

Modelo sobre los cuartos de final de la Copa del Rey

García García, Aurelio

Núñez Morales, Laura

GECO, grupo A

Índice

1. Motivación del trabajo
2. Modelos
3. Datos
4. Contraste de significación individual
5. Modelos logit
 - Modelo con todas las variables explicativas
 - Efectos marginales
 - Odds ratio
 - Modelo con la variable explicativa goles
 - Conclusiones de los modelos logit
6. Bibliografía

1. Motivación del trabajo

Hemos escogido este trabajo porque nos ha parecido interesante la situación del fútbol español este año. Como sabemos, esta temporada 2013-2014 ha irrumpido un equipo en la liga española el Atlético de Madrid rompiendo la tradicional liga de dos equipos, Real Madrid C.F. y F.C. Barcelona. Partiendo de este hecho, hemos decidido estudiar qué relación existe entre la clasificación de los equipos en la liga y si estos equipos han estado en cuartos de final de la Copa del Rey, es decir, si los primeros equipos en la liga coinciden con los componentes de los cuartos, a fin de observar distintas cuestiones como son la regularidad de los equipos, si estos han tenido un mejor inicio de temporada que final de ésta, etc.

2. Modelos

Para realizar nuestro estudio hemos escogido el modelo de José Eduardo Almaraz Reséndez de mayo de 2011 sobre el fútbol mexicano¹. Así pues, su modelo es el siguiente:

$$\text{Fin} = \text{cte} + \beta_1 \text{Edad} + \beta_2 \text{Edad}^2 + \beta_3 \text{TE} + \beta_4 \text{Tarj} + \beta_5 \text{Gol} + \beta_6 \text{PC} + \beta_7 \text{SC} + \beta_8 \text{TC} + \varepsilon$$

cte: Constante

Edad: Edad media de los jugadores

Edad²: Edad media de los jugadores elevada al cuadrado, la cual representa el máximo de años que puede permanecer un jugador de fútbol en activo.

TE: Títulos del entrenador.

Tarj: Amonestaciones totales de cada equipo ponderadas por 0'5 las tarjetas amarillas y por 1 las tarjetas rojas.

Gol: Goles del equipo, representados a través de la diferencia entre los goles a favor del club y los goles en contra recibidos.

PC: Primer clasificado en la liga

SC: Segundo clasificado en la liga

TC: Tercer clasificado en la liga

¹ (Reséndez, 2011)

En nuestro caso, hemos aplicado este modelo a la Liga BBVA de fútbol con algunas modificaciones, quedando éste de la siguiente forma:

$$\text{Fin} = \text{cte} + \beta_1 \text{Edad} + \beta_2 \text{Tarj} + \beta_3 \text{Gol} + \varepsilon$$

cte: Es la constante

Edad: Edad media de los jugadores

Tarj: Amonestaciones totales de cada equipo ponderadas por 0'5 las tarjetas amarillas y por 1 las tarjetas rojas.

Gol: Goles del equipo, representados a través de la diferencia entre los goles a favor del club y los goles en contra recibidos.

Como podemos apreciar, hemos decidido suprimir algunas variables de este modelo, pues hemos considerado o bien que no eran relevantes desde nuestro punto de visto o bien por realizar un estudio diferente, tanto en el modelo como en la liga, que pueda ser más interesante para nuestro público objetivo.

3. Datos

A continuación exponemos los datos que hemos obtenido sobre la Liga BBVA, de diversas fuentes que serán concretadas en nuestra bibliografía.

| Equipo ² | Puntos ² | Goles a favor ² | Goles en contra ² | Diferencial de goles | Edad media del equipo ³ | Tarjetas amarillas ⁴ | Tarjetas rojas ⁴ | Llega a cuartos de final |
|---------------------|---------------------|----------------------------|------------------------------|----------------------|------------------------------------|---------------------------------|-----------------------------|--------------------------|
| At. de Madrid | 90 | 77 | 26 | 51 | 25'7 | 99 | 2 | 1 |
| FC Barcelona | 87 | 100 | 33 | 67 | 26'7 | 73 | 1 | 1 |
| Real Madrid | 87 | 104 | 38 | 66 | 25'7 | 75 | 2 | 1 |
| At. Club de Bilbao | 70 | 66 | 39 | 27 | 26'0 | 82 | 4 | 1 |
| Sevilla FC | 63 | 69 | 52 | 17 | 25'4 | 130 | 8 | 0 |
| Villareal CF | 59 | 60 | 44 | 16 | 25'3 | 77 | 3 | 0 |

² (Real Federación Española de Fútbol, 2014)

³ (El Desmarque Portal Deportivo SL, 2014)

⁴ (Liga de Fútbol Profesional, 2014)

| | | | | | | | | |
|---------------------|----|----|----|-----|------|-----|---|---|
| Real Sociedad | 49 | 62 | 55 | 7 | 26'2 | 76 | 1 | 1 |
| Valencia CF | 49 | 51 | 53 | -2 | 25'5 | 105 | 4 | 0 |
| RC Celta de Vigo | 48 | 49 | 54 | -5 | 25'9 | 89 | 4 | 0 |
| Levante UD | 45 | 35 | 43 | -8 | 28'6 | 116 | 9 | 1 |
| Málaga CF | 43 | 39 | 46 | -7 | 26'7 | 134 | 9 | 0 |
| Rayo Vallecano | 43 | 46 | 80 | -34 | 25'6 | 107 | 9 | 0 |
| Getafe CF | 42 | 35 | 54 | -19 | 28'2 | 136 | 8 | 0 |
| RCD Espanyol | 42 | 41 | 51 | -10 | 25'6 | 104 | 7 | 1 |
| Granada CF | 41 | 32 | 56 | -24 | 26'2 | 112 | 9 | 0 |
| Elche CF | 40 | 30 | 50 | -20 | 28'4 | 124 | 7 | 0 |
| UD Almería | 40 | 43 | 71 | -28 | 26'7 | 106 | 5 | 0 |
| Club At. Osasuna | 39 | 32 | 62 | -30 | 27'9 | 92 | 8 | 0 |
| Real Valladolid CF | 36 | 38 | 60 | -22 | 27'7 | 105 | 1 | 0 |
| Real Betis Balompié | 25 | 36 | 78 | -42 | 26'5 | 126 | 9 | 0 |

4. Contraste de significación individual

A fin de ver que variables de nuestro modelo son significativas, hemos realizado el contraste de significación individual. De tal modo:

```
Archivo  Editar  Contrastes  Guardar  Gráficos  Análisis

Modelo 3: estimaciones MCO
utilizando las 20 observaciones 1-20
Variable dependiente: y

VARIABLE          COEFICIENTE          DESV.TÍP.          ESTAD T          VALOR P
const             -1,48281             2,31780            -0,640           0,53139
goles              0,00933665          0,00355722         2,625            0,01839 **
Edad              0,0865695           0,0887796          0,975            0,34402
tarj              -0,00810919         0,00887901         -0,913           0,37464

Media de la var. dependiente = 0,35
Desviación típica de la var. dependiente. = 0,48936
Suma de cuadrados de los residuos = 2,27716
Desviación típica de los residuos = 0,377256
R-cuadrado = 0,499525
R-cuadrado corregido = 0,405687
Estadístico F (3, 16) = 5,32322 (valor p = 0,00978)
Log-verosimilitud = -6,65073
Criterio de información de Akaike (AIC) = 21,3015
Criterio de información Bayesiano de Schwarz (BIC) = 25,2844
Criterio de Hannan-Quinn (HQC) = 22,079

Sin considerar la constante, el valor p más alto fue el de la variable 4 (tarj)
```

Basándonos en esta tabla podemos apreciar que la única variable significativa es goles, cuyo coeficiente no muestra que ante un aumento unitario del diferencial de goles, ceteris paribus, la probabilidad de entrar en cuartos de la Copa del Rey es de 0'00933665

Ahora bien, como la mayor parte de las variables son no significativas, vamos a comprobar si estamos ante un problema de multicolinealidad.

Factores de inflación de varianza (VIF)

Mínimo valor posible = 1.0

Valores mayores que 10.0 pueden indicar un problema de colinealidad

| | | |
|----|-------|-------|
| 2) | goles | 1,728 |
| 3) | Edad | 1,205 |
| 4) | tarj | 1,717 |

$VIF(j) = 1/(1 - R(j)^2)$, donde $R(j)$ es el coeficiente de correlación múltiple entre la variable j y las demás variables independientes

Propiedades de la matriz $X'X$:

norma-1 = 104874,35

Determinante = 1,3111656e+010

Número de condición recíproca = 2,4336182e-007

Como vemos, no existen problemas de colinealidad.

Así pues, en el siguiente apartado vamos a comparar nuestro modelo logit con todas las variables que hemos considerado y otro en el que únicamente trabajaremos con la variables goles y la constante, pues es la primera de estas es la única variable que hemos obtenido como significativa.

4. Modelos logit

Modelo con todas las variables explicativas

```
Archivo  Editar  Contrastes  Guardar  Gráficos  Análisis
Iteración 0: log-verosimilitud = -13,2122063541
Iteración 1: log-verosimilitud = -7,51238770241
Iteración 2: log-verosimilitud = -6,67120252104
Iteración 3: log-verosimilitud = -6,48339558263
Iteración 4: log-verosimilitud = -6,46833766823
Iteración 5: log-verosimilitud = -6,46821032391
Iteración 6: log-verosimilitud = -6,46821031361

Se alcanzó la convergencia después de 7 iteraciones

Modelo 1: estimaciones Logit
utilizando las 20 observaciones 1-20
Variable dependiente: y

VARIABLE      COEFICIENTE      DESV.TÍP.      ESTAD T      PENDIENTE
                (en media)
const          -27,0526          21,5896         -1,253
goles           0,0964141        0,0560768       1,719         0,0207762
Edad            1,11968          0,847270        1,322         0,241279
tarj           -0,0599703        0,0646054       -0,928        -0,0129230

Media de y = 0,350
Número de casos 'correctamente predichos' = 18 (90,0%)
f(beta'x) en la media de las variables independientes = 0,215
Pseudo R-cuadrado de McFadden = 0,500483
Log-verosimilitud = -6,46821
Contraste de razón de verosimilitudes: Chi-cuadrado(3) = 12,9614 (valor p 0,004721)
Criterio de información de Akaike (AIC) = 20,9364
Criterio de información Bayesiano de Schwarz (BIC) = 24,9193
Criterio de Hannan-Quinn (HQC) = 21,7139

                Predicho
                0      1
Observado 0    12    1
            1     1    6
```

- **Efectos marginales**

Para ver la variación que hay en la probabilidad de entrar en cuartos de final de la Copa del Rey ante una variación unitaria en los goles (quinta columna), edad media (sexta columna) y amonestaciones (séptima columna) de cada equipo, hemos calculado los efectos marginales en el modelo logit con todas las variables.

| Equipo ⁵ | Diferencial de goles | Edad media del equipo ⁶ | Tarjetas ponderadas | Efectos marginales de los goles | Efectos marginales de la edad | Efectos marginales de las tarjetas |
|---------------------|----------------------|------------------------------------|---------------------|---------------------------------|-------------------------------|------------------------------------|
| At. de Madrid | 51 | 25'7 | 51'5 | 0,37493388 1 | 7,13439795 E-14 | -7,52245056E-01 |
| FC Barcelona | 67 | 26'7 | 37'5 | 0,06729763 1 | 9,60409695 E-13 | -5,90386746E+00 |
| Real Madrid | 66 | 25'7 | 38'5 | 0,22660321 | 8,09113100 E-13 | -2,25306841E+00 |
| At. Club de Bilbao | 27 | 26'0 | 45 | 0,26837045 7 | 8,87638111 E-15 | -1,90526027E-01 |
| Sevilla FC | 17 | 25'4 | 73 | 0,00148731 2 | 5,16482322 E-16 | -7,40323843E-02 |
| Villareal CF | 16 | 25'3 | 41'5 | 2,30289517 7 | 3,54624408 E-15 | -6,69574001E-02 |
| Real Sociedad | 7 | 26'2 | 39 | 0,09272779 | 1,63782603 E-15 | -7,15856880E-02 |
| Valencia CF | -2 | 25'5 | 56'5 | 0,02555125 9 | 2,09496942 E-16 | -4,13303097E-02 |
| RC Celta de Vigo | -5 | 25'9 | 48'5 | 2,22848135 2 | 2,56887263 E-16 | -4,22566452E-02 |
| Levante UD | -8 | 28'6 | 67 | 0,06555833 9 | 5,76010234 E-17 | -1,42467084E-01 |
| Málaga CF | -7 | 26'7 | 76 | 0,04995807 8 | 3,58954753 E-17 | -5,11051840E-02 |
| Rayo Vallecano | -34 | 25'6 | 62'5 | 0,01157313 5 | 5,32022094 E-18 | -2,01882525E-02 |
| Getafe CF | -19 | 28'2 | 76 | 0,09102979 6 | 1,04684772 E-17 | -6,31571242E-02 |
| RCD Espanyol | -10 | 25'6 | 59 | 0,02240264 7 | 7,82474346 E-17 | -4,28254861E-02 |
| Granada CF | -24 | 26'2 | 65 | 0,07850057 3 | 1,26460371 E-17 | -6,62569947E-02 |
| Elche CF | -20 | 28'4 | 69 | 0,06712530 6 | 1,47685831 E-17 | -6,69605450E-02 |
| UD Almería | -28 | 26'7 | 58 | 0,26464552 8 | 1,31139015 E-17 | -5,55416541E-02 |
| Club At. Osasuna | -30 | 27'9 | 54 | 0,06555213 1 | 1,37873708 E-17 | -4,12755487E-02 |
| Real Valladolid CF | -22 | 27'7 | 53'5 | 0,06590653 2 | 3,23728009 E-17 | -4,60328944E-02 |
| Real Betis Balompié | -42 | 26'5 | 72 | 0,05348327 | 1,28181311 E-18 | -5,61654280E-02 |

⁵ (Real Federación Española de Fútbol, 2014)

⁶ (El Desmarque Portal Deportivo SL, 2014)

- **Odds ratio**

Para analizar cuantas veces es más probable que ocurra el suceso “entrar en cuartos de final” del suceso “no entrar en cuartos de final”, hemos calculado el odds ratio.

Así pues, necesitamos la media de nuestras variables:

- Media de la edad: 26,525
- Media del diferencial de goles: 0
- Media de tarjetas: 57,15

De tal modo, la probabilidad de entrar en cuartos de final es:

$$P_i = \exp(-27,0526 + 0,0964141 * 0 + 1,11968 * 26,525 - 0,0599703 * 57,15) / (1 - 27,0526 + 0,0964141 * 0 + 1,11968 * 26,525 - 0,0599703 * 57,15) = 0,10063$$

Una vez obtenido P_i , podemos hallar el odds ratio:

$$\text{Odds ratio} = P_i / (1 - P_i) = 0,10063 / (1 - 0,10063) = 0,11189058$$

Modelo con la variable explicativa goles

```

Archivo  Editar  Contrastes  Guardar  Gráficos  Análisis
Iteración 0: log-verosimilitud = -13,4024871184
Iteración 1: log-verosimilitud = -8,17460426290
Iteración 2: log-verosimilitud = -7,76128663247
Iteración 3: log-verosimilitud = -7,73089728441
Iteración 4: log-verosimilitud = -7,73059155643
Iteración 5: log-verosimilitud = -7,73059151774

Se alcanzó la convergencia después de 6 iteraciones

Modelo 2: estimaciones Logit
utilizando las 20 observaciones 1-20
Variable dependiente: y

VARIABLE      COEFICIENTE      DESV.TÍP.      ESTAD T      PENDIENTE
                                     (en media)
const          -0,764454         0,642884       -1,189
goles          0,0691895         0,0318540      2,172         0,0149975

Media de y = 0,350
Número de casos 'correctamente predichos' = 15 (75,0%)
f(beta'x) en la media de las variables independientes = 0,217
Pseudo R-cuadrado de McFadden = 0,402994
Log-verosimilitud = -7,73059
Contraste de razón de verosimilitudes: Chi-cuadrado(1) = 10,4367 (valor p 0,001235)
Criterio de información de Akaike (AIC) = 19,4612
Criterio de información Bayesiano de Schwarz (BIC) = 21,4526
Criterio de Hannan-Quinn (HQ) = 19,8499

      Predicho
      0      1
Observado 0  11  2
           1   3  4
  
```

Conclusiones de los modelos logit:

- Como vemos, tanto el p valor del contraste de razón de verosimilitud de un modelo como del otro son menores que 0'05, luego rechazamos la hipótesis nula de que todos los coeficientes del modelo son 0, exceptuando la constante. Por tanto, ambos modelos son globalmente significativos.
- Para comprobar la bondad del ajuste utilizamos el R^2 de McFadden, que nos muestra que ninguno de los dos modelos tiene una buena bondad del ajuste, pues ambos valores sobrepasan el intervalo 0'2-0'4 que sería el límite para una buena bondad del ajuste. No obstante hemos de mencionar que el dato del R^2 de McFadden del segundo modelo es muy cercano al valor límite (0'405994).
- Los criterios de Akaike y el bayesiano de Schwarz de los modelos logit son menores en el segundo modelo, luego podemos afirmar que el segundo modelo es mejor para nuestra causa.
- A pesar de los criterios de Akaike y el bayesiano de Schwarz nos dicen que es preferible el segundo modelo, podemos ver que el modelo con todas las variables a realizado una mejor predicción, 90'0% frente a 75'0%.

6. Bibliografía

El Desmarque Portal Deportivo SL. (2014). *La plantilla del Sevilla, segunda más joven de la Liga*. Recuperado el 05 de 2014, de <http://www.eldesmarque.com/sevilla-futbol-club/55811-La-plantilla-del-Sevilla,-segunda-m%C3%A1s-joven-de-la-Liga>

Liga de Fútbol Profesional. (2014). *Liga de Fútbol Profesional*. Recuperado el 05 de 2014, de <http://www.lfp.es/estadisticas/liga-bbva/amarillas/>

Real Federación Española de Fútbol. (2014). *Real Federación Española de Fútbol*. Recuperado el 05 de 2014, de <http://www.rfef.es/competiciones/futbol-masculino/clasificacion/?j=38>

Reséndez, J. E. (2011). *Probando la maldición del superlíder y otras conjeturas del fútbol mexicano utilizando un modelo probit*. Recuperado el 05 de 2014, de https://docs.google.com/file/d/0B4JmvXJLYvRFMGRmMjM1MmUtNmRjNS00YjBhLWFkNTMtMTVjMmM0YmY2NDQy/edit?hl=en_US&pli=1

Diario AS, S.L. (2010). Diario AS, S.L. Recuperado el 05 de 2014 de http://resultados.as.com/resultados/futbol/copa_del_rey/2013_2014/calendario/cuartos_a

Gujarati, D. N. (2004). *Econometría*. México: 4ª ed., reimp.

Chica Olmo, J. (2014). *Modelos de elección discreta*.

Sánchez González, C. (2013). *Econometría I. Tema 3: El criterio de ajuste Mínimos Cuadrados Ordinarios*