

Grau en Estadística

Títol: Supervivència dels jugadors de “La Liga”

Autor: Dídac Figueras Miró

Director: Eva Boj Del Val

Departament: Matemàtica Econòmica, Financera i Actuarial

Convocatòria: 1 de setembre de 2020



RESUM I PARAULES CLAU

Aquest treball recull l'anàlisi de supervivència dels jugadors de futbol que van participar a "La Liga", primera divisió de futbol espanyola, des del principi de la temporada 2009/2010. En aquest cas, quan ens referim a supervivència no parlem de vida, sinó al temps que aguanten formant part de la competició, per això l'esdeveniment fallida és el fet d'abandonar la competició. Durant l'estudi, primer s'ha observat com afecta cada variable predictora al temps de supervivència dels jugadors, en segon lloc, s'ha utilitzat el model de Cox per trobar quines variables són les millors predictors del temps que duren els jugadors abans d'abandonar la competició.

Paraules clau: Temps de supervivència, risc de fallida, esdeveniment, model de Cox, probabilitat de supervivència, hipòtesi de proporcionalitat i nivell de significació.

CLASSIFICACIÓ AMS

62Nxx Survival analysis and censored data

62N01 Censored data models

62N03 Testing

SUMARI

INTRODUCCIÓ.....	5
LA MOSTRA.....	6
METODOLOGIA.....	7
COS DEL TREBALL.....	8
I. ANÀLISI DESCRIPTIU	8
1. Dades qualitatives.....	8
2. Dades quantitatives	8
II. CORBES DE SUPERVIVÈNCIA	18
1. Dades qualitatives.....	18
2. Dades quantitatives	20
III. MODELITZACIÓ.....	30
1. Hipòtesi de proporcionalitat	30
2. Model de Cox	32
3. Model ampliat per les variables que no compleixen la hipòtesi de riscos proporcionals	35
4. Model de Cox estratificat.....	36
5. Model de Cox ampliat	37
CONCLUSIONS	40
BIBLIOGRAFIA	41
ANNEXOS.....	42
1. Competències vinculades al TFG del Grau en Estadística.....	42
2. Codis de R	42

INTRODUCCIÓ

Els motius pels quals m'interesso per aquest tema són dos, per una banda, volia treballar més l'anàlisi de supervivència, per l'altra, no volia caure en temes recurrents, així que vaig optar per aplicar-ho a una de les aficions, actualment, més importants en aquesta societat.

Les hipòtesis a demostrar són:

- El fet de formar part del FC Barcelona o de l'Athletic Club són motiu per tenir un major temps de supervivència donada la seva política de fitxatges i de cantera.
- Tenint en compte que és difícil adaptar-se a l'estil de joc del país, es suposa més temps de supervivència en els jugadors espanyols.
- Jugar en un dels tres equips que han quedat últims a la classificació (CD Tenerife, Valladolid CF i Xerez CD) augmenta el risc de fallida, ja que la majoria de jugadors descendeixen amb els equips a segona divisió.
- S'espera que canviar d'equip sigui motiu d'augmentar el temps de supervivència, ja que al final de l'estudi la meitat dels equips inicials havien descendit a segona divisió.

Els objectius marcats per aquets estudi són dos:

- Conèixer com afecten les diferents variables al temps de supervivència.
- Trobar quin és el millor model predictor del temps de supervivència dels jugadors.

LA MOSTRA

La mostra escollida són els 537 jugadors que van participar de 'La Liga' de la temporada 2009/2010 i el temps d'estudi va des de l'inici d'aquesta fins la finalització el mercat d'estiu posterior a la temporada 2018/2019. D'aquests 537 veurem que 519 han patit l'esdeveniment mentre 18 tenen temps censurats degut a que encara es mantenen en actiu o van deixar la competició quan ja s'havia acabat el temps d'estudi.

Les variables contemplades en l'estudi són les següents:

- Qualitatives: Posició (Atacant, Migcampista, Defensa o Porter), Equip (FC Barcelona, Real Madrid, València CF, Sevilla FC, RCD Mallorca, Getafe CF, Villareal CF, Athletic Club, Atlético de Madrid, Deportivo de la Coruña, RCD Espanyol, Osasuna, UD Almería, Real Zaragoza, Sporting de Gijón, Racing de Santander, Málaga CF, CD Tenerife, Real Valladolid o Xerez CD), Un equip fix (Sí o No) i Nacionalitat (45 països diferents d'Europa, Amèrica, Àfrica o Àsia).

- Quantitatives (Mitjana per temporada): Partits jugats, Partits titular, Partits complets, Partits suplent, Partits substituït, Convocat sense jugar, Minuts, Targetes grogues, Targetes vermelles, Gols, Minuts/gol, Gols penal, Gols pròpia, Gols encaixats, Partits guanyats, Partits empatats, Partits perduts, Temps de supervivència i Lligues guanyades.

La variable temps de supervivència està mesurada en períodes de mercat, 2 cada any, donat que és el moment en que es donen les sortides dels jugadors.

Les fonts de dades utilitzades són les webs <https://www.transfermarkt.es> i <https://www.bdfutbol.com/es>.

METODOLOGIA

El primer pas és fer un anàlisi descriptiu bàsic per veure com es distribueixen les dades, també s'utilitzen gràfics per millorar la visualització d'aquestes.

En segon lloc, amb les variables quantitatives ja agrupades per nivells, s'estudia com afecta cada variable al temps de supervivència dels jugadors mitjançant la llibreria "survival" d'R.

El següent pas abans de començar a modelitzar les dades és comprovar la hipòtesi de riscos proporcionals, donat que utilitzarem el mètode de Cox i es requereix que en les variables incloses al model no es rebutgi aquest supòsit.

Amb les variables resultants es comprova quines són bones predictores del temps de supervivència amb el test de Wald

A continuació es proposen diferents models de Cox amb les variables no descartades i es comparen amb el criteri d'Akaike per decidir quin és el més preferible.

També s'ha utilitzat el model de Cox estratificat, per comprovar si es pot fer un bon model estratificant per les variables que no compleixen amb la hipòtesi de riscos proporcionals.

En últim lloc, s'ha utilitzat el model de Cox ampliat per crear un model que inclou les variables que compleixen amb la hipòtesi de riscos proporcionals i la interacció amb el temps de les que no la compleixen.

Recordem que les dades s'han extret de les webs <https://www.transfermarkt.es> i <https://www.bdfutbol.com/es>.

El software utilitzat per realitzar l'estudi és Rstudio.

Finalment, recordar que al llarg de tot l'estudi, el nivell de significació utilitzat, sempre és 0.05.

COS DEL TREBALL

I. ANÀLISI DESCRIPTIU

1. Dades qualitatives

Posició (Pos): Dels 537 jugadors 37 són porters, 194 defenses, 185 mig-campistes i 121 davanters.

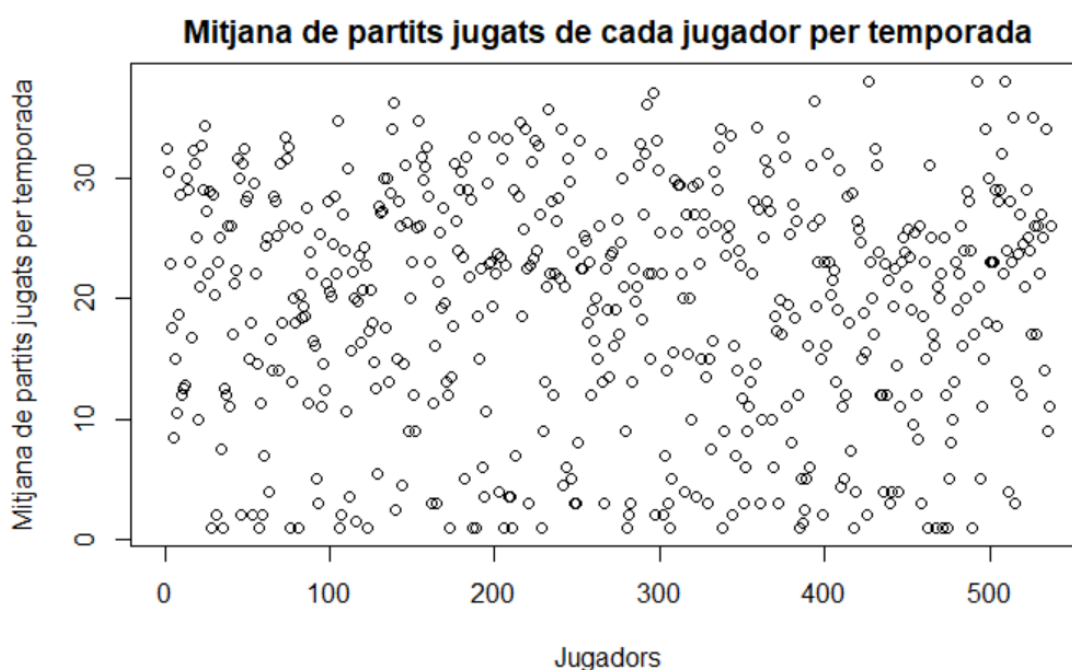
Equip (Team): Hi ha 20 equips, ja esmentats en l'apartat de la mostra. Aquests tenen entre 22 i 32 jugadors a la plantilla, sent el Club Deportivo Tenerife qui en té menys i el Club Atlético de Madrid qui en té més.

Equip fix (Fix): La majoria de jugadors, un 77% aproximadament, formen part del mateix club durant tot el període d'estudi.

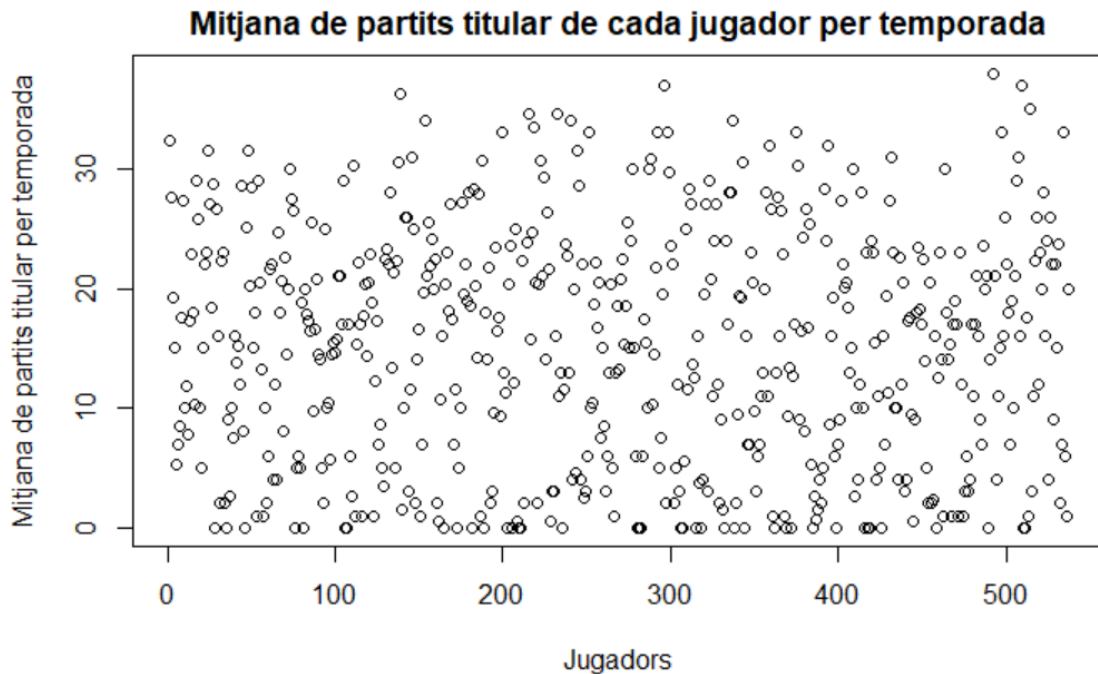
Nacionalitat (Nac): De les 45 nacionalitats que hi ha en la competició, una majoria és l'espanyola, un 65% aproximadament, seguida de l'argentina, brasilenya, portuguesa, uruguaiana i francesa.

2. Dades quantitatives

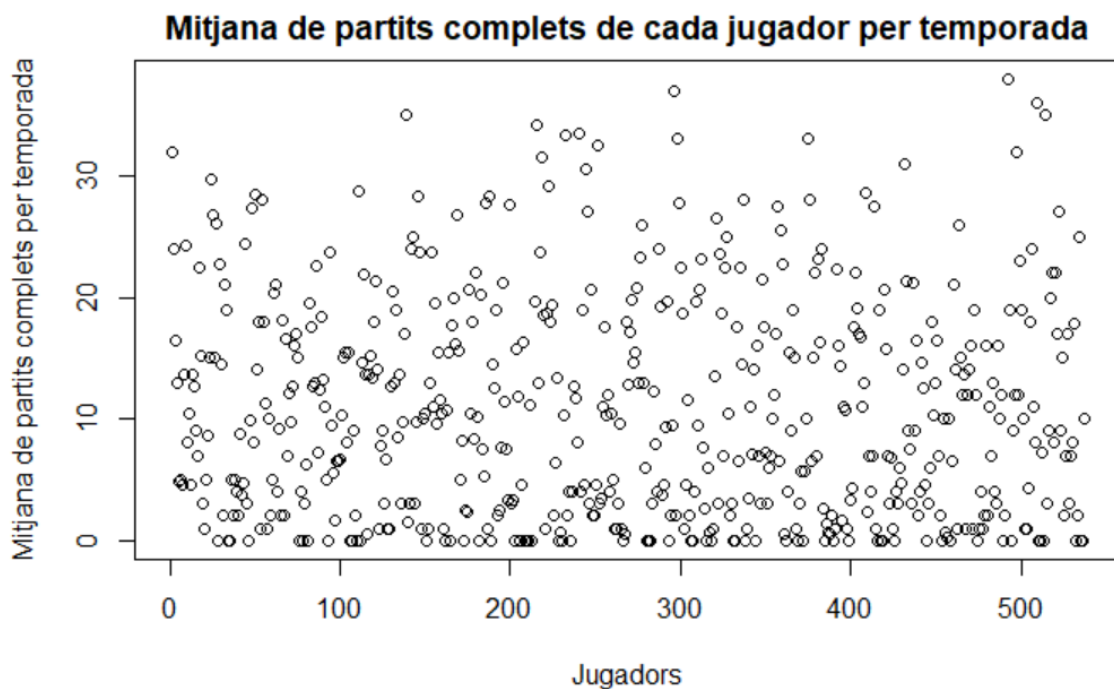
Partits jugats(PJ): El nombre de partits per temporada pot estar entre 1 i 38 que és el nombre total de partits que es juguen en aquesta competició. Es pot destacar que el tercer quartil té un rang molt menor que els altres i com es mostra el gràfic es deu a una alta densitat entre els 20 i 30 partits, a part es pot comprovar que només un nombre molt baix de jugadors es troba entre els 35 i 38 partits.



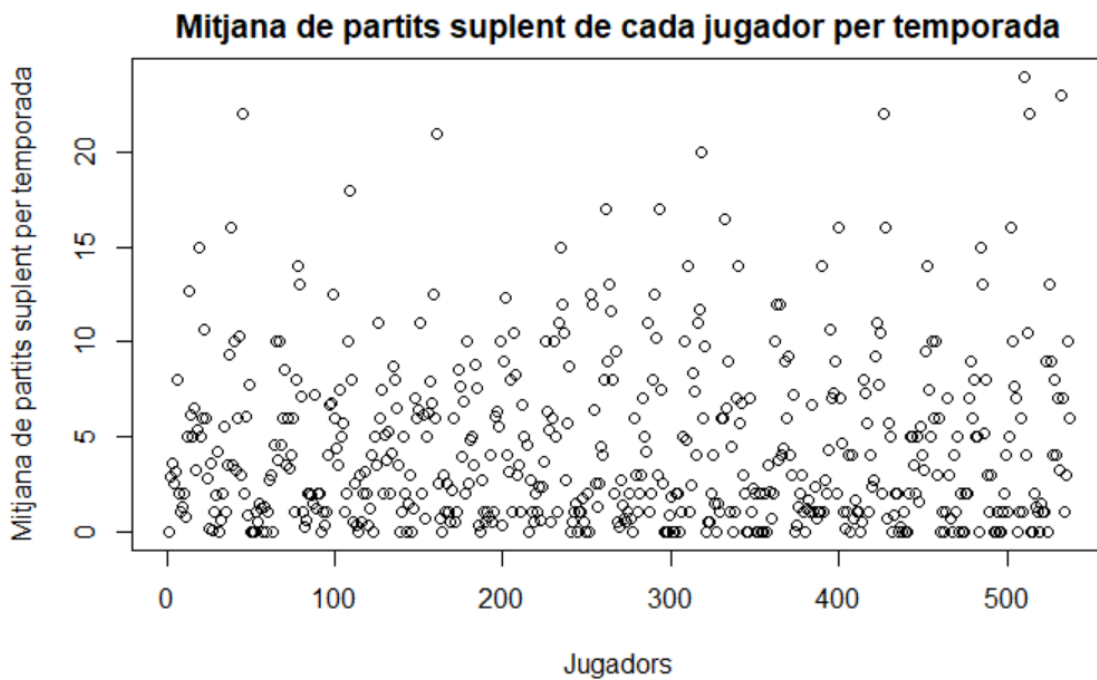
Partits titular (PT): En partits titular el mínim és 0, ja que hi ha un grup nombrós de jugadors que només ha jugat de suplents. Hi ha un sol jugador que ha disputat tots els partits de titular, el que no és gens freqüent en aquest esport, es pot comprovar que hi ha molt pocs jugadors que hagin disputat més de 30 partits de titulars.



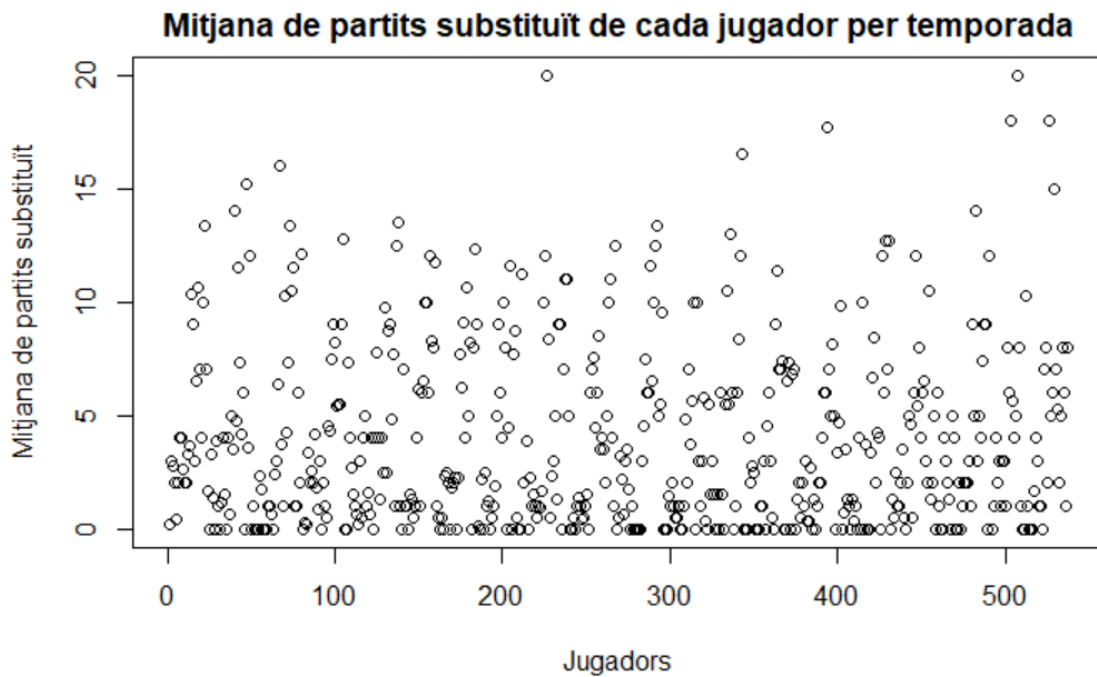
Partits Complets (PC): En partits complets la densitat de jugadors disminueix a mesura que augmenta el nombre de partits, arribant només un jugadors als 38 partits complets i tenint un grup nombrós a zero partits.



Partits suplent (PS): En partits suplent la majoria de jugadors juguen pocs o cap partit, concretament el 75% dels jugadors en disputen un màxim de 6.5 de mitjana cada temporada, per contra set jugadors en disputen entre 20 i 24 que és el màxim.

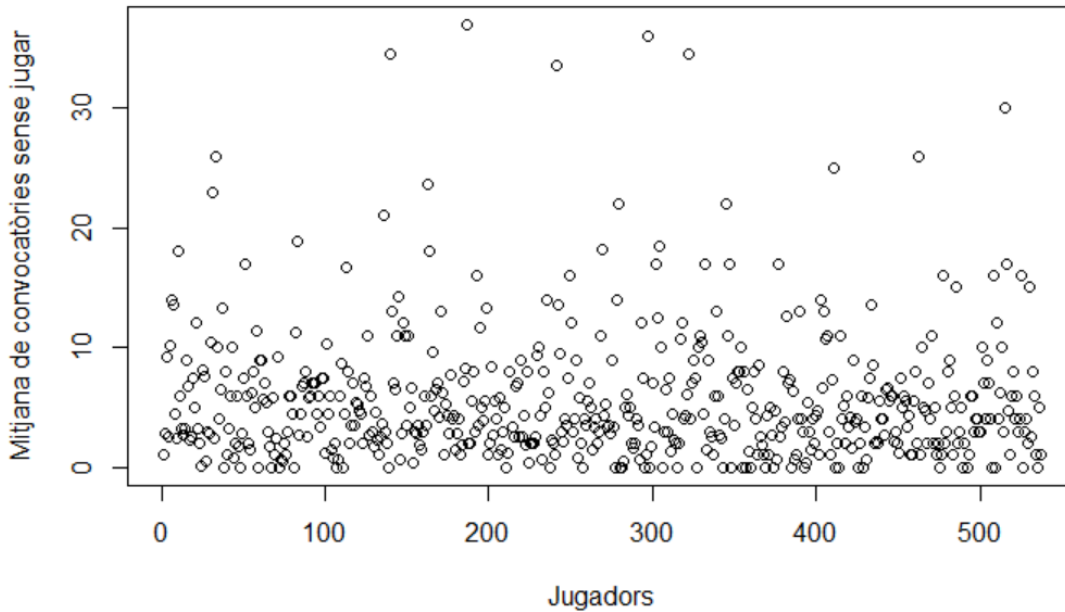


Partits substituït (PX): En partits substituït els resultats són similars als de partits suplent. Hi ha una forta concentració de jugadors d'entre 0 i 6 partits, però en canvi n'hi ha dos que arriben a les 20 substitucions.



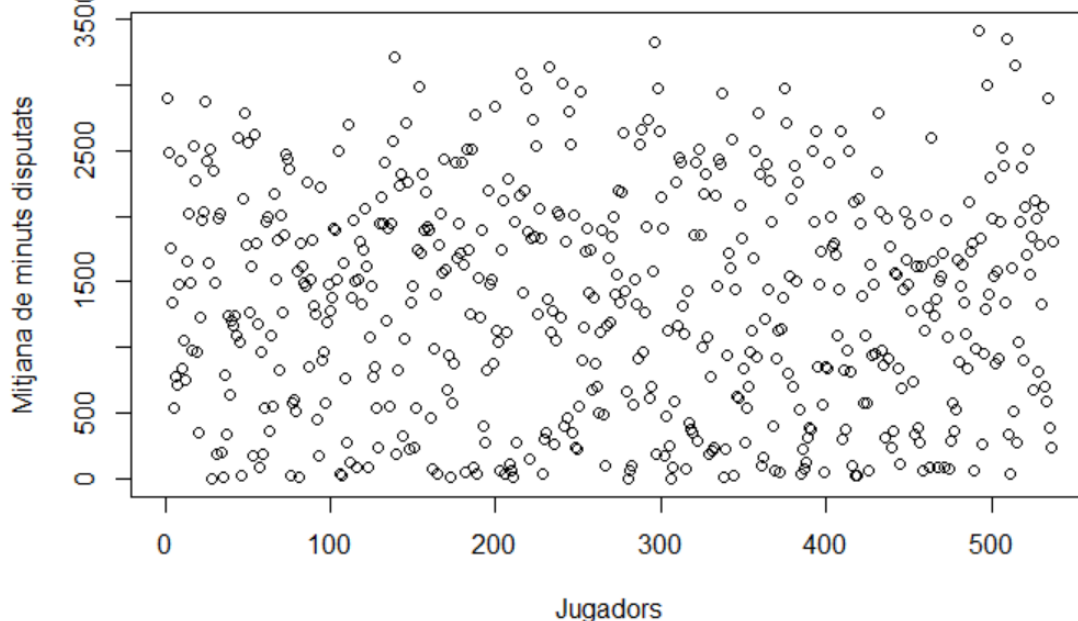
Convocatòries sense jugar (C): En Convocat sense jugar s'observa la majoria de jugadors entre els 0 i 10 partits amb algunes excepcions que poden arribar als 37. Segurament els jugadors que tenen més convocatòries són els porters reserves que sempre van als partits i només juguen si el titular es lesiona o rendeix poc.

Mitjana de convocatòries sense jugar de cada jugador per temporada

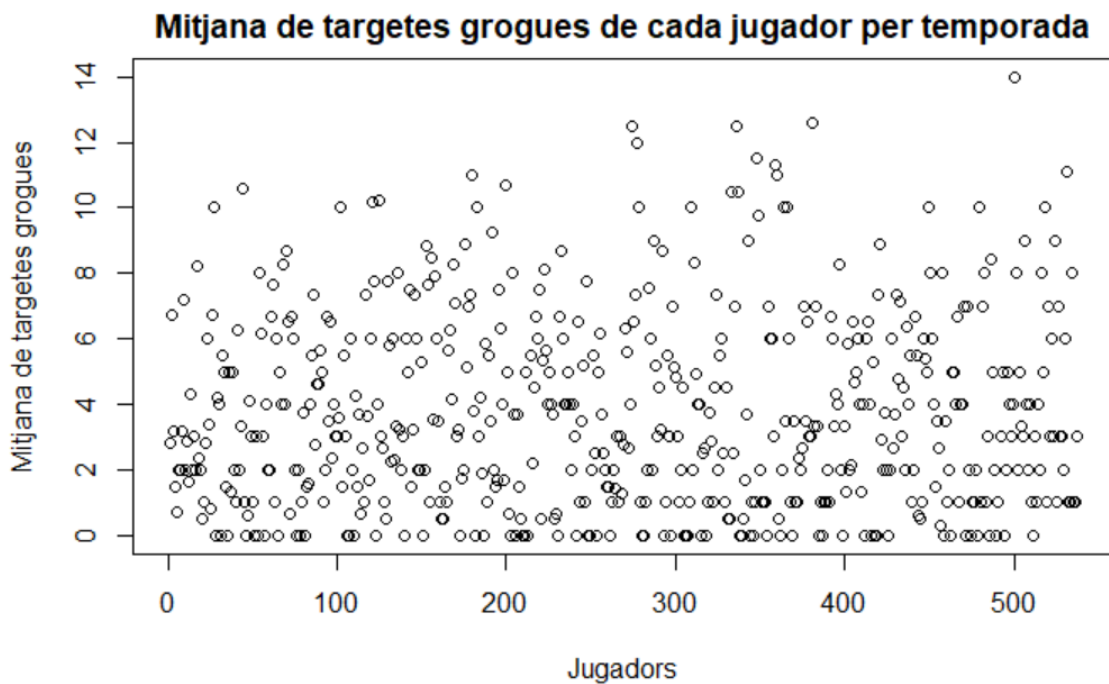


Minuts disputats (Min): En minuts es pot veure que hi ha jugadors que han jugat des d'un sol minut, que és el mínim per formar part de la competició, fins a 3420, que és la suma dels minuts totals dels 38 partits de la temporada. Com es pot veure en el gràfic la mostra és bastant uniforme, però a partir de 2500 minuts es pot veure com disminueix molt el nombre de jugadors que hi arriben, sent només un el que arriba a disputar la totalitat dels minuts que es juguen cada temporada.

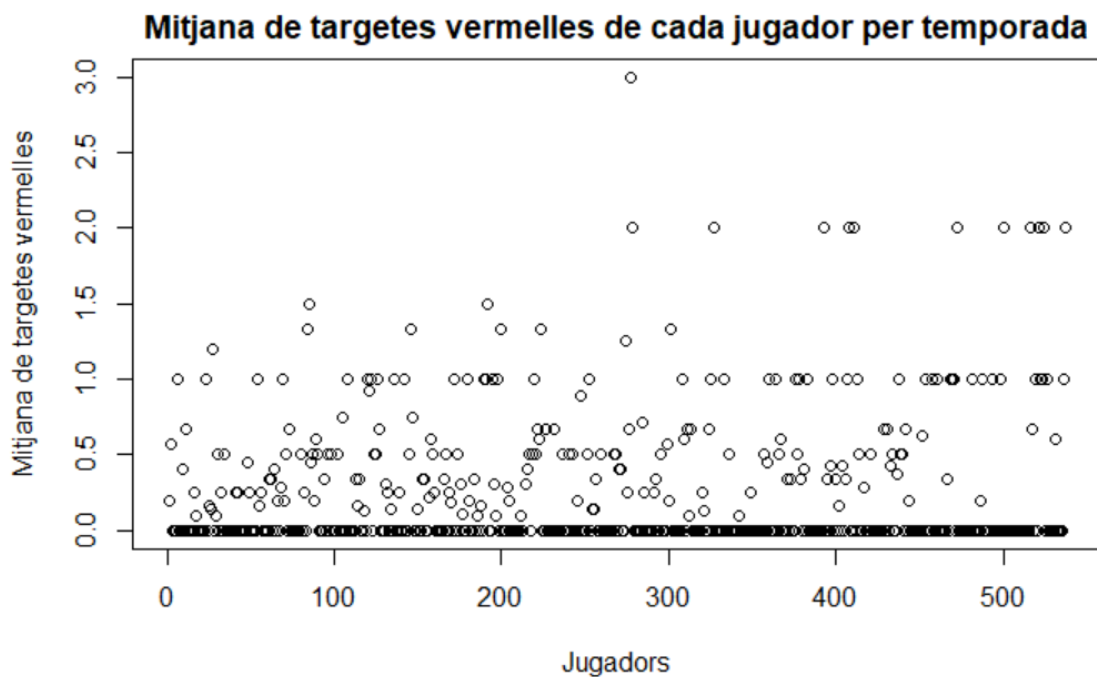
Mitjana de minuts disputats de cada jugador per temporada



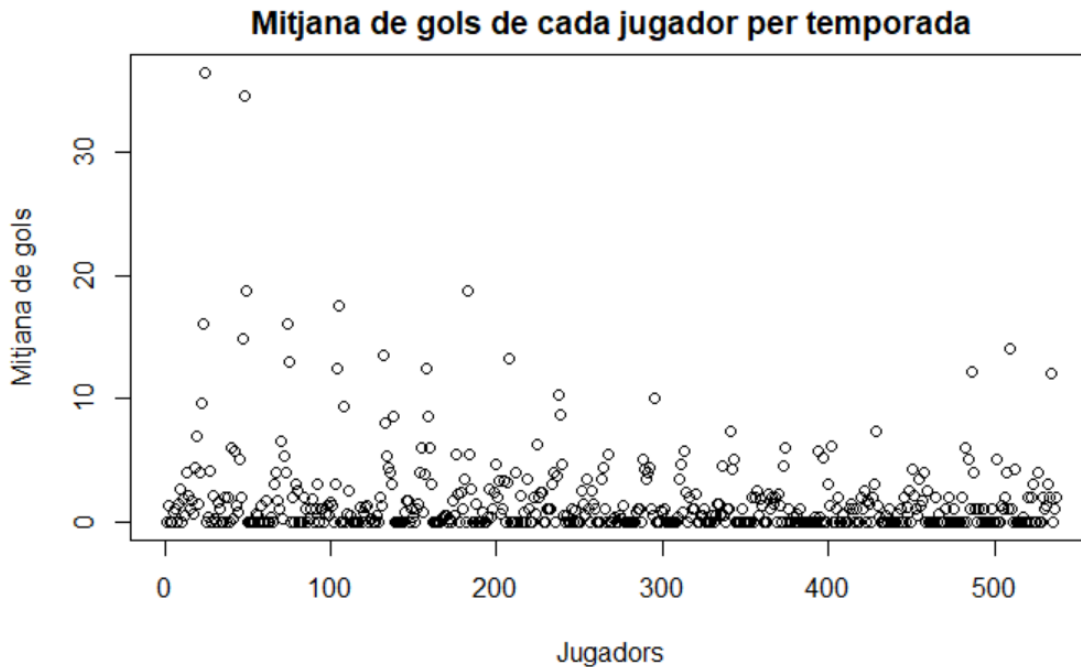
Targetes grogues (TG): En targetes grogues es pot veure que hi ha una forta concentració de jugadors entre les 0 i 4 targetes, i a partir de les 8 targetes cau en picat el nombre de jugadors, arribant un sol a les 14.



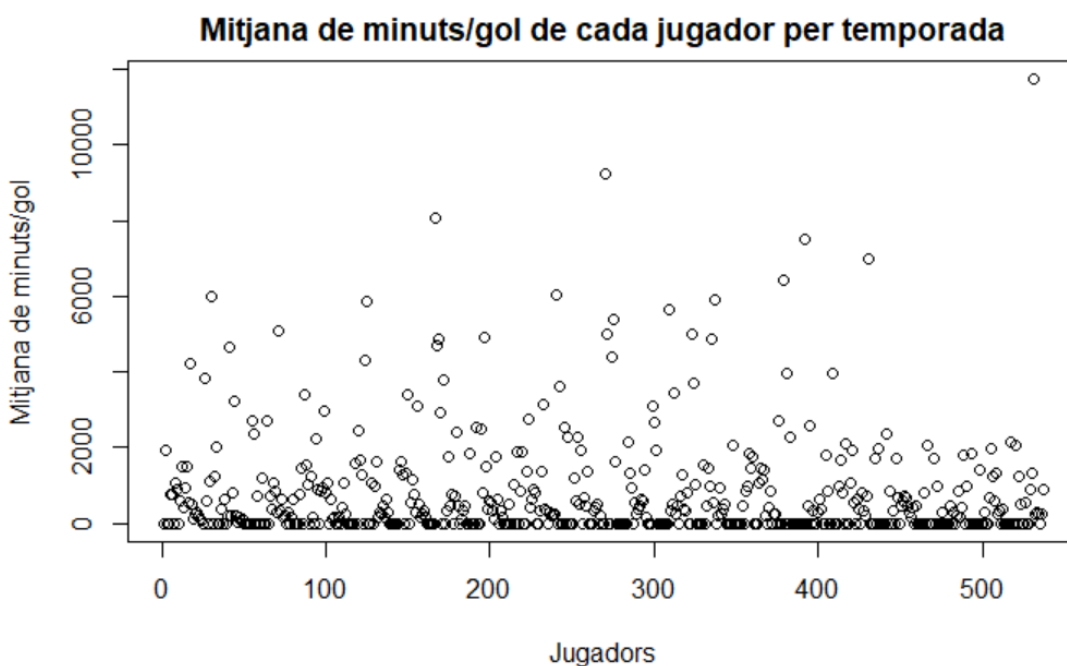
Targetes vermelles (TV): En targetes vermelles la majoria de jugadors no en tenen cap, tot i així, hi ha un grup nombrós de jugadors que es troben entre zero i una targeta per temporada. A part destacar que dels pocs jugadors que estan per sobre d'una targeta, n'hi ha un que arriba a les 3.



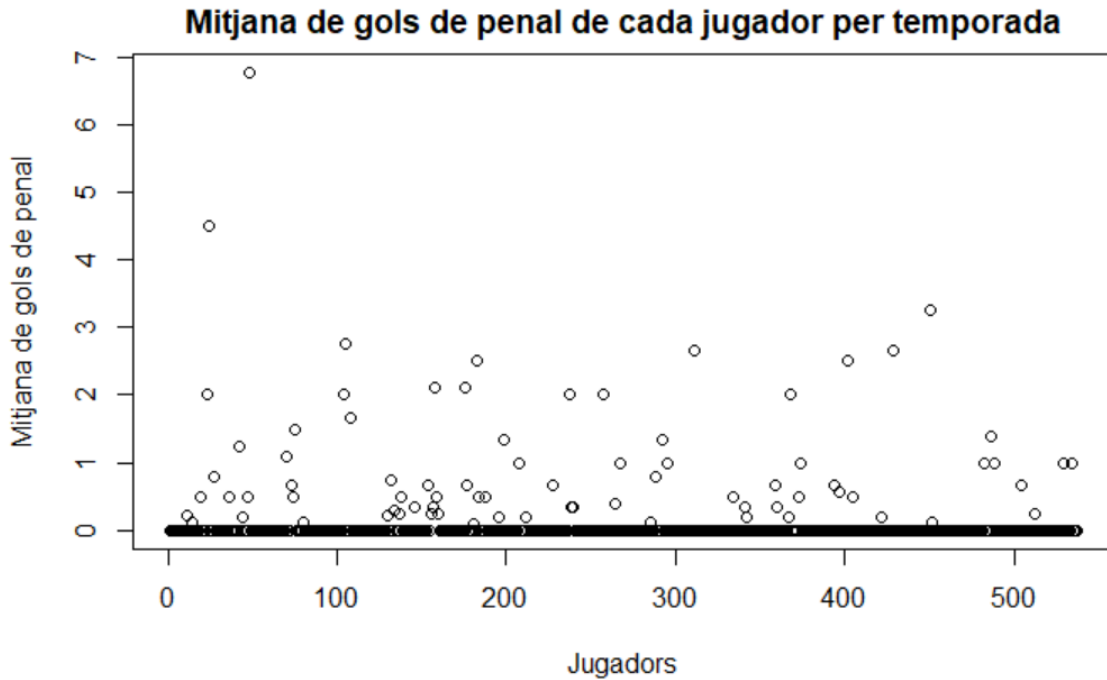
Gols (G): En gols es comprova que quasi la meitat de jugadors no marca cap gol i un 75% no passa dels 2, però hi ha un grup reduït de jugadors que pot arribar a marcar entre 10 i 20 gols per temporada. A part es pot veure dos jugadors que destaquen amb molta diferència amb mitjanes superiors a 30 gols, sense cap dubte Lionel Messi i Cristiano Ronaldo.



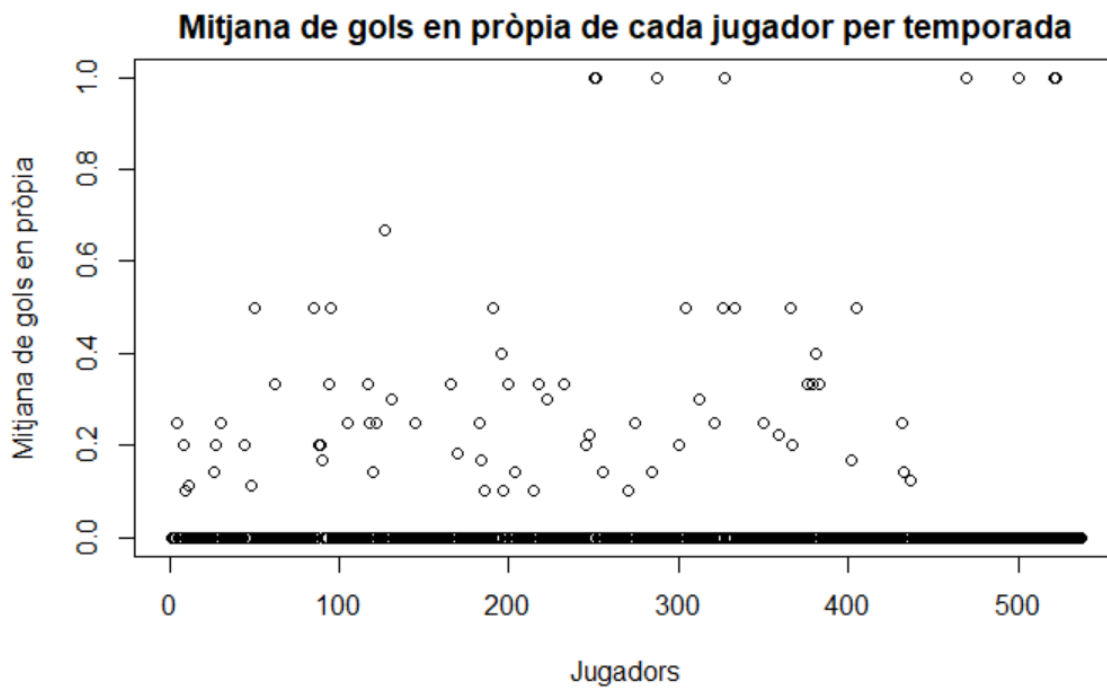
Minuts per gol (MinG): En minuts/gol es pot veure que hi ha un grup nombrós de jugadors que tenen una mitjana de 0 minuts, degut a que no han marcat cap gol en la competició. Un altre grup bastant important marca entre cada 0 i 2000 minuts disputats, a partir d'aquí hi ha un grup reduït de jugadors que tarda més de 2000 minuts de joc en marcar un gol de mitjana.



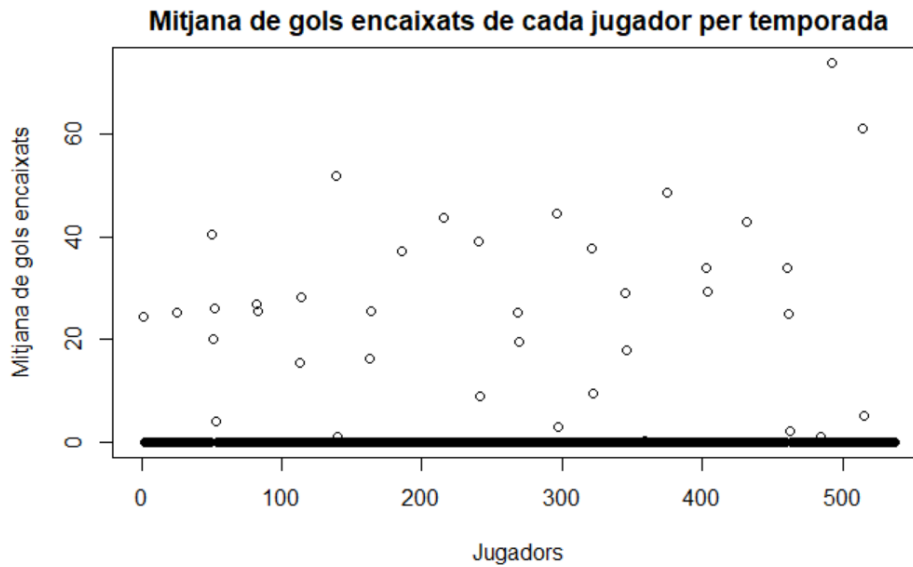
Gols de penal (GP): En gols de penal molt pocs jugadors tenen una mitjana superior a 0 perquè solen ser els mateixos jugadors els que sempre els llencen. Dels jugadors que han marcat la majoria no passa de 3 gols per temporada, excepte dos jugadors amb una mitjana de 4.5 i 6.78 gols.



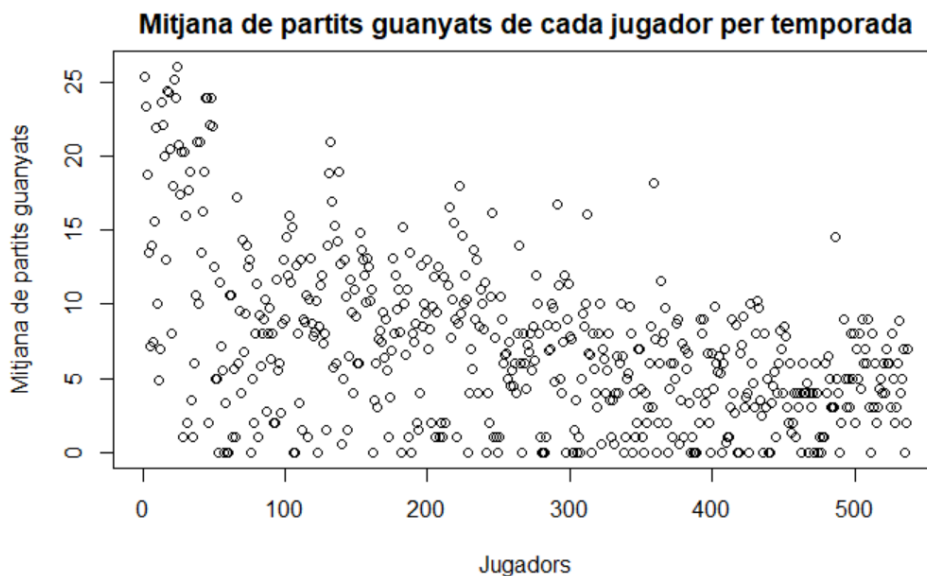
Gols en pròpia (GPP): En gols en pròpia molt pocs jugadors n'han marcat i els que ho han fet no pasen dels 0.5 de mitjana per temporada. Excepte un grup molt reduït de 9 jugadors que han arribat a marcar 1 gol en pròpia de promig per temporada.



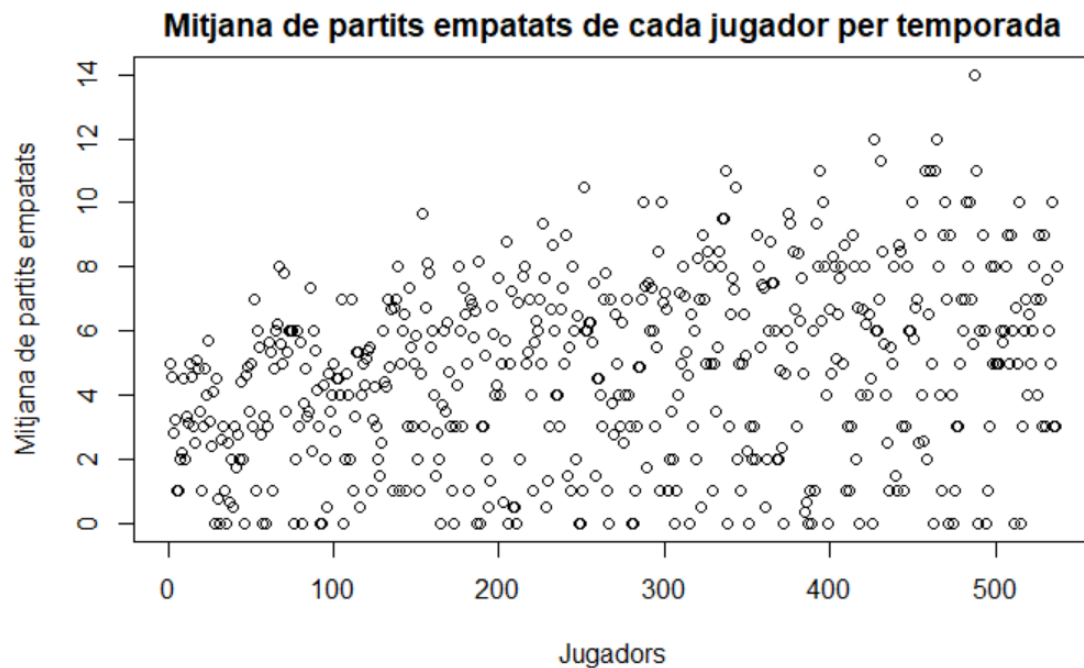
Gols encaixats (GE): En gols encaixats pocs jugadors en tenen, degut a que s'ha de jugar de porter per encaixar gols, a part és una posició que normalment no requereix canvis durant els partits. Es pot veure que la concentració més important, obviant els que en tenen 0, es troba entre 20 i 40 gols encaixats, tot i així n'hi ha que en poden encaixar més, arribant un porter als 74 gols encaixats.



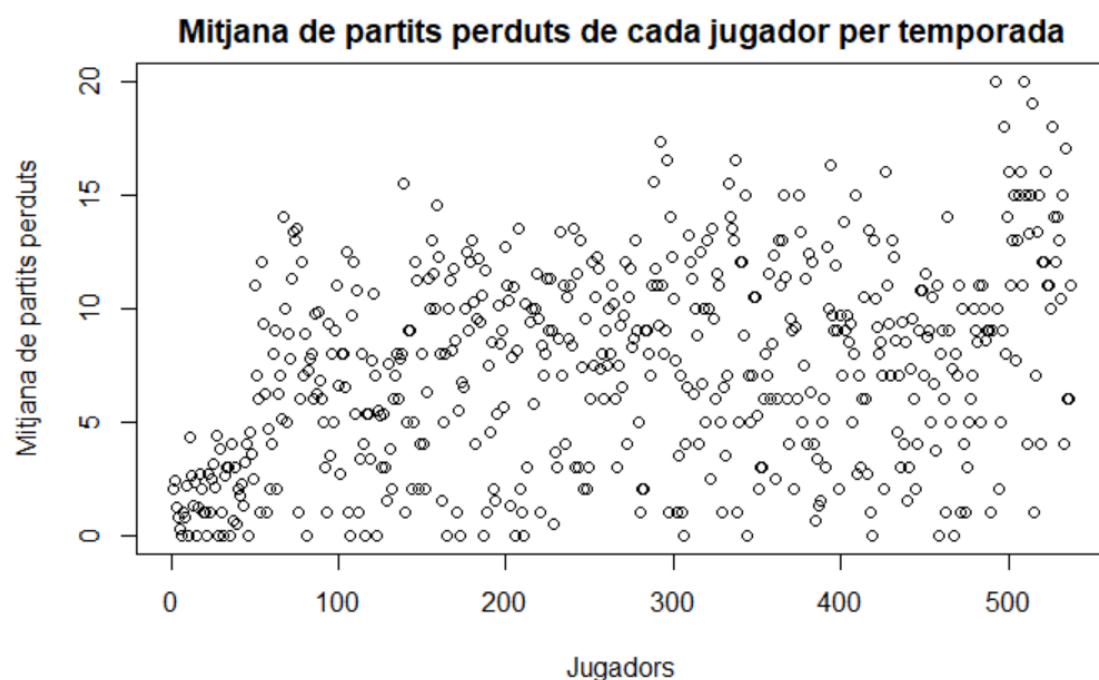
Partits guanyats (PG): En partits guanyats es poden diferenciar tres punts de concentració, segurament influït per l'ordre en que s'han introduït les dades a la base de dades. Es van introduint les dades per equips en funció de l'ordre de la classificació de la temporada 2009/2010, començant pel FCB i acabant pels descendits. Per tant hi ha de menor a major concentració de jugadors, un primer grup concentrat entre els 20 i 25 partits guanyats (FCB i RM, equips que han dominat la lliga durant la última dècada), un segon entre 10 i 15 partits (equips que s'han mantingut molts anys a la part superior de la classificació amb certa regularitat) i un tercer grup d'entre 0 i 10 partits (equips que en general han baixat a segona divisió a conseqüència dels seus mals resultats).



Partits empatats (PE): En partits empatats també es poden veure diferents resultats en funció de la classificació de la temporada inicial, però en aquest cas són els jugadors d'equips pitjor classificats els que sumen més partits empatats i no hi ha grups tan diferenciats, l'evolució és més lineal.



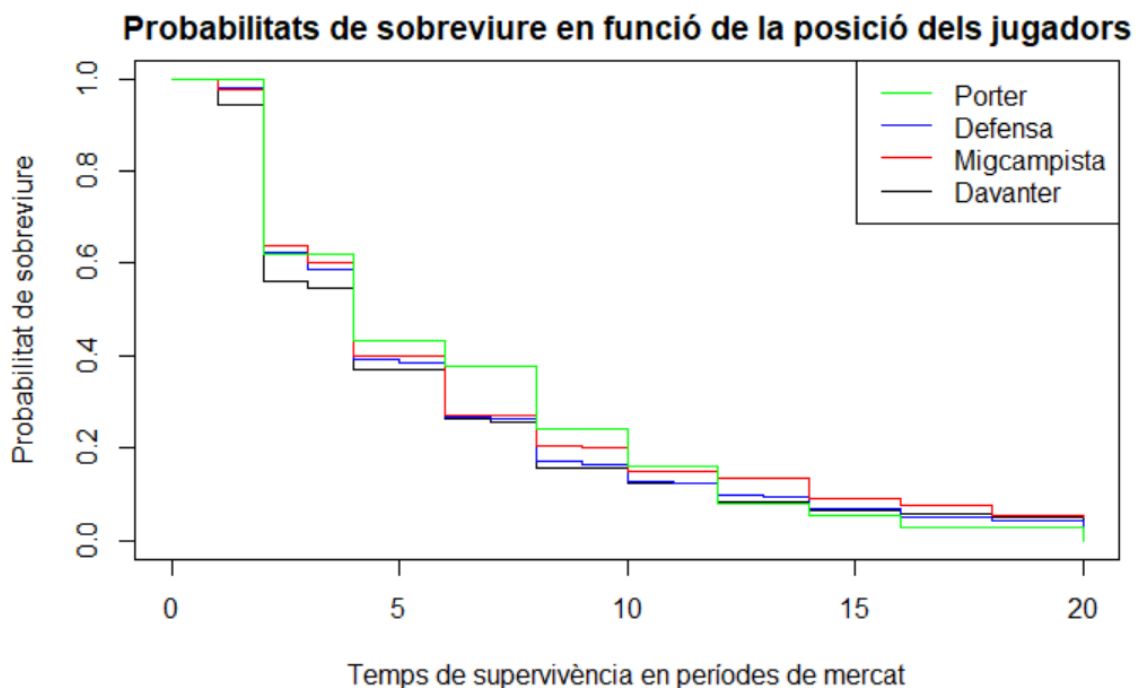
Partits perduts (PP): En partits perduts, com era d'esperar, el gràfic és similar al de partits guanyats però amb l'ordre invers. Els primers equips no tenen cap jugador amb una mitjana superior a 5 derrotes per temporada, els jugadors d'equips que van descendir durant la primera temporada tenen un número de derrotes superior a la resta de jugadors i finalment els que queden al mig en la classificació tenen un volum de partits perduts bastant homogeni.



II. CORBES DE SUPERVIVÈNCIA

1. Dades qualitatives

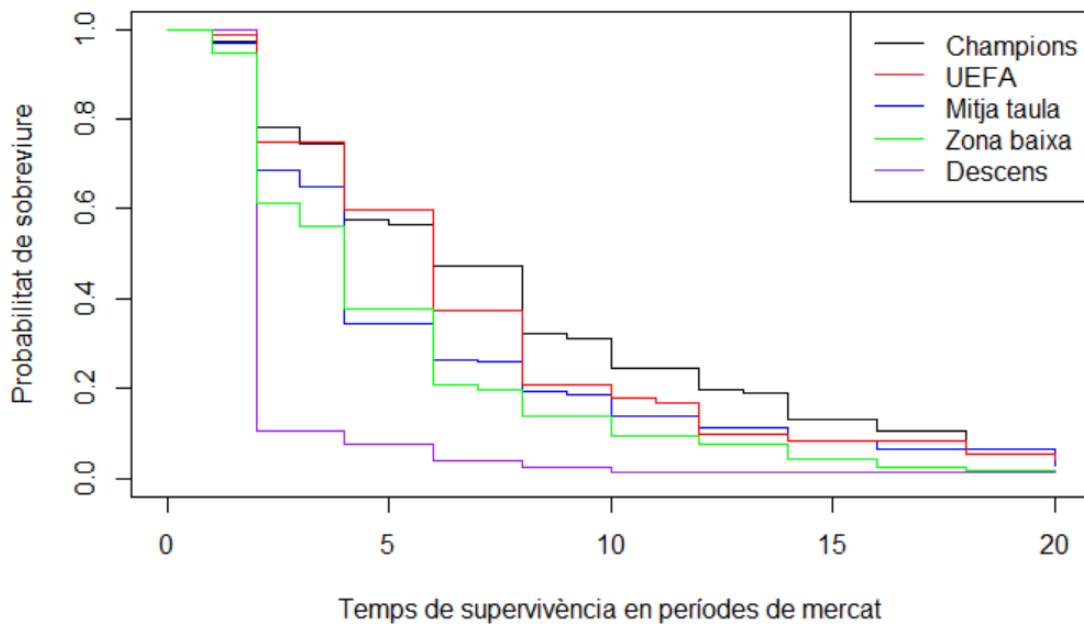
Posició: La probabilitat de supervivència és bastant similar en totes les posicions, però si es mira molt detalladament alguna petita diferència es pot trobar. Fins el període 12 els porters tenen una lleugera superioritat en la corba de supervivència, però en aquest període passa a ser la pitjor. Per altra banda davanters i defenses tenen corbes pràcticament iguals i la corba dels migcampistes es superior a aquestes dos, però les diferències són mínimes.



Equip: Els jugadors que formen part d'equips que acaben en la zona de descens tenen una probabilitat molt baixa de continuar formant part de la competició, mentre els que estan en equips de Champions tenen més probabilitats de sobreviure envers la resta de jugadors durant la majoria de períodes. En general les probabilitats de sobreviure guarden relació amb la posició en que queden els seus equips en la classificació de la competició, com millors resultats més probabilitats de sobreviure.

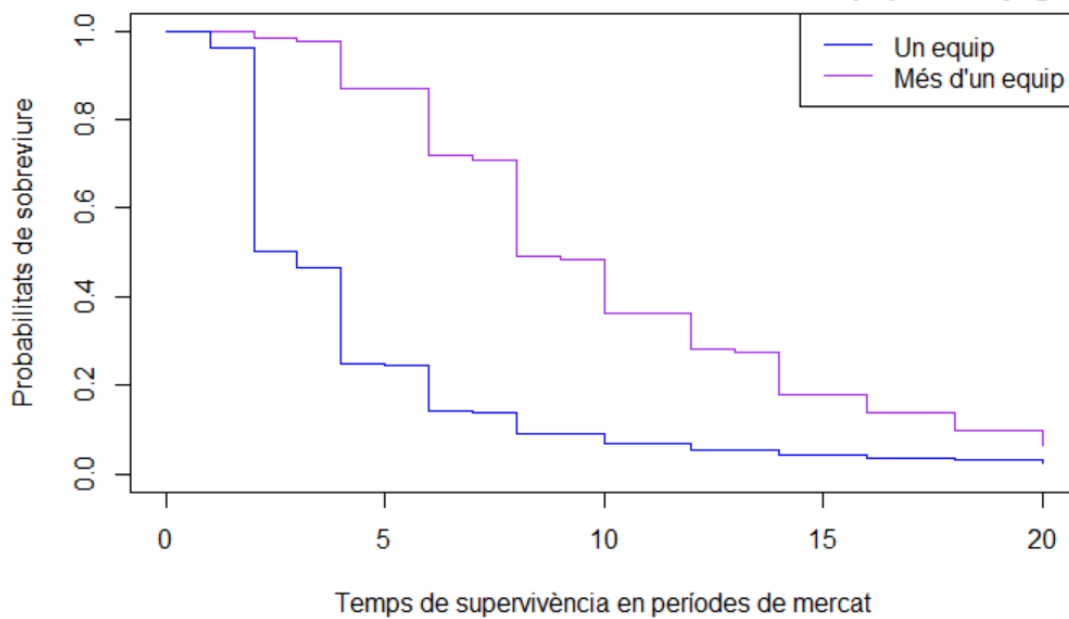
També destacar que les probabilitats de sobreviure disminueixen més durant els períodes parells (mercat d'estiu).

Probabilitats de sobreviure en funció dels equips dels dels jugadors

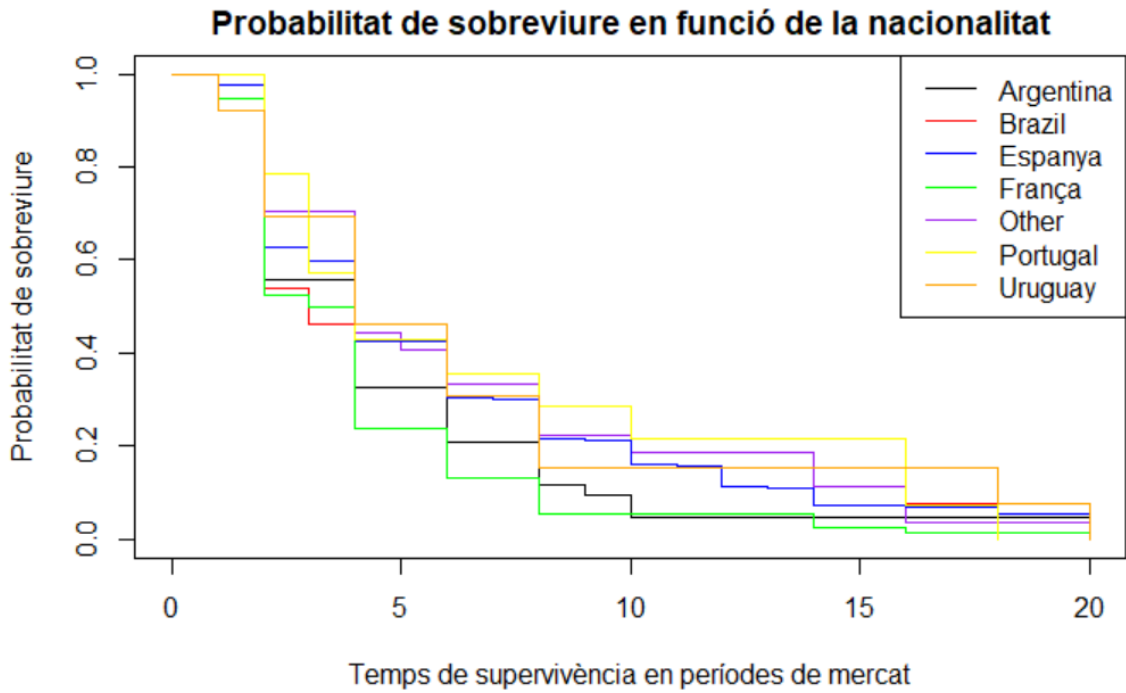


Equip fix: Els jugadors que es mantenen en un mateix equip tenen menys probabilitats de sobreviure, el qual és d'esperar, donat que cada any descendeixen tres equips i amb aquests la majoria dels seus jugadors si no canvien d'equip.

Probabilitat de sobreviure en funció del nombre d'equips dels jugadors

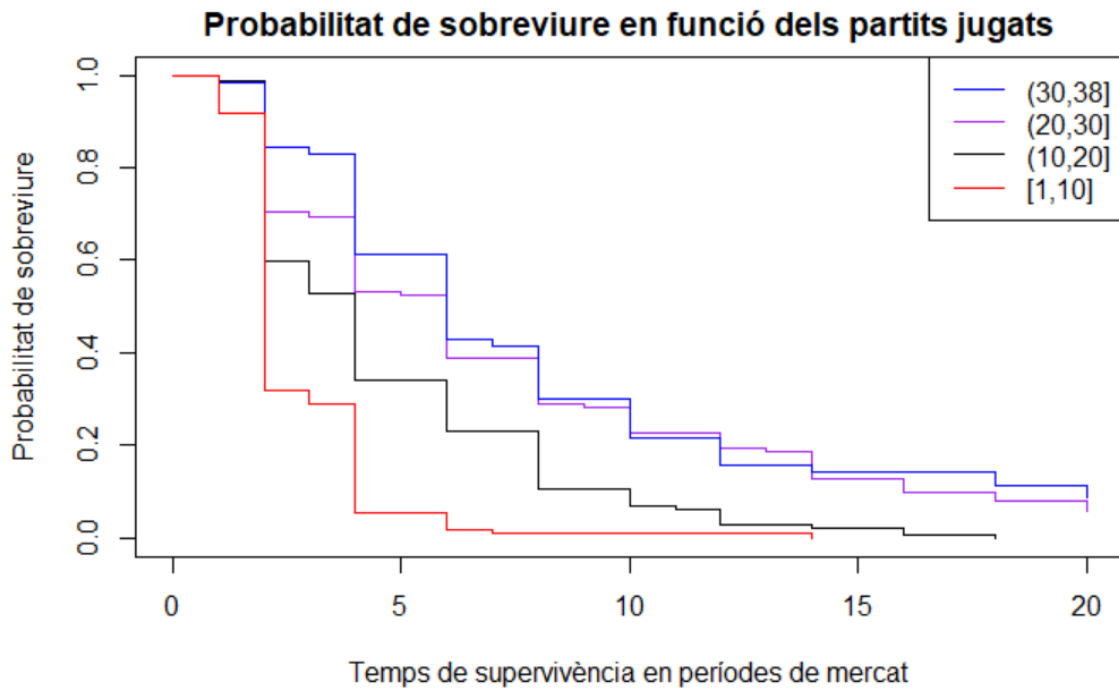


Nacionalitat: En general la probabilitat de sobreviure és similar per totes les nacionalitats amb els seus alts i baixos, excepte França i Argentina que destaquen negativament.

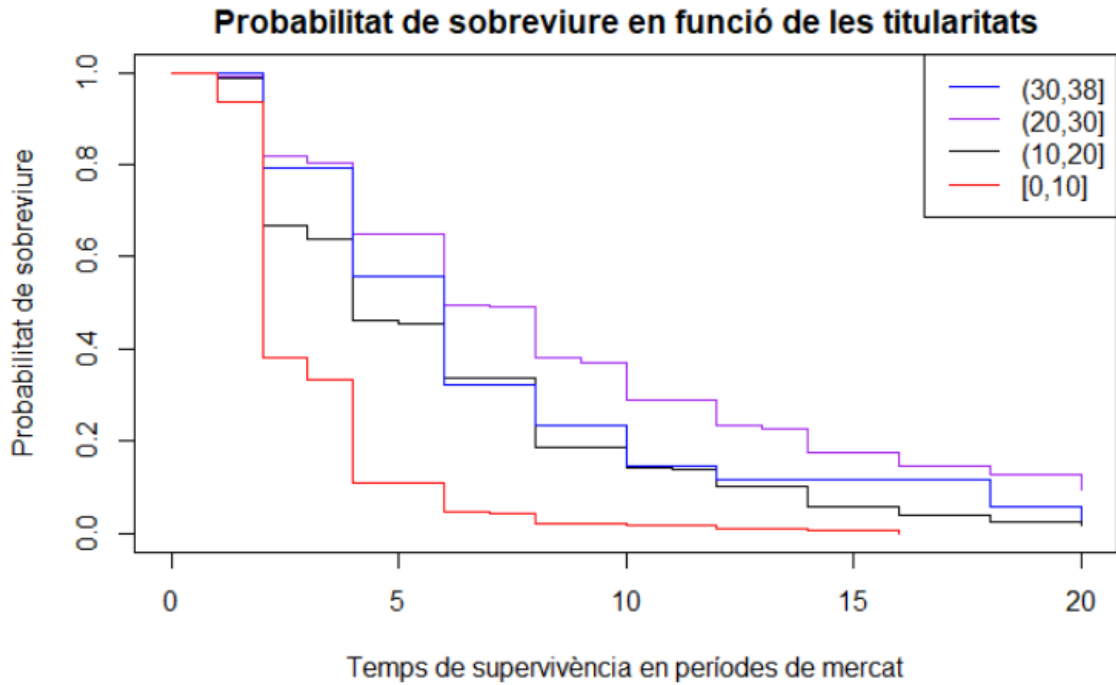


2. Dades quantitatives

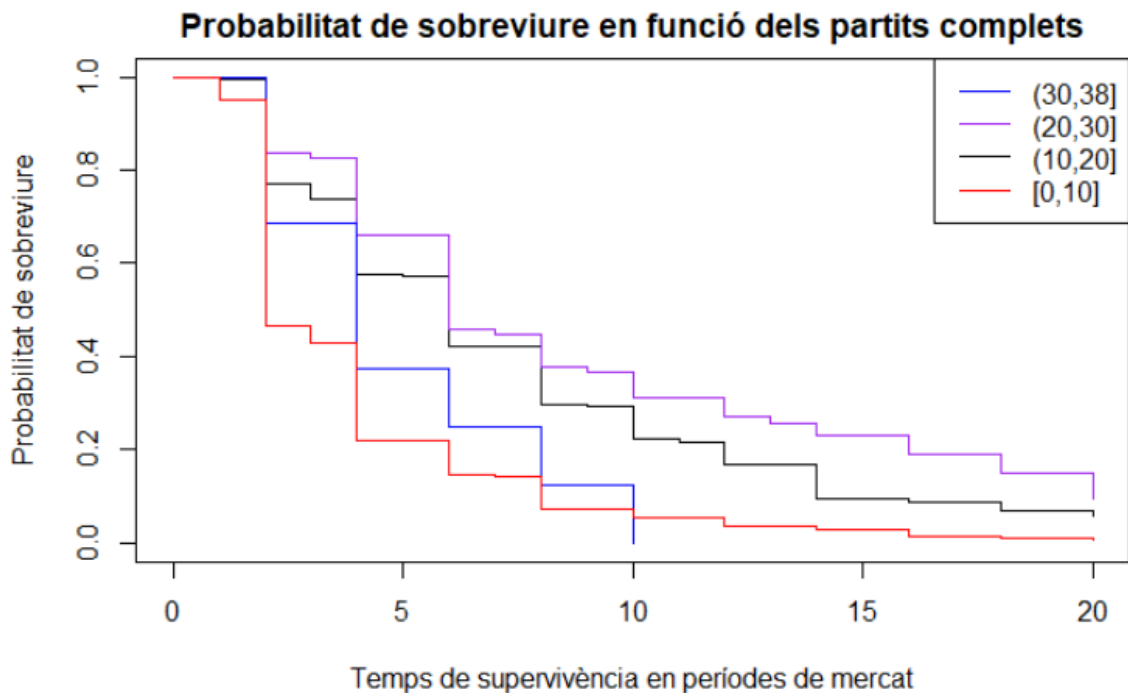
Partits jugats: Es pot veure que la probabilitat de sobreviure és major com més partits es disputen per temporada, tot i que a partir dels 20 partits la diferència que es va guanyant és molt baixa.



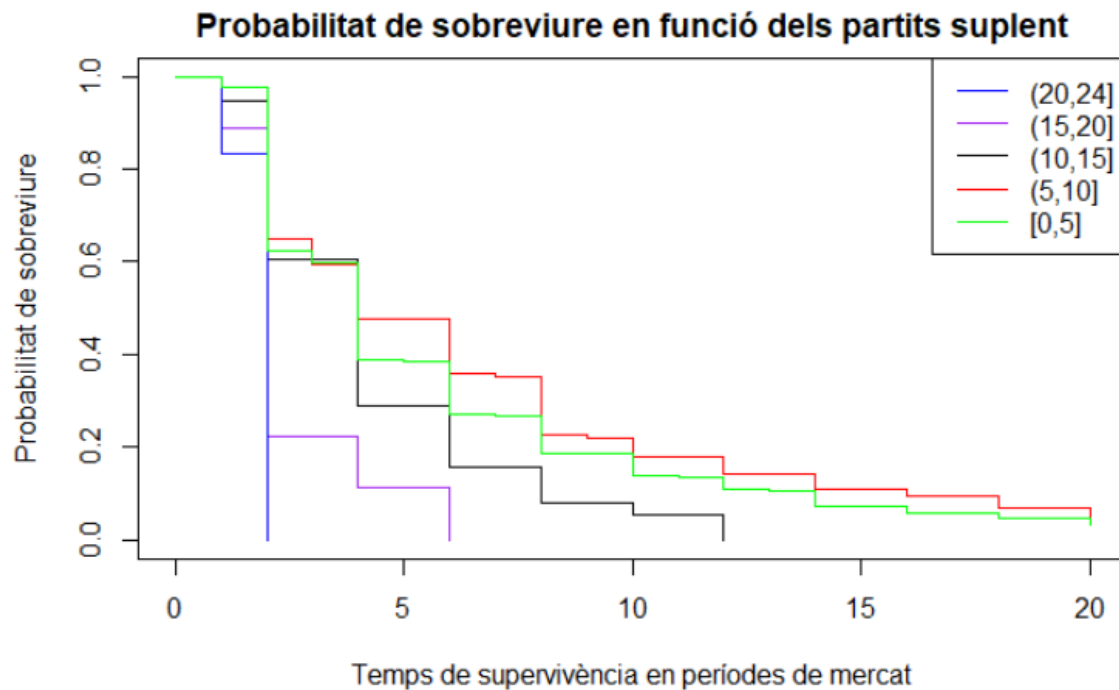
Partits titular: Una vegada més els jugadors amb menys partits, en aquest cas de titular, són els que tenen pitjor probabilitat de sobreviure. Cal destacar que, tot i seguir la mateixa tendència que en partits jugats, els jugadors que són més de 30 vegades titulars no tenen millors resultats que els que juguen entre 20 i 30 partits de titular.



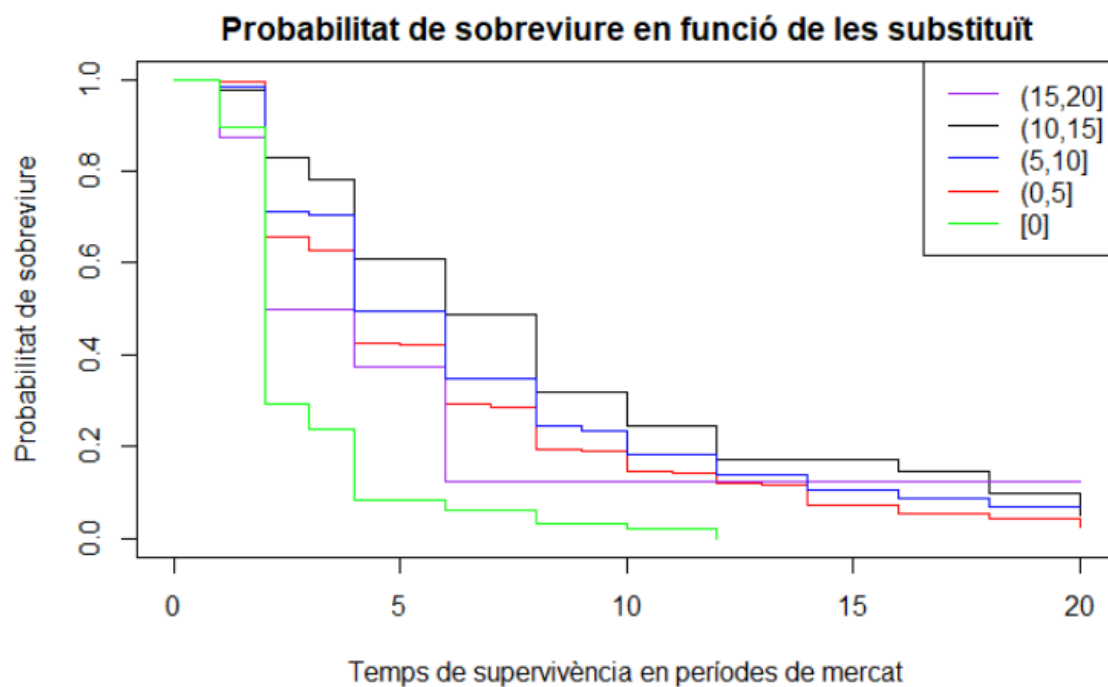
Partits complets: Com més partits complets millor probabilitat de supervivència tenen els jugadors, excepte els jugadors que superen els 30 partits que com a màxim arriben al període 10, és a dir a la meitat del temps d'estudi.



Partits suplent: Els jugadors amb més partits substituïts tenen menor probabilitat de sobreviure, destacant que els que sumen més de 20 substitucions no passen del primer any en la competició, mentre els que tenen 10 o menys substitucions de mitjana tenen major probabilitat d'arribar al final de l'estudi sense patir l'esdeveniment de fallida.

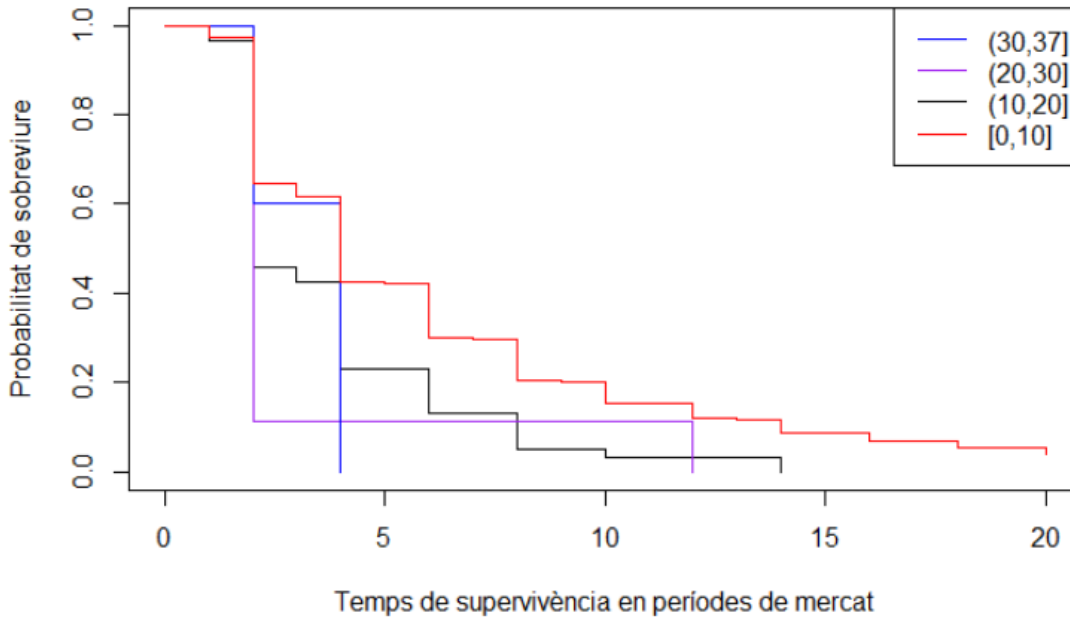


Partits substituït: Els jugadors que no han estat substituïts tenen la pitjor probabilitat de supervivència, seguits dels que n'han fet més de 15, tot i que a partir del període sis es manté la seva probabilitat de sobreviure fins al final de l'estudi, sent el grup que té més probabilitat de sobreviure en els últims dos períodes. Pel que fa als altres tres grups com més substitucions han realitzat major probabilitat de sobreviure tenen.



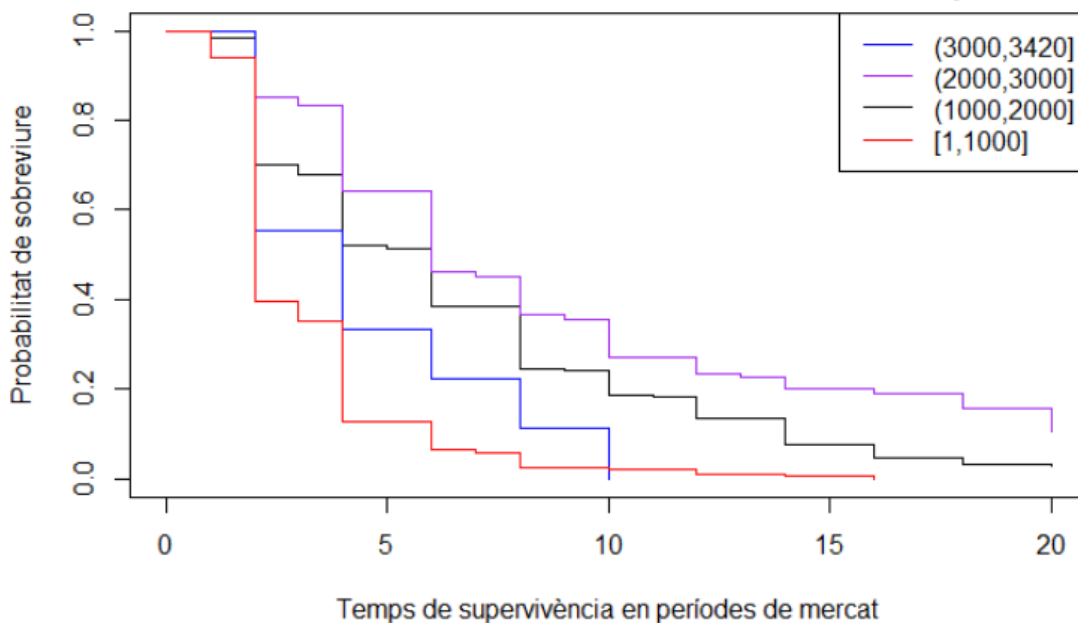
Convocatòries sense jugar: Els jugadors que tenen 10 o menys convocatòries sense jugar són els que tenen millor probabilitat de supervivència, seguits dels que en tenen més de 10 fins a 20. Els jugadors que en tenen més de 20 fins a 30 pateixen una caiguda de la probabilitat de supervivència molt gran en el període 2, però després es manté estable fins al 12, en canvi els que passen de 30 convocatòries sense jugar cau la probabilitat de sobreviure fins a zero al període 4.

Probabilitat de sobreviure en funció de les convocatòries sense jugar



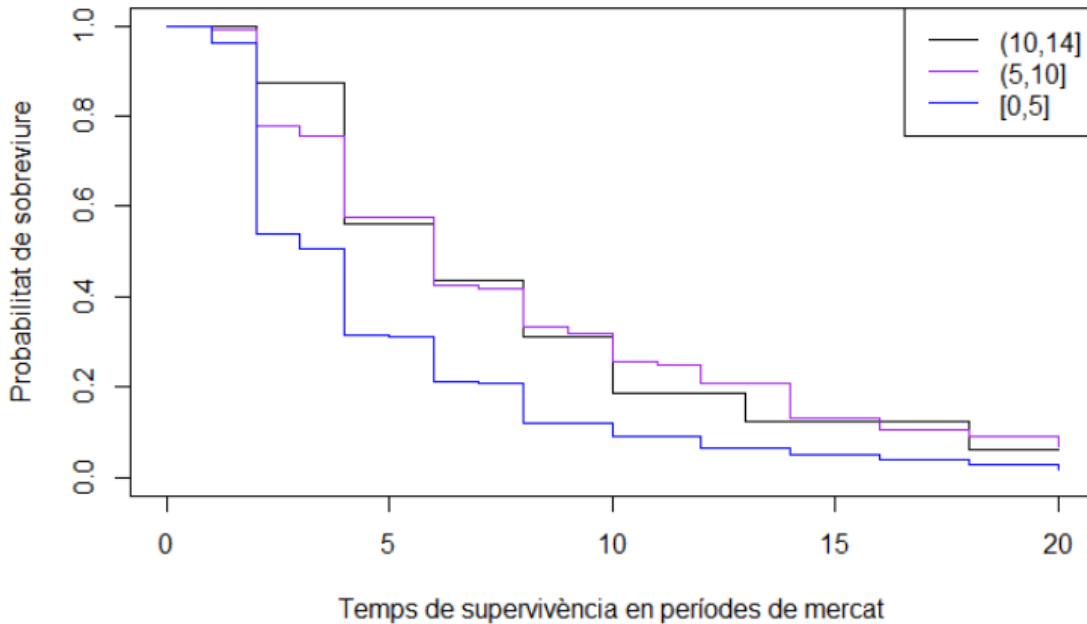
Minuts disputats: Els jugadors que han disputat més de 2000 minuts fins a 3000 tenen la millor probabilitat de supervivència, seguits dels que n'han disputat més de 1000 fins 2000. Els jugadors que menys partits han disputat tenen menys probabilitats de supervivència en la majoria dels períodes, però els primer d'arribar a pobabilitat 0 són els jugadors que han disputat més de 3000 minuts.

Probabilitat de sobreviure en funció dels minuts disputats



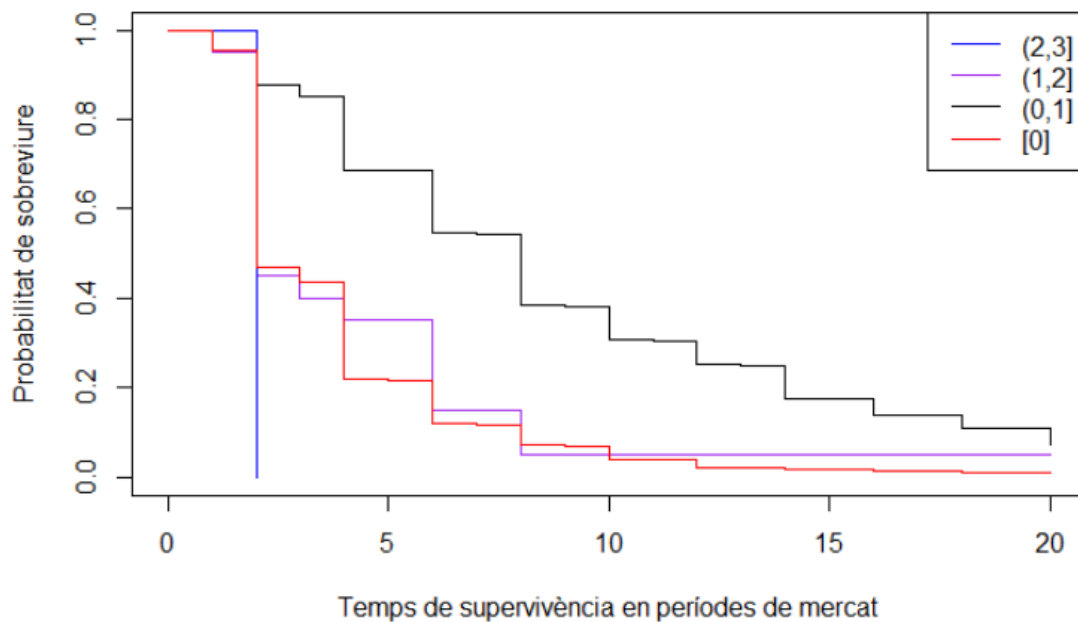
Targetes grogues: Els jugadors amb menys targetes grogues tenen pitjor probabilitat de supervivència que els altres, a partir de les 5 targetes grogues tots tenen una probabilitat de sobreviure similar.

Probabilitat de sobreviure en funció de les targetes grogues rebudes

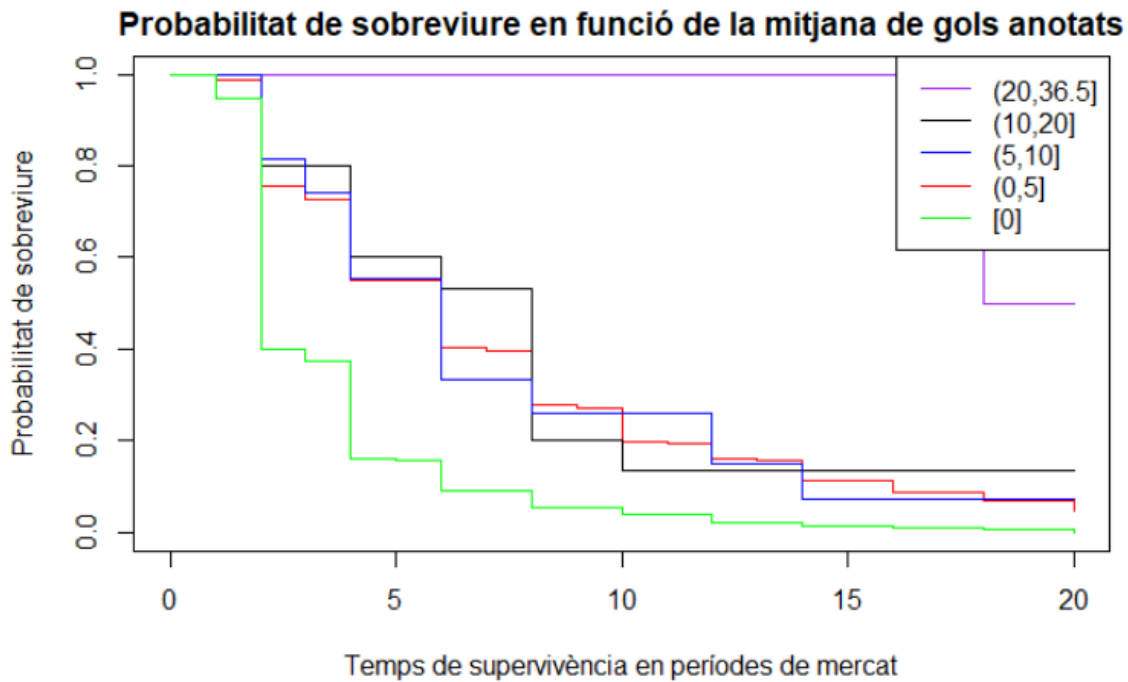


Targetes vermelles: La probabilitat de supervivència, dels jugadors amb una mitjana de targetes vermelles superior a 2, cau fins a zero en el segon període, mentre que els jugadors que tenen més de 0 i fins a una targeta de mitjana són els que tenen major probabilitat de sobreviure. Els altres dos grups de jugadors tenen probabilitats similars de supervivència.

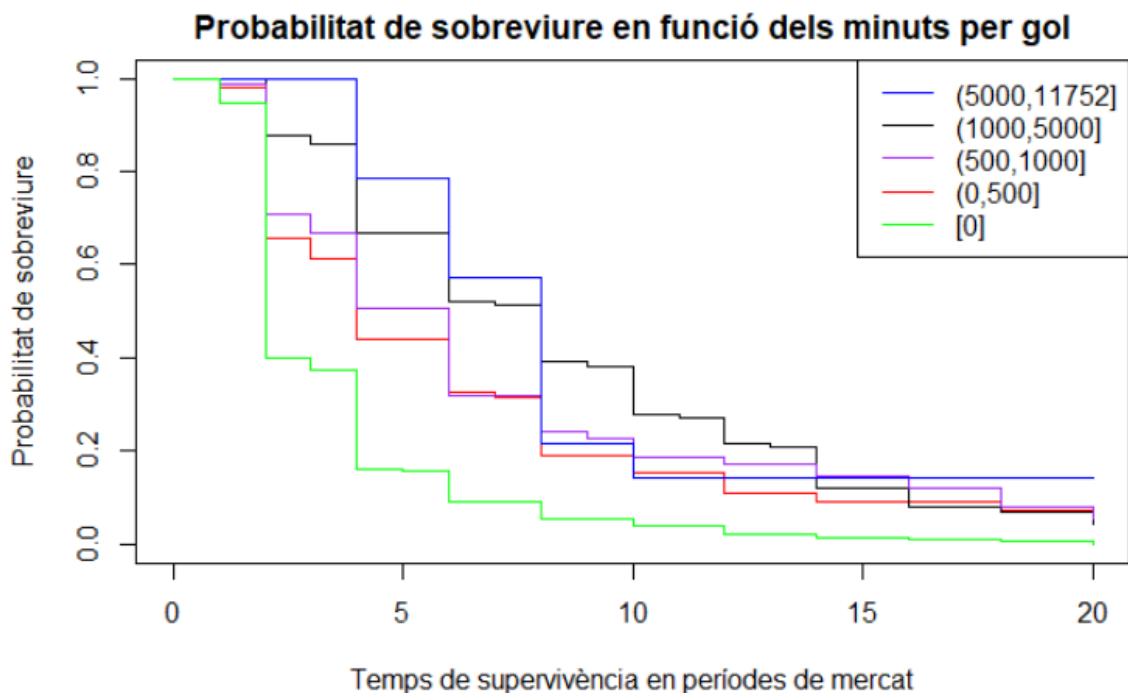
Probabilitat de sobreviure en funció de les targetes vermelles rebudes



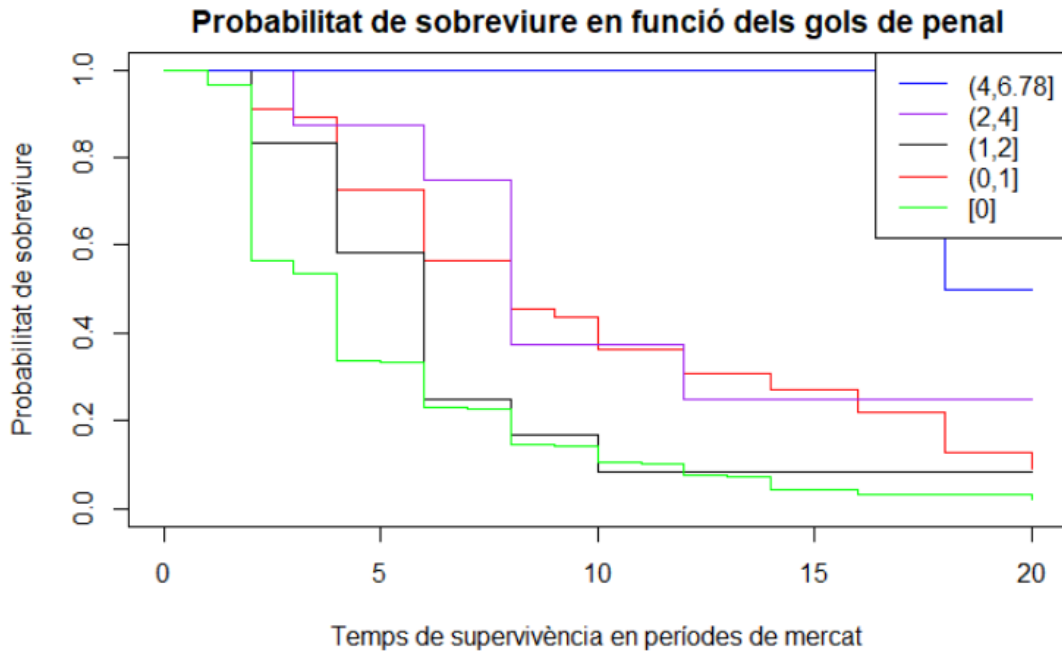
Gols: Els jugadors que no han marcat cap gol en la competició són els que tenen menor probabilitat de supervivència, mentre la resta tenen probabilitats similars al llarg del temps, excepte els jugadors que marquen més de 20 gols que tenen una probabilitat de supervivència molt superior a tota la resta.



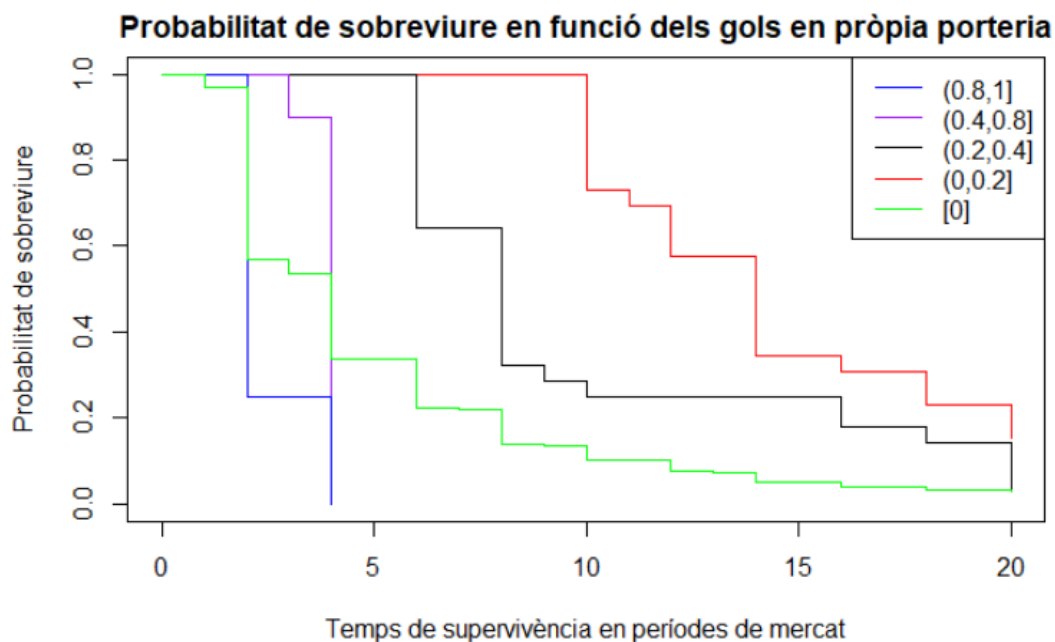
Minuts per gol: Des del principi la probabilitat de supervivència és major en aquells jugadors que més minuts tarden en marcar i disminueix a mesura que baixen els minuts per gol. Però a partir del període vuit, les probabilitats de supervivència es comencen a igualar, excepte en els jugadors que no han marcat cap gol que segueixen per sota de la resta.



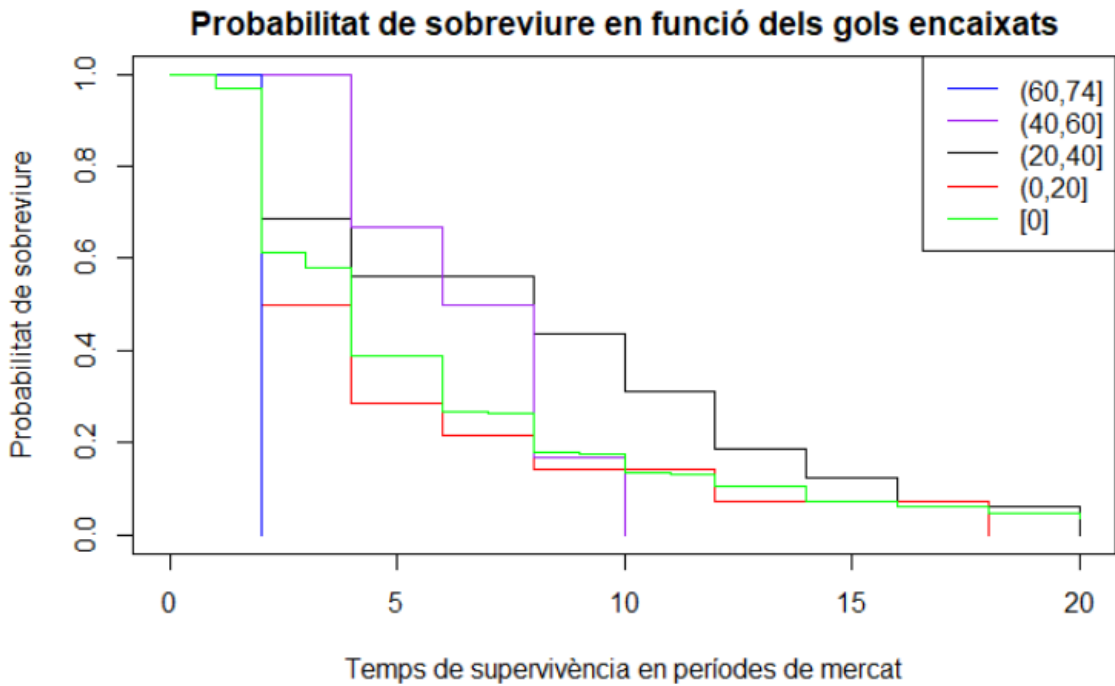
Gols de penal: Els jugadors amb més de 4 gols de penal tenen una probabilitat de supervivència molt alta en comparació a la resta, mentre els que no n'han marcat cap tenen la probabilitat més baixa, seguits dels que n'han marcat més d'un fins a 2. Les corbes de supervivència dels jugadors que han marcat més de 0 fins a 1 penal de mitjana i més de 2 fins a 4, es creuen fins a 6 vegades al llarg del temps.



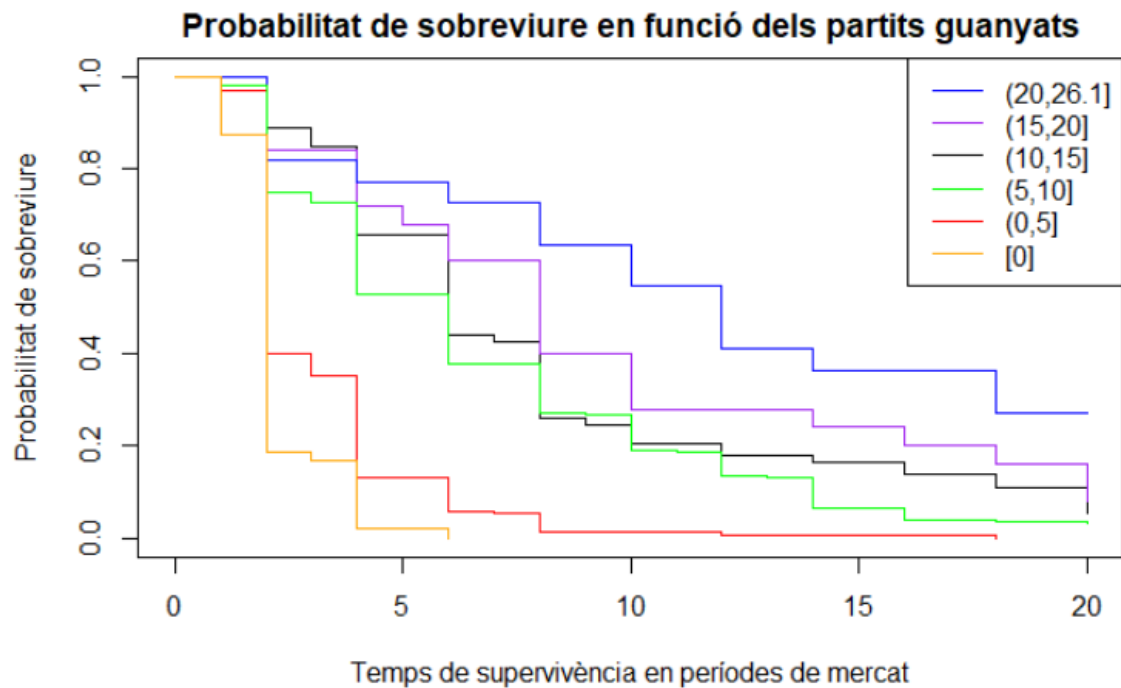
Gols en pròpia: Els jugadors amb la probabilitat de supervivència més baixa són els que han marcat més de 0.8 gols en pròpia per temporada, seguit dels que n'han marcat més de 0.4 fins 0.8, tot i que aquests tenen major probabilitat que els que no n'han marcat cap fins el període quatre. A partir del període quatre els que segueixen tenint menys probabilitats són els que no n'han marcat cap, seguit dels que n'han marcat més de 0.2 fins 0.4, quedant amb millor corba de supervivència els que n'han marcat més de 0 fins a 0.2.



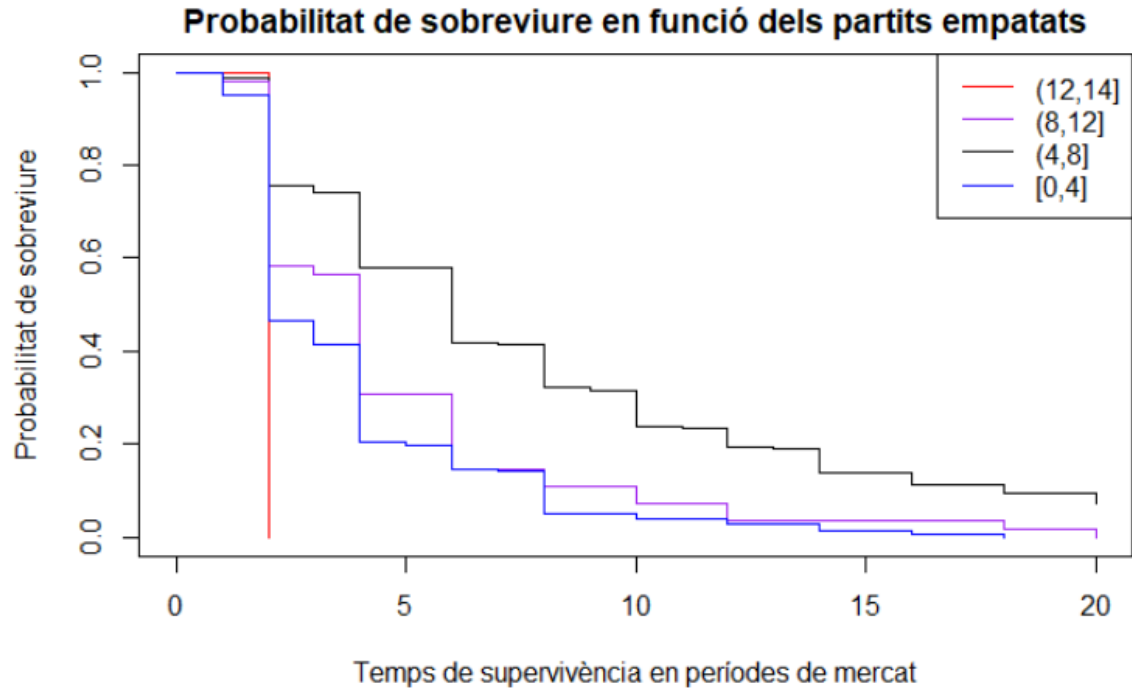
Gols encaixats: Els jugadors que han encaixat més de 60 gols són els primers en caure fins a la probabilitat 0, en el període dos, seguits dels que n'han encaixat més de 40 fins 60, però aquest són els que tenen major probabilitat de sobreviure fins el període sis. Pel que fa als altres comencen de probabilitat més baixa a més alta, els que han encaixat més de 0 fins 20 gols, cap gol encaixat i més de 20 fins 40 gols, però a partir del període vuit es comencen a igualar.



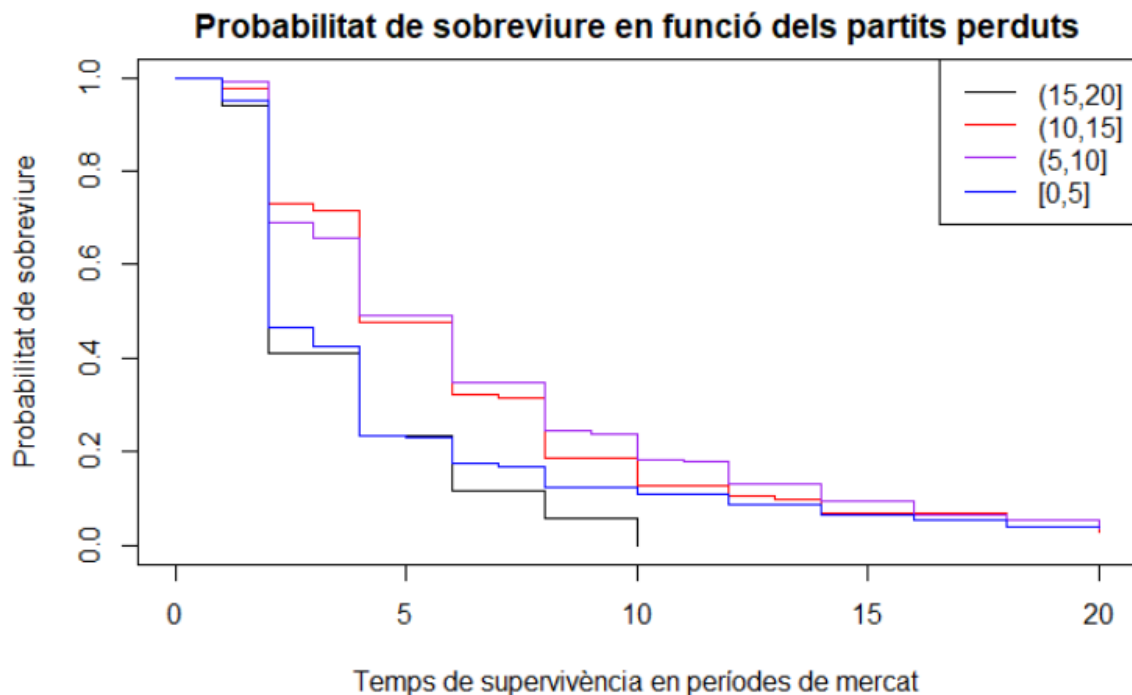
Partits guanyats: La probabilitat de supervivència dels jugadors disminueix a mesura que els jugadors han guanyat menys partits.



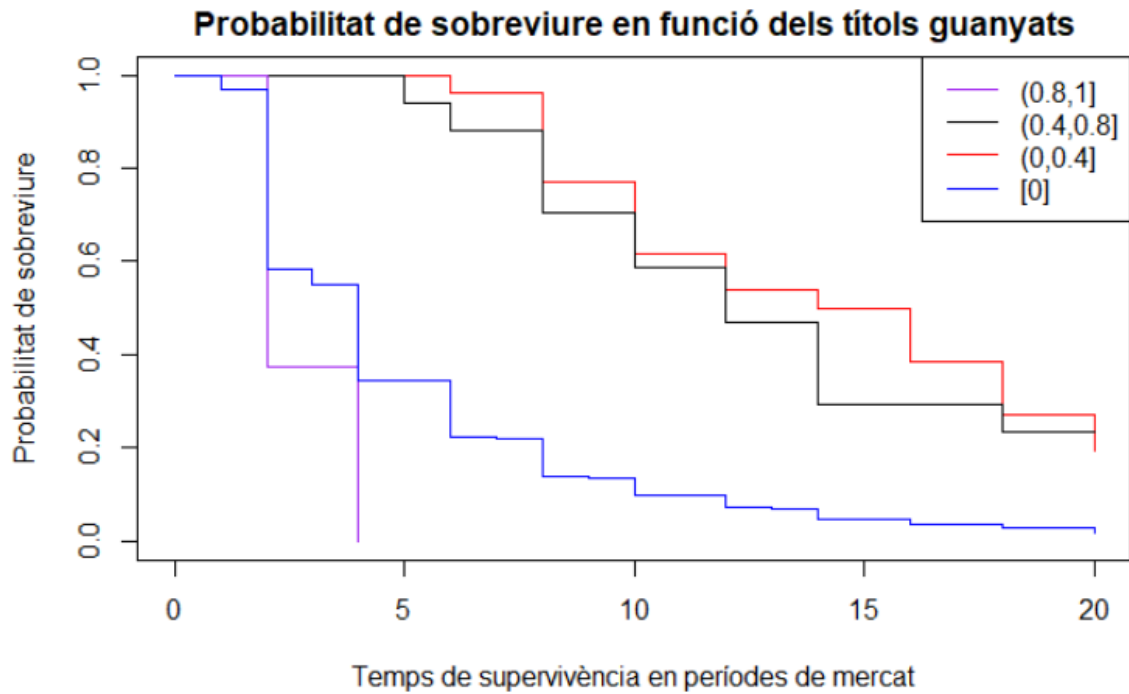
Partits empatats: La probabilitat de supervivència dels jugadors que han empatat més de 12 partits cau a 0 en el període 2, mentre que els jugadors que n'han empatat més 4 fins 8 són els que tenen una probabilitat més alta. En mig queden els jugadors que han empatat entre 0 i 4 partits i els que n'han empatat més de 8 fins 12, sent els segons qui tenen una probabilitat de supervivència major, però per poc.



Partits perduts: Els jugadors amb més de 15 partits perduts i els que en tenen un màxim de 5 són els que tenen pitjor probabilitat de sobreviure, però la probabilitat d'aquests últims, a partir del període 6, comença a apropar-se a la dels jugadors restants arribant a igualar-la.



Lligues guanyades: Els jugadors amb pitjor probabilitat de supervivència són els que tenen major percentatge de lligues guanyades, seguits dels que no n'han guanyat cap. Els jugadors que tenen un percentatge de campionats superior a 0 fins a 0.4 són els que tenen major probabilitat de sobreviure, mentre els que tenen més de 0.4 fins 0.8 estan lleugerament per sota.



III. MODELITZACIÓ

1. Hipòtesi de proporcionalitat

Per iniciar l'anàlisi amb el model de Cox, s'ha de comprovar la hipòtesi de proporcionalitat de les diferents variables que s'empraran.

```
> cox.zph(coxph(Y~PJ, data=cdata))
      rho chisq  p
PJ 0.00855 0.043 0.836
> cox.zph(coxph(Y~PT, data=cdata))
      rho chisq  p
PT 0.0595  2.3 0.129
> cox.zph(coxph(Y~PC, data=cdata))
      rho chisq  p
PC 0.134 12.6 0.000393
> cox.zph(coxph(Y~PS, data=cdata))
      rho chisq  p
PS -0.0479 1.76 0.185
> cox.zph(coxph(Y~PX, data=cdata))
      rho chisq  p
PX 0.0552 1.98 0.159
> cox.zph(coxph(Y~C, data=cdata))
      rho chisq  p
C 0.0538 1.53 0.216
> cox.zph(coxph(Y~Min, data=cdata))
      rho chisq  p
Min 0.0623 2.49 0.114
> cox.zph(coxph(Y~TG, data=cdata))
      rho chisq  p
TG 0.122 9.67 0.00187
> cox.zph(coxph(Y~TV, data=cdata))
      rho chisq  p
TV 0.0231 0.509 0.475
> cox.zph(coxph(Y~G, data=cdata))
      rho chisq  p
G 0.0945 6.98 0.00824
> cox.zph(coxph(Y~MinG, data=cdata))
      rho chisq  p
MinG 0.274 50.7 1.08e-12
> cox.zph(coxph(Y~GP, data=cdata))
      rho chisq  p
GP 0.0926 6.08 0.0137
> cox.zph(coxph(Y~GPP, data=cdata))
      rho chisq  p
GPP 0.0419 1.76 0.184
> cox.zph(coxph(Y~GE, data=cdata))
      rho chisq  p
GE 0.0458 1.19 0.275
> cox.zph(coxph(Y~PG, data=cdata))
      rho chisq  p
PG 0.143 14.5 0.000138
> cox.zph(coxph(Y~PE, data=cdata))
      rho chisq  p
PE 0.0122 0.118 0.731
> cox.zph(coxph(Y~PP, data=cdata))
      rho chisq  p
PP 0.136 12.6 0.000394
> cox.zph(coxph(Y~Win, data=cdata))
      rho chisq  p
Win 0.0217 0.383 0.536
> cox.zph(coxph(Y~Nac, data=cdata))
      rho chisq  p
NacALE -0.029962 4.65e-01 0.495
NacANG -0.029962 4.65e-01 0.495
NacARG -0.044946 1.04e+00 0.307
NacAUS -0.034581 6.20e-01 0.431
NacBOS -0.029962 4.65e-01 0.495
NacBRA -0.039969 8.28e-01 0.363
NacBUR 0.002071 2.22e-03 0.962
NacCAM -0.059622 1.85e+00 0.174
NacCAV -0.031334 5.08e-01 0.476
NacCHE -0.006625 2.27e-02 0.880
NacCHI -0.034581 6.20e-01 0.431
NacCOL -0.006030 1.88e-02 0.891
NacCRO -0.026404 3.61e-01 0.548
NacDIN -0.005739 1.71e-02 0.896
NacERI -0.005739 1.71e-02 0.896
NacESP -0.045880 1.09e+00 0.297
NacFRA -0.055675 1.61e+00 0.205
NacGHA -0.005739 1.71e-02 0.896
NacGUI -0.031060 4.99e-01 0.480
NacHOL -0.045616 1.07e+00 0.300
NacHUN -0.005739 1.71e-02 0.896
NacING -0.029962 4.65e-01 0.495
NacIRA -0.004543 1.07e-02 0.918
NacISR -0.031334 5.08e-01 0.476
NacITA 0.000525 1.43e-04 0.990
NacJAP -0.029962 4.65e-01 0.495
NacLIT -0.005739 1.71e-02 0.896
NacMAL -0.016238 1.36e-01 0.712
NacMAR -0.037831 7.40e-01 0.390
NacMEX -0.041438 8.92e-01 0.345
NacMRR -0.005739 1.71e-02 0.896
NacNIG -0.019070 1.88e-01 0.664
NacPAR -0.029962 4.65e-01 0.495
NacPER -0.005739 1.71e-02 0.896
NacPOR -0.034781 6.27e-01 0.428
NacRUM -0.034581 6.20e-01 0.431
NacSEN -0.029962 4.65e-01 0.495
NacSER -0.039530 8.07e-01 0.369
NacSUD -0.029962 4.65e-01 0.495
NacSUE -0.029962 4.65e-01 0.495
NacTUN -0.005739 1.71e-02 0.896
NacUCR -0.029962 4.65e-01 0.495
NacURU -0.039893 8.25e-01 0.364
NacVEN -0.057534 1.72e+00 0.190
GLOBAL NA 2.02e+01 0.999
> cox.zph(coxph(Y~Pos, data=cdata))
      rho chisq  p
PosD 0.0473 1.162 0.281
PosM 0.0193 0.193 0.660
PosP 0.0484 1.218 0.270
GLOBAL NA 1.902 0.593
> cox.zph(coxph(Y~Team, data=cdata))
      rho chisq  p
TeamATC -0.02867 4.27e-01 0.513580
TeamATM -0.08332 3.63e+00 0.056705
TeamDEP 0.03064 4.88e-01 0.484639
TeamESP -0.01010 5.31e-02 0.817769
TeamFCB -0.00259 3.47e-03 0.953014
TeamGET 0.01695 1.49e-01 0.699804
TeamMAL 0.00192 1.92e-03 0.965051
TeamMLL -0.01782 1.65e-01 0.684633
TeamOSA 0.00116 6.98e-04 0.978916
TeamRDS -0.07181 2.69e+00 0.100704
TeamRMA -0.04945 1.28e+00 0.258257
TeamSDG 0.03897 7.92e-01 0.373569
TeamSEV 0.04005 8.35e-01 0.360871
TeamTEN -0.04778 1.17e+00 0.278618
TeamVAL 0.04157 8.94e-01 0.344344
TeamVIL 0.05324 1.47e+00 0.224765
TeamVLL -0.05550 1.58e+00 0.209280
TeamXER -0.13298 9.14e+00 0.002503
TeamZAR -0.03421 6.09e-01 0.435215
GLOBAL NA 4.63e+01 0.000448
> cox.zph(coxph(Y~Fix, data=cdata))
      rho chisq  p
FixS -0.394 68.3 1.44e-16
```

Observant els resultats obtinguts es descarta la hipòtesi de riscos proporcionals en les variables Partits complets, Targetes grogues, Gols, Minuts per gol, Partits guanyats, Partits perduts i Equip fix. A part en els nivells Xerez i global de la variable Equip també s'observa un p-valor significatiu, però es suposa que compleix la hipòtesi de riscos proporcionals i més endavant, en l'apartat de Cox Ampliat, es comprovarà multiplicant la variable per una funció de temps. El que si que es farà, és un test de raó de verisimilitud per contrastar si els coeficients són significatius o no conjuntament.

```

Call:
coxph(formula = Y ~ Team, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
TeamATC -0.596030  0.550995  0.281828 -2.115 0.034441 *
TeamATM -0.119367  0.887482  0.274408 -0.435 0.663563
TeamDEP  0.002329  1.002332  0.278518  0.008 0.993327
TeamESP -0.239204  0.787254  0.281101 -0.851 0.394794
TeamFCB -0.852682  0.426270  0.300218 -2.840 0.004508 **
TeamGET -0.686414  0.503378  0.299891 -2.289 0.022086 *
TeamMAL -0.082650  0.920673  0.276331 -0.299 0.764864
TeamMLL -0.297674  0.742543  0.292287 -1.018 0.308474
TeamOSA -0.03631  0.966928  0.286194 -0.118 0.906455
TeamRDS -0.106993  0.898532  0.280998 -0.381 0.703382
TeamRMA -0.625161  0.535175  0.296346 -2.110 0.034896 *
TeamSDG -0.005977  0.994040  0.278937 -0.021 0.982903
TeamSEV -0.237072  0.788934  0.272612 -0.870 0.384501
TeamTEN  1.034983  2.815059  0.297902  3.474 0.000512 ***
TeamVAL -0.496999  0.608353  0.286493 -1.735 0.082782 .
TeamVIL -0.155378  0.856091  0.286491 -0.542 0.587578
TeamVLL  1.036208  2.818508  0.274722  3.772 0.000162 ***
TeamXER  0.695850  2.005414  0.293079  2.374 0.017583 *
TeamZAR  0.123058  1.130950  0.274335  0.449 0.653744
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
TeamATC    0.5510    1.8149    0.3171    0.9573
TeamATM    0.8875    1.1268    0.5183    1.5196
TeamDEP    1.0023    0.9977    0.5807    1.7302
TeamESP    0.7873    1.2702    0.4538    1.3658
TeamFCB    0.4263    2.3459    0.2367    0.7678
TeamGET    0.5034    1.9866    0.2797    0.9061
TeamMAL    0.9207    1.0862    0.5357    1.5824
TeamMLL    0.7425    1.3467    0.4187    1.3168
TeamOSA    0.9669    1.0342    0.5518    1.6943
TeamRDS    0.8985    1.1129    0.5180    1.5585
TeamRMA    0.5352    1.8685    0.2994    0.9566
TeamSDG    0.9940    1.0060    0.5754    1.7173
TeamSEV    0.7889    1.2675    0.4624    1.3461
TeamTEN    2.8151    0.3552    1.5700    5.0473
TeamVAL    0.6084    1.6438    0.3470    1.0666
TeamVIL    0.8561    1.1681    0.4883    1.5010
TeamVLL    2.8185    0.3548    1.6450    4.8291
TeamXER    2.0054    0.4987    1.1291    3.5618
TeamZAR    1.1309    0.8842    0.6606    1.9362

Concordance= 0.654 (se = 0.017 )
Likelihood ratio test= 97.85 on 19 df,  p=1e-12
Wald test              = 110 on 19 df,  p=8e-15
Score (logrank) test = 122.3 on 19 df,  p=<2e-16

```

Es pot comprovar que en el test de raó de verisimilitut el p-valor és inferior a 0.05, per tant, no es pot rebutjar que el conjunt de coeficients no sigui significatiu. Per altra banda observant els coeficients dels equips significatius tenen sentit. En els tres equips que descendeixen el primer any el coeficient és significatiu i a part augmenta el risc de fallida, mentre en equips com el FC Barcelona o l'Athletic Club, coneguts per apostar per jugadors de la seva cantera abans que per fitxar, tenen coeficients negatius, altres equips que també són significatius són el Getafe CF i el Real Madrid i coincideixen amb els anteriors sent motiu de menor risc de fallida.

2. Model de Cox

En primer lloc s'ha calculat quines variables són bones predictors pel temps de supervivència:

```
Call:
coxph(formula = Y ~ PT, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
PT -0.062875  0.939061  0.005532 -11.37 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
PT    0.9391    1.065    0.9289    0.9493

Concordance= 0.694 (se = 0.014 )
Likelihood ratio test= 133.7 on 1 df,  p=<2e-16
Wald test              = 129.2 on 1 df,  p=<2e-16
Score (logrank) test = 134.3 on 1 df,  p=<2e-16
```

```
Call:
coxph(formula = Y ~ PJ, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
PJ -0.058074  0.943580  0.005228 -11.11 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
PJ    0.9436    1.06    0.934    0.9533

Concordance= 0.67 (se = 0.014 )
Likelihood ratio test= 119.6 on 1 df,  p=<2e-16
Wald test              = 123.4 on 1 df,  p=<2e-16
Score (logrank) test = 128.5 on 1 df,  p=<2e-16
```

```
Call:
coxph(formula = Y ~ PS, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
PS  0.02030  1.02051  0.01203  1.687  0.0915 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
PS    1.021    0.9799    0.9967    1.045

Concordance= 0.486 (se = 0.017 )
Likelihood ratio test= 2.77 on 1 df,  p=0.1
Wald test              = 2.85 on 1 df,  p=0.09
Score (logrank) test = 2.85 on 1 df,  p=0.09
```

```
Call:
coxph(formula = Y ~ PX, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
PX -0.05617  0.94538  0.01242 -4.521 6.15e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
PX    0.9454    1.058    0.9226    0.9687

Concordance= 0.611 (se = 0.017 )
Likelihood ratio test= 21.98 on 1 df,  p=3e-06
Wald test              = 20.44 on 1 df,  p=6e-06
Score (logrank) test = 20.59 on 1 df,  p=6e-06
```

```
Call:
coxph(formula = Y ~ C, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
C  0.035543  1.036182  0.007981  4.453 8.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
C    1.036    0.9651    1.02    1.053

Concordance= 0.529 (se = 0.017 )
Likelihood ratio test= 16.96 on 1 df,  p=4e-05
Wald test              = 19.83 on 1 df,  p=8e-06
Score (logrank) test = 19.87 on 1 df,  p=8e-06
```

```
Call:
coxph(formula = Y ~ Min, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
Min -7.260e-04  9.993e-01  6.422e-05 -11.3 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
Min    0.9993    1.001    0.9991    0.9994

Concordance= 0.694 (se = 0.014 )
Likelihood ratio test= 132.5 on 1 df,  p=<2e-16
Wald test              = 127.8 on 1 df,  p=<2e-16
Score (logrank) test = 132.5 on 1 df,  p=<2e-16
```

```
Call:
coxph(formula = Y ~ TV, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
TV -0.1802    0.8351    0.1328 -1.356  0.175
---
exp(coef) exp(-coef) lower .95 upper .95
TV    0.8351    1.197    0.6437    1.083

Concordance= 0.611 (se = 0.014 )
Likelihood ratio test= 1.91 on 1 df,  p=0.2
Wald test              = 1.84 on 1 df,  p=0.2
Score (logrank) test = 1.84 on 1 df,  p=0.2
```

```
Call:
coxph(formula = Y ~ GP, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
GP -0.3566    0.7000    0.1032 -3.457 0.00045 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
GP    0.7    1.429    0.5719    0.8569

Concordance= 0.573 (se = 0.011 )
Likelihood ratio test= 17.75 on 1 df,  p=3e-05
Wald test              = 11.95 on 1 df,  p=5e-04
Score (logrank) test = 12.44 on 1 df,  p=4e-04
```

```
Call:
coxph(formula = Y ~ GPP, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
GPP -0.7574    0.4689    0.3948 -1.918  0.0551 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
GPP    0.4689    2.133    0.2163    1.017

Concordance= 0.569 (se = 0.01 )
Likelihood ratio test= 4.08 on 1 df,  p=0.04
Wald test              = 3.68 on 1 df,  p=0.06
Score (logrank) test = 3.68 on 1 df,  p=0.05
```

```
Call:
coxph(formula = Y ~ GE, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
GE -0.003761  0.996246  0.005533 -0.68  0.497
---
exp(coef) exp(-coef) lower .95 upper .95
GE    0.9962    1.004    0.9855    1.007

Concordance= 0.508 (se = 0.008 )
Likelihood ratio test= 0.48 on 1 df,  p=0.5
Wald test              = 0.46 on 1 df,  p=0.5
Score (logrank) test = 0.46 on 1 df,  p=0.5
```



```
Call:
coxph(formula = Y ~ PE, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
PE -0.11363  0.89259  0.01935  -5.873 4.27e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
PE    0.8926    1.12    0.8594    0.9271

Concordance= 0.599 (se = 0.016 )
Likelihood ratio test= 34.64 on 1 df,  p=4e-09
Wald test              = 34.5 on 1 df,  p=4e-09
Score (logrank) test = 34.75 on 1 df,  p=4e-09
```

```
Call:
coxph(formula = Y ~ Win, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
Win -1.4705  0.2298  0.3372  -4.36 1.3e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
Win    0.2298    4.351    0.1187    0.4451

Concordance= 0.566 (se = 0.01 )
Likelihood ratio test= 25.07 on 1 df,  p=6e-07
Wald test              = 19.01 on 1 df,  p=1e-05
Score (logrank) test = 19.99 on 1 df,  p=8e-06
```

```
Call:
coxph(formula = Y ~ Team, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
TeamATC -0.596030  0.550995  0.281828  -2.115 0.034441 *
TeamATM -0.119367  0.887482  0.274408  -0.435 0.663563
TeamDEP  0.002329  1.002332  0.278518  0.008 0.993327
TeamESP  0.239204  0.787254  0.281101  -0.851 0.394794
TeamFCB  -0.852682  0.426270  0.300218  -2.840 0.004508 **
TeamGET  -0.686414  0.503378  0.299891  -2.289 0.022086 *
TeamMAL  -0.082650  0.920673  0.276331  -0.299 0.764864
TeamMLL  -0.297674  0.742543  0.292287  -1.018 0.308474
TeamOSA  -0.033631  0.966928  0.286194  -0.118 0.906455
TeamRDS  -0.106993  0.898532  0.280998  -0.381 0.703382
TeamRMA  -0.625161  0.535175  0.296346  -2.110 0.034896 *
TeamSDG  -0.005977  0.994040  0.278937  -0.021 0.982903
TeamSEV  -0.237072  0.788934  0.272612  -0.870 0.384501
TeamTEN  1.034983  2.815059  0.297902  3.474 0.000512 ***
TeamVAL  -0.496999  0.608353  0.286493  -1.735 0.082782 .
TeamVIL  -0.155378  0.856091  0.286491  -0.542 0.587578
TeamVLL  1.036208  2.818508  0.274722  3.772 0.000162 ***
TeamXER  0.695850  2.005414  0.293079  2.374 0.017583 *
TeamZAR  0.123058  1.130950  0.274335  0.449 0.653744
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
TeamATC    0.5510    1.8149    0.3171    0.9573
TeamATM    0.8875    1.1268    0.5183    1.5196
TeamDEP    1.0023    0.9977    0.5807    1.7302
TeamESP    0.7873    1.2702    0.4538    1.3658
TeamFCB    0.4263    2.3459    0.2367    0.7678
TeamGET    0.5034    1.9866    0.2797    0.9061
TeamMAL    0.9207    1.0862    0.5357    1.5824
TeamMLL    0.7425    1.3467    0.4187    1.3168
TeamOSA    0.9669    1.0342    0.5518    1.6943
TeamRDS    0.8985    1.1129    0.5180    1.5585
TeamRMA    0.5352    1.8685    0.2994    0.9566
TeamSDG    0.9940    1.0060    0.5754    1.7173
TeamSEV    0.7889    1.2675    0.4624    1.3461
TeamTEN    2.8151    0.3552    1.5700    5.0473
TeamVAL    0.6084    1.6438    0.3470    1.0666
TeamVIL    0.8561    1.1681    0.4883    1.5010
TeamVLL    2.8185    0.3548    1.6450    4.8291
TeamXER    2.0054    0.4987    1.1291    3.5618
TeamZAR    1.1309    0.8842    0.6606    1.9362

Concordance= 0.654 (se = 0.017 )
Likelihood ratio test= 97.85 on 19 df,  p=1e-12
Wald test              = 110 on 19 df,  p=8e-15
Score (logrank) test = 122.3 on 19 df,  p=<2e-16
```

```
Call:
coxph(formula = Y ~ Pos, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
PosD -0.04271  0.95819  0.11765  -0.363 0.717
PosM -0.13125  0.87700  0.11936  -1.100 0.271
PosP -0.10296  0.90217  0.18878  -0.545 0.585

      exp(coef) exp(-coef) lower .95 upper .95
PosD    0.9582    1.044    0.7609    1.207
PosM    0.8770    1.140    0.6941    1.108
PosP    0.9022    1.108    0.6232    1.306

Concordance= 0.52 (se = 0.017 )
Likelihood ratio test= 1.41 on 3 df,  p=0.7
Wald test              = 1.41 on 3 df,  p=0.7
Score (logrank) test = 1.41 on 3 df,  p=0.7
```

Observant els resultats del test de Wald es pot veure que les variables Partits suplent, Targetes vermelles, Gols en pròpia porta, Gols encaixats, Posició i Nacionalitat no són bones predictors, mentre les que tenen p-valors més significatius són Partits jugats, Partits titular i Equip.

Es proven diferents models de Cox i es comparen utilitzant els criteris d' Akaike.

```

Call:
coxph(formula = Y ~ Team + PJ * PT, data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
TeamATC -0.6537801  0.5200761  0.2817979 -2.320 0.020339 *
TeamATM -0.0424378  0.9584501  0.2754138 -0.154 0.877541
TeamDEP -0.0074799  0.9925480  0.2787699 -0.027 0.978594
TeamESP -0.1932633  0.8242649  0.2823420 -0.685 0.493659
TeamFCB -0.9069405  0.4037576  0.3016590 -3.007 0.002643 ***
TeamGET -0.6940219  0.4995628  0.3014983 -2.302 0.021340 *
TeamMAL -0.2278303  0.7962594  0.2773148 -0.822 0.411328
TeamMLL -0.2925194  0.7463808  0.2928998 -0.999 0.317939
TeamOSA -0.0285052  0.9718972  0.2863763 -0.100 0.920711
TeamRDS -0.1258521  0.8817452  0.2816450 -0.447 0.654986
TeamRMA -0.5058452  0.6029957  0.2970854 -1.703 0.088625 .
TeamSDG -0.2191559  0.8031965  0.2805531 -0.781 0.434710
TeamSEV -0.2191617  0.8031919  0.2745384 -0.798 0.424701
TeamTEN  1.1117307  3.0396144  0.2996838  3.710 0.000208 ***
TeamVAL -0.4251863  0.6536480  0.2887601 -1.472 0.140898
TeamVIL -0.0892687  0.9145998  0.2876384 -0.310 0.756295
TeamVLL  0.9803665  2.6654331  0.2761674  3.550 0.000385 ***
TeamXER  0.9250831  2.5220779  0.2946422  3.140 0.001691 **
TeamZAR  0.1227709  1.1306253  0.2748501  0.447 0.655104
PJ       -0.0337227  0.9668396  0.0121115 -2.784 0.005364 ***
PT       -0.1141075  0.8921620  0.0213722 -5.339 9.34e-08 ***
PJ:PT    0.0025123  1.0025154  0.0006039  4.160 3.18e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
TeamATC  0.5201  1.9228  0.2994  0.9035
TeamATM  0.9585  1.0434  0.5586  1.6444
TeamDEP  0.9925  1.0075  0.5747  1.7141
TeamESP  0.8243  1.2132  0.4740  1.4335
TeamFCB  0.4038  2.4767  0.2235  0.7293
TeamGET  0.4996  2.0018  0.2767  0.9020
TeamMAL  0.7963  1.2559  0.4624  1.3712
TeamMLL  0.7464  1.3398  0.4204  1.3252
TeamOSA  0.9719  1.0289  0.5544  1.7037
TeamRDS  0.8817  1.1341  0.5077  1.5314
TeamRMA  0.6030  1.6584  0.3368  1.0794
TeamSDG  0.8032  1.2450  0.4635  1.3920
TeamSEV  0.8032  1.2450  0.4690  1.3756
TeamTEN  3.0396  0.3290  1.6894  5.4690
TeamVAL  0.6536  1.5299  0.3712  1.1512
TeamVIL  0.9146  1.0934  0.5205  1.6072
TeamVLL  2.6654  0.3752  1.5513  4.5798
TeamXER  2.5221  0.3965  1.4157  4.4932
TeamZAR  1.1306  0.8845  0.6597  1.9376
PJ       0.9668  1.0343  0.9442  0.9901
PT       0.8922  1.1209  0.8556  0.9303
PJ:PT    1.0025  0.9975  1.0013  1.0037

Concordance= 0.769 (se = 0.012 )
Likelihood ratio test= 253.8 on 22 df, p=<2e-16
Wald test              = 274 on 22 df, p=<2e-16
Score (logrank) test = 310.4 on 22 df, p=<2e-16

```

El model amb menor AIC és $Y \sim \text{Team} + \text{PJ} \times \text{PT}$ amb un valor de 5402.714, per tant segons el criteri d'Akaike preferim aquest per davant dels altres.

3. Model ampliat per les variables que no compleixen la hipòtesi de riscos proporcionals

Per escollir les variables que s'incorporaran al model estratificat i després en l'ampliat s'utilitza un Cox ampliat, amb cada variable i la seva interacció amb el temps.

```
Call:
coxph(formula = Y ~ PC + tPC, data = maPC)

      coef exp(coef) se(coef)      z      p
PC -0.080252  0.922883  0.009905 -8.102 5.39e-16
tPC  0.005139  1.005152  0.001525  3.370 0.000752

Likelihood ratio test=101.6 on 2 df, p=< 2.2e-16
n= 2971, number of events= 519
```

```
Call:
coxph(formula = Y ~ G + tG, data = maG)

      coef exp(coef) se(coef)      z      p
G -0.110997  0.894941  0.028391 -3.910 9.24e-05
tG  0.004864  1.004875  0.002680  1.815  0.0695

Likelihood ratio test=30.77 on 2 df, p=2.08e-07
n= 2971, number of events= 519
```

```
Call:
coxph(formula = Y ~ TG + tTG, data = maTG)

      coef exp(coef) se(coef)      z      p
TG -0.197588  0.820708  0.028811 -6.858 6.98e-12
tTG  0.010372  1.010426  0.004151  2.498  0.0125

Likelihood ratio test=76.84 on 2 df, p=< 2.2e-16
n= 2971, number of events= 519
```

```
Call:
coxph(formula = Y ~ MInG + tMInG, data = maMInG)

      coef exp(coef) se(coef)      z      p
MInG -4.716e-04  9.995e-01  6.807e-05 -6.928 4.27e-12
tMInG  2.853e-05  1.000e+00  5.639e-06  5.059 4.22e-07

Likelihood ratio test=73.03 on 2 df, p=< 2.2e-16
n= 2971, number of events= 519
```

```
Call:
coxph(formula = Y ~ PG + tPG, data = maPG)

      coef exp(coef) se(coef)      z      p
PG -0.172060  0.841928  0.017115 -10.053 < 2e-16
tPG  0.007908  1.007940  0.002210  3.578 0.000346

Likelihood ratio test=179.8 on 2 df, p=< 2.2e-16
n= 2971, number of events= 519
```

```
Call:
coxph(formula = Y ~ PP + tPP, data = maPP)

      coef exp(coef) se(coef)      z      p
PP -0.080349  0.922794  0.018456 -4.354 1.34e-05
tPP  0.011188  1.011251  0.003274  3.417 0.000633

Likelihood ratio test=19.46 on 2 df, p=5.939e-05
n= 2971, number of events= 519
```

```
Call:
coxph(formula = Y ~ nFix + tFix, data = maFix)

      coef exp(coef) se(coef)      z      p
nFix -2.64163  0.07124  0.25655 -10.297 < 2e-16
tFix  0.23576  1.26587  0.03117  7.563 3.95e-14

Likelihood ratio test=179.6 on 2 df, p=< 2.2e-16
n= 2971, number of events= 519
```

Un cop plantejats els 7 models ampliat es descarta la variable Gols, perquè la seva interacció amb la funció temps no és significativa (p-valor de tG > 0.05), per tant, compleix amb la hipòtesi de riscos proporcionals.

4. Model de Cox estratificat

```

Call:
coxph(formula = Y ~ Team + PJ * PT + strata(Fix), data = cdata)

n= 537, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
TeamATC -0.8820613  0.4139288  0.2896579 -3.045 0.002325 **
TeamATM  0.1669677  1.1817161  0.2772474  0.602 0.547019
TeamDEP  0.0403948  1.0412217  0.2790413  0.145 0.884898
TeamESP -0.0337586  0.9668049  0.2856969 -0.118 0.905939
TeamFCB -1.1068553  0.3305970  0.3082483 -3.591 0.000330 ***
TeamGET -0.4441695  0.6413567  0.3048143 -1.457 0.145066
TeamMAL -0.3555806  0.7007665  0.2811340 -1.265 0.205940
TeamMLL -0.2170942  0.8048541  0.2938425 -0.739 0.460021
TeamOSA -0.0653825  0.9367091  0.2872207 -0.228 0.819927
TeamRDS -0.0389913  0.9617590  0.2823975 -0.138 0.890183
TeamRMA -0.7930278  0.4524727  0.3074042 -2.580 0.009887 **
TeamSDG -0.3669179  0.6928665  0.2817802 -1.302 0.192868
TeamSEV -0.0784598  0.9245392  0.2752928 -0.285 0.775640
TeamTEN  1.3325627  3.7907455  0.3025394  4.405 1.06e-05 ***
TeamVAL -0.1611382  0.8511744  0.2909945 -0.554 0.579750
TeamVIL  0.1220264  1.1297839  0.2884310  0.423 0.672245
TeamVLL  1.1092639  3.0321255  0.2794530  3.969 7.21e-05 ***
TeamXER  1.0368331  2.8202714  0.2969520  3.492 0.000480 ***
TeamZAR  0.2300281  1.2586353  0.2766808  0.831 0.405757
PJ       -0.0186194  0.9815528  0.0114158 -1.631 0.102884
PT       -0.0753804  0.9273907  0.0207186 -3.638 0.000274 ***
PJ:PT    0.0012173  1.0012181  0.0006039  2.016 0.043813 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
TeamATC    0.4139    2.4159    0.2346    0.7303
TeamATM    1.1817    0.8462    0.6863    2.0347
TeamDEP    1.0412    0.9604    0.6026    1.7991
TeamESP    0.9668    1.0343    0.5523    1.6925
TeamFCB    0.3306    3.0248    0.1807    0.6049
TeamGET    0.6414    1.5592    0.3529    1.1656
TeamMAL    0.7008    1.4270    0.4039    1.2158
TeamMLL    0.8049    1.2425    0.4525    1.4316
TeamOSA    0.9367    1.0676    0.5335    1.6447
TeamRDS    0.9618    1.0398    0.5530    1.6728
TeamRMA    0.4525    2.2101    0.2477    0.8265
TeamSDG    0.6929    1.4433    0.3988    1.2037
TeamSEV    0.9245    1.0816    0.5390    1.5858
TeamTEN    3.7907    0.2638    2.0951    6.8588
TeamVAL    0.8512    1.1748    0.4812    1.5056
TeamVIL    1.1298    0.8851    0.6419    1.9884
TeamVLL    3.0321    0.3298    1.7534    5.2435
TeamXER    2.8203    0.3546    1.5759    5.0473
TeamZAR    1.2586    0.7945    0.7318    2.1648
PJ         0.9816    1.0188    0.9598    1.0038
PT         0.9274    1.0783    0.8905    0.9658
PJ:PT     1.0012    0.9988    1.0000    1.0024

Concordance= 0.785 (se = 0.014 )
Likelihood ratio test= 233.7 on 22 df, p=<2e-16
Wald test              = 237.6 on 22 df, p=<2e-16
Score (logrank) test = 263.4 on 22 df, p=<2e-16

```

En aquesta aplicació es pot veure que hi ha moltes variables que no compleixen amb la hipòtesis de riscos proporcionals, per això seria exagerat tenir tants estrats i finalment es decideix no contemplar el model estratificat com un model final. Tot i així, destaca el model $Y \sim \text{Team} + \text{PJ} \times \text{PT} + \text{strata}(\text{Fix})$ donant un major poder predictiu segons els tres tests globals, a part indica que els predictors són significatius conjuntament.

5. Model de Cox ampliat

```

Call:
coxph(formula = Y ~ Team + PJ * PT + tFix + tPC + tTG + tMinG +
      tPG + tPP, data = model)

n= 2971, number of events= 519

      coef exp(coef) se(coef)      z Pr(>|z|)
TeamATC -1.228e+00  2.929e-01  3.088e-01 -3.976  7.01e-05 ***
TeamATM  2.663e-01  1.305e+00  2.826e-01  0.942  0.346038
TeamDEP  3.871e-02  1.039e+00  2.800e-01  0.138  0.890050
TeamESP -5.485e-02  9.466e-01  2.883e-01 -0.190  0.849133
TeamFCB -4.716e-01  6.240e-01  3.410e-01 -1.383  0.166654
TeamGET -6.540e-01  5.200e-01  3.055e-01 -2.141  0.032303 *
TeamMAL -5.173e-01  5.962e-01  2.866e-01 -1.805  0.071080 .
TeamMALL -1.905e-01  8.265e-01  2.962e-01 -0.643  0.520175
TeamOSA -1.111e-03  9.989e-01  2.887e-01 -0.004  0.996929
TeamRDS -1.264e-01  8.812e-01  2.831e-01 -0.447  0.655070
TeamRMA -6.161e-02  9.403e-01  3.353e-01 -0.184  0.854232
TeamSDG -3.281e-01  7.203e-01  2.834e-01 -1.158  0.246870
TeamSEV -9.565e-02  9.088e-01  2.766e-01 -0.346  0.729530
TeamTEN  1.126e+00  3.083e+00  3.014e-01  3.736  0.000187 ***
TeamVAL -2.377e-02  9.765e-01  2.941e-01 -0.081  0.935586
TeamVIL  9.222e-02  1.097e+00  2.894e-01  0.319  0.750037
TeamVLL  1.139e+00  3.124e+00  2.804e-01  4.062  4.86e-05 ***
TeamXER  1.292e+00  3.640e+00  2.991e-01  4.320  1.56e-05 ***
TeamZAR  1.979e-01  1.219e+00  2.792e-01  0.709  0.478287
PJ      -1.407e-02  9.860e-01  1.472e-02 -0.956  0.339020
PT      -1.325e-01  8.759e-01  2.359e-02 -5.617  1.94e-08 ***
tFix    -1.254e-01  8.821e-01  1.680e-02 -7.464  8.40e-14 ***
tPC     4.926e-03  1.005e+00  1.474e-03  3.341  0.000833 ***
tTG     -3.737e-03  9.963e-01  3.379e-03 -1.106  0.268707
tMinG   -1.763e-05  1.000e+00  4.274e-06 -4.124  3.72e-05 ***
tPG     -1.182e-02  9.883e-01  2.801e-03 -4.220  2.44e-05 ***
tPP     6.791e-03  1.007e+00  4.229e-03  1.606  0.108314
PJ:PT   2.583e-03  1.003e+00  6.424e-04  4.021  5.79e-05 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
TeamATC  0.2929  3.4144  0.1599  0.5365
TeamATM  1.3051  0.7662  0.7501  2.2707
TeamDEP  1.0395  0.9620  0.6004  1.7996
TeamESP  0.9466  1.0564  0.5379  1.6658
TeamFCB  0.6240  1.6026  0.3198  1.2174
TeamGET  0.5200  1.9231  0.2857  0.9463
TeamMAL  0.5962  1.6774  0.3400  1.0454
TeamMALL 0.8265  1.2099  0.4625  1.4772
TeamOSA  0.9989  1.0011  0.5672  1.7590
TeamRDS  0.8812  1.1348  0.5060  1.5347
TeamRMA  0.9403  1.0635  0.4873  1.8142
TeamSDG  0.7203  1.3884  0.4133  1.2552
TeamSEV  0.9088  1.1004  0.5284  1.5629
TeamTEN  3.0834  0.3243  1.7080  5.5664
TeamVAL  0.9765  1.0241  0.5487  1.7380
TeamVIL  1.0966  0.9119  0.6218  1.9339
TeamVLL  3.1241  0.3201  1.8031  5.4129
TeamXER  3.6405  0.2747  2.0255  6.5430
TeamZAR  1.2189  0.8204  0.7052  2.1067
PJ       0.9860  1.0142  0.9580  1.0149
PT       0.8759  1.1417  0.8363  0.9173
tFix     0.8821  1.1336  0.8536  0.9117
tPC      1.0049  0.9951  1.0020  1.0078
tTG      0.9963  1.0037  0.9897  1.0029
tMinG    1.0000  1.0000  1.0000  1.0000
tPG      0.9883  1.0119  0.9828  0.9937
tPP      1.0068  0.9932  0.9985  1.0152
PJ:PT    1.0026  0.9974  1.0013  1.0039

Concordance= 0.816 (se = 0.013 )
Likelihood ratio test= 350.4 on 28 df,  p=<2e-16
Wald test              = 362.9 on 28 df,  p=<2e-16
Score (logrank) test = 410.4 on 28 df,  p=<2e-16

```

Després de veure els p-valors de les variables del nou model, se'n descarten dues perquè són superiors a 0.05, la interacció de Targetes grogues amb la funció de temps i la interacció de Partits perduts amb la funció de temps. És cert que Partits jugats tampoc té un p-valor significatiu, però la seva interacció amb Partits titular sí, per tant, es mantindrà.

Finalment, el model ampliat resultant és:

```
Call:
coxph(formula = Y ~ Team + PJ * PT + tFix + tPC + tMirG + tPG,
      data = model)

n= 2971, number of events= 519

              coef exp(coef)    se(coef)      z Pr(>|z|)
TeamATC -1.151e+00  3.162e-01  3.000e-01 -3.837 0.000124 ***
TeamATM  2.252e-01  1.253e+00  2.814e-01  0.800 0.423618
TeamDEP  5.839e-02  1.060e+00  2.793e-01  0.209 0.834389
TeamESP -3.790e-02  9.678e-01  2.868e-01 -0.132 0.894864
TeamFCB -6.413e-01  5.266e-01  3.196e-01 -2.006 0.044827 *
TeamGET -6.552e-01  5.194e-01  3.055e-01 -2.144 0.032002 *
TeamMAL -5.350e-01  5.857e-01  2.844e-01 -1.881 0.059923 .
TeamMLL -1.800e-01  8.353e-01  2.949e-01 -0.610 0.541694
TeamOSA  5.173e-03  1.005e+00  2.880e-01  0.018 0.985670
TeamRDS -1.265e-01  8.812e-01  2.829e-01 -0.447 0.654810
TeamRMA -1.421e-01  8.675e-01  3.325e-01 -0.428 0.668991
TeamSDG -3.442e-01  7.088e-01  2.822e-01 -1.220 0.222625
TeamSEV -1.191e-01  8.877e-01  2.762e-01 -0.431 0.666344
TeamTEN  1.126e+00  3.085e+00  3.011e-01  3.741 0.000183 ***
TeamVAL -7.174e-02  9.308e-01  2.930e-01 -0.245 0.806610
TeamVIL  6.406e-02  1.066e+00  2.889e-01  0.222 0.824496
TeamVLL  1.125e+00  3.081e+00  2.791e-01  4.032 5.53e-05 ***
TeamXER  1.287e+00  3.621e+00  2.986e-01  4.309 1.64e-05 ***
TeamZAR  1.831e-01  1.201e+00  2.772e-01  0.660 0.508985
PJ       -3.394e-03  9.966e-01  1.272e-02 -0.267 0.789548
PT       -1.434e-01  8.664e-01  2.274e-02 -6.306 2.86e-10 ***
tFix     -1.198e-01  8.871e-01  1.584e-02 -7.561 4.00e-14 ***
tPC       5.252e-03  1.005e+00  1.449e-03  3.624 0.000291 ***
tMirG    -1.913e-05  1.000e+00  4.050e-06 -4.724 2.32e-06 ***
tPG      -1.448e-02  9.856e-01  2.316e-03 -6.255 3.98e-10 ***
PJ:PT    2.850e-03  1.003e+00  6.212e-04  4.588 4.47e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) Lower .95 upper .95
TeamATC    0.3162      3.1622    0.1756    0.5694
TeamATM    1.2525      0.7984    0.7215    2.1743
TeamDEP    1.0601      0.9433    0.6132    1.8327
TeamESP    0.9628      1.0386    0.5489    1.6890
TeamFCB    0.5266      1.8989    0.2815    0.9853
TeamGET    0.5194      1.9255    0.2854    0.9452
TeamMAL    0.5857      1.7074    0.3354    1.0226
TeamMLL    0.8353      1.1972    0.4686    1.4890
TeamOSA    1.0052      0.9948    0.5716    1.7677
TeamRDS    0.8812      1.1348    0.5061    1.5342
TeamRMA    0.8675      1.1527    0.4522    1.6644
TeamSDG    0.7088      1.4109    0.4076    1.2324
TeamSEV    0.8877      1.1265    0.5166    1.5255
TeamTEN    3.0847      0.3242    1.7096    5.5659
TeamVAL    0.9308      1.0744    0.5241    1.6531
TeamVIL    1.0662      0.9379    0.6053    1.8780
TeamVLL    3.0810      0.3246    1.7830    5.3240
TeamXER    3.6211      0.2762    2.0167    6.5019
TeamZAR    1.2009      0.8327    0.6975    2.0678
PJ          0.9966      1.0034    0.9721    1.0218
PT          0.8664      1.1542    0.8286    0.9059
tFix       0.8871      1.1273    0.8500    0.9151
tPC        1.0053      0.9948    1.0024    1.0081
tMirG      1.0000      1.0000    1.0000    1.0000
tPG        0.9856      1.0146    0.9812    0.9901
PJ:PT      1.0029      0.9972    1.0016    1.0041

Concordance= 0.816 (se = 0.013 )
Likelihood ratio test= 347.2 on 26 df, p=<2e-16
Wald test = 365.7 on 26 df, p=<2e-16
Score (logrank) test = 409.1 on 26 df, p=<2e-16
```

Un cop tenim el model analitzem quin efecte tenen les variables sobre el risc de fallida:

Per la variable Equip 6 dels 20 nivells són significatius, sent l'Athletic Club, el FC Barcelona i el Getafe CF els equips que redueixen el risc de fallida dels jugadors 3.1622, 1.8989 i 1.9255 vegades respectivament, en comparació amb els jugadors que formen part del UD Almería, nivell inclòs en l'intercept. A part, els jugadors que formen part del CD Tenerife, Valladolid CF i Xerez CD tenen un risc de fallida superior 3.0847, 3.0810 i 3.6211 vegades respectivament en comparació amb els jugadors del UD Almería.

Per calcular l'efecte de la interacció entre les variables Partits jugats i Partits titular es descarta l'ús del coeficient de Partits jugats, perquè no és significatiu, mentre si que ho són el de Partits titular i el conjunt de les dues variables. El valor en que augmenta o disminueix el rati de risc surt del càlcul $\exp(-1.434e-01 \times PT + 2.850e-03 \times PT \times PJ)$, per això es fixarà la variable Partits jugats en 38 que és el màxim de partits que es poden disputar i s'observa el rati de risc per Partits titular. El coeficient resultant és $\exp(-$

0.0351), per tant per cada Partit titular disputat amb PJ=38 disminueix el risc de fallida 1.0357 vegades. Cal destacar, que $-1.434e-01 + 2.850e-03 \times PJ$ sempre serà negatiu, donat que per ser positiu PJ hauria de ser 50.3158 i és més de 38, per tant es pot concloure que sigui quin sigui el nombre de partits jugats el coeficient conjunt sempre serà menor que 1.

L'exponencial del coeficient de la interacció de Minuts per gol amb la funció de temps és 1, per tant no afecta al risc de fallida.

Per la interacció de Equip fix amb la funció de temps es pot veure que els jugadors que es mantenen en el seu equip no els afecta el pas del temps, mentre els que canvien d'equip disminueix el risc de fallida 1.1273 quan $t=1$, però per altres temps aquesta disminució varia, per exemple per $t=10$ $\exp(-(-1.198e-01 \times 10))=3.3135$ o per $t=20$ $\exp(-(-1.198e-01 \times 20))=10.9792$.

Per la interacció de Partits complets amb la funció de temps, el risc de fallida augmenta 1.0053 vegades per cada partit complet quan $t=1$, mentre en altres temps com $t=10$ o $t=20$ augmenta 1.0539 i 1.1108 respectivament per cada partit complet.

Finalment, la interacció de Partits guanyats amb la funció de temps, disminueix en 1.0146 vegades el risc de fallida dels jugadors quan $t=1$ per cada partit guanyat, per altres temps com $t=10$ o $t=20$ disminueix per cada partit guanyat 1.1558 o 1.3359 respectivament.

CONCLUSIONS

El Model escollit és $Y \sim \text{Team} + \text{PJ} \times \text{PT} + \text{tFix} + \text{tPC} + \text{tMinG} + \text{tPG}$. Després de veure com afecten els coeficients del model es pot afirmar que les condicions que disminueixen més el risc de fallida són: formar part de l'Athletic Club, el FC Barcelona o el Getafe CF, jugar molts partits de titular, però amb el mínim de partits jugats possible, canviar d'equip i guanyar partits al llarg del temps i finalment jugar el mínim de partits complets al llarg del temps.

Veure que formar part del FC Barcelona o de l'Athletic Club ajuda a disminuir el risc de fallida era esperable tenint en compte la seva política de fitxatges i les canteres. Per altra banda estar en un dels equips que descendeix el primer any, òbviament l'augmenta severament. Destacar que el fet de formar part del Getafe CF disminueixi el risc de fallida ha estat un descobriment

També s'esperava que canviar d'equip fos quelcom positiu, degut a que al final de l'estudi la meitat dels equips havien descendit a segona divisió junt amb la majoria dels seus jugadors. De la mateixa manera que té sentit que a més partits guanyats i disputats de titular també disminueix aquest risc de fallida.

No s'esperava que completar partits fos un motiu per augmentar el risc de fallida, tot i que li donaria sentit el fet que els equips importants d'Europa s'emporten els jugadors més bons dels equips més petits o amb menys aspiracions i que, per tant, juguen més.

Finalment, destacar que a diferència de com s'esperava els jugadors espanyols no tenen un temps de supervivència major que els jugadors estrangers.

BIBLIOGRAFIA

Eva Boj del Val (2017). *El modelo de regresión de Cox*.

David G. Kleinbaum i Mitchel Klein (1996). *Survival Analysis*.

ANNEXOS

1. Competències vinculades al TFG del Grau en Estadística

Estadística descriptiva

Anàlisi de supervivència

Programació en R

2. Codis de R

```
# Importar dades
```

```
data <- read.csv2("treball.csv", header = TRUE, sep=';')
```

```
IND <- c(1:537)
```

```
data <- data.frame(data, IND)
```

```
# Esdeveniment
```

```
table(data$Status)
```

```
# Anàlisi descriptiu
```

```
## Dades qualitatives
```

```
table(data$Pos)
```

```
table(data$Team)
```

```
table(data$Fix)
```

```
table(data$Nac)
```

```
## Dades quantitatives
```

```
summary(data$PJ)
```

```
plot(data$IND, data$PJ, ylab="Mitjana de partits jugats per temporada",  
xlab="Jugadors", main="Mitjana de partits jugats de cada jugador per temporada")
```

```
summary(data$PT)
```

```
plot(data$IND, data$PT, ylab="Mitjana de partits titular per temporada",  
xlab="Jugadors", main="Mitjana de partits titular de cada jugador per temporada")
```

```
summary(data$PC)
```

```
plot(data$IND, data$PC, ylab="Mitjana de partits complets per temporada",  
xlab="Jugadors", main="Mitjana de partits complets de cada jugador per temporada")
```

```
summary(data$PS)
```

```
plot(data$IND, data$PS, ylab="Mitjana de partits suplent per temporada",  
xlab="Jugadors", main="Mitjana de partits suplent de cada jugador per temporada")
```

```
summary(data$PX)
```

```
plot(data$IND, data$PX, ylab="Mitjana de partits substituït", xlab="Jugadors",  
main="Mitjana de partits substituït de cada jugador per temporada")
```

```
summary(data$C)
```

```
plot(data$IND, data$C, ylab="Mitjana de convocatòries sense jugar", xlab="Jugadors",  
main="Mitjana de convocatòries sense jugar de cada jugador per temporada")
```

```
summary(data$Min)
```

```
plot(data$IND, data$Min, ylab="Mitjana de minuts disputats", xlab="Jugadors",  
main="Mitjana de minuts disputats de cada jugador per temporada")
```

```
summary(data$TG)
```

```
plot(data$IND, data$TG, ylab="Mitjana de targetes grogues", xlab="Jugadors",  
main="Mitjana de targetes grogues de cada jugador per temporada")
```

```
summary(data$TV)
```

```
plot(data$IND, data$TV, ylab="Mitjana de targetes vermelles", xlab="Jugadors",  
main="Mitjana de targetes vermelles de cada jugador per temporada")
```

```
summary(data$G)
```

```
plot(data$IND, data$G, ylab="Mitjana de gols", xlab="Jugadors", main="Mitjana de  
gols de cada jugador per temporada")
```

```
summary(data$MinG)
```

```
plot(data$IND, data$MinG, ylab="Mitjana de minuts/gol", xlab="Jugadors",  
main="Mitjana de minuts/gol de cada jugador per temporada")
```

```
summary(data$GP)
```

```
plot(data$IND, data$GP, ylab="Mitjana de gols de penal", xlab="Jugadors",  
main="Mitjana de gols de penal de cada jugador per temporada")
```

```
summary(data$GPP)
```

```
plot(data$IND, data$GPP, ylab="Mitjana de gols en pròpia", xlab="Jugadors",  
main="Mitjana de gols en pròpia de cada jugador per temporada")
```

```
summary(data$GE)
```

```
plot(data$IND, data$GE, ylab="Mitjana de gols encaixats", xlab="Jugadors",  
main="Mitjana de gols encaixats de cada jugador per temporada")
```

```
summary(data$PG)
```

```
plot(data$IND, data$PG, ylab="Mitjana de partits guanyats", xlab="Jugadors",  
main="Mitjana de partits guanyats de cada jugador per temporada")
```

```
summary(data$PE)
```

```
plot(data$IND, data$PE, ylab="Mitjana de partits empatats", xlab="Jugadors",
main="Mitjana de partits empatats de cada jugador per temporada")
```

```
summary(data$PP)
```

```
plot(data$IND, data$PP, ylab="Mitjana de partits perduts", xlab="Jugadors",
main="Mitjana de partits perduts de cada jugador per temporada")
```

```
table(data$Win)
```

```
plot(data$IND, data$Win, ylab="Mitjana de lligues per temporada", xlab="Jugadors",
main="Mitjana de lligues guanyades de cada jugador per temporada")
```

```
table(data$Time)
```

```
plot(data$IND, data$Time, ylab="Períodes de supervivència", xlab="Jugadors",
main="Temps de supervivència dels jugadors partíceps de l'estudi")
```

```
# Preparar dades per l'anàlisi
```

```
ndata <- read.csv2("treball2.csv", header = TRUE, sep=';')
```

```
IND <- c(1:537)
```

```
ndata <- data.frame(ndata, IND)
```

```
# Observació de la corba de supervivència per cada variable
```

```
library(survival)
```

```
survpos <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$Pos)
```

```
plot(survpos, lty=c("solid", "solid", "solid", "solid"), col=c("black", "blue", "red",
"green"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de
sobreviure", main="Probabilitats de sobreviure en funció de la posició dels jugadors")
```

```
legend("topright", c("Porter", "Defensa", "Migcampista", "Davanter"), lty=c("solid",
"solid", "solid", "solid"), col=c("green", "blue", "red", "black"))
```

```
survteam <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$Team)
```

```
plot(survteam, lty=c("solid", "solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red", "green"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitats de sobreviure en funció dels equips dels dels jugadors")
```

```
legend("topright", c("Champions", "UEFA", "Mitja taula", "Zona baixa", "Descens"), lty=c("solid", "solid", "solid", "solid", "solid"), col=c("black", "red", "blue", "green", "purple"))
```

```
survfix <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$Fix)
```

```
plot(survfix, lty=c("solid", "solid"), col=c("purple", "blue"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitats de sobreviure", main="Probabilitat de sobreviure en funció del nombre d'equips dels jugadors")
```

```
legend("topright", c("Un equip", "Més d'un equip"), lty=c("solid", "solid"), col=c("blue", "purple"))
```

```
survnac <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$Nac)
```

```
plot(surnac, lty=c("solid", "solid", "solid", "solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red", "green", "yellow", "orange"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció de la nacionalitat")
```

```
legend("topright", c("Argentina", "Brazil", "Espanya", "França", "Other", "Portugal", "Uruguay"), lty=c("solid", "solid", "solid", "solid", "solid", "solid", "solid"), col=c("black", "red", "blue", "green", "purple", "yellow", "orange"))
```

```
survpj <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$PJ)
```

```
plot(survpj, lty=c("solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció dels partits jugats")
```

```
legend("topright", c("(30,38]", "(20,30]", "(10,20]", "[1,10]"), lty=c("solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red"))
```

```
survpt <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$PT)
```

```
plot(survpt, lty=c("solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció de les titularitats")
```

```
legend("topright", c("(30,38]", "(20,30]", "(10,20]", "[0,10]"), lty=c("solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red"))
```

```
survpc <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$PC)
```

```
plot(survpc, lty=c("solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció dels partits complets")
```

```
legend("topright", c("(30,38]", "(20,30]", "(10,20]", "[0,10]"), lty=c("solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red"))
```

```
survps <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$PS)
```

```
plot(survps, lty=c("solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red", "green"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció dels partits suplent")
```

```
legend("topright", c("(20,24]", "(15,20]", "(10,15]", "(5,10]", "[0,5]"), lty=c("solid", "solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red", "green"))
```

```
survpx <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$PX)
```

```
plot(survpx, lty=c("solid", "solid", "solid", "solid", "solid"), col=c("red", "black", "purple", "blue", "green"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció de les substituït")
```

```
legend("topright", c("(15,20]", "(10,15]", "(5,10]", "(0,5]", "[0]"), lty=c("solid", "solid", "solid", "solid", "solid"), col=c("purple", "black", "blue", "red", "green"))
```

```
survc <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$C)
```

```
plot(survc, lty=c("solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció de les convocatòries sense jugar")
```

```
legend("topright", c("(30,37]", "(20,30]", "(10,20]", "[0,10]"), lty=c("solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red"))
```

```
survmin <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$Min)
```

```
plot(survmin, lty=c("solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció dels minuts disputats")
```

```
legend("topright", c("(3000,3420]", "(2000,3000]", "(1000,2000]", "[1,1000]"), lty=c("solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red"))
```

```
survtg <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$TG)
```

```
plot(survtg, lty=c("solid", "solid", "solid"), col=c("black", "purple", "blue"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció de les targetes grogues rebudes")
```

```
legend("topright", c("(10,14]", "(5,10]", "[0,5]"), lty=c("solid", "solid", "solid"), col=c("black", "purple", "blue"))
```

```
survtv <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$TV)
```

```
plot(survtv, lty=c("solid", "solid", "solid", "solid"), col=c("black", "purple", "blue", "red"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció de les targetes vermelles rebudes")
```

```
legend("topright", c("(2,3]", "(1,2]", "(0,1]", "[0]"), lty=c("solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red"))
```

```
survg <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$G)
```

```
plot(survg, lty=c("solid", "solid", "solid", "solid", "solid"), col=c("red", "black", "purple", "blue", "green"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció de la mitjana de gols anotats")
```

```
legend("topright", c("(20,36.5]", "(10,20]", "(5,10]", "(0,5]", "[0]"), lty=c("solid", "solid", "solid", "solid", "solid"), col=c("purple", "black", "blue", "red", "green"))
```

```
survmg <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$MinG)
```

```
plot(survmg, lty=c("solid", "solid", "solid", "solid", "solid"), col=c("red", "black", "purple", "blue", "green"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció dels minuts per gol")
```



```
legend("topright", c("(5000,11752]", "(1000,5000]", "(500,1000]", "(0,500]", "[0]"),
lty=c("solid", "solid", "solid", "solid", "solid"), col=c("blue", "black", "purple", "red",
"green"))
```

```
survgp <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$GP)
```

```
plot(survgp, lty=c("solid", "solid", "solid", "solid", "solid"), col=c("red", "black",
"purple", "blue", "green"), xlab="Temps de supervivència en períodes de mercat",
ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció dels gols
de penal")
```

```
legend("topright", c("(4,6.78]", "(2,4]", "(1,2]", "(0,1]", "[0]"), lty=c("solid", "solid",
"solid", "solid", "solid"), col=c("blue", "purple", "black", "red", "green"))
```

```
survgpp <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$GPP)
```

```
plot(survgpp, lty=c("solid", "solid", "solid", "solid", "solid"), col=c("red", "black",
"purple", "blue", "green"), xlab="Temps de supervivència en períodes de mercat",
ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció dels gols
en pròpia porteria")
```

```
legend("topright", c("(0.8,1]", "(0.4,0.8]", "(0.2,0.4]", "(0,0.2]", "[0]"), lty=c("solid",
"solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red", "green"))
```

```
survge <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$GE)
```

```
plot(survge, lty=c("solid", "solid", "solid", "solid", "solid"), col=c("red", "black",
"purple", "blue", "green"), xlab="Temps de supervivència en períodes de mercat",
ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció dels gols
encaixats")
```

```
legend("topright", c("(60,74]", "(40,60]", "(20,40]", "(0,20]", "[0]"), lty=c("solid",
"solid", "solid", "solid", "solid"), col=c("blue", "purple", "black", "red", "green"))
```

```
survpg <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$PG)
```

```
plot(survpg, lty=c("solid", "solid", "solid", "solid", "solid", "solid"), col=c("red", "black",
"purple", "blue", "green", "orange"), xlab="Temps de supervivència en períodes de
mercat", ylab="Probabilitat de sobreviure", main="Probabilitat de sobreviure en funció
dels partits guanyats")
```

```
legend("topright", c("(20,26.1]", "(15,20]", "(10,15]", "(5,10]", "(0,5]", "[0]"),
lty=c("solid", "solid", "solid", "solid", "solid", "solid"), col=c("blue", "purple",
"black", "green", "red", "orange"))
```

```

survpe <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$PE)

plot(survpe, lty=c("solid", "solid", "solid", "solid"), col=c("red", "black", "purple",
"blue"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de
sobreviure", main="Probabilitat de sobreviure en funció dels partits empatats")

legend("topright", c("(12,14]", "(8,12]", "(4,8]", "[0,4]"), lty=c("solid", "solid", "solid",
"solid"), col=c("red", "purple", "black", "blue"))

survpp <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$PP)

plot(survpp, lty=c("solid", "solid", "solid", "solid"), col=c("red", "black", "purple",
"blue"), xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de
sobreviure", main="Probabilitat de sobreviure en funció dels partits perduts")

legend("topright", c("(15,20]", "(10,15]", "(5,10]", "[0,5]"), lty=c("solid", "solid", "solid",
"solid"), col=c("black", "red", "purple", "blue"))

survw <- survfit(Surv(ndata$Time,ndata$Status==1)~ndata$Win)

plot(survw, lty=c("solid", "solid", "solid", "solid"), col=c("red", "black", "purple", "blue"),
xlab="Temps de supervivència en períodes de mercat", ylab="Probabilitat de
sobreviure", main="Probabilitat de sobreviure en funció dels títols guanyats")

legend("topright", c("(0.8,1]", "(0.4,0.8]", "(0,0.4]", "[0]"), lty=c("solid", "solid", "solid",
"solid"), col=c("purple", "black", "red", "blue"))

## Hipòtesis de proporcionalitat

cdata <- read.csv2("treball.csv", header = TRUE, sep=';')
IND <- c(1:537)
cdata <- data.frame(cdata, IND)

Y=Surv(cdata$Time,cdata$Status==1)
cox.zph(coxph(Y~PJ, data=cdata))
cox.zph(coxph(Y~PT, data=cdata))

```

```
cox.zph(coxph(Y~PC, data=cdata))
cox.zph(coxph(Y~PS, data=cdata))
cox.zph(coxph(Y~PX, data=cdata))
cox.zph(coxph(Y~C, data=cdata))
cox.zph(coxph(Y~Min, data=cdata))
cox.zph(coxph(Y~TG, data=cdata))
cox.zph(coxph(Y~TV, data=cdata))
cox.zph(coxph(Y~G, data=cdata))
cox.zph(coxph(Y~MinG, data=cdata))
cox.zph(coxph(Y~GP, data=cdata))
cox.zph(coxph(Y~GPP, data=cdata))
cox.zph(coxph(Y~GE, data=cdata))
cox.zph(coxph(Y~PG, data=cdata))
cox.zph(coxph(Y~PE, data=cdata))
cox.zph(coxph(Y~PP, data=cdata))
cox.zph(coxph(Y~Win, data=cdata))
cox.zph(coxph(Y~Pos, data=cdata))
cox.zph(coxph(Y~Team, data=cdata))
cox.zph(coxph(Y~Fix, data=cdata))
cox.zph(coxph(Y~Nac, data=cdata))
```

```
summary(coxph(Y~Team, data=cdata))
```

```
## Model de Cox
```

```
summary(coxph(Y~PJ, data=cdata))
summary(coxph(Y~PT, data=cdata))
summary(coxph(Y~PS, data=cdata))
summary(coxph(Y~PX, data=cdata))
```

```
summary(coxph(Y~C, data=cdata))
summary(coxph(Y~Min, data=cdata))
summary(coxph(Y~TV, data=cdata))
summary(coxph(Y~GP, data=cdata))
summary(coxph(Y~GPP, data=cdata))
summary(coxph(Y~GE, data=cdata))
summary(coxph(Y~PE, data=cdata))
summary(coxph(Y~Win, data=cdata))
summary(coxph(Y~Pos, data=cdata))
summary(coxph(Y~Team, data=cdata))
summary(coxph(Y~Nac, data=cdata))
```

```
summary(coxph(Y~PJ, data=cdata))
summary(coxph(Y~PT, data=cdata))
summary(coxph(Y~Team, data=cdata))
summary(coxph(Y~PJ+PT, data=cdata))
summary(coxph(Y~PJ*PT, data=cdata))
summary(coxph(Y~PJ+Team, data=cdata))
summary(coxph(Y~PT+Team, data=cdata))
summary(coxph(Y~Team+PJ*PT, data=cdata))
```

```
AIC(coxph(Y~PJ, data=cdata))
AIC(coxph(Y~PT, data=cdata))
AIC(coxph(Y~Team, data=cdata))
AIC(coxph(Y~PJ+PT, data=cdata))
AIC(coxph(Y~PJ*PT, data=cdata))
AIC(coxph(Y~PJ+Team, data=cdata))
AIC(coxph(Y~PT+Team, data=cdata))
AIC(coxph(Y~Team+PJ*PT, data=cdata))
```

Model ampliat de les variables que no compleixen l'hipòtesis de riscos
proporcionals

```
sort(cdata$Time[cdata$Status==1])  
cuts <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,16,18,20)
```

```
maPC<-survSplit(Surv(Time, Status)~.,data=cdata,cut=cuts, end="Time",  
event="Status",start="start",id="id")
```

```
maPC$tPC=maPC$PC*maPC$Time
```

```
Y <- Surv(maPC$start, maPC$Time, maPC$Status)
```

```
modPC <- coxph(Y~PC+tPC, data=maPC)
```

```
modPC
```

```
maTG<-survSplit(Surv(Time, Status)~.,data=cdata,cut=cuts, end="Time",  
event="Status",start="start",id="id")
```

```
maTG$tTG=maTG$TG*maTG$Time
```

```
Y <- Surv(maTG$start, maTG$Time, maTG$Status)
```

```
modTG <- coxph(Y~TG+tTG, data=maTG)
```

```
modTG
```

```
maG<-survSplit(Surv(Time, Status)~.,data=cdata,cut=cuts, end="Time",  
event="Status",start="start",id="id")
```

```
maG$tG=maG$G*maG$Time
```

```
Y <- Surv(maG$start, maG$Time, maG$Status)
```

```
modG <- coxph(Y~G+tG, data=maG)
```

```
modG
```

```
maMinG<-survSplit(Surv(Time, Status)~.,data=cdata,cut=cuts, end="Time",  
event="Status",start="start",id="id")
```

```
maMinG$tMinG=maMinG$MinG*maMinG$Time
```

```

Y <- Surv(maMinG$start, maMinG$Time, maMinG$Status)
modMinG <- coxph(Y~MinG+tMinG, data=maMinG)
modMinG

maPG<-survSplit(Surv(Time, Status)~.,data=cdata,cut=cuts, end="Time",
event="Status",start="start",id="id")
maPG$tPG=maPG$PG*maPG$Time
Y <- Surv(maPG$start, maPG$Time, maPG$Status)
modPG <- coxph(Y~PG+tPG, data=maPG)
modPG

maPP<-survSplit(Surv(Time, Status)~.,data=cdata,cut=cuts, end="Time",
event="Status",start="start",id="id")
maPP$tPP=maPP$PP*maPP$Time
Y <- Surv(maPP$start, maPP$Time, maPP$Status)
modPP <- coxph(Y~PP+tPP, data=maPP)
modPP

maFix<-survSplit(Surv(Time, Status)~.,data=cdata,cut=cuts, end="Time",
event="Status",start="start",id="id")
maFix$nFix[maFix$Fix=='S']<-0
maFix$nFix[maFix$Fix=='N']<-1
maFix$tFix=maFix$nFix*maFix$Time
Y <- Surv(maFix$start, maFix$Time, maFix$Status)
modFix <- coxph(Y~nFix+tFix, data=maFix)
modFix

## Models estratificats

Y=Surv(cdata$Time,cdata$Status==1)

```

```

summary(coxph(Y~Team+PJ*PT+strata(PC), data=cdata))
summary(coxph(Y~Team+PJ*PT+strata(TG), data=cdata))
summary(coxph(Y~Team+PJ*PT+strata(MinG), data=cdata))
summary(coxph(Y~Team+PJ*PT+strata(PG), data=cdata))
summary(coxph(Y~Team+PJ*PT+strata(PP), data=cdata))
summary(coxph(Y~Team+PJ*PT+strata(Fix), data=cdata))

```

```
## Model ampliat
```

```

model<-survSplit(Surv(Time, Status)~.,data=cdata,cut=cuts, end="Time",
event="Status",start="start",id="id")
model$nFix[model$Fix=='S']<-0
model$nFix[model$Fix=='N']<-1
model$tFix=model$nFix*model$Time
model$tPC=model$PC*model$Time
model$tTG=model$TG*model$Time
model$tMinG=model$MinG*model$Time
model$tPG=model$PG*model$Time
model$tPP=model$PP*model$Time
Y <- Surv(model$start, model$Time, model$Status)
summary(coxph(Y~Team+PJ*PT+tFix+tPC+tTG+tMinG+tPG+tPP, data=model))

summary(coxph(Y~Team+PJ*PT+tFix+tPC+tMinG+tPG, data=model))

```