$

donde la variable binaria  $C_{i,k}$  toma el valor 1 si y solo si el partido  $k$  del equipo  $i$  se juega en el continente al que pertenece  $i$  e  $i$  no es local.

- Modelo con factor de confederación (MRLnFc):

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{local} \cdot 1_{\{i \text{ es local}\}} + \beta_{visita} \cdot 1_{\{i \text{ es visita}\}} + \beta_{AFC} * AFC_i + \beta_{OFC} * OFC_i + \beta_{CAF} * CAF_i + \beta_{CNMBL} * CNMBL_i + \beta_{UEFA} * UEFA_i.$$

Aquí, las variables binarias  $AFC_i$ ,  $CAF_i$ ,  $OFC_i$ ,  $CNMBL_i$ ,  $UEFA_i$  indican si el equipo pertenece o no a la respectiva confederación (se omite la confederación CONCACAF, que tomamos como confederación de referencia).

El primero de estos modelos se considera para tener un punto de referencia respecto al poder predictivo del ranking FIFA por sí solo. El segundo de los modelos se considera para establecer la influencia de la localía a nivel de confederación, mientras que el tercero busca medir la influencia de considerar variables asociadas a la confederación de los equipos (recordar que el procedimiento actual de ranking incluye dicho tipo de variable).

Para estimar los modelos, recolectamos datos sobre la totalidad de partidos jugados entre los años 2005 y Octubre del 2013, entre selecciones adultas de fútbol asociadas a alguna de las seis confederaciones continentales que son miembros de FIFA. La base de datos resultante incluye 8049 partidos, cuya caracterización se presenta en la Tabla 5.

Los modelos de regresión logística fueron implementados en el software *R Studio*. La estimación de los parámetros fue realizada de acuerdo al método de máxima verosimilitud, que intenta encontrar los  $k + 1$  parámetros para los cuales la probabilidad de observar los datos sea la más alta posible, donde  $k$

Tabla 5: Partidos de Fútbol de Selecciones nacionales adultas 2005 - Oct 2013.

	Amistosos	Clasificatorias	Fase Final de Confederación	Torneo Copa del Mundo	Total
Con localía	2716	3079	189	10	5994
Cancha neutral	1170	282	485	118	2055
Total	3886	3361	674	128	8049

Tabla 6: Resultados modelos de regresión.

Variable	MR		MRLn		MRLnLc		MRLnFc	
	Coficiente	Error	Coficiente	Error	Coficiente	Error	Coficiente	Error
$\beta_0$	0,3071	0,0274	0,3019	0,0519	0,1478	0,0895	0,1478	0,0895
$\beta_r$	0,0131	0,0003	0,014	0,0003	0,014	0,0003	0,014	0,0003
$\beta_{local}$			0,3463	0,0614	0,5005	0,0954	0,3601	0,0954
$\beta_{visita}$			-0,5058	0,0645	-0,3517	0,0974	-0,467	0,0974
$\beta_c$					0,2049	0,0975		
$\beta_{AFC}$							0,0611	0,06
$\beta_{OFC}$							0,477	0,2012
$\beta_{CNMBL}$							0,3136	0,1021
$\beta_{UEFA}$							-0,1454	0,0631
$\beta_{CAF}$							0,1705	0,0633

es la cantidad de características de un equipo al momento del partido. Los resultados se muestran en la Tabla 6.

En los resultados del modelo MR, el signo positivo obtenido para el parámetro  $\beta_r$  indica que a un equipo con mejor ranking se le asocia una mayor probabilidad de ganar que el equipo adversario. Por otra parte, el signo y valor del intercepto ( $\beta_0$ ) indica que ante partidos con equipos de similar ranking, es más probable que el resultado sea un triunfo para alguno de los equipos a que el resultado sea un empate. Adicionalmente, al comparar ambos estimadores se puede apreciar que las probabilidades de empatar o ganar de un equipo se igualan ( $V_{i,k} = 0$ ) cuando el rival está  $\frac{\beta_0}{\beta_r} \approx 23$  unidades arriba en el ranking.

Los resultados para modelo MRLn permiten inferir que las tres variables de este modelo son estadísticamente significativas (debido a que el intervalo de confianza de todos los parámetros a un 95 % de confianza no incluyen el valor 0). En cuanto a la variable  $\beta_{local}$ , se puede ver que su signo es positivo, en línea con los resultados anteriores en la literatura que han mostrado la importancia de este factor. De la misma forma, vemos que la variable  $\beta_{visita}$  toma un valor negativo, e incluso en módulo mayor que la variable  $\beta_{local}$ . Esto no es significativo si consideramos que en cada partido que no se juega en terreno neutral, hay un equipo local y uno visita. En este sentido, lo importante acá es

la magnitud de la diferencia entre estas variables. Al respecto, podemos usar dicha diferencia para estimar a cuánta diferencia de ranking equivale el factor localía. Esta relación se puede obtener resolviendo la siguiente ecuación:

$$\beta_{local} - \beta_{visita} = 2 * x * \beta_r \Rightarrow x = 30,4. \quad (8)$$

que se deduce igualando la ecuación de (6) para ambos rivales.

Este resultado indica que los partidos más parejos dentro del ámbito de selecciones adultas se producen cuando el equipo local tiene una fuerza de 30 unidades más baja en comparación al equipo visitante.

Comparamos los distintos modelos descritos arriba utilizando el *Criterio de Información Bayesiano* (BIC).<sup>2</sup> Nuestros resultados se presentan en la Tabla 7, donde se puede apreciar que el modelo MRLn obtiene un mejor desempeño (menor BIC) que el modelo MR.

Tabla 7: Resumen Resultados Modelos propuestos.

Modelo	Log Verosimilitud	N° Param.	AIC	BIC
MR	-7.704	2	15.412	15.426
MRLn	-7.409	4	14.828	14.856
MRLnLc	-7.407	5	14.825	14.860
MRLnFc	-7.391	9	14.800	14.862

Los resultados de la estimación del modelo MRLnLc, muestran que la variable  $\beta_c$  es una variable significativa y posee signo positivo, por lo cual un equipo que juega un partido en su continente tiene mayor probabilidad de ganar que su equipo contrincante, si este contrincante es de otro continente. Sin embargo, notamos que el efecto de localía continental es considerablemente menor que el efecto de localía nacional. Además, si bien la variable  $\beta_c$  es significativa, la Tabla 7 muestra que el aumento de log verosimilitud del modelo MRLnLc no compensa la agregación de un parámetro extra por sobre el modelo MRLn, por lo que el modelo MRLn continúa siendo el que mejor ajusta a los datos. Los resultados de la estimación del modelo MRLnFc muestran que la  $\beta_{AFC}$  no es significativa, es decir, estadísticamente no es distinta de 0 por lo tanto es igual a la variable omitida, que en este caso es  $\beta_{CONCACAF}$ , por lo cual países de ambas confederaciones se comportan de manera similar. Por otro lado países de la CONMEBOL, OFC y CAF presentan un mejor desempeño

<sup>2</sup>Este índice está dado por  $BIC = -2 \cdot \ln(L) + k \cdot \ln(n)$ , donde  $L$  es la verosimilitud asociada con el MLE,  $k$  el número de parámetros estimados por el modelo, y  $n$  el número de observaciones. El valor del BIC aumenta ya sea por la cantidad de variables o por la varianza no explicada de la variable dependiente, para más detalles ver, por ejemplo, [11].

que CONCACAF, mientras que países de la UEFA tienen peores resultados <sup>3</sup>. El BIC de MRLnFc es mayor que el de los dos modelos anteriores.

En conclusión, el mejor modelo parece ser el MRLn, en cuanto a que es el que mejor explica los datos con el mínimo número de parámetros. Al ser la importancia del factor de localía una variable tan determinante en el resultado de un partido, esperaríamos entonces que fuera un factor tomado en cuenta al calcular el puntaje que un partido contribuye al rating de un equipo.

---

## 4. Propuestas de mejora y resultados

---

El ranking ideal construido en la Sección 2 lo utilizamos en este trabajo sólo como referencia. Consideramos que el ranking a ser implementado en la práctica por FIFA debe ser construido a través de una fórmula sencilla, tal cual se hace actualmente, de modo que pueda ser fácilmente interpretado por los aficionados al fútbol.

### 4.1. Propuestas de Mejora al Procedimiento de Ranking

A partir de las deficiencias develadas y los resultados obtenidos por los modelos en la sección anterior, procedemos a generar distintas propuestas de mejoras.

**P1: Incluir localía y omitir confederación.** Los resultados obtenidos en la sección anterior dan cuenta de la alta importancia de la variable localía nacional y de la relativamente baja importancia de la variable confederación. Intentando capturar esto proponemos eliminar la variable  $C$  de la fórmula de puntaje actual (3) y ajustar el factor de fuerza del rival de acuerdo a la prescripción dada por (8). Esto es,  $P_{i,k} = M_{i,k} \cdot I_{i,k} \cdot T_{i,k}$ , donde

$$T_{i,k} = \begin{cases} \max\{(186 - R_{j_k,t_k}, 50)\} & \text{si } i \text{ juega de local} \\ \max\{(216 - R_{j_k,t_k}, 50)\} & \text{si } i \text{ juega de visita} \\ \max\{(201 - R_{j_k,t_k}, 50)\} & \text{si } i \text{ juega en estadio neutral} \end{cases} \quad (9)$$

---

<sup>3</sup>Si bien algunos países pertenecientes a la UEFA han tenido grandes logros en torneos de mayor envergadura (como Alemania e Italia), la confederación tiene en total 55 países, entre los cuales se encuentran países que consistentemente han ocupado los últimos escalafones del ranking FIFA (tales como Malta, San Marino y Liechtenstein) lo cual ayuda a explicar los resultados

Esta fórmula impone una penalización respecto a la fórmula actual equivalente a 30 puestos del ranking, cuando un equipo juega de local.

**P2: Omitir partidos amistosos.** Como alternativa para mitigar lo perjudicial de los partidos amistosos, proponemos no considerar los partidos amistosos para efecto de cálculo de ratings. Esto elimina los incentivos a no jugar partidos amistosos (los que son claramente importantes, dado que permiten probar nuevos jugadores y experimentar nuevas estrategias). En términos concretos proponemos redefinir  $G_{i,s}(t)$  en (4) como el conjunto de partidos *no amistosos* jugados por un equipo entre las fechas  $t - s$  y  $t - s + 1$ , para todo  $s$ .

**P3: Modificación en los puntos por victoria.** Para mitigar las grandes diferencias en puntaje obtenido en un partido si se empata con los mejores del mundo en comparación a si se obtiene un triunfo ante equipos de mediano o bajo ranking, proponemos otorgar sólo dos puntos por victoria en vez de tres en la fórmula (9). De esta manera, un empate con equipos del top ten del ranking otorga una cantidad de puntos similar o incluso superior a derrotar a equipos de ranking mediocre. En términos concretos, proponemos utilizar

$$M_{i,k} = \begin{cases} 2 & \text{si } Y_{i,k} = \text{victoria} \\ 1 & \text{si } Y_{i,k} = \text{empate} \\ 0 & \sim \end{cases} \quad (10)$$

Notamos que esta propuesta se encuentra respaldada de forma parcial por los resultados en [3] los que muestran que bajo dicha asignación de puntajes, el ranking esperado al final del torneo será el mismo que el obtenido aplicando una variación al modelo probabilístico de [2], el cual deriva rankings mediante comparaciones de a pares (muy similar a regresiones logísticas)

**P4: No depreciar puntos.** Para corregir las deficiencias causadas por el desajuste de fechas en la calendarización de los torneos importantes de las distintas confederaciones, proponemos omitir la depreciación de puntaje que actualmente utiliza el ranking FIFA. En términos concretos proponemos utilizar

$$\alpha_s = \begin{cases} 1 & s \in \{1, 2, 3, 4\} \\ 0 & \sim \end{cases} \quad (11)$$

Al no depreciar los puntajes por año, el puntaje de un equipo en cualquier momento reflejará su desempeño en los últimos cuatro años, abarcando

todos los torneos de confederaciones posibles en los que un país puede participar.

## 4.2. Resultados

En nuestros experimentos, testeamos las propuestas de la sección anterior en forma incremental. Denotamos a la propuesta **P1-X** aquella que incluye las propuestas **P1** hasta la **PX**, inclusive. La Tabla 8 muestra los 16 primeros puestos de los rankings propuestos y el ranking FIFA, todos ellos en Octubre del año 2013 (fecha utilizada por la FIFA para sembrar los cabezas de serie del mundial más reciente).

Tabla 8: Rankings obtenidos mediante las cuatro propuestas analizadas más el ranking FIFA en Octubre de 2013.

	<b>P1-1</b>	<b>P1-2</b>	<b>P1-3</b>	<b>P1-4</b>	FIFA
1	España	Brasil	Brasil	Brasil	España
2	Argentina	España	España	España	Alemania
3	Alemania	Argentina	Argentina	Argentina	Argentina
4	Uruguay	Alemania	Alemania	Alemania	Colombia
5	Colombia	Estados Unidos	Chile	Uruguay	Bélgica
6	Costa de Marfil	Chile	Italia	Holanda	Uruguay
7	Italia	Holanda	Uruguay	Chile	Suiza
8	Estados Unidos	Uruguay	Estados Unidos	Costa de Marfil	Holanda
9	Brasil	Italia	Holanda	Italia	Italia
10	Bélgica	Costa de Marfil	Colombia	Japón	Inglaterra
11	Holanda	Colombia	Costa de Marfil	Estados Unidos	Brasil
12	Ghana	Bélgica	Bélgica	Inglaterra	Chile
13	Chile	Ghana	Inglaterra	Colombia	Estados Unidos
14	México	Japón	Ghana	Corea del Sur	Portugal
15	Suiza	México	Francia	Grecia	Grecia
16	Inglaterra	Francia	Japón	México	Bosnia

Es obviamente imposible aunar opiniones para dilucidar cual entre estas cuatro propuestas y la que actualmente usa la FIFA provee un *mejor* ranking. Una medida objetiva de comparación es la distancia de cada uno de estos rankings respecto al ranking ideal de referencia desarrollado en la Sección . Como medida de distancia utilizamos el error cuadrático medio (MSE) respecto al ranking de referencia:

$$MSE_j = \frac{1}{N} \sum_{i \in N} \left( R_i^j(\text{oct}, 2013) - R_i^*(\text{oct}, 2013) \right)^2, \quad (12)$$

donde  $MSE_j$  representa el error cuadrático medio asociado a la propuesta **P1-j** y  $R_i^j(\cdot)$  denota el ranking del equipo  $i$  bajo la propuesta **P1-j**, (el superíndice

\* denota el ranking de referencia), y  $N$  es la cantidad de países incluidos en el modelo. Mientras menor sea el MSE del ranking, mejor la propuesta.

La Tabla 9 muestra la comparación de las propuestas en términos de sus MSEs, los cuales revelan que la cuarta propuesta es la de mejor desempeño. Le sigue el ranking FIFA, con un error cuadrático medio 15.6% mayor que nuestra mejor propuesta. Las restantes propuestas muestran peor desempeño, especialmente la número 1, que se aleja en 41.9% del ajuste de la número 4. Concluimos entonces que, si bien el ranking FIFA presenta un aceptable desempeño, es posible mejorar dicho ranking utilizando el conjunto de nuestras propuestas.

Tabla 9: Desempeño del ranking FIFA y de las propuestas.

Propuesta	MSE
Ranking FIFA (Octubre 2013)	66,25
<b>P1</b> (localía incluida y confederación omitida)	81,31
<b>P1-2</b> (+ partidos amistosos omitidos)	72,87
<b>P1-3</b> (+ modificación en los puntos por victoria)	69,87
<b>P1-4</b> (+ sin depreciación de puntaje)	57,31

---

## 5. Formación de grupos en sorteo de la Copa del Mundo

---

El principal uso del ranking FIFA es determinar a los sembrados para el sorteo de la fase de grupos con la cual comienza la fase final de la Copa del Mundo. Es importante recordar que en esta etapa, los equipos sólo se enfrentan a equipos del mismo grupo. Analizamos a continuación el proceso de formación de estos grupos, y proponemos un esquema más balanceado mediante una formulación de programación entera.

### 5.1. Deficiencias del sistema actual de sorteo

Bajo el sistema actual, los grupos del Mundial a los cuales los países son sorteados se forman mediante una elección aleatoria. Los 32 equipos participantes en la Copa Mundial son divididos en 4 *pots*. El pot 1 contiene a los equipos que serán sembrados (o cabezas de serie) de los grupos: el país organizador y los siete de mejor ranking entre los restantes equipos. Los otros tres

pots contienen equipos de acuerdo a zonas geográficas. Para el Mundial Brasil 2014, el pot 2 contuvo a los equipos sudamericanos no sembrados y a los equipos africanos (7 en total), el pot 3 a los equipos norteamericanos y asiáticos (8 en total), y el pot 4 a los equipos europeos no sembrados (9 en total). Para que todos los pots contuvieran ocho equipos, uno de los equipos del pot 4 fue asignado mediante sorteo al pot 2. La Tabla 10 muestra la configuración de los pots resultantes.

Tabla 10: Pots de equipos y ranking FIFA en la antesala al Mundial 2014 (entre paréntesis el ranking relativo, asignándole a Brasil el ranking relativo 1 por ser el país anfitrión).

Pot 1	Pot 2	Pot 3	Pot 4
11 Brasil (1)	9 Italia (10)	13 USA (13)	8 Holanda (9)
1 España (2)	12 Chile (12)	24 México (23)	10 Inglaterra (11)
2 Alemania (3)	17 Costa de Marfil (17)	31 Costa Rica (24)	14 Portugal (14)
3 Argentina (4)	22 Ecuador (21)	34 Honduras (27)	15 Grecia (15)
4 Colombia (5)	23 Ghana (22)	44 Japón (28)	16 Bosnia (16)
5 Bélgica (6)	32 Argelia (25)	49 Irán (29)	18 Croacia (18)
6 Uruguay (7)	33 Nigeria (26)	56 Corea (30)	19 Rusia (19)
7 Suiza (8)	59 Camerún (32)	57 Australia (31)	21 Francia (20)

Luego, los grupos son formados sorteando la asignación de los equipos, de tal manera que cada uno de los ocho grupos contiene exactamente un equipo de cada pot. Existe una restricción adicional que impide que 3 equipos europeos y 2 equipos sudamericanos formen parte del mismo grupo. Si bien este sistema promueve la diversidad geográfica dentro de cada grupo, perjudica a algunos equipos, específicamente a aquellos con mejor ranking dentro de su propio pot que con alta probabilidad serán situados en grupos más difíciles. Por ejemplo, en 2014 Estados Unidos se vio altamente afectado con este sistema ya que compartió el pot 3 con equipos mucho más débiles, con los cuales estaba impedido de enfrentarse en la fase de grupos. Además Estados Unidos estaba obligado a enfrentar a un cabeza de serie y a otro equipo europeo del pot 4 (cuyos ranking varían entre 8 y 21). Otro país perjudicado fue Italia, el sorteado para pasar del pot 4 al 2, que en consecuencia quedó compartiendo pot con países de ranking muy inferior, como lo son Argelia, Nigeria y Camerún. La configuración final de grupos del Mundial 2014 se muestra en la Tabla 11. El número que antecede al país es su Ranking FIFA respectivo en el mes de Octubre del 2013, mes en el cual se determinaron los cabezas de serie. Entre paréntesis aparece el ranking relativo entre las 32 selecciones mundialistas. Brasil por ser el país anfitrión obtiene el primer lugar.

Como medida de dificultad de un grupo, utilizamos la suma de los rankings

Tabla 11: Cuadro Final Mundial Brasil 2014.

Grupo	Equipos			
A	11 Brasil (1)	18 Croacia (18)	24 México (23)	59 Camerún (32)
B	1 España (2)	8 Holanda (9)	12 Chile (12)	57 Australia (31)
C	4 Colombia (5)	15 Grecia (15)	17 Costa de Marfil (17)	44 Japón (28)
D	6 Uruguay (7)	31 Costa Rica (24)	10 Inglaterra (11)	9 Italia (10)
E	7 Suiza (8)	22 Ecuador (21)	21 Francia (20)	34 Honduras (27)
F	3 Argentina (4)	16 Bosnia (16)	49 Irán (29)	33 Nigeria (26)
G	2 Alemania (3)	14 Portugal (14)	23 Ghana (22)	13 USA (13)
H	5 Bélgica (6)	32 Argelia (25)	19 Rusia (19)	56 Corea (30)

de los equipos que lo conforman. Cuando menor es la suma, mayor es el nivel de dificultad. La Tabla 12 muestra esta medida para los grupos del Mundial 2014, revelando un gran desbalance entre los diferentes grupos. De acuerdo a la suma de rankings, los grupos B, D y G son los de mayor dificultad, mientras que los grupo H, A, E y F son los más débiles. Notar que Italia y Estados Unidos quedaron en los grupos más fuertes (Italia finalmente quedó eliminado en primera ronda, mientras que Estados Unidos clasificó segundo en su grupo).

Tabla 12: Suma de Ranking relativo y Ranking FIFA para los grupos del Mundial 2014.

Grupo	A	B	C	D	E	F	G	H
Suma Rank. Rel.	74	54	65	52	76	75	52	80
Suma Rank. FIFA	112	78	80	56	84	101	52	112

## 5.2. Modelo de Programación Entera para balancear los grupos

Nuestra propuesta es intentar corregir el desbalance en la dificultad de los distintos grupos, mediante un modelo de programación lineal entera. La formulación del modelo se presenta a continuación. En dicho modelo, la variable de decisión es en qué grupo se asigna cada país participante. Mas allá de las restricciones lógicas (todos los grupos tienen 4 países, los países son asignados a un solo grupo, etc), el modelo restringe el número de países de una misma confederación asignados a un grupo, y considera como función objetivo el minimizar la diferencia entre la máxima y mínima suma de rankings de los integrantes de un grupo.

### 5.2.1. Formulación matemática

#### Conjuntos

$G$ : grupos.

$C$ : confederaciones.

$I$ : equipos.

$S$ : equipos sembrados ( $S \subset I$ ).

$J_c$ : equipos de la confederación  $c$  ( $J_c \subset I, c \in C$ ).

#### Parámetros

$R_i$ : ranking del equipo  $i$  ( $i \in I$ ).

$M_c$ : máximo número de equipos de la confederación  $c$  que pueden pertenecer a un mismo grupo ( $c \in C$ ).

#### Variables de Decisión

$$x_{ig} = \begin{cases} 1 & \text{si el equipo } i \text{ es asignado al grupo } g \\ 0 & \text{si no} \end{cases}$$

$w_{min}$ : menor de las sumas de los rankings de los equipos dentro de un mismo grupo.

$w_{max}$ : mayor de las sumas de los rankings de los equipos dentro de un mismo grupo.

#### Restricciones

$$\sum_{g \in G} x_{ig} = 1 \quad \forall i \in I \quad (13)$$

$$\sum_{i \in I} x_{ig} = 4 \quad \forall g \in G \quad (14)$$

$$\sum_{i \in S} x_{ig} = 1 \quad \forall g \in G \quad (15)$$

$$\sum_{i \in J_c} x_{ig} \leq M_c \quad \forall g \in G, c \in C \quad (16)$$

$$w_{min} \leq \sum_{i \in I} R_i x_{ig} \quad (17)$$

$$w_{max} \geq \sum_{i \in I} R_i x_{ig} \quad (18)$$

$$x_{ig} \in \{0,1\}, w_{min} \geq 0, w_{max} \geq 0 \quad \forall i \in I, g \in G \tag{19}$$

### Función Objetivo

$$\text{mín } f = w_{max} - w_{min} \tag{20}$$

Las restricciones (13) aseguran que todo equipo sea asignado a exactamente un grupo. Las restricciones (14) establecen que cada grupo contiene a 4 equipos, mientras que las restricciones (15) aseguran que uno de estos equipos sea sembrado. Las restricciones (16) imponen la cota superior en el número de equipos de una misma confederación que pueden ser asignados a un mismo grupo (el Mundial 2014, se permitió a lo más un equipo de cada confederación, excepto la europea para la cual se permitieron dos). Las restricciones (17)-(18) ayudan a calcular las sumas mínima y máxima de los rankings de los equipos dentro de un mismo grupo, mientras que la función objetivo (20) busca minimizar la diferencia entre estas dos sumas.

### 5.3. Resultados

El modelo fue implementado en AMPL/CPLEX y su resolución tomó unos pocos segundos para todas las instancias consideradas. Las tablas 13 y 14 muestran grupos obtenidos utilizando el ranking FIFA y el ranking relativo, respectivamente.

Tabla 13: Grupos obtenidos por el modelo utilizando el ranking FIFA.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8
Alemania	España	Suiza	Uruguay	Bélgica	Brasil	Colombia	Argentina
Chile	Irán	Ecuador	Honduras	Australia	México	Corea	Japón
Camerún	Costa de Marfil	Argelia	Ghana	USA	Nigeria	Italia	Costa Rica
Inglaterra	Bosnia	Francia	Rusia	Holanda	Grecia	Portugal	Croacia
<b>Fuerza</b>	<b>83</b>	<b>83</b>	<b>82</b>	<b>83</b>	<b>83</b>	<b>83</b>	<b>83</b>

Tabla 14: Grupos obtenidos por el modelo utilizando el ranking relativo.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8
Brasil	Uruguay	Argentina	Suiza	España	Colombia	Bélgica	Alemania
Camerún	Corea	Australia	Chile	Irán	Costa Rica	Ecuador	Rusia
Italia	Francia	Bosnia	Japón	USA	Nigeria	Argelia	Honduras
México	Holanda	Grecia	Croacia	Ghana	Inglaterra	Portugal	Costa de Marfil
<b>Fuerza</b>	<b>66</b>	<b>66</b>	<b>66</b>	<b>66</b>	<b>66</b>	<b>66</b>	<b>66</b>

En ambos casos, el nivel de dificultad de los grupos obtenidos es considerablemente más parejo que los grupos obtenidos por el sorteo FIFA. Esto queda claramente reflejado en las figuras 1 y 2, que grafican la suma del ranking FIFA y relativo de los equipos en cada grupo, respectivamente. Usando el ranking FIFA, el modelo obtiene un valor objetivo óptimo  $f^* = 1$ , mientras que en los grupos sorteados en la realidad la diferencia entre los grupos de mayor y menor suma de rankings fue de 66. Usando el ranking relativo, el modelo obtiene un valor objetivo óptimo  $f^* = 0$ , mientras que en los grupos sorteados en la realidad la diferencia entre los grupos de mayor y menor suma de rankings fue de 28.

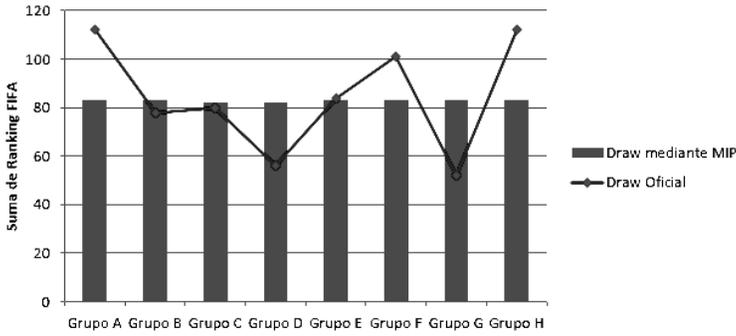


Figura 1: Suma del ranking de los equipos en cada grupo obtenidos por el modelo usando el ranking FIFA.

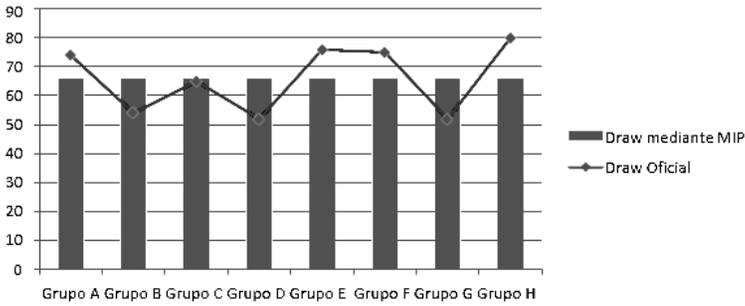


Figura 2: Suma del ranking de los equipos en cada grupo obtenidos por el modelo usando el ranking relativo.

---

## 6. Conclusiones

---

Este artículo ha revelado falencias importantes en la metodología actual del ranking FIFA, y genera a su vez, un conjunto de propuestas de mejoras basadas en los resultados de modelos matemáticos desarrollados para este estudio y aplicados a datos de partidos internacionales. Primero, nuestro estudio empírico sugiere que el puntaje entregado en un partido no refleja de buena manera cuán meritorio fue para cada equipo obtenerlo. Si bien la metodología de FIFA entrega mayor puntaje por jugar contra equipos con mejor ranking, no hace referencia alguna a otras variables tan relevantes, como lo es por ejemplo el factor localía. Por otro lado, el ranking actual permite a países sacar ventaja evitando jugar partidos amistosos en el año previo a la designación de los cabezas de series del Mundial. Esto debido principalmente a la escasa entrega de puntos de este tipo de partidos en comparación con cualquier otro tipo de formato, combinado con la fórmula de promediar los puntajes de los juegos disputados. Nuestro trabajo empírico muestra como este factor favoreció a países como Bélgica, Suiza y Colombia para ser sembrados en el Mundial 2014. Otro factor que estudiamos es el de la depreciación del promedio de puntos de los últimos 4 años. Este sistema genera disparidad entre países de distintos continentes debido a que no todos juegan su torneo continental en el mismo año, por ende, ante igualdad de resultados en los demás torneos, es probable que los equipos de confederaciones que jueguen sus torneos más cerca del próximo mundial se vean beneficiados.

Mediante modelos de regresión logística multinomial, y el uso intensivo de datos, analizamos la importancia de distintas variables en el resultado de partidos de fútbol entre selecciones. Los resultados muestran lo fundamental que es la variable localía, ya que entrega una gran ventaja al equipo que posea dicho atributo. Este factor acompañado con la diferencia de Ranking entre equipos fueron los de mejor ajuste a la data histórica. De nuestros resultados, el efecto de jugar de local es comparable con tener una diferencia de ranking de 30 puestos a favor. Otros factores, como la pertenencia a una determinada confederación y la localía continental tienen mucho menor poder predictivo.

Con el fin de mejorar el ranking FIFA actual, generamos propuestas que modifican su metodología. Estas modificaciones consisten en incluir la variable localía, eliminar la variable confederación, omitir partidos amistosos, reestructurar los puntos por partido ganado y cambiar el sistema de depreciación de puntaje por año. Para testear el desempeño de estas modificaciones y com-

pararlas con la metodología actual, utilizamos el criterio de error cuadrático medio y un ranking referencial creado en base a simulaciones. Los resultados muestran que la incorporación de estas modificaciones en conjunto mejoran considerablemente el desempeño del ranking FIFA, cuyo error cuadrático medio fue 15,6% mayor que el de nuestra mejor propuesta.

Finalmente, mediante un modelo de programación lineal entera mixta, logramos generar grupos que mantienen los criterios geográficos de la FIFA y a la vez son más parejos en nivel de dificultad. Aplicando el modelo a los datos del Mundial 2014 y utilizando como medida de dificultad de un grupo la suma de los rankings de los equipos que lo componen, logramos formar grupos mucho más balanceados que los grupos sorteados por la FIFA.

## Referencias

- [1] S. Anderson, A. De Palma, y J. Thisse. *Discrete choice theory of product differentiation*. MIT press, 1992.
- [2] R. Bradley y M. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.
- [3] R. Davidson. On extending the bradley-terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329):317–328, 1970.
- [4] M. Dixon y M. Robinson. A birth process model for association football matches. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):523–538, 1998.
- [5] P. Downward y M. Jones. Effects of crowd size on referee decisions: Analysis of the fa cup. *Journal of sports sciences*, 25(14):1541–1545, 2007.
- [6] G. Durán, M. Guajardo, y D. Sauré. Scheduling the south american qualifiers to the 2018 fifa world cup by integer programming. *European Journal of Operational Research*, 262(3):1109–1115, 2017.
- [7] D. Dyte, S. Clarke, et al. A ratings based poisson model for world cup soccer simulation. *Journal of the Operational Research society*, 51(8):993–998, 2000.
- [8] FIFA. 2014 fifa world cup brazil technical report and statistics. Technical report, 2014.

- [9] FIFA. 2014 fifa world cup brazil television audience report. Technical report, 2014.
- [10] FIFA. Fifa ranking, marzo 2017.
- [11] J. Friedman, T. Hastie, y R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.
- [12] J. Guyon. Rethinking the fifa world cup final draw. *Journal of Quantitative Analysis in Sports*, 11(3):169–182, 2015.
- [13] J. Lasek, Z. Szlávik, y S. Bhulai. The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, 1(1):27–46, 2013.
- [14] M. Maher. Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118, 1982.
- [15] I. McHale, S. Davies, et al. Statistical analysis of the effectiveness of the fifa world rankings. *Statistical thinking in sports*, páginas 77–90, 2007.
- [16] R. Pollard. Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, 4(3):237–248, 1986.
- [17] H. Rue y O. Salvesen. Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):399–418, 2000.
- [18] P. Scarf y M. Yusof. A numerical study of tournament structure and seeding policy for the soccer world cup finals. *Statistica Neerlandica*, 65(1):43–57, 2011.
- [19] A. Suzuki, L. Salasar, J. Leite, y F. Louzada-Neto. A bayesian approach for predicting match outcomes: the 2006 (association) football world cup. *Journal of the Operational Research Society*, 61(10):1530–1539, 2010.

