



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

EVENTOS SECUENCIALES Y SU EFECTO EN EL FÚTBOL: EVIDENCIA EMPÍRICA
EN PREDICCIONES Y CALENDARIZACIÓN DE CAMPEONATOS

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

MATÍAS RUBÉN MOLINA CORTEZ

PROFESOR GUÍA:
DANIEL SCHWARTZ PERLROTH

MIEMBROS DE LA COMISIÓN:
DENIS SAURÉ VALENZUELA
GUILLERMO DURÁN

Este trabajo ha sido parcialmente financiado por Corporación ISCI

SANTIAGO DE CHILE
2017

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: MATÍAS RUBÉN MOLINA CORTEZ
FECHA: 31 DE MAYO 2017
PROF. GUÍA: DANIEL SCHWARTZ PERLROTH

EVENTOS SECUENCIALES Y SU EFECTO EN EL FÚTBOL: EVIDENCIA EMPÍRICA EN PREDICCIONES Y CALENDARIZACIÓN DE CAMPEONATOS

Durante los últimos 20 años se ha observado un incremento en el interés por aplicar *analytics* (*análisis de datos*) en la toma de decisiones en deporte. Un ejemplo reciente es la planificación del *fixture* para las Eliminatorias al Mundial de Rusia 2018 en la CONMEBOL en la que se utilizaron modelos de optimización. En la presente memoria se analizan dos casos a la toma de decisiones vinculadas al fútbol: (1) predicciones de resultados si se cuenta con información secuencial y (2) orden de partidos en el *fixture* de un torneo y su impacto en la posición final de un equipo.

Es por ello que se realizan dos experimentos, en el primero, a través de una plataforma online de *crowdsourcing*, se invita por medio de incentivos a un grupo de personas a que participen en el pronóstico de partidos para la Eurocopa realizada en Francia. Se contrastan dos grupos: uno solo tiene una instancia para hacer las predicciones de todos los partidos a considerar, mientras que el otro grupo debe efectuar predicciones todos los días. Por lo tanto, en este escenario, se examina cómo la presencia o ausencia de decisiones secuenciales afecta la certeza en predicciones de partidos de fútbol.

En el segundo experimento, se aprovecha el orden aleatorio de la secuencia de partidos que se da en ciertos campeonatos de fútbol, con el objetivo de medir el impacto de la dificultad del equipo rival a comienzos del torneo (*experimento natural*). Se utiliza los datos recogidos de los partidos de la Liga Inglesa de fútbol en un período de 10 años (2005-2015). A través de técnicas econométricas se evalúa el efecto de la dificultad inicial, controlando por factores inherentes a cada equipo y temporada.

En término de conclusiones, en el primer experimento se encuentra que no existen diferencias significativas entre los dos grupos, aun cuando un grupo tenía mayor información al momento de hacer predicciones. En el segundo experimento, se infiere que la dificultad de los primeros partidos de una temporada, medida a partir de las posiciones de los equipos rivales en temporadas pasadas, afecta significativamente la posición final de un equipo, es decir, los equipos que enfrentan mayor dificultad en estos partidos tienen más chances de ver afectada negativamente su posición final. Además, al analizar la heterogeneidad de este efecto, no se encuentran diferencias dependiendo si la dificultad inicial del rival es asignada a mejores o peores equipos, según sus posiciones en temporadas pasadas.

Finalmente, para investigaciones futuras se propone expandir el estudio con muestras de mayor tamaño, para el primer experimento, y al mismo tiempo, buscar formas de evitar sesgos de selección entre grupos experimentales. En el segundo experimento se sugiere incluir más temporadas y realizar comparaciones con otras ligas de fútbol.

A la gente del sur

Agradecimientos

Quiero agradecer a todos quienes han sido parte en mi formación, sobre todo a aquellos que han estado conmigo en mi formación como persona, proceso que nunca termina y en el cual espero seguir teniendo tan buena compañía como la que he tenido hasta ahora. Quiero agradecer también a todas y cada una de las personas que de una u otra forma me han hecho ser quien soy.

También quiero agradecer a mi familia, en especial a mi mamá, papá y hermana. Ya que ellos siempre, pero siempre han estado ahí a pesar de la distancia. Dar gracias por el apoyo incondicional brindado por ellos, no importaba que decisión fuera ellos siempre me apoyaron. A mi mamá por siempre creer en mí y darme su cariño, a mi papá por siempre empatizar conmigo y darme sus consejos sabios y a mi hermana por creer en mí.

Agradecer también a mis abuelos, tíos, primos por estar ahí disponible ante cualquier inconveniente y ayudarme en lo que necesitaba.

Quiero agradecer al Instituto Alemán de Puerto Montt por su formación y las posibilidades que me otorgó de conocer Alemania. A Thomas por ser uno de mis mejores amigos, a la Cata, a la Lele, y la Min por su amistad. También al Diaxo, Krebs, Neumann, Martin, a Juanfra y Álvaro por el buen tiempo que pasamos en el colegio.

Quiero agradecer a mi segunda familia, la Burschenschaft Andinia, por ser los primeros que me acogieron al llegar a Santiago, donde pude hacer tantas amistades, pasarla bien y también lograr desarrollarme como persona y mantener el idioma alemán, que tanto me gusta. En especial a Juanjo por enseñarme con cada consejo y experiencia de vida, a Jorge por ser mi primer amigo en la Andinia, a Guille por apoyarme, a Christian por su amistad y compañía para ver los partidos de la U, a Stefan, Juancris, Felipe, Rainer, Boni por creer en mí cuando me tocó ser Sprecher, a cada miembro de la institución por hacerme parte de esta gran hermandad. También agradecer a la Kity por su buena onda y confianza en mis aptitudes.

También agradecer a mis amigos en Alemania por su acogida, a Simon, Thilo, Wichmann, a Lucho, a Niko Beller, a Benni y a muchos más que me hicieron vivir una experiencia increíble en tierras germanas.

Quiero agradecer a mi primer grupo de amigos de la U, a la Lore, Pablo y la Cami, por esta larga amistad que hemos logrado y por el apoyo que siempre me brindaron.

Quiero agradecer también a la Comunidad Internacional de Beauchef, en especial a Juan, a Pablo, Michelle y Viviana por el buen grupo que se formó y la motivación con la cual pudimos fundar esta comunidad muy importante para la Facultad.

Quiero agradecer a la Feria Empresarial por los grandes momentos que viví los cuales me permitieron aportar con un granito de arena en la formación de profesionales en esta Universidad. En especial a los que me acogieron en el grupo TI, a Cristián, al Andrés y al Vixon. Por su puesto también agradecer a Marta y a Maceta por el grupo que se formó con el cual pudimos darle un sello al área. Además de agradecer a mi Directorio 2015, a Pablo, a Lorca, Cortines, Coni y al Andrés por esas reuniones semanales infaltables. Por último, al área de Logística 2016, a la Feña, la Anto, la Kari, el Feño y a Leyton por esos con pocas horas de sueño para poder levantar la Feria y por esa alegría de siempre.

Quiero agradecer a mis compañeros de departamento (Encomenderos), a Lalo y Jiménez por estar en la última parte de mi carrera universitaria siempre con buena onda y siendo muy buenos amigos, escuchando mis problemas y mis éxitos. Quiero agradecer a mi profesor guía Daniel Schwartz, por la dedicación, paciencia, disposición y confianza que me entregó. Asimismo, de agradecer a Andrés por ayudarme en mi trabajo a realizar el experimento, la cual significo gran parte de mi memoria y un aprendizaje enorme para mí. Por su puesto al profesor Denis y profesor Guillermo por sus comentarios y feedback que me permitieron elaborar un mejor trabajo.

Quiero agradecer en especial a Loreto por apoyarme, acompañarme hasta las tantas de la noche en la oficina y sacar adelante este trabajo de memoria, a Lorca por siempre creer en mí y ayudarme cuando lo necesitaba, y Hans por confiar en mis aptitudes por el apoyo sobre todo en este último año. Gracias por apoyarme incondicionalmente y por supuesto a la Sofí que me ayudo en la redacción de la memoria, de lo contrario no lo hubiera logrado. Quiero agradecer a mis profesores de la sección de Marketing por sus críticas que me permitieron ir aprendiendo y sus consejos para sacar adelante este trabajo.

Finalmente quiero agradecer a la gente de la oficina 501, Renzo, Caro, Loreto y Pablo por su ayuda.

Tabla de Contenido

1. Introducción	1
1.1. Antecedentes Generales	1
1.2. Contexto de la Investigación	2
1.3. Toma de decisiones en apuestas del fútbol	3
1.4. Planteamiento del Problema y Justificación	4
1.5. Objetivos	6
1.5.1. Objetivo General	6
1.5.2. Objetivos Específicos	6
1.6. Alcances	6
1.7. Resultados Esperados	7
1.8. Marco Conceptual	7
1.8.1. Diseño de experimentos	8
1.8.2. Experimento Natural	8
1.8.3. Attrition	8
1.8.4. Políticas closed-loop	9
1.8.5. Análisis de poder estadístico	9
2. Experimento online para decisiones anticipadas y secuenciales	11
2.1. Metodología	11
2.1.1. Encuesta online	11
2.2. Desarrollo Metodológico	16
2.2.1. Descripción de la muestra	16
2.2.2. Análisis de Resultados	20
2.2.3. Propuesta de soluciones Attrition	25
2.2.4. Regresiones y análisis de Heterogeneidad	28
2.2.5. Conclusiones y Aprendizajes	32
3. Experimento Natural para el orden de la secuencia de eventos	34
3.1. Metodología	34
3.2. Desarrollo Metodológico	38
3.2.1. Descripción de la muestra	38
3.2.2. Propuesta de Modelos y Análisis de Resultados	46
3.2.3. Heterogeneidad	52
3.2.4. Conclusiones y Aprendizajes	58
4. Conclusiones Generales y Trabajos Futuros	60

4.1. Conclusiones Generales	60
4.2. Trabajos Futuros	61
Bibliografía	62
A. Antecedentes Generales	65
A.1. Contexto de la Investigación	65
B. Metodología	66
B.1. Cuestionario online	66
B.2. Cuotas de los sitios de apuestas online	79
B.3. Caracterización del attrition del experimento online	84
B.4. Modelo alternativo para el attrition	85
B.5. Descripción de las 10 temporadas	85
B.6. Tipos de tabla de análisis	87
B.7. Descripción muestra experimento natural	91
C. Desarrollo Metodológico	92
C.1. Resultados de los modelos propuestos para análisis de la Liga Inglesa	92
D. Regresión con heterogeneidad considerando 4 temporadas anteriores	111

Índice de Tablas

2.1. Partidos de la Eurocopa por día	14
2.2. Cantidad de participantes por cada grupo según nivel de fanatismo	17
2.3. Cantidad de participantes según fanatismo, género y grupo	18
2.4. Cantidad de participantes según cantidad de partidos a ver, género y grupo .	19
2.5. Cantidad de partidos por día	20
2.6. Promedio de aciertos acumulados por grupo	21
2.7. P-valores (%) según diferencia de attrition de los participantes	21
2.8. Promedios de aciertos considerando 9 días	22
2.9. Promedio de aciertos considerando 6 días	22
2.10. Simulación de ambos grupos según casa de apuesta <i>Bet 365</i> por día	23
2.11. Resultados y pronósticos según casa de apuesta <i>Bet 365</i> para el día 7 del experimento	23
2.12. Predicciones según <i>Bwin</i> por día y grupo	24
2.13. Comparación de medias entre el algoritmo y resultados del experimento para los primeros 6 días.	24
2.14. Comparación de medias entre el algoritmo y resultados del experimento para los primeros 9 días.	24
2.15. Resultados de la regresión para 6 días considerando solo la variable indepen- diente grupo	25
2.16. Resultados de la regresión para 9 días considerando solo la variable indepen- diente grupo	25
2.17. Contraste de predicción del grupo One Shot con y sin attrition	26
2.18. Contraste del Grupo One Shot sin attrition y asumiendo que los missing values del Grupo day by Day son todos aciertos	26
2.19. Contraste del Grupo One Shot sin attrition y asumiendo que los missing values del Grupo day by Day son todos erróneos	26
2.20. Cantidad de participantes que abandonaron el estudio del grupo Day by Day	27
2.21. Aciertos según contraste del Grupo One Shot sin attrition y asumiendo solu- ción del modelo 2	27
2.22. Cantidad de fanáticos según grupo experimental	28
2.23. Predicciones promedio por fanatismo y grupos	29
2.24. Resultados de la regresión considerando heterogeneidad del fanatismo para 6 días	29
2.25. Resultados de la regresión considerando heterogeneidad del fanatismo para 9 días	29

2.26. Resultados de la regresión considerando heterogeneidad del fanatismo (de forma continua) para 6 días	30
2.27. Resultados de la regresión considerando heterogeneidad del fanatismo (de forma continua) para 9 días	30
2.28. Resultados de la regresión considerando heterogeneidad educación para 6 días	30
2.29. Resultados de la regresión considerando heterogeneidad educación para 9 días	31
2.30. Resultados de la regresión considerando heterogeneidad “cantidad partidos a ver” para 6 días	31
2.31. Resultados de la regresión considerando heterogeneidad “cantidad partidos a ver” para 9 días	31
3.1. Información cruzada por cantidad de temporadas jugadas por equipo.	39
3.2. Goles en promedio según localía y tiempo del partido	40
3.3. Valores de estimadores considerando 3 temporadas anteriores comparando variables discretas con variables continuas	46
3.4. Cuadro comparativo que considera 4 temporadas anteriores, variables discretas con variables continuas	47
3.5. Valores de estimadores respecto a distinta cantidad de temporadas	48
3.6. Estimadores según variable dependiente posición a la fecha numero 20 con sus efectos fijos	48
3.7. Estimadores según variable dependiente posición a la fecha numero 20 según desempeño últimas 4 temporadas	49
3.8. Estimadores según variable dependiente puntaje a la fecha numero 20 según desempeño temporada anterior y dos temporadas anteriores	49
3.9. Estimadores según variable dependiente puntaje a la fecha numero 20 según desempeño 4 y 5 temporadas anteriores	49
3.10. Estimadores según variable dependiente posición al final de la temporada, según desempeño de local y de visita	50
3.11. Estimadores según variable dependiente posición al final de la temporada, según desempeño de local y de visita, solo considerando 1 temporada anterior	51
3.12. Estimadores según dificultad de manera discreta considerando los 6 últimos equipos de la tabla acumulada (considerando las 2 temporadas anteriores) . .	51
3.13. Estimadores según dificultad de manera discreta considerando los 6 últimos equipos de la tabla acumulada (considerando las 2 temporadas anteriores) . .	52
3.14. Resultados de los estimadores de la regresión, con heterogeneidad fija (temporadas 2000 a 2005)	53
3.15. Resultados de los estimadores de la regresión con dificultad discreta considerando heterogeneidad para 7 equipos	55
3.16. Resultados de los estimadores de la regresión con dificultad discreta considerando heterogeneidad para 4 equipos (Big Four)	56
3.17. Resultados de los estimadores de la regresión con dificultad discreta considerando heterogeneidad dinámica	57
B.1. Cuotas de los partidos del sitio Bet365 hasta el 4 de Junio	80
B.2. Cuotas de los partidos del sitio Bet365 del 5 de Junio hasta el 9 de Junio . .	81
B.3. Cuotas de los partidos del sitio Bwin hasta el 4 de Junio	82
B.4. Cuotas de los partidos del sitio Bwin del 5 de Junio hasta el 9 de Junio . . .	83

B.5. Promedio de predicciones segun modelo alternativo	85
B.6. Resultados según las 10 temporadas	86
B.7. Ejemplo de clasificación final equipos Liga Inglesa temporada 2005/2006 . .	88
B.8. Ejemplo de clasificación según partidos de local equipos Liga Inglesa tempo- rada 2005/2006	89
B.9. Ejemplo de clasificación según partidos de visita equipos Liga Inglesa tempo- rada 2005/2006	90
C.1. Resultados de los estimadores con efectos fijos de la regresión considerando la temporada anterior, con dificultad continua	93
C.2. Resultados de los estimadores con efectos fijos de la regresión considerando la temporada anterior, con dificultad discreta	94
C.3. Resultados de los estimadores con efectos fijos de la regresión considerando dos temporadas anteriores, con dificultad continua	95
C.4. Resultados de los estimadores con efectos fijos de la regresión considerando dos temporadas anteriores, con dificultad discreta	96
C.5. Resultados de los estimadores con efectos fijos de la regresión considerando tres temporadas anteriores, con dificultad continua	97
C.6. Resultados de los estimadores con efectos fijos de la regresión considerando tres temporadas anteriores, con dificultad discreto	98
C.7. Resultados de los estimadores con efectos fijos de la regresión considerando cuatro temporadas anteriores, con dificultad continua	99
C.8. Resultados de los estimadores con efectos fijos de la regresión considerando cuatro temporadas anteriores, con dificultad discreto	100
C.9. Resultados de los estimadores con efectos fijos de la regresión considerando cinco temporadas anteriores, con dificultad continuo	101
C.10. Resultados de los estimadores con efectos fijos de la regresión considerando cinco temporadas anteriores, con dificultad discreto	102
C.11. Resultados de los estimadores con efectos fijos de la regresión considerando 1 temporada anterior según localía, con dificultad continua	103
C.12. Resultados de los estimadores con efectos fijos de la regresión considerando 1 temporada anterior según localía, con dificultad discreta (5 primeros equipos)	104
C.13. Resultados de los estimadores con efectos fijos de la regresión considerando 2 temporadas anteriores según localía, con dificultad continua	105
C.14. Resultados de los estimadores con efectos fijos de la regresión considerando 2 temporadas anteriores según localía, con dificultad discreta (5 primeros equipos)	106
C.15. Resultados de los estimadores con efectos fijos de la regresión considerando 1 temporada anterior, con dificultad discreta (5 últimos equipos)	107
C.16. Resultados de los estimadores con efectos fijos de la regresión considerando 1 temporada anterior según localía, con dificultad discreta (5 últimos equipos)	108
C.17. Resultados de los estimadores con efectos fijos de la regresión considerando 5 temporadas fijas, con dificultad continua	109
C.18. Resultados de los estimadores con efectos fijos de la regresión considerando 5 temporadas fijas, con dificultad discreta	110

D.1. Resultados de los estimadores de la regresión con dificultad discreta considerando heterogeneidad dinámica (según 4 temporadas anteriores)	111
---	-----

Índice de Ilustraciones

1.1. Ejemplo de cuotas en una casa de apuesta	3
2.1. Análisis de poder estadístico a través de <i>GPower</i>	12
2.2. Diagrama del funcionamiento del experimento	14
2.3. Histograma según edad de los participantes	16
2.4. Histograma de participantes según nivel de fanatismo por género	17
2.5. Histograma de participantes según nivel de fanatismo por grupo	18
2.6. Participación por día de la gente por grupo	19
2.7. Aciertos por día y grupo	20
2.8. Acierto acumulado por día según grupo	21
2.9. Promedio de predicciones según modelo 2	28
3.1. Varianza entre posición final y posición en fecha determinada	40
3.2. Promedio dificultad según cada temporada, usando la dificultad promedio de los primeros 7 partidos	41
3.3. Cambio de posición según dificultad de primeros 7 partidos para equipos que han participado las 10 temporadas	42
3.4. Cambio de posición según dificultad por temporada de primeros equipos	43
3.5. Cambio de posición según dificultad por temporada para el segundo grupo.	43
3.6. Tendencia de los equipos que participaron 3 o 4 temporadas.	44
3.7. Tendencia de los equipos que participaron 1 o 2 temporadas.	45
3.8. Contraste del análisis entre variables continuas y discretas	47
3.9. Posición final de la temporada según dificultad discreta por tipo de equipo	54
3.10. Contraste del análisis entre variables continuas y discretas con heterogeneidad	55
3.11. Tendencia de equipos, dependiendo si pertenecen al grupo Big Four o no	56
3.12. Tendencia de equipos según heterogeneidad dinámica propuesta	58
A.1. Reparto de Ingresos de la temporada 2015/2016 - Premier League	65
B.1. Proporción de gente que abandonó experimento por género	84
B.2. Proporción de gente que abandonó experimento según educación	84
B.3. Proporción de gente que abandonó experimento según fanatismo	85
D.1. Tendencia de los equipos según heterogeneidad dinámicas considerando 4 temporadas anteriores	112

Capítulo 1

Introducción

1.1. Antecedentes Generales

Durante el último período se ha observado un incremento en el interés por aplicar *analytics* (*análisis de datos*) a la toma de decisiones en deportes [1]. Indicio de esto es la inclusión de proyectos relacionados al uso de *analytics* en variados deportes, como el tenis, el golf y el básquetbol. Un ejemplo de este interés por aplicar matemática y estadísticas en los deportes lo otorga el *Journal of Sports Sciences* que, desde 1983, viene publicando investigaciones [2] que proveen a atletas y entrenadores importantes elementos de cada deporte, tales como el poder mejorar rendimiento usando técnicas como regresiones lineales múltiples, las regresiones logísticas, entre otros.

El fútbol también ha sido considerado dentro de estos análisis. Por ejemplo, se ha examinado el efecto de localía para las rondas finales de la UEFA Champions League [3], y también se ha estructurado vía optimización la conformación de grupos para las Copas del Mundo con el fin de balancear los grupos según el ranking de los equipos [4].

Por otro lado, también se han hecho estudios para entender de qué manera pueden tener ventaja los equipos al enfrentar un partido. A través de los estudios de [5] y [6] se comprobó que más de la mitad de los encuentros de la Liga Inglesa eran ganados por los equipos que jugaron de local. Esto último puede deberse a componentes externos como el de jugar en un estadio lleno [7], el conocimiento del terreno de juego [8], y/o el viaje que tienen que realizar los equipos [9].

Además, otras características que han sido objeto de estudio son: cómo las estrategias (i.e. la manera de juego que propone el entrenador) de los equipos pueden influir en el resultado [10], y que un equipo juegue adicionalmente un campeonato internacional afecte el rendimiento en el torneo local [11], entre otros.

Otro análisis matemático dentro del contexto del fútbol ha sido la calendarización de los partidos [12]. En particular, se han utilizado métodos de programación entera, uso de heurísticas o la utilización de *constraint programming* (programación con restricciones). La

finalidad de estos métodos es contar con una propuesta de calendarización con la finalidad de que sea un torneo atractivo. Algunos factores a considerar serán que los clásicos se juegan en instancias finales del torneo y que los equipos no desembolsen grandes sumas de dinero por concepto de viajes para jugar en otras ciudades, como se muestra en el trabajo [13] realizado para el torneo de la liga chilena.

Dentro de esta área de aplicación, destaca el hecho que en la lista de finalistas al prestigioso premio *Franz Edelman* [14], se encontraba un proyecto sobre la calendarización del fútbol chileno, el cual fue gestionado por académicos del Departamento de Ingeniería Civil Industrial de la Universidad de Chile [15]. Dicho estudio provocó el interés de la CONMEBOL para usar el método [16] en las clasificatorias rumbo a la Copa del Mundo Rusia 2018. Sin embargo, un dato no menor es el bajo número de federaciones de fútbol que utilizan programación matemática para confeccionar la calendarización de partidos (v.g. Bélgica, Brasil y Noruega).

En el presente trabajo de título se estudia si existe otro componente que afecte el resultado final de la Liga Inglesa, en específico si el orden en el *fixture* altera el resultado final (por ejemplo, la posición en la tabla o el puntaje obtenido al final de la temporada). Otra arista presente en esta memoria son las decisiones secuenciales en apuestas de fútbol aplicado en un estudio online.

1.2. Contexto de la Investigación

El fútbol hoy en día es el deporte que tiene más seguidores ("*fans*") en el mundo. Un ejemplo de esto último es que en la final del torneo más importante de clubes, i.e. la UEFA Champions League en Lisboa a finales de Mayo del 2016, más de 400 millones de personas apreciaron el partido entre Real Madrid y Atlético de Madrid [17]. Más aún, el fútbol es el deporte que registra mayores movimientos de dinero, cuyas tres principales fuentes son: (i) *merchandising* y *sponsorio*, (ii) ingresos por derechos de transmisión de los partidos e (iii) ingresos relacionados con el día del partido, como la venta de entradas y abonos por temporada [18].

Cabe destacar que las 5 ligas más grandes del mundo (inglesa, alemana, española, italiana y francesa) juntas superan los €15 Millones [19] durante la temporada 2015/2016. De estas, la liga más fuerte en términos económicos es la Premier League de Inglaterra, con una facturación anual de 5.300 millones de dólares. Le siguen la Bundesliga de Alemania, la Liga de España, la serie A de Italia y la Liga 1 de Francia. Pero la distancia entre estas ligas y la Premier League es exorbitante. De hecho, la competición de Alemania (que es la segunda en importancia) tuvo ingresos por 3.100 millones de dólares. Es más, la liga de Inglaterra supera la suma de ingresos que se obtendría de España e Italia conjuntamente. Y otro dato interesante de destacar es que en el último tiempo estas cifras han aumentado a tal punto que ahora 6 clubes de la liga inglesa superan las ganancias de los 20 equipos de la temporada 1991/1992. Por ende, y a partir de estas cifras, la Premier League es considerada la liga más importante en términos económicos. En el Anexo A.1 se observa cómo fueron distribuidos los dineros en la liga inglesa de la temporada 2015/2016.

Además, en el informe de Deloitte del año 2015 se menciona cuáles son los 10 equipos del mundo que mueven más dinero. El primer lugar, por décimo año consecutivo, lo ocupa el Real Madrid con una facturación anual de 750 millones de dólares. Luego, aparece el Manchester United; con 700 millones; en tercer lugar, está el Bayern Munich con 660 millones. El Barcelona ocupa el cuarto lugar con una cifra muy similar y la quinta posición es para el Paris Saint-Germain con una facturación de 645 millones de dólares. El resto del Top-10 lo completan el Manchester City, Chelsea, Arsenal, Liverpool y Juventus.

Año a año los equipos invierten grandes sumas de dinero para poder competir y lograr sus respectivos objetivos tanto en el ámbito local como también internacional. Ahora bien, esta práctica involucra a todos los equipos, incluso a aquellos que durante su historia han tenido escasos triunfos, pues les es fundamental mantener sus categorías.

1.3. Toma de decisiones en apuestas del fútbol

Los resultados de los eventos deportivos tienen como característica principal que siempre existe incertidumbre respecto al resultado final. Dado esta situación es que han surgido los sitios de apuestas. Estos elaboran cuotas para cada posible resultado: empate, gana local o gana visita. La menor de las cuotas asociada al resultado significa que es más probable que se otorgue uno de los 3 resultados anteriormente mencionados. Y estas cuotas usan información pasada para saber cuál de los 3 eventos es más probable, como se muestra en la figura 1.1, un ejemplo de los primeros partidos de la Euro 2016.

Fri 10 Jun		1	X	2
15:00	France v Romania	1.28	4.75	13.00
Sat 11 Jun		1	X	2
09:00	Albania v Switzerland	5.50	3.50	1.66
12:00	Wales v Slovakia	2.62	3.00	2.87
15:00	England v Russia	1.85	3.30	4.50

Figura 1.1: Ejemplo de cuotas en una casa de apuesta

Fuente: Sitio de apuesta Bet365

Es tanta la participación de la gente en este tipo de juegos que en España se movieron €2.627 millones en sitios de apuestas entre los meses de Enero y Junio del 2016 (donde solo se consideraban eventos deportivos). De estas, un 70% corresponden a apuestas de fútbol. Como dato del dinero que mueven los sitios de apuestas es que los 20 equipos de la Premier están patrocinados por una casa de apuestas [20].

Por estas razones, es difícil que las personas que apuestan en eventos deportivos logren acertar en la mayoría de los casos, debido a la incertidumbre antes mencionada. Además, en este ámbito pueden haber sesgos que perjudiquen la decisión final. Por ejemplo: que esté jugando el equipo favorito, claramente condiciona la decisión por querer que gane ese equipo. Pensar que, si se ganan cierta cantidad de partidos consecutivos, el equipo va a lograr ganar

el campeonato. Por lo que pronosticar de manera rentable es una actividad al alcance de un grupo reducido, debido a las desventajas del margen, los medios y la información disponible.

Desde un punto normativo, la capacidad de acierto en estas predicciones debiese aumentar si se cuenta con mayor información relevante. Sin embargo, el uso de la información puede sesgar las decisiones de un punto descriptivo. Por cual, nacen interrogantes de si (1) es mejor o no apostar según la información que uno dispone, (2) si es apropiado o no agregar información de noticias y/o (3) si es o no conveniente considerar las cuotas de los sitios de apuestas. Respecto a esto último, estudios han demostrado que la capacidad de las personas en sus predicciones son inferiores a una fórmula predictiva, a menos que ocurra un evento muy raro que cambie el resultado, lo cual lo denominan como “regla de la pierna rota” [21]. Estos sucesos pueden ser pocos probables, pero si pueden darse en fútbol.

1.4. Planteamiento del Problema y Justificación

Esta memoria se enfoca en este tipo de decisiones secuenciales aplicadas al fútbol, tanto a decisiones de apuestas como al calendario del *fixture*. En primer lugar, se examina si el “principal” (ente que organiza el calendario del juego, o *fixture*) puede afectar un resultado final (i.e. al concluir la temporada) al alterar la secuencia con que se establece el *fixture* de partidos de una liga. Actualmente, esto no está considerado en la forma de establecer el *fixture* de una liga.

En específico, se pretende examinar si el orden puede provocar que el resultado al final de la temporada sea distinto, es decir, que la posición final del equipo dependa del orden de los partidos. Por ejemplo: jugar con equipos que tienen mejor desempeño que el promedio de los equipos en la liga o jugar con equipos que tengan regular rendimiento en la liga en los primeros partidos. Por ende, dada estas posibles situaciones, surge la interrogante de si enfrentar equipos difíciles al comienzo motiva el rendimiento para al inicio de la temporada con el fin de lograr un mejor resultado y posición al final de la temporada o, en caso contrario, que perjudique el resultado al final por el hecho de iniciar la temporada con malos resultados, desmotivando al equipo y llegando, incluso, a despedir al técnico por los malos resultados. Esto trae un impacto significativo para los equipos sobre todo pensando en aquellos que se encuentran en situaciones límites, por ejemplo, clasificar o no a un torneo internacional o de lograr la permanencia en la serie de honor.

Este factor de dificultad no estaría afectando solo a la posición final de la temporada sino que también, por ejemplo, lo podemos observar en la clasificación a torneos internacionales. El monto de los premios no deja de ser menor, a saber, € 12.7 millones por participar en la UEFA Champions League [22] y € 2.6 millones por competir en Europa League [23], aumentando considerablemente los montos al ir avanzando en la fase de grupos.

Por tanto, construir un marco de mejores decisiones puede optimizar el resultado final v.g: flexibilizar la estrategia de un equipo dependiendo si parte en casa, ajustar las decisiones de inversión cuando hay decisiones secuenciales o tener un campeonato más justo para todos los equipos, si este es uno de los objetivos que se persigue en la calendarización de un torneo.

Es importante recalcar que, para saber si el orden de la calendarización del campeonato afecta en el resultado final de cada equipo, se hará uso de aleatoriedad con que se establecen las secuencias de partidos en una temporada. Es decir, se considerará un experimento natural en el que el tratamiento corresponde a la dificultad del comienzo de la liga.

La aleatoriedad está dada por las reglas para realizar la calendarización, las cuales influyen para todos los equipos por igual y permite que la planificación sea distinta para cada temporada. En este caso, el tratamiento aleatorio elimina potenciales sesgos de selección, y se considera experimento natural ya que factores de variación no corresponden al experimentador [24].

Una segunda temática relacionada a la secuencia de eventos se relaciona a la temporalidad de decisiones de apuestas en el fútbol. A modo general, el *paper* [25] contempla algunos ejemplos sobre cómo cambian las preferencias según la temporalidad con que se separan los eventos. Las decisiones cambian según el nivel de importancia que otorga la gente y lo impaciente que pueden ser las personas por el hecho de obtener resultados lo más rápido posible. Un agente puede variar cuándo y/o en qué orden se toman las decisiones y considerar cómo afectan su resultado final. Sin embargo, la evidencia previa señala que las personas se alejan de las decisiones normativas al momento de enfrentarse a un evento secuencial. Por ejemplo, esto ocurre en el caso que proponen Tversky y Kahnemann en el cual comparan 2 situaciones, donde la única diferencia entre ambas es que en un caso la decisión consiste en 2 etapas y en el otro sólo se considera una [26]. A priori uno pensaría que son las mismas opciones y, por lo tanto, las personas decidirán de igual manera. Pero ocurre todo lo contrario: la gente determina de manera opuesta, pudiendo concluirse, así, que las personas ignoran la primera etapa del juego (fenómeno denominado *isolation effect*). Por ende, las decisiones de este tipo se diferencian con el hecho de considerar a los eventos por separado, sin tomar en cuenta lo anterior o lo que viene o de agrupar los eventos para la toma de decisión.

En particular, en esta memoria se examinan decisiones y resultados de apuestas para partidos de fútbol considerando si la decisión se toma antes de cada evento o antes de "todos" los eventos futuros. Desde un punto de vista normativo, las decisiones consecutivas son preferibles ya que aportan más información a cada decisión subsiguiente. De hecho, el enfoque de programación dinámica [27] aplicado a la toma secuencial de decisiones bajo incertidumbre señala que el uso de políticas *closed-loop* domina al uso de políticas *open-loop*, y que la ventaja del primer conjunto de políticas proviene de su habilidad para ajustar las decisiones en función de la información que se recolecta a través del tiempo (ejemplo: trabajar en un inventario y recibir demanda aleatoria).

Considerando lo anterior, la literatura de toma de decisiones en múltiples áreas de aplicación se ha desarrollado en torno al supuesto de que las personas utilizan políticas *closed-loop*, y que hacen uso eficiente de la información a medida que esta se recolecta, siendo capaces de procesar dicha información y estimar las consecuencias acumulativas o finales que ella trae en el largo plazo. Por otro lado, bibliografía previa ha constatado que las personas sufren de sesgos en la toma de decisiones [26], y que no pueden seguir el enfoque anteriormente expuesto, como es el caso de *feedback* continuo, que podría estar agregando información innecesaria o sesgando la decisión vía otros canales (v.g. las emociones) [28]. Este tipo de decisiones, son objeto de investigación a través de un experimento online, en el cual se les pide a las

personas realizar predicciones compatibles con incentivos para una serie de partidos en un torneo internacional de fútbol, contrastando dos grupos: un grupo haciendo predicciones de forma acumulada y el otro grupo pronosticando partido a partido con la incorporación de *feedback* de los resultados transcurridos.

A pesar de que esta memoria se centra en el contexto del fútbol, sus resultados pueden conducir, en forma más amplia, al entendimiento sobre cómo las personas toman decisiones para eventos consecutivos y sobre el efecto del orden de éstos dependiendo de su dificultad.

1.5. Objetivos

1.5.1. Objetivo General

Determinar el efecto en el resultado final del orden de una secuencia de eventos y el efecto de contar con *feedback* en predicciones para cada evento de la secuencia.

1.5.2. Objetivos Específicos

1. Determinar cómo la acumulación de información y la temporalidad, a través de políticas *open-loop* y *closed-loop*, influyen en la toma de decisiones secuenciales.
2. Determinar si el efecto de añadir *feedback* al grupo de tratamiento difiere según características de los individuos.
3. Contrastar las predicciones entre las personas y las casas de apuestas.
4. Determinar si un comienzo difícil o fácil en la calendarización de un torneo, asignado en forma aleatoria, reduce el rendimiento final de un equipo.
5. Analizar si el efecto del orden de la calendarización afecta de igual manera a todos los equipos.

1.6. Alcances

El trabajo de memoria se limita a 2 experimentos: un experimento online y un experimento natural.

Para el primer caso, se limita al diseño de un experimento online, contenido en una plataforma de *crowdsourcing*, llamada *Prolific*. Este tipo de experimento es compatible con incentivos, aprovechando la coyuntura de eventos deportivos (en este caso, la Eurocopa 2016). Se limita a realizar predicciones durante 10 días y, para ello, las personas interesadas tienen

que declarar que participarán todos los días podrán inscribirse. Ahora bien, este estudio está circunscrito a un número reducido debido al presupuesto con que se cuenta para pagarles por participar. Además, dada las condiciones del experimento, sólo existe validez interna para este estudio, por lo que si se quiere ahondar en el tema se recomienda realizar un experimento de campo para lograr la validez externa.

El otro estudio, que consiste en un experimento natural, se realiza a través de recopilación de datos de 10 temporadas de solo una liga, la Liga Inglesa (desde 2005/2006 hasta 2014/2015) con distintas medidas de dificultad en diferentes intervalos de tiempo. Sin embargo, cabe destacar que este efecto no necesariamente es idéntico a otras ligas europeas.

Finalmente, esta memoria está enfocada a decisiones y eventos ligados al fútbol, por lo que variaciones en los resultados aplicados a otras situaciones o deportes quedan supeditados a nuevos estudios.

1.7. Resultados Esperados

En este proyecto se espera evaluar el diseño del experimento, además de realizar un análisis descriptivo y entender el comportamiento de la gente a través del experimento online. Conjuntamente, se pretende analizar el *attrition* (abandono de participantes) [29] y ver de qué manera se podría disminuir esta tasa de deserción con otras técnicas (por ejemplo: cambiando el incentivo), con el fin de poder realizar un estudio a futuro, disponiendo, obviamente, con más información sobre este tipo de experimentos (en plataformas web).

En cuanto al análisis estadístico, **(i)** se espera saber cuál es la diferencia de acierto de los pronósticos entre los 2 grupos que se contrastan en el torneo de la Eurocopa 2016, **(ii)** en caso de existir un *attrition* significativo, proponer modelos de mejora, **(iii)** agregar heterogeneidad al análisis, y **(iv)** contrastar estos resultados.

Para el caso de la Liga Inglesa, se tiene la expectativa de concluir si es que el orden de los partidos afecta el desempeño al final de la temporada, o si una sucesión de partidos difíciles puede efectivamente perjudicar o favorecer el rendimiento al final de la temporada, además de poder inferir si la forma en que sea realiza el calendario favorece a ciertos equipos o no. Para conseguirlo, se propone agregar heterogeneidad al análisis, con el fin de identificar a qué grupo de equipos les afecta más el orden de los partidos o si afecta a todos por igual. Conjuntamente, se quiere comprobar si cambiando los modelos con distintas variables dependientes, se podría llegar a concluir lo mismo; con el fin de determinar si el orden afecta que un equipo mantenga su categoría o no, o su chance de clasificar a una copa internacional.

1.8. Marco Conceptual

Para contextualizar de mejor manera el presente estudio se introducen los siguientes tópicos:

1.8.1. Diseño de experimentos

Los enfoques experimentales son herramientas muy valiosas pues pueden ser usado para determinar si existe causalidad entre ciertas variables y el comportamiento de las personas. En ese sentido, supera a otros métodos ya que permiten ver más allá de la correlación que pueda existir entre variables [30, 31]. El diseño de experimentos busca testear de manera precisa cómo diversos atributos o factores influyen en un determinado comportamiento, debido a que el investigador crea las condiciones y maneja las variables de interés.

1.8.2. Experimento Natural

Un experimento natural es un estudio empírico en el que las condiciones experimentales se determinan por la naturaleza o por otros factores fuera del control del experimentador. Por lo tanto, los experimentos naturales son los estudios observacionales y no son controlados en el sentido tradicional de un experimento aleatorio. Los experimentos naturales son más útiles cuando se ha producido un cambio claramente definida y grande en el tratamiento a una subpoblación claramente definida (y ningún cambio a una subpoblación comparables), por lo que los cambios en las respuestas pueden ser plausiblemente atribuido al cambio en tratamientos [24].

Experimentos naturales son considerados para los diseños de estudio cada vez que la experimentación controlada es difícil, como en muchos problemas de la epidemiología y la economía [24].

Un factor importante a considerar, que afecta significativamente, es el medio ambiente donde se realizan los experimentos, además de otras influencias que pudiesen existir. En los experimentos de campo se utiliza una situación y se altera la variable independiente, contrastando 2 o más grupos. La desventaja de este tipo de experimentos es que sufren de la posibilidad de contaminación: las condiciones experimentales se pueden controlar con más precisión y certeza en el laboratorio, lo cual no es lo óptimo para entender el comportamiento de las personas.

La importancia de la asignación aleatoria radica en que permite la comparación de sujetos teóricamente idénticos y que han sido sometidos a una única variación.

1.8.3. Attrition

El efecto de *attrition* o la deserción de los participantes en un experimento es un suceso que provoca valores perdidos, induciendo una inferencia sesgada de los datos. Este efecto ocurre en experimentos donde la gente no está dispuesta a participar en una encuesta rutinaria, perdiendo el investigador el rastro de la muestra que ha sido tratada. Aquí, él observa algunos datos de los participantes y el resto serán valores perdidos (*missing values*).

El problema de la existencia de *missing values* es que existen distintas acepciones, to-

das ellas susceptibles de ser aplicadas, dificultándose la labor de llegar a una interpretación correcta. Entonces, dado este problema, el *attrition* obliga a realizar suposiciones sobre los datos perdidos, lo que puede inducir a error por el hecho de que el *attrition* se encuentra relacionado con los posibles resultados del experimento, i.e. el análisis de los datos puede producir estimaciones sesgadas sobre el efecto promedio del tratamiento (*average treatment effect*).

Lo ideal sería que las personas con resultados faltantes tengan, en promedio, los mismos resultados que los datos recolectados.

Otro factor importante del *attrition* es que reduce el tamaño efectivo de la muestra y, por lo tanto, aumenta los errores estándar. Debido a esto último, se busca siempre una muestra más grande cuando existe la posibilidad de este fenómeno.

Uno de los métodos a aplicar, como posible solución, sería el de explorar los peores escenarios, rellenando con valores más extremos posibles.

Otro método para hacer frente al *attrition* sería reunir (recolectar) más datos de los sujetos que faltan.

Pero la mejor solución sería diseñar e implementar estudios que minimizan el desgaste o eliminados por completo.

1.8.4. Políticas closed-loop

Las políticas *closed-loop* es un término usado en la programación dinámica, la cual se caracteriza por ser una forma para tomar decisiones que tiene el fin de minimizar costos. La principal característica es que considera decisiones de varias etapas. En efecto, la decisión incorpora información de los eventos pasados, por lo cual la decisión se retroalimenta del pasado (se considera el hecho de que no existe un costo asociado por cada decisión). Este tipo de política se contrapone con la política *open-loop*, la cual permite tomar solo una decisión la que se define al principio, la cual será afectada en todas las etapas correspondientes sin poder realizar un cambio. Por lo general la gente toma decisiones con el modelo *closed-loop*.

Dentro de las las políticas de closed-loop podemos hacer mención a otros factores tales como los sesgos en la toma de decisiones secuenciales o las emociones, que afectan las decisiones finales [32].

1.8.5. Análisis de poder estadístico

El poder estadístico depende principalmente del criterio de significancia, del tamaño del muestreo y del tamaño del efecto. El cual es usado para determinar el tamaño de la muestra y, evitar cometer el error de tipo II (no rechazar hipótesis nula cuando era falsa) [33].

A través del poder estadístico se puede calcular la muestra mínima que garantice de manera

razonable el poder encontrar un determinado efecto dentro de la población muestral.

Capítulo 2

Experimento online para decisiones anticipadas y secuenciales

2.1. Metodología

2.1.1. Encuesta online

Para examinar el efecto de decisiones secuenciales, se utiliza un experimento compatible con incentivos, aprovechando la coyuntura de eventos deportivos. A saber, se invita a algunas personas a participar de un estudio online a través de la plataforma *crowdsourcing* como *Prolific Academic* (otro similar es *Amazon Mechanical Turk*). A través del análisis de poder estadístico, se calcula el tamaño de la muestra para que garantice de manera razonable encontrar un determinado efecto dentro de la población muestral. En la cual se contrastan dos grupos, asumiendo que los participantes de un grupo tienen un 30% de aciertos en las predicciones y la diferencia entre ellos es de un 14% (valor ajustado por restricciones presupuestarias), considerando además un *attrition* de 7% por cada grupo. Por lo que se estimó una muestra de 400 personas, el análisis se muestra en la Figura 2.1 que se realizó mediante el software *GPower*.

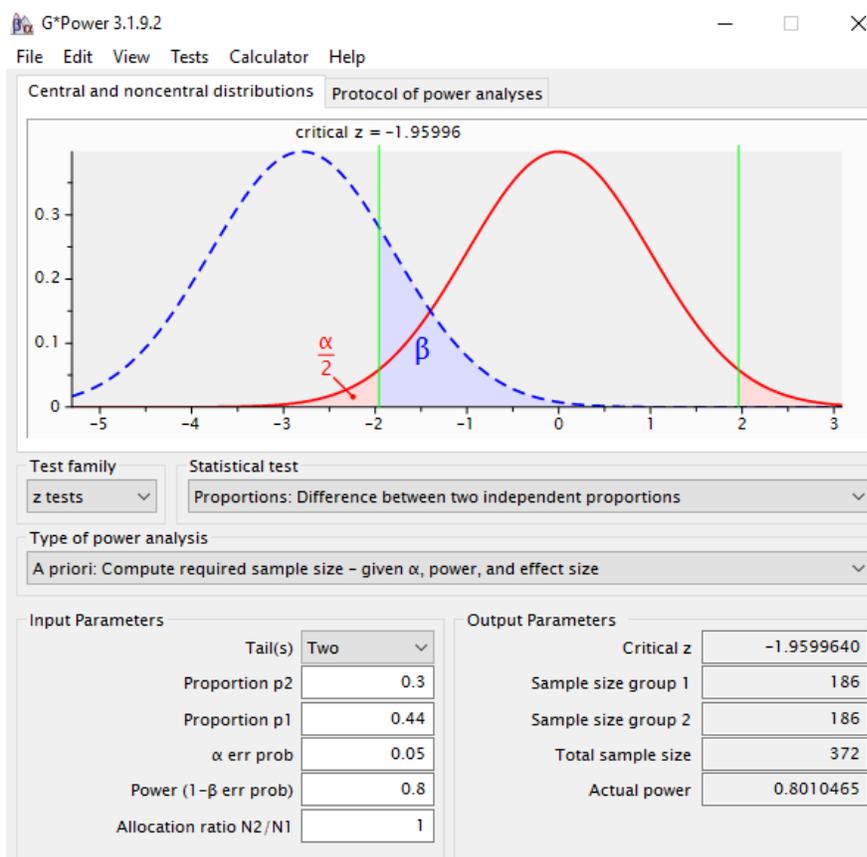


Figura 2.1: Análisis de poder estadístico a través de *GPower*

Fuente: elaboración propia

Este diseño experimental incluirá estímulos económicos para los participantes como una forma de motivarlos a colaborar. Serán premiados con *bonuses* (que la plataforma de *crowdsourcing* permite) cuando sus predicciones se cumplan. Además, aquellos participantes que logren mejor predicción a lo largo de todo el experimento recibirán, además, un *bonus* especial detallado más adelante.

El experimento es llevado a cabo a través de un cuestionario online en inglés desarrollado a partir de la herramienta *Qualtrics* (software para elaboración de encuestas online, con múltiples aplicaciones en investigación).

En la encuesta, los participantes responden en primer lugar preguntas de filtro para saber si están dispuestos o no a participar por 9 días seguidos en el estudio, si son mayores de 18 años y si entendieron todas las instrucciones del estudio. En forma preliminar (al momento de aceptar las condiciones), los participantes reciben £0.50 por inscribirse. Luego, continúan preguntas de caracterización, cuántos partidos están dispuesto a ver con una escala de Likert ((i) “*All matches*”, (ii) “*most of the matches*”, (iii) “*half of the matches*”, (iv) “*only a few matches*”, (v) “*none of them*”), a qué género corresponde, la edad, el nivel educacional, país de residencia, cuántos partidos estiman que iban a ver y qué tan fanático se consideraban del fútbol con una escala de Likert ((i) “*Dislike a great deal*”, (ii) “*Dislike a moderate amount*”, (iii) “*Dislike a little*”, (iv) “*Neither like nor dislike*”, (v) “*Like a little*”, (vi) “*Like a moderate*

amount”, (vii) “*Like a great deal*”). El cuestionario utilizado se encuentra disponible en la sección B.1 de Anexos.

Luego de las preguntas, se realiza una asignación aleatoria entre 2 grupos, denominados One Shot (OS) y Day by Day (DbD).

Durante los días anteriores al inicio de la Eurocopa los participantes, dependiendo de su asignación tenían que proceder de distinta manera: el grupo One Shot tenía que realizar sus predicciones para todos los partidos de una vez, teniendo la posibilidad de visitar los sitios web de apuestas que fueron recomendados en el cuestionario (v.g.: Bet365 y Bwin). Los participantes de este grupo, contaron con todo el tiempo necesario para realizar sus predicciones. En cambio, los del grupo Day by Day tuvieron la opción de consultar el calendario de los partidos, sin posibilidad de realizar predicciones.

A contar del 9 de junio, es decir, un día antes de iniciarse la Eurocopa, mediante la plataforma, los participantes recibieron diariamente un mensaje con un link de la encuesta y donde Qualtrics filtraba según el grupo asignado.

A los del grupo One Shot se les preguntó si habían visto el partido anterior; y como el campeonato estaba ad portas de comenzar, a cada uno se les envió mensajes recordándoles sus apuestas y que el torneo empezaría al día siguiente.

Por el contrario, a las personas del grupo Day by Day se les pidió pronosticar qué equipos ganarían, perderían y/o empatarían en los partidos a disputarse en el próximo día. Para terminar, el pago a los participantes se realizaría cada dos días, detallándose de la siguiente manera:

- Se realizaban 3 pagos “buenos” cuya suma ascendía a £4 por partido acertado. Para ello, se asignaban al azar 3 participantes y se comprobaba si acertaron o no.
- Luego se realizaba un pago “medio” de £0.4 por partido acertado. Eran pagados a 30 personas al azar sin considerar a las 3 personas del pago “bueno”.
- Por último, la gente que no fue elegida al azar se les pagaba £0.2 por haber participado el día anterior en el estudio (sin necesidad de haber acertado).

De esta forma los participantes podían recibir £0-£100 si ganan en las predicciones (i.e. si tienen correcta la predicción, se les paga £4 por partido acertado; y, como son 23 partidos, se concluye que en el mejor de los casos pueden hasta £92). Los partidos a predecir se encuentran en la tabla 2.1:

El estudio contempla una duración de entre 1 y 5 minutos diarios, a realizarse por 9 días seguidos. Las personas tendrán la opción de participar desde sus computadores o dispositivos electrónicos personales.

Es menester aclarar que los incentivos económicos otorgados son, exclusivamente, para disminuir la tasa de deserción. Además, a los participantes les llegará un reporte con los resultados obtenidos en la fecha anterior y para el caso de los que predicen una vez se les recordará por quienes apostaron.

	Fecha (Junio)	Partido 1	Partido 2	Partido 3
1	Viernes 10	Francia vs Rumania		
2	Sábado 11	Albania vs Suiza	Gales vs Eslovaquia	Inglaterra vs Rusia
3	Domingo 12	Turquía vs Croacia	Polonia vs Irlanda del Norte	Alemania vs Ucrania
4	Lunes 13	España vs Rep. Checa	Rep. De Irlanda vs Suecia	Bélgica vs Italia
5	Martes 14	Austria vs Hungría	Portugal vs Islandia	
6	Miércoles 15	Rusia vs Eslovaquia	Francia vs Albania	
7	Jueves 16	Inglaterra vs Gales	Ucrania vs Irlanda del Norte	Alemania vs Polonia
8	Viernes 17	Italia vs Suecia	Rep. Checa vs Croacia	España vs Turquía
9	Sábado 18	Bélgica vs Rep. De Irlanda	Islandia vs Hungría	Portugal vs Austria

Tabla 2.1: Partidos de la Eurocopa por día

Fuente: elaboración propia

Por último, los pagos se realizan cada 2 días para que los participantes no pierdan el interés en este experimento. (Figura 2.2 muestra el diagrama)

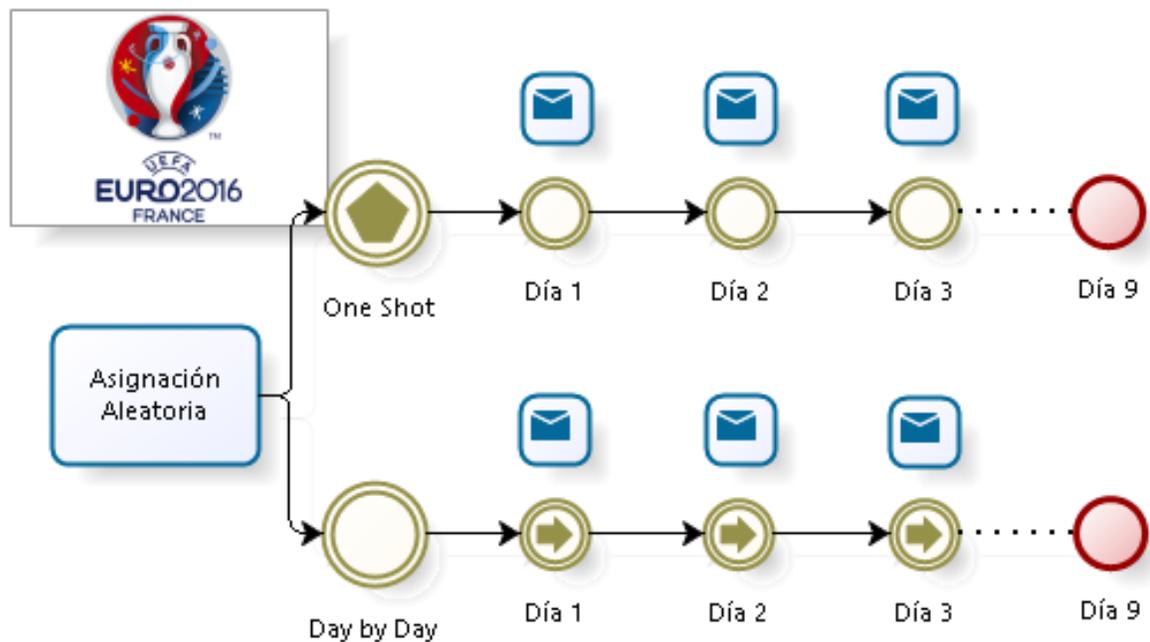


Figura 2.2: Diagrama del funcionamiento del experimento

Fuente: elaboración propia

Además, durante el experimento se recopilan la información de las casas de apuestas (específicamente de dos: Bet 365 y Bwin) antes del inicio de la Eurocopa y durante el torneo (B.2). La idea es probar, con los datos obtenidos de los participantes, si en vez de haber usado sus decisiones basadas en distintos factores, sólo hubiesen utilizado la información proporcionada por las casas de apuestas para, luego, comparar si esta diferencia es estadísticamente distinta y, asimismo, si la acumulación de información para las casas de apuestas puede ser perjudicial para la predicción de partidos.

Luego de la obtención, limpieza de datos y comparación con las casas de apuestas, se procede a realizar regresiones lineales, cuyo propósito sería indicar cuáles eran las variables independientes significativas para explicar la variable dependiente.

En relación a la variable dependiente, se considera la cantidad de partidos total acertados en el período de tiempo determinado por el estudio.

Las variables independientes son, en un orden no taxativo: **(i)** grupo experimental al que pertenece, **(ii)** cantidad de tiempo que se utilizó para realizar la predicción, **(iii)** si vio el partido anterior, **(iv)** cómo realizó la predicción (mirando páginas webs, intuición, etc.), **(v)** si es hombre o mujer, **(vi)** edad, **(vii)** si es fanático del fútbol, entre otros.

El siguiente paso es analizar si existe heterogeneidad considerando alguna de las variables ya mencionadas en el párrafo anterior (fanáticos del fútbol, cantidad de partidos que iban a ver, nivel educacional, edad, entre otros).

Por otro lado, también es importante determinar si existe o no el fenómeno de *attrition* en el estudio ya que, dada su característica de que tener que ser respondido por un horizonte de tiempo limitado, está latente el hecho de que la gente no quiera seguir participando en el estudio. Por esta razón, como se ha venido sosteniendo reiteradamente, se establecieron incentivos económicos y mensajes recordatorios para prevenir la deserción.

En el caso que el *attrition* haya sido un elemento que altere el estudio, se compararán las predicciones considerando hasta ciertos días donde no hubo una cantidad determinante de gente que haya abandonado del estudio.

Luego de identificar si el *attrition* es estadísticamente significativo para cada día, se verifica si la gente que dejó de responder tiene las mismas características que las personas que sí respondieron.

Dado que este efecto sesga la conclusión, se corregirá el *attrition* a través de 2 métodos: el primer método es asumir que la diferencia de gente que no contestó tuvo las mejores predicciones o, en su defecto, las peores predicciones posibles. El otro método es considerar que la diferencia de la gente afectada por *attrition* tenga el mismo comportamiento que la gente que contestó del grupo contrario, por lo que se asume la misma distribución en los casos extremos. Y esto con el fin de analizar cuál de los grupos donde se realizó tratamiento es más efectivo para efectuar predicciones, dada las correcciones propuestas.

Por último, se redactan las conclusiones y análisis de lo aprendido de la ejecución de lo mencionado anteriormente.

2.2. Desarrollo Metodológico

2.2.1. Descripción de la muestra

En el estudio participaron 387 personas, de las cuales 9 de ellas ingresaron a las 2 condiciones (One Shot y Day by Day), teniendo que ser descartadas para el análisis de datos. Por lo tanto, la muestra para trabajar contó con 378 personas. De estas personas, la mayoría pertenecía a Estados Unidos (103 personas) y el Reino Unido (88 personas).

El promedio de edad es de 29 años con una desviación estándar de 9 años. En el siguiente gráfico se muestra, un histograma de la edad de los participantes.

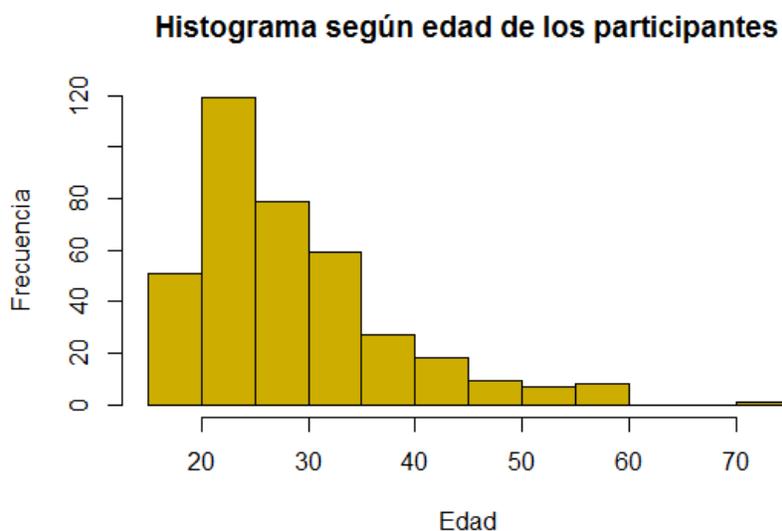


Figura 2.3: Histograma según edad de los participantes

Fuente: elaboración propia

Además, en el estudio participaron en su mayoría hombres (254), lo que corresponde a un 66% de la muestra.

Otro aspecto relevante a considerar fue que la mayoría de los participantes se declaró fanática del fútbol y este fanatismo fue reflejado por los dos grupos de igual manera.

	Dislike a great deal (27)	Dislike a moderate amount (28)	Dislike a Little (29)	Neither like nor Dislike (30)	Like a Little (31)	Like a moderate amount (32)	Like a great deal (33)
One Shot	4	0	6	29	50	49	53
Day by Day	8	8	6	30	52	49	43

Tabla 2.2: Cantidad de participantes por cada grupo según nivel de fanatismo

Fuente: elaboración propia

Asimismo, como se puede observar de la Figura 2.2, existe mayor fanatismo de los hombres que las mujeres y, por otro lado, el fanatismo entre los grupos experimentales es equitativo (ver 2.4).

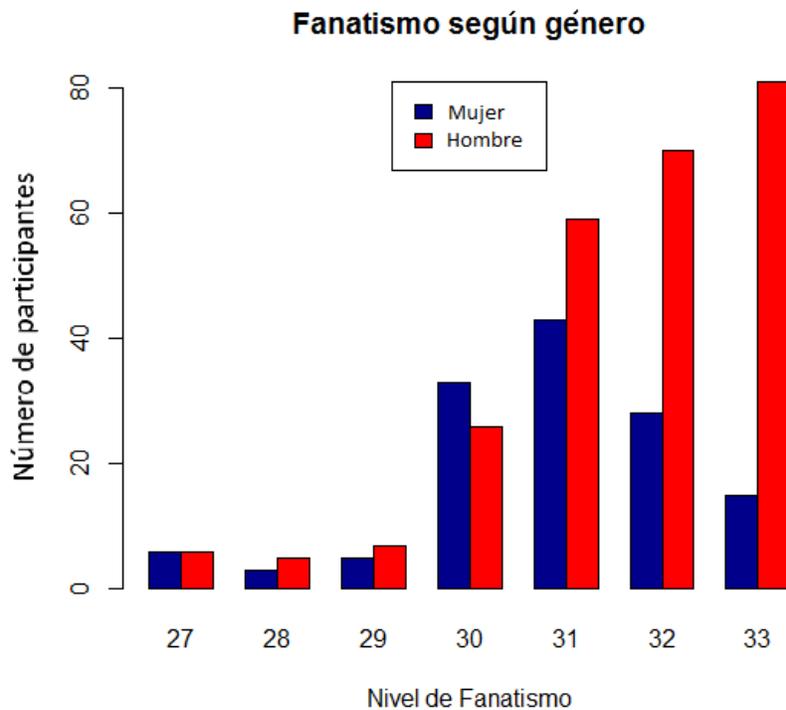


Figura 2.4: Histograma de participantes según nivel de fanatismo por género

Fuente: elaboración propia

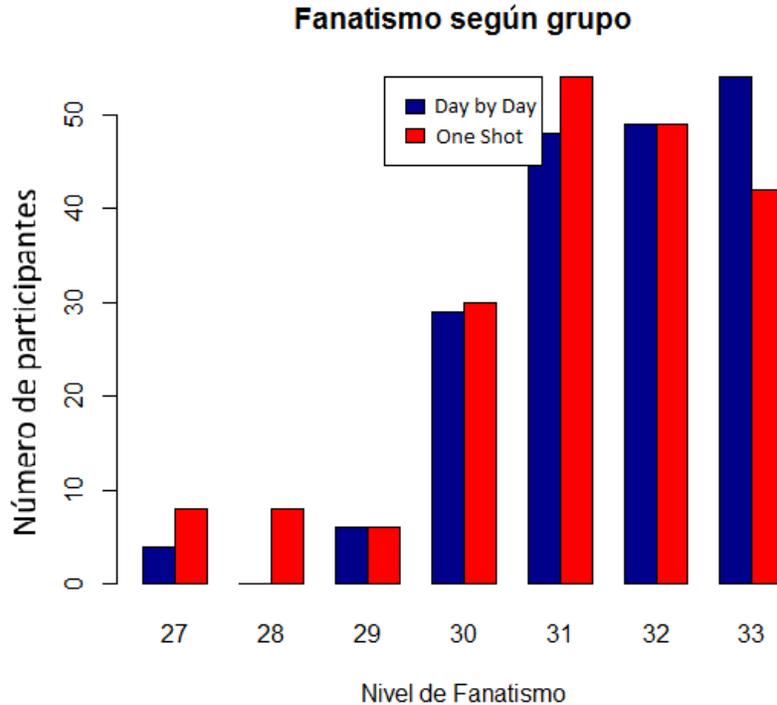


Figura 2.5: Histograma de participantes según nivel de fanatismo por grupo

Fuente: elaboración propia

Dado que existen 7 opciones y no tantos participantes, se prefiere separar la categoría de los fanáticos por 2 tipos, las cuales se denominan fanáticos y no fanáticos. Donde el criterio para pertenecer a los fanáticos es que considera, “Like a moderate amount” y “Like a great deal” y el resto son los no fanáticos. Con estos cambios el resumen de la información se muestra en la siguiente tabla (tabla 2.3):

	One Shot			Day by day		
	Hombre	Mujer	Total	Hombre	Mujer	Total
Fanático	72	20	92	76	22	98
No Fanático	56	46	102	44	42	86
	128	66	194	120	64	184

Tabla 2.3: Cantidad de participantes según fanatismo, género y grupo

Fuente: elaboración propia

Además también se obtuvo la información de los participantes de cuantos partidos iban a mirar para la Eurocopa en Francia. Los cuales muestran una tendencia similar a la de los fanáticos a pesar de que muy poco iban a ver todos los partidos del campeonato.

	One Shot			Day by Day		
	Hombre	Mujer	Total	Hombre	Mujer	Total
All matches	11	6	17	11	3	14
Most of the matches	38	10	48	38	6	44
Half of the matches	14	6	20	18	7	25
Only a few matches	31	21	52	34	24	58
None of them	34	23	57	19	24	43
	128	66	194	120	64	184

Tabla 2.4: Cantidad de participantes según cantidad de partidos a ver, género y grupo

Fuente: elaboración propia

Por otra parte, es importante destacar el porcentaje de gente que participó en este estudio, considerando la eventual existencia del fenómeno del *attrition* y, por consiguiente, perjudicar el estudio. En la Figura 2.6 exhibe el porcentaje de participación por día de los participantes.

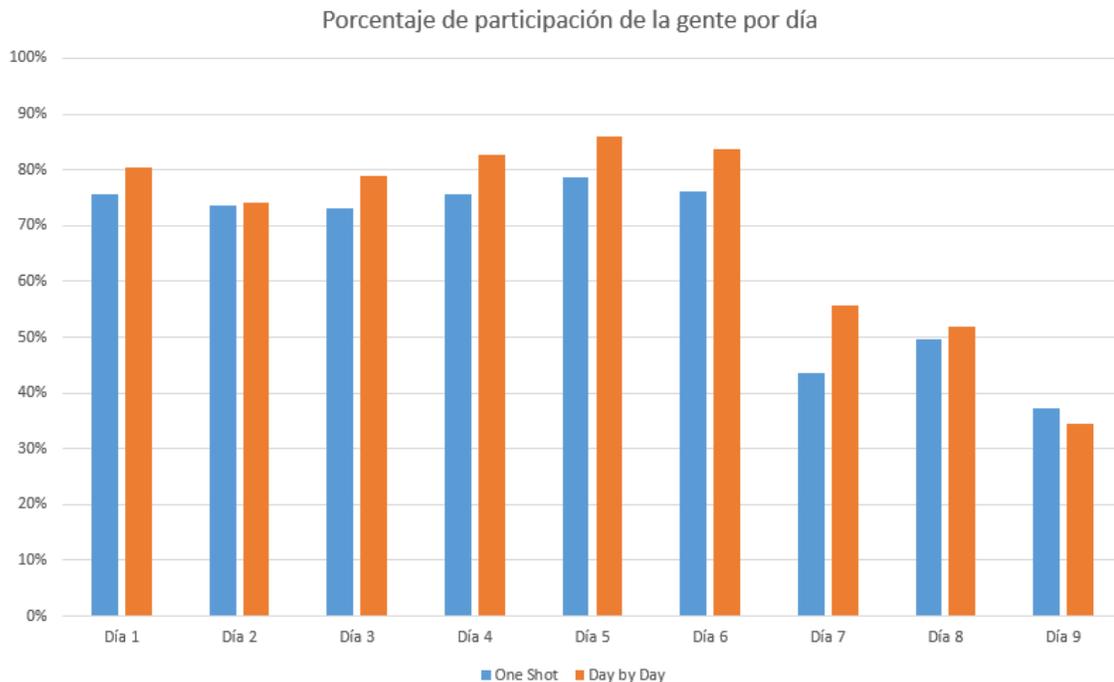


Figura 2.6: Participación por día de la gente por grupo

Fuente: elaboración propia

Lo cual refleja que tal fenómeno estuvo presente y que en los primeros 6 días actuó de manera constante y considerando los días 7 al 9 el *attrition* fue más significativo. Esto último se puede explicar porque la plataforma Prolific tuvo problema estos días por lo cual perjudicó la función de la plataforma para mandar mensajes de recordatorio a los participantes. Un *attrition* que afecte a ambos grupos no es preocupante – el diseño experimental fue realizado pensando en esto. Sin embargo, el problema que se genera al haber desgaste (*attrition*) es

preocupante siempre cuando esté correlacionado a variables no observables que produzcan endogeneidad en la variable de tratamiento.

Dado el número de personas que participaron en el estudio, se calculó el porcentaje de acierto de los 2 grupos: Grupo One Shot y Grupo Day by Day.

Cabe recordar que se realizaron predicciones por 9 días de la Eurocopa, pero no todos los días hubo la misma cantidad de partidos. En la siguiente tabla se muestra la cantidad de partidos por día.

Días	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9
Cantidad de partidos	1	3	3	3	2	2	3	3	3

Tabla 2.5: Cantidad de partidos por día

Fuente: elaboración propia

2.2.2. Análisis de Resultados

Los resultados obtenidos, representados a través de los siguientes gráficos, muestran los porcentajes de acierto que tuvieron cada grupo, tanto el acumulado como el desagregado:

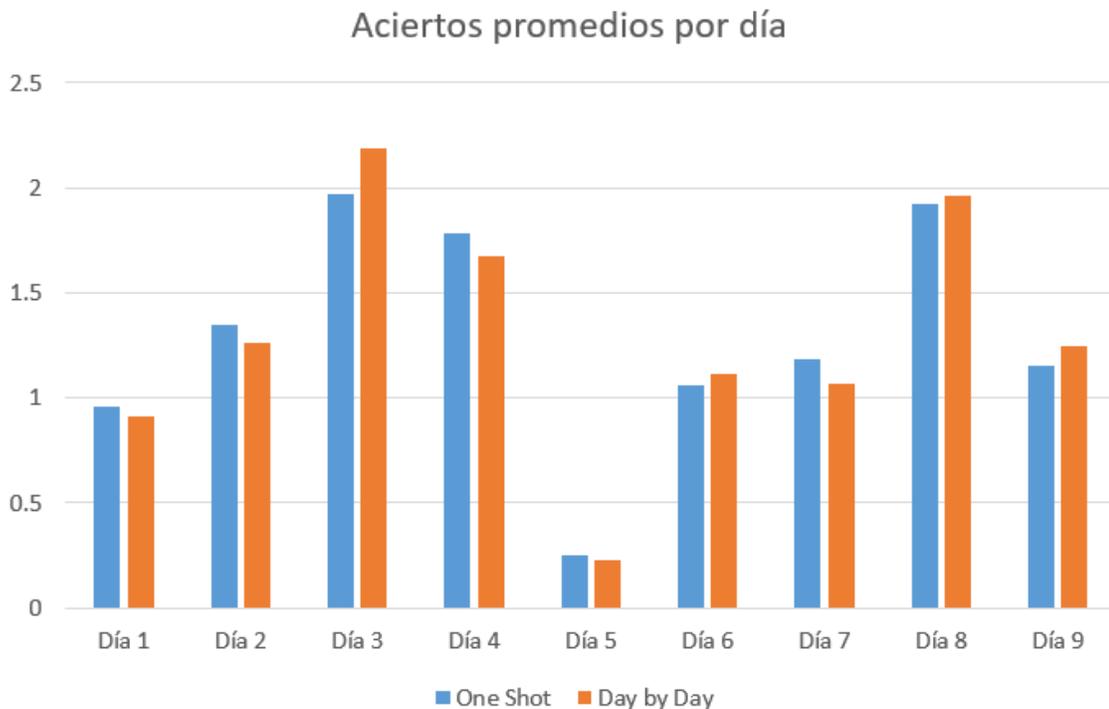


Figura 2.7: Aciertos por día y grupo

Fuente: elaboración propia

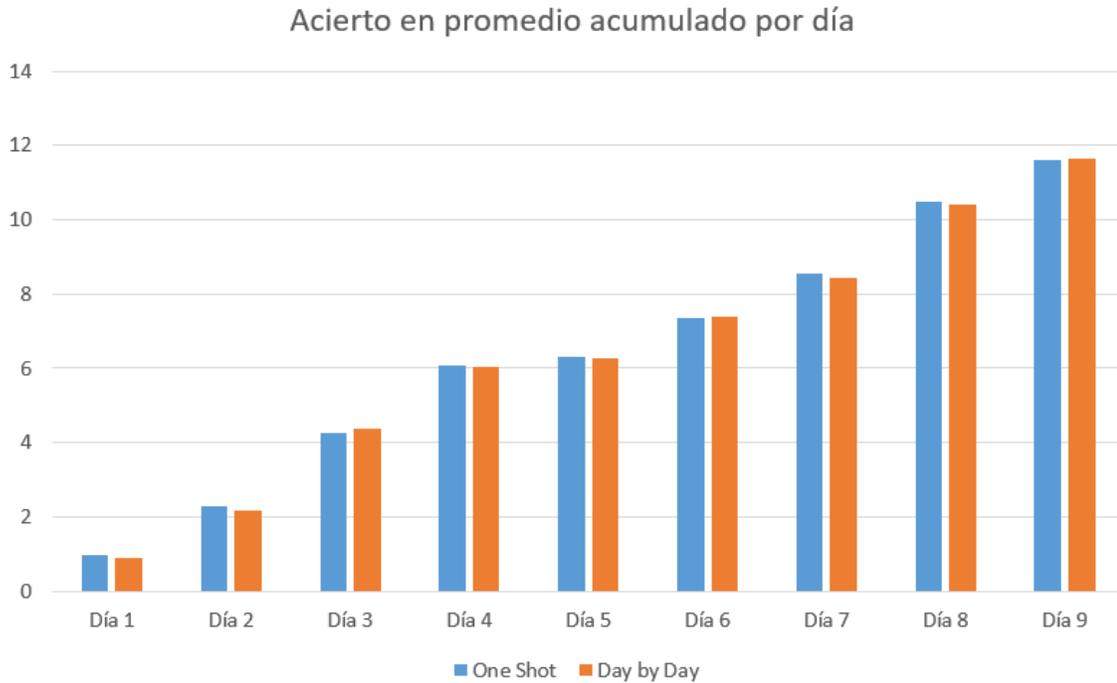


Figura 2.8: Acierto acumulado por día según grupo

Fuente: elaboración propia

En ambos gráficos (Figura 2.7 y 2.8) se puede apreciar de que no existen grandes diferencias, de hecho en el gráfico 7 se muestra que casi no existen diferencia y si existiese, esta es mínima. Para comprobarlo se calcula el promedio de aciertos de forma acumulada se puede ver en la siguiente tabla (2.6), que no existe mucha diferencia, donde los resultados para ambos grupos es:

One Shot	11.622
Day by Day	11.651

Tabla 2.6: Promedio de aciertos acumulados por grupo

Fuente: elaboración propia

Por el contrario donde el gráfico de la Figura 2.7 muestra que en algunos días existió más diferencia que en otros. Por esta misma razón se analiza si las diferencias de los resultados son estadísticamente significativas, calculando los p-valores, expuestos en la siguiente tabla:

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9
p-valor	11.53	31.85	3.18*	18.57	66.89	25.39	18.27	74.27	42.54

Tabla 2.7: P-valores (%) según diferencia de attrition de los participantes

Fuente: elaboración propia

Dado los resultados obtenidos por la diferencia de medias por días, se concluye que solamente para el día 3 existe una diferencia estadísticamente distinta en las predicciones de los individuos.

Por esta misma razón, y tomando en consideración que el efecto del *attrition* está sesgando los resultados de los pronósticos, pues la falta de gente para un grupo perjudica o favorece el promedio de aciertos; es que se realizó un análisis con sólo los datos de las personas que contestaron todos los días (9 días).

En relación a lo anterior, como resultado se obtiene un total de 87 personas ya que, como queda en evidencia en el gráfico de la Figura 2.6 , los últimos tres días la participación bajó considerablemente debido a que hubo una falla en la plataforma y provocó el no envío de los mensajes. Ergo, de estos 87 participantes, 43 pertenecen al grupo One Shot y 44 son del Day by Day.

Acto seguido, se realiza una comparación de los aciertos promedios de ambos grupos, según se muestra en la siguiente tabla:

One Shot	11.744
Day by Day	11.591

Tabla 2.8: Promedios de aciertos considerando 9 días

Fuente: elaboración propia

No existe diferencia estadísticamente significativa entre grupos con los resultados presentados en la Tabla 2.8).

Por último, dado lo mencionado anteriormente sobre los problemas en la plataforma por los últimos 3 días de los experimentos; se decide realizar un análisis con la gente que participa todos los días por los primeros 6 días con el fin de poder realizar un análisis con una mayor cantidad de participantes.

Se obtiene una muestra de 213 personas que participaron los primeros 6 días sin excepción, de las cuales 101 personas pertenecen al grupo One Shot y 112 personas corresponden al grupo Day by Day.

Al comparar los promedios de aciertos para estos días, se obtuvieron los siguientes resultados (tabla 2.9):

One Shot	7.406
Day by Day	7.336

Tabla 2.9: Promedio de aciertos considerando 6 días

Fuente: elaboración propia

A pesar de que el grupo One Shot tuvo mejores pronósticos, en promedio no se puede

aún formular conclusión alguna respecto a la cantidad de aciertos promedios, ya que no son estadísticamente distintos.

Al contrastar los resultados obtenidos de ambos grupos se pretende analizar qué hubiese pasado si los participantes del experimento online hubiesen utilizado la información proporcionada por los sitios web de apuestas *Bet 365* o *Bwin* – en otras palabras, se simulan los resultados de ambos grupos en base a estas casas de apuestas. Formulada esta hipótesis, se asume que en esta simulación no existe el fenómeno de *attrition*, por lo que no es un factor a considerar.

Para el caso del sitio de apuestas de *Bet 365*, los resultados se ilustran en la siguiente tabla (Número 2.10):

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9	Total 6/9 días ¹
One Shot	1	1	3	1	0	1	1*	2	1	7/11
Day by Day	1	1	3	1	0	1	0*	2	1	7/10

(Los números en las celdas representan la cantidad de aciertos según la simulación)

Tabla 2.10: Simulación de ambos grupos según casa de apuesta *Bet 365* por día

Fuente: elaboración propia

En virtud de las cifras obtenidas, queda en evidencia que la información acumulada por el sitio de apuestas perjudica la decisión del grupo Day by Day. Así, según los resultados, se deduce que es conveniente predecir antes de todos los partidos. Dada que la diferencia de los pronósticos se muestra en el día 7, se evidencia cual fue la discrepancia en la siguiente tabla:

Partido	Resultado	Pronóstico OS	Pronóstico DbD
Inglatera vs Gales	2 - 1	Gana Inglaterra	Gana Gales
Ucrania vs Irlanda del Norte	0 - 2	Gana Ucrania	Gana Ucrania
Alemania vs Polonia	0 - 0	Gana Alemania	Gana Alemania

Tabla 2.11: Resultados y pronósticos según casa de apuesta *Bet 365* para el día 7 del experimento

Fuente: elaboración propia

Donde la única diferencia en los pronósticos fue para el primer partido señalado en la Tabla 2.11.

¹El concepto “X/Y” significa que el primer número (X) corresponde al total de aciertos para los 6 días y el número siguiente (Y) corresponde a los aciertos considerando los 9 días.

Por otro lado, a partir de los números de las apuestas del sitio *Bwin*, se llegaron a resultados opuestos: para la casa de apuestas *Bwin* la información recopilada permitió aventajar al grupo Day by Day por sobre el de One Shot. Donde el One Shot asumió que en el partido entre Albania y Suiza ganaría Albania (el resultado fue 1-0 a favor de Suiza).

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9	Total 6/9 días
One Shot	1	1*	3	1	0	1	1	2	1	7/11
Day by Day	1	2*	3	1	0	1	1	2	1	8/12

Tabla 2.12: Predicciones según *Bwin* por día y grupo

Fuente: elaboración propia

El detalle sobre los resultados de los sitios de apuestas se puede revisar en la sección Anexos: “Cuotas de los sitios de apuestas online”.

Luego se procede a analizar si la diferencia entre los aciertos hecho por los participantes y los algoritmos es significativa o no a través de un test T. En las tablas 2.13 y 2.14 se muestran las medias de los pronósticos según el experimento, los cuales se contrastan según la casa de apuesta, por la cantidad de días considerados.

	Bet 365	Bwin
One Shot	7.366071***	7.366071 ***
Day by Day	7.405941 ***	7.405941 ***

Significativos al nivel 0,10 - **Significativos al nivel 0,05 y ***Significativos al nivel 0,01

Tabla 2.13: Comparación de medias entre el algoritmo y resultados del experimento para los primeros 6 días.

Fuente: elaboración propia

	Bet 365	Bwin
One Shot	11.59091***	11.59091***
Day by Day	11.74419 **	11.74419 ***

Significativos al nivel 0,10 - **Significativos al nivel 0,05 y ***Significativos al nivel 0,01

Tabla 2.14: Comparación de medias entre el algoritmo y resultados del experimento para los primeros 9 días.

Fuente: elaboración propia

Dado los p-valores podemos concluir que las diferencias son estadísticamente distintas, en todos los casos de los 6 días de pronósticos favorecen a los participantes que contestaron

durante estos días y la diferencia es estadísticamente significativa. Luego para el caso de los 9 días de apuestas, *Bet 365* tiene menos aciertos que la gente del experimento, en la casa de apuesta *Bwin* también es el caso pero exceptuando cuando es el caso del Day by Day, donde es mejor la casa de apuesta.

Luego como primera regresión de este estudio es observar si la cantidad de acierto de partidos depende de si las personas pertenecen al grupo One Shot o al grupo Day by Day. Los resultados obtenidos son los siguientes, mostrados en la tabla 2.15:

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	7.3661	0.1337	55.10	0.0000
Grupo	0.0399	0.1941	0.21	0.8375

Tabla 2.15: Resultados de la regresión para 6 días considerando solo la variable independiente grupo

Fuente: elaboración propia

Al respecto se puede observar que no existen diferencias entre grupos.

Ahora se continúa con el análisis para la gente que participó constantemente en los 9 días. Al igual que el caso anterior se procede a realizar la regresión obteniendo los resultados que se reflejan en la tabla 2.16:

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	11.5909	0.2385	48.59	0.0000
Grupo	0.1533	0.3393	0.45	0.6526

Tabla 2.16: Resultados de la regresión para 9 días considerando solo la variable independiente grupo

Fuente: elaboración propia

2.2.3. Propuesta de soluciones Attrition

2.2.3.1. Modelo 1: Agregar valores extremos

Dado que en el experimento existe el fenómeno *attrition* se plantean las siguientes propuestas de solución (en Anexo sección Caracterización del attrition del experimento online se puede apreciar la caracterización de las personas que abandonaron el experimento online).

La primera solución toma en cuenta la gente que no participó ciertos días y se asume que ellos hubiesen acertado correctamente sus predicciones.

En contraste, para el segundo caso se asume que las personas que no contestaron iban a tener todas las predicciones incorrectas.

Cabe recordar que se cuentan con todas las predicciones realizadas por las personas que pertenecen al grupo One Shot, así que, al aplicar los casos extremos, sólo se realiza para las personas del Day by Day. En primera instancia, se muestra una tabla comparativa de los participantes del One Shot según los que contestaban todos los días ya que, si ellos respondían todos los días, se validaba la predicción que habían realizado, en caso contrario, no se contabilizaba.

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9	Suma
One Shot con attrition	0.958	1.345	1.971	1.783	0.248	1.056	1.180	1.926	1.154	11.622
One Shot sin attrition	0.943	1.289	1.954	1.737	0.253	1.041	1.119	1.819	1.077	11.231

Tabla 2.17: Contraste de predicción del grupo One Shot con y sin attrition

Fuente: elaboración propia

Luego, al no considerar el *attrition* en el One Shot y al complementar con los valores extremos para el Day by Day, se obtiene lo siguiente (tabla 2.18):

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9	Suma
One Shot	0.943	1.289	1.954	1.737	0.253	1.041	1.119	1.819	1.077	11.232
Day by Day	0.929	1.707	2.353	1.908	0.467	1.255	1.929	2.456	2.391	15.397

Tabla 2.18: Contraste del Grupo One Shot sin attrition y asumiendo que los missing values del Grupo day by Day son todos aciertos

Fuente: elaboración propia

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9	Suma
One Shot	0.943	1.289	1.954	1.737	0.253	1.041	1.119	1.819	1.077	11.232
Day by Day	0.728	0.941	1.734	1.386	0.196	0.929	0.592	0.728	0.435	7.668

Tabla 2.19: Contraste del Grupo One Shot sin attrition y asumiendo que los missing values del Grupo day by Day son todos erróneos

Fuente: elaboración propia

Para este caso, se observa que los cambios son excesivos. Para la primera tabla, eso sí,

aún sigue siendo mejor la predicción del One Shot. Esto quiere decir que, para el día que hay solo un partido, éste no afecta; pero, para los otros días, la diferencias son más grandes exceptuando los casos de los días 5 y 6.

2.2.3.2. Modelo 2: Considerando que attrition se comporta similar a quienes más fallan en el grupo One Shot

En el siguiente modelo de solución para el *attrition* se usan los datos del One Shot y se calcula la cantidad de gente que no respondió del grupo Day by Day por cada día. Luego, se ordenan de manera ascendente las predicciones del One Shot. Por último, se asigna la predicción de los que no respondieron según la lista ascendente del Grupo One Shot. De esta manera, se asume que las predicciones de los que no contestaron hubieran sido igual que las peores predicciones que el grupo One Shot.

En la tabla a continuación (tabla 2.20 se muestra la cantidad de participantes que abandonaron por día (cantidad de participantes del grupo Day by Day son 184).

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9
Cant.	37	47	38	32	25	30	82	88	120

Tabla 2.20: Cantidad de participantes que abandonaron el estudio del grupo Day by Day

Fuente: elaboración propia

Luego, al contrastar los resultados de la misma manera que la del modelo 1, se obtiene una tabla comparativa por día con los valores promedios:

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9	Suma
One Shot	0.943	1.288	1.953	1.737	0.252	1.041	1.118	1.819	1.077	11.232
Day by Day	0.869	1.005	1.771	1.521	0.195	1.005	0.880	0.869	0.875	8.994

Tabla 2.21: Aciertos según contraste del Grupo One Shot sin attrition y asumiendo solución del modelo 2

Fuente: elaboración propia

Al respecto, se puede apreciar que perjudica los pronósticos del grupo Day by Day. De hecho, hay algunos que son mejores que en otros días, se puede apreciar en el gráfico por día (Gráfico 2.9).

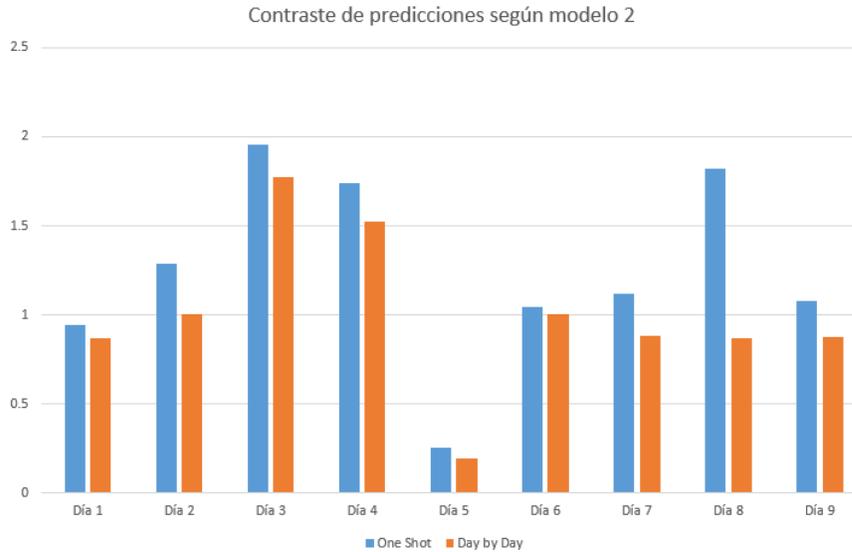


Figura 2.9: Promedio de predicciones según modelo 2

Fuente: elaboración propia

En la sección Anexo se muestra otra propuesta, el cual soluciona el *attrition* asignándole un valor de manera uniforme a los que no participaron, en ambos grupos experimentales (Anexo B.4).

2.2.4. Regresiones y análisis de Heterogeneidad

Para poder indagar de mejor manera los resultados, se consideró el nivel de fanatismo declarado por los mismos participantes del estudio para analizar el comportamiento según los grupos y si efectivamente la variable fanatismo es importante al momento de realizar las predicciones.

Como se mencionó anteriormente, existen distintos niveles a pesar del número reducido de participantes. Debido a esta razón es que se agruparon en 2 categorías: los no fanáticos y los fanáticos. Esto quiere decir que en el grupo de los fanáticos se encuentran las personas que contestaron sólo la opción “*Like a great deal*”, y el resto en el grupo no fanáticos. Lo anterior quedó distribuido de la siguiente forma:

	One Shot	Day by Day
Fanático	92	98
No Fanático	102	86
Total	194	184

Tabla 2.22: Cantidad de fanáticos según grupo experimental

Fuente: elaboración propia

Luego, se realiza un cuadro comparativo con las predicciones acumuladas por grupo:

	One Shot	Day by Day
Fanático	11.768	11.957
No Fanático	11.486	11.218

Tabla 2.23: Predicciones promedio por fanatismo y grupos

Fuente: elaboración propia

Ahora, se procede con las regresiones agregando la heterogeneidad. Este estudio se puede desglosar por 4 categorías: por sexo, por fanatismo, por educación y por la cantidad de partidos que los participantes declararon al iniciar el estudio. En primera instancia, se inicia con los casos de la gente que respondió 6 y 9 días para luego continuar con las regresiones de las propuestas de mejora del *attrition*. Considerando las variables independientes, grupos, variable de heterogeneidad (fanatismo, género y cantidad de partidos a observar) y sus interacciones:

En la siguiente tabla se muestran los resultados de los estimadores y su significancia para el caso que sólo haya respondido para los primeros 6 días:

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	7.1277	0.2061	34.59	0.0000
Grupo	0.2057	0.2818	0.73	0.4663
Fanático	0.4108	0.2705	1.52	0.1304
Grupo:Fanático	-0.2548	0.3906	-0.65	0.5150

Tabla 2.24: Resultados de la regresión considerando heterogeneidad del fanatismo para 6 días

Fuente: elaboración propia

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	11.4706	0.3880	29.56	0.0000
Grupo	0.2567	0.5166	0.50	0.6206
fanatico	0.1961	0.4953	0.40	0.6932
Grupo:fanatico	-0.1614	0.6953	-0.23	0.8170

Tabla 2.25: Resultados de la regresión considerando heterogeneidad del fanatismo para 9 días

Fuente: elaboración propia

Al analizar los resultados, estos no reflejan valores estadísticamente significativos pero, a pesar de esto, existe una tendencia negativa al pertenecer al grupo One Shot y al ser fanático.

En cambio, al considerar la variable de fanatismo como variable continua se obtienen los siguientes resultados:

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	0.9008	3.1905	0.28	0.7780
Grupo	2.8326	4.1176	0.69	0.4923
Fanatismo	0.2046	0.1009	2.03	0.0438
Grupo:Fanatismo	-0.0864	0.1311	-0.66	0.5104

Tabla 2.26: Resultados de la regresión considerando heterogeneidad del fanatismo (de forma continua) para 6 días

Fuente: elaboración propia

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	0.8245	5.8640	0.14	0.8885
Grupo	9.4875	7.2047	1.32	0.1915
Fanatismo	0.3391	0.1845	1.84	0.0697
Grupo:Fanatismo	-0.2931	0.2282	-1.28	0.2027

Tabla 2.27: Resultados de la regresión considerando heterogeneidad del fanatismo (de forma continua) para 9 días

Fuente: elaboración propia

Las cuales muestran que el factor fanatismo es importante y estadísticamente significativo al momento de realizar predicciones, pero en ambas regresiones también se muestra que el factor de interacción entre grupo y fanatismo tiene pendiente negativa, por lo que es bueno ser fanático para las predicciones pero que es mejor que un fanático pertenezca al grupo Day by Day en vez del grupo One Shot. De lo contrario, estaría perjudicando la predicción.

Por otra parte, se agrega otra variable independiente: la educación. Esta también fue discretizada en 1 y 0, en el cual la gente que pertenece al grupo “*Graduate degree (MA, PhD, MBA, etc)*” o “*Undergraduate college degree (BS, BA)*”, se le asigna 1, y el resto se evalúa como 0. Así, se obtienen los siguientes estimadores:

	Estimate	Error Estándar	valor t	Pr(> t)
(Intercepto)	7.1957	0.2083	34.55	0.0000
Grupo	0.4359	0.3096	1.41	0.1607
Educación	0.2892	0.2713	1.07	0.2877
Grupo:Educación	-0.6509	0.3972	-1.64	0.1028

Tabla 2.28: Resultados de la regresión considerando heterogeneidad educación para 6 días

Fuente: elaboración propia

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	11.6176	0.2726	42.61	0.0000
Grupo	-0.0370	0.3948	-0.09	0.9255
Educación	-0.1176	0.5719	-0.21	0.8375
Grupo:Educación	0.7037	0.7869	0.89	0.3738

Tabla 2.29: Resultados de la regresión considerando heterogeneidad educación para 9 días

Fuente: elaboración propia

Con respecto a los resultados obtenidos, las pendientes son opuestas al observar las predicciones entre 6 y 9 días. Por lo que, se puede decir que con tener un nivel educacional más alto y ser del grupo One Shot mejora las predicciones al aumentar los días, en el caso de considerar solo la personas que contestaron los 9 días. Ya que al observar la otra tabla (tabla 2.28) se muestra lo contrario. En conclusión el nivel educacional no es un factor relevante al tomar en cuenta si se pertenece al grupo One Shot o al grupo Day by day. Sin embargo, un factor importante en la regresión es el factor grupo, ya que esta aumenta al pertenecer al grupo One Shot con un p-valor menor 0.16, para el caso de los 6 días.

Por último se realiza la regresión para ambos casos (6 y 9 días), considerando la cantidad de partidos que declararon que iban a ver. A continuación se muestra en las siguientes tablas:

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	7.7722	0.3824	20.32	0.0000
Grupo	-0.3001	0.5464	-0.55	0.5834
‘cant partidos a ver’	-0.1213	0.1070	-1.13	0.2582
Grupo:‘cant partidos a ver’	0.1023	0.1497	0.68	0.4953

Tabla 2.30: Resultados de la regresión considerando heterogeneidad “cantidad partidos a ver” para 6 días

Fuente: elaboración propia

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	12.1605	0.6322	19.24	0.0000
Grupo	0.1970	0.8769	0.22	0.8228
‘cant partidos a ver’	-0.1728	0.1777	-0.97	0.3335
Grupo:‘cant partidos a ver’	-0.0225	0.2513	-0.09	0.9288

Tabla 2.31: Resultados de la regresión considerando heterogeneidad “cantidad partidos a ver” para 9 días

Fuente: elaboración propia

Las cuales muestran que no existen resultados estadísticamente significativos, y que el hecho de declarar la cantidad de partidos a mirar aporta negativamente al momento de realizar la predicción. También la interacción entre grupo y la variable de heterogeneidad

cambia de signo al considerar más días, pero sus desviaciones estándares son muy grandes con respecto al estimador para poder concluir.

Luego se procede a realizar las regresiones usando los datos obtenidos por los modelos de solución al *attrition*. Este procedimiento se realiza para el modelo 2 y el 3 el cual se muestra en Anexo B.4. No se considera el modelo 1 porque tiene valores muy extremos. Cabe destacar que al proponer los modelos se considera de qué manera se puede equilibrar el *attrition*, luego en la implementación de estas propuestas, al momento de reemplazar los *missing values* por los valores según la regla del modelo, no se distinguen a qué género corresponde y tampoco nivel de fanatismo ni nivel educacional.

Por lo que al aplicar heterogeneidad a los modelos propuestos, los resultados tienen sesgos ya que no se realiza de manera equitativa y pareja para todos los tipos de participantes. Por lo que las regresiones que se muestran a continuación no consideran la heterogeneidad, pero si la variable Grupo que es de interés en el estudio.

Resultados Modelo 2

	Estimador	Error Estándar	valor t	Pr(> t)
(Intercepto)	9.7446	0.1782	54.68	0.0000
Grupo	1.4874	0.2488	5.98	0.0000

Los resultados muestran que al reemplazar los valores de *attrition* del grupo Day by Day, estos perjudican de manera significativa. Por esta misma razón es que la regresión muestra que es mejor pertenecer al grupo One Shot al momento de apostar.

Resultados Modelo 3

	Estimate	Error Estándar	valor t	Pr(> t)
(Intercepto)	8.7440	0.2670	32.75	0.0000
Grupo	-0.7272	0.3723	-1.95	0.0516

Para este caso se considera de manera negativa, estadísticamente significativa (p-valor<0.1). Por lo que en este caso en particular se puede interpretar que al considerar más información favorece al momento de realizar apuestas en resultados deportivos. Por ende se concluye que considerando el *attrition* no es posible concluir que un grupo sea superior a otro.

2.2.5. Conclusiones y Aprendizajes

En primer lugar, se observa que el *attrition* resulta ser un tema relevante y por consiguiente es necesario que se considere en este tipo de experimentos. Debido a esto, los incentivos que se utilizan para que los participantes no pierdan continuidad al momento de participar en todo el estudio adquieren importancia. Es por ello, que mandar mensajes a cada participante de manera personalizada parece ser la opción adecuada, y así, poder disipar de manera drástica el fenómeno del *attrition*. Sin embargo, mandar mensajes de manera personalizada puede producir sesgos a ciertos participantes, e incluso que los sujetos no sean tratados de la misma manera en el desarrollo del experimento.

Lo anterior ocurre debido a que, por un lado, puede haber gente comprometida que no requiera los recordatorios y por otro lado, que exista gente que necesite más de un recordatorio para que no exista abandono de participantes durante el estudio, que además puede producir sesgos al momento de ser tratada. Otro sesgo que se produce por el *attrition* - en el que subyace la importancia del trabajo de memoria - se evidencia cuando se desconocen las características de las personas que abandonan el estudio de ambos grupos. El cual produce un problema al comparar los dos grupos, ya que las muestras se encuentran desbalanceadas, vale decir, existe la posibilidad de que los participantes que abandonaron el experimento subieran o disminuyeran el porcentaje de acierto de ese grupo.

Por el contrario al momento de levantar conclusiones referentes al género resulta complejo sintetizar un análisis de información ya que las cantidades entre hombres y mujeres son dispares. No obstante, respecto a las predicciones por un algoritmo o casas de apuestas, se pudo observar que existe un contraste con respecto a las dos casas de apuestas. Sin embargo no se esclarece cuál de las dos maneras es mejor para realizar las apuestas, producto de que una casa de apuesta demostraba que era mejor predecir una vez al inicio. En contraste, para la otra casa de apuesta era mejor considerar más información para aumentar la cantidad de aciertos.

A su vez, en todas las situaciones modeladas en la presente memoria, al momento de comparar con los pronósticos de los participantes, el algoritmo aplicado en el experimento no demostró un incremento en el número de aciertos, por lo que no se pudo observar la ventaja del algoritmo frente a las decisiones humanas. Esto ratifica el hecho de que es difícil poder predecir resultados deportivos.

De la misma forma, la significancia de los datos al agregar heterogeneidad y considerar 6 días en vez de los 9 días reflejan más estadísticamente. Lo que puede estar condicionado por la cantidad de participantes que tuvo ese grupo en específico (213 vs 87 personas). Por esta razón, se consideraron más los resultados que evidencian un contraste entre los participantes que respondieron durante un periodo de 6 días y quienes respondieron en el estudio durante los 9 días.

Por último, al analizar los datos con distintos modelos muestran que el grupo Day by Day no tuvo un mejor desempeño respecto al grupo One shot. Además, es indispensable señalar que los resultados obtenidos no demuestran lo que en la literatura se refleja en relación a las políticas *closed-loop*.

Capítulo 3

Experimento Natural para el orden de la secuencia de eventos

3.1. Metodología

Para el segundo experimento, se analizarán los resultados de los partidos de la Liga Inglesa (English Premier League) de los últimos años, debido a que la Federación [34] realiza el *fixture*, año a año, de forma aleatoria, (*Natural Experiment*) [31], con algunas reglas generales [35]), que influyen a todos los equipos del torneo por igual.

Por lo tanto, el objetivo es encontrar un patrón que pueda explicar el motivo por el cual a veces los equipos pueden obtener muy buenos resultados en una temporada y en otra muy nefastos. A priori, se estima que este efecto puede estar producido por el orden en que son asignados los partidos, distribuyéndose estos con equipos de mejor rendimiento al comienzo, al final, a mitad de temporada o uniformemente en las 38 fechas de la liga.

La base de datos fue obtenida del sitio oficial [36] de la liga inglesa y del sitio web alemán donde están los valores de los clubes (Transfermarkt). Se emplearán, conjuntamente, los datos de las temporadas que median desde 2005/2006 hasta 2014/2015.

Los principales datos a considerar son:

- los resultados de los partidos
- qué equipo jugó de local y de visita
- la fecha del partido
- la cantidad de público en cada partido
- cantidad de tarjetas amarillas por partido
- cantidad de tarjetas rojas por partido

Antes de iniciar con los rendimientos de los equipos según los partidos difíciles al inicio de temporada, se realiza un análisis exploratorio y descriptivo de la muestra. Luego, se calculan

las variables dependientes a testear, puntaje total al final de la temporada, puntaje total a mitad de la temporada (puntaje en la fecha 20), posición al final de la temporada y posición a mitad de la temporada (posición a mitad de la temporada). Después, se proponen distintos criterios para clasificar si el partido con un equipo específico es difícil o no (también si es o no fácil), de la siguiente manera:

El **primer modelo** se define según la cantidad de años anteriores a considerar. Se considera en primera instancia solo la temporada pasada, en segunda instancia 2 temporadas anteriores y así sucesivamente hasta 5 temporadas anteriores. Luego para cada caso (son 5 casos, porque son 5 años) se califica al rival según su dificultad dependiendo en qué posición terminó la temporada anterior. En el caso de considerar más de una temporada, se realiza una tabla acumulada sumando los puntajes obtenidos de los equipos en dichas temporadas, para luego poder realizar un ranking y otorgar a cada equipo una posición. Por lo tanto para estos casos se tiene la dificultad de manera continua. Además se agrega la dificultad de manera discreta para cada uno de estos 5 casos. La definición a esta variable se obtiene a partir de una binaria: 1 si el equipo pertenece a los 5 primeros de la tabla y 0 si no. Este criterio fue escogido porque los 5 primeros de la Liga Inglesa son los que clasifican a copas europeas (UEFA Champions League y Europa League).

Como **segundo modelo** se realiza el mismo análisis anteriormente mencionado separando la tabla de posiciones según la tabla de desempeño por localía (la tabla de clasificación final se puede separar según el desempeño de local o visita que tuvo esa temporada). En este caso si el número es menor es más difícil. V.g., si Chelsea FC juega de local contra FC Arsenal, entonces hay que buscar en qué posición está FC Arsenal en la tabla de desempeño de visita. Un ejemplo de las tablas formadas se muestra en la sección Tipos de tabla de análisis, considerando también la temporada 2005/2006 como referencia. Este segundo criterio se realiza tomando en cuenta la temporada anterior y las 2 temporadas anteriores.

Como **tercer modelo** se considera si el rival pertenece a los últimos 5 de la tabla de posiciones. El criterio en este caso es que tan ‘fácil’ fueron los partidos seleccionados. Se recuerda que los últimos 3 al final de un campeonato descienden a la segunda división de la Liga Inglesa, por lo que si el equipo rival no está en la tabla de posiciones significa que recién ascendió, por lo cual se le considera con valor 1. Este criterio se aplica al considerar la tabla final de la temporada anterior, según criterio de rendimiento de localía y visita, y también para el caso de 2 temporadas anteriores. En la tabla B.7, se muestra un ejemplo de como se establece la tabla de análisis para la temporada 2005/2006.

Finalmente, como **cuarto modelo** se realiza el análisis de dificultad considerando la tabla acumulada entre las temporadas 2000/2001 hasta el 2004/2005. realizando los modelos de manera discreta como continua.

Por último, realizando un esquema de lo anteriormente mencionado, se presenta de la siguiente manera.

- Modelo 1: Dificultad por agregación de temporadas
 - Según rendimiento del rival de la temporada pasada
 - * Posición del rival en la tabla (variable continua)

- * Si pertenece a los primeros 5 equipos de la tabla (variable discreta)
- Según rendimiento del rival de las 2 temporadas pasadas
 - * Posición del rival en la tabla (variable continua)
 - * Si pertenece a los primeros 5 equipos de la tabla (variable discreta)
- Según rendimiento del rival de las 3 temporadas pasadas
 - * Posición del rival en la tabla (variable continua)
 - * Si pertenece a los primeros 5 equipos de la tabla (variable discreta)
- Según rendimiento del rival de las 4 temporadas pasadas
 - * Posición del rival en la tabla (variable continua)
 - * Si pertenece a los primeros 5 equipos de la tabla (variable discreta)
- Según rendimiento del rival de las 5 temporadas pasadas
 - * Posición del rival en la tabla (variable continua)
 - * Si pertenece a los primeros 5 equipos de la tabla (variable discreta)
- Modelo 2: Dificultad por rendimiento de local y visita
 - Según rendimiento del rival de la temporada pasada según localía y visita
 - * Posición del rival en la tabla según si es local o visita (variable continua)
 - * Si pertenece a los primeros 5 equipos de la tabla según tabla de local y visita (variable discreta)
 - * Según rendimiento del rival de 2 temporadas anteriores según localía y visita
 - * Posición del rival en la tabla según si es local o visita (variable continua)
 - * Si pertenece a los primeros 5 equipos de la tabla según tabla de local y visita (variable discreta)
- Modelo 3: Criterio considerando a últimos equipos de la tabla (fácil)
 - Según rendimiento del rival de la temporada pasada
 - * Si pertenece a los últimos 5 equipos de la tabla (variable discreta)
 - * Si pertenece a los últimos 5 equipos de la tabla según tabla de local y visita (variable discreta)
 - Según rendimiento del rival de 2 temporadas pasadas
 - * Si pertenece a los últimos 5 equipos de la tabla (variable discreta)
 - * Si pertenece a los últimos 5 equipos de la tabla según tabla de local y visita (variable discreta)
- Modelo 4: Considerando datos entre temporadas 2000/2001 y 2004/2005
 - Según rendimiento del rival de las 5 temporadas
 - * Posición del rival en la tabla (variable continua)
 - * Si pertenece a los primeros 5 equipos de la tabla (variable discreta)

Luego de formar la base de datos con las distintas propuestas de dificultad del partido más las variables dependientes, se definen métricas que van a incluir cierta cantidad de partidos, por lo que se proponen las siguientes como variables independientes:

- Primeros 3 partidos

- Primeros 5 partidos
- Primeros 7 partidos
- Partidos de la fecha 19, 20 y 21
- Partidos de la fecha 18, 19 y 20
- Partidos de la fecha 17, 18, 19 y 20
- Partidos de la fecha 16, 17, 18, 19 y 20
- Partidos de la fecha 16, 17, 18, 19, 20 y 21
- Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)
- Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)
- Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)
- Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)
- Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)
- Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)
- Últimos 3 partidos de la temporada (fechas 36, 37 y 38)
- Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)
- Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)

Luego, se realizan regresiones lineales con la finalidad de determinar cuál o cuáles son los valores que afectan directamente a las distintas variables dependientes propuestas anteriormente, agregándole otras variables independientes (v.g.: la localía de los partidos).

La primera regresión a analizar es la siguiente:

$$y_{eq,temp} = \beta_0 + \beta_1 DIF_{eq,temp} + \varepsilon_{eq,temp} \quad (3.1)$$

La variable $DIF_{eq,temp}$ se define de distintas maneras, según el modelo y la cantidad de partidos a considerar. Esto nos permite analizar cuáles son las tendencias de las regresiones. Por ejemplo: entre más difícil juegue un equipo peor posición termina al final de la temporada.

Luego, se añaden los efectos fijos de los equipos y las temporadas a la primera regresión. Los equipos en sí pueden estar relacionados con la variable dependiente, por lo que considerar este efecto fijo mejorará la interpretación del resultado. Se define la regresión de la siguiente manera:

$$y_{eq,temp} = \beta_0 + \beta_1 DIF_{eq,temp} + \lambda DumEq_{eq} + \mu DumTemp_{temp} + \varepsilon_{eq,temp} \quad (3.2)$$

La variable $DumEq_{temp}$ y $DumTemp_{temp}$ son binarias según el equipo y temporada a considerar respectivamente. Donde λ y μ corresponden a vectores que contienen los estimadores para cada equipo y temporada.

Además, se analiza en cuál de todas las secuencias propuestas anteriormente influyen en las variables independientes. Una vez que se logre encontrar cuál de todas las secuencias

influye más, se proponen otras métricas, con el fin de realizar un análisis más detallado y de sensibilidad para este caso.

La tercera y última regresión que se propone para este estudio es agregar a la segunda regresión la variable $LOC_{eq,temp}$, la cual significa cuántos de los partidos a analizar corresponden a los jugadores de local. En su mayoría la cantidad de partidos a considerar son impares, por lo que podría haber un efecto al jugar 4 de local y 3 visita (en el caso de considerar 7 partidos). Por lo que la regresión queda de la siguiente manera, asumiendo que se aplica el efecto fijo tanto de la temporada como del equipo.

$$y_{eq,temp} = \beta_0 + \beta_1 DIF_{eq,temp} + \beta_2 LOC_{eq,temp} + \beta_3 DIF_{eq,temp} * LOC_{eq,temp} + \varepsilon_{eq,temp} \quad (3.3)$$

Por último, se agrega al análisis heterogeneidad de los equipos, como criterio, los resultados obtenidos entre los años 2000 y 2005, a fin de confeccionar una tabla acumulada con los puntajes y dividir los equipos en 3 categorías según se encuentren en los primeros 8 puestos, 8 equipos en los últimos puestos y el resto. En otras palabras, equipos con un muy buen desempeño, equipos débiles y equipos promedios respectivamente. A fin de analizar si es que el efecto altera a todos por igual o no. También se realiza una distinción del comportamiento entre los 4 equipos grandes de la Premier League que dominaron la década del 2000, los también denominados Big Four [37], quienes eran: Arsenal, Chelsea, Liverpool y Manchester United.

Con respecto a lo anterior, también se formulan regresiones con el fin de encontrar si existe una interacción entre el tratamiento que sería la dificultad de partidos según la cantidad de partidos y la caracterización del equipo, agregando la variable $TipoDeEquipo$. Por lo que la regresión se formula de la siguiente manera asumiendo el efecto fijo de temporada y equipos:

$$y_{eq,temp} = \beta_0 + \beta_1 DIF_{eq,temp} + \beta_2 TipoDeEquipo + \beta_3 DIF_{eq,temp} * TipoDeEquipo + \varepsilon_{eq,temp} \quad (3.4)$$

La variable independiente $TipoDeEquipo$ puede estar definida de dos maneras, por ejemplo, si el equipo pertenece a los Big Four o si pertenece a la tabla formada según los resultados por los 5 años anteriores a los datos analizados. Por otro lado, una manera diferente de visualizar la variable $TipoDeEquipo$, es que dependiera también del equipo y temporada, es decir, la variable se va ajustando según los resultados de las 5 temporadas anteriores de manera dinámica por lo que la variable quedaría formulada de la siguiente manera $TipoDeEquipo_{eq,temp}$.

3.2. Desarrollo Metodológico

3.2.1. Descripción de la muestra

Para este caso se consideraron 10 temporadas que van desde los años 2005/2006 hasta 2014/2015, abarcando una cantidad total de 380 partidos con 36 equipos distintos (Anexo:

Equipos de la Premier League entre 2005/2006 hasta 2014/15). Adicionando las 5 temporadas anteriores (2000/2001 hasta 2004/2005) que serán utilizadas para realizar los análisis a las 10 temporadas. En cada temporada se jugó con un formato de 20 equipos cuya modalidad era de dos partidos: uno de local y otro de visita. Hubo 1783 victorias para los equipos locales, 951 empates y 1066 victorias de los equipos que jugaron de visita (Anexo B.6, para ver el detalle).

Como todos los equipos no estuvieron presentes en estas 10 temporadas, se confecciona la tabla 3.1, la cual muestra qué equipos participaron cierta cantidad de temporadas. Cabe recordar que los últimos 3 equipos en la clasificación de cada temporada tienen que jugar la categoría inferior de la siguiente temporada (Championship).

10	9	8	7	5
Aston Villa	AFC Sunderland	Wigan Athletic	Blackburn Rovers	FC Portsmouth
Chelsea FC	FC Fulham		Bolton Wanderers	
FC Arsenal	Newcastle United		Stoke City	
FC Everton	West Ham United		West Bromwich Albion	
FC Liverpool				
Manchester City				
Manchester United				
Tottenham Hotspur				

4	3	2	1
Birmingham City	FC Reading	Charlton Athletic	Cardiff City
FC Middlesbrough	FC Southampton	Crystal Palace	Derby County
Hull City	Norwich City	FC Burnley	FC Blackpool
Swansea City	Queens Park Rangers		FC Watford
	Wolverhampton Wanderers		Leicester City
			Sheffield United

Tabla 3.1: Información cruzada por cantidad de temporadas jugadas por equipo.

Fuente: elaboración propia

Cabe destacar que en promedio existen más goles de los locales que los equipos que juegan

de visita, y esto podría deberse a que los equipos de local realizan más tiros durante los partidos. Esto último puede apreciarse en el siguiente gráfico donde se comparan los tiros que realizan los equipos según su localía y la cantidad de tiros que van directo al arco, los cuales están calculados por fecha. Por último, existen en promedio más goles durante la segunda parte que los primeros 45 minutos en los partidos tanto para los equipos que juegan de local como los de visita, como se evidencia en la tabla 3.2.

Localía\Parte del partido	1er tiempo	2do tiempo
Equipo de Local	0.69	0.85
Equipo de Visita	0.49	0.63

Tabla 3.2: Goles en promedio según localía y tiempo del partido

Fuente: elaboración propia

Con respecto a la posiciones que los equipos van obteniendo en cada fecha según los resultados y cómo éstas van cambiando fecha tras fecha, se calcula la varianza que existe entre la posición de una fecha determinada y la posición al final de la temporada de un equipo. Con estos resultados se calcula el promedio de cada fecha con lo que se obtiene el siguiente gráfico, Figura 3.1.

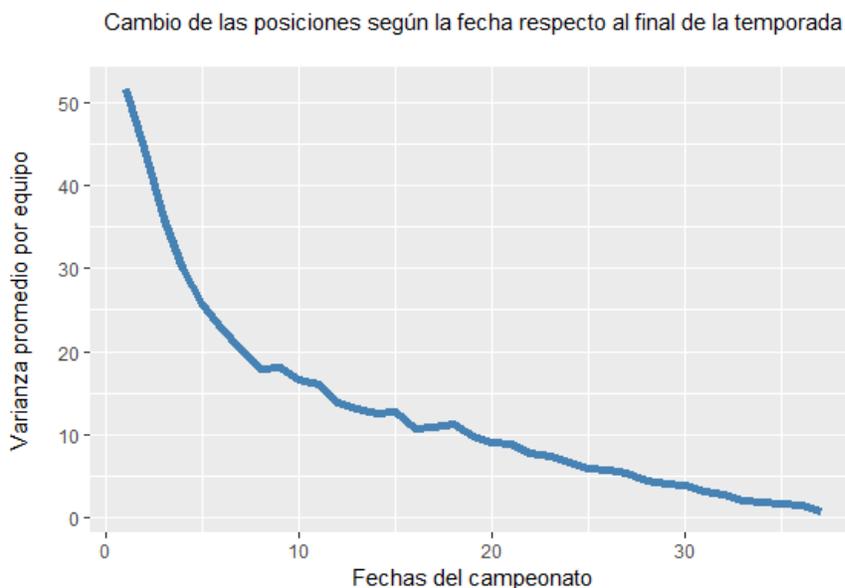


Figura 3.1: Varianza entre posición final y posición en fecha determinada

Fuente: elaboración propia

En el gráfico 3.1 se muestra que entre más se acerca el fin de la temporada, más difícil es cambiar la posición final. Lo cual tiene sentido ya que los puntos que se obtienen en las últimas fechas son marginales respecto a los que se tienen en las últimas fechas. En consecuencia, se puede deducir que los primeros partidos son más relevantes para terminar en lo más alto de la tabla final del campeonato.

Antes de iniciar con los modelos de regresión, se analiza si la dificultad según el criterio, posición final del equipo rival de la temporada anterior es igual en todos los casos o no, usando el promedio de 7 partidos. A continuación, se muestra el gráfico 3.2, que refleja si los equipos reciben la misma “dificultad” en todas las temporadas según los primeros 7 partidos (la dificultad se calcula en promedio).

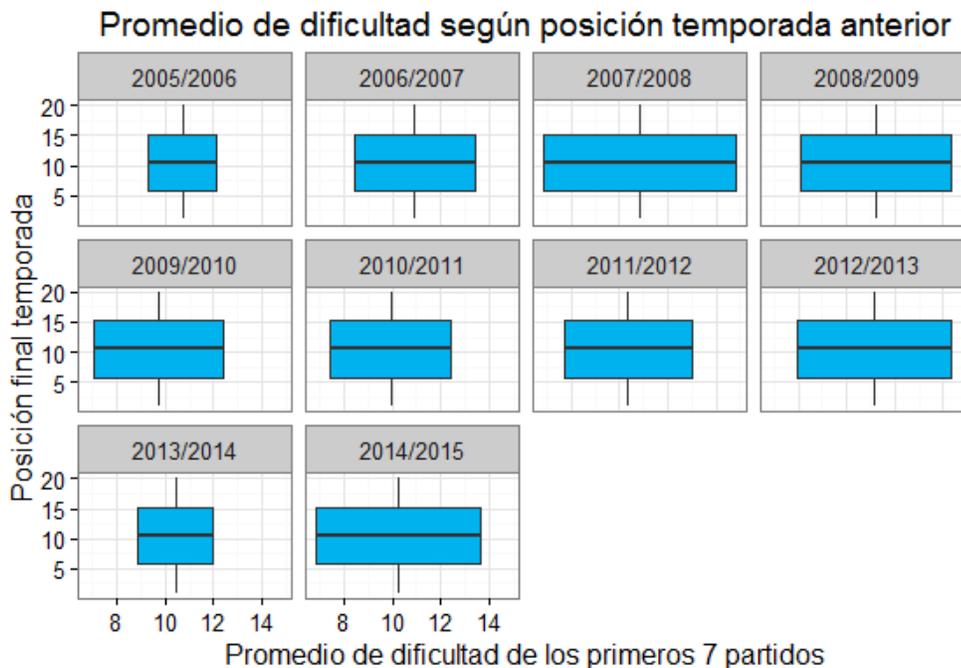


Figura 3.2: Promedio dificultad según cada temporada, usando la dificultad promedio de los primeros 7 partidos

Fuente: elaboración propia

Lo anterior refleja que existen algunas temporadas que tienen más varianza que otras, por lo que muestra la aleatoriedad del sorteo del *fixture* en todas las temporadas. En un caso particular, queda en evidencia que existió poca varianza en promedio en las temporadas 2005/2006 y 2013/2014.

Por otro lado, se analiza el comportamiento que tienen los equipos que han participado en las 10 temporadas en contraste con los equipos que no. Lo anterior queda reflejado en los tres gráficos que se incorporarán luego de la siguiente breve explicación de éstos: el primero muestra a los equipos que siempre participaron, el segundo considera los que participaron entre 9 y 5 temporadas, y el tercer gráfico muestra el resto.

El gráfico 3.3 revela que equipos como el Tottenham Hotspur y el Manchester United no les influye la dificultad en sus primeros 7 partidos, ya que siempre terminan cerca de las mismas posiciones; por el contrario de lo que ocurre con el Aston Villa. El gráfico exhibe las posiciones finales de estos equipos por temporada.

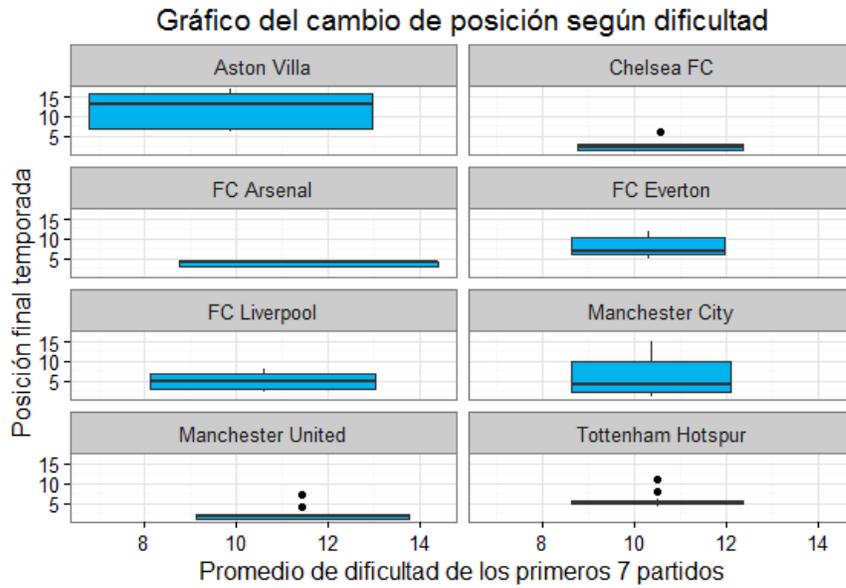


Figura 3.3: Cambio de posición según dificultad de primeros 7 partidos para equipos que han participado las 10 temporadas

Fuente: elaboración propia

Otra forma de apreciar los datos es ir identificando en el mismo gráfico las progresiones que han tenido los equipos a lo largo del tiempo, apreciándose en el gráfico 3.4, por ejemplo, para los casos de los equipos de Manchester United y Tottenham Hotspur se identifican los *outliers* de qué temporada son. Además, de lo irregular que son las campañas del Aston Villa, observándose una pendiente negativa, lo cual explica que entre más difíciles sean los primeros 7 partidos, peor resultado tendrá en la tabla final.

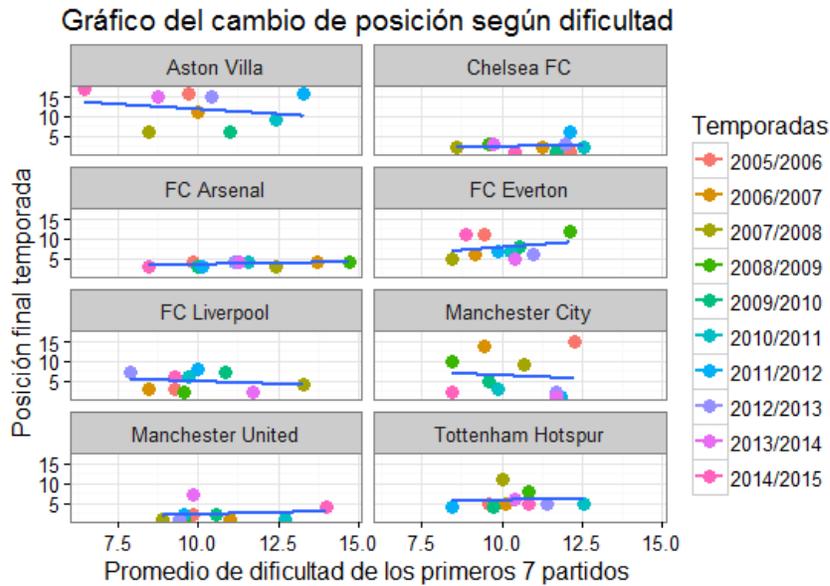


Figura 3.4: Cambio de posición según dificultad por temporada de primeros equipos

Fuente: elaboración propia

Por otro lado, existen los equipos que participaron entre 9 y 5 temporadas, los cuales se muestran en los gráficos de a continuación.

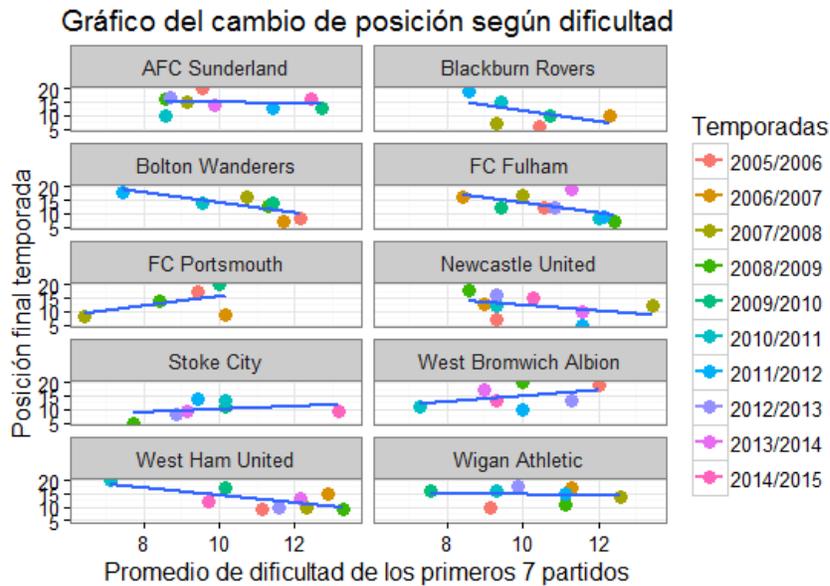


Figura 3.5: Cambio de posición según dificultad por temporada para el segundo grupo.

Fuente: elaboración propia

Como se puede apreciar, se muestra más varianza y distintos comportamientos de los equipos, exceptuando el caso del Wigan Athletic que tiene el mismo comportamiento a pesar

del promedio difícil que fue otorgado en las 7 primeras fechas. Por otro lado, aparece que el FC Portsmouth tiene una tendencia que entre más difícil sea el promedio mejor posición obtiene al final de la temporada, muy por el contrario de lo que sucede con Blackburn Rovers y West Ham United.

Por último, se muestran 2 gráficos el comportamiento del resto de los equipos. Dado lo observado adicionalmente por los que participaron al menos 4 temporadas, se pueden ver distintas tendencias, como por ejemplo, que a ciertos equipos le vaya mejor que a otros por ostentar la calidad de tal y que es necesario considerar el efecto fijo en la regresiones lineales, tanto como para los equipos y temporadas, que vienen en la siguiente parte de esta memoria. Ergo, considerar el efecto fijo en las regresiones es relevante para el análisis.

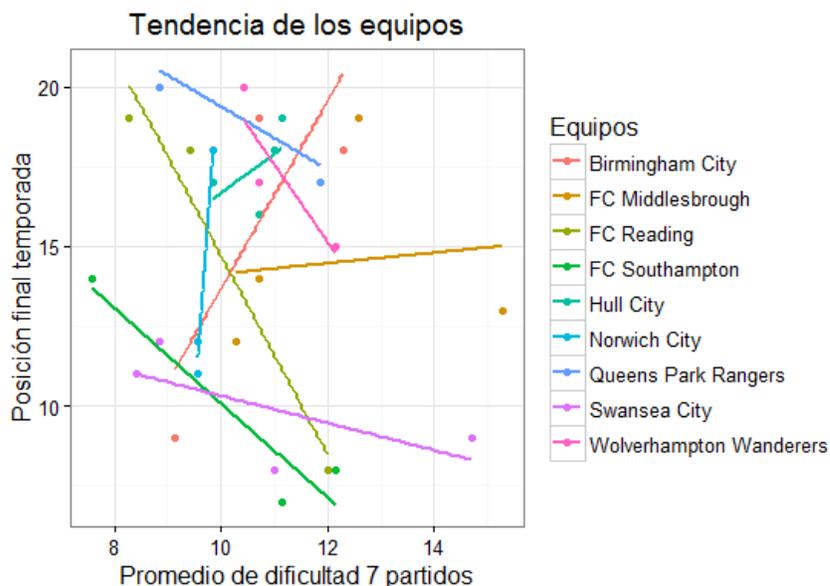


Figura 3.6: Tendencia de los equipos que participaron 3 o 4 temporadas.

Fuente: elaboración propia

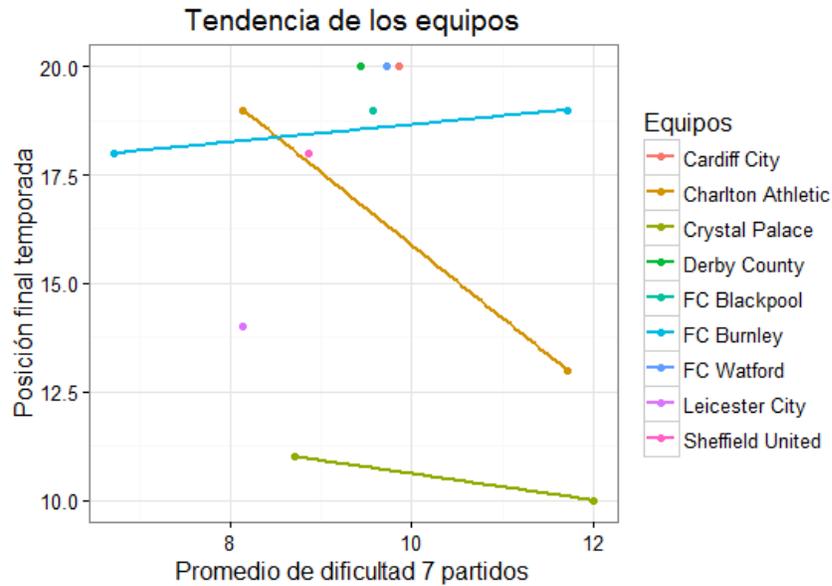


Figura 3.7: Tendencia de los equipos que participaron 1 o 2 temporadas.

Fuente: elaboración propia

En estos últimos casos se obtiene el aprendizaje de los equipos que recién logran ascender a la primera división de la Liga Inglesa con los equipos recién ascendidos. Durante las 10 temporadas a analizar, sólo en una temporada no descendió un equipo recién ascendido y en ningún caso de las 10 temporadas descendieron los 3 equipos recién ascendidos. Se puede pensar que son los rivales fáciles a vencer, pero su comportamiento no demuestra aquello. De hecho, está el caso de equipos que luego de descender una temporada, logran ascender a la temporada siguiente y obtienen resultados finales a mitad de la tabla como, por ejemplo es el caso del Newcastle United que en la temporada 2008/2009 desciende y en la siguiente temporada logra el ascenso para terminar en la posición 12 en la temporada 2010/2011 (en el Anexo B.7, se pueden ver 5 de estos casos).

3.2.2. Propuesta de Modelos y Análisis de Resultados

Al realizar las diversas propuestas de dificultad con las distintas variables dependientes mencionadas anteriormente, se puede observar lo siguiente según cada modelo anteriormente citado en la sección Experimento Natural para el orden de la secuencia de eventos (los detalles de los resultados se pueden observar en Resultados de los modelos propuestos para análisis de la Liga Inglesa en Anexos).

3.2.2.1. Modelo 1: Dificultad por agregación de temporadas

Al considerar la variable dependiente posición final del equipo, se puede apreciar en las regresiones que, al aumentar la agregación de temporadas a la formación de la tabla a analizar, se muestra que la variable discreta es importante en los primeros partidos (primeros 3, 5 y 7 partidos de la temporada). Incluso, en los casos que considera 3 y 4 temporadas anteriores el nivel de significancia, para los primeros 3 partidos con su efecto fijo, es de 0,01. Como se grafica en la siguiente tabla:

Variable independiente \ Regresiones	Modelo con variables continuas		Modelo con variables discretas	
	Con efecto fijo	Con efecto fijo y partidos de local	Con efecto fijo	Con efecto fijo y partidos de local
Intercepto	15.9320 (1.6456)***	15.8278 (1.8048)***	13.9464 (1.4023)***	13.8597 (1.5816)
Primeros 3 partidos	-0.1048 (0.0761)	-0.1049 (0.0763)	1.1087 (0.3862)**	1.1086 (0.3874)**
Cantidad de partidos de local		0.0778 (0.5443)		0.0640 (0.5336)

Entre paréntesis la desviación estándar.

. p < 0,10 , * p < 0,05 , ** p < 0,01, *** p < 0,001

Tabla 3.3: Valores de estimadores considerando 3 temporadas anteriores comparando variables discretas con variables continuas

Fuente: elaboración propia

El contraste de estos 2 tipos de modelo se puede apreciar a continuación en los siguientes gráficos:

Por consiguiente, a través de los resultados obtenidos se puede interpretar que es necesario contar con una historia de los equipos analizados ya que, si un equipo “bueno” tiene una “mala” temporada, esto no se verá reflejado en la tabla dada la cantidad de datos disponibles (más de una temporada). Esto se realiza con el fin de no tener casos fuera de serie y evitar sesgos que puedan haber ocurrido en una temporada en específico. Por estas razones, es que se usa el promedio de las 3 o 4 temporadas anteriores.

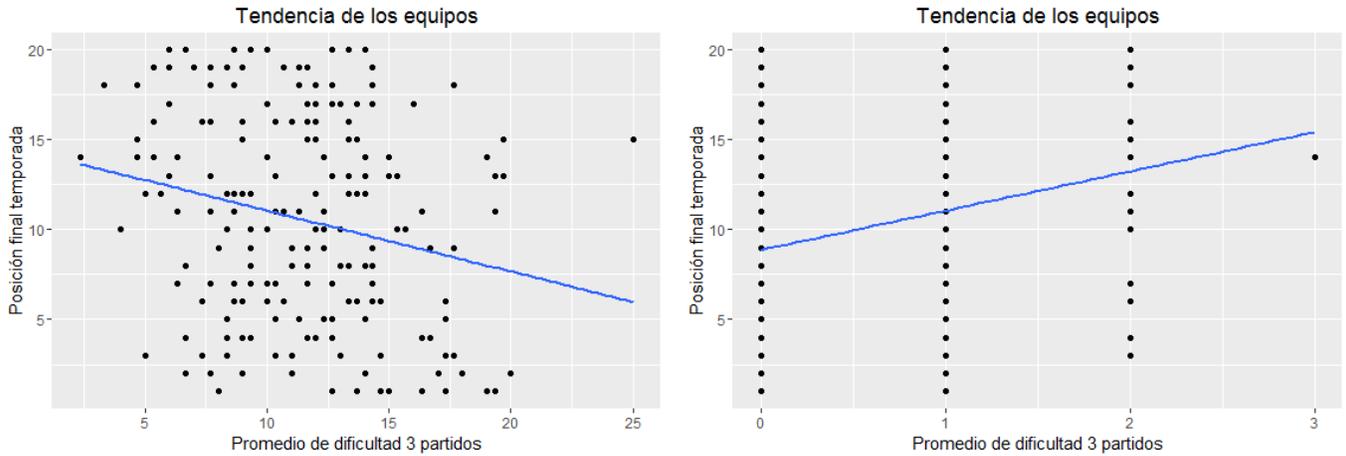


Figura 3.8: Constraste del análisis entre variables continuas y discretas

Fuente: elaboración propia

Variable independiente \ Regresiones	Modelo con variables continuas		Modelo con variables discretas	
	Con efecto fijo	Con efecto fijo y partidos de local	Con efecto fijo	Con efecto fijo y partidos de local
Intercepto	15.7190 (1.6417)***	15.6248 (1.8059)***	13.97084 (1.40716)***	13.7782 (1.5937)***
Primeros 3 partidos	-0.0851 (0.0742)	-0.0851 (0.0744)	1.05139 (0.38668)**	1.0562 (0.3883)**
Cantidad de partidos de local		0.0694 (0.5453)		0.1394 (0.5355)

Entre paréntesis la desviación estándar.

. p < 0,10 , * p < 0,05 , ** p < 0,01, *** p < 0,001

Tabla 3.4: Cuadro comparativo que considera 4 temporadas anteriores, variables discretas con variables continuas

Fuente: elaboración propia

Por otro lado, al considerar la variable dependiente puntaje final obtenido al término de la temporada se llega a visualizar resultados estadísticamente significativos, tomando en cuenta las 4 y 5 temporadas anteriores donde destacan los últimos 3 partidos de la primera fase con un p-valor menor a 0,05; el que se demuestra en la tabla 3.5:

Para este caso, se interpreta que los puntos obtenidos al final de la primera fase son importantes, a pesar de que falten otras 19 fechas. Esto deja abierta la posibilidad de realizar el análisis para torneos más breves.

Variable independiente \ Regresiones	Modelo 4 temporadas anteriores	Modelo 5 temporadas anteriores
Intercepto	34.472 (4.348)***	34.295 (4.384)***
Últimos 3 partidos primera fase	0.368 (0.185)*	0.367 (0.183)*

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.5: Valores de estimadores respecto a distinta cantidad de temporadas

Fuente: elaboración propia

De la variable dependiente “posición a la fecha número 20” destaca tanto la temporada anterior como 2 temporadas anteriores, tanto como el caso continuo como el caso discreto en los últimos 3 partidos de la primera fase, los cuales son estadísticamente significativos. Agregándose también el criterio que considera las últimas 4 temporadas sólo en el criterio continuo. Estos se reflejan en las tablas 3.6 y 3.7:

Variable independiente \ Regresiones	Modelo 1 temporada anterior		Modelo 2 temporadas anteriores	
	Variable continua	Variable discreta	Variable continua	Variable discreta
Intercepto	16.7893 (1.9380)***	13.946 (1.671)***	16.7795 (1.8733)***	13.7580 (1.6791)***
Últimos 3 partidos de la primera fase	-0.2417 (0.1107)*	0.805 (0.442).	-0.2427 (0.0993)*	0.9238 (0.4555)*

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.6: Estimadores según variable dependiente posición a la fecha numero 20 con sus efectos fijos

Fuente: elaboración propia

Como se puede observar, los resultados ratifican lo interesante que es considerar análisis para torneos que tengan menos fechas.

Por último, al considerar la variable dependiente “puntaje obtenido a mitad de temporada” son estadísticamente significativos de manera continua para los 4 primeros casos exceptuando el de las últimas 5 temporadas. Se muestran los más significativos estadísticamente a continuación (tabla 3.8 y 3.9):

De estos análisis se puede discernir que la variable “cantidad de partidos de local” se

Variable independiente \ Regresiones	Modelo con efecto fijo
Intercepto	16.2251 (1.8770)***
Últimos 3 partidos de la primera fase	-0.1494 (0.0797).

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.7: Estimadores según variable dependiente posición a la fecha numero 20 según desempeño últimas 4 temporadas

Fuente: elaboración propia

Variable independiente \ Regresiones	Modelo 1 temporada anterior	Modelo 2 temporadas anteriores
Intercepto	16.4209 (2.8415)***	15.9250 (2.7366)***
Últimos 3 partidos de la primera fase	0.2790 (0.1623).	0.3345 (0.1451)*

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.8: Estimadores según variable dependiente puntaje a la fecha numero 20 según desempeño temporada anterior y dos temporadas anteriores

Fuente: elaboración propia

Variable independiente \ Regresiones	Modelo 1 temporada anterior	Modelo 2 temporadas anteriores
Intercepto	16.4209 (2.8415)***	15.9250 (2.7366)***
Últimos 3 partidos de la primera fase	0.2790 (0.1623).	0.3345 (0.1451)*

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.9: Estimadores según variable dependiente puntaje a la fecha numero 20 según desempeño 4 y 5 temporadas anteriores

Fuente: elaboración propia

jugaron dentro del análisis no son importantes para los casos anteriormente analizados. Por ende, se puede concluir que la dificultad está descrita por el valor que obtiene el rival y no por si se juega de local o visita como uno tendería a concluir con respecto a la bibliografía (v.g. *home-advantage*). Por otro lado, los partidos que son importantes o interesantes a tomar en cuenta son los primeros y los últimos 3 de la primera fase. Más aun, en relación a esto

último, sería interesante realizar un análisis para torneos que tengan menos partidos por equipo. Además, cabe agregar que la variable continua al considerar la variable dependiente “posición al final de la temporada” tiene pendiente negativa por lo que explica que también se puede terminar al final de la temporada según los partidos, por el contrario, la variable discreta explica que tan mal puede terminar el equipo al final de la temporada. Luego, la variable discreta es útil para los casos de inicio de temporada al observar su posición al final de la temporada. En cambio, para la variable continua se puede interpretar que los puntos obtenidos son esenciales y explican el puntaje que vamos a obtener. Con respecto a las otras 2 variables dependientes, se puede inferir que en un campeonato corto las últimas fechas son relevantes tanto para su puntaje y su posición final.

3.2.2.2. Modelo 2: Dificultad por rendimiento de local y visita

En el caso de separar la tabla final de posiciones según desempeño de los equipos cuando juegan de local y de visita, tanto como al considerar 1 temporada anterior y 2 temporadas anteriores destaca el caso continuo para los primeros 7 primeros partidos de manera estadísticamente significativa (p-valor<0,05).

Para el caso discreto coinciden en los 2 casos los primeros 5 partidos. Además, destaca el caso de los primeros 7 partidos al considerar sólo 1 temporada anterior. Todas estas conclusiones son referentes a la variable dependiente “posición al final de la temporada”. Al tomar en cuenta los últimos 3 partidos de la primera fase destaca que sólo la temporada anterior sea estadísticamente significativa tanto en la variable continua como discreta.

En el resto de las variables dependientes, en tanto, no se encuentran resultados concluyentes. A continuación, se muestra en las siguientes tablas (3.10 y 3.11) los datos más destacados:

Variable independiente \ Regresiones	Modelo 1 temporada anterior		Modelo 2 temporadas anteriores	
	Variable continua	Variable discreta	Variable continua	Variable discreta
Intercepto	18.0431 (1.9897)***	13.6260 (1.4937)***	18.0220 (2.0072)***	12.441 (2.470)***
Primeros 7 partidos de la temporada	-0.3572 (0.1545)*	0.5646 (0.2738)*	-0.3147 (0.1392)*	1.054 (0.930) ¹

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.10: Estimadores según variable dependiente posición al final de la temporada, según desempeño de local y de visita

Fuente: elaboración propia

¹Al considerar de manera discreta los 6 primeros equipos en vez de los 5, la variable de los primeros 7 partidos tiene un valor beta estimado de 0,48494 (0,25060) con un p-valor menor a 0,1.

Variable independiente \ Regresiones	Caso Continuo	Caso Discreto
Intercepto	16.0081 (1.5823)***	13.9822 (1.4623)***
Últimos 3 partidos de la primera fase	-0.1501 (0.0881).	0.6677 (0.3790).

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.11: Estimadores según variable dependiente posición al final de la temporada, según desempeño de local y de visita, solo considerando 1 temporada anterior

Fuente: elaboración propia

3.2.2.3. Modelo 3: Criterio considerando a últimos equipos de la tabla (“fácil”)

Otra forma de apreciar la dificultad es analizando si los equipos enfrentan a quienes que tienen peor rendimiento y/o los equipos recién ascendidos a la división de honor. Los datos muestran que no se pueden obtener resultados estadísticamente significativos al momento de considerar una temporada anterior (tanto como su clasificación final como la tabla según desempeño de local y visita).

Por otro lado, al considerar las 2 temporadas anteriores y seleccionar los últimas 6 como equipos fáciles en la tabla acumulada, los datos se interpretan de la siguiente manera: si jugar con equipos recién ascendidos (al menos 3) y equipos con mal desempeño respecto a las últimas 2 temporadas, influyen en el resultado final. Luego, en las regresiones se destaca en general en las 4 variables dependientes para los últimos 3 partidos de la primera fase y los últimos 5 también de la primera. Los resultados sobre los últimos 3 partidos de la primera fase se muestran a continuación:

Variable independiente \ Variable Dependiente	Variable Dependiente Posición final	Variable Dependiente Puntaje final
Intercepto	15.3 (1.42)***	37.079 (3.858)***
Últimos 3 partidos de la primera fase	-0.792 (0.418).	2.514 (1.136)*

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.12: Estimadores según dificultad de manera discreta considerando los 6 últimos equipos de la tabla acumulada (considerando las 2 temporadas anteriores)

Fuente: elaboración propia

Por último, al separar la tabla acumulada según el desempeño de las 2 temporadas anteriores por rendimiento de local y de visita (considera equipos que tuvieron mal rendimiento la temporada anterior y equipos recién ascendidos), destaca el caso de los últimos 5 partidos de la primera fase para 3 variables dependientes exceptuando la variable dependiente “posición

Variable independiente \ Variable dependiente	Posición Intermedio	Puntaje Intermedio
Intercepto	15.0819 (1.6717)***	17.9344 (2.4163)***
Últimos 3 partidos de la primera fase	-0.8867 (0.4923).	1.7375 (0.7116)*

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.13: Estimadores según dificultad de manera discreta considerando los 6 últimos equipos de la tabla acumulada (considerando las 2 temporadas anteriores)

Fuente: elaboración propia

a mitad de temporada”, resultados que son estadísticamente significativos al 0,1% (Ver en Anexos C.1).

En síntesis, jugar contra equipos llamados “débiles” no son importante para los primeros partidos, pero donde sí se muestra un efecto es al considerar los últimos partidos de la primera fase.

3.2.2.4. Modelo 4: Considerando datos entre temporadas 2000/2001 y 2004/2005

Con respecto a los resultados del presente modelo no se logra encontrar un patrón como el de los casos anteriormente mencionados. Los resultados de las variables de interés tienen más varianza que el mismo estimador por lo que lograr una conclusión no es trivial. Solo en 2 casos se muestra un resultado significativo: (i) para el caso de la variable continua se refleja en los primeros 5 partidos de la segunda fase para la posición al final de los equipos y (ii) para el caso discreto destaca los partidos de la fecha 19, 20 y 21 para el puntaje final de la temporada. Para este último, nos explica que es necesario tener retroalimentación de los resultados para temporadas pasadas y que el comportamiento de los equipos va cambiando a medida que se obtienen buenos y/o malos resultados, y no tomando en cuenta las temporadas fijas. Existen casos donde los equipos tienen más recursos y pueden invertir más dinero como el caso de Manchester City o que un equipo logre contratar a un entrenador muy exitoso y que logre dirigir muchas temporadas, como lo fue el caso del entrenador Alex Ferguson, en el Manchester United. Estos casos mostrados son entre varios posibles que pueden ocurrir, por lo que es importante considerar temporadas anteriores y no temporadas fijas pues pueden existir sesgos.

3.2.3. Heterogeneidad

Al realizar las distintas regresiones con los efectos fijos, se observa que algunos equipos les influye más su posición final o puntaje final según el equipo que son. Por lo que un análisis de este tipo podría permitir identificar qué patrón afecte más que a otros y, además, resolver

la interrogante si es que la dificultad altera de la misma manera a todos los equipos.

Así, en primera instancia, se segrega en los 3 grupos mencionados en la sección de metodología, obtenidos a partir de los resultados de la tabla acumulada entre las temporadas 00/01 y 04/05. Donde el grupo 2 considera los primeros 8 equipos de dicha tabla, los últimos 8 conforman el grupo 0 y el resto conforma el grupo 1. Cabe destacar que dentro de las 5 temporadas anteriores se consideran 29 equipos, de los cuales 4 no están en las siguientes 10 temporadas (Leeds United, Ipswich Town, Coventry City y Bradford City) y, además, dentro de las 10 temporadas se agregan 11 equipos. Por lo que, al no estar dentro de los 29 de equipos de clasificación, estos 11 equipos se consideran que están en el grupo 0 en primera instancia. Luego, los resultados en las regresiones no son significativos; pero, al clasificar los equipos, se muestra una tendencia que al ser de los de más arriba de la tabla mencionada, les permite terminar mejor en la tabla de posiciones al final de la temporada pues el p-valor del tipo de equipo tiene un valor de 0,132. Por otro lado, con respecto a la interacción de las 2 variables mencionadas anteriormente, no existe claridad estadísticamente dado que el estimador es muy pequeño y la desviación estándar grande con un p-valor de 0,836.

	Modelo con heterogeneidad
Intercepto	17.5216 (2.2009)***
Primeros 3 partidos	1.2173 (0.6524).
Heterogeneidad	-3.5782 (2.3606)
Interacción: dificultad y tipo de equipo	-0.0956 (0.4622)

Entre paréntesis la desviación estándar.

. p <0,10 , * p <0,05 , ** p <0,01, *** p <0,001

Tabla 3.14: Resultados de los estimadores de la regresión, con heterogeneidad fija (temporadas 2000 a 2005)

Fuente: elaboración propia

Los resultados también se pueden apreciar de manera gráfica, fabricándose una línea de tendencia para cada equipo:

Lo anterior muestra que, en general, los equipos con mejor rendimiento durante las 5 temporadas mantienen la tendencia de buen rendimiento la cual se refleja en las 10 siguientes, pero también existen equipos que empeoraron su desempeño en estas 10 temporadas incluso terminando en posiciones muy cercanas al descenso (bajo posición 15). Por lo cual, los equipos de factor 2 tienden a tener más varianza. Se destaca que la tendencia, tanto de los de tipo 1 como 2, tienen una misma dirección, pero desde un punto de referencia distinto, lo cual es diferente para los equipos de tipo 0, ya que se podría decir que los equipos que ascienden y juegan contra “difíciles” al inicio de la temporada no afecta en gran medida ya que, por estar en una nueva división, todos los partidos son igual de difíciles para mantener la categoría de honor.

Además, a partir de los resultados se puede interpretar que agregar la variable “tipo de

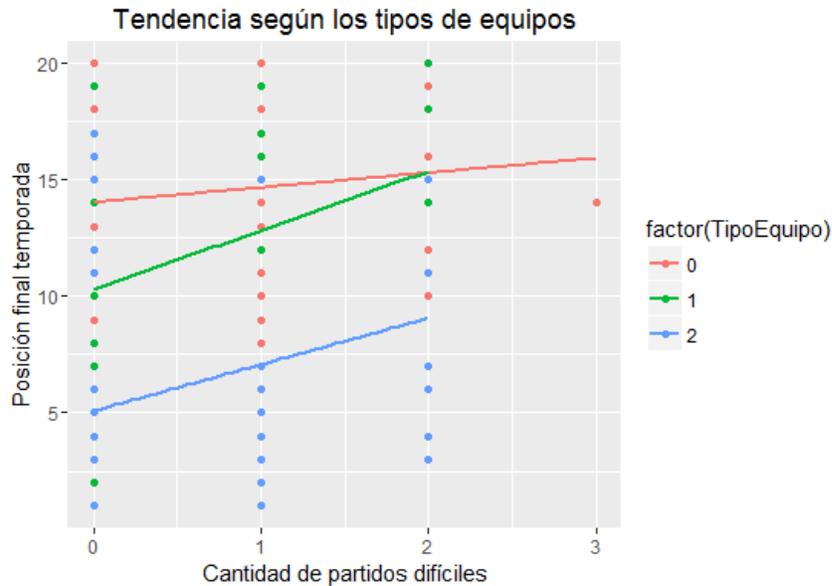


Figura 3.9: Posición final de la temporada según dificultad discreta por tipo de equipo

Fuente: elaboración propia

equipo”, de manera categórica, no es la adecuada ya que los equipos pueden pasar por buenas temporadas, crisis, o fracasos, los cuales pueden explicar que en 10 temporadas hayan existido 36 equipos participando. A modo de ejemplo, se encuentran el Stoke City y Wigan Athletic, equipos que no jugaron ningún partido entre las temporadas 2000/2001 y 2004/2005 (ambos equipos jugaron en la segunda división de Inglaterra, *Football League Championship*) pero en las 10 temporadas siguientes participaron en 7 y 8 respectivamente.

Dado que el criterio anterior no permite considerar los equipos objetivamente según sus rendimientos, se propone otra norma para crear la variable “tipo de equipo” con el fin de analizar si existe una cierta cantidad de partidos que afecte más a otro tipo de equipos, tomando casos en cuenta como Stoke City y Wigan Athletic. El nuevo criterio es, si pertenece a los primeros 7 equipos de esta lista o no, y se considera de tal manera porque son los que pueden clasificar a copas europeas. Este caso se puede dar ya que, si de los 5 que clasifican directo, se encuentra el campeón de Copa de Liga y el campeón de la FA Cup (las cuales otorgan un cupo cada una a la Europa League), se clasifican los que están en el puesto 6 y 7 al final de la Liga Inglesa. Los equipos seleccionados son: Manchester United, Arsenal, Liverpool, Chelsea, Aston Villa, Newcastle United y Tottenham Hotspur. Los cuales demuestran, en la tabla y en el gráfico, que son determinantes para obtener una mejor posición con respecto al resto de los equipos.

En el gráfico 3.10 se puede apreciar que en los equipos elegidos también existen algunos que tienen mal rendimiento dentro de las 10 temporadas, como es el caso de Newcastle United y Aston Villa. Asimismo, se dejan equipos fuera de esta selecta lista como es el caso de equipos como Everton y Manchester City. Pero demuestra que existe una brecha entre ambos grupos, muy significativa, que los distancia en promedio alrededor de nueve puestos y que la variable de los primeros 3 partidos afecta de igual forma a todos los equipos, pero la posición final de la temporada tiene una referencia distinta para cada tipo de equipo.

	Modelo considerando 7 primeros equipos
Intercepto	13.6400 (1.4048)***
Primeros 3 partidos	1.6106 (0.4819)**
Heterogeneidad (tipo de equipo)	-7.8934 (1.6731)***
Interacción: dificultad y tipo de equipo	-1.4232 (0.8265).

Entre paréntesis la desviación estándar.

. p < 0,10 , * p < 0,05 , ** p < 0,01, *** p < 0,001

Tabla 3.15: Resultados de los estimadores de la regresión con dificultad discreta considerando heterogeneidad para 7 equipos

Fuente: elaboración propia

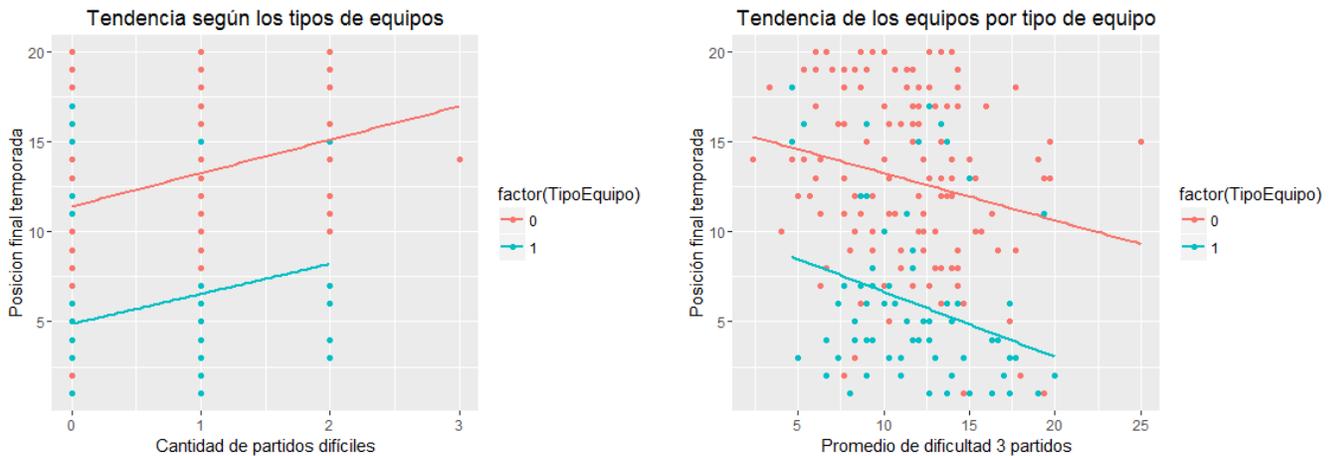


Figura 3.10: Contraste del análisis entre variables continuas y discretas con heterogeneidad

Fuente: elaboración propia

Luego, dentro de los 7 equipos anteriormente mencionados, se encuentran los denominados “Big Four”. Al considerar la heterogeneidad de sí pertenece al selecto grupo de los 4 grandes, los denominados “Big Four” obtienen los siguientes resultados (se aplica al modelo que considera las 3 temporadas anteriores con el modelo discreto):

La tabla 3.16 nos muestra que pertenecer a este grupo es determinante para la posición final, incluso mejora al menos 11 puestos de los 13 o 14 que empiezan en promedio todos los equipos. Pero la dificultad en el tratamiento para estos equipos no es significativa por lo que no se puede concluir que la dificultad afecte más a algunos equipos que a otros. Otra forma de mirar el resultado es a través del gráfico 3.11.

Aquí, el factor 1 representa a los mencionados 4 equipos, apreciándose que tienen menos varianza en su posición y el resto, en cambio, varía mucho más y, además, que sus posiciones

	Modelo considerando "Big Four"
Intercepto	13.8993 (1.4071)***
Primeros 3 partidos	1.2313 (0.4334)**
Tipo de Equipo: Big Four	-11.9241 (1.6692)***
Interacción: dificultad y tipo de equipo (Big Four)	-0.6356 (1.0122)

Entre paréntesis la desviación estándar.

. p < 0,10 , * p < 0,05 , ** p < 0,01, *** p < 0,001

Tabla 3.16: Resultados de los estimadores de la regresión con dificultad discreta considerando heterogeneidad para 4 equipos (Big Four)

Fuente: elaboración propia

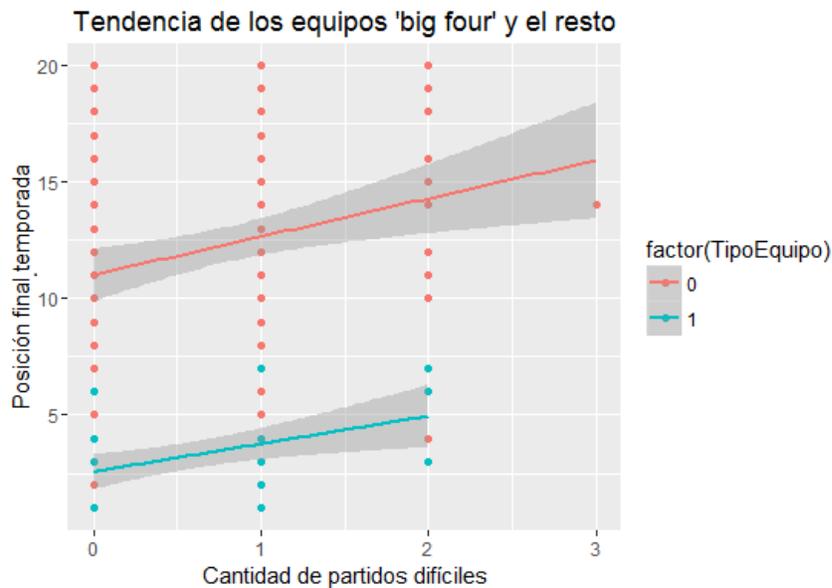


Figura 3.11: Tendencia de equipos, dependiendo si pertenecen al grupo Big Four o no

Fuente: elaboración propia

cambian de manera significativa.

Por lo anteriormente mencionado se puede señalar que los equipos “grandes” y exitosos en las últimas temporadas están definidos por su historia. Pero, al respecto, solo destacaron 7 equipos según los criterios establecidos. A modo recordatorio, los equipos que logran llegar a la serie de honor lo hacen por periodos breves y otros por varias temporadas. Por tanto, es interesante poder analizar y modelar estos casos. Para aquello se propone emplear que la variable “tipo de equipo” se vaya ajustando en el tiempo, o sea de manera heterogeneidad dinámica, Para el caso particular se al consideranr 4 y 3 temporadas anteriores a la que

es objeto de análisis, aplicado al caso de los primeros 3 partidos con variable dependiente “posición final al término de la temporada”, el cual destacó por significancia ($p < 0,01$). Y las tablas construidas se dividen en 3 grupos del mismo tamaño, siendo los primeros equipos los que tienen rendimiento sobresaliente, el segundo grupo con rendimiento normal y el tercero con rendimiento irregular.

Luego, al tomar en cuenta las 3 temporadas anteriores a la temporada a analizar destaca de manera significativa la variable *tipo de equipo* (sin considerar la interacción entre la dificultad y el tipo de equipo), pero al agregar la interacción con la dificultad, ésta deja de ser estadísticamente significativa, lo cual se muestra en la siguiente tabla.

	Modelo efecto fijo y heterogeneidad dinámica
Intercepto	14.22085 (1.45904) ^{***}
Primeros 3 partidos	1.24468 (0.61745) [*]
Tipo de Equipo	-0.88934 (0.70624)
Interacción dificultad y tipo de equipo	-0.17230 (0.49514)

Entre paréntesis la desviación estándar.

. $p < 0,10$, * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 3.17: Resultados de los estimadores de la regresión con dificultad discreta considerando heterogeneidad dinámica

Fuente: elaboración propia

Los tres grupos se diferencian en el punto de partida de sus posiciones, en promedio por más de 1 puesto, lo que se refleja en el siguiente gráfico donde se separan por los 3 grupos, pero la diferencia de posiciones entre equipos se muestra más atenuada sobre todo entre el grupo uno y dos. Sin embargo, no existe en esta tabla evidencia de que la dificultad afecte más a ciertos equipos que a otros. Por otro lado, pertenecer a cierto tipo de equipo puede mejorar en promedio la posición final de la temporada. Lo anterior se puede apreciar de manera gráfica de la siguiente manera:

Como se puede ver en el gráfico las tendencias son distintas, pero con pendiente positiva las 3 rectas. Cabe destacar que para los equipos del grupo 1 se puede ver que tiene mayor pendiente, con respecto a los grupos 0 y 2. Al respecto, del grupo 0 señalamos que son los recién ascendidos a la serie de honor, los cuales al estar en primera división su importancia es mantener la categoría, por lo que no influye tanto la posición según sus rivales, como sí es el caso del grupo 1, que evidencia que llevan más temporadas en la Liga Inglesa, por lo que a ellos les influye contra qué equipos les toca jugar. Esto puede ser causado por el conocimiento de los planteamientos de los partidos o, para el caso del grupo 0, el desconocimiento de cómo juegan o la motivación que tienen los equipos recién ascendidos al lograr buenos resultados. Con respecto a los equipos del grupo 2, estos tienen más estabilidad por sus buenos resultados, por lo que es difícil que ellos reciban tantos cambios en promedio de posición.

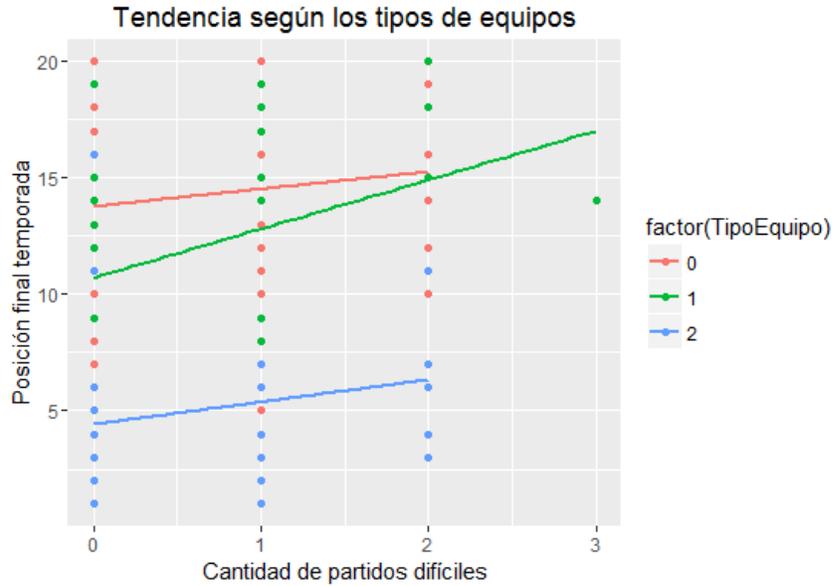


Figura 3.12: Tendencia de equipos según heterogeneidad dinámica propuesta

Fuente: elaboración propia

Por último, se realiza el mismo análisis anterior, pero para el caso de considerar las últimas 4 temporadas con el criterio discreto definido en la sección Anexos D.1.

3.2.4. Conclusiones y Aprendizajes

Con respecto a las conclusiones, se destaca que hay muchas maneras de modelar la dificultad en un campeonato de fútbol, por lo que es complejo establecer cuál es la más indicada. Sin embargo, al mismo tiempo es importante observar si el resultado o la dirección del resultado cambia según el criterio con el que se decidió modelar la dificultad. Asimismo, al haber aplicado distintas formas en el desarrollo de la memoria, se concluye que modelar la dificultad de manera discreta evidencia resultados estadísticamente significativos, en contraste, al modelamiento continuo.

Además, por un lado, es conveniente agrupar por temporadas, al menos tres, para poder explicar en qué situaciones y/o circunstancias el rival tiene buen rendimiento o no. Esto permite evitar sesgos en caso de que hubiese un problema específico en el equipo dentro de una temporada determinada. Por otro lado, las regresiones señalan que existe una importancia notoria en los primeros partidos, vale decir, cuando la variable dependiente es la posición final. Por ende, se interpreta que efectivamente los primeros partidos influyen en la posición final de los equipos, ya que la posición final al término de la temporada es la más importante para los equipos, debido a que esta señala que equipos juegan torneos internacionales, quienes descienden y quién es campeón del torneo.

Adicionalmente debido a la cantidad de partidos y puntos que existen, se puede concluir que el efecto de la dificultad de los partidos en el tramo final de la temporada no tiene un

efecto relevante para la posición final del equipo, ya que el cambio de posición y de puntos es prácticamente marginal. Al mismo tiempo, el efecto de la temporada no es determinante para efectos de la posición final. A pesar de que a través del análisis se sintetiza que en algunos casos si son importantes los últimos partidos de la primera fase, es por ello que se espera que este efecto pueda ser mayor para torneos cortos (menos fechas).

Al formular la dificultad según desempeño de localía y visita, los resultados no muestran datos estadísticamente significativos, aunque la intuición dice que entre más detallado sea la forma de modelar la dificultad, se lograrían conclusiones más precisas. Lo anterior es consecuencia de que la dificultad de un partido depende del rendimiento en general del equipo y no según su desempeño de local o visita. Otra conclusión importante es que no se puede saber si rivales fáciles son aquellos que son los recién ascendidos, esto se pudo ver reflejado al observar cuando el enfrentamiento es contra equipos “fáciles”.

Con respecto al agregar la variable tipo de equipo a la regresión se concluye que el efecto de dificultad afecta a todos por igual, pero la diferencia radica en que por ser de un tipo de equipo que tiene mejor rendimiento según su historia, tiende a terminar en mejor posición.

Por último, al crear la variable tipo de equipo de manera dinámica, es decir, la categoría cambia según resultados anteriores, no existe mayor variabilidad en los equipos. Por lo que, en definitiva, es mejor definir de manera dinámica según las temporadas anteriores y no según ciertas temporadas de manera categórica. Sin embargo, al considerarlo de la manera mencionada puede que exista endogeneidad en el análisis.

Capítulo 4

Conclusiones Generales y Trabajos Futuros

4.1. Conclusiones Generales

En síntesis, para el primer experimento que no existen diferencias estadísticamente significativas al contrastar ambos grupos. Aunque, en las regresiones en la mayoría de los casos tiende a favorecer al grupo que pronostica una sola vez al comienzo (grupo One Shot). Por otro lado, al aplicar heterogeneidad cambia la tendencia en algunos casos, por lo que se concluye en algunos casos que al ser fanático es preferible pertenecer al grupo Day by Day que al grupo One Shot para mejorar la predicción.

En relación a los algoritmos que en la presente memoria fueron las casas de apuestas (*Bet365* y *Bwin*). Las diferencias en las predicciones no marcaron tendencia que favorece o no al realizar un contraste con los participantes, por lo que los resultados no son concluyentes. Asimismo, al considerar los dos grupos experimentales usando solo predicciones de los algoritmos tampoco hubo un mejor grupo que otro para ambos casos, ratificando la complejidad de realizar este tipo de predicciones.

Para finalizar, con el primer experimento, al aplicar los modelos de solución para el *attrition*, no queda demostrado que sea mejor un grupo que otro. Por lo que contrasta con lo mencionado al inicio de la memoria sobre las políticas *closed-loop*.

Para el segundo experimento, con respecto al output final de los equipos, el más importante es la posición al final de la temporada, ya que estas definen qué sucede con cada equipo para la temporada que viene, vale decir, descender, campeónar y/o participar en copas europeas.

Debido a la cantidad de partidos que existe y la cantidad de puntos que se puede obtener adicional de los que ya se disponen, efecto de la dificultad de los partidos no tiene incidencia al final de la temporada, debido a que no influye la dificultad al final del campeonato, ya que el cambio de posición y de puntos es marginal. Por lo tanto, el efecto de la temporada no es determinante para definir la posición final del equipo.

Así, el equipo que recibe la secuencia es un factor muy importante, por lo que el efecto del orden en el calendario afecta principalmente a los equipos que compiten por mantener la categoría, ya que lograr un puesto más, logra el objetivo del equipo. Debido a esto es que el hecho de ser equipos de ciertas categorías, no establece que pertenecer a un equipo débil este salga más afectado, pues la dificultad afecta por igual a todos los equipos. Además, con respecto al agregar la variable “tipo de equipo” al análisis, lo destacable es el hecho de que tiene que ser dinámica ya que los equipos pasan por temporadas y periodos, y al menos es necesario considerar 3 temporadas o 4 temporadas al momento de modelar. Con esto se evita el hecho de que un equipo tenga una temporada “mala” y sesgue el estudio. Por último, se espera que para los torneos con menos fechas, sean otras fechas las que influyen en el resultado final y no las mismas que en el estudio realizado en la Liga Inglesa.

4.2. Trabajos Futuros

A raíz de los dos experimentos realizados en la presente memoria, se proponen, para el experimento online para decisiones anticipadas secuenciales, (i) realizar un experimento con otra situación considerando, además, un número más grande de participantes y (ii) evitar con más herramientas el caso del *attrition* a través de recordatorios vía correos electrónicos. Por otro lado, se sugiere analizar las correlaciones de los aciertos de los participantes del Day by Day, si el acierto o desacierto en rondas previas influye en responder al día siguiente los pronósticos. Otro punto importante en el experimento, es analizar el caso hipotético de que los participantes, que no hayan recibido el mensaje, tengan un efecto causal.

En referencia, al experimento natural para el orden de la secuencia de eventos, se propone que se realice el estudio en otras ligas, con el fin de observar si el comportamiento es similar en las 5 ligas principales de Europa, como ejemplo. Por otro lado, un análisis importante sería ver si el efecto encontrado en esta memoria es similar si el torneo tuviese menos fechas.

En seguida, esta conclusión puede variar en cierto grado cuando los torneos sean más cortos. Como por ejemplo es el caso del campeonato de fútbol chileno donde sólo existen 15 fechas para determinar al campeón del torneo en contraste con los 38 de la liga inglesa o las 34 de la liga alemana. Otro factor importante a considerar son los incentivos que existen al respecto, como el hecho de que en la última fecha para torneos cortos sean más emocionantes por la diferencia de puntos, contra el hecho de que en los torneos largos ya se sepa quién sale campeón.

Bibliografía

- [1] Michael J Fry and Jeffrey W Ohlmann. Introduction to the special issue on analytics in sports, part i: General sports applications, 2012.
- [2] Alan Nevill, Greg Atkinson, and Mike Hughes. Twenty-five years of sport performance research in the Journal of Sports Sciences. *Journal of sports sciences*, 26(4):413–426, 2008.
- [3] Manuel JA Eugster, Jan Gertheiss, Sebastian Kaiser, et al. Having the second leg at home—advantage in the UEFA Champions League knockout phase. *Journal of Quantitative Analysis in Sports*, 7(1):1–11, 2011.
- [4] Philip A Scarf and Muhammad Mat Yusof. A numerical study of tournament structure and seeding policy for the soccer World Cup Finals. *Statistica Neerlandica*, 65(1):43–57, 2011.
- [5] Richard Pollard. Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, 4(3):237–248, 1986.
- [6] Sion Thomas, Colin Reeves, and Simon Davies. An analysis of home advantage in the English Football Premiership. *Perceptual and motor skills*, 99(3 suppl):1212–1216, 2004.
- [7] Barry Schwartz and Stephen F Barsky. The home advantage. *Social forces*, pages 641–661, 1977.
- [8] James C Moore and Jody Brylinsky. Facility familiarity and the home advantage. *Journal of Sport Behavior*, 18(4):302, 1995.
- [9] Andrew Pace and Albert V Carron. Travel and the home advantage. *Canadian Journal of Sport Sciences*, 1992.
- [10] Ricardo Santos. Optimal Soccer Strategies. *Economic Inquiry*, 52(1):183–200, 2014.
- [11] Carlos Lago-Peñas, Ezequiel Rey, Joaquín Lago-Ballesteros, Luis Casáis, and Eduardo Domínguez. The influence of a congested calendar on physical performance in elite soccer. *The Journal of Strength & Conditioning Research*, 25(8):2111–2117, 2011.
- [12] Rasmus V Rasmussen and Michael A Trick. Round robin scheduling—a survey. *European Journal of Operational Research*, 188(3):617–636, 2008.

- [13] Rodrigo Wolf. Programación matemática en la confección de fixtures del fútbol chileno. *Universidad de Chile*, pages 1–3, 2010. Tesis de Ingeniero Civil Industrial y Magister en Gestión de Operaciones. Facultad de Ciencias Físicas y Matemáticas.
- [14] Informs. Franz Edelman Award for Achievement in Operations Research and the Management Sciences [en línea]. <<https://www.informs.org/Recognize-Excellence/Franz-Edelman-Award>>, 2016. [Consulta: 28 Mayo 2016].
- [15] La Tercera. Anfp finalista premio internacional [en línea]. <<http://www.latercera.com/noticia/deportes/2016/01/656-666192-9-la-anfp-finalista-del-prestigioso-premio-franz-edelman.shtml>>, 2016. [Consulta: 29 Mayo 2016].
- [16] Guillermo Durán, Mario Guajardo, Jaime Miranda, Denis Sauré, Sebastián Souyris, Andres Weintraub, and Rodrigo Wolf. Scheduling the Chilean soccer league by integer programming. *Interfaces*, 37(6):539–552, 2007.
- [17] Europa Press. La final de la Champions ya es más grande que la Super Bowl [en línea]. <<http://www.europapress.es/economia/noticia-final-champions-ya-mas-grande-super-bowl-20160527123838.html>>, 2016. [Consulta: 15 Noviembre 2016].
- [18] Consultoría Deloitte. Football Money League. *Deloitte Consultoría, Sport Business Group*, 2014.
- [19] Consultoría Deloitte. Deloitte’s Annual Review of Football Finance [en línea]. <<http://www2.deloitte.com/uk/en/pages/sports-business-group/articles/annual-review-of-football-finance.html>>, 2016. [Consulta: 25 Noviembre 2016].
- [20] POR Matatoros. Las apuestas deportivas mueven 5.000 millones al año en España [en línea]. <<http://www.expansion.com/directivos/deporte-negocio/2016/10/15/57ffcba446163f55278b4629.html>>, 2016. [Consulta: 2 Diciembre 2016].
- [21] Paul E Meehl. Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. *American Psychological Association*, pages 24–25, 1954.
- [22] UEFA. 2016/17 Champions League revenue distribution [en línea]. <<http://www.uefa.com/uefachampionsleague/news/newsid=2398575.html>>, 2016. [Consulta: 19 Noviembre 2016].
- [23] UEFA. 2016/17 Europa League revenue distribution [en línea]. <<http://www.uefa.com/uefaeuropaleague/news/newsid=2398584.html>>, 2016. [Consulta: 19 Noviembre 2016].
- [24] John DiNardo. Natural experiments and quasi-natural experiments. In *Microeconomics*, pages 139–153. Springer, 2010.
- [25] George F Loewenstein and Dražen Prelec. Preferences for sequences of outcomes. *Psychological review*, 100(1):91, 1993.

- [26] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, pages 263–291, 1979.
- [27] Dimitri P Bertsekas, Dimitri P Bertsekas, Dimitri P Bertsekas, and Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, MA, 1995.
- [28] Joseph P Simmons, Leif D Nelson, Jeff Galak, and Shane Frederick. Intuitive biases in choice versus estimation: Implications for the wisdom of crowds. *Journal of Consumer Research*, 38(1):1–15, 2011.
- [29] Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation*. WW Norton, 2012.
- [30] John A List, Sally Sadoff, and Mathis Wagner. So you want to run an experiment, now what? some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14(4):439–457, 2011.
- [31] John A List. An introduction to field experiments in economics. *Journal of Economic Behavior & Organization*, 70(3):439–442, 2009.
- [32] Milagros Ezquerro and Jose María Buceta. Estilo de procesamiento de la información y toma de decisiones en competiciones deportivas: Las dimensiones rapidez y exactitud cognitivas. *Análise Psicológica*, 19(1):37–50, 2001.
- [33] Paul D Ellis. *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press, 2010.
- [34] European Professional Football Leagues. How the fixtures are produced EPL [en línea]. <http://www.epfl-europeanleagues.com/how_the_fixtures_are_produced.htm>. [Consulta: 3- Julio 2016].
- [35] Mail Online. Premier league fixtures: how the list is compiled 2017 [en línea]. <<http://www.dailymail.co.uk/sport/football/article-3127659/Premier-League-fixtures-Q-Sportsmail-reveals-exactly-schedule-compiled.html>>, 2017. [Consulta: 30 Agosto 2016].
- [36] English Premier League. Premier League Football Scores, Results & Season Archives [en línea]. <<http://www.premierleague.com/results>>. [Consulta: 29 de Mayo 2016].
- [37] BBC Sport. Power of top four concerns Keegan [en línea]. <http://news.bbc.co.uk/sport2/hi/football/teams/n/newcastle_united/7384247.stm>, 2008. [Consulta: 30 Noviembre 2016].

Anexos A

Antecedentes Generales

A.1. Contexto de la Investigación

2015/16 PAYMENTS TO CLUBS							
	UK LIVE	EQUAL SHARE	FACILITY FEES	MERIT PAYMENT	OVERSEAS TV	CENTRAL COMMERCIAL	TOTAL PAYMENT
LEICESTER CITY	15	21,924,800	12,521,698	24,848,100	29,415,848	4,509,152	93,219,598
ARSENAL	27	21,924,800	21,496,762	23,605,695	29,415,848	4,509,152	100,952,257
TOTTENHAM HOTSPUR	21	21,924,800	17,009,230	22,363,290	29,415,848	4,509,152	95,222,320
MANCHESTER CITY	25	21,924,800	20,000,918	21,120,885	29,415,848	4,509,152	96,971,603
MANCHESTER UNITED	26	21,924,800	20,748,840	19,878,480	29,415,848	4,509,152	96,477,120
SOUTHAMPTON	12	21,924,800	10,277,932	18,636,075	29,415,848	4,509,152	84,763,807
WEST HAM UNITED	15	21,924,800	12,521,698	17,393,670	29,415,848	4,509,152	85,765,168
LIVERPOOL	23	21,924,800	18,505,074	16,151,265	29,415,848	4,509,152	90,506,139
STOKE CITY	10	21,924,800	8,782,088	14,908,860	29,415,848	4,509,152	79,540,748
CHELSEA	22	21,924,800	17,757,152	13,666,455	29,415,848	4,509,152	87,273,407
EVERTON	18	21,924,800	14,765,464	12,424,050	29,415,848	4,509,152	83,039,314
SWANSEA CITY	10	21,924,800	8,782,088	11,181,645	29,415,848	4,509,152	75,813,533
WATFORD	10	21,924,800	8,782,088	9,939,240	29,415,848	4,509,152	74,571,128
WEST BROMWICH ALBION	10	21,924,800	8,782,088	8,696,835	29,415,848	4,509,152	73,328,723
CRYSTAL PALACE	10	21,924,800	8,782,088	7,454,430	29,415,848	4,509,152	72,086,318
AFC BOURNEMOUTH	10	21,924,800	8,782,088	6,212,025	29,415,848	4,509,152	70,843,913
SUNDERLAND	13	21,924,800	11,025,854	4,969,620	29,415,848	4,509,152	71,845,274
NEWCASTLE UNITED	16	21,924,800	13,269,620	3,727,215	29,415,848	4,509,152	72,846,635
NORWICH CITY	10	21,924,800	8,782,088	2,484,810	29,415,848	4,509,152	67,116,698
ASTON VILLA	11	21,924,800	9,530,010	1,242,405	29,415,848	4,509,152	66,622,215
ALL FIGURES IN £		438,496,000	260,904,868	260,905,050	588,316,960	90,183,040	1,638,805,918

Figura A.1: Reparto de Ingresos de la temporada 2015/2016 - Premier League

Fuente:Diario Online MARCA

Anexos B

Metodología

B.1. Cuestionario online

Page 0)

Please read the following information before you start: This study examines how people make predictions. In particular, we will use the upcoming 2016 UEFA Euro football (soccer) tournament. Important: It does not matter (at all) if you follow this particular sport or tournament, and how you come up with predictions. In addition to today's survey, you will be asked to answer a short survey every day, and you can earn daily bonuses (between £0.2 and £12 per day). For this reason, if you want to participate in this study you have to commit some time to answer a daily short survey for one week. There will be no personal benefit from participating, other than your payment. The risks associated with participation in this study are no greater than those ordinarily encountered in daily life or during mild physical activity, such as when surfing the Internet. Please be assured that your responses will be kept confidential. If you have any questions, feel free to contact us. By clicking the Next button below, you acknowledge that your participation is voluntary and you may choose to terminate your participation in the study at any time and for any reason.

Pag 1)

Please answer the following questions before starting the survey:

	Yes	No
I am 18 or older	<input type="radio"/>	<input type="radio"/>
I have read and understand the information above	<input type="radio"/>	<input type="radio"/>
I want to participate in this study	<input type="radio"/>	<input type="radio"/>

Pag 2)

[En esta página se le pide el código de participación para registrar al participante]

Please, enter your Prolific Participant ID.

Pag 3)

In this page we will explain the basics of the study, and ask a few questions to know whether you can participate in this study. Basics of the study and bonuses: We will ask you to enter predictions for the upcoming football matches of the tournament (winner/draw). You can get any information from the Web for your predictions. We will use the matches of the first week of the tournament starting on 10/June/2016. In addition to the questions you will answer today, for this study to work, we need people to answer a few daily questions about the matches of that week (it does not matter whether you are following the tournament). This can take less than 1 minute per day and it can be answered from any device. Surveys' link will be sent from 9/June to 17/June, and we will send out daily reminders to make everything easier. After the matches of one day are played, we will randomly pick 3 people and pay them £4 per match correctly predicted (winner or draw). Randomly pick other 30 people and pay them £0.4 per match correctly predicted, and all the rest of participants will get £0.2 per day. All these bonuses will be paid only if the participant has answered the daily survey the previous day (between £0.2 and £12 per day)*. The purpose of these bonuses is to encourage you to answer the daily survey and predict as best as you can. We will send out reports with daily and accumulated earnings, which will be paid every two days as bonuses. People who did not answer a daily survey will be notified the bonus they missed, but they can still continue their participation and earn bonuses the following days.

** We expect 400 participants using two studies. Please click next to check if you will be able to participate in this study.*

Pag 4)

Can you commit to answer the daily short surveys? (Surveys' link will be sent from 9/June to 17/June)

- Yes, I will answer the daily short surveys
- No, I do not have the time

How many matches are you planning to watch from the 2016 UEFA Euro football (soccer) tournament?

- All matches
- Most of the matches
- Half of the matches
- Only a few matches
- None of them

Pag 5)

In addition, please answer these couple of questions about you:

How old are you?

What is your gender?

- Female
- Male

How much do you like football (soccer) as sport?

	Dislike a great deal	Dislike a moderate amount	Dislike a little	Neither like nor dislike	Like a little	Like a moderate amount	Like a great deal
How much do you like football (soccer) as sport?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What is your educational background?

- Some high school, incomplete
- High school diploma
- Some college, incomplete
- Undergraduate college degree (BS, BA)
- Graduate degree (MA, PhD, MBA, etc)

Which is your country of Residence?

Pag 6)

Thank you for participating in this study. Please create a username that you will use for the daily short surveys. Use an easy to recall word (at least 6 characters):

Pag 7)

Remember that every day you will receive a short survey with questions about the tournament, and a report describing how you are doing with your predictions. The subject of the message will be "Prediction for 2016 UEFA Euro football". Please check that it is not in your spam folder.

Pag 8.1) [Este mensaje era sólo para los participantes que se encontraban en el grupo One Shot]

Now, we will ask you to predict some matches of the group stage. At the end of the matches and after a few questions is the Prolific code to receive today's payment. Please click Next to start.

Pag 8.2) [Este mensaje era sólo para los participantes que se encontraban en el grupo One Shot]

Please make your predictions for the following matches. You can take as much time as you want to make your predictions. Feel free to consult betting Websites or news about the tournament for your predictions. Just to remind you about the bonuses: After the matches of one day are played, we will randomly pick 3 people and pay them £4 per match correctly predicted (winner or draw). Randomly pick other 30 people and pay them £0.4 per match correctly predicted, and all the rest of participants will get £0.2 per day. All these bonuses will be paid only if the participant has answered the daily survey the previous day. The purpose of these bonuses is to encourage you to answer the daily survey and predict as best as you can. We will send out reports with daily and accumulated earnings, which will be paid every two days as bonuses. People who did not answer a daily survey will be notified the bonus they missed, but they can still continue their participation and earn bonuses the following days.

Pag 8.3) [Este mensaje era sólo para los participantes que se encontraban en el grupo One Shot]

The following matches are ordered by date, indicating the teams playing and their group. You have to predict as best as you can the winner or draw. For example, if it says "A vs B," you can indicate that A wins ("Win 1"), or B wins ("Win 2"), or the match ends in draw (there is no extra time in the group stage). The best teams in each group advance to the knockout phase. All the matches are played in France.

Day #1 Friday 10 June 2016

	Win 1	Draw	Win 2
France vs Romania (Group A)	O	O	O

Day #2 Saturday 11 June 2016

	Win 1	Draw	Win 2
Albania vs Switzerland (A)	O	O	O
Wales vs Slovakia (B)	O	O	O
England vs Russia (B)	O	O	O

Day #3 Sunday 12 June 2016

	Win 1	Draw	Win 2
Turkey vs Croatia (D)	O	O	O
Poland vs Northern Ireland (C)	O	O	O
Germany vs Ukraine (C)	O	O	O

Day #4 Monday 13 June 2016

	Win 1	Draw	Win 2
Turkey vs Croatia (D)	O	O	O
Poland vs Northern Ireland (C)	O	O	O
Germany vs Ukraine (C)	O	O	O

Day #5 Tuesday 14 June 2016

	Win 1	Draw	Win 2
Austria vs Hungary (F)	O	O	O
Portugal vs Iceland (F)	O	O	O

Day #6 Wednesday 15 June 2016

	Win 1	Draw	Win 2
Russia vs Slovakia (B)	O	O	O
France vs Albania (A)	O	O	O

Day #7 Thursday 16 June 2016

	Win 1	Draw	Win 2
England vs Wales (B)	O	O	O
Ukraine vs Northern Ireland (C)	O	O	O
Germany vs Poland (C)	O	O	O

Day #8 Friday 17 June 2016

	Win 1	Draw	Win 2
Italy vs Sweden (E)	O	O	O
Czech Republic vs Croatia (D)	O	O	O
Spain vs Turkey (D)	O	O	O

Day #9 Saturday 18 June 2016

	Win 1	Draw	Win 2
Belgium vs Republic of Ireland (E)	O	O	O
Iceland vs Hungary (F)	O	O	O
Portugal vs Austria (F)	O	O	O

Did you finish your predictions? Please click Next. Remember that your bonus will depend on your predictions.

Pag 8.4) [Este mensaje era sólo para los participantes que se encontraban en el grupo One Shot]

Please rank as best as you can how you made your predictions from the following items. Drag and drop only the ones that you actually used, ranking them from the ones that you used the most for your predictions.

I used for my predictions
<input type="checkbox"/> I used my football expertise
<input type="checkbox"/> I read some news about the tournament
<input type="checkbox"/> I used betting websites
<input type="checkbox"/> I entered predictions randomly

In addition, please describe as best as you can how you made your predictions.

Pag 8.5) [Este mensaje era sólo para los participantes que se encontraban en el grupo One Shot]

Remember that this Friday 10th June you will start receiving a daily message asking you to answer a short survey (after the first match). You can use your username: $q : //QID225/ChoiceTextEntr$ (código que guarda el nombre de usuario que el participante eligió y se lo muestra para recordárselo)

Pag 9.1) [Este mensaje era sólo para los participantes que se encontraban en el grupo Day by Day]

Now, you will see the schedule that we will use for your predictions. You don't have to predict any match today, but this is the format we will use in the daily surveys. Your predictions will start on Thursday 9th June, before the first match. At the end of the calendar and after a few questions is the Prolific code to receive today's payment.

Pag 9.2) [Este mensaje era sólo para los participantes que se encontraban en el grupo Day by Day]

Please remember that in a few days you will make predictions for the following matches. You can take as much time as you want looking at the calendar. Feel free to consult betting Websites or news about the tournament for your predictions. Just to remind you about the bonuses: After the matches of one day are played, we will randomly pick 3 people and pay them £4 per match correctly predicted (winner or draw). Randomly pick other 30 people and pay them £0.4 per match correctly predicted, and all the rest of participants will get £0.2 per day. All these bonuses will be paid only if the participant has answered the daily survey the previous day. The purpose of these bonuses is to encourage you to answer the daily survey and predict as best as you can. We will send out reports with daily and accumulated earnings, which will be paid every two days as bonuses. People who did not answer a daily survey will be notified the bonus they missed, but they can still continue their participation and earn bonuses the following days.

Pag 9.3) [Este mensaje era sólo para los participantes que se encontraban en el grupo Day by Day]

The following matches are ordered by date, indicating the teams playing and their group. You have to predict as best as you can the winner or draw. For example, if it says “A vs B,” you can indicate that A wins (“Win 1”), or B wins (“Win 2”), or the match ends in draw (there is no extra time in the group stage). The best teams in each group advance to the knockout phase. All the matches are played in France. Do not make any predictions today.

Day #1 Friday 10 June 2016

	Win 1	Draw	Win 2
France vs Romania (Group A)	O	O	O

Day #2 Saturday 11 June 2016

	Win 1	Draw	Win 2
Albania vs Switzerland (A)	O	O	O
Wales vs Slovakia (B)	O	O	O
England vs Russia (B)	O	O	O

Day #3 Sunday 12 June 2016

	Win 1	Draw	Win 2
Turkey vs Croatia (D)	O	O	O
Poland vs Northern Ireland (C))	O	O	O
Germany vs Ukraine (C)	O	O	O

Day #4 Monday 13 June 2016

	Win 1	Draw	Win 2
Turkey vs Croatia (D)	O	O	O
Poland vs Northern Ireland (C)	O	O	O
Germany vs Ukraine (C)	O	O	O

Day #5 Tuesday 14 June 2016

	Win 1	Draw	Win 2
Austria vs Hungary (F)	O	O	O
Portugal vs Iceland (F)	O	O	O

Day #6 Wednesday 15 June 2016

	Win 1	Draw	Win 2
Russia vs Slovakia (B)	O	O	O
France vs Albania (A)	O	O	O

Day #7 Thursday 16 June 2016

	Win 1	Draw	Win 2
England vs Wales (B)	O	O	O
Ukraine vs Northern Ireland (C)	O	O	O
Germany vs Poland (C)	O	O	O

Day #8 Friday 17 June 2016

	Win 1	Draw	Win 2
Italy vs Sweden (E)	O	O	O
Czech Republic vs Croatia (D)	O	O	O
Spain vs Turkey (D)	O	O	O

Day #9 Saturday 18 June 2016

	Win 1	Draw	Win 2
Belgium vs Republic of Ireland (E)	O	O	O
Iceland vs Hungary (F)	O	O	O
Portugal vs Austria (F)	O	O	O

Did you finish looking at this calendar? Please click Next

Pag 9.4) [Este mensaje era sólo para los participantes que se encontraban en el grupo Day by Day]

Thanks. Remember that this Thursday 9th June you will start receiving a daily message asking you to answer a short survey (before the first match). You can use your username: *q : //QID225/ChoiceTextEntryValue* (código que guarda el nombre de usuario que el participante eligió y se lo muestra para recordárselo)

Pag 10) [Preguntas para ambos grupos]

Please answer these final questions. We would appreciate your views on the study.

Were the rules about the study clear? _____

Please provide any other comments about this study: _____

Pag 11) [En caso de que los partidos se estuvieran jugando, los participantes no podían realizar sus predicciones por lo que el sitio se encontraba bloqueado]

Sorry it is not possible to enter the survey now because there is football match playing from the 2016 UEFA Euro football (soccer) tournament. Try again when the match is over.

Pag 12) [En caso de que no se estuvieran jugando partidos, la gente ingresaba de la siguiente manera]

Id o user? Please select whether you want to use your Prolific ID or username?

- Prolific ID
 - Username
-

Pag 13) [Luego se muestra un mensaje diario de los reportes de los resultados (ejemplo)]

Pag 14) [Esta página aparece para el Grupo Day by Day dependiendo del día para que

Hello pikachu

Daily Report of bonuses:

Earnings from last day: £0.4

Accumulated earnings: £1.8

Recall that earnings are paid every two days. Don't forget to answer the daily surveys, so you don't miss your bonus.

Details (last day):

You answered the last daily survey: Yes

Number of correct predictions: 0

Randomly picked for £4 per correct prediction (3 people): 0

Randomly picked for £0.4 per correct prediction (30 people): 0

Actual Results / Your Predictions (last day):

Belgium vs Northern Ireland 3-0 Prediction: Draw

Iceland vs Hungary 1-1 Prediction: Win Iceland

Portugal vs Austria 0-0 Prediction: Win Portugal

hagan la predicción para el siguiente día]

Display This Question: If Current day Is Equal to 1

Day #1 Friday 10 June 2016

	Win 1	Draw	Win 2
France vs Romania (Group A)	O	O	O

Display This Question: If Current day Is Equal to 2

Day #2 Saturday 11 June 2016

	Win 1	Draw	Win 2
Albania vs Switzerland (A)	O	O	O
Wales vs Slovakia (B)	O	O	O
England vs Russia (B)	O	O	O

Display This Question: If Current day Is Equal to 3

Day #3 Sunday 12 June 2016

	Win 1	Draw	Win 2
Turkey vs Croatia (D)	O	O	O
Poland vs Northern Ireland (C))	O	O	O
Germany vs Ukraine (C)	O	O	O

Display This Question: If Current day Is Equal to 4

Day #4 Monday 13 June 2016

	Win 1	Draw	Win 2
Turkey vs Croatia (D)	O	O	O
Poland vs Northern Ireland (C)	O	O	O
Germany vs Ukraine (C)	O	O	O

Display This Question: If Current day Is Equal to 5

Day #5 Tuesday 14 June 2016

	Win 1	Draw	Win 2
Austria vs Hungary (F)	O	O	O
Portugal vs Iceland (F)	O	O	O

Display This Question: If Current day Is Equal to 6

Day #6 Wednesday 15 June 2016

	Win 1	Draw	Win 2
Russia vs Slovakia (B)	O	O	O
France vs Albania (A)	O	O	O

Display This Question: If Current day Is Equal to 7

Day #7 Thursday 16 June 2016

	Win 1	Draw	Win 2
England vs Wales (B)	O	O	O
Ukraine vs Northern Ireland (C)	O	O	O
Germany vs Poland (C)	O	O	O

Display This Question: If Current day Is Equal to 8

Day #8 Friday 17 June 2016

	Win 1	Draw	Win 2
Italy vs Sweden (E)	O	O	O
Czech Republic vs Croatia (D)	O	O	O
Spain vs Turkey (D)	O	O	O

Display This Question: If Current day Is Equal to 9

Day #9 Saturday 18 June 2016

	Win 1	Draw	Win 2
Belgium vs Republic of Ireland (E)	O	O	O
Iceland vs Hungary (F)	O	O	O
Portugal vs Austria (F)	O	O	O

Display This Question: If Current day Is Less Than or Equal to 9

Did you finish looking at this calendar? Please click Next

Pag 16) [Luego se muestra la siguiente pregunta, para el caso del One Shot también]

Did you follow or watch any of the last matches? Don't forget to click Next to submit your answers.

- All of them
- Some
- None

Pag 17) [Pregunta para el último día]

Display This Question: If Current day Is Equal to 9

Thank you for your participation. In this daily survey, you have to answer a few more questions.

Pag 18) [Sólo para el grupo Day by Day]

The last few days you have been predicting several matches, please rank as best as you can how you made your predictions from the following items. Drag and drop only the ones that you actually used, ranking them from the ones that you used the most for your predictions.

I used for my predictions
_____ I used my football expertise
_____ I read some news about the tournament
_____ I used betting websites
_____ I entered predictions randomly

In addition, please describe as best as you can how you made your predictions. _____

Pag 19) [Por último preguntas de feedback para ambos grupos]

Thank you for your participation. In this daily survey, you have to answer a few more questions.

Pag 20) [Por último preguntas de feedback para ambos grupos]

How often did you receive the daily messages asking to answer the daily surveys? (from 9/June to 16/June: total 8 days). Answer as best as you can.

- 8 days (I received it every day)
- 7 days
- 6 days
- 5 days
- 4 days (half of them)
- 3 days
- 2 days
- 1 day
- I didn't receive any daily message asking to answer a daily survey

Which of the following daily emails (subjects line) did you generally receive? You can select one or more.

- "Prediction for 2016 UEFA Euro football"
- Reminder - Prediction for 2016 UEFA Euro football"
- "Last chance to answer: Prediction for 2016 UEFA Euro football"
- None of them

Please indicate if you had any issues receiving the daily messages _____

Please write any comment about this study. This will help us for future versions. _____

Pag 21) [Preguntaba que se mostraba al One Shot un día antes de empezar el campeonato]

Hello xxxx This is a just a reminder that tomorrow the 2016 UEFA Euro football will start with the football match: France vs Romania. Your prediction was: xxxxxx The bonus report will be sent tomorrow

B.2. Cuotas de los sitios de apuestas online

<i>Matches</i>	0 (3 de Junio)			1 (9 de Junio)			2 (10 de Junio)			3 (11 de Junio)			4 (12 de Junio)		
	<i>Win</i>	<i>Draw</i>	<i>Lose</i>	<i>Win</i>	<i>Draw</i>	<i>Lose</i>	<i>Win</i>	<i>Draw</i>	<i>Lose</i>	<i>Win</i>	<i>Draw</i>	<i>Lose</i>	<i>Win</i>	<i>Draw</i>	<i>Lose</i>
France vs Rumania	1.28	4.75	13	1.3	5.5	13	1.36	5	11						
Albania vs Switzerland	5.5	3.5	1.66	5.5	3.4	1.83	5.75	3.4	1.8	5	3.2	1.95			
Wales vs Slovakia	2.62	3	2.87	2.87	3	2.87	3	3	2.8	2.87	3	2.87			
England vs Russia	1.85	3.3	4.5	1.9	3.4	4.75	1.9	3.4	4.75	1.85	3.5	5			
Turkey vs Croatia	3.75	3.1	2.1	3.9	3.25	2.15	3.9	3.25	2.15	3.75	3.25	2.2	4	3.25	2
Poland vs Northern Ireland	1.66	3.5	5.5	1.72	3.6	6	1.72	3.6	6	1.72	3.6	6	1.75	3.5	6
Germany vs Ukraine	1.5	3.8	7.5	1.57	4	7	1.6	4	6.5	1.6	4	6.5	1.55	4.2	7
Spain vs Czech Republic	1.5	3.8	7.5	1.44	4.2	7.5	1.5	4.33	8	1.5	4.33	8	1.5	4.33	8
Rep of Ireland vs Sweden	3	3	2.5	3.4	3.1	2.45	3.4	3.1	2.45	3.4	3.1	2.45	3.4	3.1	2.45
Belgium vs Italy	2.5	3	3	2.6	3.1	3.1	2.6	3.1	3.1	2.6	3.1	3.1	2.6	3	3.25
Austria vs Hungary	1.7	3.5	5.25	1.72	3.7	5.75	1.72	3.7	5.75	1.72	3.7	5.75	1.72	3.7	5.75
Portugal vs Iceland	1.61	3.6	6	1.6	3.7	7.5	1.6	3.7	7.5	1.6	3.7	7.5	1.6	3.7	7.5
Russia vs Slovakia	2	3.2	4	2.1	3.2	3.6	2.1	3.2	3.6	2.1	3.2	3.6	2.3	3.2	3.6
Romania vs Switzerland	4	3.3	1.95	4	3.3	1.95	3.6	3.3	2.05	3.4	3.3	2.15	3.4	3.2	2.37
France vs Albania	1.18	6.5	15	1.2	6	15	1.2	6	15	1.2	6	15	1.22	6.5	19
England vs Wales	1.61	3.75	5.5	1.61	3.75	6	1.61	3.75	4.2	1.61	3.75	6	1.61	3.8	7
Ukraine vs Northern Ireland	1.8	3.5	4.5	1.7	3.75	5	1.75	3.5	5	1.75	3.5	5	1.75	3.5	5
Germany vs Poland	1.53	4	6	1.57	4	5.75	1.57	3.8	5.75	1.57	4	5.75	1.57	4	5.75
Italy vs Sweden	1.9	3.4	4	2	3.1	4.2	2	3.1	4.2	2	3.1	4.2	2	3.1	4.2
Czech Republic vs Croatia	3.4	3.4	2.1	3.6	3.5	2	3.6	3.5	2	3.6	3.5	2	3.6	3.5	2
Spain vs Turkey	1.45	4.25	7	1.5	4.2	8	1.53	3.8	6.5	1.53	3.8	6.5	1.53	3.8	6.5
Belgium vs Rep of Ireland	1.55	4	6	1.53	4	6	1.57	4	5.75	1.57	4	5.75	1.57	4	5.75
Iceland vs Hungary	2.5	3.1	2.9	2.3	3.1	3.2	2.3	3.1	3.2	2.3	3.1	3.2	2.3	3.1	3.2
Portugal vs Austria	2.05	3.3	3.6	2.1	3.25	3.5	2.1	3.25	3.5	2.1	3.25	3.5	2.05	3.2	3.75

Tabla B.1: Cuotas de los partidos del sitio Bet365 hasta el 4 de Junio

Fuente: elaboración propia

<i>Matches</i>	5 (13 de Junio)			6 (14 de Junio)			7 (15 de Junio)			8 (16 de Junio)			9 (17 de Junio)		
	<i>Win</i>	<i>Draw</i>	<i>Lose</i>												
Spain vs Czech Republic	1.44	4.5	9												
Rep of Ireland vs Sweden	3.6	3.1	2.35												
Belgium vs Italy	2.6	3	3.25												
Austria vs Hungary	1.72	3.7	5.75												
Portugal vs Iceland	1.57	3.9	7.5	1.44	4.33	10									
Russia vs Slovakia	2.3	3.25	3.5	2.4	3.25	3.3	2.4	3.25	3.3						
Romania vs Switzerland	3.4	3.2	2.37	3.25	3.2	2.5	3.3	3.1	2.5						
France vs Albania	1.22	6.5	19	1.22	6.5	19	1.25	6	17						
England vs Wales	1.57	4	7	1.57	4	7	1.57	4	7	4	2.87	2.2			
Ukraine vs Northern Ireland	1.57	4	7	1.57	4	7	1.57	4	7	1.75	3.4	6.25			
Germany vs Poland	1.57	4.33	6.5	1.57	4.33	6.5	1.57	4.33	6.5	1.57	4.2	6.5			
Italy vs Sweden	2	3.1	4.2	1.83	3.5	5.25	1.83	3.5	5.25	1.85	3.3	5.5	1.9	3.3	5
Czech Republic vs Croatia	3.75	3.5	1.95	4.75	3.5	1.9	4.75	3.5	1.9	5	3.6	1.83	4.75	3.4	1.9
Spain vs Turkey	1.44	4.33	7	1.44	4.5	9	1.44	4.5	9	1.44	4.5	9	1.44	4.33	9.5
Belgium vs Rep of Ireland	1.57	4	5.75	1.72	3.8	5.5	1.72	3.8	5.5	1.72	3.8	5.5	1.75	3.75	5.5
Iceland vs Hungary	2.3	3.1	3.2	2.37	3.1	3.1	2.6	3.25	3	2.6	3.25	3	2.62	3.1	3.1
Portugal vs Austria	2.05	3.2	3.75	1.83	3.25	4.75	1.83	3.6	5	1.8	3.6	5.25	1.8	3.6	5.25

Tabla B.2: Cuotas de los partidos del sitio Bet365 del 5 de Junio hasta el 9 de Junio

Fuente: elaboración propia

<i>Matches</i>	0 (3 de Junio)			1 (9 de Junio)			2 (10 de Junio)			3 (11 de Junio)			4 (12 de Junio)		
	<i>Win</i>	<i>Draw</i>	<i>Lose</i>	<i>Win</i>	<i>Draw</i>	<i>Lose</i>	<i>Win</i>	<i>Draw</i>	<i>Lose</i>	<i>Win</i>	<i>Draw</i>	<i>Lose</i>	<i>Win</i>	<i>Draw</i>	<i>Lose</i>
France vs Rumania	1.3	4.75	12	1.3	5.25	13.5	1.36	4.75	11.5						
Albania vs Switzerland	5.25	3.4	1.75	5.25	3.4	1.85	5.5	3.4	1.8	5	3.3	1.9			
Wales vs Slovakia	2.55	3	3	2.85	3	2.85	3	3	2.75	2.8	3	2.9			
England vs Russia	1.87	3.4	4.4	1.9	3.4	4.75	1.91	3.3	5	1.85	3.4	5.25			
Turkey vs Croatia	3.4	3.25	2.2	3.9	3.25	2.15	3.8	3.3	2.15	3.8	3.3	2.15	4	3.1	2
Poland vs Northern Ireland	1.67	3.6	5.5	1.7	3.6	6	1.7	3.6	6	1.7	3.6	6	1.72	3.6	6
Germany vs Ukraine	1.48	4.2	7	1.57	4.1	6.75	1.57	4.1	6.75	1.57	4.1	6.75	1.55	4.1	7
Spain vs Czech Republic	1.5	4	7	1.48	4.33	8	1.5	4.25	7.75	1.5	4.25	7.75	1.5	4.25	7.75
Rep of Ireland vs Sweden	3	3.1	2.5	3.4	3.1	2.4	3.4	3.1	2.4	3.4	3.1	2.4	3.4	3.1	2.4
Belgium vs Italy	2.5	3.1	3	2.6	3.1	3.1	2.6	3.1	3.1	2.6	3.1	3.1	2.6	3.1	3.1
Austria vs Hungary	1.72	3.6	5	1.7	3.7	5.75	1.7	3.7	5.75	1.7	3.7	5.75	1.7	3.7	5.75
Portugal vs Iceland	1.6	3.6	6.25	1.6	3.7	7.25	1.6	3.7	7.25	1.6	3.7	7.25	1.6	3.8	7.25
Russia vs Slovakia	1.91	3.4	4.1	2	3.25	3.9	2.15	3.25	3.4	2.15	3.25	3.4	2.35	3.3	3.3
Romania vs Switzerland	4	3.4	1.91	4.1	3.3	1.95	4.1	3.3	1.95	3.7	3.2	2.15	3.5	3.25	2.3
France vs Albania	1.2	6.5	13.5	1.2	6.5	13.5	1.2	6	16	1.22	6	16	1.22	6.5	17.5
England vs Wales	1.75	3.5	4.9	1.62	3.7	5.75	1.62	3.7	5.75	1.62	3.7	5.75	1.6	4.1	6.25
Ukraine vs Northern Ireland	1.91	3.5	4	1.72	3.6	5	1.72	3.6	5	1.72	3.6	5	1.72	3.6	5
Germany vs Poland	1.53	4.1	6	1.57	4	5.75	1.57	4	5.75	1.57	4	5.75	1.57	4	5.75
Italy vs Sweden	2	3.3	3.9	2	3.3	3.9	2	3.3	3.9	2	3.3	3.9	2	3.3	3.9
Czech Republic vs Croatia	3.2	3.25	2.25	3.6	3.5	2	3.6	3.5	2	3.6	3.5	2			
Spain vs Turkey	1.55	4	6	1.53	4	6.5	1.53	4	6.5	1.53	4	6.5			
Belgium vs Rep of Ireland	1.6	3.9	5.5	1.6	3.9	5.5	1.6	3.9	5.5	1.6	3.9	5.5	1.6	3.9	5.5
Iceland vs Hungary	2.3	3.2	3.2	2.3	3.2	3.2	2.3	3.2	3.2	2.3	3.2	3.2	2.3	3.2	3.2
Portugal vs Austria	2.25	3.3	3.2	2.1	3.3	3.5	2.1	3.3	3.5	2.1	3.3	3.5	2.1	3.3	3.5

Tabla B.3: Cuotas de los partidos del sitio Bwin hasta el 4 de Junio

Fuente: elaboración propia

<i>Matches</i>	5 (13 de Junio)			6 (14 de Junio)			7 (15 de Junio)			8 (16 de Junio)			9 (17 de Junio)		
	<i>Win</i>	<i>Draw</i>	<i>Lose</i>												
Spain vs, Czech Republic	1.42	4.5	9.5												
Rep of,Ireland vs Sweden	3.8	3.1	2.25												
Belgium,vs Italy	2.55	3	3.3												
Austria,vs Hungary	1.75	3.6	5.5												
Portugal,vs Iceland	1.57	3.8	7.5	1.44	4.4	9									
Russia vs,Slovakia	2.35	3.3	3.3	2.4	3.25	3.3	2.35	3.25	3.4						
Romania,vs Switzerland	3.5	3.25	2.3	3.3	3.2	2.4	3.2	3.2	2.45						
France vs,Albania	1.22	6.5	17.5	1.22	6.5	16.5	1.25	6	17						
England,vs Wales	1.57	4.1	6.5	1.57	4.1	6.5	1.57	4.1	6.5						
Ukraine,vs Northern Ireland	1.55	4	7.25	1.57	3.9	7.25	1.57	3.9	7.25	1.75	3.4	6			
Germany,vs Poland	1.57	4.25	6.25	1.57	4.25	6.25	1.57	4.2	6.5	1.6	4.2	6			
Italy vs,Sweden	2	3.3	3.9	1.83	3.5	5	1.83	3.5	5	1.85	3.4	5	1.85	3.3	5.25
Czech,Republic vs Croatia	3.8	3.5	1.95	4.5	3.6	1.87	4.5	3.6	1.87	4.75	3.6	1.83	4.75	3.5	1.87
Spain vs,Turkey	1.48	4.2	7	1.42	4.6	9	1.42	4.6	9	1.42	4.6	9	1.44	4.4	9.25
Belgium,vs Rep of Ireland	1.6	3.9	5.5	1.7	4	5.25	1.7	3.8	5.5	1.7	3.8	5.5	1.75	3.7	5.25
Iceland,vs Hungary	2.3	3.2	3.2				2.5	3.2	2.9	2.6	3.2	3	2.6	3.1	3.1
Portugal,vs Austria	2.1	3.3	3.5				1.8	3.6	4.4	1.78	3.7	5	1.78	3.6	5.25

Tabla B.4: Cuotas de los partidos del sitio Bwin del 5 de Junio hasta el 9 de Junio

Fuente: elaboración propia

B.3. Caracterización del attrition del experimento online

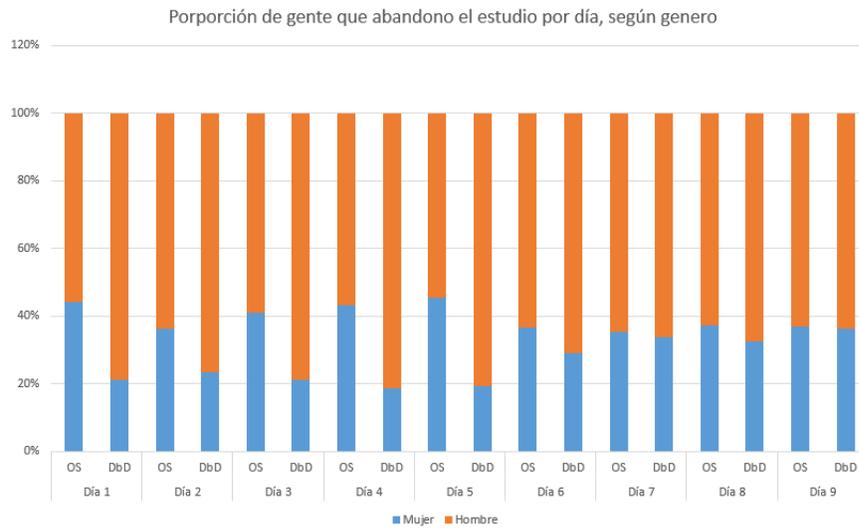


Figura B.1: Proporción de gente que abandonó experimento por género

Fuente: elaboración propia

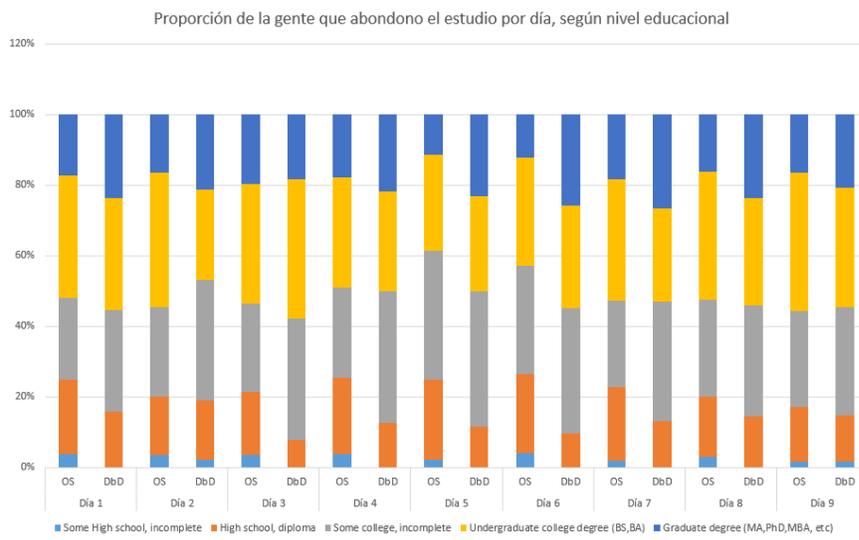


Figura B.2: Proporción de gente que abandonó experimento según educación

Fuente: elaboración propia

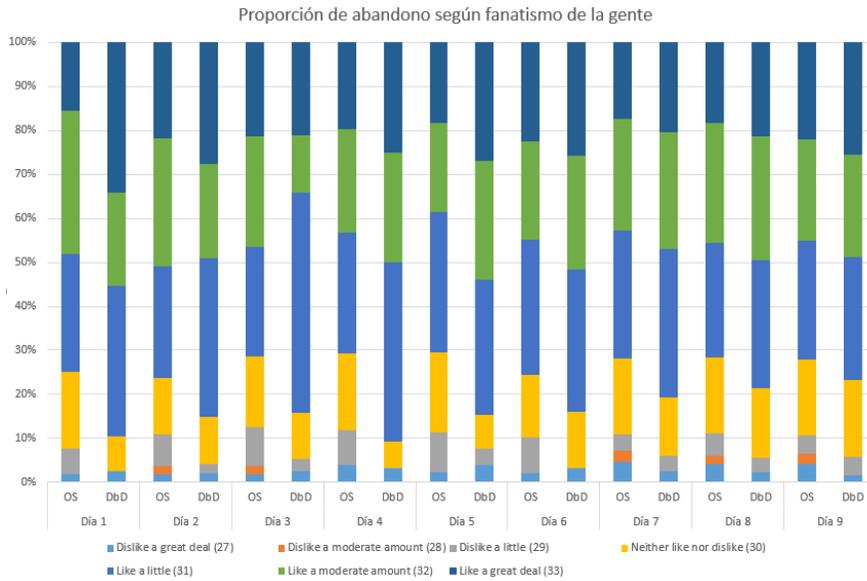


Figura B.3: Proporción de gente que abandonó experimento según fanatismo

Fuente: elaboración propia

B.4. Modelo alternativo para el attrition

	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7	Día 8	Día 9	Suma
One Shot	0.835	1.108	1.549*	1.448	0.309	0.915*	0.793	1.193	0.741*	8.893
Day by Day	0.831	1.071	1.839*	1.475	0.266	1.013*	0.817	0.831	1.25*	9.397

Las medias son estadísticamente significativos con un p-valor menor a 0.05

Tabla B.5: Promedio de predicciones segun modelo alternativo

Fuente: elaboración propia

B.5. Descripción de las 10 temporadas

Temporada/ Resultados	05/06	06/07	07/08	08/09	09/10	10/11	11/12	12/13	13/14	14/15
Victorias de Local	192	182	176	173	193	179	171	166	179	172
Empates	77	98	100	97	96	111	93	108	78	93
Victorias de Visita	111	100	104	110	91	90	116	106	123	115

Tabla B.6: Resultados según las 10 temporadas

Fuente: elaboración propia

B.6. Tipos de tabla de análisis

ID	Pos	Equipo	PJ	PG	PE	PP	GF	GC	DG	Pts	Diffcil	Fácil
2005/2006	1	Chelsea FC	38	29	4	5	72	22	50	91	1	0
2005/2006	2	Manchester United	38	25	8	5	72	34	38	83	1	0
2005/2006	3	FC Liverpool	38	25	7	6	57	25	32	82	1	0
2005/2006	4	FC Arsenal	38	20	7	11	68	31	37	67	1	0
2005/2006	5	Tottenham Hotspur	38	18	11	9	53	38	15	65	1	0
2005/2006	6	Blackburn Rovers	38	19	6	13	51	42	9	63	0	0
2005/2006	7	Newcastle United	38	17	7	14	47	42	5	58	0	0
2005/2006	8	Bolton Wanderers	38	15	11	12	49	41	8	56	0	0
2005/2006	9	West Ham United	38	16	7	15	52	55	-3	55	0	0
2005/2006	10	Wigan Athletic	38	15	6	17	45	52	-7	51	0	0
2005/2006	11	FC Everton	38	14	8	16	34	49	-15	50	0	0
2005/2006	12	FC Fulham	38	14	6	18	48	58	-10	48	0	0
2005/2006	13	Charlton Athletic	38	13	8	17	41	55	-14	47	0	0
2005/2006	14	FC Middlesbrough	38	12	9	17	48	58	-10	45	0	0
2005/2006	15	Manchester City	38	13	4	21	43	48	-5	43	0	1
2005/2006	16	Aston Villa	38	10	12	16	42	55	-13	42	0	1
2005/2006	17	FC Portsmouth	38	10	8	20	37	62	-25	38	0	1
2005/2006	18	Birmingham City	38	8	10	20	28	50	-22	34	0	1
2005/2006	19	West Bromwich Albion	38	7	9	22	31	58	-27	30	0	1
2005/2006	20	AFC Sunderland	38	3	6	29	26	69	-43	15	0	1

Tabla B.7: Ejemplo de clasificación final equipos Liga Inglesa temporada 2005/2006

Fuente: Sitio oficial Premier League

ID	Equipo	PG	PE	PP	Ptos	Pos	Diffcil	Fácil
2005/2006	Chelsea FC	18	1	0	55	1	1	0
2005/2006	FC Liverpool	15	3	1	48	2	1	0
2005/2006	FC Arsenal	14	3	2	45	3	1	0
2005/2006	Manchester United	13	5	1	44	4	1	0
2005/2006	Blackburn Rovers	13	3	3	42	5	1	0
2005/2006	FC Fulham	13	2	4	41	6	0	0
2005/2006	Tottenham Hotspur	12	5	2	41	6	0	0
2005/2006	Newcastle United	11	5	3	38	8	0	0
2005/2006	Bolton Wanderers	11	5	3	38	8	0	0
2005/2006	West Ham United	9	3	7	30	10	0	0
2005/2006	Manchester City	9	2	8	29	11	0	0
2005/2006	FC Everton	8	4	7	28	12	0	0
2005/2006	Charlton Athletic	8	4	7	28	12	0	0
2005/2006	FC Middlesbrough	7	5	7	26	14	0	0
2005/2006	Aston Villa	6	6	7	24	15	0	0
2005/2006	Wigan Athletic	7	3	9	24	15	0	0
2005/2006	Birmingham City	6	5	8	23	17	0	1
2005/2006	FC Portsmouth	5	7	7	22	18	0	1
2005/2006	West Bromwich Albion	6	2	11	20	19	0	1
2005/2006	AFC Sunderland	1	4	14	7	20	0	1

Tabla B.8: Ejemplo de clasificación según partidos de local equipos Liga Inglesa temporada 2005/2006

Fuente: Sitio oficial Premier League

ID	Equipo	PG	PE	PP	Ptos	Pos	Diffcil	Fácil
2005/2006	Manchester United	12	3	4	39	1	1	0
2005/2006	Chelsea FC	11	3	5	36	2	1	0
2005/2006	FC Liverpool	10	4	5	34	3	1	0
2005/2006	Wigan Athletic	8	3	8	27	4	1	0
2005/2006	West Ham United	7	4	8	25	5	1	0
2005/2006	Tottenham Hotspur	6	6	7	24	6	0	0
2005/2006	FC Everton	6	4	9	22	7	0	0
2005/2006	FC Arsenal	6	4	9	22	7	0	0
2005/2006	Blackburn Rovers	6	3	10	21	9	0	0
2005/2006	Newcastle United	6	2	11	20	10	0	0
2005/2006	FC Middlesbrough	5	4	10	19	11	0	0
2005/2006	Charlton Athletic	5	4	10	19	11	0	0
2005/2006	Aston Villa	4	6	9	18	13	0	0
2005/2006	Bolton Wanderers	4	6	9	18	13	0	0
2005/2006	FC Portsmouth	5	1	13	16	15	0	0
2005/2006	Manchester City	4	2	13	14	16	0	1
2005/2006	Birmingham City	2	5	12	11	17	0	1
2005/2006	West Bromwich Albion	1	7	11	10	18	0	1
2005/2006	AFC Sunderland	2	2	15	8	19	0	1
2005/2006	FC Fulham	1	4	14	7	20	0	1

Tabla B.9: Ejemplo de clasificación según partidos de visita equipos Liga Inglesa temporada 2005/2006

Fuente: Sitio oficial Premier League

B.7. Descripción muestra experimento natural

Birmingham City desciende en la temporada 2007/2008 y en la temporada 2009/2010 logra posición 9 en la tabla de posiciones de la Premier League.

Newcastle United y West Bromwich Albion descienden en la temporada 2008/2009 y en la temporada 2010/2011 logran posiciones 12 y 11 en la tabla de posiciones de la Premier League respectivamente.

West Ham United desciende en la temporada 2010/2011 y en la temporada 2012/2013 logra posición 10 en la tabla de posiciones de la Premier League.

Queens Park Rangers desciende en la temporada 2012/2013 y en la temporada 2014/2015 logra posición 20 en la tabla de posiciones de la Premier League y desciende nuevamente. Es el único caso de los 5 donde ocurre nuevamente después de descender.

Anexos C

Desarrollo Metodológico

C.1. Resultados de los modelos propuestos para análisis de la Liga Inglesa

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad de temporada	Pje mitad de temporada
Primeros 3 partidos	-0.1622 (0.0992)	0.375 (0.271)	0.0162 (0.1176)	0.0285 (0.1715)
Primeros 5 partidos	-0.2604 (0.0982).	0.437 (0.370)	-0.1550 (0.1598)	0.2291 (0.2329)
Primeros 7 partidos	-0.3759 (0.1666)*	0.636 (0.459)	-0.2944 (0.1977)	0.42225 (0.28831)
Partidos de la fecha 19, 20 y 21	-0.0521 (0.964)	0.218 (0.263)		
Partidos de la fecha 18, 19 y 20	0.0531 (0.1022)	-0.0679 (0.2790)		
Partidos de la fecha 17, 18, 19 y 20	-0.1474 (0.1198)	0.396 (0.327)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.0551 (0.1425)	0.153 (0.389)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.1384 (0.1542)	0.470 (0.420)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.1662 (0.0946).	0.370 (0.259)	-0.2417 (0.1107)*	0.2790 (0.1623).
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.0317 (0.1447)	0.0168 (0.3948)	-0.1914 (0.1695)	0.1057 (0.2480)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.1354 (0.1750)	-0.0448 (0.4782)	-0.181 (0.206)	0.13169 (0.30037)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0201 (0.0900)	0.222 (0.245)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.1261 (0.1284)	0.363 (0.350)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	-0.1019 (0.1773)	0.241 (0.484)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.0707 (0.1076)	0.081 (0.294)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.214 (0.145)	0.728 (0.393).		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.0506 (0.1860)	0.329 (0.507)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.1: Resultados de los estimadores con efectos fijos de la regresión considerando la temporada anterior, con dificultad continua

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad de temporada	Pje mitad de temporada
Primeros 3 partidos	0.5597 (0.4163)	-1.257 (1.138)	-0.0187 (0.4924)	-0.0095 (0.7179)
Primeros 5 partidos	0.5404 (0.3426)	-1.053 (0.938)	0.3776 (0.4049)	-0.5886 (0.5901)
Primeros 7 partidos	0.6070 (0.3131).	-1.071 (0.860)	0.4486 (0.370)	-0.7749 (0.5396)
Partidos de la fecha 19, 20 y 21	-0.1241 (0.4421)	-0.173 (1.206)		
Partidos de la fecha 18, 19 y 20	0.000914 (0.412)	-0.189 (1.124)		
Partidos de la fecha 17, 18, 19 y 20	0.3957 (0.3458)	-1.048 (0.943)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.0537 (0.3349)	0.1703 (0.9137)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.1146 (0.3333)	0.115 (0.909)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	0.4603 (0.3780)	-0.915 (1.033)	0.805 (0.07033).	-0.7068 (0.6436)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.0165 (0.3576)	-0.0461 (0.9754)	0.2809 (0.4199)	0.0333 (0.6130)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	-0.0730 (0.3032)	-0.155 (0.827)	0.48642 (0.35439)	-0.2546 (0.5194)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.2339 (0.4379)	-0.0033 (1.1956)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	0.1000 (0.3603)	-0.272 (0.983)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.1161 (0.3404)	-0.221 (0.929)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.3330 (0.4235)	-1.2729 (1.1529)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	0.667 (0.349).	-2.253 (0.947)*		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	0.0981 (0.3058)	-0.751 (0.832)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.2: Resultados de los estimadores con efectos fijos de la regresión considerando la temporada anterior, con dificultad discreta

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	-0.1382 (0.0935)	0.2739 (0.25558)	0.01072 (0.11069)	0.0156 (0.1614)
Primeros 5 partidos	-0.1895 (0.1217)	0.331 (0.334)	-0.1249 (0.1439)	0.169 (0.210)
Primeros 7 partidos	-0.2765 (0.1501).	0.413 (0.413)	-0.1764 (0.1779)	0.2023 (0.2597)
Partidos de la fecha 19, 20 y 21	-0.0308 (0.0877)	0.115 (0.239)		
Partidos de la fecha 18, 19 y 20	0.00887 (0.09494)	0.0878 (0.2589)		
Partidos de la fecha 17, 18, 19 y 20	-0.1653 (0.1094)	0.511 (0.2977).		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.0511 (0.1367)	0.158 (0.373)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.1154 (0.1473)	0.355 (0.402)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.1584 (0.0852)	0.4172 (0.2325).	-0.2427 (0.0993)*	0.3345 (0.1451)*
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.0826 (0.1389)	0.166 (0.379)	-0.173 (0.163)	0.160 (0.238)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.0147 (0.1645)	0.197 (0.448)	-0.23099 (0.19252)	0.2851 (0.3033)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0721 (0.0877)	0.29663 (0.23844)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.1484 (0.1211)	0.4110 (0.3303)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	-0.0529 (0.1672)	-0.0103 (0.4561)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.0416 (0.097)	0.160 (0.264)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.132 (0.131)	0.553 (0.356)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.0263 (0.1774)	0.312 (0.483)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.3: Resultados de los estimadores con efectos fijos de la regresión considerando dos temporadas anteriores, con dificultad continua

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	0.854 (0.397)*	-1.667 (1.092)	0.4604 (0.4729)	-0.67593 (0.68933)
Primeros 5 partidos	0.6913 (0.3343)*	-1.294 (0.919)	0.4987 (0.3965)	-0.675 (0.578)
Primeros 7 partidos	0.6851 (0.2937)*	-1.261 (0.809)	0.417 (0.350)	-0.65179 (0.50956)
Partidos de la fecha 19, 20 y 21	-0.0340 (0.4252)	0.173 (1.160)		
Partidos de la fecha 18, 19 y 20	0.2062 (0.4113)	-0.500 (1.122)		
Partidos de la fecha 17, 18, 19 y 20	0.6454 (0.3534).	-1.572 (0.966)		
Partidos de la fecha 16, 17, 18, 19 y 20	0.2708 (0.3579)	-0.712 (0.976)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	0.2586 (0.3606)	-0.737 (0.983)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	0.6809 (0.3887).	-1.296 (1.066)	0.9238 (0.4555)*	-0.8423 (0.6695)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	0.2268 (0.3550)	-0.721 (0.968)	0.4930 (0.4161)	-0.3561 (0.6088)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.0935 (0.3106)	-0.546 (0.846)	0.497 (0.363)	-0.4391 (0.5315)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0112 (0.4168)	-0.234 (1.137)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	0.281 (0.328)	-0.438 (0.897)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.1981 (0.3127)	-0.218 (0.854)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	-0.0122 (0.4109)	-0.606 (1.120)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	0.394 (0.345)	-1.601 (0.937).		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.1288 (0.3019)	-0.116 (0.824)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.4: Resultados de los estimadores con efectos fijos de la regresión considerando dos temporadas anteriores, con dificultad discreta

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	-0.1048 (0.0761)	0.243 (0.208)	-0.0118 (0.0899)	0.03186 (0.13117)
Primeros 5 partidos	-0.1137 (0.1007)	0.195 (0.275)	-0.0844 (0.1187)	0.1057 (0.1731)
Primeros 7 partidos	-0.1684 (0.1288)	0.280 (0.353)	-0.1234 (0.1520)	0.12702 (0.22180)
Partidos de la fecha 19, 20 y 21	-0.0557 (0.0731)	0.131 (0.199)		
Partidos de la fecha 18, 19 y 20	0.000948 (0.0754)	0.016 (0.206)		
Partidos de la fecha 17, 18, 19 y 20	-0.1508 (0.0912)	0.417 (0.249)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.0743 (0.1125)	0.205 (0.307)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.1265 (0.1195)	0.357 (0.326)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.1396 (0.0713)	0.366 (0.195)	-0.1679 (0.0838)	0.2633 (0.1219)*
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.1107 (0.1112)	0.265 (0.304)	-0.1473 (0.1307)	0.2043 (0.1906)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.0295 (0.1366)	0.177 (0.372)	-0.1205 (0.1604)	0.2194 (0.2336)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0538 (0.0712)	0.1936 (0.1940)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.0756 (0.0953)	0.152 (0.260)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	-0.0264 (0.1319)	-0.110 (0.360)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.00861 (0.07893)	0.159 (0.215)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.088 (0.108)	0.37611 (0.29350)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.0107 (0.1460)	0.212 (0.398)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.5: Resultados de los estimadores con efectos fijos de la regresión considerando tres temporadas anteriores, con dificultad continua

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	1.1087 (0.3862)**	-2.644 (1.060)*	0.525 (0.464)	-0.9162 (0.6755)
Primeros 5 partidos	0.7795 (0.3344)*	-1.701 (0.918).	0.5924 (0.3973)	-0.9380 (0.5784)
Primeros 7 partidos	0.70168 (0.28720)	-1.424 (0.790).	0.3874 (0.3428)	-0.7201 (0.4984)
Partidos de la fecha 19, 20 y 21	0.1307 (0.4427)	-0.187 (1.208)		
Partidos de la fecha 18, 19 y 20	0.1506 (0.3911)	-0.341 (1.067)		
Partidos de la fecha 17, 18, 19 y 20	0.4448 (0.3434)	-0.973 (0.938)		
Partidos de la fecha 16, 17, 18, 19 y 20	0.491 (0.334)	0.1585 (0.9118)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	0.0648 (0.3375)	0.183 (0.921)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	0.430 (0.386)	-0.730 (1.056)	0.7191 (0.4521)	-0.5298 (0.6632)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.1113 (0.3532)	0.320 (0.963)	0.2081 (0.4151)	0.2012 (0.6055)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	-0.1870 (0.3042)	0.353 (0.830)	0.2751 (0.3575)	0.0730 (0.5221)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	0.1957 (0.4181)	-0.495 (1.141)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	0.386 (0.341)	-0.570 (0.934)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.3855 (0.3249)	-0.392 (0.890)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	-0.00516 (0.42248)	-0.496 (1.152)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	0.2437 (0.3515)	-1.522 (0.953)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.2352 (0.3073)	-0.0708 (0.8399)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.6: Resultados de los estimadores con efectos fijos de la regresión considerando tres temporadas anteriores, con dificultad discreto

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	-0.0851 (0.0742)	0.154 (0.203)	-0.0145 (0.0876)	0.0107 (0.1277)
Primeros 5 partidos	-0.0760 (0.0972)	0.0636 (0.2657)	-0.0748 (0.1144)	0.0669 (0.1669)
Primeros 7 partidos	-0.1035 (0.1237)	0.0659 (0.3380)	-0.0874 (0.1456)	0.0456 (0.2124)
Partidos de la fecha 19, 20 y 21	-0.0640 (0.0706)	0.140 (0.193)		
Partidos de la fecha 18, 19 y 20	-0.00092 (0.072)	0.0157 (0.1955)		
Partidos de la fecha 17, 18, 19 y 20	-0.1473 (0.0867).	0.425 (0.236)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.0534 (0.1080)	0.214 (0.294)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.1164 (0.1156)	0.372 (0.315)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.1276 (0.0677).	0.368 (0.185)*	-0.1494 (0.0797).	0.25069 (0.11572)*
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.0752 (0.1076)	0.279 (0.293)	-0.1253 (0.1263)	0.212 (0.184)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.0594 (0.1076)	0.143 (0.352)	-0.0895 (0.1515)	0.198 (0.220)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0818 (0.0690)	0.211 (0.188)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.1074 (0.0923)	0.189 (0.252)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	-0.0712 (0.1277)	-0.0519 (0.3486)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.0198 (0.0768)	0.137 (0.209)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.0681 (0.1043)	0.3131 (0.2837)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.0132 (0.1405)	0.214 (0.383)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.7: Resultados de los estimadores con efectos fijos de la regresión considerando cuatro temporadas anteriores, con dificultad continua

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	1.05139 (0.38668)**	-2.299 (1.064)*	0.5839 (0.4631)	-0.8730 (0.6750)
Primeros 5 partidos	0.77570 (0.3268)*	-1.478 (0.900)	0.5907 (0.3883)	-0.7777 (0.5669)
Primeros 7 partidos	0.71901 (0.28143)*	-1.317 (0.777).	0.366 (0.337)	0.6293 (0.4900)
Partidos de la fecha 19, 20 y 21	-0.305 (0.419)	0.8504 (1.1417)		
Partidos de la fecha 18, 19 y 20	-0.1684 (0.3942)	0.199 (1.076)		
Partidos de la fecha 17, 18, 19 y 20	0.3392 (0.3494)	-0.878 (0.953)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.2001 (0.3431)	0.410 (0.936)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.1371 (0.3328)	0.3610 (0.9078)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	0.4592 (0.3980)	-0.933 (1.088)	0.6089 (0.4675)	-0.4620 (0.6843)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.1658 (0.3549)	0.146 (0.969)	0.2948 (0.4169)	-0.0077 (0.6088)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	-0.2949 (0.3019)	0.627 (0.825)	0.1736 (0.3559)	0.1849 (0.5190)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0540 (0.4185)	0.242 (1.142)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	0.1061 (0.3248)	0.0987 (0.8863)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.1201 (0.3061)	0.149 (0.835)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	-0.1812 (0.4127)	-0.368 (1.126)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.0127 (0.3459)	-0.9273 (0.9407)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.2674 (0.3077)	-0.0865 (0.8414)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.8: Resultados de los estimadores con efectos fijos de la regresión considerando cuatro temporadas anteriores, con dificultad discreto

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	-0.0824 (0.0736)	0.144 (0.201)	-0.0101 (0.0834)	-0.0123 (0.1267)
Primeros 5 partidos	-0.0699 (0.0987)	0.0331 (0.2697)	-0.0647 (0.1161)	0.0286 (0.1695)
Primeros 7 partidos	-0.105 (0.123)	0.0468 (0.3353)	-0.0855 (0.1444)	-0.00433 (0.2108)
Partidos de la fecha 19, 20 y 21	-0.0568 (0.0710)	0.106 (0.194)		
Partidos de la fecha 18, 19 y 20	0.0032 (0.0722)	-0.0015 (0.1971)		
Partidos de la fecha 17, 18, 19 y 20	-0.1430 (0.0867)	0.439 (0.236)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.0353 (0.1082)	0.171 (0.278)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.0973 (0.1164)	0.325 (0.317)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.1164 (0.0672).	0.367 (0.183)*	-0.1316 (0.0791).	0.2360 (0.1776)*
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.0633 (0.1040)	0.277 (0.283)	-0.1156 (0.1221)	0.2199 (0.1776)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.0824 (0.1217)	0.0805 (0.3325)	-0.0699 (0.1432)	0.180 (0.208)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0778 (0.0677)	0.1761 (0.1850)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.1182 (0.0905)	0.185 (0.248)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	-0.0936 (0.1237)	-0.0499 (0.3379)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.0106 (0.0758)	0.162 (0.206)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.0629 (0.1021)	0.317 (0.278)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.0012 (0.1369)	0.201 (0.373)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.9: Resultados de los estimadores con efectos fijos de la regresión considerando cinco temporadas anteriores, con dificultad continuo

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	0.7300 (0.3930).	-1.399 (1.078)	0.1485 (0.4671)	-0.2570 (0.6809)
Primeros 5 partidos	0.7026 (0.3425)*	-1.225 (0.942)	0.3881 (0.4070)	-0.4407 (0.5941)
Primeros 7 partidos	0.6009 (0.2880)*	-1.054 (0.792)	0.266 (0.343)	-0.4095 (0.4996)
Partidos de la fecha 19, 20 y 21	-0.498 (0.425)	1.171 (1.161)		
Partidos de la fecha 18, 19 y 20	-0.3490 (0.4286)	0.825 (1.170)		
Partidos de la fecha 17, 18, 19 y 20	0.2066 (0.3757)	-0.555 (1.025)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.3267 (0.3522)	0.960 (0.960)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.3461 (0.3429)	0.890 (0.936)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	0.4035 (0.3975)	-0.883 (1.086)	0.8238 (0.4643)	-0.7417 (0.6812)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.138 (0.3559)	0.326 (0.979)	0.356 (0.421)	-0.0076 (0.6155)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	-0.2184 (0.3079)	0.642 (0.840)	0.3952 (0.3612)	-0.0288 (0.5286)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.2966 (0.4100)	0.817 (1.118)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	0.0865 (0.3290)	0.183 (0.898)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.0496 (0.3065)	0.448 (0.835)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.1814 (0.4167)	-1.299 (1.133)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	0.222 (0.342)	-1.612 (0.926)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.1588 (0.3031)	-0.453 (0.827)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.10: Resultados de los estimadores con efectos fijos de la regresión considerando cinco temporadas anteriores, con dificultad discreto

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	-0.1676 (0.0974).	0.319 (0.267)	0.0669 (0.1155)	-0.028 (0.1685)
Primeros 5 partidos	-0.2009 (0.1313)	0.252 (0.36)	-0.0282 (0.1555)	0.0564 (0.2267)
Primeros 7 partidos	-0.357 (0.155)*	0.56 (0.426)	-0.175 (0.184)	0.229 (0.269)
Partidos de la fecha 19, 20 y 21	-0.0544 (0.0906)	0.2648 (0.2466)		
Partidos de la fecha 18, 19 y 20	0.0183 (0.0979)	0.0394 (0.2671)		
Partidos de la fecha 17, 18, 19 y 20	-0.1524 (0.1152)	0.468 (0.314)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.0763 (0.1381)	0.231 (0.377)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.1504 (0.1509)	0.5414 (0.4107)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.1501 (0.0881).	0.359 (0.241)	-0.2183 (0.1031)	0.2449 (0.1512)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.041 (0.1307)	0.0038 (0.3566)	-0.1703 (0.1531)	0.0904 (0.22399)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.1284 (0.1592)	-0.218 (0.435)	-0.1163 (0.1874)	0.0379 (0.2735)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.001 (0.093)	0.24 (0.253)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.1008 (0.1247)	0.358 (0.34)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	-0.0949 (0.1669)	0.266 (0.455)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.0817 (0.0996)	0.178 (0.272)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.1497 (0.1471)	0.851 (0.397)*		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	0.0364 (0.1773)	0.459 (0.482)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.11: Resultados de los estimadores con efectos fijos de la regresión considerando 1 temporada anterior según localía, con dificultad continua

Fuente: elaboración propia

Metricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	0.6732 (0.4142)	-1.435 (1.134)	0.0562 (0.4912)	-0.2824 (0.7158)
Primeros 5 partidos	0.5832 (0.3264).	-1.180 (0.895)	0.4852 (0.3858)	-0.781 (0.562)
Primeros 7 partidos	0.5646 (0.2738)*	-1.013 (0.753)	0.3105 (0.3254)	-0.5103 (0.4740)
Partidos de la fecha 19, 20 y 21	-0.024 (0.4090)	0.1488 (1.1156)		
Partidos de la fecha 18, 19 y 20	0.15628 (0.3862)	-0.494 (1.053)		
Partidos de la fecha 17, 18, 19 y 20	0.6088 (0.3396).	-1.623 (0.927).		
Partidos de la fecha 16, 17, 18, 19 y 20	0.3185 (0.3460)	-0.891 (0.944)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	0.3051 (0.3365)	-0.853 (0.918)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	0.6677 (0.379).	-1.490 (1.037)	0.724 (0.446)	-0.7948 (0.6532)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	0.1955 (0.3604)	-0.619 (0.983)	0.316 (0.4234)	-0.2163 (0.6182)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	-0.0342 (0.3001)	-0.0563 (0.8187)	0.4168 (0.3513)	-0.4081 (0.5134)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	0.0972 (0.4244)	-0.981 (1.155)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	0.287 (0.3657)	-0.918 (0.863)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.3541 (0.2905)	-0.926 (0.793)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	-0.0134 (0.4257)	-0.475 (1.161)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	0.319 (0.346)	-1.3834 (0.9391)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.2348 (0.3070)	0.152 (0.839)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.12: Resultados de los estimadores con efectos fijos de la regresión considerando 1 temporada anterior según localía, con dificultad discreta (5 primeros equipos)

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	-0.1483 (0.0908)	0.3293 (0.2483)	0.0363 (0.1076)	0.02798 (0.15694)
Primeros 5 partidos	-0.225 (0.12).	0.4278 (0.3291)	-0.1022 (0.1424)	0.169 (0.208)
Primeros 7 partidos	-0.315 (0.139)*	0.5695 (0.3834)	-0.1475 (0.1660)	0.205 (0.242)
Partidos de la fecha 19, 20 y 21	-0.0077 (0.0896)	0.0915 (0.2442)		
Partidos de la fecha 18, 19 y 20	0.0408 (0.0939)	-0.0052 (0.2563)		
Partidos de la fecha 17, 18, 19 y 20	-0.1168 (0.1106)	0.3548 (0.3014)		
Partidos de la fecha 16, 17, 18, 19 y 20	0.0122 (0.1359)	-0.0123 (0.3707)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.0293 (0.1526)	0.153 (0.416)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.1143 (0.0866)	0.289 (0.236)	-0.1965 (0.1012)	0.2777 (0.1476).
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.0325 (0.135)	0.040 (0.368)	-0.1277 (0.1585)	0.1421 (0.2312)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.09 (0.1628)	-0.0057 (0.4444)	-0.1370 (0.1913)	0.474 (0.288)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0386 (0.0928)	0.240 (0.250)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.0727 (0.1210)	0.203 (0.330)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.0370 (0.1681)	-0.2563 (0.4581)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.0656 (0.0942)	0.0835 (0.2572)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.0799 (0.1306)	0.456 (0.355)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.0170 (0.1693)	0.375 (0.461)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.13: Resultados de los estimadores con efectos fijos de la regresión considerando 2 temporadas anteriores según localía, con dificultad continua

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	1.00368 (1.22886)	-3.7566 (3.3457)	0.8567 (1.4465)	-2.5110 (2.1015)
Primeros 5 partidos	1.6120 (1.0581)	-4.9207 (2.8807).	1.245 (1.249)	-3.7914 (1.8017)
Primeros 7 partidos	1.054 (0.930)	-2.514 (2.538)	-0.4231 (1.0971)	-1.001 (1.598)
Partidos de la fecha 19, 20 y 21	1.255 (1.470)	-4.630 (4.002)		
Partidos de la fecha 18, 19 y 20	2.203 (1.358)	-7.951 (3.681)*		
Partidos de la fecha 17, 18, 19 y 20	1.344 (1.225)	-3.395 (3.342)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.6352 (1.0558)	0.814 (2.883)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.12691 (1.0717)	2.135 (2.932)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	0.777 (1.226)	-1.473 (3.347)	1.529 (1.438)	-1.2612 (2.1023)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.3203 (0.8770)	0.088 (2.393)	0.1817 (1.0316)	-0.395 (1.504)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	-0.1915 (0.6787)	0.00823 (1.85189)	0.07319 (0.79827)	-0.710 (1.162)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	0.33301 (1.48992)	-3.592 (4.055)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.548 (1.269)	1.441 (3.462)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.564 (1.123)	-0.973 (3.064)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	-0.123 (1.392)	-0.606 (3.797)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.0618 (0.9840)	0.2735 (2.6841)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	0.50715 (0.79075)	-2.055 (2.154)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.14: Resultados de los estimadores con efectos fijos de la regresión considerando 2 temporadas anteriores según localía, con dificultad discreta (5 primeros equipos)

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	0.0535 (0.3937)	-0.0371 (1.0740)	0.4828 (0.4613)	-0.4894 (0.6738)
Primeros 5 partidos	0.0175 (0.3172)	-0.376 (0.865)	0.0691 (0.3729)	-0.1585 (0.5436)
Primeros 7 partidos	-0.3918 (0.2950)	0.890 (0.806)	-0.4128 (0.3473)	0.5838 (0.5064)
Partidos de la fecha 19, 20 y 21	-0.4058 (0.3868)	1.840 (1.049).		
Partidos de la fecha 18, 19 y 20	0.1102 (0.3965)	-0.1491 (1.0819)		
Partidos de la fecha 17, 18, 19 y 20	-0.3449 (0.3489)	0.808 (0.952)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.1704 (0.3234)	0.271 (0.883)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.3371 (0.3039)	1.1416 (0.8271)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.6663 (0.3823).	1.479 (1.046)	-0.6007 (0.4514)	0.9797 (0.6571)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.1935 (0.3266)	0.075 (0.892)	-0.2820 (0.3838)	0.1996 (0.5603)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.1457 (0.2817)	-0.0737 (0.7690)	0.0556 (0.3315)	-0.0752 (0.4832)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0826 (0.3637)	1.038 (0.989)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.2279 (0.3268)	0.801 (0.891)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	-0.0826 (0.3637)	0.129 (0.815)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.2627 (0.4469)	0.104 (1.220)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.500 (0.350)	1.3037 (0.9558)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.1923 (0.3234)	0.287 (0.883)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.15: Resultados de los estimadores con efectos fijos de la regresión considerando 1 temporada anterior, con dificultad discreta (5 últimos equipos)

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	-0.0914 (0.4035)	-0.0065 (1.1009)	0.60647 (0.47207)	-0.5439 (0.6905)
Primeros 5 partidos	-0.1076 (0.3207)	-0.388 (0.875)	0.1377 (0.3771)	-0.1202 (0.5499)
Primeros 7 partidos	-0.3892 (0.2821)	0.4568 (0.7735)	-0.2818 (0.3330)	0.344 (0.486)
Partidos de la fecha 19, 20 y 21	-0.1829 (0.3851)	0.9276 (1.0485)		
Partidos de la fecha 18, 19 y 20	0.2457 (0.4025)	-0.390 (1.099)		
Partidos de la fecha 17, 18, 19 y 20	-0.1031 (0.3573)	0.463 (0.974)		
Partidos de la fecha 16, 17, 18, 19 y 20	0.0950 (0.3366)	-0.158 (0.918)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.0670 (0.3010)	0.496 (0.820)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.4094 (0.3918)	1.089 (1.069)	-0.495 (0.461)	0.6012 (0.6723)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.0707 (0.3276)	-0.081 (0.894)	-0.318 (0.384)	0.23305 (0.56135)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.2057 (0.2693)	-0.3365 (0.7356)	0.1459 (0.3171)	-0.2383 (0.4622)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	0.1065 (0.3960)	0.359 (1.080)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	0.0552 (0.3373)	-0.0589 (0.9202)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.3361 (0.3077)	-0.892 (0.839)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.1274 (0.4332)	1.105 (1.179)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.323 (0.364)	1.2480 (0.9900)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.0369 (0.3272)	0.6714 (0.8910)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.16: Resultados de los estimadores con efectos fijos de la regresión considerando 1 temporada anterior según localía, con dificultad discreta (5 últimos equipos)

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pje mitad temporada
Primeros 3 partidos	-0.0936 (0.0677)	0.150 (0.185)	-0.0643 (0.0800)	0.0594 (0.1167)
Primeros 5 partidos	-0.1106 (0.0906)	0.140 (0.248)	-0.1378 (0.1064)	0.152 (0.156)
Primeros 7 partidos	-0.0878 (0.1094)	0.067 (0.299)	-0.0489 (0.1288)	-0.0098 (0.1879)
Partidos de la fecha 19, 20 y 21	-0.0396 (0.0633)	0.0255 (0.1728)		
Partidos de la fecha 18, 19 y 20	0.0258 (0.0653)	-0.0657 (0.1782)		
Partidos de la fecha 17, 18, 19 y 20	-0.0564 (0.0762)	0.221 (0.208)		
Partidos de la fecha 16, 17, 18, 19 y 20	0.0977 (0.0947)	-0.126 (0.259)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	0.0166 (0.1047)	0.0634 (0.2857)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	-0.0439 (0.0597)	0.216 (0.162)	-0.0632 (0.0709)	0.1414 (0.1018)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	0.0504 (0.0851)	0.058 (0.233)	-0.0569 (0.1001)	0.1065 (0.1459)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	0.1664 (0.1074)	-0.113 (0.295)	-0.0808 (0.271)	0.156 (0.185)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.0929 (0.0591)	0.1743 (0.1620)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	-0.186 (0.081)*	0.280 (0.224)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	-0.1491 (0.1059)	0.0799 (0.2906)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	0.08008 (0.06451)	-0.0277 (0.1768)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.0130 (0.0849)	0.168 (0.231)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	0.0402 (0.1119)	0.0967 (0.3051)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.17: Resultados de los estimadores con efectos fijos de la regresión considerando 5 temporadas fijas, con dificultad continua

Fuente: elaboración propia

Métricas /Variable dependiente	Posición final	Puntaje final	Pos mitad temporada	Pos mitad temporada
Primeros 3 partidos	0.4664 (0.4135)	-1.0519 (1.1294)	0.2629 (0.4878)	-0.6173 (0.7100)
Primeros 5 partidos	0.4185 (0.3441)	-0.900 (0.940)	0.3246 (0.4058)	-0.635 (0.591)
Primeros 7 partidos	0.3906 (0.3061)	-0.812 (0.837)	0.0569 (0.3618)	-0.2754 (0.5270)
Partidos de la fecha 19, 20 y 21	-0.682 (0.421)	2.4051 (1.1417)*		
Partidos de la fecha 18, 19 y 20	-0.3711 (0.4223)	1.016 (1.152)		
Partidos de la fecha 17, 18, 19 y 20	0.0521 (0.3849)	0.0456 (1.0500)		
Partidos de la fecha 16, 17, 18, 19 y 20	-0.4406 (0.3600)	1.144 (0.983)		
Partidos de la fecha 16, 17, 18, 19, 20 y 21	-0.4300 (0.335)	1.1832 (0.9098)		
Últimos 3 partidos de la primera fase (fechas 17, 18 y 19)	0.2392 (0.4314)	-0.805 (1.176)	0.6527 (0.5051)	-0.597 (0.739)
Últimos 5 partidos de la primera fase (fechas 15, 16, 17, 18 y 19)	-0.3036 (0.3732)	0.1944 (1.0200)	0.3329 (0.4389)	-0.1783 (0.6409)
Últimos 7 partidos de la primera fase (fechas 13, 14, 15, 16, 17, 18 y 19)	-0.5005 (0.3136)	1.083 (0.858)	0.2457 (0.3713)	0.0998 (0.5419)
Primeros 3 partidos de la segunda fase (fechas 20, 21 y 22)	-0.1041 (0.4313)	0.809 (1.175)		
Primeros 5 partidos de la segunda fase (fechas 20, 21, 22, 23 y 24)	0.57087 (0.35589)	-0.608 (0.978)		
Primeros 7 partidos de la segunda fase (fechas 20, 21, 22, 23, 24, 25 y 26)	0.3299 (0.3185)	0.145 (0.872)		
Últimos 3 partidos de la temporada (fechas 36, 37 y 38)	-0.472 (0.415)	0.780 (1.134)		
Últimos 5 partidos de la temporada (fechas 34, 35, 36, 37 y 38)	-0.129 (0.329)	-0.186 (0.899)		
Últimos 7 partidos de la temporada (fechas 32, 33, 34, 35, 36, 37 y 38)	-0.3394 (0.3001)	0.158 (0.822)		

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla C.18: Resultados de los estimadores con efectos fijos de la regresión considerando 5 temporadas fijas, con dificultad discreta

Fuente: elaboración propia

Anexos D

Regresión con heterogeneidad considerando 4 temporadas anteriores

	Modelo efecto fijo y heterogeneidad
Intercepto	14.0385 (1.4595)***
Primeros 3 partidos	1.1460 (0.6493).
Tipo de Equipo	-0.2930 (0.7259)
Interacción: dificultad y tipo de equipo	-0.0634 (0.4846)

Entre paréntesis la desviación estándar.

. Significativo al 0,10 , * Signif. al 0,05 , ** Signif. al 0,01, *** Signif. al 0,001

Tabla D.1: Resultados de los estimadores de la regresión con dificultad discreta considerando heterogeneidad dinámica (según 4 temporadas anteriores)

Fuente: elaboración propia

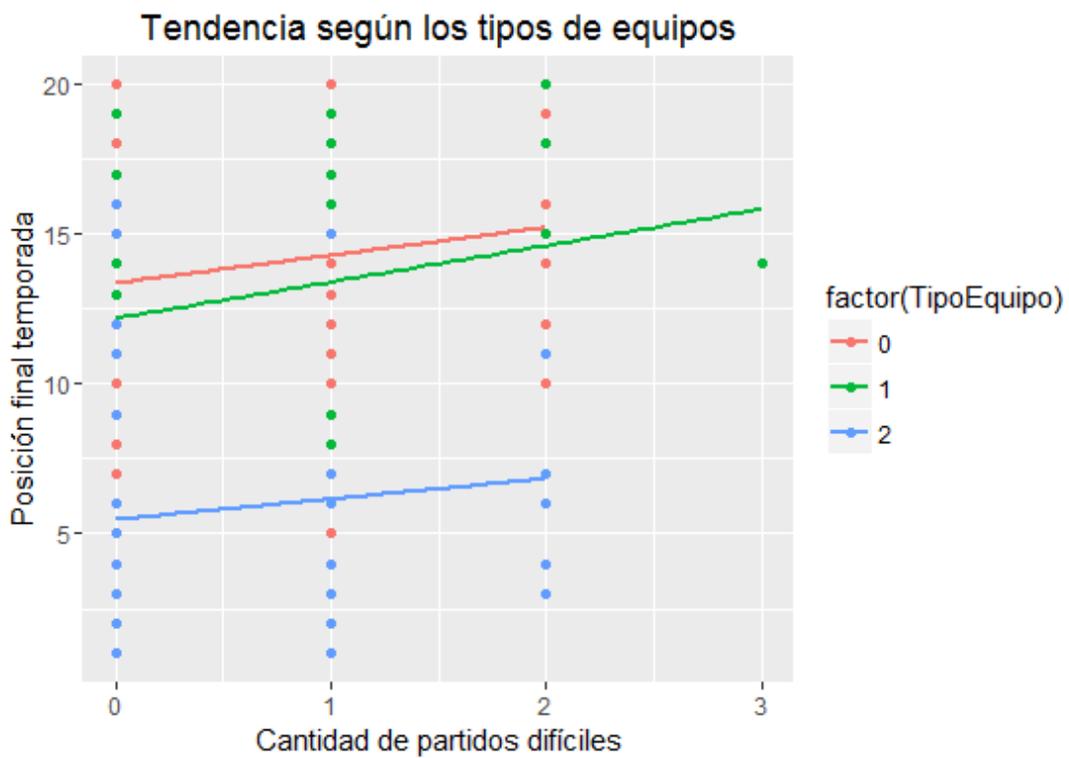


Figura D.1: Tendencia de los equipos según heterogeneidad dinámicas considerando 4 temporadas anteriores

Fuente: elaboración propia