

Grau en Estadística

Títol: Model predictiu del resultat d'un partit de futbol:
desenvolupament del model i capacitat predictiva

Autor: Glòria Trulls Noguera

Director: Erik Cobo Valeri

Departament: Estadística i Investigació Operativa (UPC)

Convocatòria: Juny 2018



Dedicatòria:

Al Carles, per ensenyar-me el que realment és l'esperit de superació i no rendir-se mai.

Gràcies per estar sempre al meu costat, per tots els ànims que em vas donar dia rere dia durant la realització d'aquest Treball de Final de Grau, i de segur, m'has continuat donant fins al dia d'avui.

No cal que afegixi res més, ja que com em deies: "tu ja m'entens".

Agraïments:

Aquest Treball de Final de Grau no hauria estat possible sense l'ajuda de l'Erik Cobo. Per això vull agrair l'acceptació de ser el tutor d'aquest treball i que m'hagi guiat i aconsellat durant tot el procés del projecte. La seva gran implicació, col·laboració i el continu seguiment han estat fonamentals pel desenvolupament d'aquest treball.

També agrair al Jordi Cortés, per la seva disponibilitat i la col·laboració que ha aportat en algun punt del treball.

Finalment, agrair a totes les persones que m'han donat suport i motivació durant la realització del treball, especialment a la meva família i amics.

A tots vosaltres, moltes gràcies.

Resum:

La Quiniela és un joc d'atzar relacionat amb el món del futbol que, principalment, consisteix en marcar si d'un partit es creu que acabarà en victòria local *1*, empataran *X* o guanyarà l'equip visitant *2*.

L'objectiu del nostre treball és arribar a un model lineal generalitzat per estimar el resultat de cada partit de les jornades de La Primera Divisió Espanyola, en forma de *1X2* tal com els resumeix *La Quiniela*.

A partir dels models GAM amb família multinomial predirem la nostra resposta multinomial *1X2*. Seguidament validarem aquest model; estudiant la col·linealitat, el calibrat i la discriminació del model. D'aquesta forma arribarem a encertar un 51,18% dels partits en la mostra d'entrenament i un 54,94% en la mostra de validació.

Paraules claus: Model Multinomial, Capacitat de Predicció, Probabilitat victòria local, Probabilitat empat, Probabilitat victòria visitant, La Quiniela.

Classificació AMS: 62J12 Generalized linear models.

Abstract

La Quiniela (Football Pools) is a football chance game that mainly consists of guessing the results of the weekly fixtures: 1 for a home win, X for a drawn and 2 for an away win.

The goal of our study is to get a generalized lineal model in order to estimate the weekly fixtures of Spanish Premier League as they are presented in the Quiniela 1X2.

We will predict the results 1X2 based on GAM models and multinomial family. We will validate this model by studying the model's collinearity, calibration and discrimination. Following this method we will be able to match 51,18% of the results of the training sample and 54,94% of the validation sample.

Keywords: Multinomial Model, Predictive Capacity, Home win probability, Drawn probability, Away win probability, *La Quiniela* (Football Pools).

AMS Classification: 62J12 Generalized linear models.

Índex

I. INTRODUCCIÓ.....	1
1. LA QUINIELA	1
2. OBJECTIUS	1
II. METODOLOGIA.....	3
1. VARIABLE RESPOSTA	3
2. VARIABLES PREDICTORES.....	3
3. SELECCIÓ I VALIDACIÓ DEL MODEL PREDICTIU	6
3.1. <i>Col·linealitat</i>	<i>7</i>
3.2. <i>Calibratge</i>	<i>8</i>
3.3. <i>Discriminació</i>	<i>8</i>
III. RESULTATS	10
1. DESCRIPCIÓ DE LES VARIABLES	10
1.1. <i>Descripció univariant.....</i>	<i>10</i>
1.2. <i>Descripció bivariant.....</i>	<i>12</i>
2. SELECCIÓ DEL MODEL I DE LES VARIABLES INCLOSES	17
3. VALIDACIÓ DEL MODEL PREDICTIU	18
3.1. <i>Col·linealitat:</i>	<i>18</i>
3.2. <i>Calibratge</i>	<i>19</i>
3.3. <i>Discriminació</i>	<i>20</i>
4. MOSTRA VALIDACIÓ.....	25
IV. CONCLUSIONS	27
V. DISCUSSIÓ	28
1.1. FORMES D' APOSTES	28
1.2. PREMIS	29
BIBLIOGRAFIA.....	30
ANNEXOS	31
1. DESCRIPCIÓ UNIVARIANT	31
2. DESCRIPCIÓ BIVARIANT	62
3. FORMULES DELS MODELS	90
3. DISCRIMINACIÓ	91
4. CODI R	96

I. INTRODUCCIÓ

La Primera divisió espanyola té una audiència mitjana de 9.767.936 espectadors per jornada, més d'un 20% sobre el total de la població. El futbol és un tema de debat, comentat per una gran part de la nostra societat. Per això, volem estudiar la capacitat d'un model lineal generalitzat per predir el resultat de cada partit tal com els resumeix *La Quiniela*, en 1X2.

1. La Quiniela

Loterías y Apuestas del Estado va crear l'any 1946 *La Quiniela* per legalitzar les apostes sobre els resultats dels partits de futbol d'Espanya.

Avui dia, tot i que amb l'arribada d'internet han aparegut un gran nombre d'apostes esportives diferents, *La Quiniela* segueix sent el joc d'atzar relacionat amb el futbol més conegut en la nostra societat. Aquest joc d'atzar consisteix a encertar el resultat de 14 partits de futbol marcant, 1, si es creu que guanyarà l'equip local, X, si empataran o si guanyarà l'equip visitant, 2. També hi ha el partit número 15, on s'ha de marcar quants gols es pensa que farà cada un dels dos equips que disputen el partit: 0,1,2 o M —més de dos gols—. Aquest últim partit només es tindrà en compte si s'han encertat els 14 primers resultats.

La Quiniela més coneguda és aquella formada per equips de la Primera i Segona Divisió Espanyola. Actualment, també n'hi ha per partits de Champions, Copa del Rei,...

En les jornades de la Lliga Espanyola, es pot apostar com a màxim fins divendres, provocant que els partits que es juguen el divendres segur que no entraran en *La Quiniela* de la jornada. S'intenta que dels 14 partits de resposta 1X2, 7 siguin de Primera Divisió i 7 de Segona.

2. Objectius

El nostre principal objectiu és arribar a obtenir un model predictiu dels resultats, de forma 1X2, dels partits de la Primera Divisió Espanyola, també coneguda com *LaLiga*. Per això, utilitzarem els partits de la jornada 10 fins a l'última jornada 38, ja que creiem que a partir d'aquesta jornada els resultats seran més predictius i obtindrem una millor estimació.

Així doncs, en finalitzar aquest treball es vol comptar amb un bon model de resposta multinomial que hagi superat les validacions adients i sigui fiable a l'hora de predir partits de *LaLiga* de la jornada 10 fins a la 38.

II. METODOLOGIA

Per poder arribar al nostre objectiu, s'han utilitzat els coneixements assolits en el Grau d'Estadística. Les assignatures d'Introducció a la Informàtica i Programació, Models Lineals Generalitzats, Econometria i Probabilitat i Processos Estocàstics han estat unes eines de gran utilitat.

Tot i això, per tal de predir una variable que pot prendre tres valors ($1X2$) ens ha estat necessari fer una cerca sobre aquest tipus de models predictius, la seva forma de validació i interpretació. D'aquesta manera, la metodologia seguida ha estat la següent:

1. Variable Resposta

La variable resposta que volem arribar a predir és el resultat en què finalitzarà cada un dels partits de les jornades de *LaLiga*, els possibles resultats són els següents:

- 1: Victòria de l'equip local.
- X: Empat.
- 2: Victòria de l'equip visitant.

2. Variables predictores

Per obtenir les variables predictores del nostre model, que no té cap pretensió de ser interpretat causalment, disposem d'una base de dades¹ que conté la següent informació: Temporada, Jornada, Equip local, Equip visitant, Gols de l'equip local, Gols de l'equip visitant. A partir d'aquesta, hem generat les nostres pròpies² variables predictores utilitzant els coneixements en el *software* estadístic *R*.

A continuació, es presenten les variables i el procediment dut a terme per crear-les:

- Posició a la classificació: Variable codificada com a *posL* per l'equip Local i *posV* per l'equip Visitant. Trobem la posició, entre 1 i 20, que ocupa cada equip en el moment abans de disputar el partit. Per generar-les, agafem els punts aconseguits fins a la jornada anterior, i els ordenem major a menor. Aquesta variable ens aporta informació sobre quina zona de la classificació es troba l'equip, i per tant del seu rendiment durant

¹ <https://selffy.com/p/NnB5/>

² La base de dades generada es facilitarà a aquelles persones que ho demanin. Tot i això, també es troba el procés de creació.

la temporada disputada.

- Punts a la classificació: Variable codificada com a $ptsL$ per l'equip local i $ptsV$ per l'equip visitant. Trobem els punts sumats fins a la jornada corresponent per a cada equip. Per generar-les, agafem els punts aconseguits fins a la jornada anterior, i se li suma 3 si l'equip guanya, 1 si empata i 0 si perd. Aquesta variable està relacionada, com l'anterior, amb el rendiment. En un principi ens pot semblar més informativa, ja que no és el mateix que s'enfrontin l'equip de la posició 1 contra el de la posició 2, quan aquest es porten 1 punt, que quan es porten 10 punts. Tot i que molt possiblement estaran molt correlacionades, de moment ens quedem amb les dues variables i més endavant decidirem quina utilitzem en el nostre model.

- Punts guanyats com a local o visitant durant la temporada: Variable codificada com a $ptsL$ per l'equip local i $ptsV$ per l'equip visitant. Per a l'equip local trobem el nombre de punts que ha aconseguït jugant com a local, i per l'equip visitant els punts guanyats jugant com a visitant durant la temporada. Per generar-les, agafem els punts aconseguits fins a la jornada anterior d'aquesta variable, i se li suma 3 si l'equip guanya, 1 empata i 0 si perd. Aquesta variable informa si un equip és molt potent jugant a casa seva, per exemple, o el contrari.

- Gols a favor durant la temporada: Variable codificada com a gfl per l'equip local i gfV per l'equip visitant. Trobem els gols fets durant la temporada. Per generar-les, en acabar els partits se sumen els gols aconseguits en el partit amb els gols que portaven abans de jugar-lo. Aquesta variable ens dóna una idea del joc ofensiu de l'equip.

- Gols en contra durant la temporada: Variable codificada com a gcL per l'equip local i gcV per l'equip visitant. Trobem els gols encaixats durant la temporada. Per generar-les, en acabar els partits se sumen els gols que li han marcat l'equip contrari en el partit amb els gols en contra que tenien abans de jugar-lo. Aquesta variable ens dóna una idea de la capacitat defensiva de l'equip.

- Partits guanyats durant la temporada: Variable codificada com a pgL per l'equip local i pgV per l'equip visitant. Trobem el nombre de partits guanyats durant la temporada. Per a generar-les, en acabar un partit se sumen 1 si l'equip l'ha guanyat i 0 altrament en la variable. És una variable important per veure la capacitat de victòria de l'equip.

- Partits empatats durant la temporada: Variable codificada com a peL per l'equip local i peV per l'equip visitant. Trobem el nombre de partits empatats durant la temporada. Per generar-les, en acabar un partit se sumen 1 si l'equip l'ha empatat i 0 altrament en la variable.

- Partits perduts durant la temporada: Variable codificada com a ppL per l'equip local i ppV per l'equip visitant. Trobem el nombre de partits perduts durant la temporada. Per

generar-les, en acabar un partit se sumen 1 si l'equip l'ha perdut i 0 altrament en la variable. És una variable que ens informa de la debilitat de l'equip.

- Partits seguits guanyats: Variable codificada com a *guanyadesL* per l'equip local i *guanyadesV* per l'equip visitant. Trobem el nombre de partits guanyats seguits des de l'última derrota o empat. Per generar-les, en acabar un partit, si l'equip ha guanyat se suma 1 a la variable que compta el nombre de partits seguits guanyats, en canvi, en cas de pèrdua o empat, la variable passa a ser 0. És una variable que pretén donar-nos una idea del funcionament de l'equip en el moment previ del partit, ja que un equip pot haver començat una temporada guanyant pocs partits, però en un moment pot haver remuntat.

- Equips provinents de la Segona Divisió Espanyola: Variable codificada com a *div2L* per l'equip local i *div2V* per l'equip visitant. Variable binària que pren valor 1 en els equips que han aconseguit pujar a Primera Divisió i 0 en els equips que ja estaven en aquesta lliga l'any anterior.

Per a cada una de les variables anteriors, es fa la resta entre el valor de l'equip local i el del visitant i es guarda en una nova variable. Per exemple, pels *Partits perduts durant la temporada*, restem els valors de les variables $ppL - ppV = pp1$. Aquestes transformacions les codifiquem amb un 1 al final del nom de la variable

Seguidament, es crea una altra nova variable, que és la variable esmentada anteriorment dividida pel nombre de partits jugats fins al moment previ de disputar el nou partit. Per exemple, si volem predir la jornada 12, pels *Partits perduts durant la temporada* la calculem de la següent forma: $\frac{ppL - ppV}{\text{última jornada disputada}} = pp2$, on l'última jornada disputada en aquest cas prendria el valor 11. Aquestes transformacions les codifiquem amb un 2 al final del nom de la variable.

D'aquesta forma el que es pretén aconseguir és la diferència entre els dos equips tenint en compte el moment de joc.

Així doncs, tenim tres formes de treballar: amb les variables originals, amb les variables restades entre l'equip local i visitant i la variable anterior dividida per l'última jornada disputada. Quan seleccionem el model, veurem quina és la forma més eficaç i còmode.

Per tenir un major coneixement de les variables anteriors i per validar que no prenen valors fora de rang o sense sentit previ, realitzem una descriptiva univariant.

Per a les variables numèriques trobem el nombre d'observacions, la mitjana, la desviació estàndard, el rang... entre d'altres. També les representem de forma gràfica a partir d'histogrames.

En el cas de les variables categòriques, trobem el percentatge d'individus que es troben en cada categoria i gràfics *pie*.

Continuant en la línia d'exploració de les variables, complementem la descriptiva univariant amb una descriptiva bivariant. A partir de gràfics *Scatterplots* i el càlcul de les correlacions.

És interessant afegir que disposem d'una base de dades que proveeix la premissa estadística d'independència entre files, ja que cada partit és independent a la resta.

3. Selecció i validació del model predictiu

Hem decidit treballar amb els models GAM (*Generalized Additive Models*) amb la família multinomial. Aquest model ajusta les probabilitats de situar-nos en cadascuna de les tres possibles situacions que ens podem trobar en un partit de futbol:

$$P(Y = 1) = \frac{e^{h_1}}{1 + e^{h_1} + e^{h_2}}$$

$$P(Y = X) = \frac{1}{1 + e^{h_1} + e^{h_2}}$$

$$P(Y = 2) = \frac{e^{h_2}}{1 + e^{h_1} + e^{h_2}}$$

On h_1 i h_2 es calculen de la següent forma per a cadascun dels partits:

$$h_z = \beta_0^z + \sum_{v=1}^p \beta_v^z * X_{v,i} \quad , i = 1, \dots, n$$

Començar amb un subíndex Z indica que tindrem diferents models per cada z. En el nostre cas en farem servir dos: un per predir 1 i un altre per predir 2.

Tot i tenir més d'un model, les probabilitats dels tres possibles resultats 1X2 sumaran sempre 1 per a cada partit, ja que així ho forcem quan fixem 1 en el cas $P(Y = X)$. Aquest és un avantatge d'utilitzar els models GAM.

Per escollir el model final, tindrem en compte les següents magnituds:

- Percentatge de partits encertats: El model ens dóna les probabilitats d'1, X i 2. Apostem pel que té una probabilitat més alta. Si el resultat del partit acaba en el valor que hem escollit el comptem com a encert, en cas contrari com a error.

- AIC (Criteri d'informació d'Akaike): $2k - 2\ln(L)$, on k és el nombre de paràmetres en el model estadístic, i L és el màxim valor de la funció de versemblança pel model estimat.

- AUC (Area Under the Curve)³: Tenim les classes c etiquetades com a 0, 1, 2 ($c-1$). Aquestes classes no han de tenir un ordre intrínsec. Això es compleix en les nostres dades, ja que si apostem a 1 i en realitat el resultat ha estat una X , aquest error té la mateixa penalització que si hagués estat un 2.

El rendiment general de la regla de classificació a la separació de les classes c és la mitjana d'això en tots els parells de classes:

$$M = \frac{2}{c(c-1)} \sum_{i < j} \hat{A}(i, j)$$

on c és el nombre total de classes i $\hat{A}(i, j) = \frac{\hat{A}(i, j) + \hat{A}(j, i)}{2}$ (mesura de la separabilitat entre classes i i j).

- Nombre de paràmetres del model: Comptem els paràmetres utilitzats en el model.

Donat que el nostre objectiu és construir un model predictiu, ens fixem principalment en quines àrees sota la corba (AUC) tenen els diferents models, és a dir, quin té una capacitat de predicció més alta. Tot i això, també tindrem en compte la simplicitat del model, fixant-nos amb el nombre de paràmetres.

Per validar el model escollit seguirem els següents passos:

3.1. Col·linealitat

Primer estudiarem la col·linealitat, a partir del Factor d'Increment de la Variància (FIV) per a cada variable x_j :

$$FIV_j = \frac{1}{1 - R_j^2}$$

on R_j^2 és el valor de R^2 obtingut en un model lineal on la variable resposta és x_j i les variables predictives són la resta de variables predictives del nostre model.

El VIF representa la proporció de la variabilitat (o variància) de la variable que és explicada per la resta de les variables predictores del model. Com és lògic, com més gran

³ Per més informació consultar al primer llibre de la bibliografia.

sigui el factor d'inflació de la variància, major serà la probabilitat que existeixi col·linealitat. Generalment es considera que hi ha col·linealitat quan el VIF és major a 5.

3.2. Calibratge

Per estudiar el calibratge modifiquem el test clàssic de Hosmer-Lemeshow per considerar la magnitud dels desajustos en lloc del p-valor, d'acord amb les noves guies de l'ASA⁴.

Per visualitzar el calibratge realitzem uns gràfics de forma pròpia, a partir el *software R*, construïts de la següent forma:

Primer dividim la mostra de la probabilitat que estudiem ($1, X$ o 2) en decils.

En les abscisses trobem el valor predit per a cada decil, és a dir, la mitjana de la probabilitat en cada grup.

En les ordenades trobem el valor observat d' $1, X$ o 2 , segons la probabilitat que estiguem estudiant, si per exemple, estudiem la probabilitat de què guanyi l'equip local, trobarem el valor observat d' 1 . Per a cada decil calculem la proporció d'encerts p_n i el seu interval de confiança al 95%:

$$IC_{95\%} = p_n \pm 1.96 \sqrt{\frac{p_n(1 - p_n)}{n}}$$

3.3. Discriminació

Finalment, estudiem la discriminació o la capacitat del model per encertar en les seves prediccions. Per a cada una de les tres probabilitats (probabilitat que guanyi l'equip local, probabilitat d'empat i probabilitat que guanyi l'equip visitant) es calcula la funció de densitat pels partits acabats en $1, X$ i 2 i s'obté el seu respectiu AUC.

A partir de la probabilitat que té un AUC més elevat, i per tant la discrimina millor, analitzarem si afegir una segona dimensió o utilitzar les tres dimensions (les tres probabilitats) augmenta la discriminació dels tres possibles resultats $1X2$. Per això,

⁴Tercera guia de la bibliografia.

realitzarem taules creuades de cada una de les dimensions i anotarem quants casos hi ha hagut d'1, X i 2. A partir d'aquests resultats, en cada cas apostarem pel valor 1X2 que tingui una màxima freqüència.

III. RESULTATS

En aquest capítol aplicarem la metodologia a les nostres dades per així obtenir els resultats.

1. Descripció de les variables

Abans de començar a treballar amb el model és important que coneguem les nostres variables i tinguem una primera idea de com es distribueixen.

1.1. Descripció univariant

Tenim una base de dades amb 5800 observacions i 55 variables.

Recollides i generades totes les variables que utilitzarem per a la creació del model, farem una anàlisi de les estadístiques⁵ d'aquestes, per així tenir-ne més coneixement i per garantir que no prenen valors fora de rang o sense sentit previ:

- Temporada: Factor que pot prendre valors 1997-98, 1998-99, 1999-00, 2000-01,... fins a 2016-17. Cada temporada té 290 observacions, que corresponen a 10 partits per cada una de les 29 últimes jornades.

- Jornada: Factor que pot contenir valors d'entre 10 a 38. Cada jornada té 200 observacions, que corresponen als 10 partits disputats per jornada durant els 20 anys que estudiem.

- Local i Visitante: Factor que pot contenir qualsevol nom dels equips que han passat per la Primera Divisió Espanyola entre la temporada 1997-98 i la 2016-17.

- GolesLocal i GolesVisitante: Variables numèriques discretes dels gols aconseguits per l'equip local i el visitant en el partit disputat. Poden prendre qualsevol valor d'entre els nombres naturals, a la base de dades la variable *GolesLocal* té rang entre [0, 10] i *GolesVisitante* [0,8].

- posL i posV: Variables numèriques discretes de la posició de l'equip local i el visitant. Poden prendre valors d'entre 1 i 20.

⁵ A l'annex es troben les taules completes i els gràfics de la descriptiva univariant.

- ptsL i ptsV: Variables numèriques discretes que poden prendre valors en el rang [0,111], on el valor màxim 111 l'obtindria un equip on a l'últim partit de la temporada (la jornada 38) ha guanyat els 37 partits jugats amb anterioritat. En el nostre cas, *ptsL* té rang [1,97] i *ptsV* [1,95].

- ptsL i ptsV: Variables numèriques discretes que poden prendre valors en el rang [0, 54], on el valor màxim l'obtindria l'equip que a l'últim partit de la temporada (la jornada 38) hagués guanyat els 37 partits jugats durant aquesta temporada. En el nostre cas, *ptsL* té rang [0,52] i *ptsV* [0,47].

- pgL i pgV: Variables numèriques discretes que poden contenir valors del rang [0,37], on el valor màxim el tindria l'equip que hagués guanyat tots els partits disputats abans de jugar l'última jornada de la temporada. En el nostre cas, tant *pgL* com *pgV* tenen rang [0,31]

- peL i peV: Variables numèriques discretes dels partits empatats durant la temporada disputada que poden contenir valors del rang [0,37]. En el nostre cas, tant *peL* com *peV* tenen rang [0,18].

- ppL i ppV: Variables numèriques discretes dels partits perduts durant la temporada disputada, poden contenir valors del rang [0,37]. En el nostre cas, tant *ppL* com *ppV* tenen rang [0,28].

- quanyadesL i quanyadesV: Variables numèriques discretes dels partits seguits guanyats des de l'última derrota o empat, poden prendre valors del rang [0,37]. En el nostre cas, *ppL* té rang [0,15] i *ppV* [0,16].

- gfL i gfV: Variables numèriques discretes dels gols a favor durant la temporada disputada que poden prendre qualsevol valor d'entre els nombres naturals, a la base de dades la variable *gfL* té rang [2, 117] i *gfV* [2,115].

- gcL i gcV: Variables numèriques discretes dels gols en contra durant la temporada disputada que poden prendre qualsevol valor d'entre els nombres naturals, a la base de dades la variable *gcL* té rang [3, 88] i *gcV* [3,89].

- div2L i div2V: Variables binàries que prenen valor 1 si l'equip prové de la Segona Divisió Espanyola i 0 en el cas contrari, a cada temporada hi ha 3 dels equips amb valor 1 i 17 amb valor 0.

- guanyatL, empat, guanyatV: Variables que contindran valor 1 o 0. Per a cada partit, una de les tres prendrà valor 1 (en el cas que hagi guanyat l'equip local, serà *guanyatL*. Si el resultat ha acabat en empat, serà la variable *empat*. Si acaba en guany de l'equip visitant, serà *guanyatV*) i els dos resultats no ocorreguts tindran valor 0.

- quiniela i quiniela2: És la variable que volem predir, *quiniela* és codificada de forma numèrica i *quiniela2* com a factor. La variable *quiniela* pren valor 0 quan un partit acaba en empat, 1 quan guanya l'equip local, i 2 quan guanya l'equip visitant. La variable *quiniela2* pren valor 1 quan un partit acaba en victòria local, X quan hi ha empat, i 2 quan guanya l'equip visitant. En la base de dades de 5800 partits, trobem que guanya l'equip local en 2830 (48,79%) partits, 1427 (26,60%) acaben en empat i 1543 (24,61%) finalitzen amb victòria visitant.

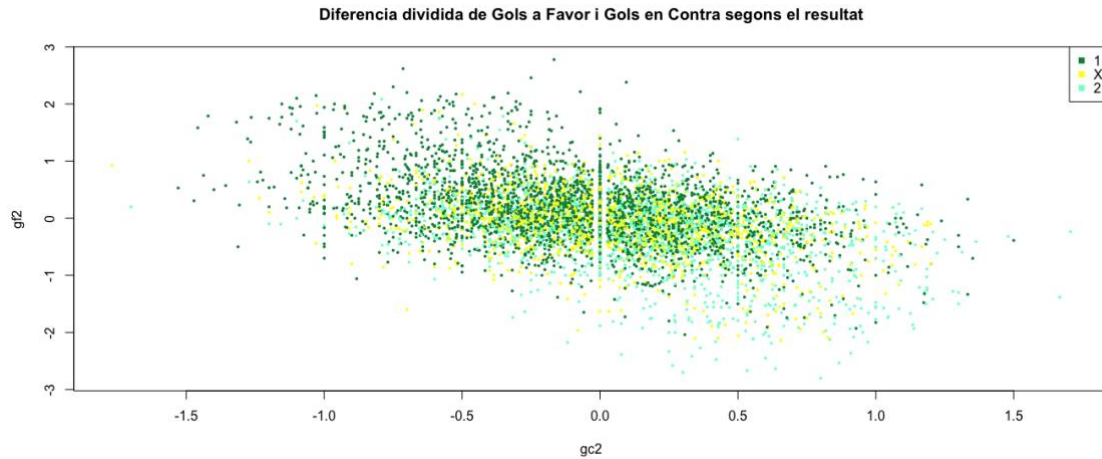
Totes aquestes variables tenen unes col·linealitats teòriques, és per això que estudiarem les correlacions. Per intentar evitar-les, també creem les variables diferències entre els valors de l'equip local i visitant, i les mateixes variables dividides per les jornades disputades.

Cal esmentar que en la base de dades no hi ha cap dada mancant, és a dir, *missing*.

1.2. Descripció bivariant

Després d'haver analitzat les nostres dades de forma univariant, és interessant també fer una descriptiva bivariant⁶. D'aquesta forma tindrem una primera idea de si les nostres variables predictores són una bona eina per discriminar els partits que finalitzen amb un resultat a favor de l'equip local, amb empat o amb victòria visitant:

⁶ A l'annex es troben la resta de gràfics.



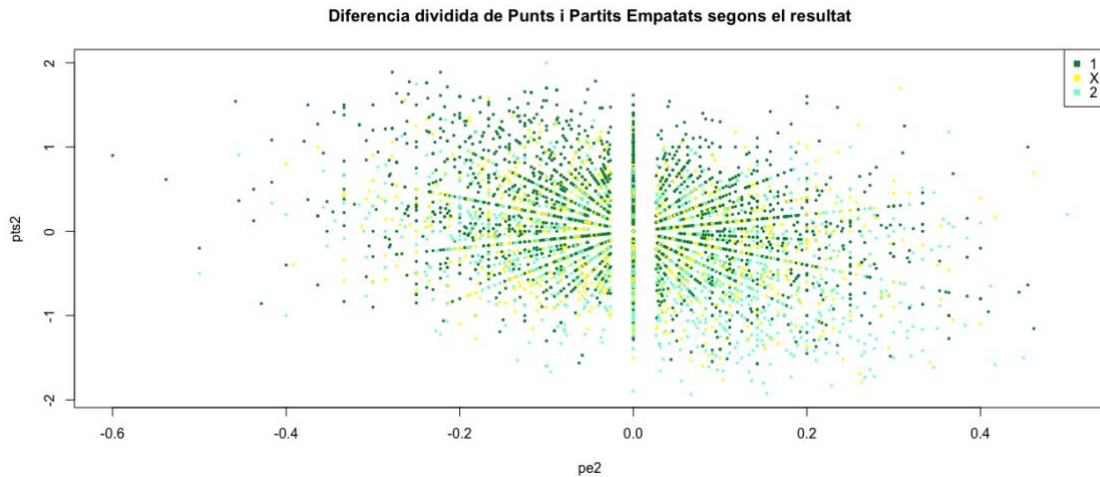
Gràfic 3.1: En l'eix de les ordenades trobem la variable $gf2$ ($\frac{\text{gols a favor equip local} - \text{gols a favor equip visitant}}{\text{última jornada disputada}}$) i en l'eix de les abscisses la variable $gc2$ ($\frac{\text{gols en contra equip local} - \text{gols en contra equip visitant}}{\text{última jornada disputada}}$). Cada un dels partits es troba marcat amb un color diferent, segons el seu resultat final sigui 1, X o 2.

En aquest primer gràfic podem observar una certa correlació negativa entre les variables que ens recull la informació dels gols a favor $gf2$ i els gols en contra $gc2$ dels dos equips que disputen el partit.

Els partits que tendeixen a acabar en victòria local són aquells en què l'equip local té un registre de gols a favor considerablement més alt que l'equip visitant i més baix en gols en contra. En el cas contrari, els partits acaben majoritàriament en victòria visitant.

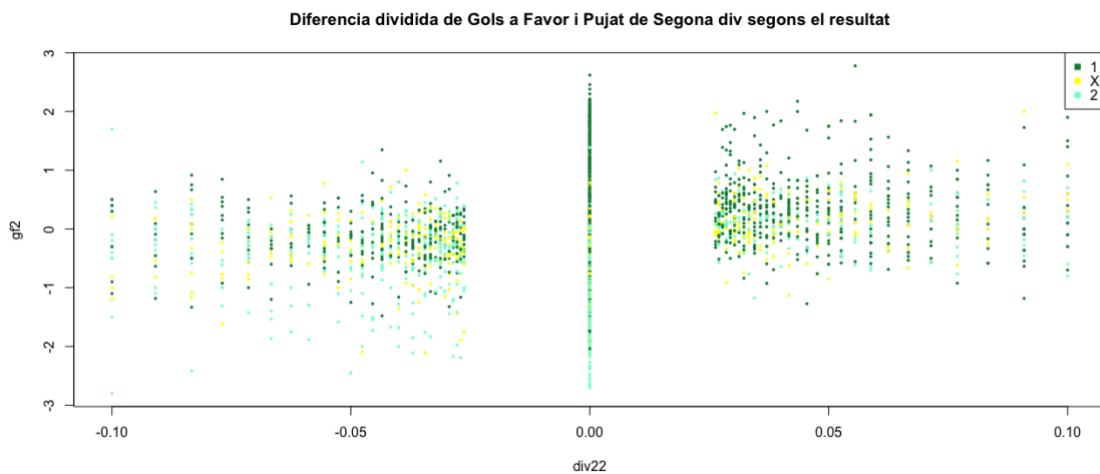
Així doncs, en el segon quadrant hi trobem un domini de punts 1 i en el quart quadrant un major nombre de punts 2. Al centre del núvol de punts no observem un domini clar per cap dels tres possibles resultats.

Pel que fa als partits finalitzats en X, els trobem repartits al llarg del núvol de punts, fet que ens dóna una primera impressió de què predir aquest tipus de partits ens serà difícil.



Gràfic 3.2: En l'eix de les ordenades trobem la variable $pts2$ ($\frac{\text{punts equip local} - \text{punts equip visitant}}{\text{última jornada disputada}}$) i en l'eix de les abscisses la variable $pe2$ ($\frac{\text{partits empatats equip local} - \text{partits empatats equip visitant}}{\text{última jornada disputada}}$). Cada un dels partits es troba marcat amb un color diferent, segons el seu resultat final sigui 1, X o 2.

Per observar si tenint la informació dels punts $pts2$ i dels partits empatats $pe2$ ens ajuda a discriminar els partits empatats de la resta de resultats, tenim l'anterior gràfic. Tot i això, es continua observant que els punts X segueixen estant repartits arreu del gràfic.



Gràfic 3.3: En l'eix de les ordenades trobem la variable $gf2$ ($\frac{\text{gols a favor equip local} - \text{gols a favor equip visitant}}{\text{última jornada disputada}}$) i en l'eix de les abscisses la variable $div22$ ($\frac{\text{equip local pujat de } 2^{\text{a}} \text{ div} - \text{equip visitant pujat de } 2^{\text{a}} \text{ div}}{\text{última jornada disputada}}$). Cada un dels partits es troba marcat amb un color diferent, segons el seu resultat final sigui 1, X o 2.

Abans de comentar el gràfic, és important aclarir que els enfrontaments entre dos equips que han fet el salt de la Segona Divisió Espanyola a la Primera Divisió i dos equips que l'any anterior ja estaven a la Primera Divisió Espanyola són aquells que tenen un valor de $div22$ igual a 0, ja que els equips que han pujat de divisió prenen valor 1 i la resta 0, llavors $div22 = \frac{1 \cdot (div2L) - 1 \cdot (div2V)}{\text{última jornada disputada}} = 0$ pels dos equips que han fet al salt a la

Primera Divisió i $div22 = \left(\frac{0 (div2L) - 0 (div2V)}{\text{última jornada disputada}} \right) = 0$ pels que ja estaven a la màxima categoria.

Tot i que $div2L$ i $div2V$ són variables binàries, hem trobat convenient fer les operacions corresponents per crear $div22$, ja que a l'inici de la lliga pot tenir més impacte el fet que un equip provingui de la segona divisió, i a mesura que es va disputant la temporada, aquesta importància disminueixi (és el que aconseguim dividint per l'última jornada disputada).

En el gràfic es veu com els punts formen unes línies, a més, aquestes són simètriques a valors positius i negatius en l'eix de les abscisses. Aquestes línies són conseqüència directa de les transformacions comentades anteriorment, provocant que cada valor (tant positiu com negatiu) formi part d'una jornada en concret i el signe ens indica quin era l'equip que ha pujat de la Segona Divisió Espanyola. Per exemple, si en la jornada 10 hi ha hagut un partit que l'equip local prové de la Segona Divisió i l'equip Visitant ja formava part de la Primera Divisió ens queda el següent valor positiu per la variable $div22 \left(\frac{1-0}{10} \right) = 0,1$.

A partir del gràfic podem observar que els enfrontaments entre dos equips que l'any anterior pertanyien a la mateixa lliga (partits que en la variable $div22$ tenen valor 0) tendeix a guanyar l'equip que té un registre de gols a favor més elevat. Això ens ho indica el fet que quan el valor de $gf2$ té un valor positiu, denominen de forma contundent els punts 1 i quan és negatiu, els punts 2.

Com hem observat a partir dels gràfics anteriors, tenim unes variables força correlacionades.

Per veure si aquest fet pot ser perillós pel nostre model, trobem quantificats els seus valors en la matriu de correlacions (només de les variables que tenen una correlació més elevada):

	pos	pts	ptsLV	pg	pp	gf	gc	pe
pos	1.00	-0.89	-0.73	-0.87	0.79	-0.70	0.63	0.23
pts	-0.89	1.00	0.81	0.98	-0.89	0.84	-0.71	-0.27
ptsLV	-0.73	0.81	1.00	0.80	-0.71	0.69	-0.57	-0.24
pg	-0.87	0.98	0.80	1.00	-0.77	0.83	-0.64	-0.47
pp	0.79	-0.89	-0.71	-0.77	1.00	-0.73	0.74	-0.20
gf	-0.70	0.84	0.69	0.83	-0.73	1.00	-0.41	-0.26
gc	0.63	-0.71	-0.57	-0.64	0.74	-0.41	1.00	-0.05
pe	0.23	-0.27	-0.24	-0.47	-0.20	-0.26	-0.05	1.00

Taula 3.1.: Correlacions ordenades de més a menys elevades, on s'han obviat les variables amb una baixa correlació. En marró es presenten les variables que s'han decidit eliminar a l'hora de construir el model per evitar la col·linealitat. En color verd trobem les correlacions baixes, en taronja les que comença'n a ser elevades però no preocupants, i en color vermell les que tenen una correlació més alta, tot i que aquestes no superen el valor ± 0.8 .

Significat variables: cada una de les variables i , en la base de dades les trobem codificades amb el nom que trobem a la taula i finalitzades amb un 2 (pos2, pts2,...). Aquest 2 ens indica que per cada variable (vari) s'ha fet la següent operació en cada individu: $\frac{varL_i - varV_i}{\text{última jornada disputada}}$. Trobem les següents variables: pos: posició, pts: punts, ptsLV: punts com a local o com a visitant (segons a quin camp jugui l'equip), pg: gols a favor, pp: partits perduts, gf: gols a favor, gc: gols en contra, pe: partits perduts.

Com es pot observar, aquestes variables presenten una col·linealitat molt elevada, i cal descartar-ne algunes a l'hora de realitzar el model. D'aquesta forma evitarem que el model sigui inestable i la variància de les estimacions dels coeficients siguin grans.

Pel que fa a les variables $pg2$ ($\frac{pgL - pgV}{\text{última jornada disputada}}$) i $pp2$ ($\frac{ppL - ppV}{\text{última jornada disputada}}$), una és gairebé una funció lineal de l'altra, on quedarien els $pe2$ ($\frac{peL - peV}{\text{jornada corresponent}}$) pel mig, i per tant, una de les dues variables és redundant. És lògic que un equip fort, on d'un nombre de jornades jugades en concret n'ha guanyat la majoria, el seu nombre de partits perduts sigui baix. Per aquesta raó, decidim eliminar $pg2$.

El mateix passa amb les variables $pts2$ ($\frac{ptsL - ptsV}{\text{última jornada disputada}}$) i $pos2$ ($\frac{posL - posV}{\text{última jornada disputada}}$). Un equip que se situa a la zona alta de la classificació tindrà una posició alta, i per tant també més punts. Així doncs, descartem la variable $pts2$ pel nostre model.

2. Selecció del model i de les variables incloses

A partir dels models GAM amb família multinomial predirem la nostra resposta multinomial, que pot prendre valors $1X2$.

La guia TRIPOD⁷ anuncia que en els estudis predictius no es faci la selecció de la mostra d'aprenentatge i de validació a l'atzar. Aconsella agafar una primera part de la mostra com a aprenentatge i la resta com a validació. La raó fonamental que ho justifica és que en un futur, quan farem servir el model, no l'utilitzarem amb una mostra aleatòria de les dades.

Així doncs, disposem de 5800 observacions. Les primeres 4930 formaran part de les dades *training*, i són les que utilitzarem a l'hora de realitzar el model. Les 870 observacions restants, que pertanyen als tres últims anys de les nostres dades, són les dades *test*, que ens serviran per veure l'ajust del nostre model.

Tenim els quatre models següents:

- Model Complet: Per a cada variable var_i utilitzem la que informa sobre l'equip local $varL_i$ i sobre l'equip visitant $varV_i$.

- Model Diferències: Per a cada variable var_i utilitzem la diferència entre la variable que fa referència a l'equip local i l'equip visitant: $varL_i - varV_i = var1_i$.

- Model Diferències i Dividit: Per a cada variable var_i utilitzem la diferència entre la variable que fa referència a l'equip local i l'equip visitant i ho dividim per l'última jornada disputada: $\frac{varL_i - varV_i}{\text{última jornada disputada}} = var2_i$.

- Model Diferències i Dividit de les variables significatives: Del model anterior, utilitzem només les variables significatives.

Cada un d'aquests models presenten les seves fórmules, en el cas del Model Diferències i Dividit de les variables significatives, són les següents:

$$h_1 = 0,687564 + 0,648253 * X_{gf2} - 0,276962 * X_{gc2}$$

$$h_2 = -0,034821 - 0,470530 * X_{pos2} - 0,569423 * X_{gf2} + 0,501434 * X_{gc2}$$

De forma similar es construeixen les fórmules en la resta de models, que es troben a l'annex.

⁷ Segona guia de la bibliografia.

Per escollir el model final, calculem les següents magnituds a partir les dades d'entrenament:

	Nº Paràmetres	% Encert	AIC	AUC
Model Complet	34	51,03%	7507,69	0,5358
Model Diferències	16	51,11%	7494,14	0,5376
Model Diferències i Dividit	16	<u>51,35%</u>	<u>7460,19</u>	<u>0,54</u>
Model Diferències i Dividit de les variables significatives	<u>5</u>	<u>51,30%</u>	7461,70	<u>0,5399</u>

Taula 3.2.: Per a cada un dels quatre models es troben els valors de les mesures, per així escollir-ne un. Per a cada mesura, se subratlla el millor valor obtingut d'entre els quatre models.

De la taula anterior ressaltem que del Model Complet de 34 paràmetres hem passat a un Model Diferències i Dividit de 16 paràmetres, on amb menys paràmetres som capaços d'augmentar el percentatge d'encert (passant de 51,03% a 51,35%) i el valor del AUC (on millora de 0,5358 a 0,54).

A partir el Model Diferències i Dividit, utilitzem les variables significatives com a selecció de variables, quedant-nos amb el Model Diferències i Dividit de les variables significatives. Aquest model amb 5 paràmetres és el més simple i presenta una àrea sota la corba (AUC) de 0,5399. Així doncs, és el que utilitzarem a l'hora de fer les prediccions.

3. Validació del model predictiu

Un cop tenim el model predictiu és essencial fer una validació, per així conèixer en quin grau les prediccions coincideixen amb les observacions.

3.1. Col·linealitat:

Per comprovar si en el nostre model hi ha presència de col·linealitat, utilitzem el Factor d'Increment de la Variància (VIF):

Variable predita per la resta de variables	VIF
gc	1,66
gf	1,95
pos	2,70

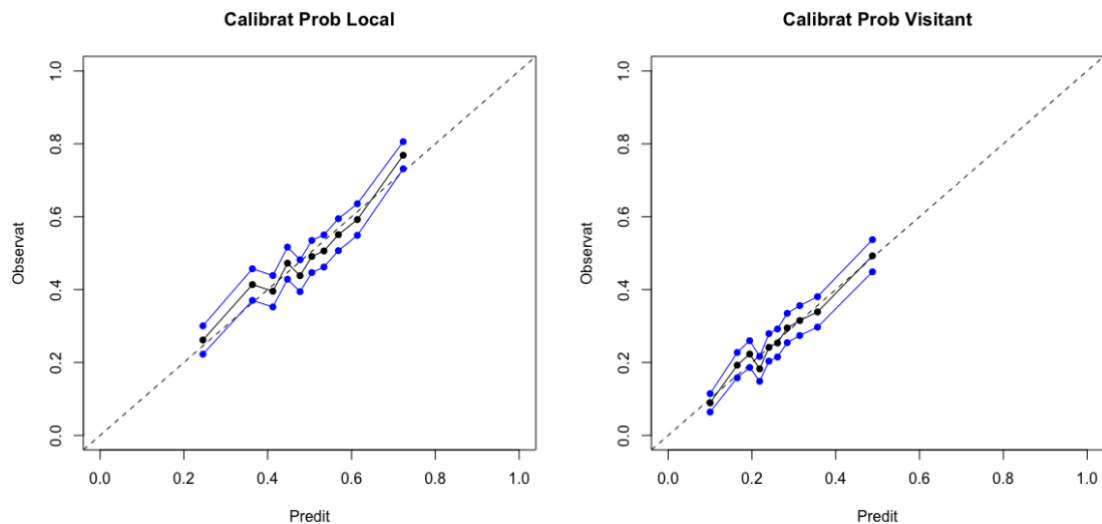
Taula 3.3.: Per a cada una de les tres variables es calcula el seu valor VIF. Significat de les variables: cada una de les variables, en la base de dades les trobem codificades amb el nom que trobem i finalitzades amb un 2 (pos2, pts2,...). Aquest 2 ens indica que són variables diferència del valor de l'equip local i el visitant, i dividides per l'última jornada disputada en el seu moment. Trobem les següents; pos: posició, gf: gols a favor, gc: gols en contra.

Observem que cap de les tres variables predictorres té un valor VIF superior a 5, on a partir d'aquest valor generalment es considera que hi ha col·linealitat. Per tant, el nostre model no té indicis de patir problemes de col·linealitat.

3.2. Calibratge

En el calibratge del model comparem el nombre predit d'esdeveniments amb el nombre observat en grups d'individus.

Utilitzem les diferències entre els observats i els predits per cada un dels decils que es proposa en Hosmer-Lemeshow, ja que ens proporciona informació útil per estudiar el calibratge, i una manera de veure-ho és a partir dels gràfics següents:



Gràfic 3.4.: En els dos gràfics trobem el calibratge de la probabilitat indicada, a l'esquerra de la probabilitat de guany de l'equip local i a la dreta de l'equip visitant. En els punts negres les abscisses són la mitjana de la probabilitat predita en el decil, i les ordenades la probabilitat d'encert en aquest. Els punts blaus tenen el mateix valor en les abscisses, i en les coordenades tenen el valor de l'interval de confiança al 95% superior i inferior de la probabilitat d'encert en el decil. A cada decil es troben 493 observacions. Com a exemple, expliquem com s'ha construït el primer decil del calibrat de la probabilitat que guanyi l'equip local. El valor de l'abscissa és la mitjana de la probabilitat de què guanyi l'equip local en el primer decil. Pel que fa als valors que es troben en les ordenades; el punt negre és la mitjana de cops que hem obtingut com a resultat 1 en el primer decil - un 26% - i els dos punts blaus són l'IC95%, que l'hem calculat de la següent forma: $\pm 1,96 * 0,26 * \sqrt{\frac{0,26*(1-0,26)}{493}} = [0'22, 0'3]$.

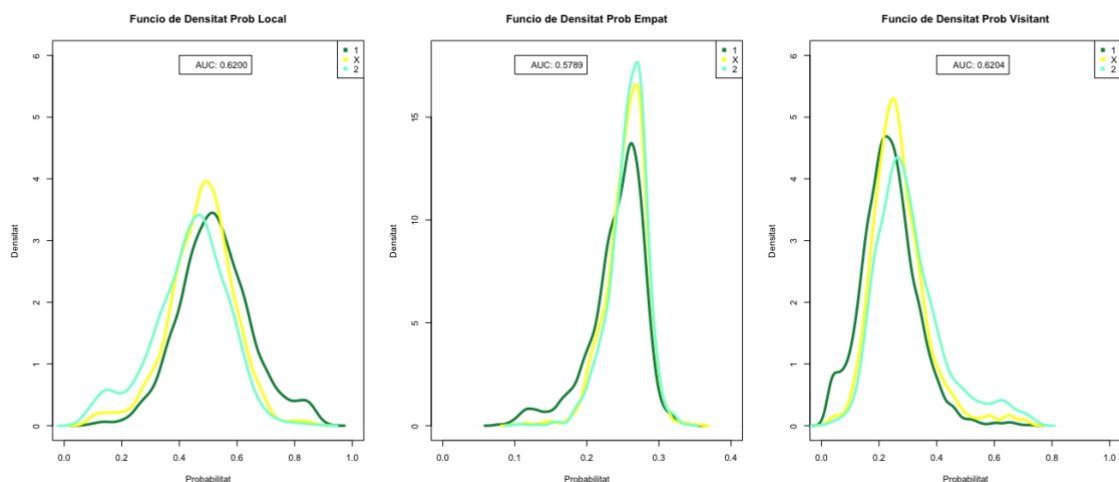
A diferència de les bandes de la regressió lineal, en aquest cas els intervals tenen una amplada similar per a cada grup, ja que la divisió s'ha fet a partir de decils, i per tant, cada grup té el mateix nombre d'observacions.

En els dos casos el calibratge és correcte, tenint sempre la recta d'igualtat dins de l'interval de confiança. Tot i això, al centre sembla que el calibratge no és perfecte, on en alguns punts, en comptes d'augmentar la probabilitat observada respecte al decil anterior, disminueix, i això crea les puntes que veiem.

3.3. Discriminació

Per avaluar el grau en què el nostre model distingeix entre individus en els quals passa cada un dels tres esdeveniments, estudiem la discriminació.

Primer, observem la funció de densitat de cadascun dels tres escenaris possibles:



Gràfic 3.5.: Per a cada probabilitat predita dels tres possibles resultats es calcula la funció de densitat dels partits acabats en 1, X i 2. També es mostra l'AUC en cada cas.

A simple vista, veiem una gran superposició en les funcions de densitat dels tres gràfics. Però, si entrem més en detall, podem notar com dels extrems podem extreure'n informació útil.

En el cas de la funció de densitat de la probabilitat de què guanyi l'equip local, observem que els partits que finalitzen guanyant l'equip visitant són els que tenen una probabilitat més baixa, i per tant la seva corba està situada més a l'esquerra. Pel que fa a l'empat, es troba clarament situat al centre, i per tant serà més complicat a l'hora de predir aquests tipus de partits. En quant els resultats acabats en victòria local, la seva corba es decanta cap a la dreta, tenint probabilitats més elevades.

Quelcom semblant passa a la funció de densitat de la probabilitat que guanyi l'equip visitant, però a la inversa: els partits en què acaba guanyant l'equip local són els que tenen una probabilitat més baixa, en canvi, els que tenen una probabilitat més elevada són els que finalitzen amb victòria visitant. Pel que fa a la funció de densitat d'empat, les tres corbes estan molt superposades, però per a valors amb una probabilitat molt baixa, la corba dels partits acabats en 1 es discrimina força respecte a la resta.

A la vista dels resultats, en un primer moment arribem a la conclusió de què la probabilitat que guanyi l'equip visitant és la que té una millor mesura de la capacitat de discriminació, tenint el valor d'AUC més alt, concretament de 0,6204. Tot i això, volem anar més enllà, i veure si afegir una segona dimensió (com seria la probabilitat de guany de l'equip local o la d'empat) ens aporta una millor discriminació. També analitzarem el resultat utilitzant les tres dimensions, i per últim veurem si fent servir la probabilitat de guany de l'equip local i la d'Empat de forma conjunta ens aporta una millor discriminació:

Prob Local, Encert = 50,97%

0-0,1	0,1-0,2	0,2-0,3	0,3-0,4	0,4-0,5	0,5-0,6	0,6-0,7	0,7-0,8	0,8-0,9	0,9-1	Total
<u>1</u>	<u>14</u>	<u>70</u>	283	679	755	395	134	78	2	4930
<u>5</u>	<u>25</u>	<u>54</u>	189	426	376	129	15	7	0	
<u>15</u>	<u>69</u>	<u>103</u>	251	439	308	93	12	3	0	

Prob Empat, Encert = 48,90%

0-0,1	0,1-0,2	0,2-0,3	0,3-0,4	0,4-1	Total
7	301	2055	48	0	4930
1	44	1154	27		
0	39	1223	31		

Prob Visitant, Encert = 51,18%

0-0,1	0,1-0,2	0,2-0,3	0,3-0,4	0,4-0,5	0,5-0,6	0,6-0,7	0,7-0,8	0,8-1	Total
173	691	1033	408	<u>80</u>	<u>15</u>	<u>10</u>	<u>1</u>	0	4930
17	245	594	267	<u>67</u>	<u>19</u>	<u>14</u>	<u>3</u>		
13	204	524	334	<u>113</u>	<u>50</u>	<u>45</u>	<u>10</u>		

Taula 3.4.: Cada una de les tres probabilitats (Probabilitat que guanyi l'equip local, probabilitat d'empat i probabilitat que guanyi l'equip visitant) s'ha dividit de 0 a 1 en deu intervals d'una llargada de 0,1.

En cada un dels intervals hem calculat els partits que realment han finalitzat en victòria local (color verd fort), els que han finalitzat en empat (color groc) i el que han acabat en victòria visitant (color verd fluix). Els casos remarcats en negreta són aquells decils on hi ha més casos en què el partit ha acabat en victòria local, i per tant, en aquests casos recomanem apostar 1. Els resultats subratllats són els que predomina la victòria visitant, i per tant recomanem apostar 2.

Es calcula el percentatge d'encert per a cada cas si s'aplica el criteri d'aposta explicat anteriorment.

Per exemple, en la primera cel·la de la primera taula (0-0,1 de la taula Prob Local) el resultat que succeeix en més partits és la victòria visitant, ja que dels 21 partits 15 (71,43%) acaben en aquest resultat, per això recomanem apostar pel 2. Es realitza aquest mateix raonament per a cada cel·la, quedant així la següent aposta: $\frac{15 + 69 + 103 + 283 + 679 + 755 + 395 + 134 + 78 + 2}{4930} = 0,5097$.

La probabilitat de què guanyi l'equip visitant és la que ens proporciona un major percentatge d'encert, sent aquest del 51,18%.

Encert: 51,18%

Probabilitat guany equip visitant	Probabilitat guany equip local										
	0- 0,1	0,1- 0,2	0,2- 0,3	0,3- 0,4	0,4- 0,5	0,5- 0,6	0,6- 0,7	0,7- 0,8	0,8- 0,9	0,9- 1	Total
0-0,1	<u>0</u>	<u>0</u>	<u>0</u>	0	0	0	0	93 10 10	78 7 3	2 0 0	173 17 13
0,1-0,2	<u>0</u>	<u>0</u>	<u>0</u>	0	0	255 111 109	395 129 93	41 5 2	0	0	691 245 204
0,2-0,3	<u>0</u>	<u>0</u>	<u>0</u>	3 0 1	530 329 324	500 265 199	0	0	0	0	1033 594 524
0,3-0,4	<u>0</u>	<u>0</u>	<u>5</u> <u>1</u> <u>2</u>	254 169 217	149 97 115	0	0	0	0	0	408 267 334
0,4-0,5	<u>0</u>	<u>1</u> <u>0</u> <u>0</u>	<u>53</u> <u>47</u> <u>80</u>	<u>26</u> <u>20</u> <u>33</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>80</u> <u>67</u> <u>113</u>
0,5-0,6	<u>0</u>	<u>3</u> <u>13</u> <u>29</u>	<u>12</u> <u>6</u> <u>21</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>15</u> <u>19</u> <u>50</u>
0,6-0,7	<u>0</u> <u>2</u> <u>5</u>	<u>10</u> <u>12</u> <u>40</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>10</u> <u>14</u> <u>45</u>
0,7-0,8	<u>1</u> <u>3</u> <u>10</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>1</u> <u>3</u> <u>10</u>
0,8-1	<u>0</u>	<u>0</u>	<u>0</u>	0	0	0	0	0	0	0	0
Total	<u>1</u> <u>5</u> <u>15</u>	<u>14</u> <u>25</u> <u>69</u>	<u>70</u> <u>54</u> <u>103</u>	283 189 251	679 426 439	755 376 308	395 129 93	134 15 12	78 7 3	2 0 0	4930

Taula 3.5.: Taula creuada de la probabilitat de guany de l'equip visitant i de l'equip local. Els valors en color verd fort són els resultats acabats en victòria local, els grocs en empat i els verds clar en victòria visitant.

En negreta trobem les files on les marginals de la probabilitat de què guanyi l'equip Visitant i dominen com a resultat la victòria de l'equip Local. En subratllat hi ha les marginals on en els majors de casos

guanya l'equip Visitant. Es troba el mateix per a les columnes, tenint en compte les marginals de la probabilitat de què guanyi l'equip local.

El càlcul del percentatge d'encert està calculat tenint en compte que en cada cel·la apostarem pel resultat que té una freqüència més gran.

Per exemple, a la cel·la de la fila 0,3-0,4 i columna 0,2-0,3 es troba en negreta i subratllada, ja que dels 1009 partits que tenen una probabilitat de guany de l'equip visitant entre 0,3 i 0,4 408 partits acaben en victòria local. En canvi, dels 227 partits que tenen una probabilitat de guany de l'equip local entre 0,2 i 0,3 103 acaben en victòria visitant. Mirem la cel·la per veure quin valor domina, en aquest cas, trobem 5 casos en què el partit finalitza en victòria local, 1 en empat i 2 en victòria visitant. Com que el resultat que més succeeix és la victòria local, apostarem per aquest en forma d'1.

Es troba un exemple d'aposta completa en el peu de la taula anterior.

A partit la taula anterior observem que utilitzar com a segona dimensió la probabilitat de victòria local no aporta més informació a la qual ja tenim en la probabilitat de victòria visitant, ja que es fa exactament la mateixa aposta, i per tant el percentatge d'encert és el mateix que en la tercera taula de la Taula 3.4, és a dir un 51,18%.

A l'annex es troben la resta de taules similars a l'anterior, pels casos on s'utilitza la probabilitat de què guanyi l'equip local i la probabilitat d'empat, la probabilitat de què guanyi l'equip visitant i la probabilitat d'empat, i finalment on s'utilitzen les tres probabilitats. En els dos primers casos la probabilitat d'encert no millora respecte a quan utilitzem només la probabilitat de què guanyi l'equip visitant. Quan s'utilitzen les tres probabilitats sí que augmenta el percentatge d'encert, sent aquest del 51,20%. Tot i això, la millora és tan petita, que per realitzar les apostes utilitzarem només la probabilitat de què guanyi l'equip Visitant i seguirem el següent criteri d'aposta:

$$Prob 2 \leq 0,4 \rightarrow Aposta 1$$

$$Prob 2 > 0,4 \rightarrow Aposta 2$$

En cap cas s'aposta X, ja que a partir a les taules anteriors hem pogut veure com en cap combinació de les variables predictores s'arriba a assolir suficient freqüència per aconsellar apostar per ell.

Així doncs, utilitzant les dades *d'entrenament* obtenim un 51,18% d'encert, i la nostra aposta quedaria de la següent forma:

	Aposta 1	Aposta X	Aposta 2	<u>Total</u>
Resultat 1	2305 (51,19%)	0	106 (24,82%)	2411
Resultat X	1123 (23,87%)	0	103 (24,13%)	1226
Resultat 2	1075 (24,94%)	0	218 (51,05%)	1293
<u>Total</u>	4503 (100%)	0	427 (100%)	4930

Taula 3.6.: Taula creuada del resultat real i l'aposta realitzada. Entre parèntesis es troba el % per columna.

En la taula anterior s'observa dels 5800 partits en 4503 (77,63%) apostem per 1, on encertem 51,19% dels partits que apostem per aquest resultat. Pel que fa a la resta de casos, apostem per 2, i d'aquesta forma encertem un 51,05% dels partits que apostem per la victòria de l'equip visitant.

4. Mostra Validació

A partir del model predictiu utilitzem les dades test per veure com s'ajusta el mateix model per a les dades *test*.

Pel que fa a la capacitat de discriminació, el valor de l'AUC de la probabilitat de què guanyi l'equip visitant augmenta en comparació amb les dades d'*entrenament*, i passa a ser de 0,7010. Així doncs, s'ajusta de forma correcta.

Tot seguit, apliquem l'estratègia d'aposta per a les dades *test*, i obtenim els següents resultats:

	Aposta 1	Aposta X	Aposta 2	<u>Total</u>
Resultat 1	394 (53,67%)	0	25 (18,38%)	419
Resultat X	174 (23,71%)	0	27 (19,85%)	201
Resultat 2	166 (22,62%)	0	84 (61,77%)	250
<u>Total</u>	734 (100%)	0	136 (100%)	870

Taula 4.1.: Taula creuada del resultat real i l'aposta realitzada. Entre parèntesis es troba el % per columna.

Els resultats són bastant semblants als que hem obtingut a partir de les dades d'*entrenament*, tot i això, en el cas de les dades *test* tenim un percentatge d'encert una

mica més elevat.

Observem que dels 870 partits en 734 (84,37%) apostem per 1, on encertem 53,67% dels partits que apostem per aquest resultat. Pel que fa a la resta de casos, apostem per 2, i d'aquesta forma encertem un 61,77% dels partits que apostem per la victòria de l'equip visitant.

D'aquesta forma, encertem 478 partits del 870, és a dir, un 54,94% dels partits.

IV. CONCLUSIONS

Predir esdeveniments esportius relacionats amb el món del futbol ha estat sempre un gran repte que poques vegades s'ha pogut assolir de forma satisfactòria.

La base del treball recau en la recollida de dades, on hem generat les nostres pròpies variables predictores.

Una de les mancances d'aquest treball ha estat la impossibilitat de recollir variables que haurien pogut estar decisives pels resultats, com per exemple: els dies passats des de l'últim partit jugat, la possessió durant la temporada, l'estat del terreny de joc, l'estat d'ànim dels jugadors i l'entrenador, lesions de jugadors importants per l'equip... entre d'altres.

D'aquesta manera, instem a que es realitzin en el futur treballs on es recullin aquestes variables, ja que seria interessant veure si tenint-les en compte el model que hem proposat millora. Tot i aquestes mancances de les quals en som conscients, creiem que hem creat una metodologia prou fiable.

A partir de la creació de la base de dades, hem portat a terme el model predictiu. Ha estat un gran repte, ja que hem emparat un model de resposta multinomial.

Un cop hem aconseguit el model multinomial per predir 1X2 l'hem validat i hem calculat la seva capacitat predictiva.

En la nostra base de dades, un 48,79% de partits acaben en victòria local. És quelcom lògic que sense tenir un model previ s'aposti sempre a 1.

A partir del nostre model hem millorat aquest punt de partida. En el cas de les dades d'aprenentatge hem obtingut un AUC del 0,6204 – utilitzant la probabilitat de què guanyi l'equip visitant – i un percentatge d'encert del 51,18% dels partits disputats. Pel que fa a la mostra de validació, hem augmentat tant el valor de l'AUC com el percentatge d'encert: l'AUC a partir les probabilitats de què guanyi l'equip visitant ha estat de 0,7010 i el percentatge d'encert de 54,94%.

Cal remarcar que donem més importància als resultats obtinguts en aquesta segona mostra, ja que la primera pot estar afectada pel sobreajustament.

V. Discussió

En aquest treball ens hem centrat a estimar les probabilitats de 1X2 dels resultats dels partits de la Primera Divisió Espanyola. Tot i això, seria interessant tenir un estudi complet sobre *La Quiniela*, on els següents passos a seguir serien els següents:

Primer de tot queda per veure la predicció pels resultats de la Segona Divisió.

Seguidament s'ha d'avaluar les opcions d'apostar de forma simple i de forma múltiple. A partir d'aquí, elaborar una estratègia d'aposta i estudiar si es podrien obtenir beneficis. Això es pot fer a partir la teoria de decisió, l'optimització de la investigació operativa o les eines d'optimització econòmica.

Per fer aquest estudi, s'ha de tenir coneixement sobre les formes d'apostes i com es reparteixen els premis en *La Quiniela*:

1.1. Formes d'apostes

A l'hora d'apostar, es pot fer de diferents formes⁸:

- Simple: per a cada un dels 14 primers partits només es marca una de les tres opcions 1X2, i els gols que es creu que faran els dos equips del partit 15 (0,1,2 o M). Té un cost de 0,75 €.

- Múltiple: en els 14 primers partits es poden fer apostes *dobles*, marcant dues opcions de les tres possibles 1X2, o *triples*, marcant les tres opcions 1X2. Es poden marcar fins a 14 partits de forma doble (sense fer-ne cap de triple) i 9 de forma triple (sense fer-ne cap de doble). També es pot marcar combinant *dobles* i *triples*. Quan s'aposta de forma múltiple, cada una de les combinacions de 1X2 simples dels 14 partits comptarà com a una aposta, tenint així:

$$\text{nombre total apostes} = 2^{\text{dobles}} * 3^{\text{triples}}$$

apostes en *La Quiniela*, on *dobles* són el nombre d'apostes dobles realitzades i *triples* les apostes triples fetes.

El cost de cada aposta és de 0,75 €, així doncs, el cost d'una aposta múltiple (en euros) serà el següent:

⁸ Es troben altres formes d'apostes al següent enllaç: <https://www.loteriasyapuestas.es/es/la-quiniela/como-se-juega/como-jugar-a-la-quiniela>

$$\text{cost} = \text{nombre total apostes} * 0,75$$

El cas del partit 15 només es pot marcar un resultat per a cadascun dels dos equips, i serà comú per a cada una de les apostes totals.

1.2. Premis

A *La Quiniela* es destina el 55% de la recaptació als premis, que es distribueixen per categories:

Categoria	Encerts	Premi
Especial	14 Encerts + el Ple al 15	7.5% de la recaptació
1 ^a	14	16% de la recaptació
2 ^a	13	7.5% de la recaptació
3 ^a	12	7.5% de la recaptació
4 ^a	11	7.5% de la recaptació
5 ^a	10	9% de la recaptació

Taula 5.1.: Per a cada nombre encert tenim el premi obtingut en euros.

En cas que cap dels participants encerti la categoria especial, s'acumula pel pot de la següent setmana en aquesta categoria.

Quan no hi ha encerts en la 1^a categoria, el seu premi passa a ser pot per a la setmana següent.

Si no hi ha encerts en la 2^a categoria, el seu premi s'acumula a la següent categoria (3^a), si tampoc n'hi ha en aquesta, el premi se suma a la categoria d'11 encerts, i si aquí tampoc hi ha encerts, s'acumula a l'última categoria. Si en la 5^a categoria tampoc hi hagués cap encertant, el total acumulat se sumaria al pot per a la setmana següent.

En cap cas un premi de determinada categoria pot ser superior a un premi d'una categoria superior, si això succeeix, se sumarà el premi d'ambdues categories i es reparteix a parts iguals entre els encertants d'ambdós.

Quan en una categoria hi ha més d'un encertant, el premi es reparteix a parts iguals entre ells.

En aquesta última temporada 2017/18 de La Primera Divisió d'Espanya s'han destinat una mitjana 2.285.382€ en premis per a cada una de les 38 jornades disputades.

BIBLIOGRAFIA

Hand, D., Till, R. (2001). *A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems* (Imperial College London). Consultat des de <https://link.springer.com/content/pdf/10.1023%2FA%3A1010920819831.pdf>

Moons, K. , Altman,G.,... (2015). *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration*. Consultat des de <http://annals.org/aim/fullarticle/2088542/transparent-reporting-multivariable-prediction-model-individual-prognosis-diagnosis-tripod-explanation>)

Wasserstein, R. (2016). *The ASA's Statement on p-Values: Context, Process, and Purpose*. Consultat des de <https://amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108#.WzV8cy0zmRt>

ANNEXOS

1. Descripció univariant

Basic descriptive statistics

<i>var</i>	<i>type</i>	<i>label</i>	<i>n</i>	<i>NA.prc</i>	<i>mean</i>	<i>sd</i>	<i>se</i>	<i>md</i>	<i>trimmed</i>	<i>range</i>	<i>skew</i>
GolesLocal	integer	GolesLocal	5800	0	1.59	1.32	0.02	1	1.45	10 (0-10)	0.94
GolesVisitante	integer	GolesVisitante	5800	0	1.12	1.12	0.01	1	0.97	8 (0-8)	1.17
posL	numeric	posL	5800	0	10.88	5.78	0.08	11	10.94	19 (1-20)	-0.07
ptsL	numeric	ptsL	5800	0	31.41	15.43	0.2	29	30.16	96 (1-97)	0.81
pjL	numeric	pjL	5800	0	23	8.37	0.11	23	23	28 (9-37)	0
pgL	numeric	pgL	5800	0	8.53	4.87	0.06	8	8.04	31 (0-31)	1.08
peL	numeric	peL	5800	0	5.81	3.04	0.04	6	5.64	18 (0-18)	0.48
ppL	numeric	ppL	5800	0	8.66	4.57	0.06	8	8.41	28 (0-28)	0.48
gfL	numeric	gfL	5800	0	30.55	15.89	0.21	28	28.93	115 (2-117)	1.27
gcL	numeric	gcL	5800	0	30.8	13.83	0.18	29	30.08	85 (3-88)	0.45
div2L	numeric	div2L	5800	0	0.84	0.36	0	1	0.93	1 (0-1)	-1.88
posV	numeric	posV	5800	0	10.7	5.76	0.08	11	10.73	19 (1-20)	-0.03
ptsV	numeric	ptsV	5800	0	31.77	15.48	0.2	30	30.52	94 (1-95)	0.8
pjV	numeric	pjV	5800	0	23	8.37	0.11	23	23	28 (9-37)	0
pgV	numeric	pgV	5800	0	8.65	4.89	0.06	8	8.16	31 (0-31)	1.05
peV	numeric	peV	5800	0	5.82	3.05	0.04	6	5.66	18 (0-18)	0.47
ppV	numeric	ppV	5800	0	8.53	4.52	0.06	8	8.29	28 (0-28)	0.48
gfV	numeric	gfV	5800	0	30.78	15.93	0.21	28	29.14	113 (2-115)	1.25
gcV	numeric	gcV	5800	0	30.53	13.68	0.18	29	29.81	86 (3-89)	0.45
div2V	numeric	div2V	5800	0	0.84	0.36	0	1	0.93	1 (0-1)	-1.88
pos1	numeric	pos1	5800	0	0.18	8.21	0.11	1	0.2	38	-0.02

										(-19-19)	
pts1	numeric	pts1	5800	0	-0.36	14.71	0.19	0	-0.38	124 (-64-60)	0.04
pj1	numeric	pj1	5800	0	0	0	0	0	0	0 (0-0)	NaN
pg1	numeric	pg1	5800	0	-0.12	5.32	0.07	0	-0.11	47 (-24-23)	0.04
pe1	numeric	pe1	5800	0	-0.01	3.1	0.04	0	-0.02	25 (-12-13)	0.06
pp1	numeric	pp1	5800	0	0.13	4.67	0.06	0	0.14	38 (-18-20)	-0.02
gf1	numeric	gf1	5800	0	-0.22	15.97	0.21	0	-0.23	162 (-81-81)	0.05
gc1	numeric	gc1	5800	0	0.28	10.92	0.14	0	0.24	103 (-53-50)	0.01
div21	numeric	div21	5800	0	0	0.53	0.01	0	0	2 (-1-1)	0
ptsL	numeric	ptsL	5800	0	19.04	9.46	0.12	18	18.48	52 (0-52)	0.52
ptsvV	numeric	ptsvV	5800	0	11.85	7.11	0.09	11	11.16	47 (0-47)	1.06
guanyadesL	numeric	guanyadesL	5800	0	0.55	1.25	0.02	0	0.27	15 (0-15)	4.12
guanyadesV	numeric	guanyadesV	5800	0	0.8	1.31	0.02	0	0.53	16 (0-16)	3.56
ptsLV	numeric	ptsLV	5800	0	7.18	8.8	0.12	7	6.93	73 (-30-43)	0.27
guanyadesLV1	numeric	guanyadesLV1	5800	0	-0.25	1.8	0.02	0	-0.25	31 (-16-15)	-0.01
pos2	numeric	pos2	5800	0	0.01	0.43	0.01	0.03	0.01	3.8 (-1.9-1.9)	0.01
pts2	numeric	pts2	5800	0	-0.02	0.6	0.01	0	-0.02	3.93 (-1.93-2)	0.02
pj2	numeric	pj2	5800	0	0	0	0	0	0	0 (0-0)	NaN
pg2	numeric	pg2	5800	0	-0.01	0.22	0	0	-0.01	1.41	0.01

										(-0.69- 0.72)	
pe2	numeric	pe2	5800	0	0	0.14	0	0	0	1.1 (-0.6- 0.5)	0.02
pp2	numeric	pp2	5800	0	0.01	0.19	0	0	0.01	1.4 (-0.77- 0.63)	-0.03
gf2	numeric	gf2	5800	0	-0.01	0.64	0.01	0	-0.01	5.58 (-2.8- 2.78)	-0.04
gc2	numeric	gc2	5800	0	0.01	0.45	0.01	0	0.01	3.48 (-1.77- 1.71)	0
div22	numeric	div22	5800	0	0	0.03	0	0	0	0.2 (-0.1- 0.1)	0.02
ptsLV2	numeric	ptsLV2	5800	0	0.29	0.35	0	0.31	0.3	2.27 (-1- 1.27)	-0.2
guanyadesLV2	numeric	guanyadesLV2	5800	0	-0.01	0.09	0	0	-0.01	1.55 (-0.8- 0.75)	-0.17
guanyatL	numeric	guanyatL	5800	0	0.49	0.5	0.01	0	0.48	1 (0-1)	0.05
empetat	numeric	empetat	5800	0	0.25	0.43	0.01	0	0.18	1 (0-1)	1.18
guanyatV	numeric	guanyatV	5800	0	0.27	0.44	0.01	0	0.21	1 (0-1)	1.06
quiniela	numeric	quiniela	5800	0	1.02	0.72	0.01	1	1.02	2 (0-2)	-0.03
quiniela2	categorical	quiniela2	5800	0	1.76	0.82	0.01	2	1.7	2 (1-3)	0.47

quiniela2

<i>value</i>	<i>N</i>	<i>raw %</i>	<i>valid %</i>
1	2830	48.79	48.79
2	1543	26.60	26.60
X	1427	24.60	24.60
missings	0	0.00	

total N=5800 · valid N=5800 ·

Temporada

<i>value</i>	<i>N</i>	<i>raw %</i>	<i>valid %</i>
1997-98	290	5.00	5.00
1998-99	290	5.00	5.00
1999-00	290	5.00	5.00
2000-01	290	5.00	5.00
2001-02	290	5.00	5.00
2002-03	290	5.00	5.00
2003-04	290	5.00	5.00
2004-05	290	5.00	5.00
2005-06	290	5.00	5.00
2006-07	290	5.00	5.00
2007-08	290	5.00	5.00
2008-09	290	5.00	5.00
2009-10	290	5.00	5.00
2010-11	290	5.00	5.00
2011-12	290	5.00	5.00
2012-13	290	5.00	5.00
2013-14	290	5.00	5.00
2014-15	290	5.00	5.00
2015-16	290	5.00	5.00
2016-17	290	5.00	5.00
missings	0	0.00	

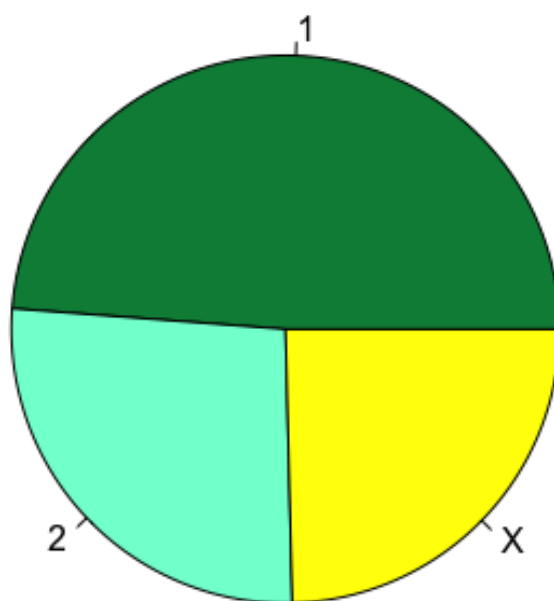
total N=5800 · valid N=5800 ·

Jornada

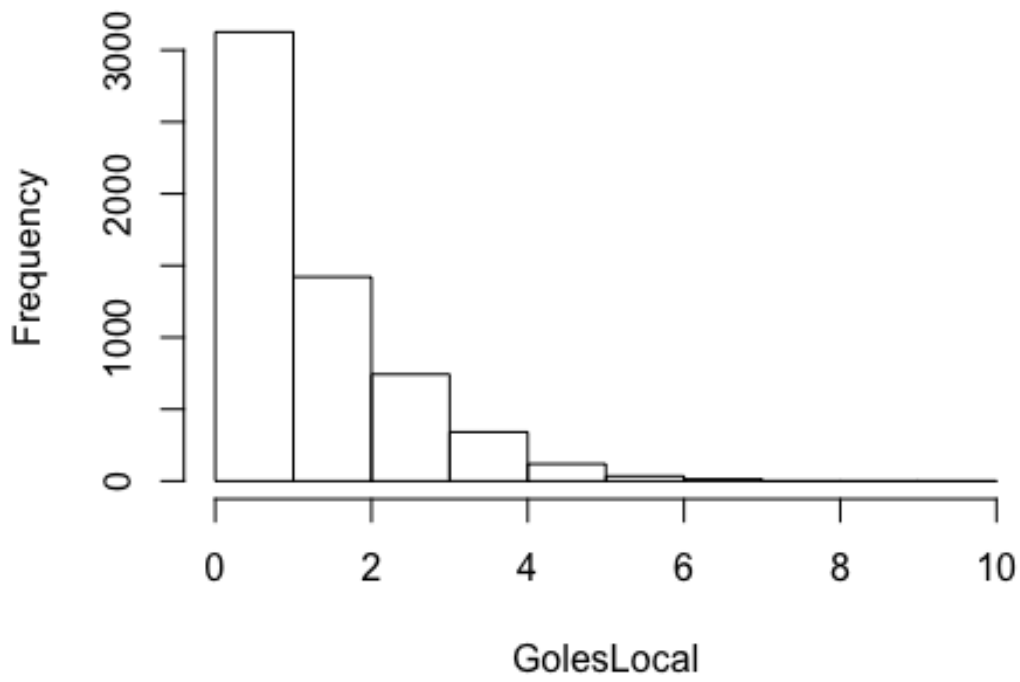
<i>value</i>	<i>N</i>	<i>raw %</i>	<i>valid %</i>
10	200	3.45	3.45
11	200	3.45	3.45
12	200	3.45	3.45
13	200	3.45	3.45
14	200	3.45	3.45
15	200	3.45	3.45
16	200	3.45	3.45
17	200	3.45	3.45
18	200	3.45	3.45
19	200	3.45	3.45
20	200	3.45	3.45
21	200	3.45	3.45
22	200	3.45	3.45
23	200	3.45	3.45
24	200	3.45	3.45
25	200	3.45	3.45
26	200	3.45	3.45
27	200	3.45	3.45
28	200	3.45	3.45
29	200	3.45	3.45
30	200	3.45	3.45
31	200	3.45	3.45
32	200	3.45	3.45
33	200	3.45	3.45
34	200	3.45	3.45
35	200	3.45	3.45
36	200	3.45	3.45
37	200	3.45	3.45
38	200	3.45	3.45
missings	0	0.00	

total N=5800 · valid N=5800

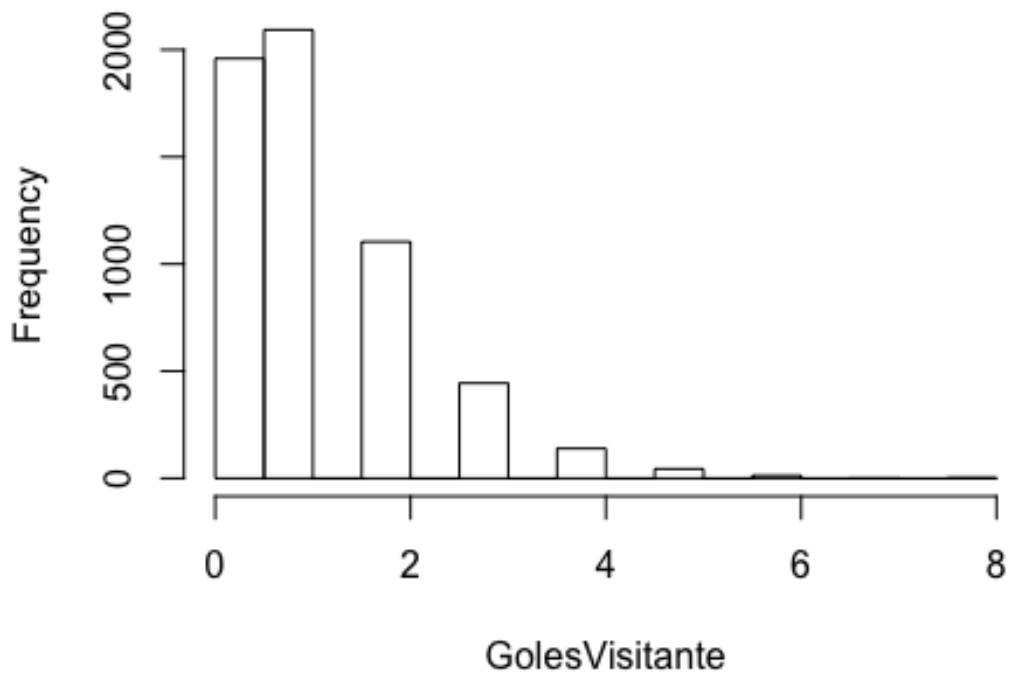
Resultats Quiniela



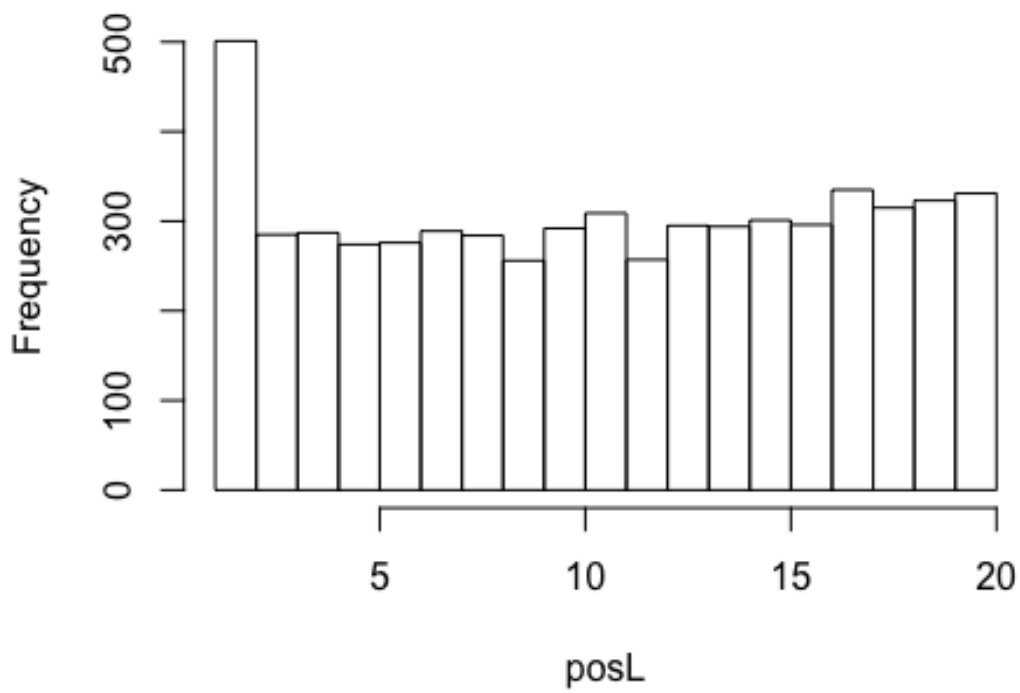
Histograma de GolesLocal



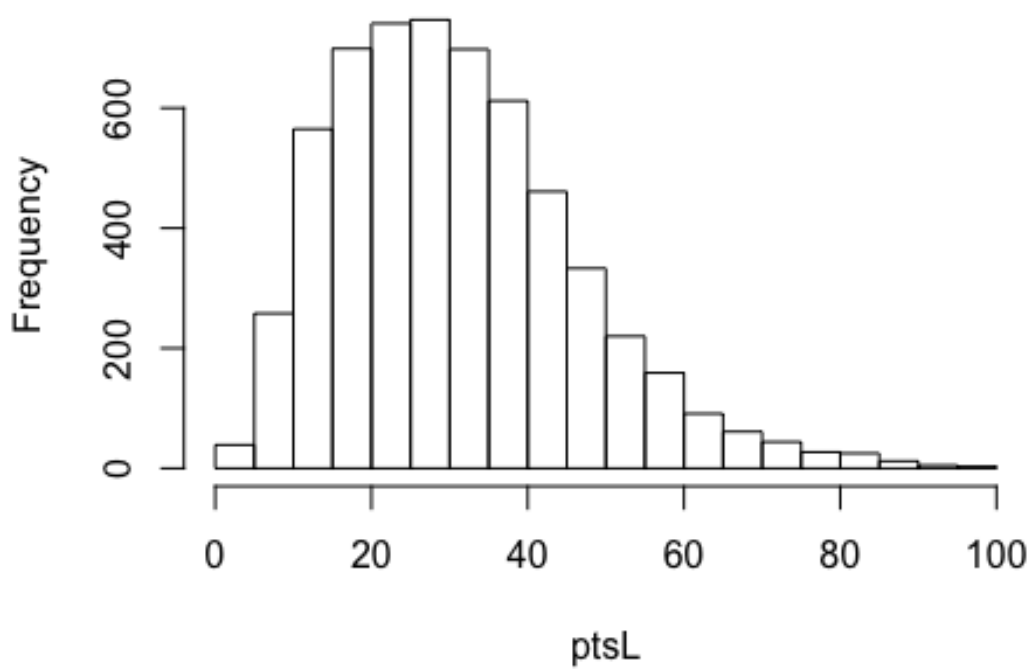
Histograma de GolesVisitante



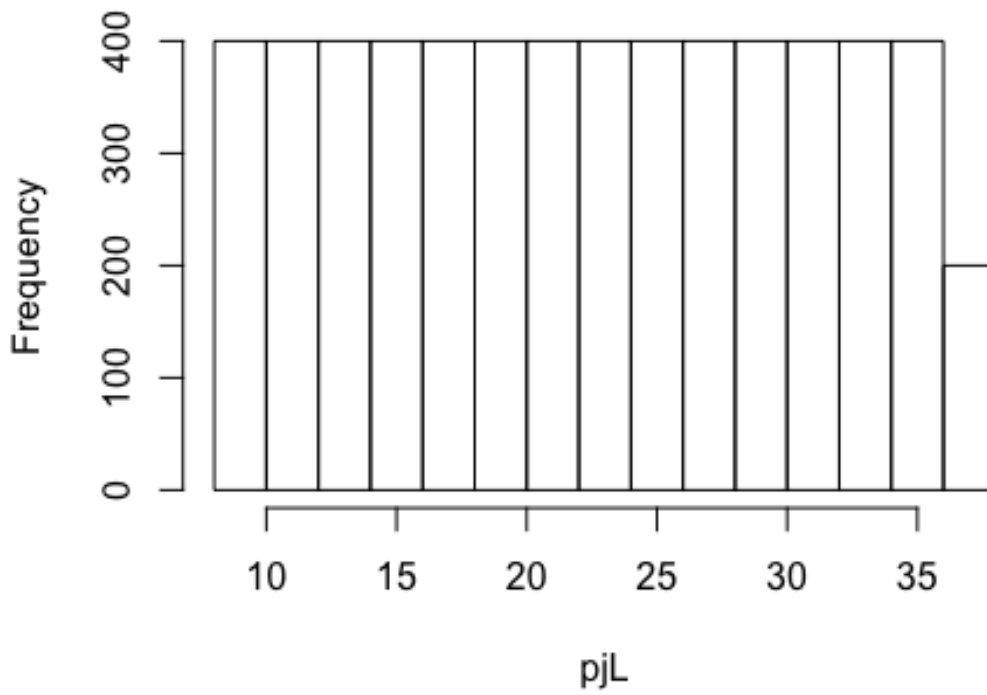
Histograma de posL



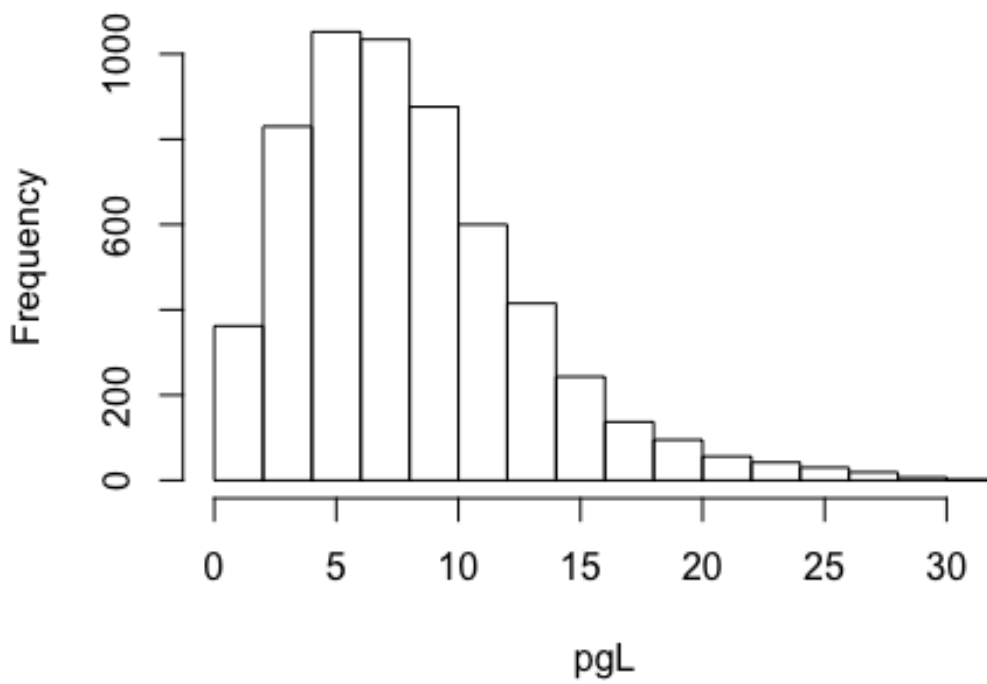
Histograma de ptsL



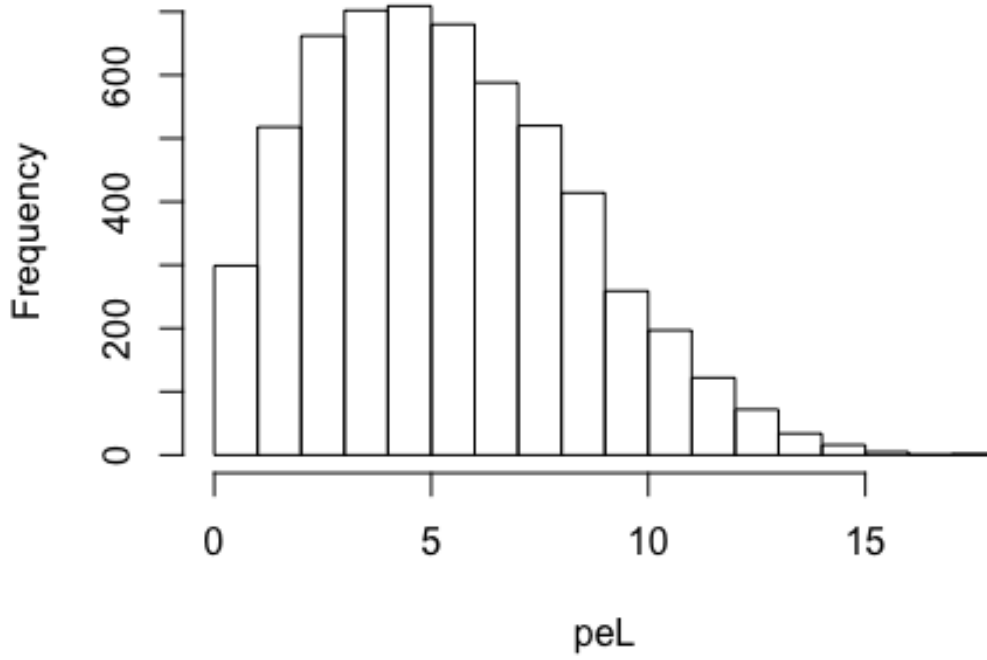
Histograma de pjL



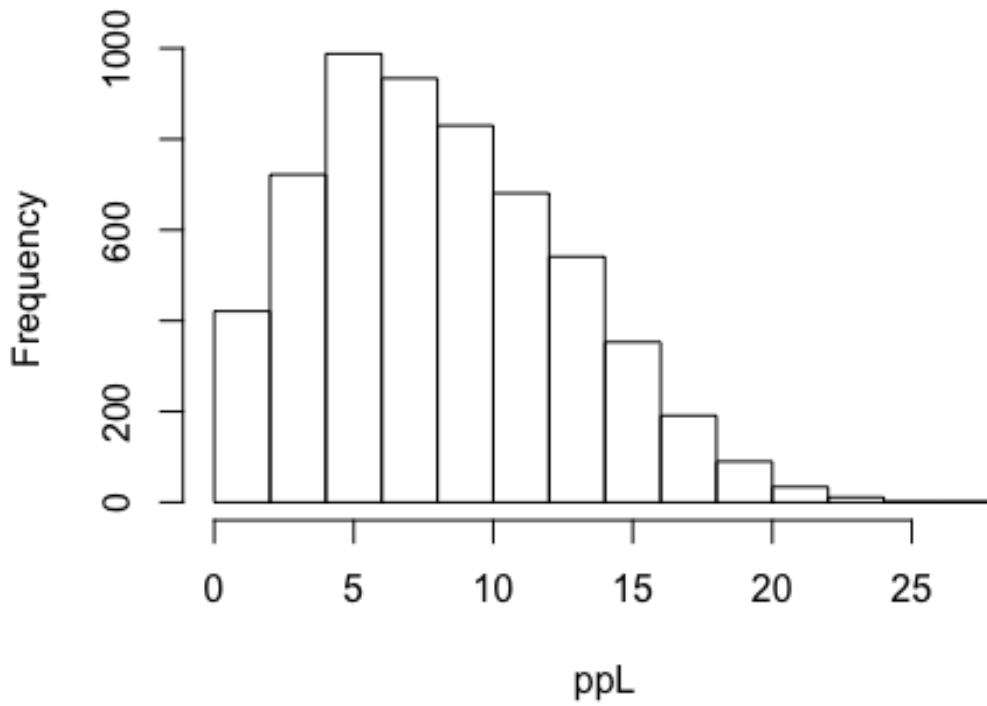
Histograma de pgL



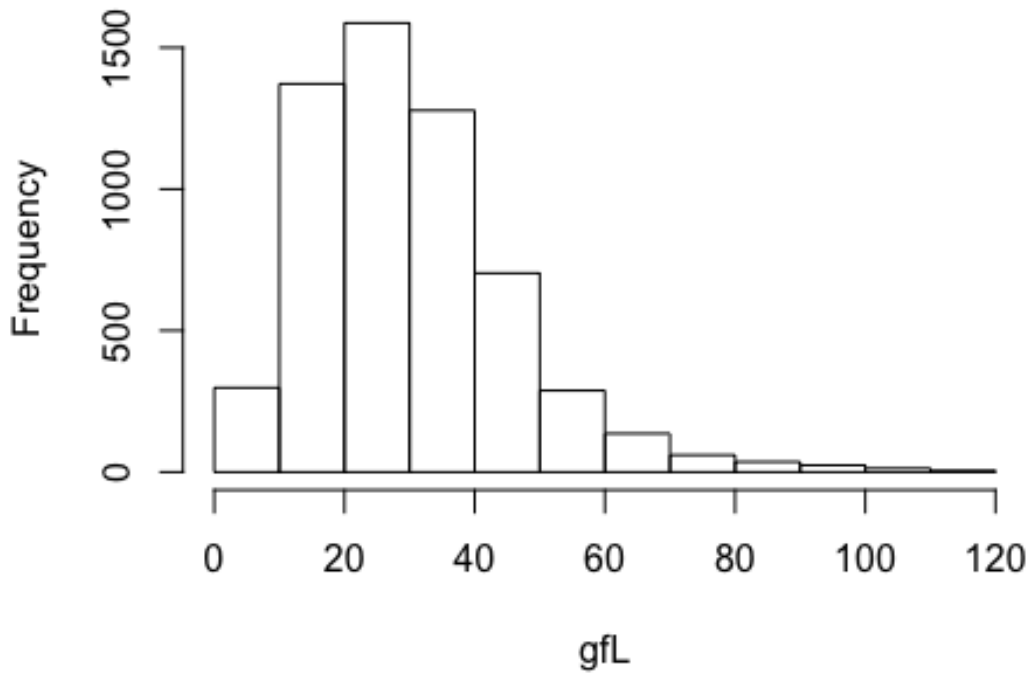
Histograma de peL



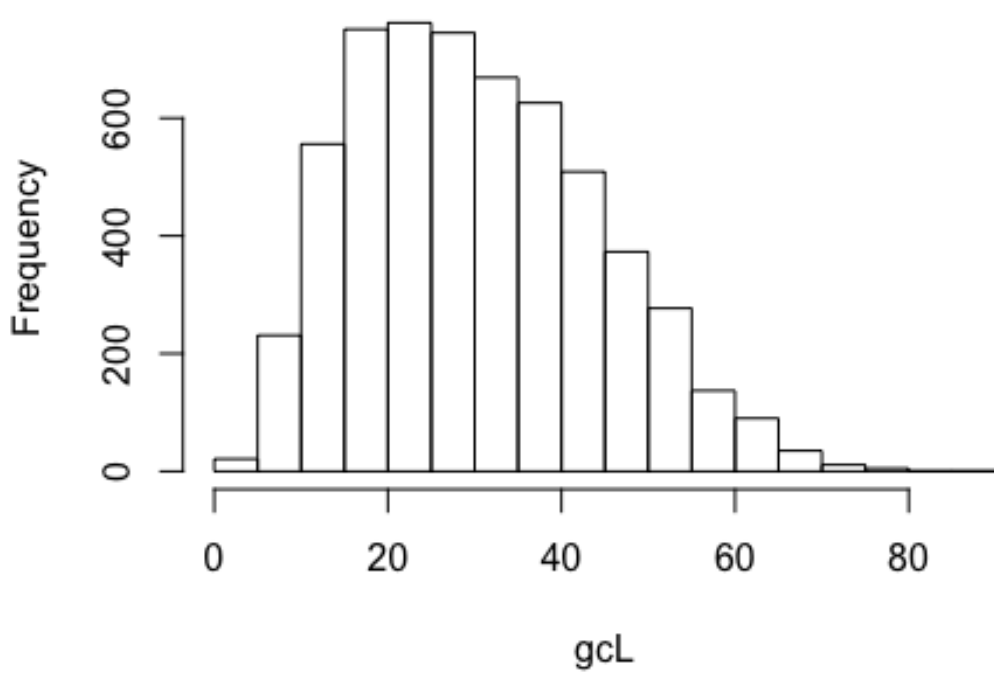
Histograma de ppL



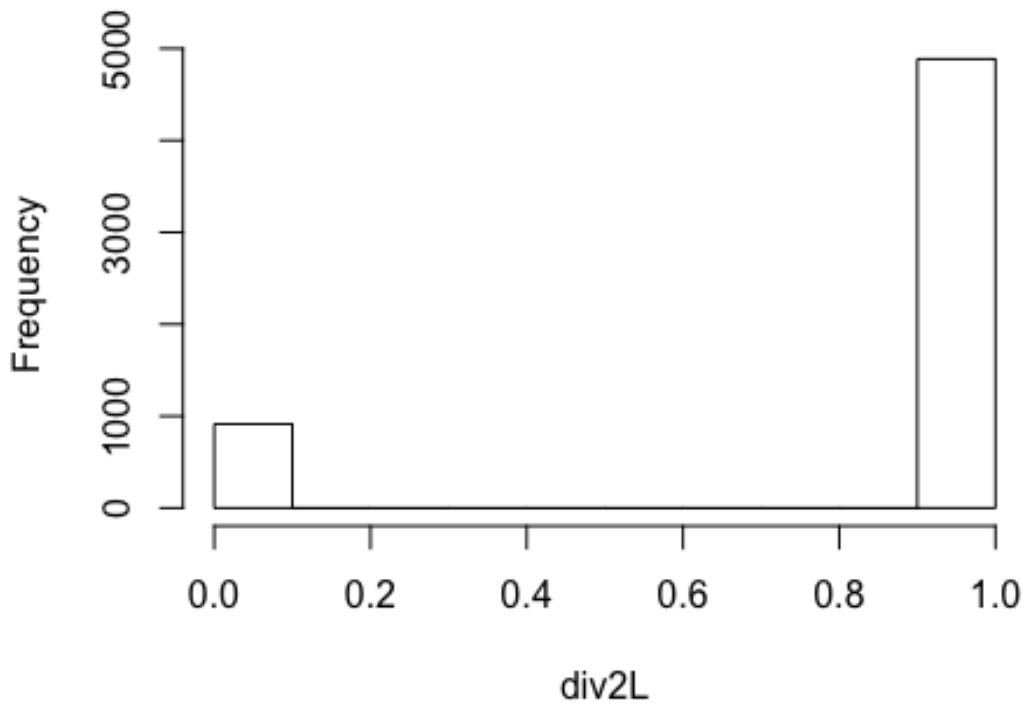
Histograma de gFL



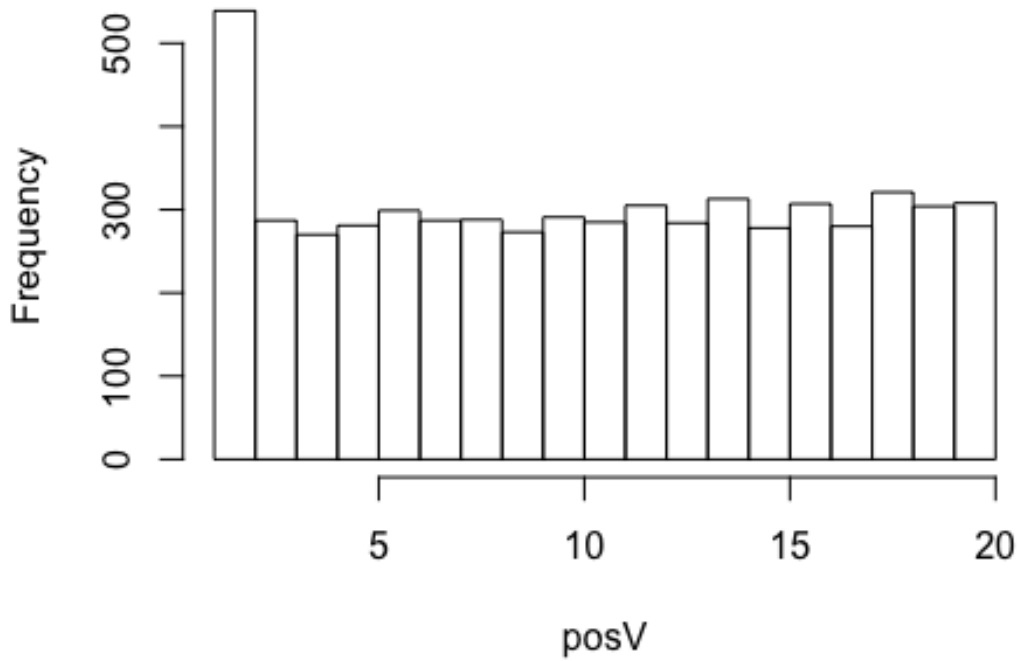
Histograma de gCL



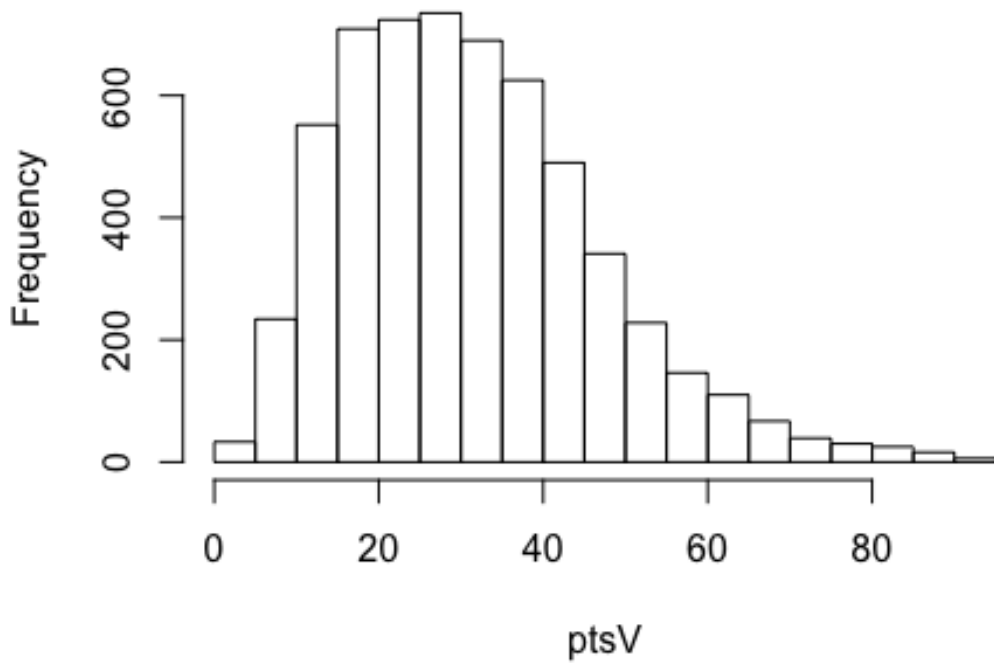
Histograma de div2L



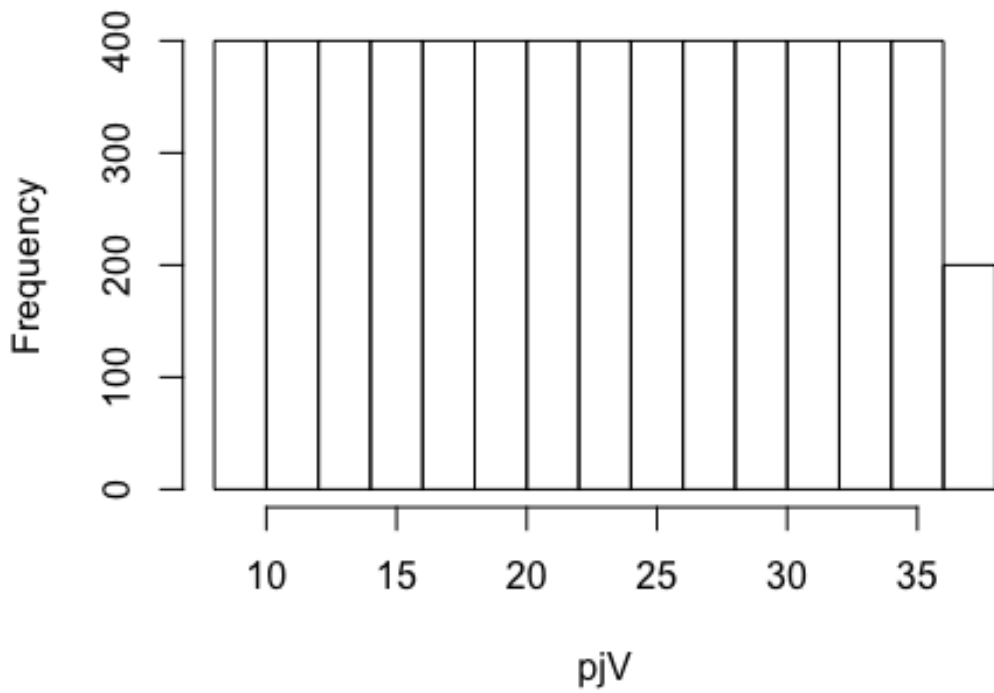
Histograma de posV



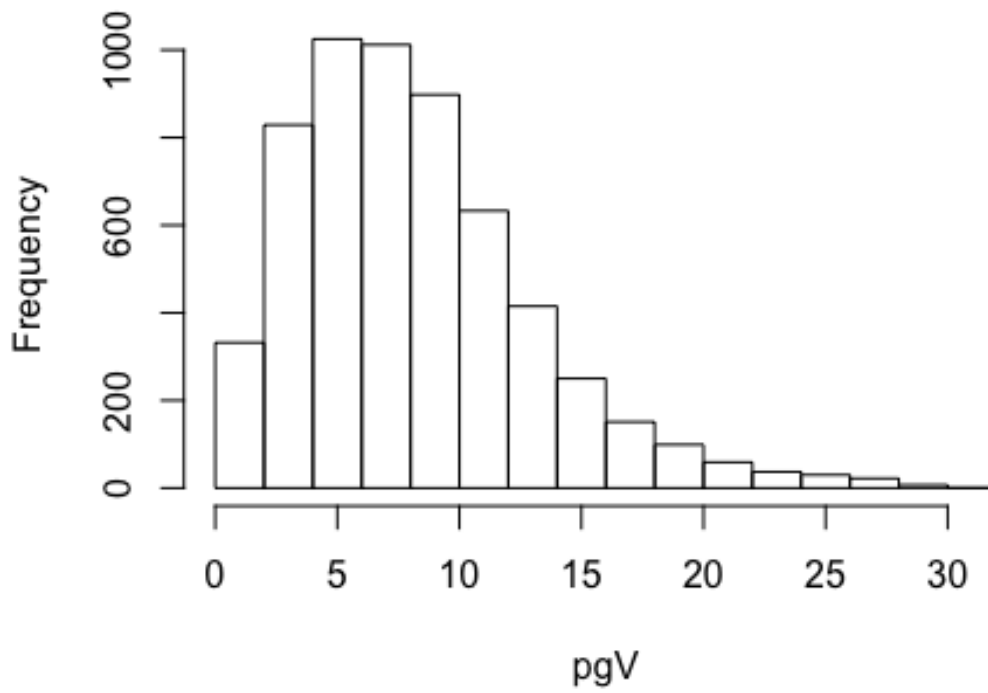
Histograma de ptsV



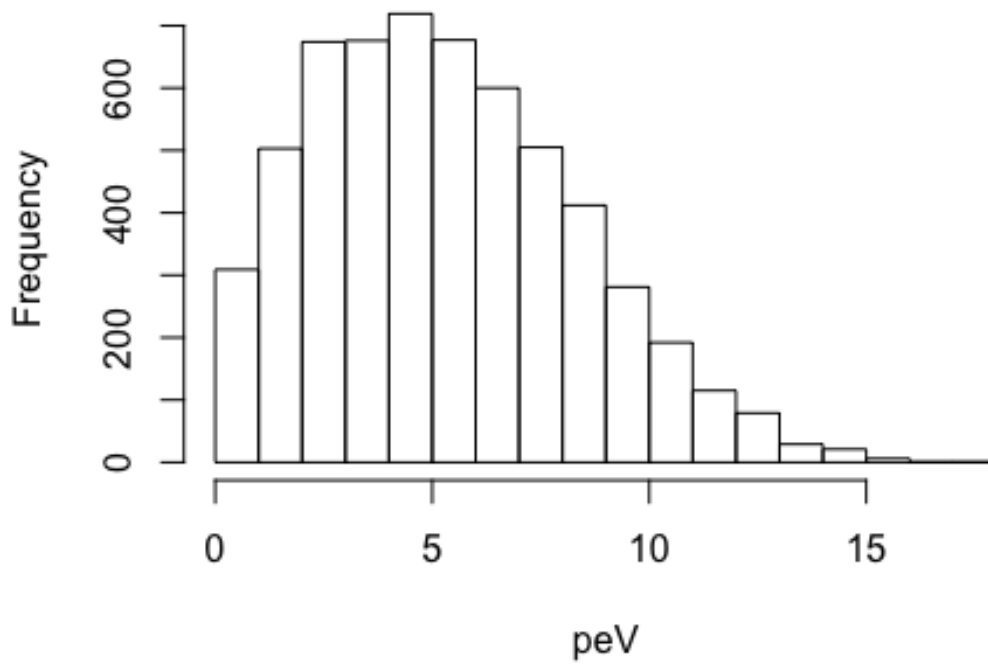
Histograma de pjV



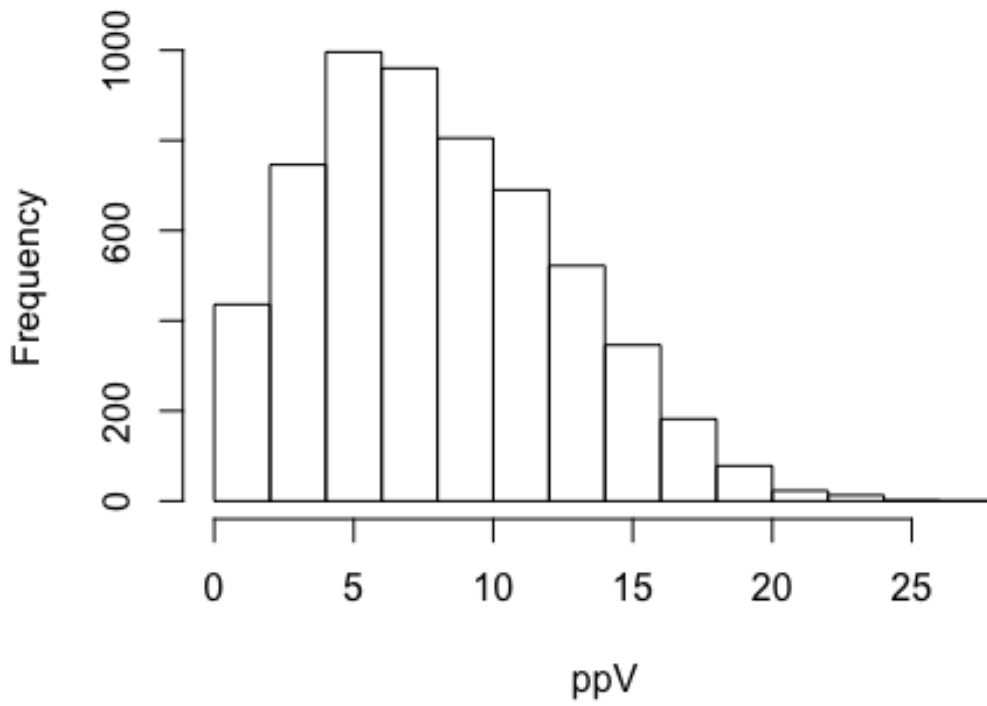
Histograma de pgV



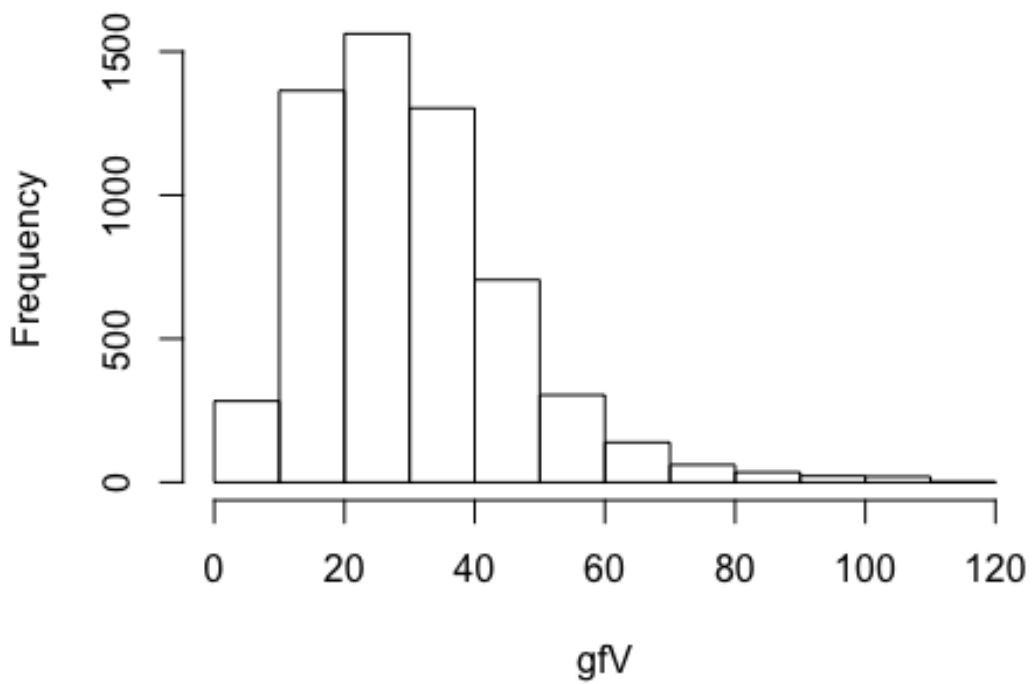
Histograma de peV



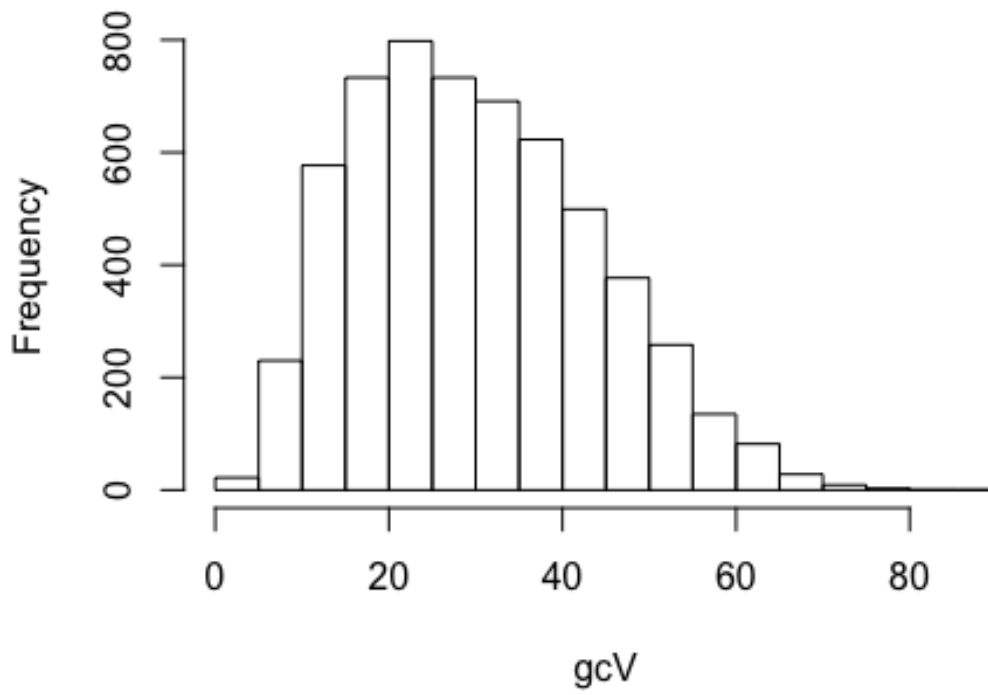
Histograma de ppV



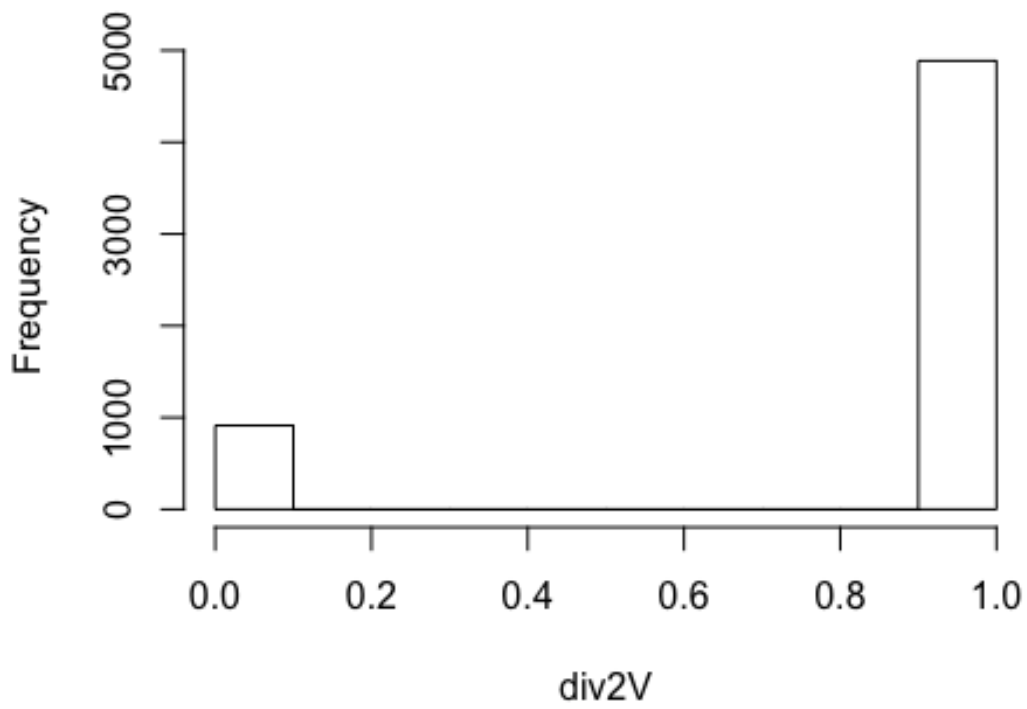
Histograma de gfV



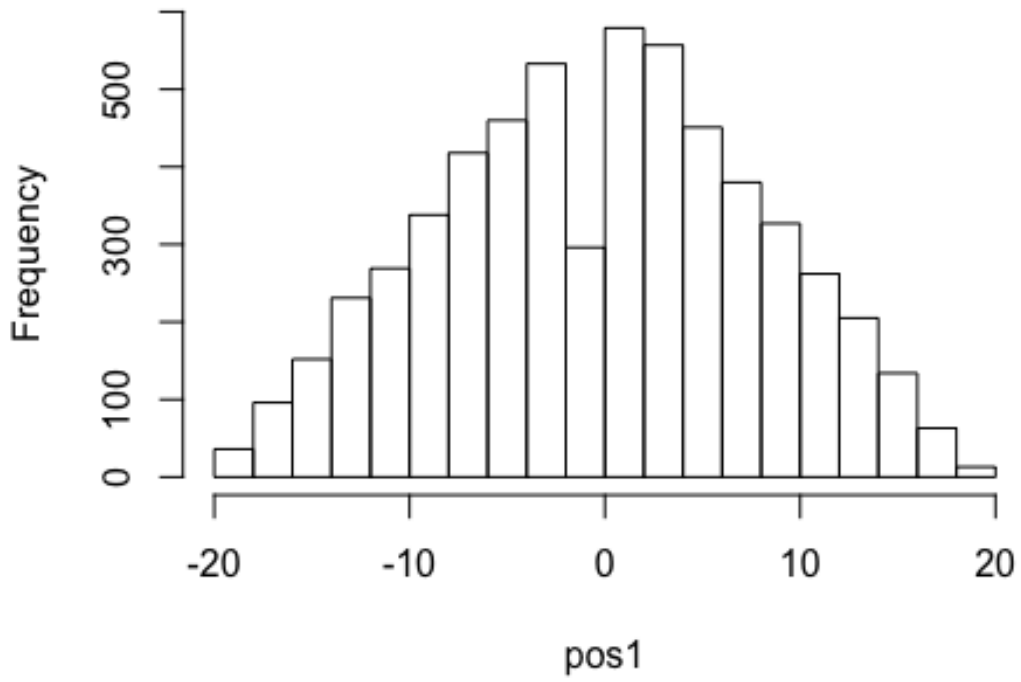
Histograma de gcV



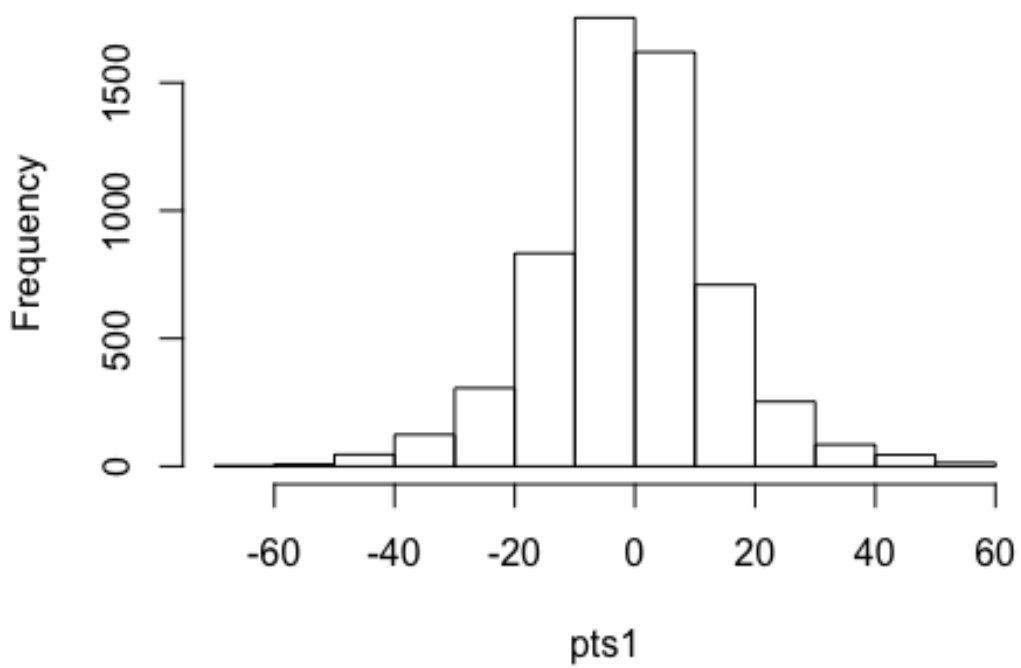
Histograma de div2V



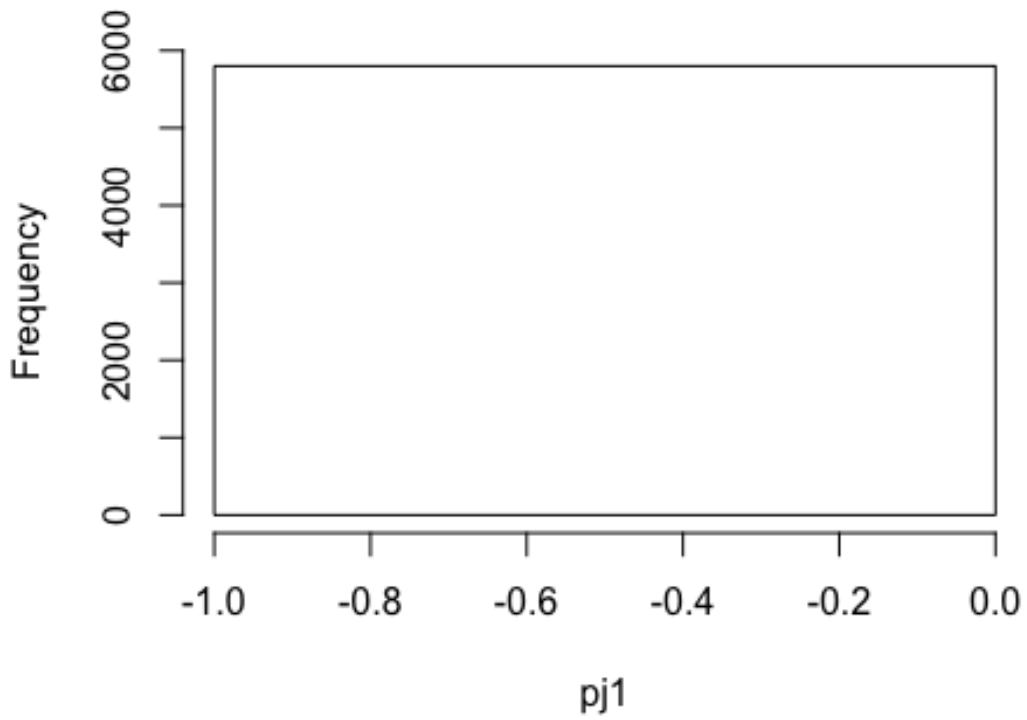
Histograma de pos1



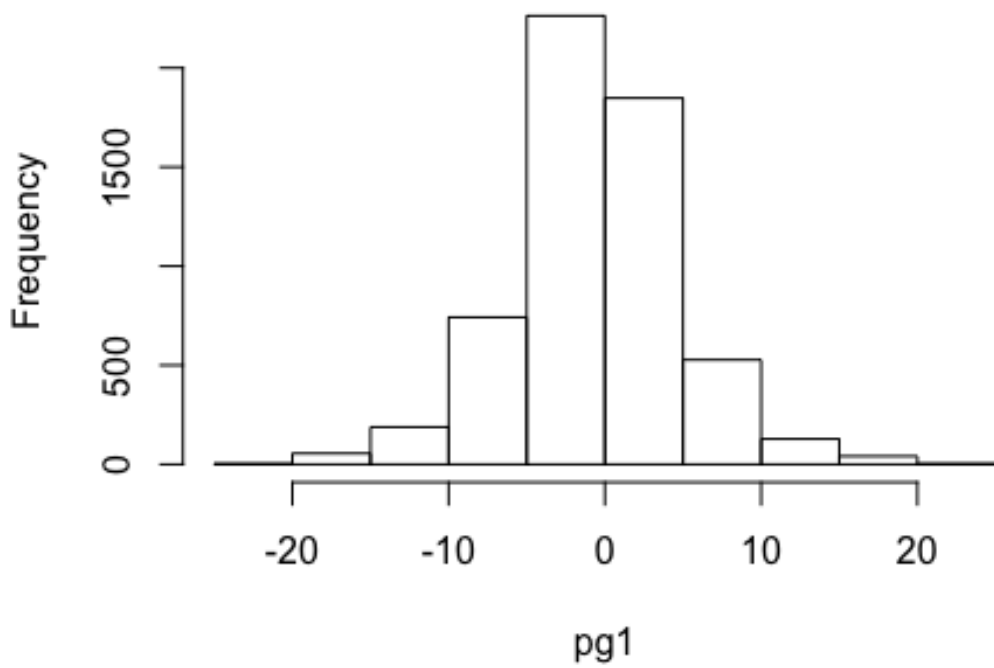
Histograma de pts1



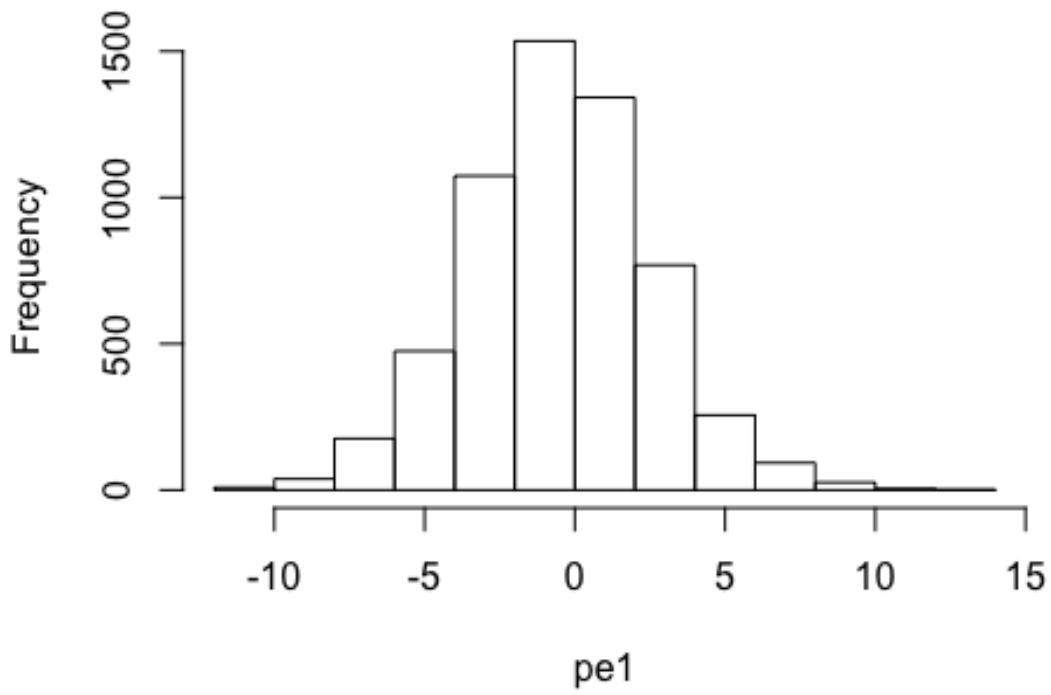
Histograma de pj1



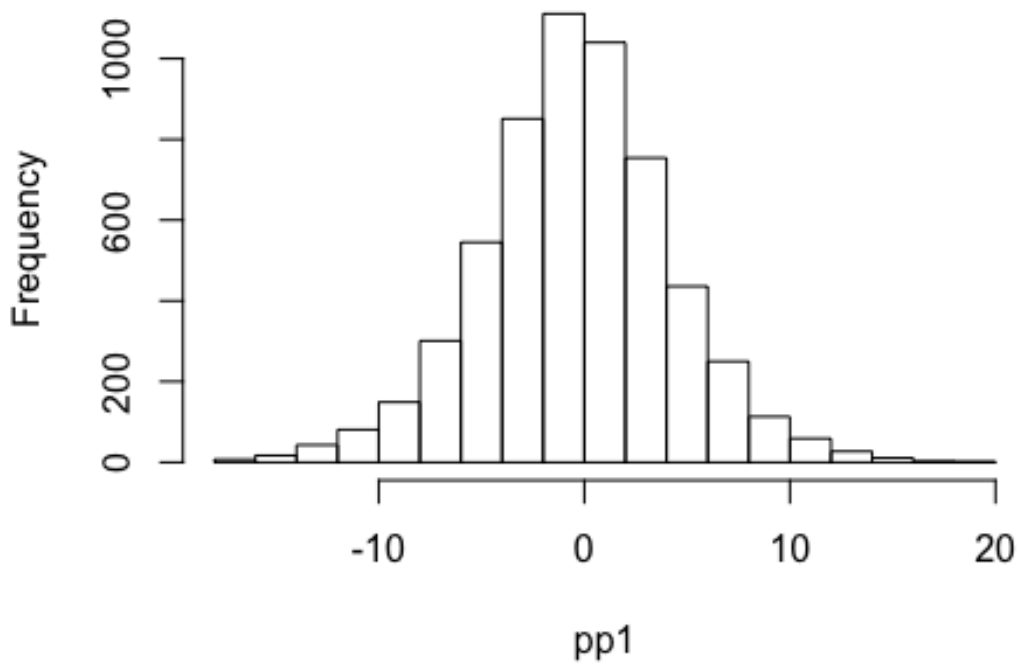
Histograma de pg1



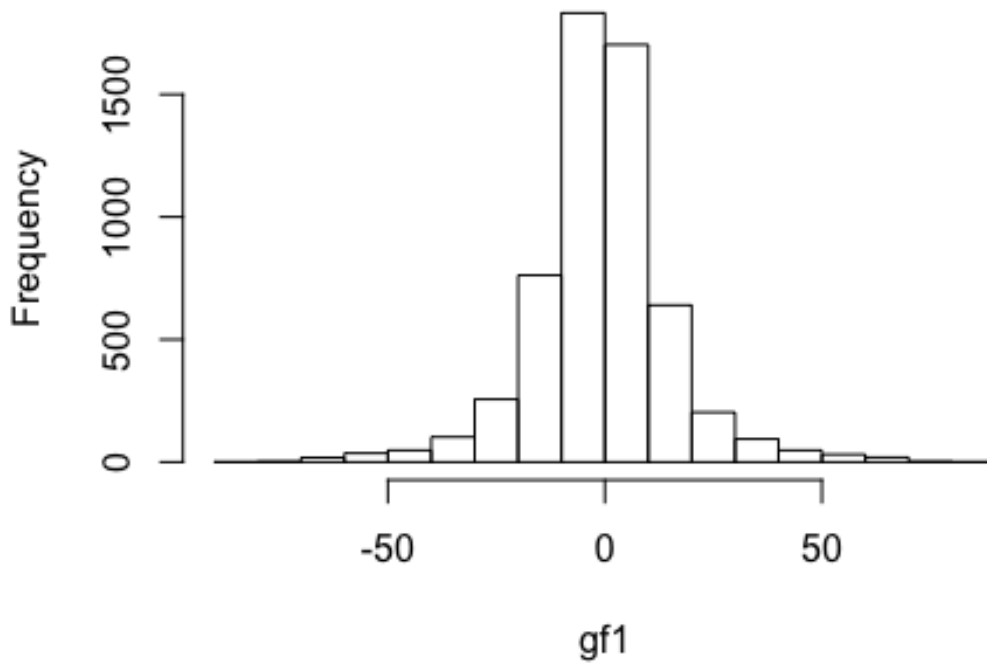
Histograma de pe1



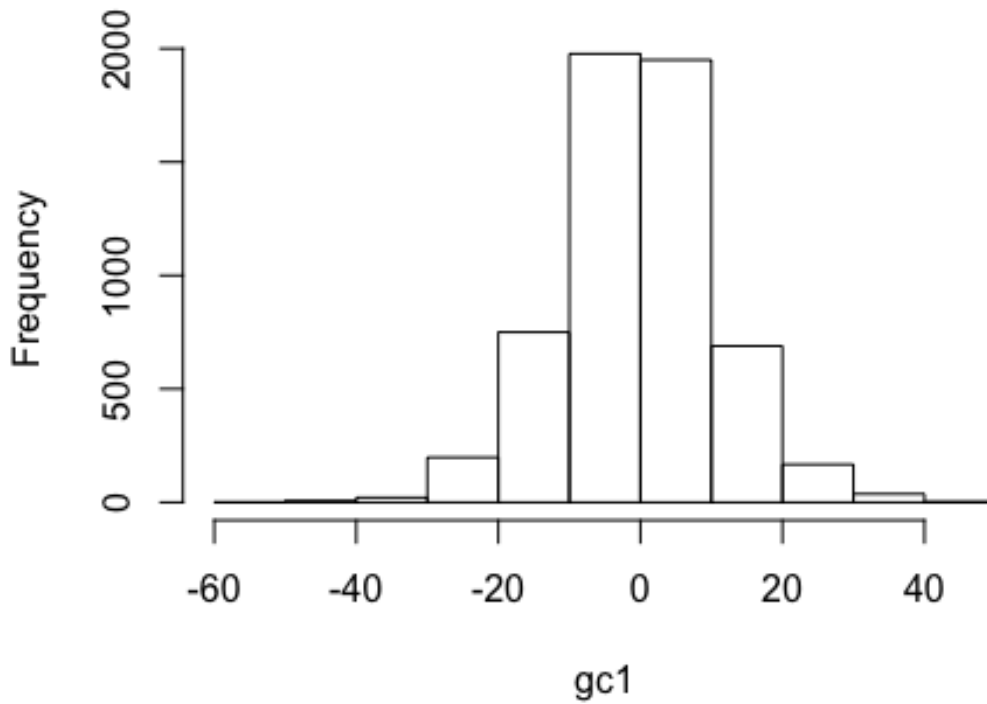
Histograma de pp1



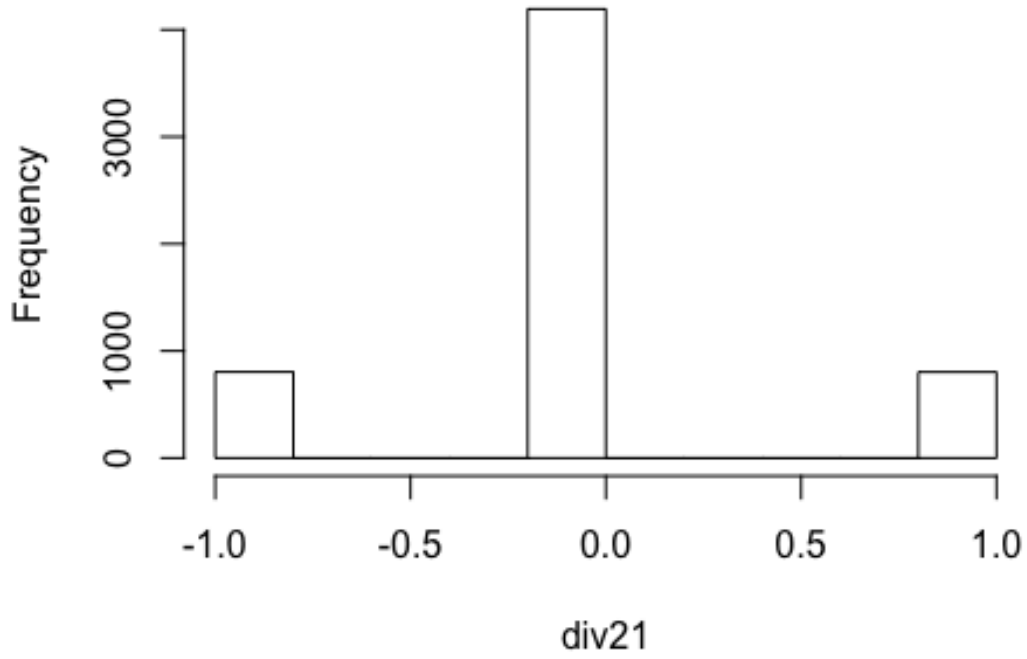
Histograma de gf1



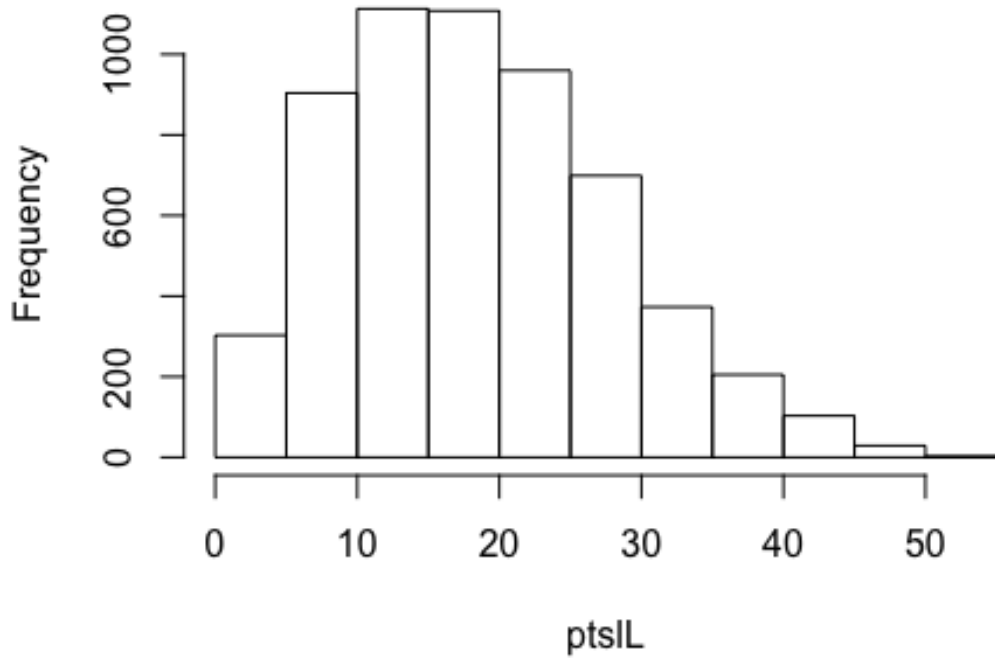
Histograma de gc1



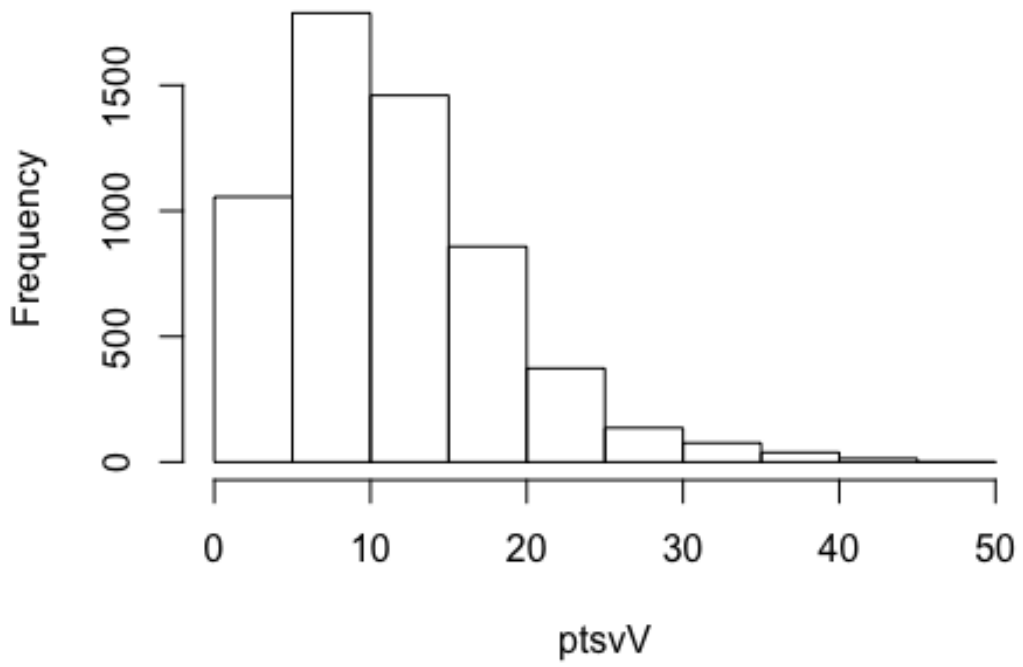
Histograma de div21



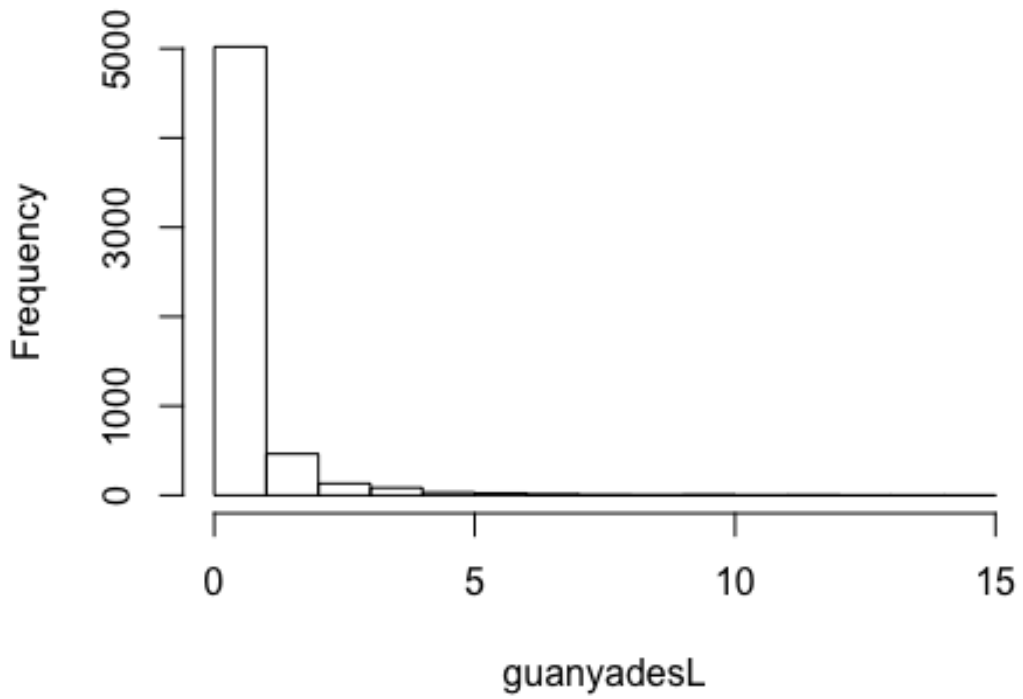
Histograma de ptsIL



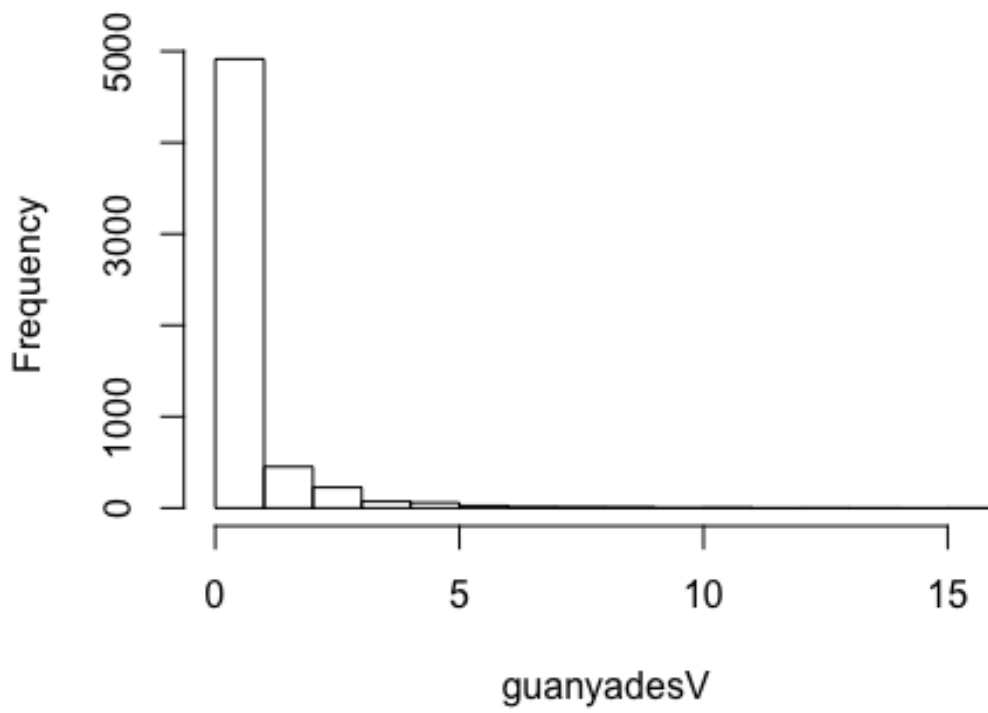
Histograma de ptsvV



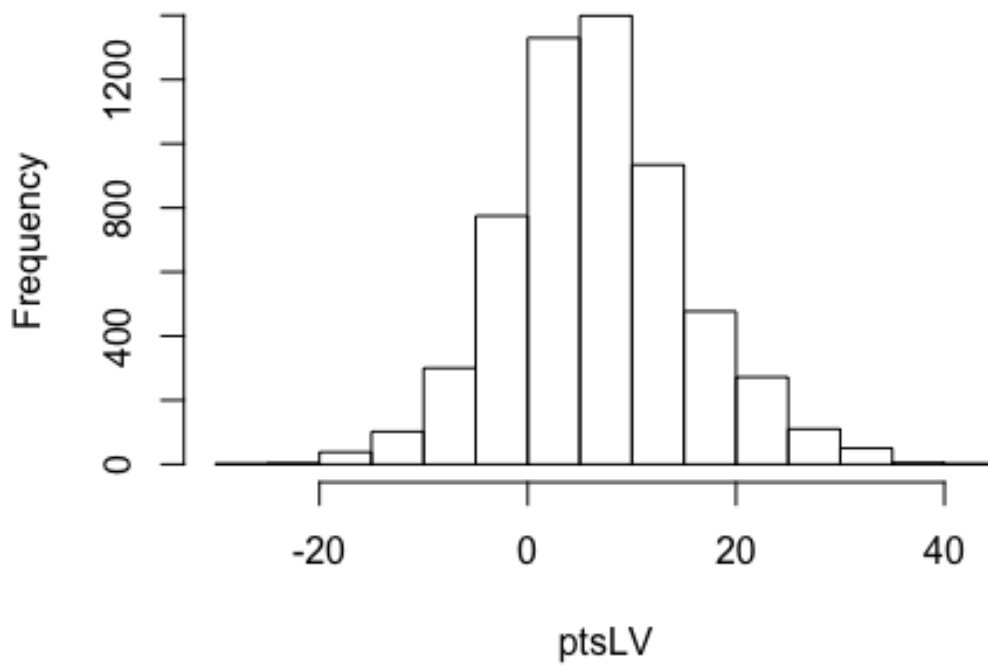
Histograma de guanyadesL



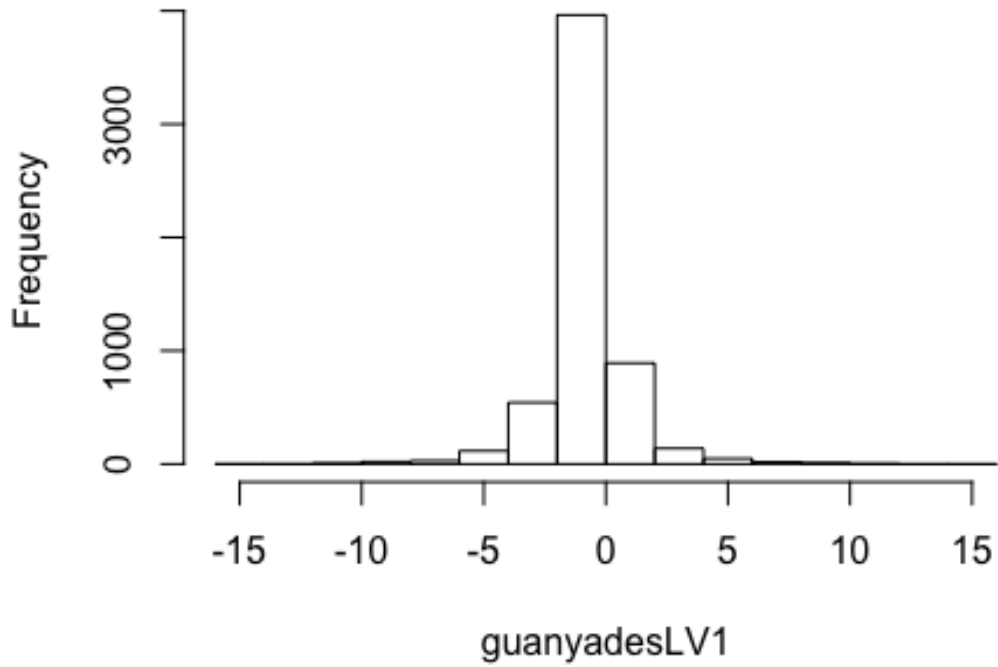
Histograma de guanyadesV



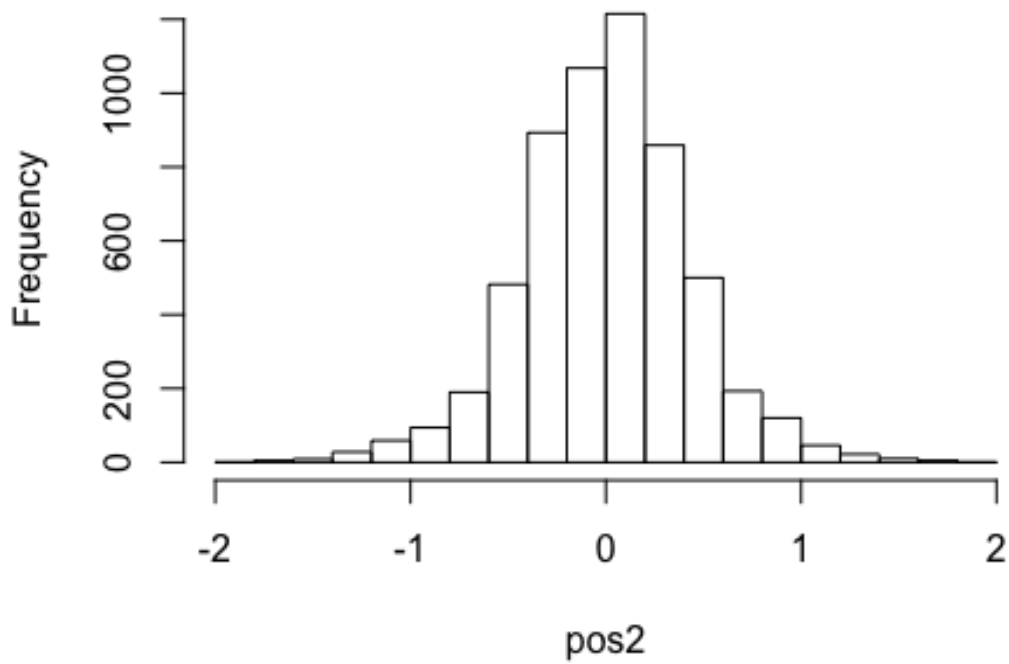
Histograma de ptsLV



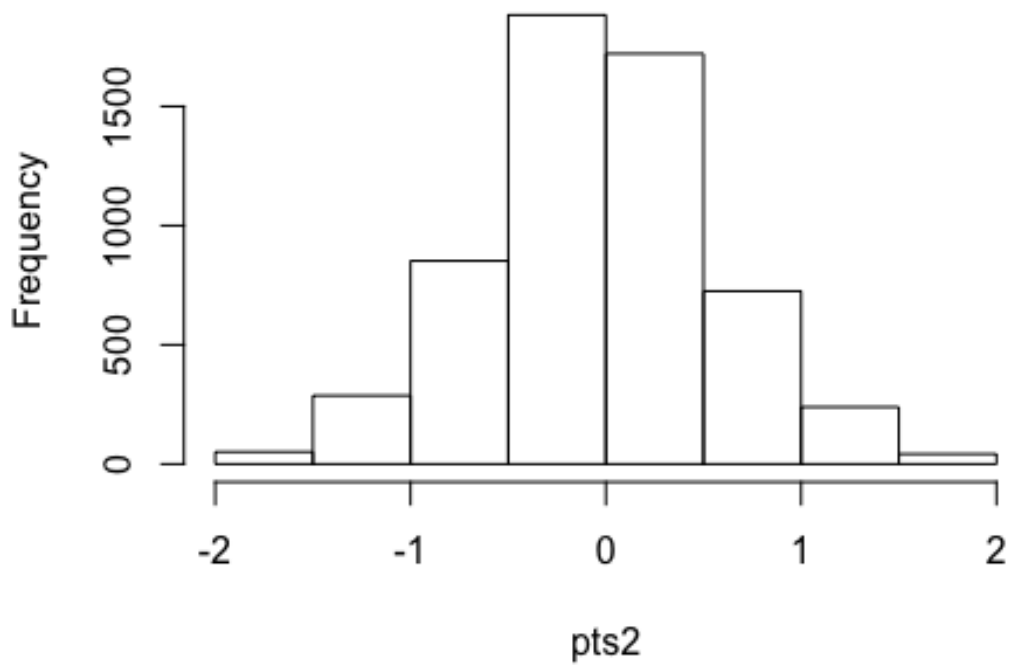
Histograma de guanyadesLV1



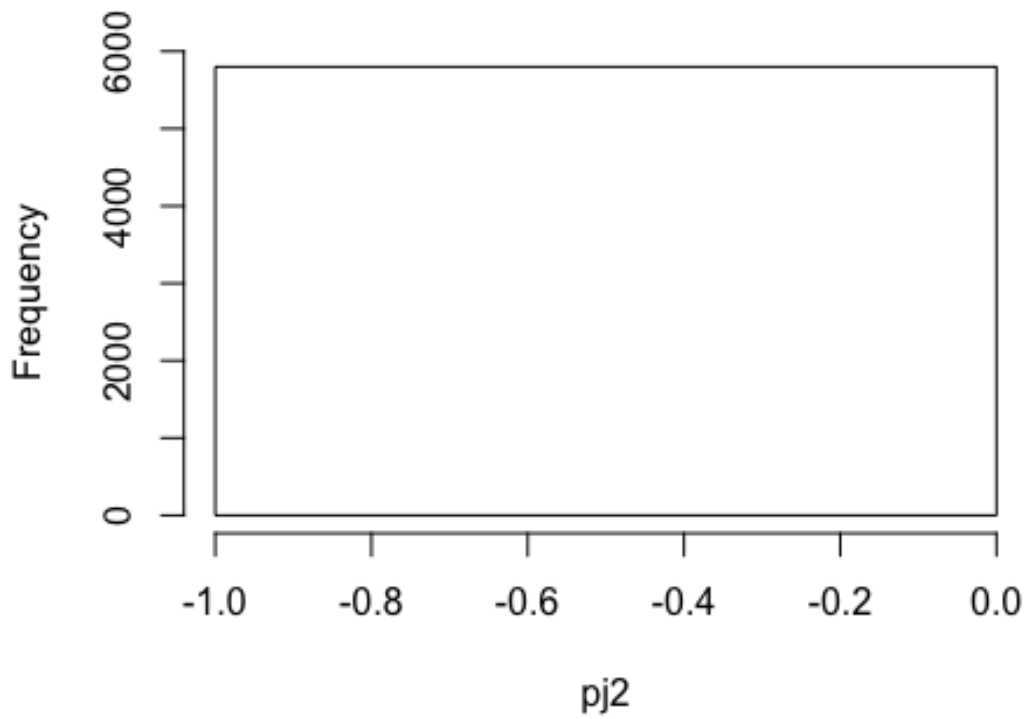
Histograma de pos2



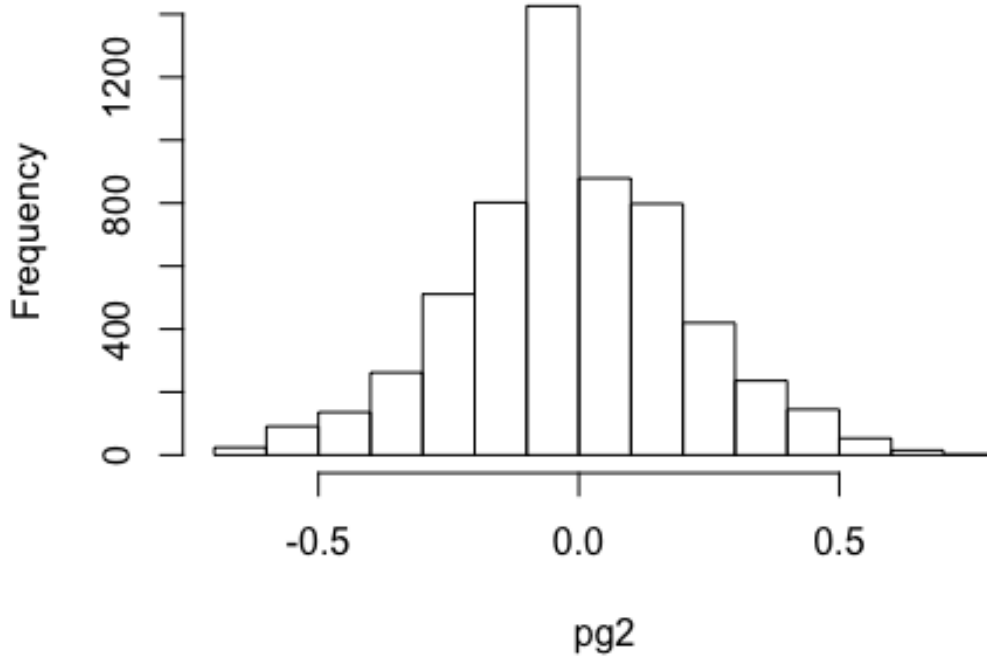
Histograma de pts2



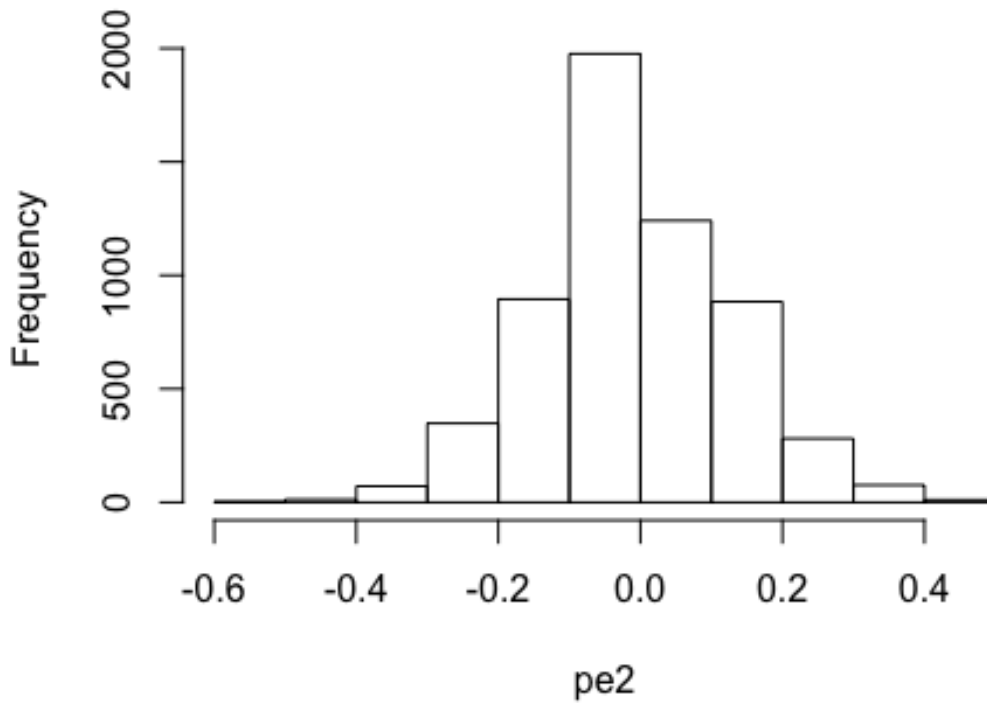
Histograma de pj2



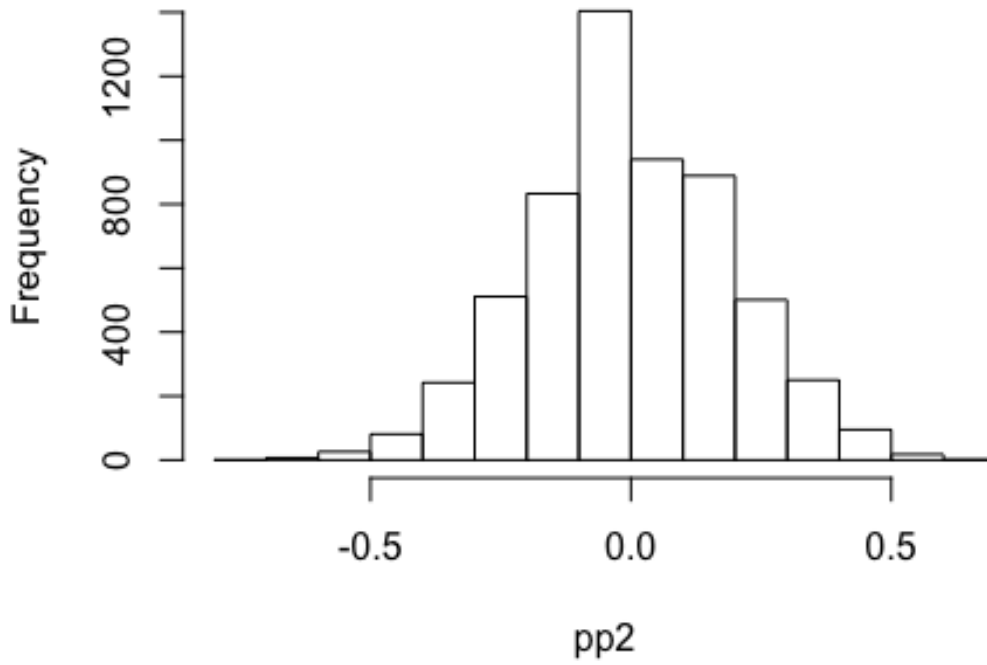
Histograma de pg2



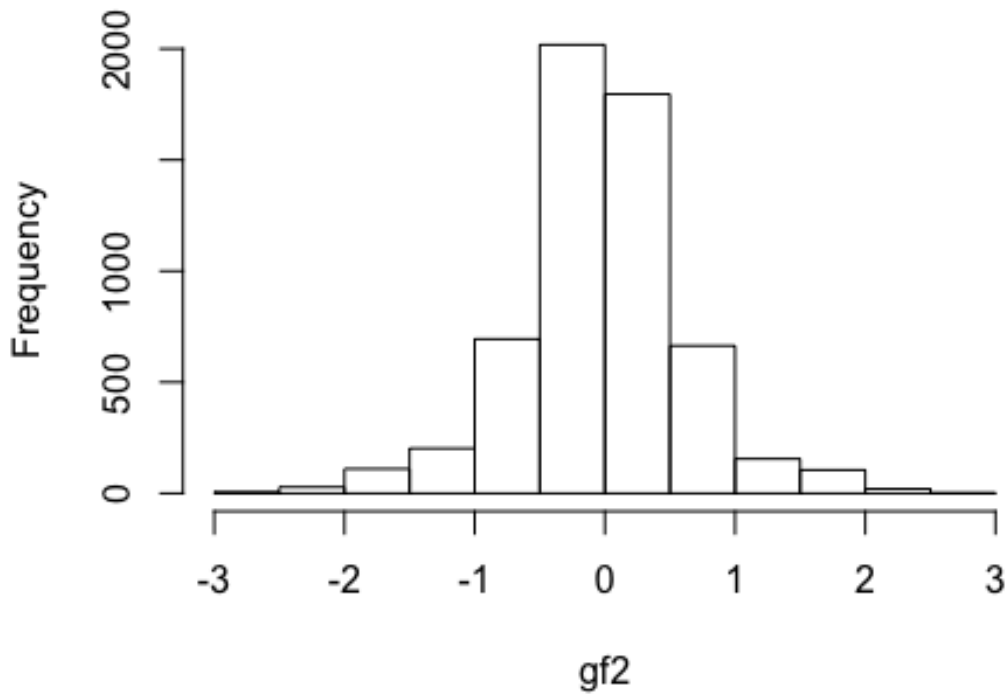
Histograma de pe2



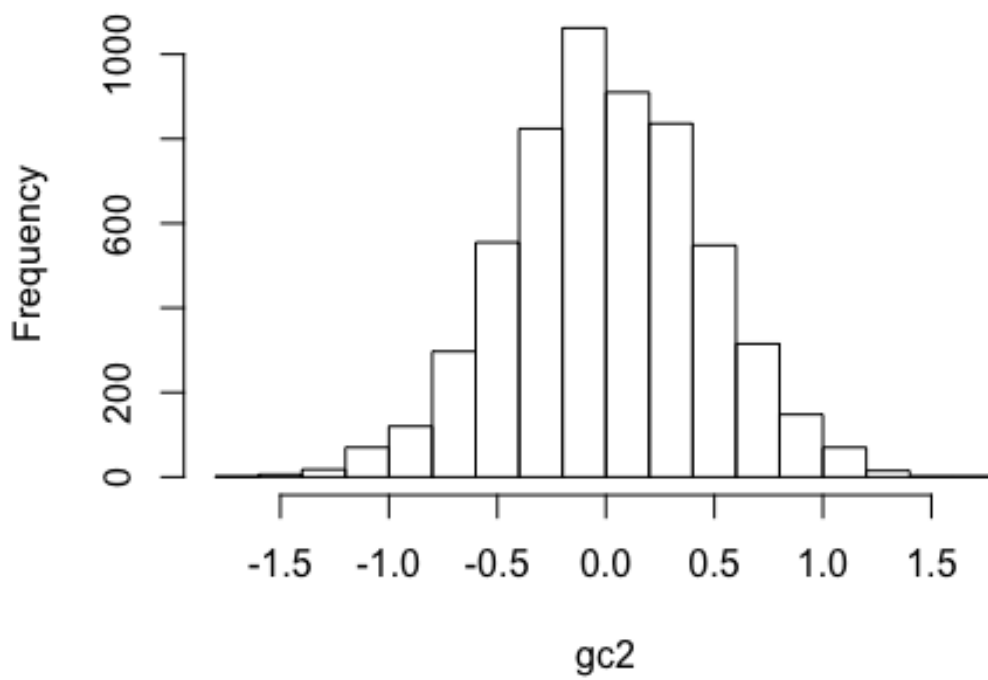
Histograma de pp2



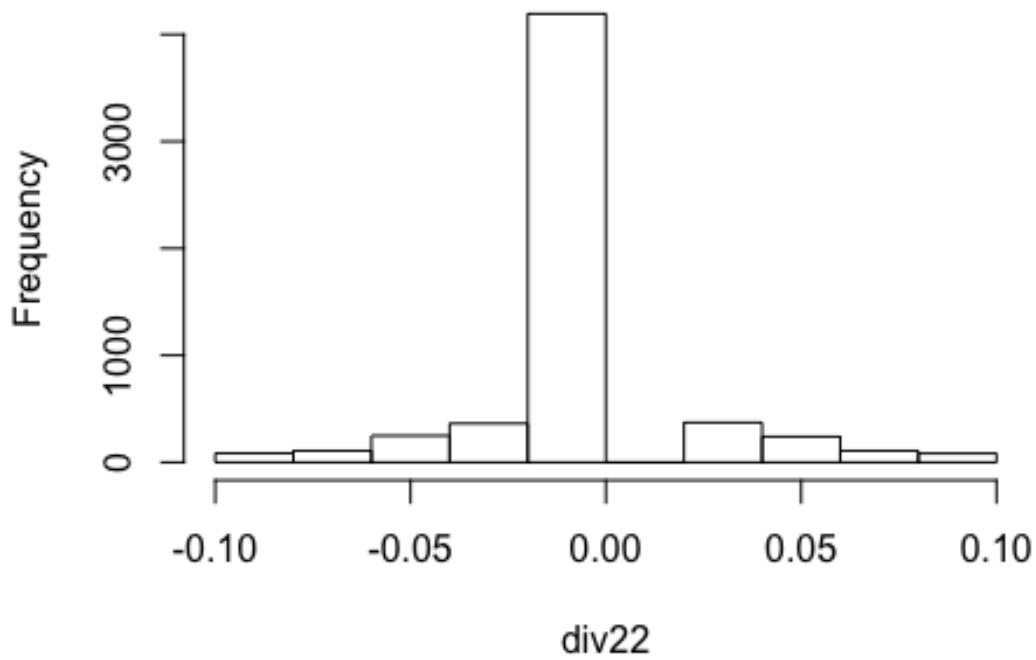
Histograma de gf2



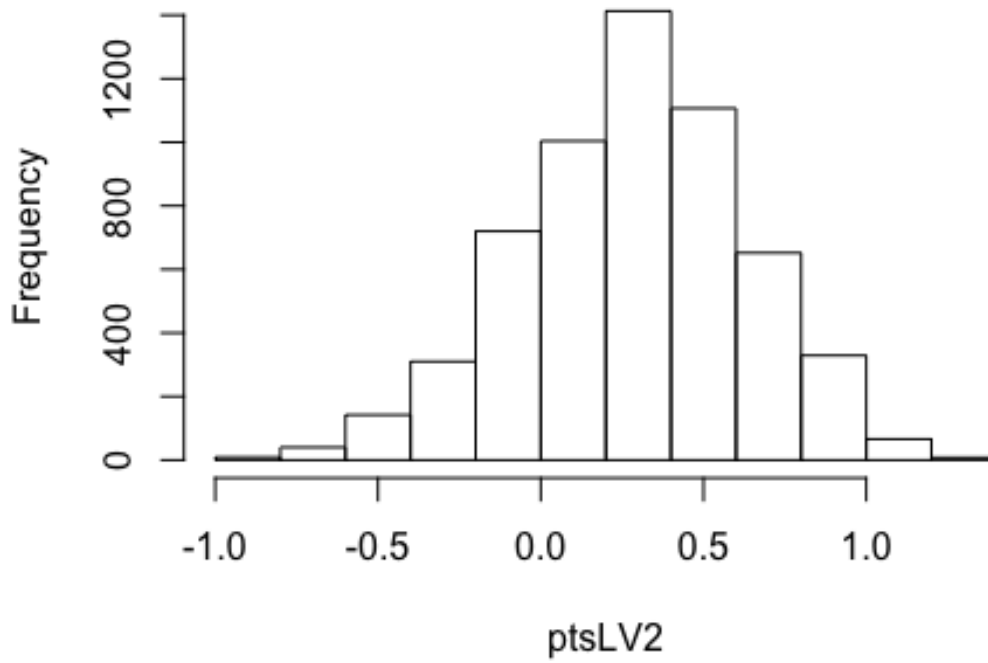
Histograma de gc2



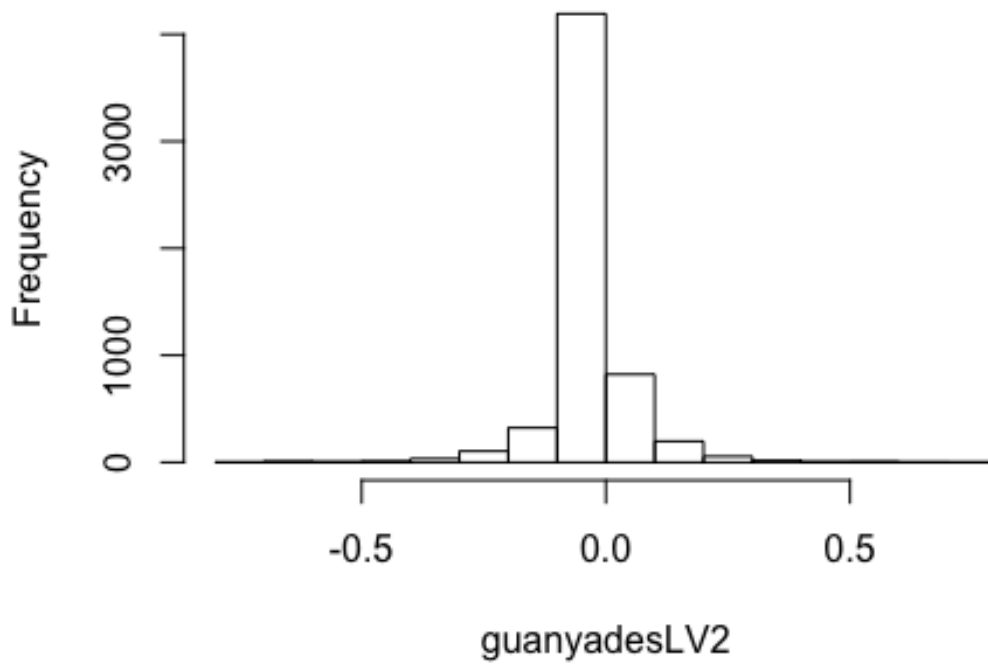
Histograma de div22



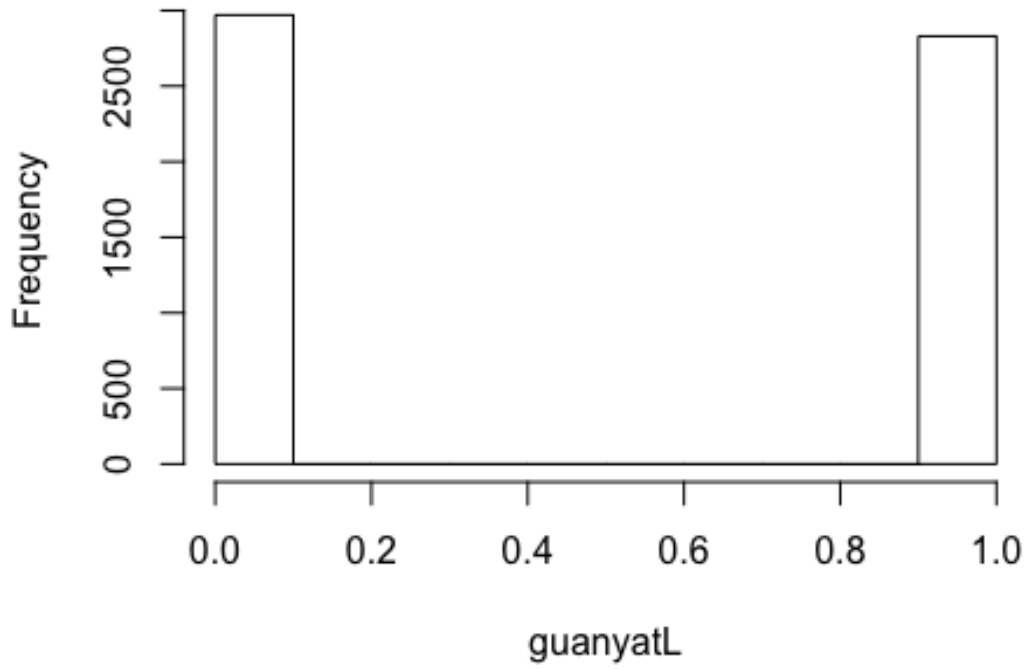
Histograma de ptsLV2



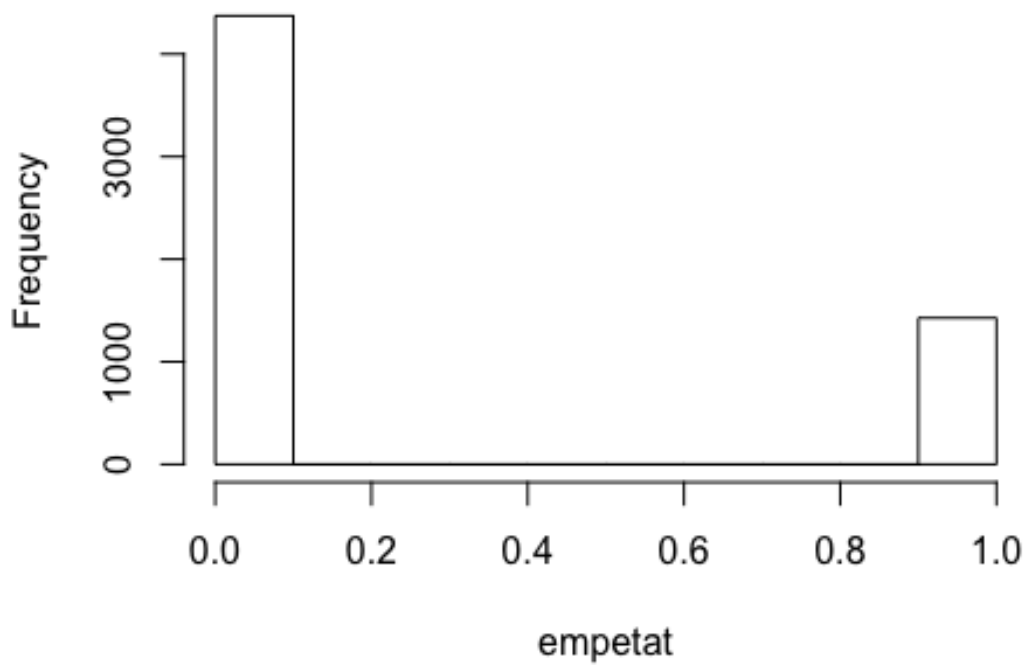
Histograma de guanyadesLV2



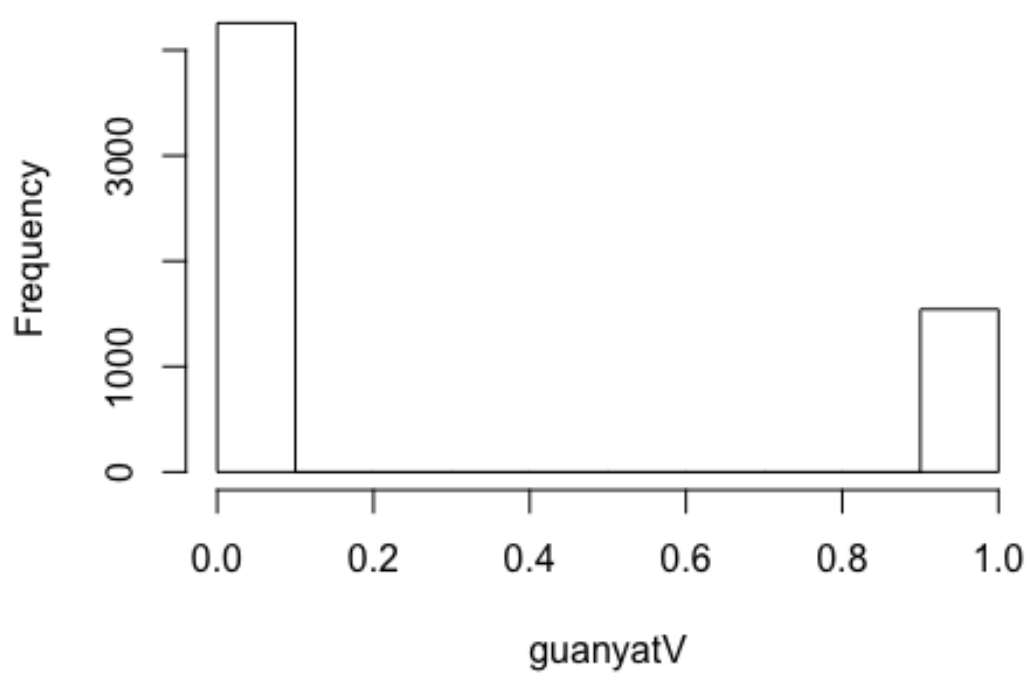
Histograma de guanyatL



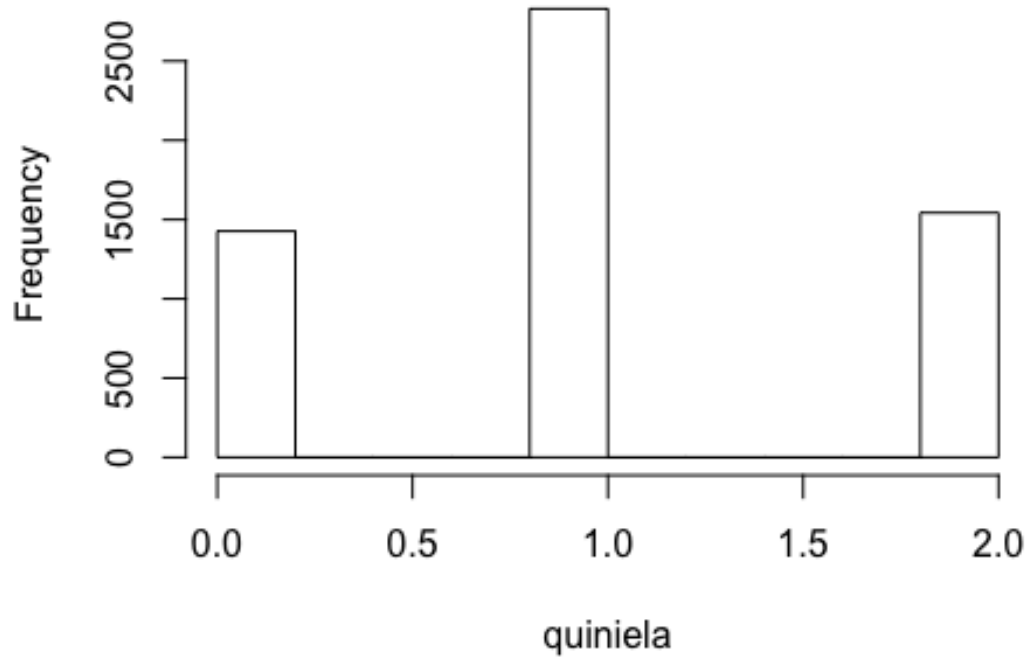
Histograma de empetat



Histograma de guanyatV

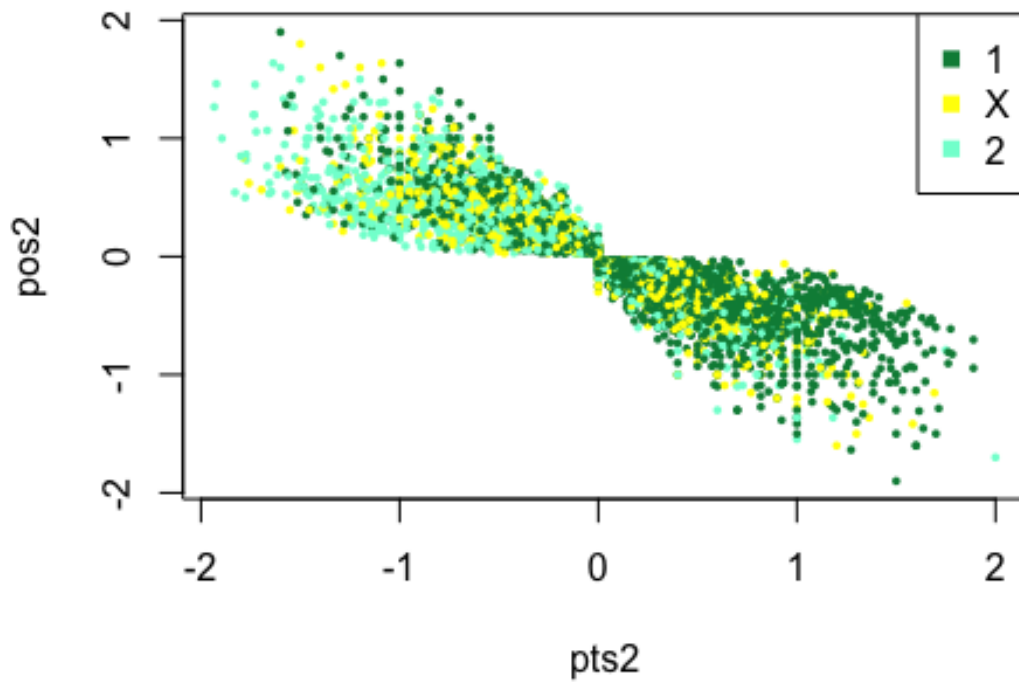


Histograma de quiniela

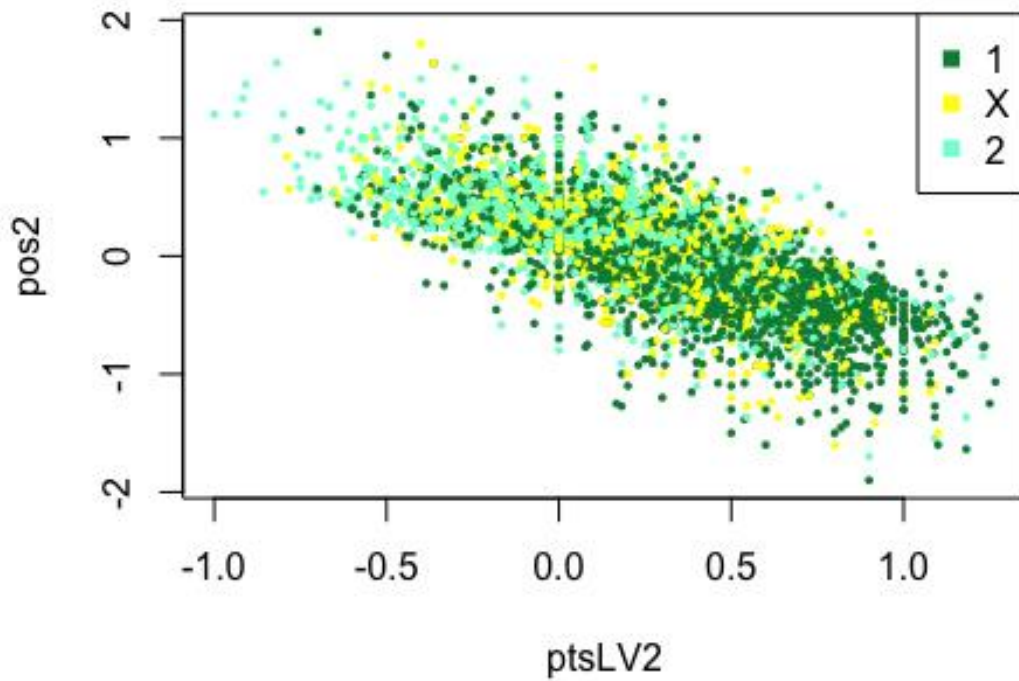


2. Descripció bivariant

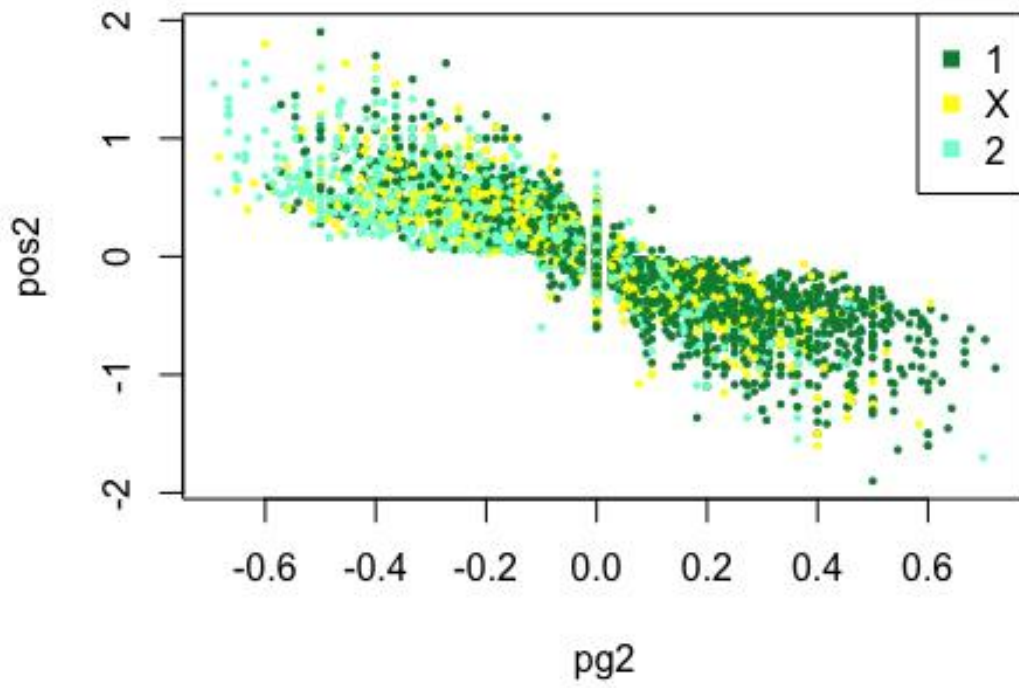
pos2 i pts2 segons el resultat



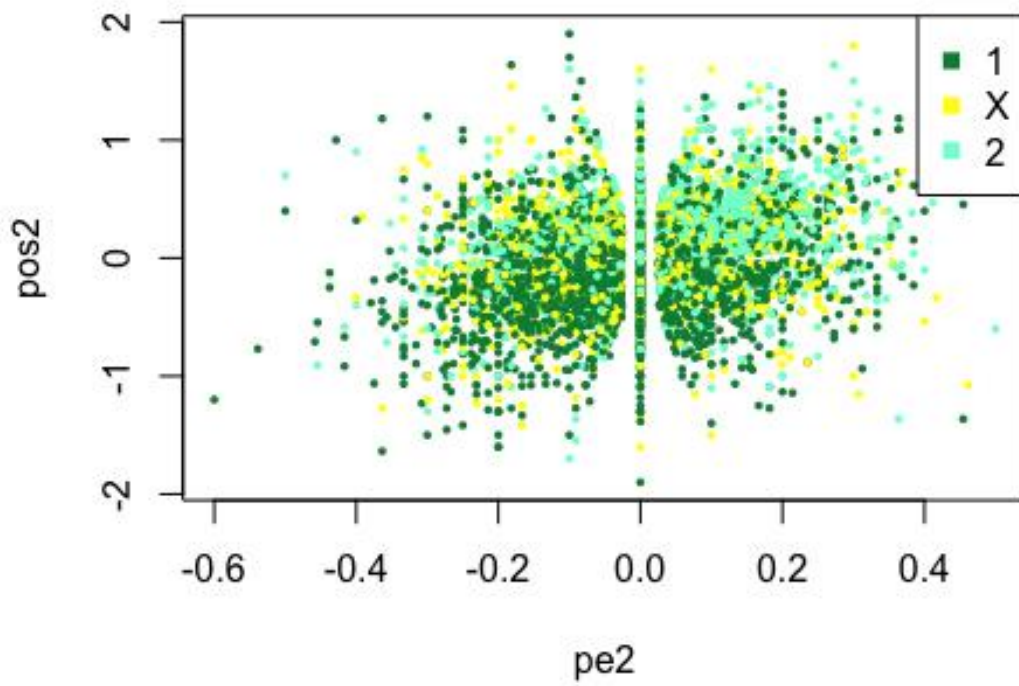
pos2 i ptsLV2 segons el resultat



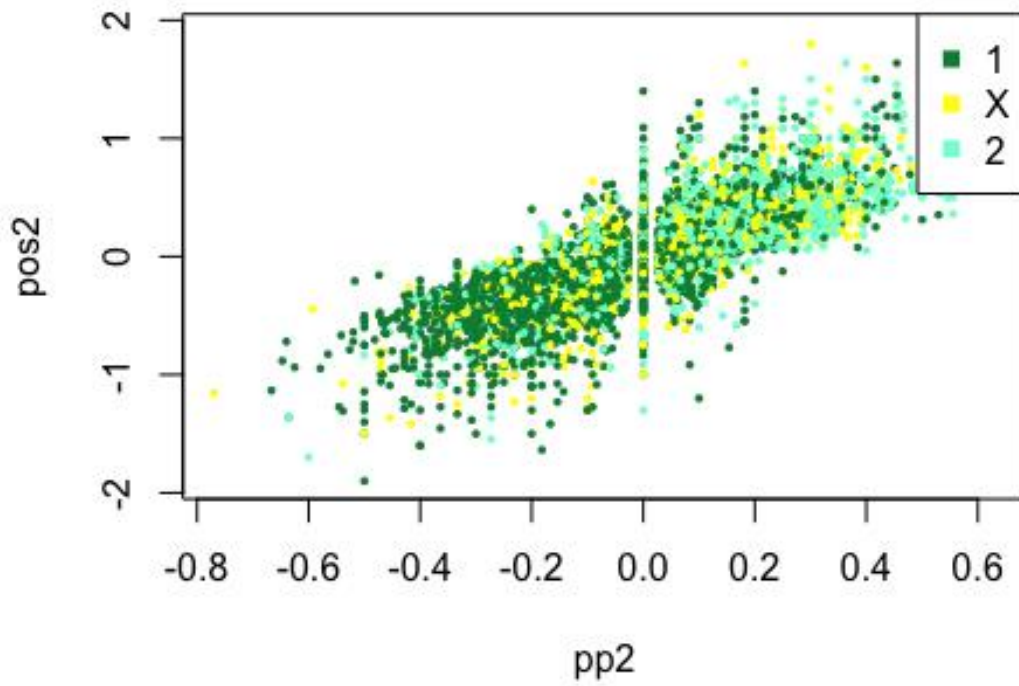
pos2 i pg2 segons el resultat



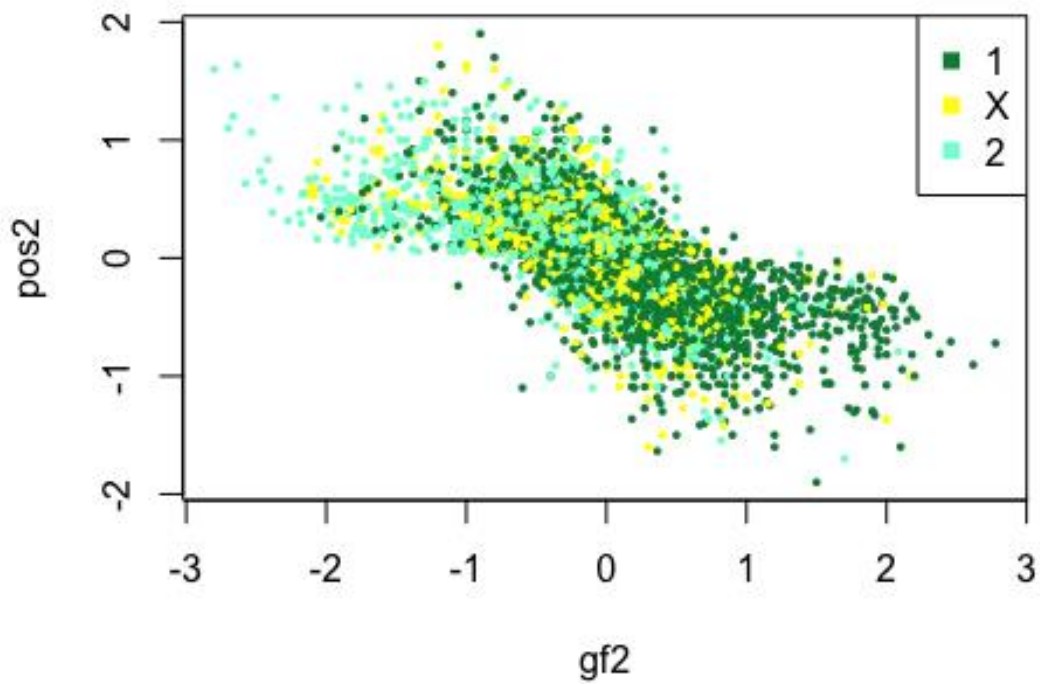
pos2 i pe2 segons el resultat



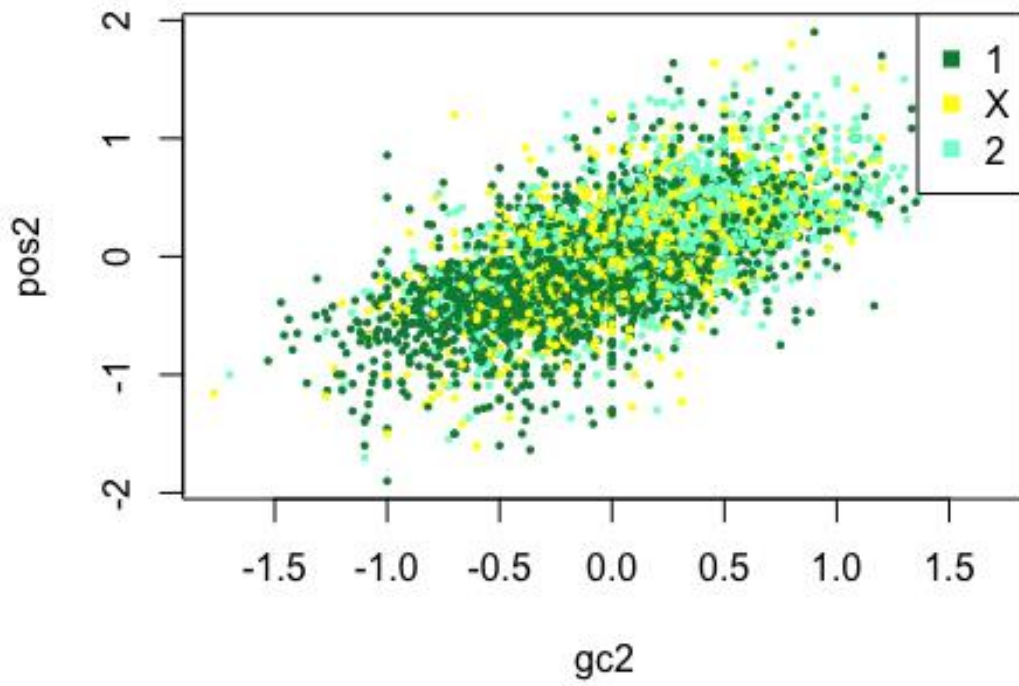
pos2 i pp2 segons el resultat



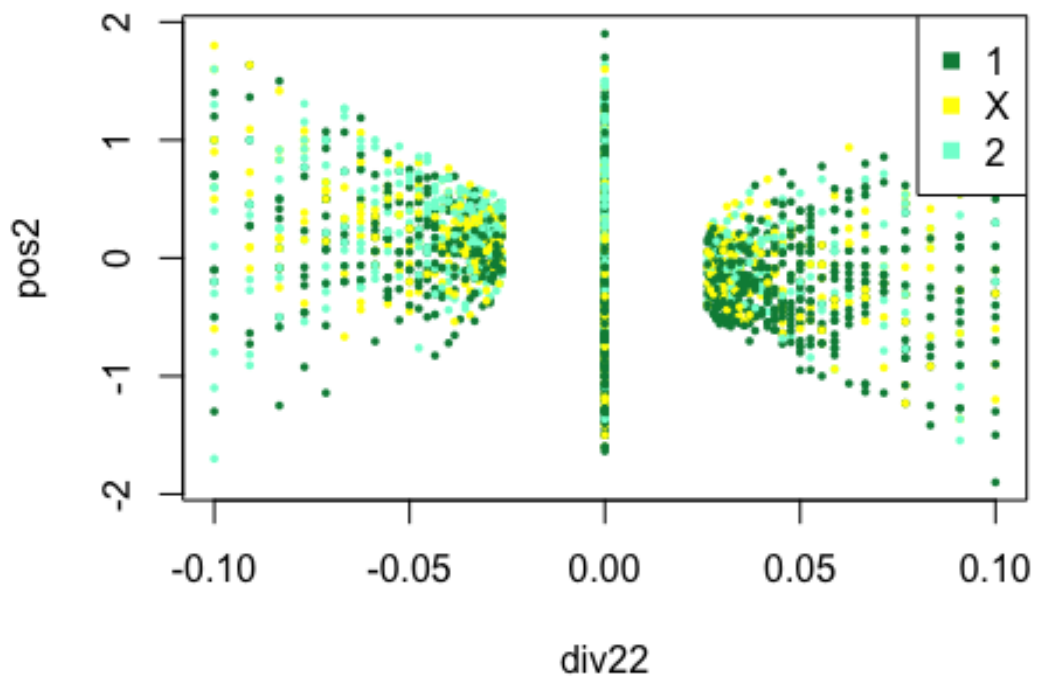
pos2 i gf2 segons el resultat



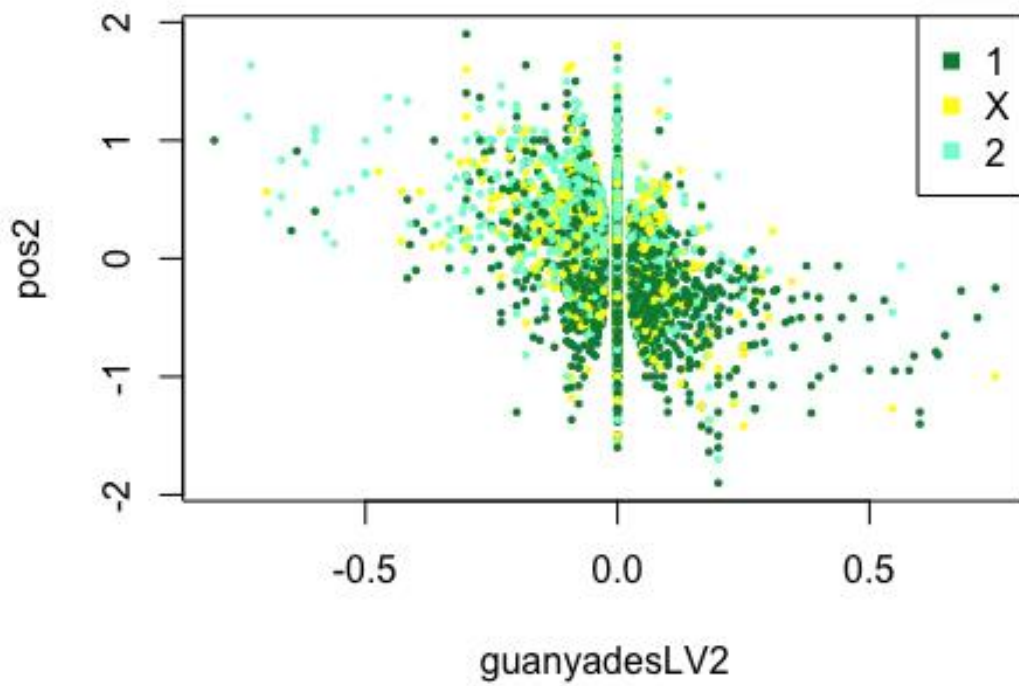
pos2 i gc2 segons el resultat



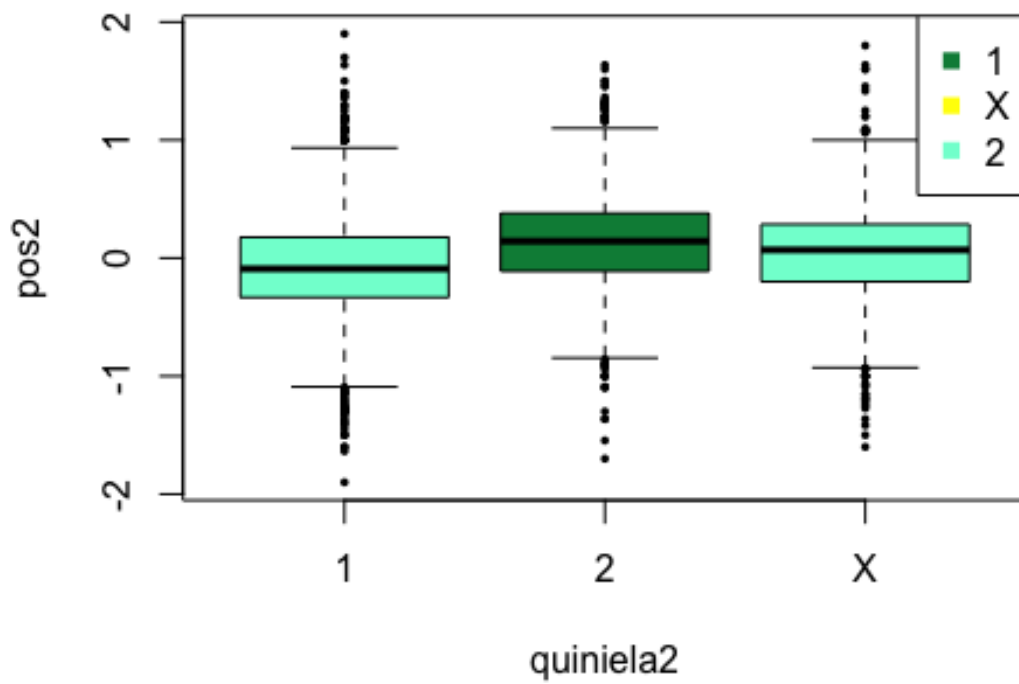
pos2 i div22 segons el resultat



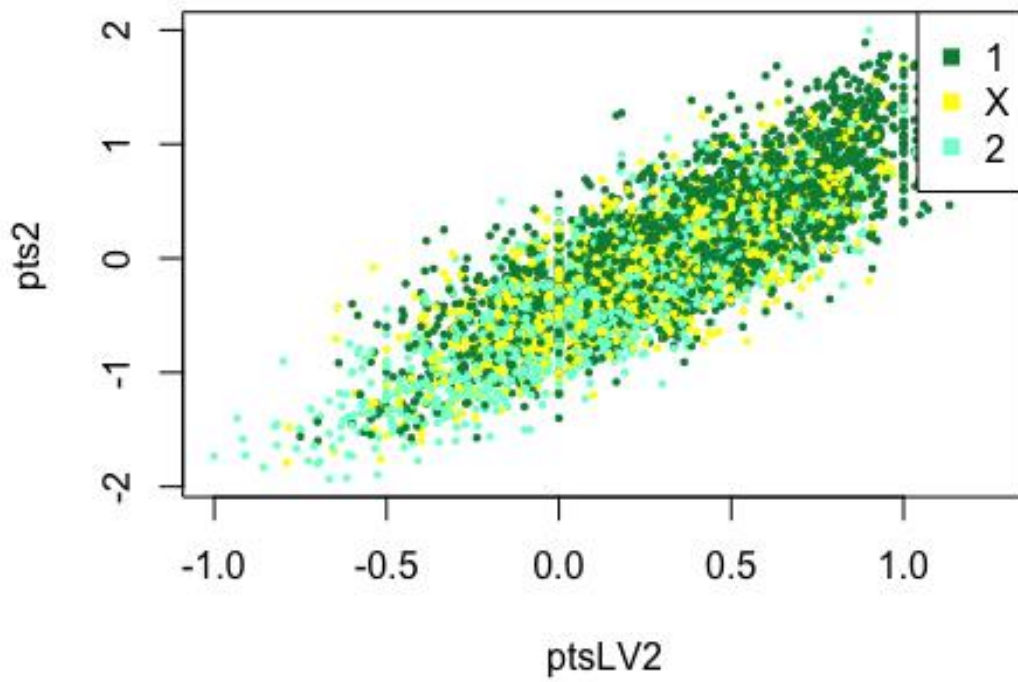
pos2 i guanyadesLV2 segons el resultat



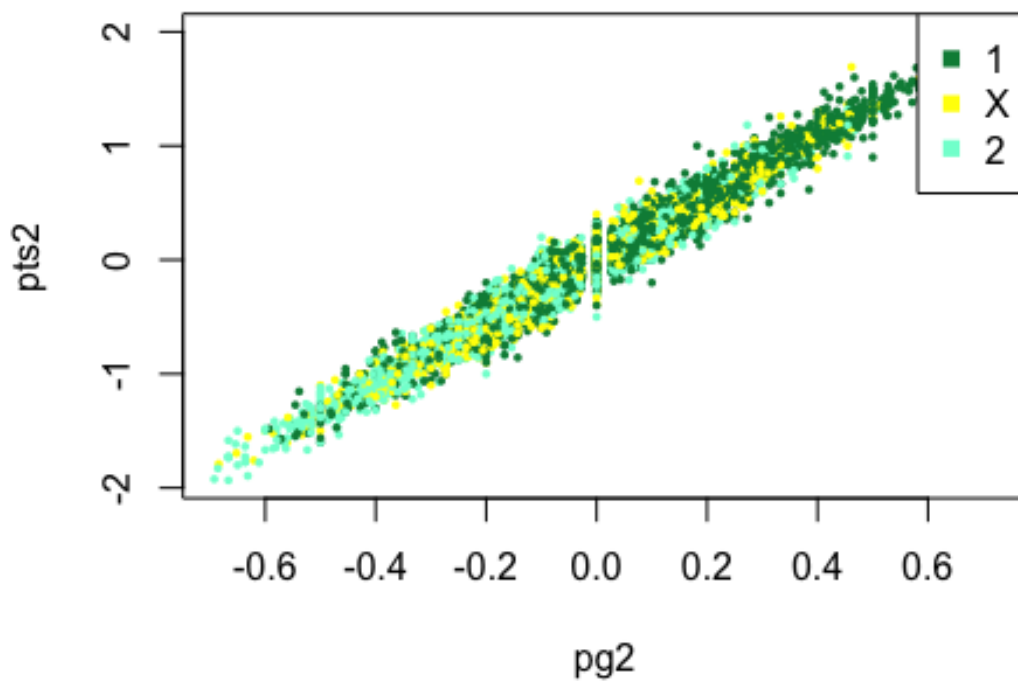
pos2 i quiniela2 segons el resultat



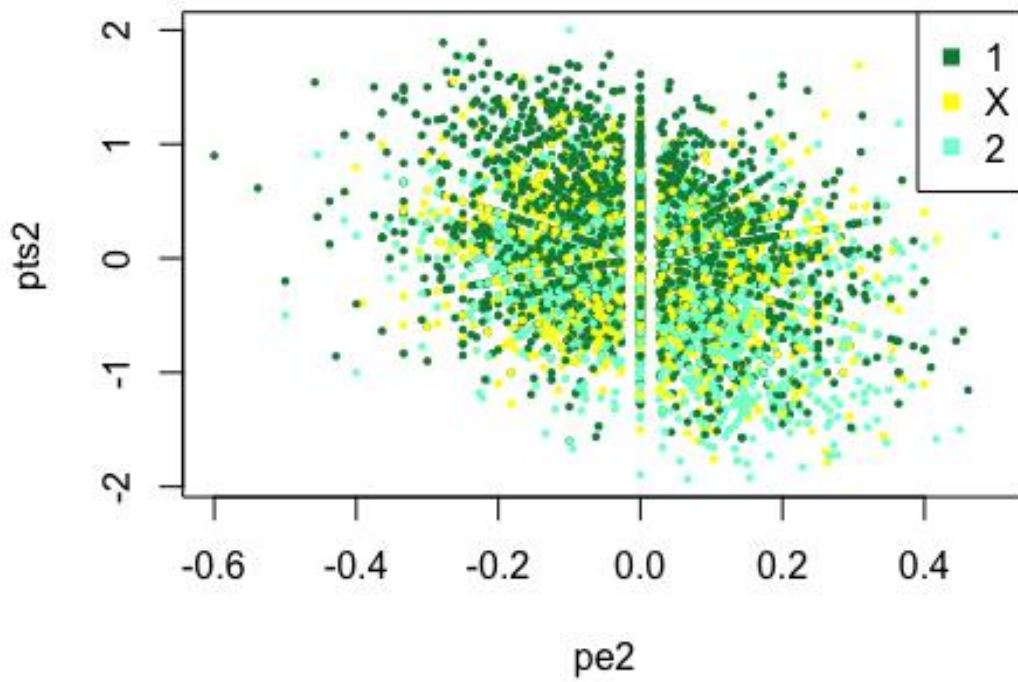
pts2 i ptsLV2 segons el resultat



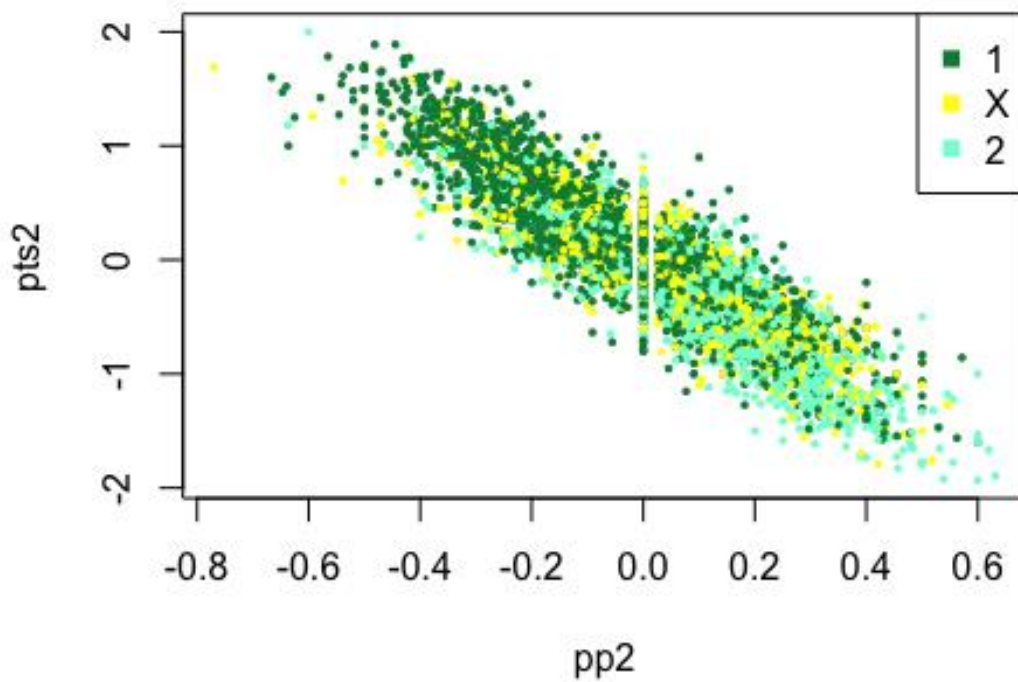
pts2 i pg2 segons el resultat



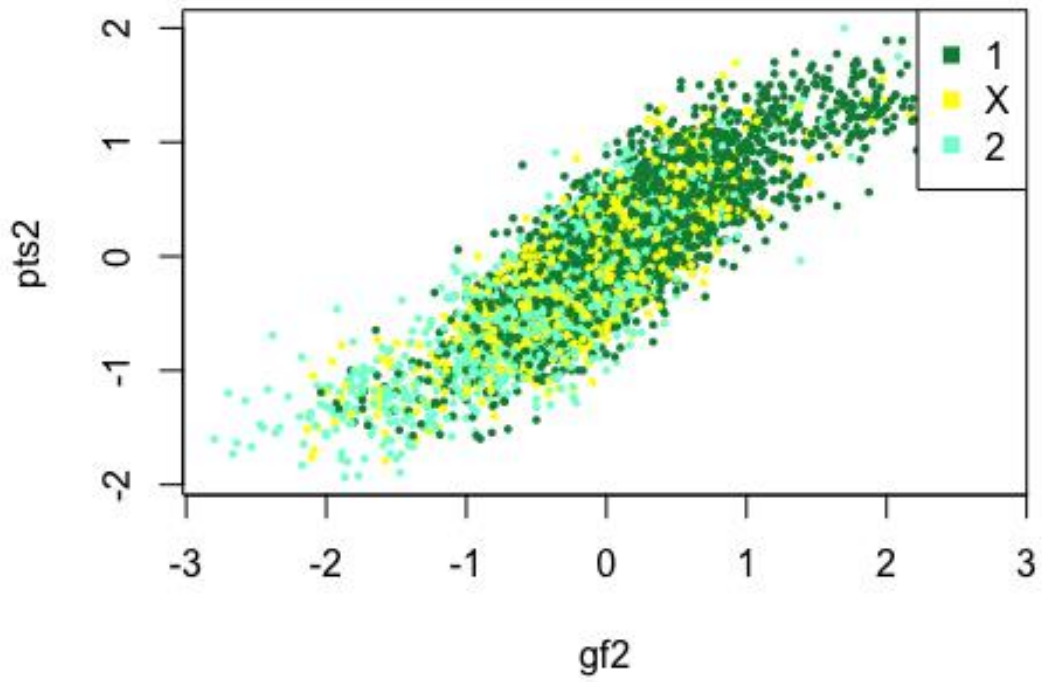
pts2 i pe2 segons el resultat



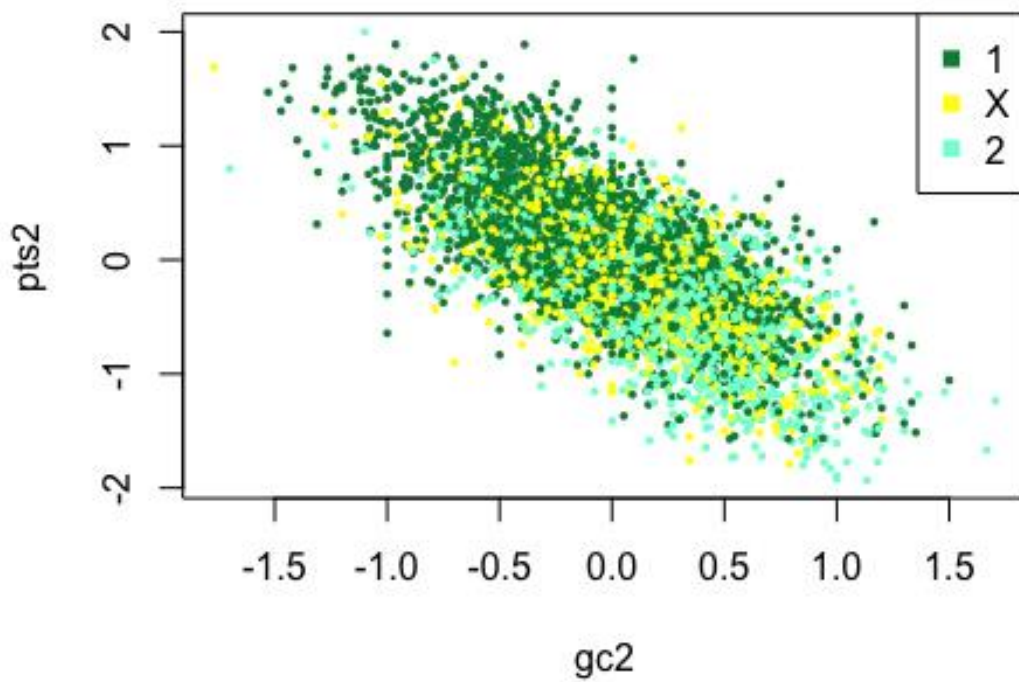
pts2 i pp2 segons el resultat



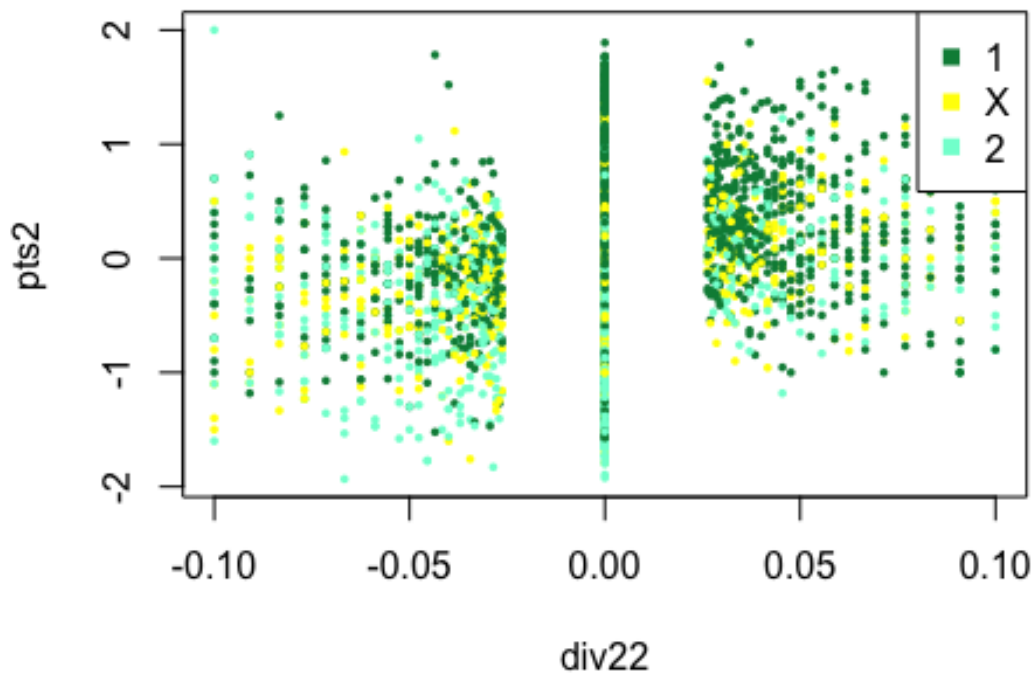
pts2 i gf2 segons el resultat



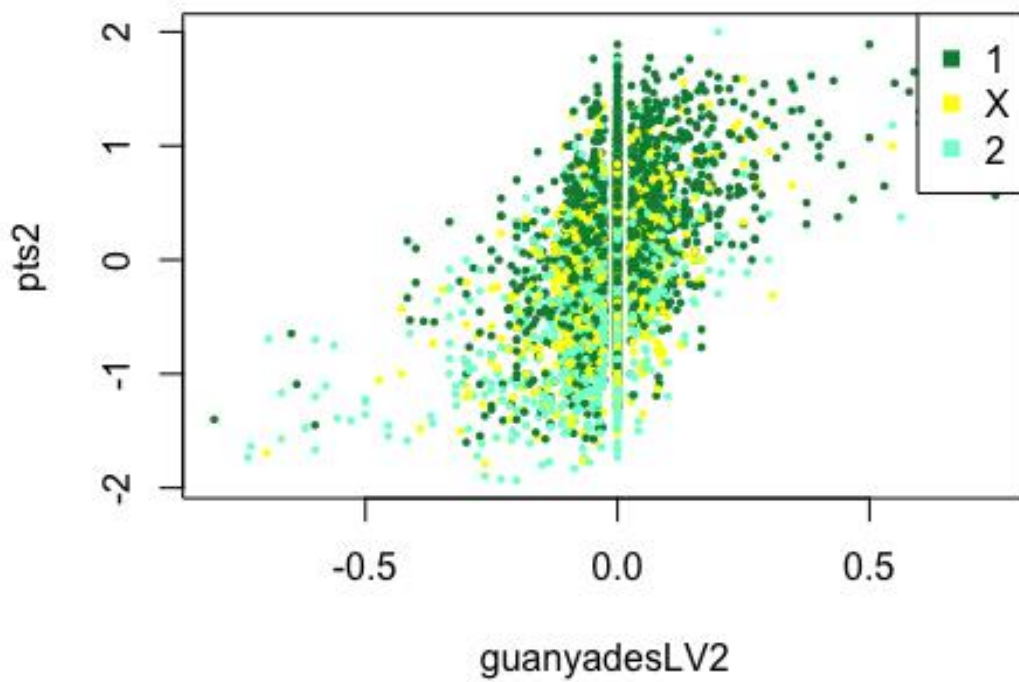
pts2 i gc2 segons el resultat



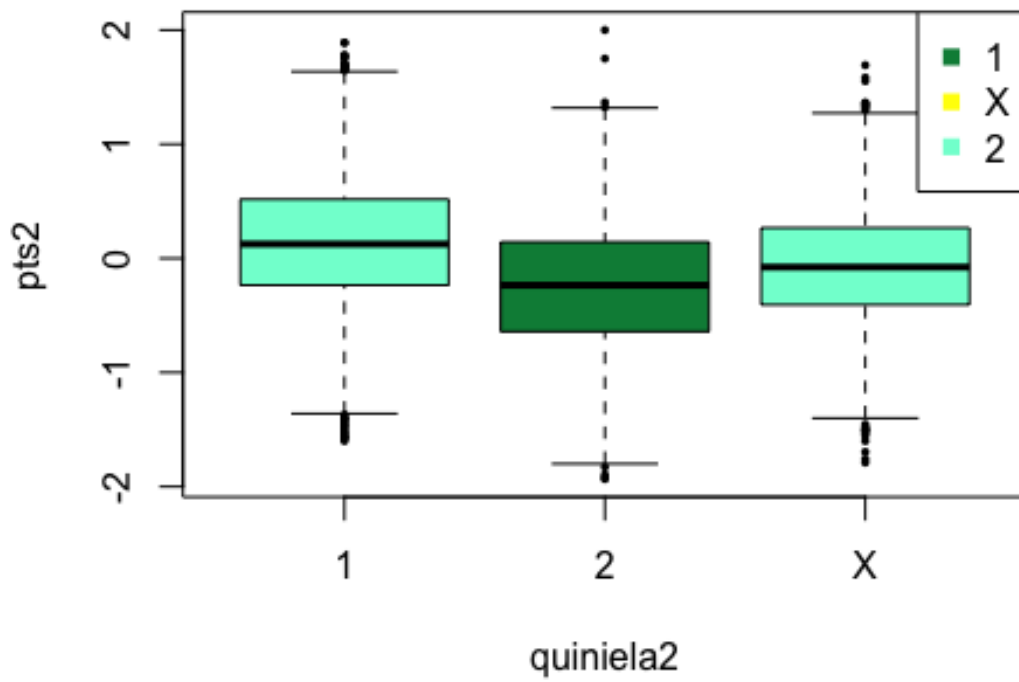
pts2 i div22 segons el resultat



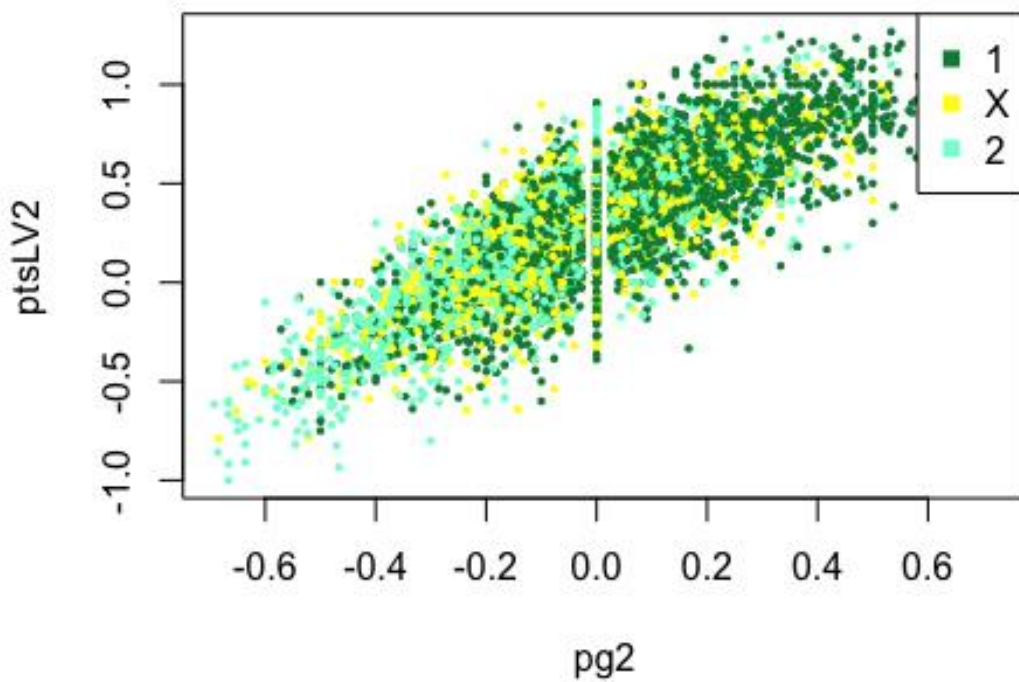
pts2 i guanyadesLV2 segons el resultat



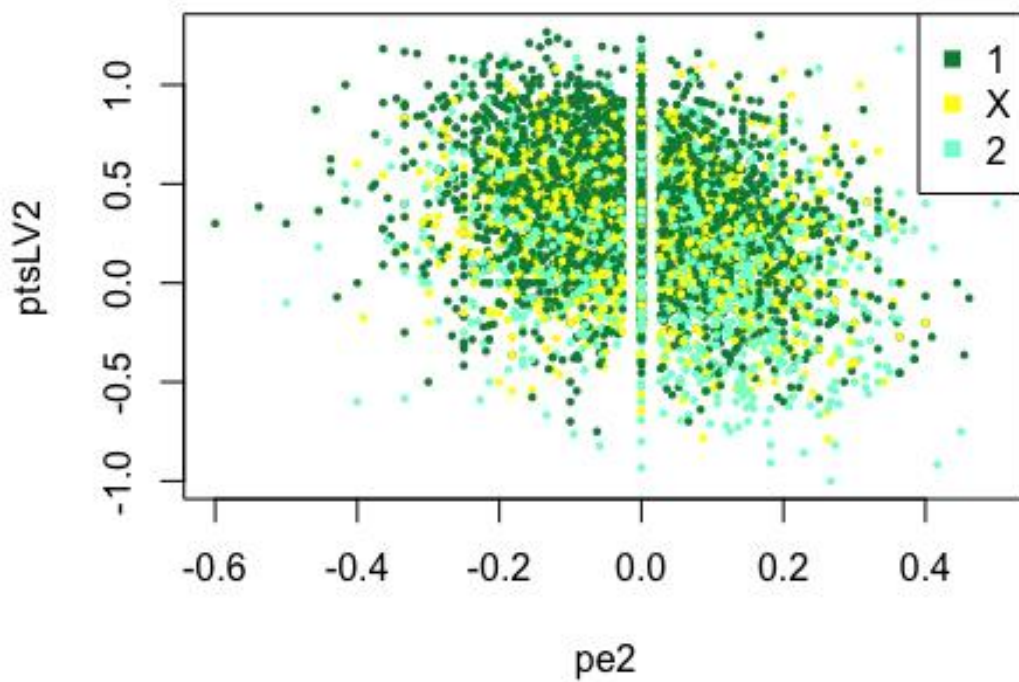
pts2 i quiniela2 segons el resultat



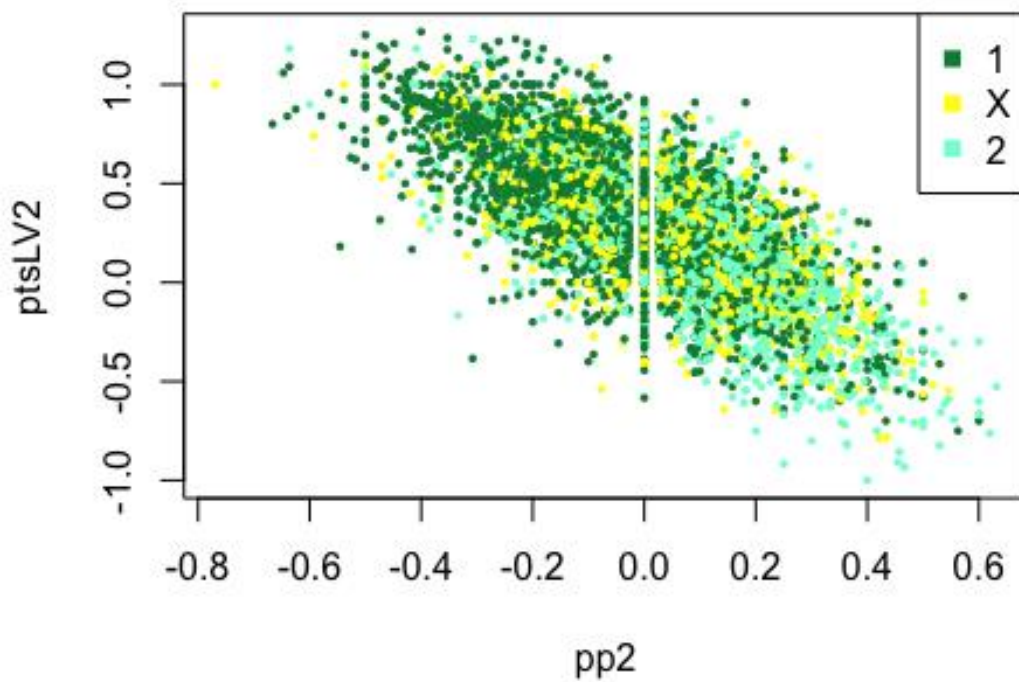
ptsLV2 i pg2 segons el resultat



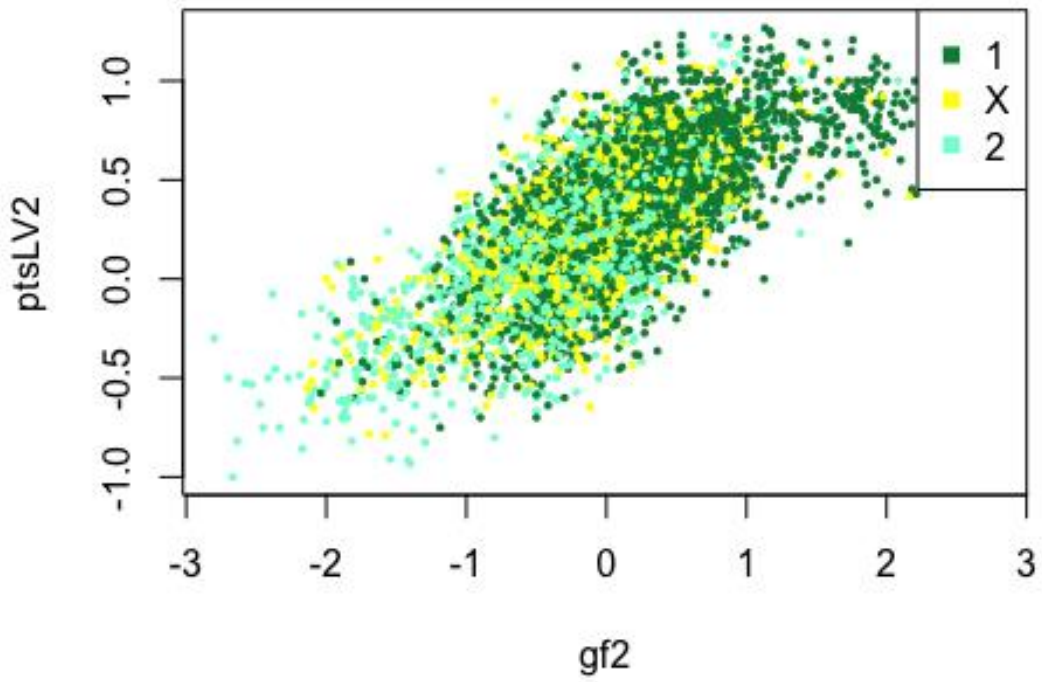
ptsLV2 i pe2 segons el resultat



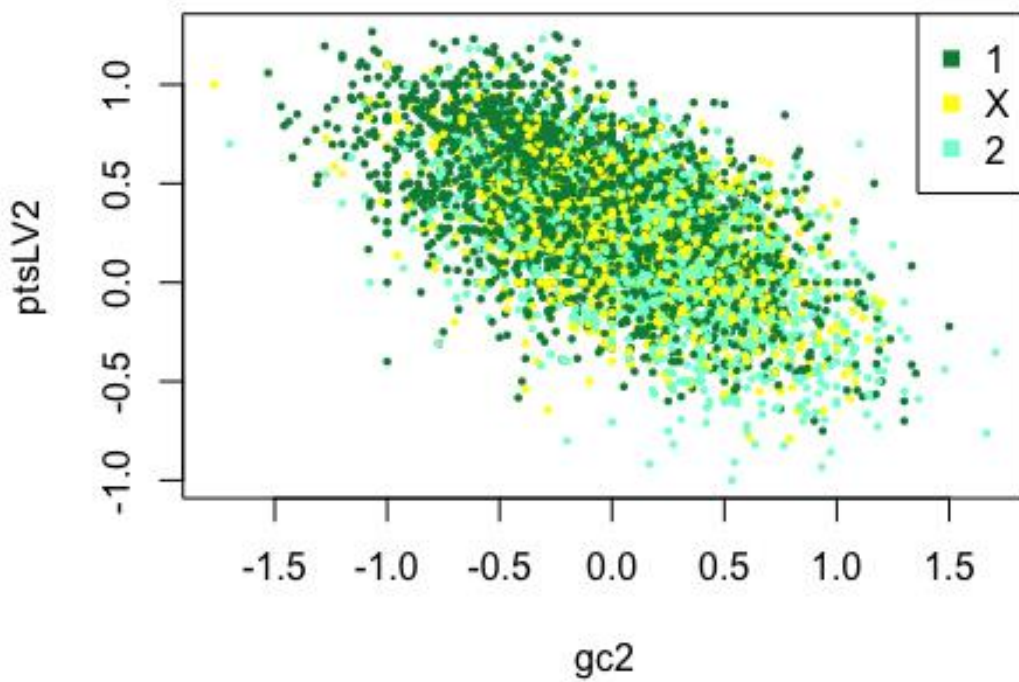
ptsLV2 i pp2 segons el resultat



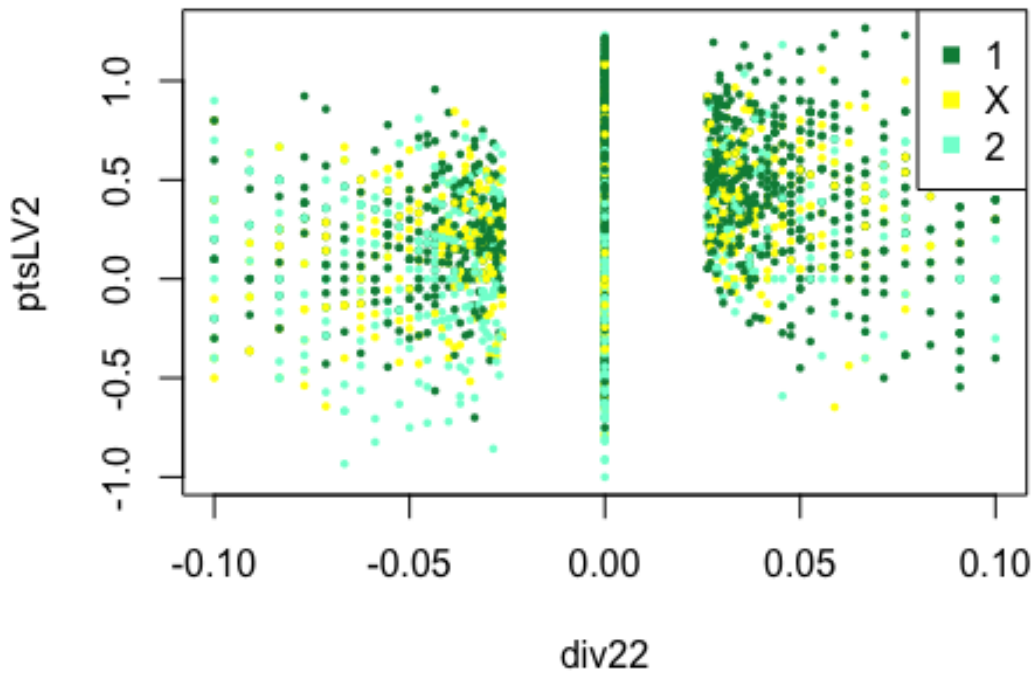
ptsLV2 i gf2 segons el resultat



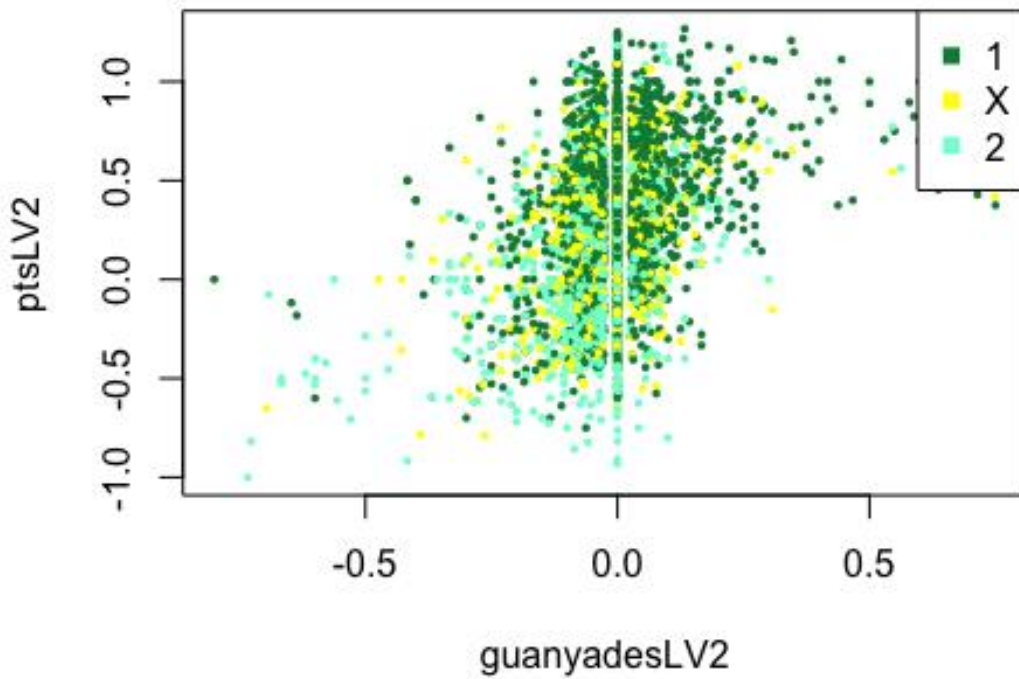
ptsLV2 i gc2 segons el resultat



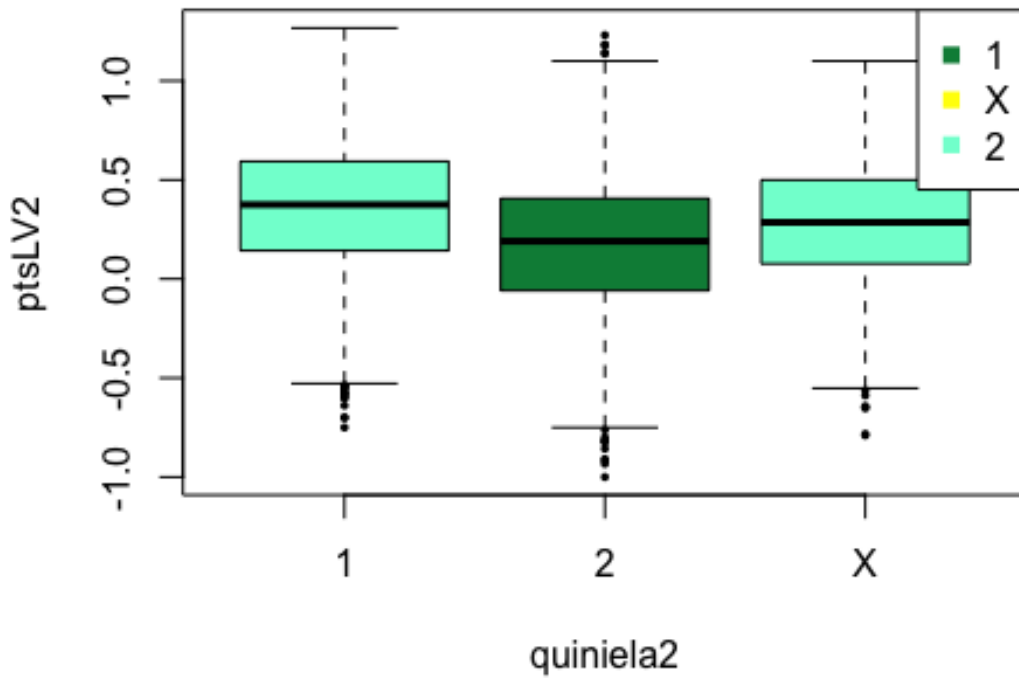
ptsLV2 i div22 segons el resultat



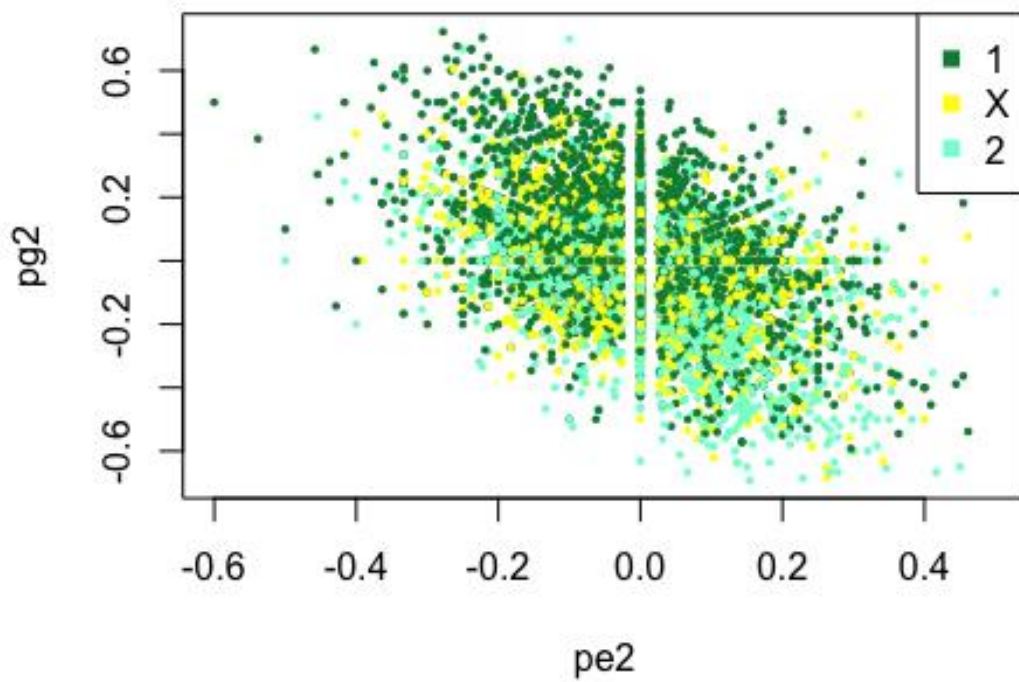
ptsLV2 i guanyadesLV2 segons el resultat



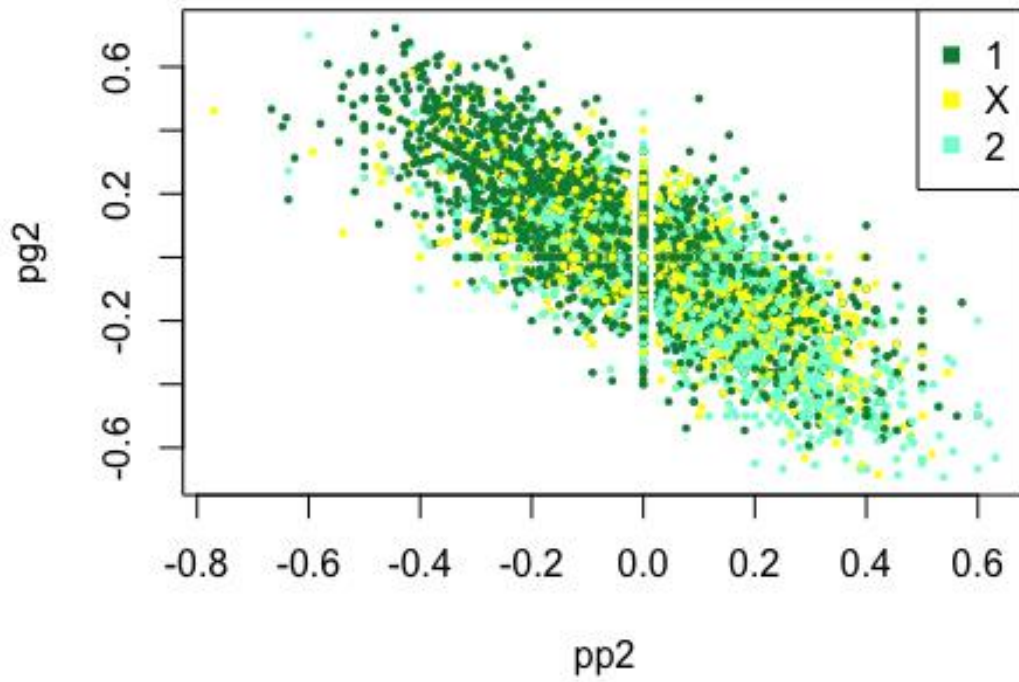
ptsLV2 i quiniela2 segons el resultat



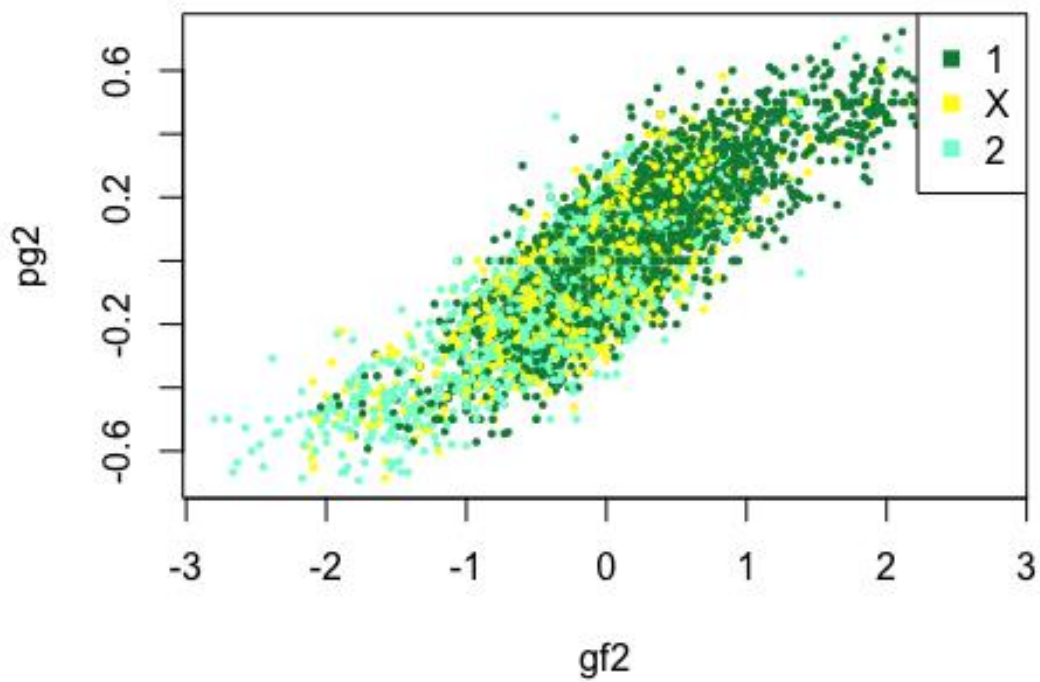
pg2 i pe2 segons el resultat



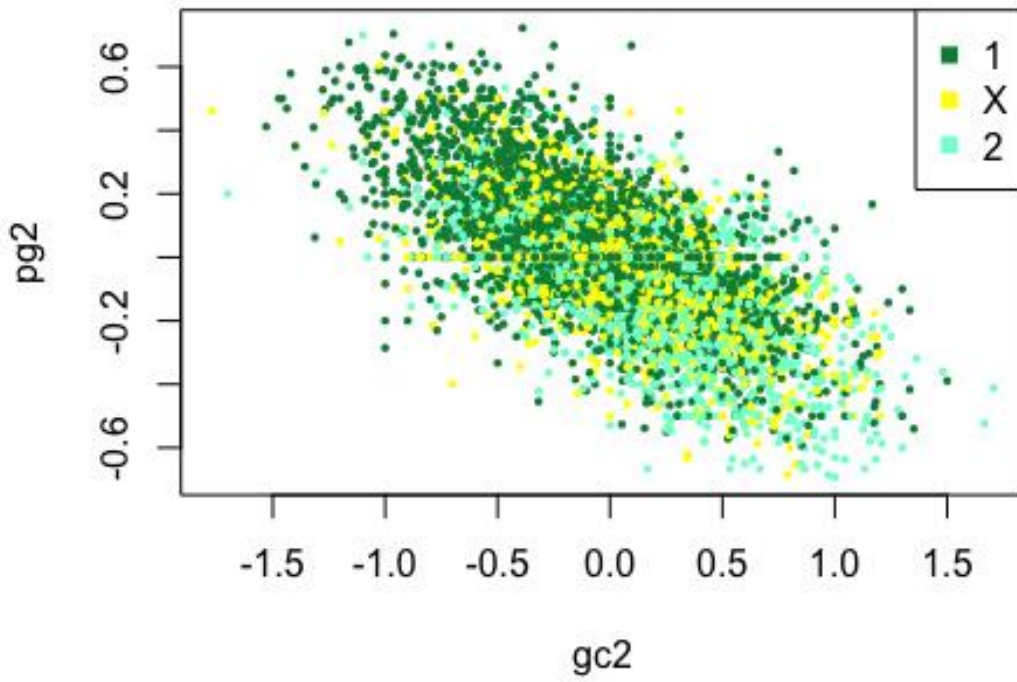
pg2 i pp2 segons el resultat



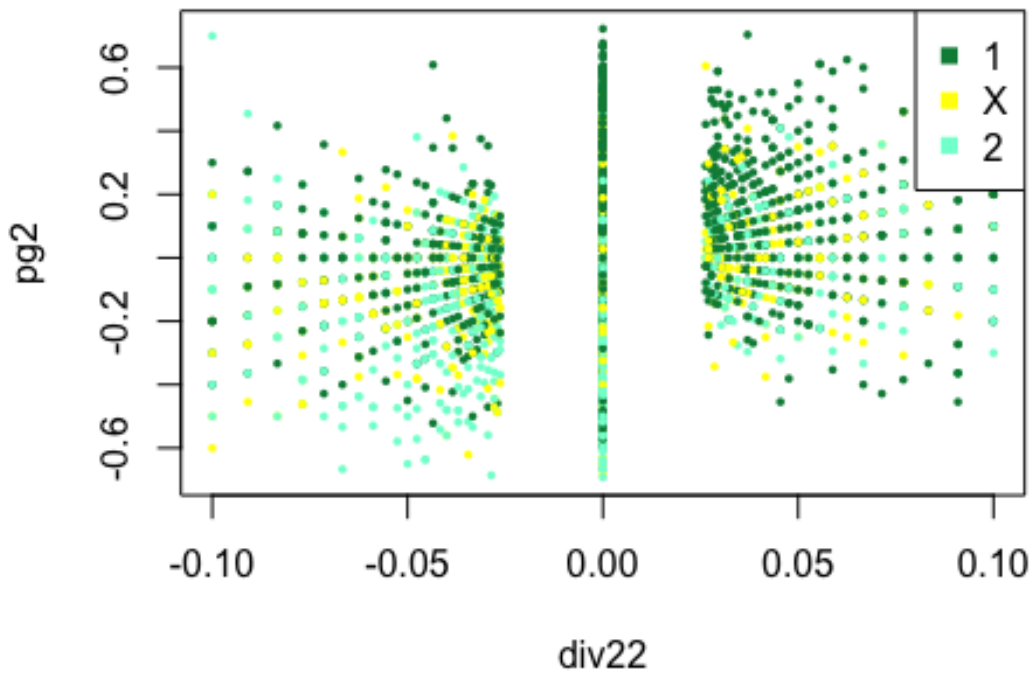
pg2 i gf2 segons el resultat



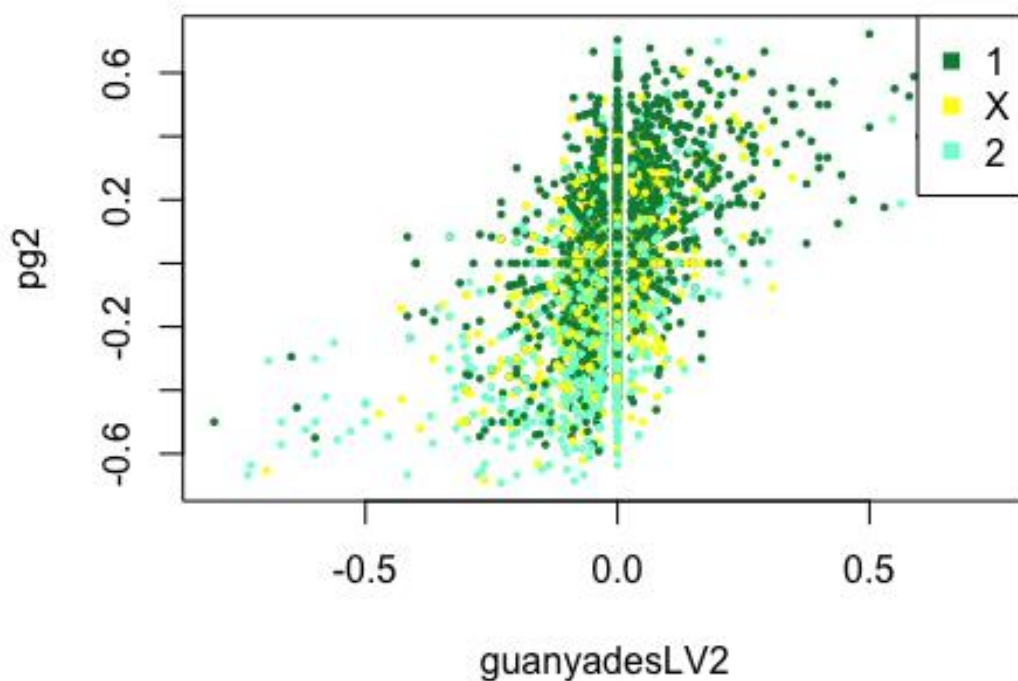
pg2 i gc2 segons el resultat



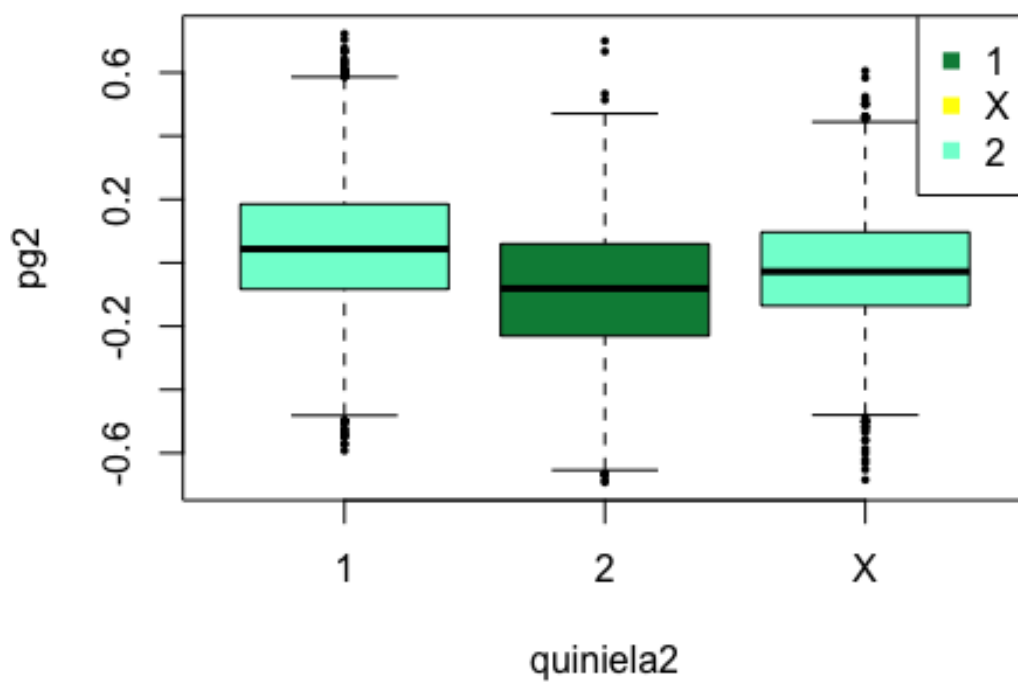
pg2 i div22 segons el resultat



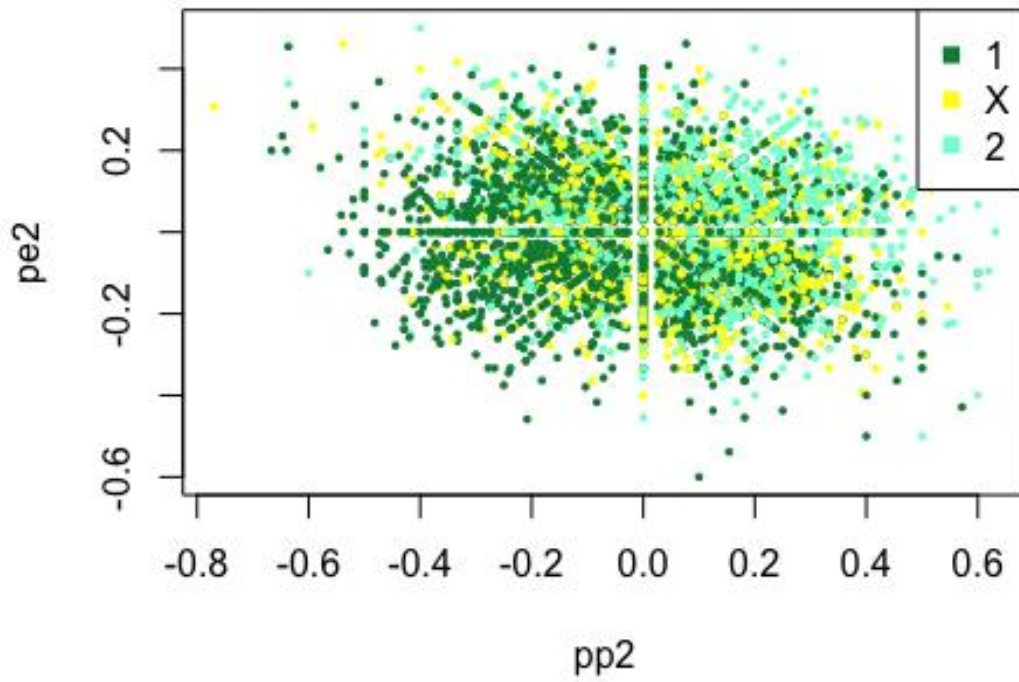
pg2 i guanyadesLV2 segons el resultat



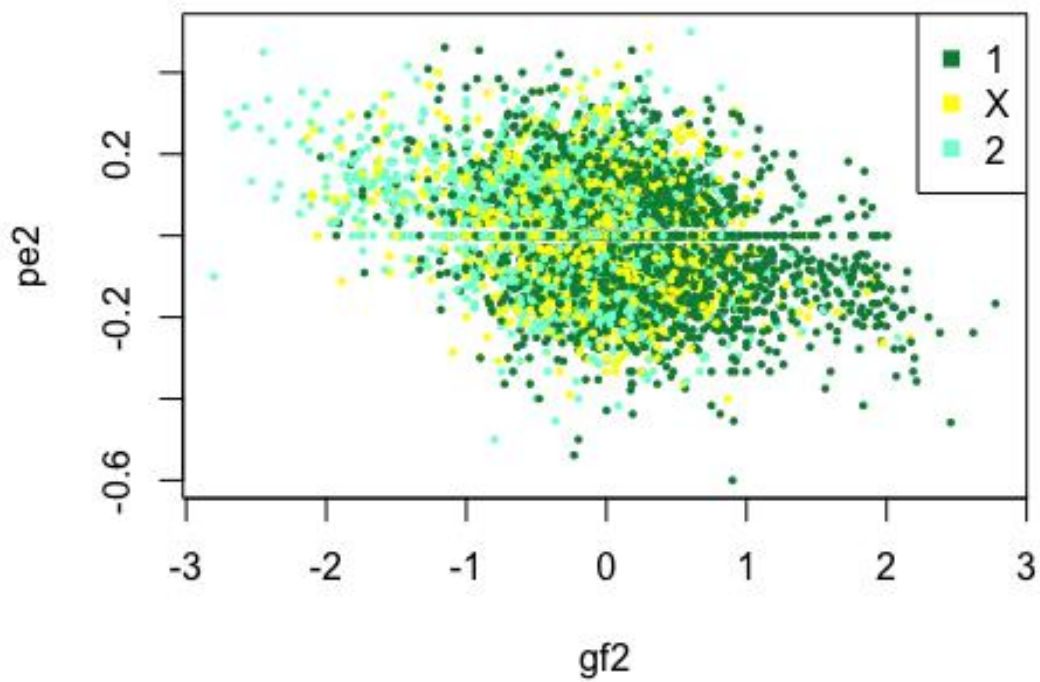
pg2 i quiniela2 segons el resultat



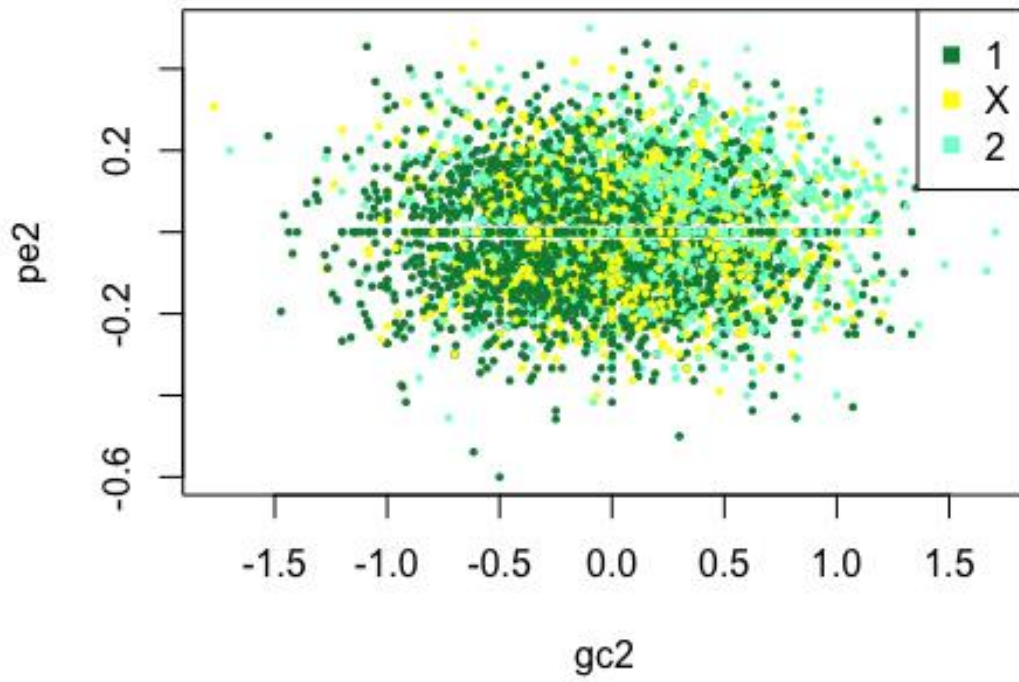
pe2 i pp2 segons el resultat



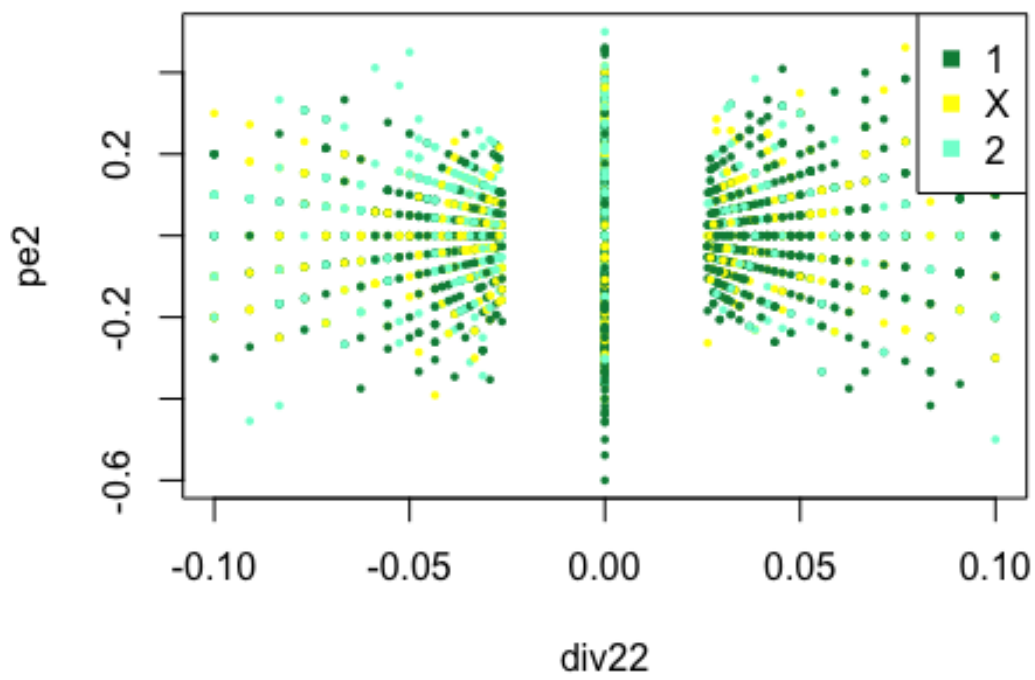
pe2 i gf2 segons el resultat



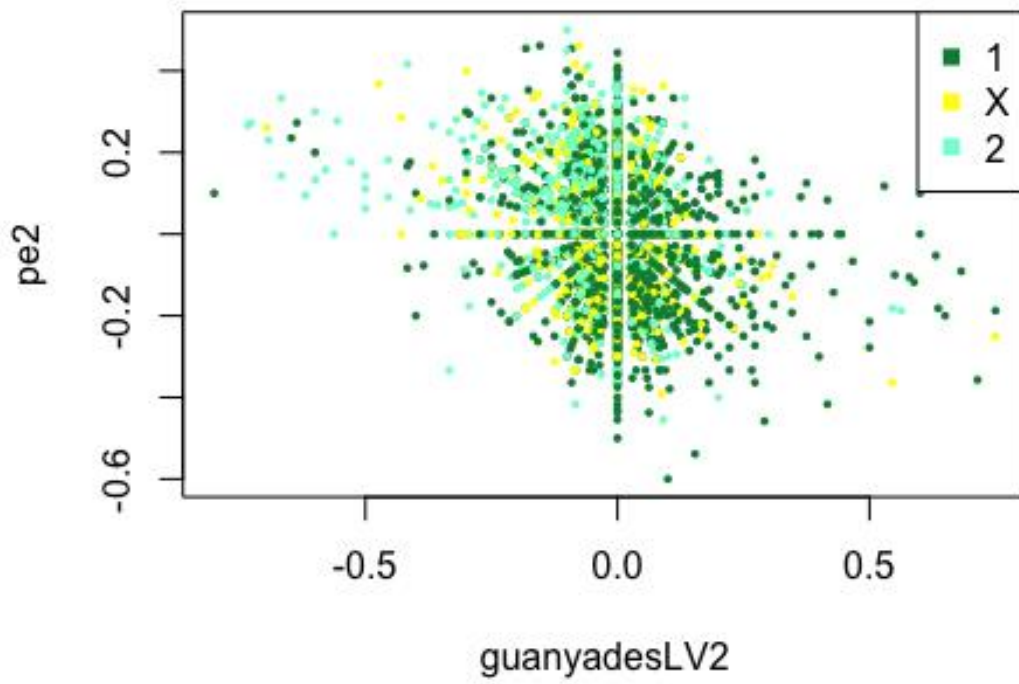
pe2 i gc2 segons el resultat



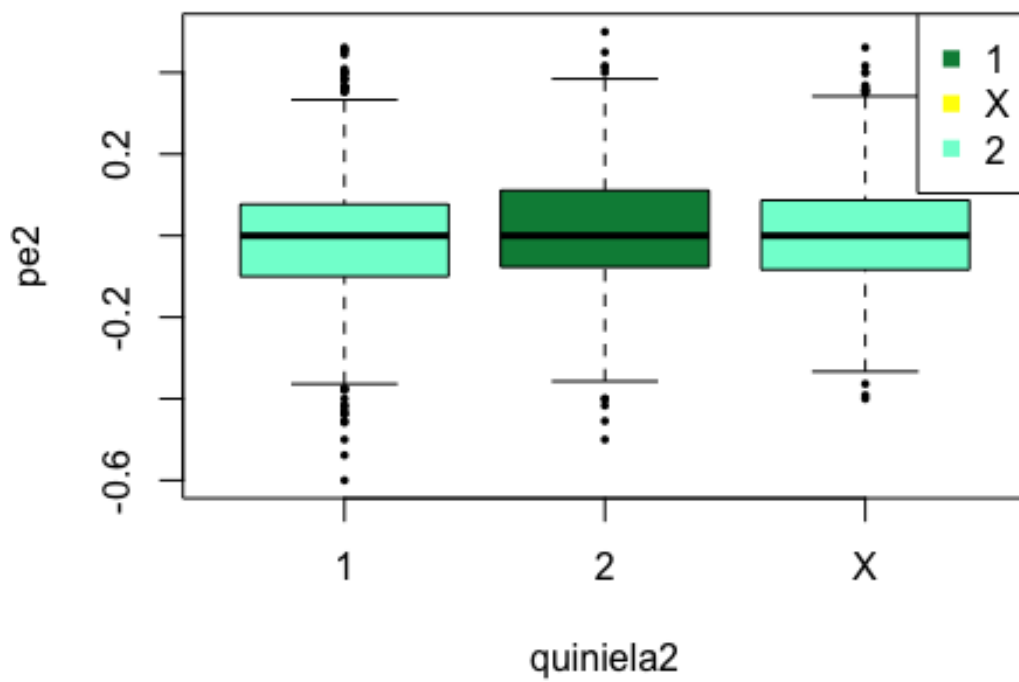
pe2 i div22 segons el resultat



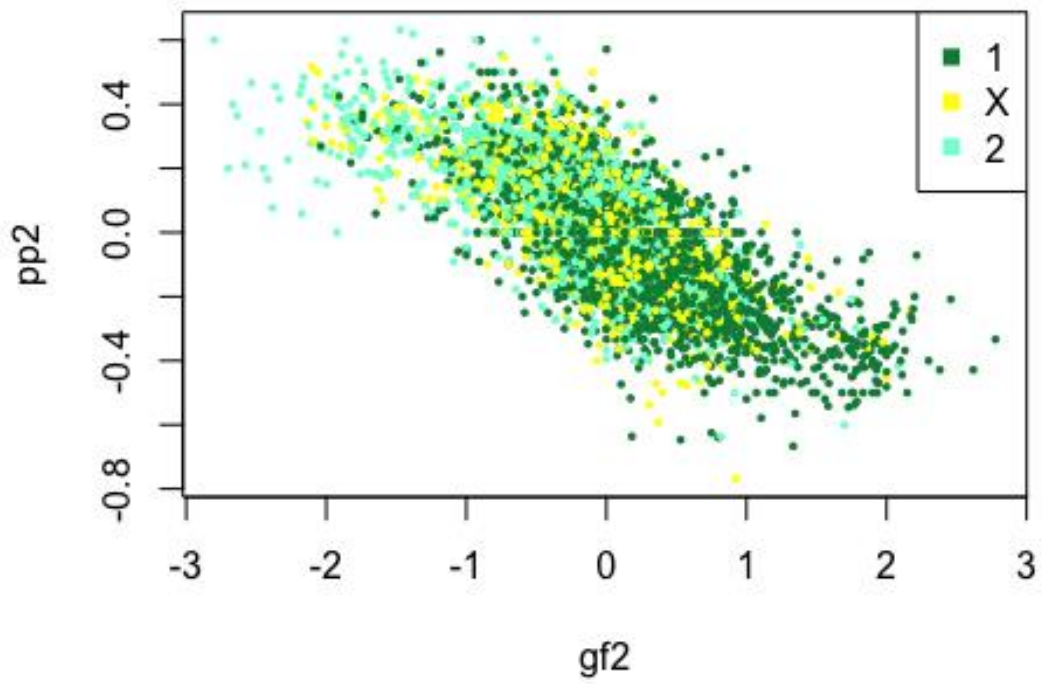
pe2 i guanyadesLV2 segons el resultat



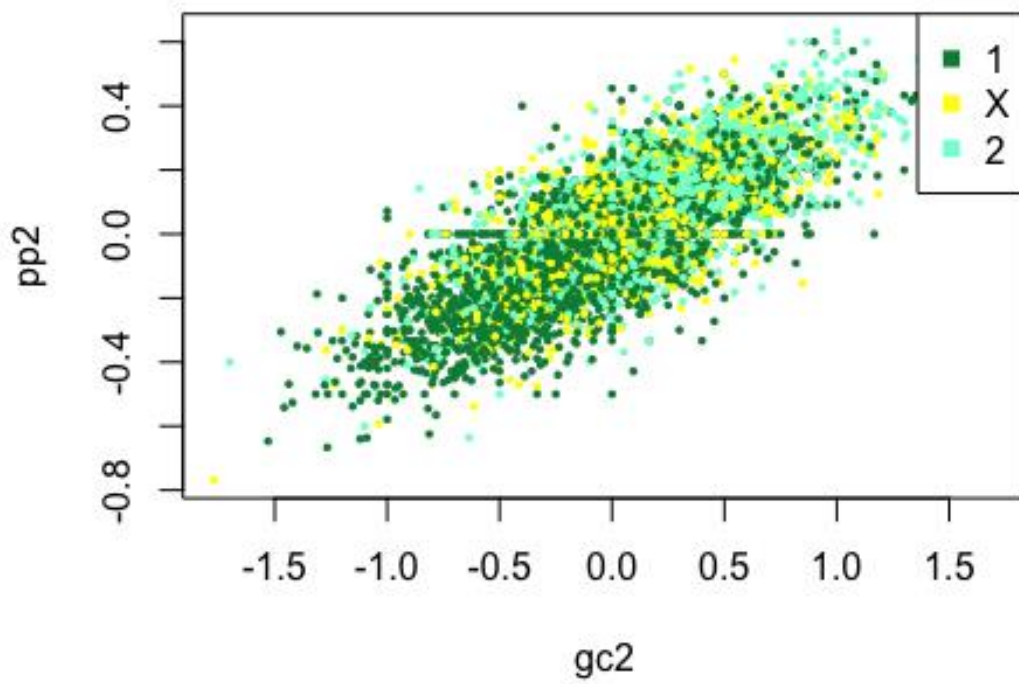
pe2 i quiniela2 segons el resultat



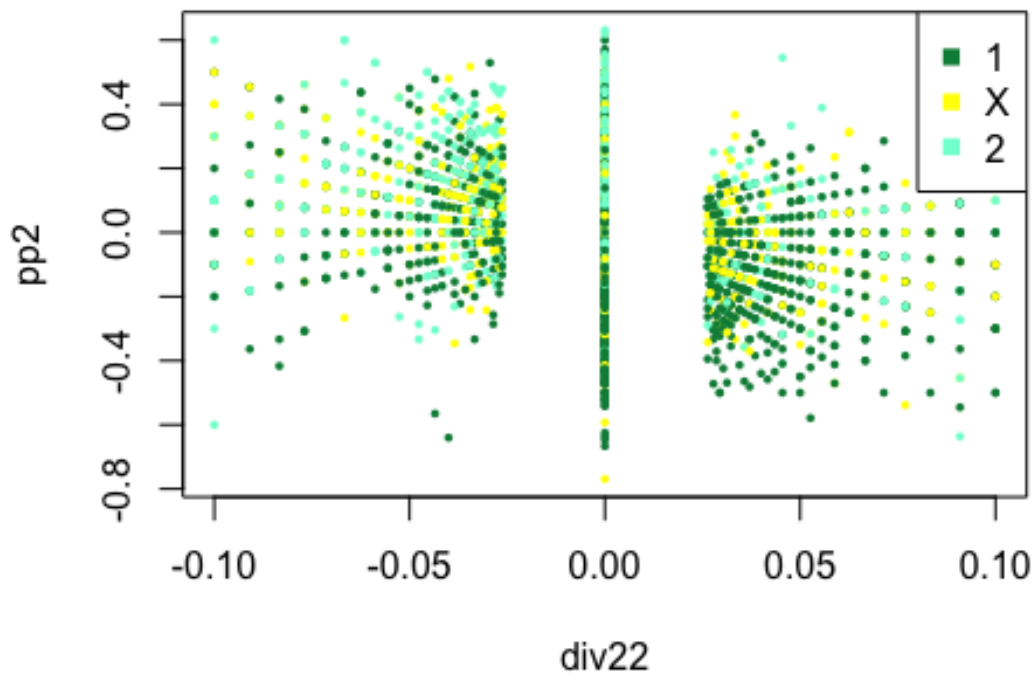
pp2 i gf2 segons el resultat



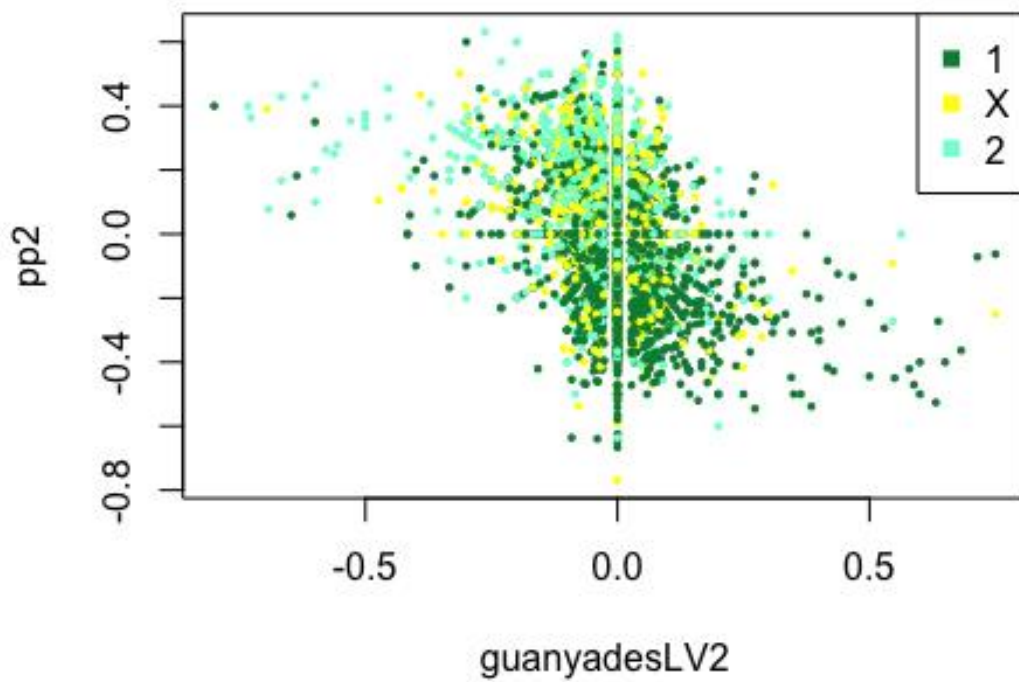
pp2 i gc2 segons el resultat



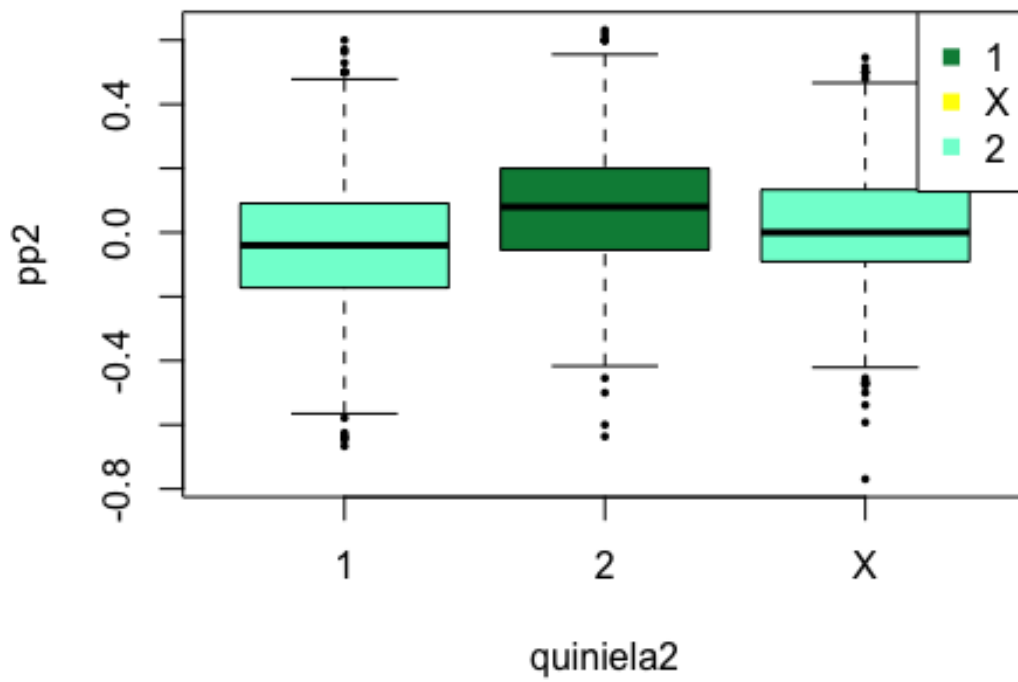
pp2 i div22 segons el resultat



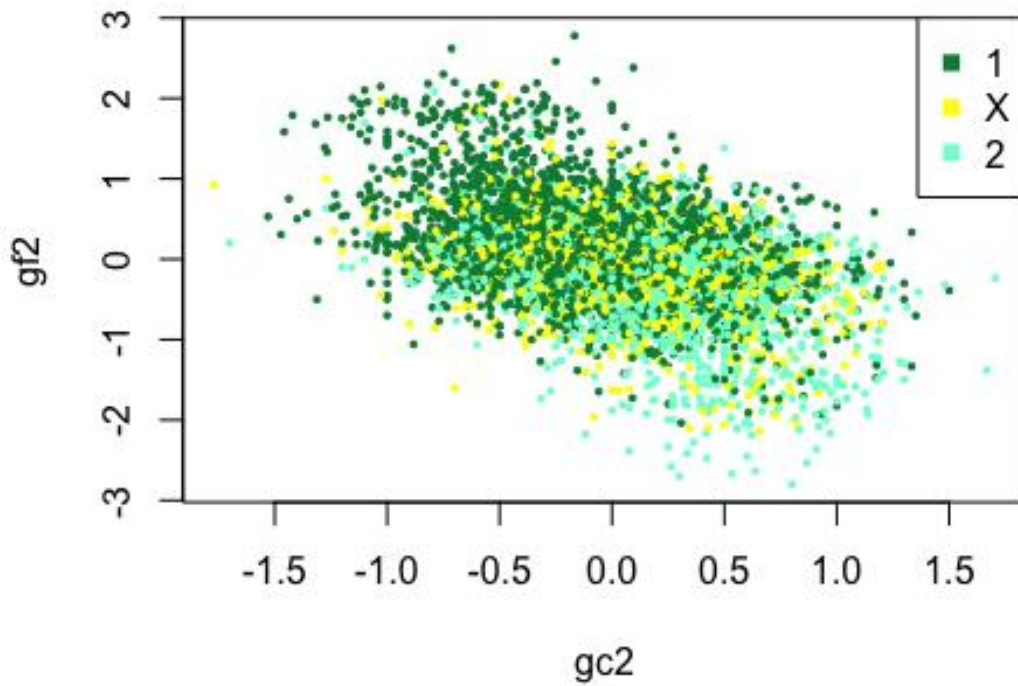
pp2 i guanyadesLV2 segons el resultat



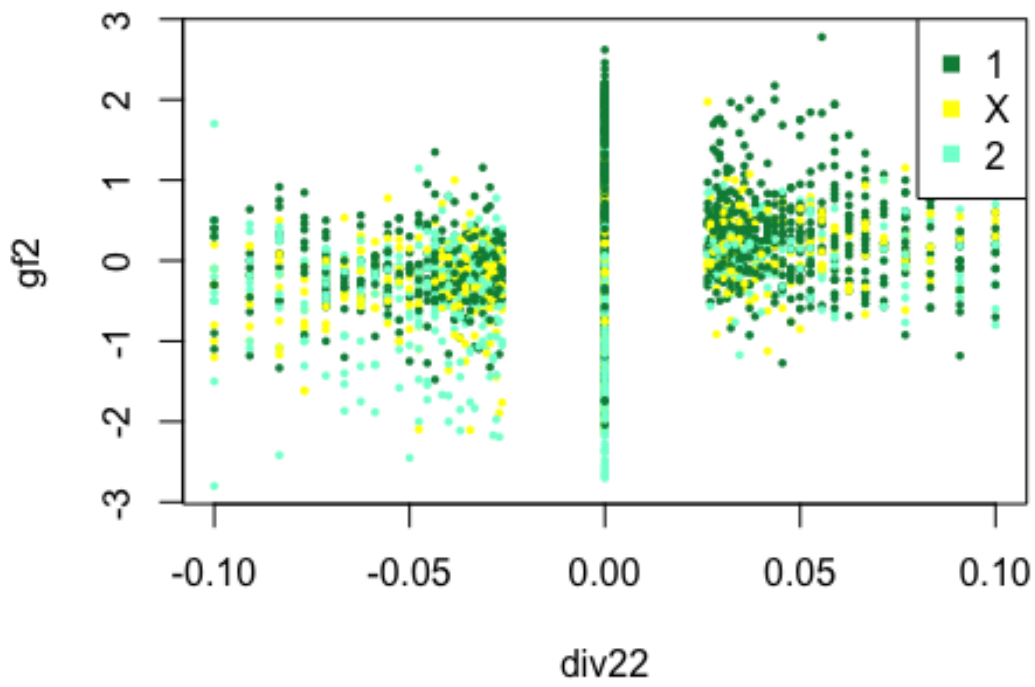
pp2 i quiniela2 segons el resultat



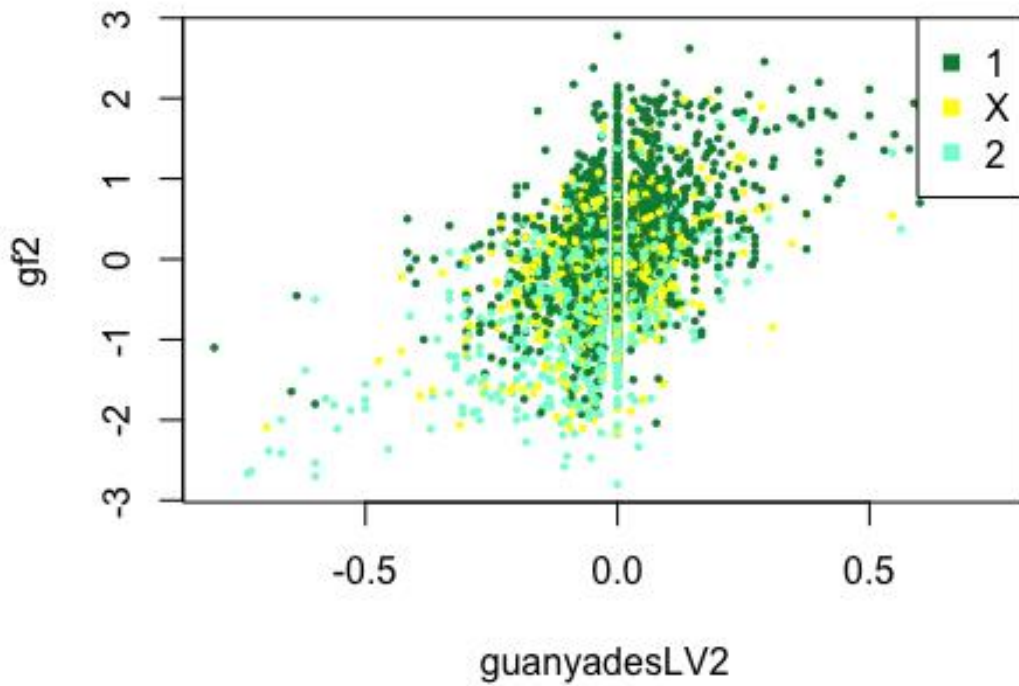
gf2 i gc2 segons el resultat



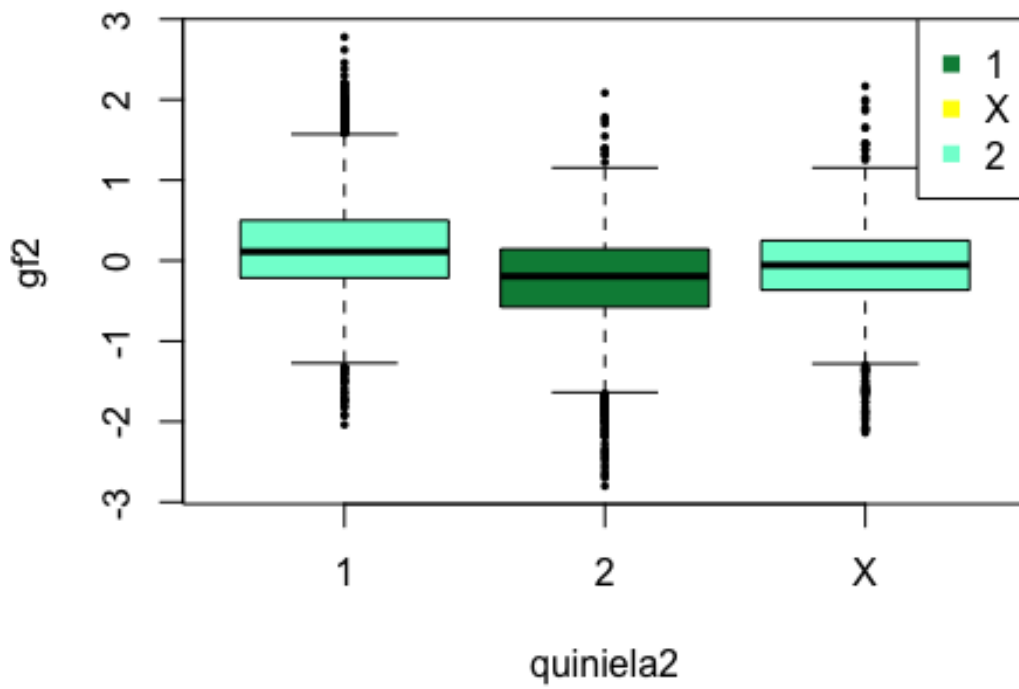
gf2 i div22 segons el resultat



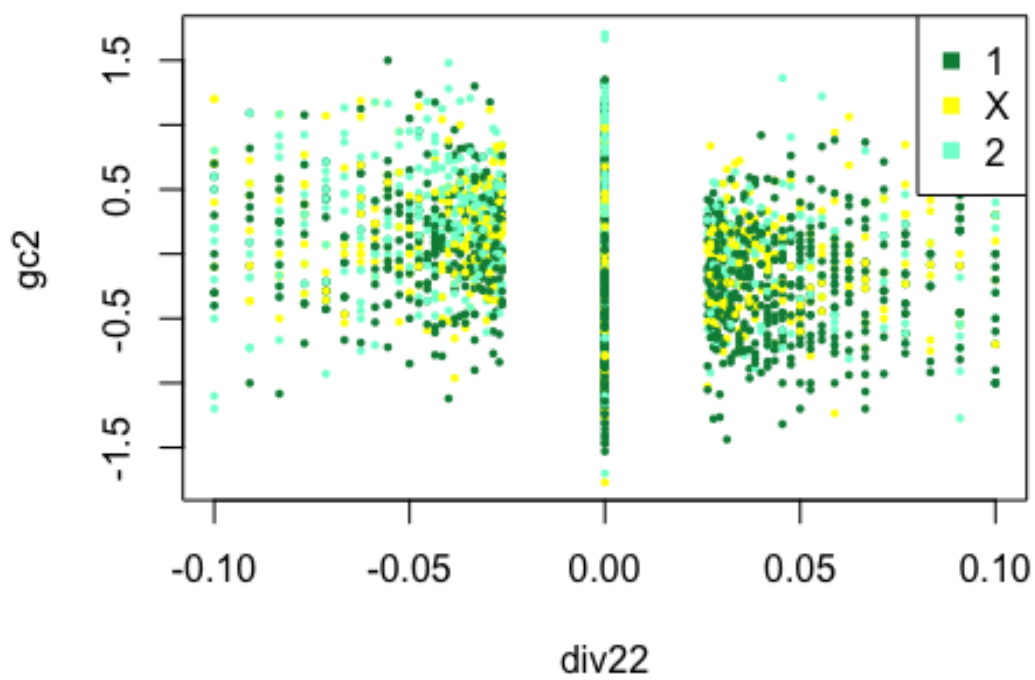
gf2 i guanyadesLV2 segons el resultat



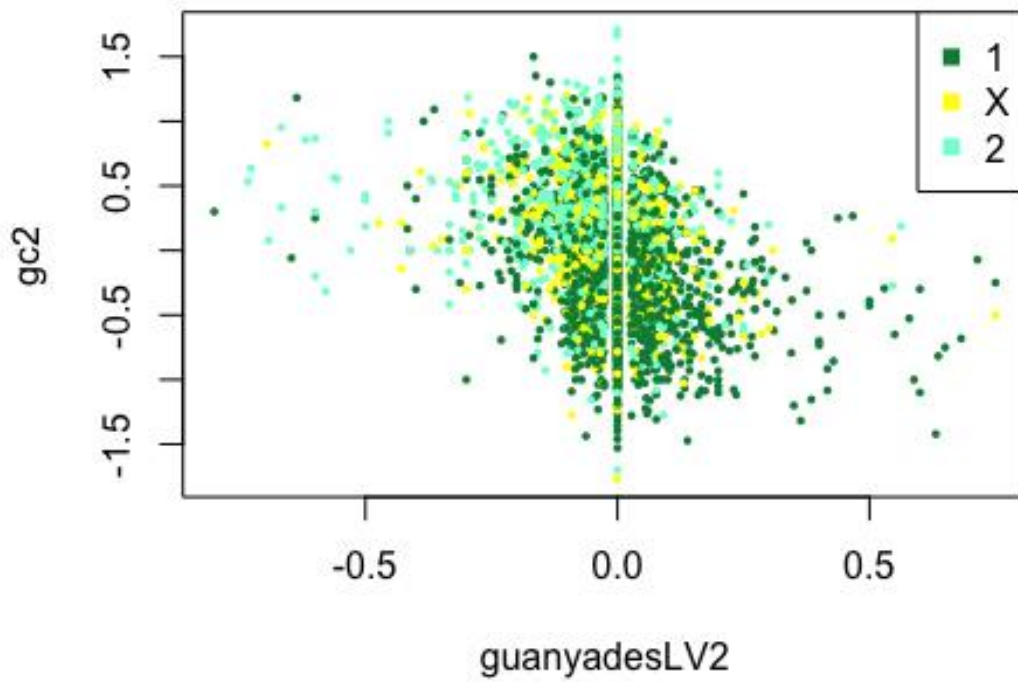
gf2 i quiniela2 segons el resultat



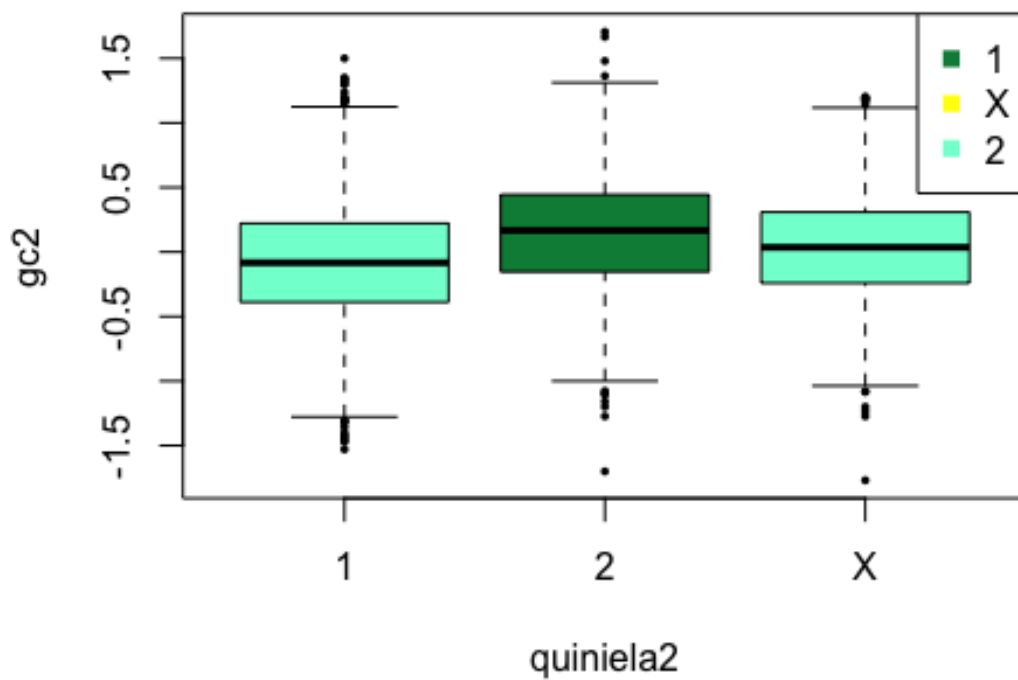
gc2 i div22 segons el resultat



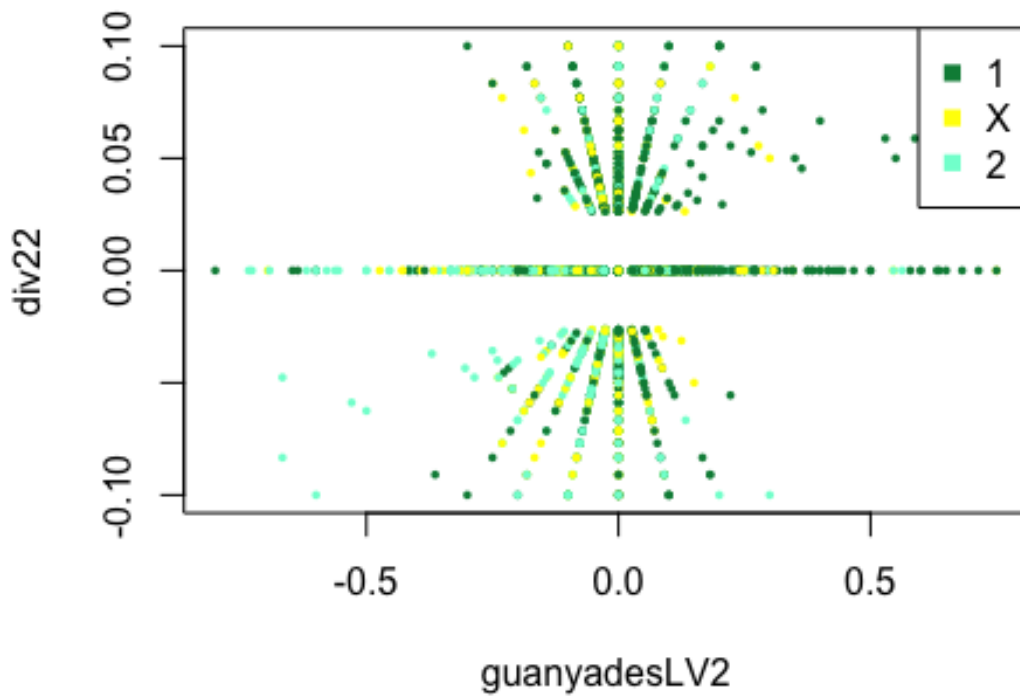
gc2 i guanyadesLV2 segons el resultat



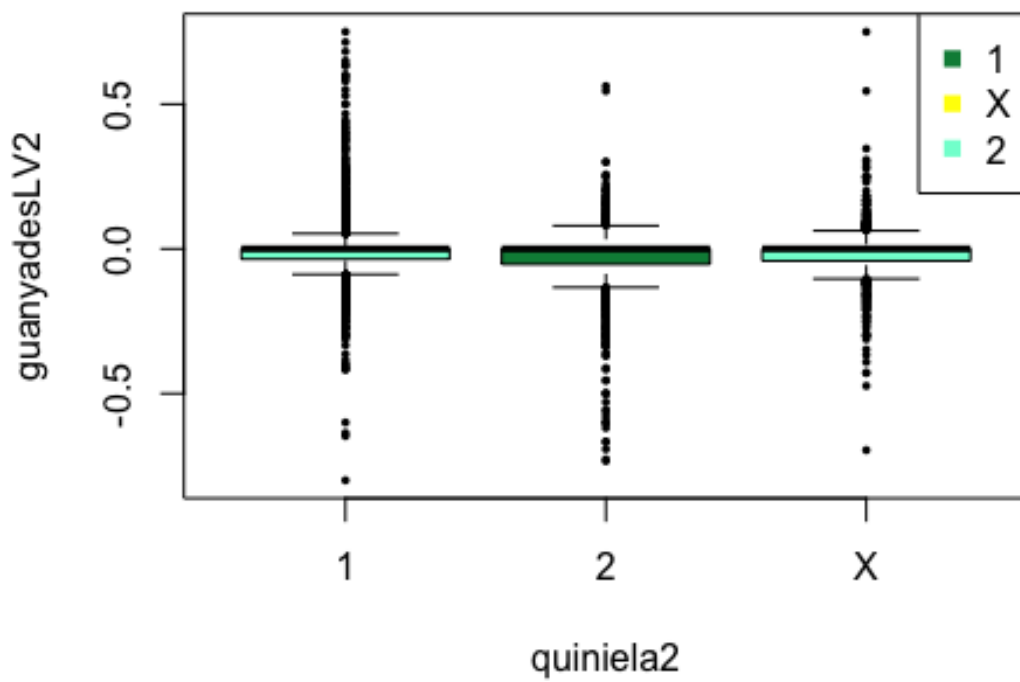
gc2 i quiniela2 segons el resultat



div22 i guanyadesLV2 segons el resultat



guanyadesLV2 i quiniela2 segons el resultat



3. Formules dels models

Model Complet:

$$\begin{aligned}h_1 = & 1,5874683 - 0,0434896 * X_{posL} - 0,0175776 * X_{ptsL} - 0,0660982 * X_{pgL} \\ & - 0,0323776 * X_{peL} + 0,0329579 * X_{gfL} - 0,0113105 * X_{gcL} \\ & + 0,0249017 * X_{div2L} + 0,0385548 * X_{guanyadesL} - 0,0155148 * X_{posV} \\ & + 0,0226991 * X_{ptsvV} - 0,0167357 * X_{peV} - 0,0225018 * X_{ppV} \\ & - 0,0224161 * X_{gfV} + 0,0151590 * X_{gcV} - 0,2299985 * X_{div2V} \\ & - 0,0511168 * X_{guanyadesV}\end{aligned}$$

$$\begin{aligned}h_2 = & 0,6230134 - 0,0498510 * X_{posL} - 0,0126844 * X_{ptsL} - 0,0585917 * X_{pgL} \\ & - 0,0101332 * X_{peL} - 0,0067987 * X_{gfL} + 0,0138313 * X_{gcL} \\ & - 0,1133916 * X_{div2L} - 0,0041304 * X_{guanyadesL} - 0,0005927 * X_{posV} \\ & + 0,0220570 * X_{ptsvV} - 0,0071818 * X_{peV} - 0,0153183 * X_{ppV} \\ & + 0,0156402 * X_{gfV} - 0,0110045 * X_{gcV} + 0,1246220 * X_{div2V} \\ & - 0,0042889 * X_{guanyadesV}\end{aligned}$$

Model Diferències:

$$\begin{aligned}h_1 = & 0,776976 - 0,012995 * X_{pos1} + 0,021631 * X_{pp1} - 0,015098 * X_{ptsLV} \\ & + 0,009395 * X_{pe1} + 0,026562 * X_{gf1} - 0,012141 * X_{gc1} + 0,130511 \\ & * X_{div21} + 0,041111 * X_{guanyadesLV1}\end{aligned}$$

$$\begin{aligned}h_2 = & 0,065410 - 0,028055 * X_{pos1} + 0,035584 * X_{pp1} - 0,012500 * X_{ptsLV} \\ & + 0,036590 * X_{pe1} - 0,011572 * X_{gf1} + 0,011945 * X_{gc1} - 0,107199 \\ & * X_{div21} - 0,003369 * X_{guanyadesLV1}\end{aligned}$$

Model Diferències i Dividit:

$$\begin{aligned}h_1 = & 0,75601 - 0,14115 * X_{pos2} - 0,49509 * X_{pp2} - 0,33613 * X_{ptsLV2} + 0,52194 \\ & * X_{pe2} + 0,63723 * X_{gf2} - 0,29346 * X_{gc2} + 2,38016 * X_{div22} + 0,07992 \\ & * X_{guanyadesLV2}\end{aligned}$$

$$\begin{aligned}h_2 = & 0,05056 - 0,74272 * X_{pos2} + 1,02900 * X_{pp2} - 0,28955 * X_{ptsLV2} + 0,91119 \\ & * X_{pe2} - 0,33846 * X_{gf2} + 0,33124 * X_{gc2} - 1,81430 * X_{div22} - 0,07788 \\ & * X_{guanyadesLV2}\end{aligned}$$

3. Discriminació

El procediment utilitzat en les següents taules és similar al de la taula 3.5 de l'apartat 3.3. Discriminació del capítol III. RESULTATS.

Encert: 51,18%

<u>Probabilitat guany equip visitant</u>	<u>Probabilitat d'empat</u>					
	0-0,1	0,1-0,2	0,2-0,3	0,3-0,4	0,4-1	Total
0-0,1	7 1 0	166 16 13	0	0	0	173 17 13
0,1-0,2	0	131 27 18	560 218 186	0	0	691 245 204
0,2-0,3	0	3 0 2	1018 587 518	12 7 4	0	1033 594 524
0,3-0,4	0	0	380 252 316	28 15 18	0	408 267 334
<u>0,4-0,5</u>	<u>0</u>	<u>0</u>	<u>72</u> <u>62</u> <u>104</u>	<u>8</u> <u>5</u> <u>9</u>	<u>0</u>	<u>80</u> <u>67</u> <u>113</u>
<u>0,5-0,6</u>	<u>0</u>	<u>0</u>	<u>15</u> <u>19</u> <u>50</u>	<u>0</u>	<u>0</u>	<u>15</u> <u>19</u> <u>50</u>
<u>0,6-0,7</u>	<u>0</u>	<u>0</u>	<u>10</u> <u>14</u> <u>45</u>	<u>0</u>	<u>0</u>	<u>10</u> <u>14</u> <u>45</u>
<u>0,7-0,8</u>	<u>0</u>	<u>1</u> <u>1</u> <u>6</u>	<u>0</u> <u>2</u> <u>4</u>	<u>0</u>	<u>0</u>	<u>1</u> <u>3</u> <u>10</u>
0,8- 1	0	0	0	0	0	0
Total	7 1 0	301 44 39	2055 1154 1223	48 27 31	0	4930

Encert: 50,97%

<u>Probabilitat d'empat</u>	<u>Probabilitat guany equip local</u>										
	0-0,1	0,1-0,2	0,2-0,3	0,3-0,4	0,4-0,5	0,5-0,6	0,6-0,7	0,7-0,8	0,8-0,9	0,9-1	Total
0-0,1	<u>0</u>	<u>0</u>	<u>0</u>	0	0	0	0	0	5	2	7
									1	0	1
									0	0	0
0,1-0,2	<u>1</u>	<u>0</u>	<u>0</u>	0	0	3	90	134	73	0	691
	<u>1</u>					0	22	15	6		245
	<u>6</u>					2	16	12	3		204
0,2-0,3	<u>0</u>	<u>13</u>	<u>58</u>	257	670	752	305	0	0	0	1033
	<u>4</u>	<u>25</u>	<u>48</u>	175	419	376	107				594
	<u>9</u>	<u>69</u>	<u>92</u>	234	436	306	77				524
0,3-0,4	<u>0</u>	<u>1</u>	<u>12</u>	<u>26</u>	9	0	0	0	0	0	408
		<u>0</u>	<u>6</u>	<u>14</u>	7						267
		<u>0</u>	<u>11</u>	<u>17</u>	3						334
0,4-1	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
Total	<u>1</u>	<u>14</u>	<u>70</u>	283	679	755	395	134	78	2	4930
	<u>5</u>	<u>25</u>	<u>54</u>	189	426	376	129	15	7	0	
	<u>15</u>	<u>69</u>	<u>103</u>	251	439	308	93	12	3	0	

Utilitzant 3 dimensions (Probabilitat d'empat $\leq 0,2$) (taula 1/3), Encert total = 51,20%

Probabilitat guany equip visitant	Probabilitat guany equip local							
	0-0,1	0,1-0,5	0,5-0,6	0,6-0,7	0,7-0,8	0,8-0,9	0,9-1	Total
0-0,1	0	0	0	0	93 10 10	78 7 3	2 0 0	173 17 13
0,1-0,2	0	0	0	90 22 16	41 5 2	0	0	131 27 18
0,2-0,3	0	0	3 0 2	0	0	0	0	3 0 2
0,3-0,7	0	0	0	0	0	0	0	0
0,7-0,8	1 1 6	0	0	0	0	0	0	1 1 6
0,8-1	0	0	0	0	0	0	0	0
Total	1 1 6	0	3 0 2	90 22 16	134 15 12	78 7 3	2 0 0	392

Utilitzant 3 dimensions ($0,2 < \text{Probabilitat d'empat} \leq 0,3$) (taula 2/3)

<u>Probabilitat guany equip visitant</u>	<u>Probabilitat guany equip local</u>								
	0-0,1	0,1-0,2	0,2-0,3	0,3-0,4	0,4-0,5	0,5-0,6	0,6-0,7	0,7-1	Total
0-0,1	0	0	0	0	0	0	0	0	0
0,1-0,2	0	0	0	0	0	255 111 109	305 107 77	0	560 218 186
0,2-0,3	0	0	0	0	521 322 321	497 265 197	0	0	1018 587 518
0,3-0,4	0	0	0	231 155 201	149 97 115	0	0	0	380 252 316
0,4-0,5	0	0	46 42 71	26 20 33	0	0	0	0	72 62 104
0,5-0,6	0	3 13 29	12 6 21	0	0	0	0	0	15 19 50
0,6-0,7	0 2 5	10 12 40	0	0	0	0	0	0	10 14 45
0,7-0,8	0 2 4	0	0	0	0	0	0	0	0 2 4
0,8-1	0	0	0	0	0	0	0	0	0
Total	0 4 9	13 25 69	58 48 92	257 175 234	670 419 436	752 376 306	305 107 77	0	4432

Utilitzant 3 dimensions (0,3 < Probabilitat d'empat ≤ 1) (taula 3/3)

<u>Probabilitat guany equip visitant</u>	<u>Probabilitat guany equip local</u>						
	0-0,1	0,1-0,2	0,2-0,3	0,3-0,4	0,4-0,5	0,5-1	Total
0-0,2	0	0	0	0	0	0	0
0,2-0,3	0	0	0	3 0 1	9 3 7	0	12 3 8
0,3-0,4	0	0	5 1 2	23 14 16	0	0	28 15 18
0,4-0,5	0	1 0 0	7 5 9	0	0	0	8 5 9
0,5-1	0	0	0	0	0	0	0
Total	0	1 0 0	12 6 11	26 14 17	9 3 7	0	106

4. Codi R

```
#####  
##### PACKAGES #####  
#####  
library(dplyr)  
library(sqldf)  
library(ROCR)  
library(mgcv)  
library(sjPlot)  
library(sjmisc)  
library(AER)  
library(car)  
library(generalhoslem)  
library(pROC)  
library(stats)
```

```
#####  
##### CREACIO BASE DE DADES #####  
#####
```

```
# Resultats Jornades
```

```
jornades <- read.csv2("tfg1.csv")
```

```
# Classificacio
```

```
jornada <- rep(0,20)
```

```
posicio <- 1:20
```

```
equip <- as.character(read.csv2("equipps.csv")[1:20,2])
```

```
pts <- rep(0,20)
```

```
ptscomL <- rep(0,20)
```

```
ptscomV <- rep(0,20)
```

```
pj <- rep(0,20)
```

```
pg <- rep(0,20)
```

```
pe <- rep(0,20)
```

```
pp <- rep(0,20)
```

```
gf <- rep(0,20)
```

```
gc <- rep(0,20)
```

classificacio

<-

```
data.frame(jornada=jornada,posicio=posicio,Local=equip,Visitante=equip,pts=pts  
          ,pj=pj,pg=pg,pe=pe,pp=pp,gf=gf,gc=gc)
```

```
posL <- rep(0,11564)
```

```
ptsL <- rep(0,11564)
```

```
ptsL <- rep(0,11564)
```

```
pjL <- rep(0,11564)
```

```
pgL <- rep(0,11564)
```

```
peL <- rep(0,11564)
```

```
ppL <- rep(0,11564)
```

```
gfL <- rep(0,11564)
```

```
gcL <- rep(0,11564)
```

```
div2L <- rep(0,11564)
```

```
posV <- rep(0,11564)
```

```
ptsV <- rep(0,11564)
```

```
ptsV <- rep(0,11564)
```

```
pjV <- rep(0,11564)
```

```
pgV <- rep(0,11564)
```

```
peV <- rep(0,11564)
```

```
ppV <- rep(0,11564)
```

```
gfV <- rep(0,11564)
```

```
gcV <- rep(0,11564)
```

```
gfV <- rep(0,11564)
```

```
gcV <- rep(0,11564)
```

```
div2V <- rep(0,11564)
```

```
pos1 <- rep(0,11564)
```

```
pts1 <- rep(0,11564)
```

```
pj1 <- rep(0,11564)
```

```
pg1 <- rep(0,11564)
```

```
pe1 <- rep(0,11564)
```

```
pp1 <- rep(0,11564)
```

```
gf1 <- rep(0,11564)
```

```
gc1 <- rep(0,11564)
```

```
gf1 <- rep(0,11564)
```

```
gc1 <- rep(0,11564)
```

```
div21 <- rep(0,11564)
```



```

ptsL <- rep(0,11564)
ptsvV <- rep(0,11564)
quini
data.frame(jornades,posL=posL,ptsL=ptsL,pjL=pjL,pgL=pgL,peL=peL,ppL=ppL,gfL=gfL,gcL=gcl
,
div2L=div2L,posV=posV,ptsV=ptsV,pjV=pjV,pgV=pgV,peV=peV,ppV=ppV,gfV=gfV,gcV=gcV,
div2V=div2V,pos1=pos1,pts1=pts1,pj1=pj1,pg1=pg1,pe1=pe1,pp1=pp1,gf1=gf1,gc1=gc1,
div21=div21,ptsL=ptsL,ptsvV=ptsvV)

partit <- 3965
jornada1 <- 1

anys <- 2
Pequip <- 205
Fequip <- 224

while (partit <=11564){
jornada1 <- 1
jornada <- rep(0,20)
posicio <- 1:20

equip <- read.csv2("equipps.csv")[Pequip:Fequip,]
equipasat <- read.csv2("equipps.csv")[:(Pequip-20):(Fequip-20),]
temporada <- as.character(read.csv2("equipps.csv")[:(Pequip-20),1])
s <- which (as.character( arrange(merge(x = equip, y = equipasat, by = "Equipo", all.x =
TRUE),classificacio.x )[, 'Temporada.y']) == temporada)
div2 <- rep(0,20)
div2[s] <- 1

equip <- as.character(read.csv2("equipps.csv")[Pequip:Fequip,3])
pts <- rep(0,20)
pj <- rep(0,20)
pg <- rep(0,20)
pe <- rep(0,20)
pp <- rep(0,20)
gf <- rep(0,20)
gc <- rep(0,20)

```

```

ptscomL <- rep(0,20)
ptscomV <- rep(0,20)
guanyades <- rep(0,20)
classificacio <-
data.frame(jornada=jornada,posicio=posicio,Local=equip,Visitante=equip,pts=pts
, pj=pj,pg=pg,pe=pe,pp=pp,gf=gf,gc=gc,div2=div2,ptscomL=ptscomL,
ptscomV=ptscomV,guanyades=guanyades)#atac=atac,mig=mig,defensa=defensa

while (jornada1 <=38 ){
  while(jornada1==quini[partit,'Jornada'] && partit <=11564){
    quini[partit,'posL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'posicio']
    quini[partit,'ptsL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'pts']
    quini[partit,'pjL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'pj']
    quini[partit,'pgL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'pg']
    quini[partit,'peL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'pe']
    quini[partit,'ppL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'pp']
    quini[partit,'gfL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'gf']
    quini[partit,'gcl'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'gc']
    quini[partit,'div2L'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'div2']
    quini[partit,'ptsL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'ptscomL']
    quini[partit,'guanyadesL'] <- merge(x = quini[partit,], y = classificacio, by = "Local", all.x =
TRUE)[,'guanyades']

    quini[partit,'posV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'posicio']

```

```

quini[partit,'ptsV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'pts']
quini[partit,'pjV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'pj']
quini[partit,'pgV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'pg']
quini[partit,'peV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'pe']
quini[partit,'ppV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'pp']
quini[partit,'gfV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'gf']
quini[partit,'gcV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'gc']
quini[partit,'div2V'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'div2']
quini[partit,'ptsvV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'ptscomV']
quini[partit,'guanyadesV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x
= TRUE)[,'guanyades']
#quini[partit,'atacV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'atac']
#quini[partit,'defensaV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x
= TRUE)[,'mig']
#quini[partit,'migV'] <- merge(x = quini[partit,], y = classificacio, by = "Visitante", all.x =
TRUE)[,'defensa']

```

```

quini[partit,'pos1'] <- quini[partit,'posL'] - quini[partit,'posV']
quini[partit,'pts1'] <- quini[partit,'ptsL'] - quini[partit,'ptsV']
quini[partit,'pj1'] <- quini[partit,'pjL'] - quini[partit,'pjV']
quini[partit,'pg1'] <- quini[partit,'pgL'] - quini[partit,'pgV']
quini[partit,'pe1'] <- quini[partit,'peL'] - quini[partit,'peV']
quini[partit,'pp1'] <- quini[partit,'ppL'] - quini[partit,'ppV']
quini[partit,'gf1'] <- quini[partit,'gfL'] - quini[partit,'gfV']
quini[partit,'gc1'] <- quini[partit,'gcL'] - quini[partit,'gcV']
quini[partit,'div21'] <- quini[partit,'div2L'] - quini[partit,'div2V']
quini[partit,'ptsvV'] <- quini[partit,'ptsvL'] - quini[partit,'ptsvV']
quini[partit,'guanyadesLV1'] <- quini[partit,'guanyadesL'] - quini[partit,'guanyadesV']

```

```

quini[partit,'pos2'] <- (quini[partit,'posL'] - quini[partit,'posV'])/jornada1
quini[partit,'pts2'] <- (quini[partit,'ptsL'] - quini[partit,'ptsV'])/jornada1
quini[partit,'pj2'] <- (quini[partit,'pjL'] - quini[partit,'pjV'])/jornada1
quini[partit,'pg2'] <- (quini[partit,'pgL'] - quini[partit,'pgV'])/jornada1
quini[partit,'pe2'] <- (quini[partit,'peL'] - quini[partit,'peV'])/jornada1
quini[partit,'pp2'] <- (quini[partit,'ppL'] - quini[partit,'ppV'])/jornada1
quini[partit,'gf2'] <- (quini[partit,'gfL'] - quini[partit,'gfV'])/jornada1
quini[partit,'gc2'] <- (quini[partit,'gcL'] - quini[partit,'gcV'])/jornada1
quini[partit,'div22'] <- (quini[partit,'div2L'] - quini[partit,'div2V'])/jornada1
quini[partit,'ptsLV2'] <- (quini[partit,'ptsL'] - quini[partit,'ptsV'])/jornada1
quini[partit,'guanyadesLV2'] <- (quini[partit,'guanyadesL'] -
quini[partit,'guanyadesV'])/jornada1

```

```
# Actualitzar classificacio
```

```
i <- 1
```

```
trobat <- 0
```

```
buscarL <- quini[partit,'Local']
```

```
buscarV <- quini[partit,'Visitante']
```

```
while (trobat < 2){
```

```
  if(levels(buscarL)[i] == buscarL){
```

```
    trobat <- trobat + 1
```

```
    iL <- which (classificacio[, 'Local'] == levels(buscarL)[i])
```

```
  }
```

```
  if(levels(buscarV)[i]==buscarV){
```

```
    trobat <- trobat + 1
```

```
    iV <- which (classificacio[, 'Visitante'] == levels(buscarV)[i])
```

```
  }
```

```
  i <- i + 1
```

```
}
```

```
classificacio[iL,'pj'] <- classificacio[iL,'pj'] + 1
```

```
classificacio[iL,'gf'] <- classificacio[iL,'gf'] + merge(x = quini[partit,],
```

```

        y = classificacio, by = "Visitante", all.x =
TRUE)['GolesLocal']
    classificacio[iL,'gc'] <- classificacio[iL,'gc'] + merge(x = quini[partit,],
        y = classificacio, by = "Visitante", all.x =
TRUE)['GolesVisitante']

    classificacio[iV,'pj'] <- classificacio[iV,'pj'] + 1
    classificacio[iV,'gf'] <- classificacio[iV,'gf'] + merge(x = quini[partit,],
        y = classificacio, by = "Visitante", all.x =
TRUE)['GolesVisitante']
    classificacio[iV,'gc'] <- classificacio[iV,'gc'] + merge(x = quini[partit,],
        y = classificacio, by = "Visitante", all.x =
TRUE)['GolesLocal']

```

```

if (quini[partit,'GolesLocal'] > quini[partit,'GolesVisitante']){
    classificacio[iL,'pg'] <- classificacio[iL,'pg'] + 1
    classificacio[iL,'pts'] <- classificacio[iL,'pts'] + 3
    classificacio[iV,'pp'] <- classificacio[iV,'pp'] + 1
    classificacio[iL,'ptscomL'] <- classificacio[iL,'ptscomL'] + 3
    classificacio[iL,'guanyades'] <- classificacio[iL,'guanyades'] + 1
    classificacio[iV,'guanyades'] <- 0

```

```

    quini[partit,'guanyatL'] <- 1
    quini[partit,'empat'] <- 0
    quini[partit,'guanyatV'] <- 0
}

```

```

if (quini[partit,'GolesLocal'] < quini[partit,'GolesVisitante']){
    classificacio[iV,'pg'] <- classificacio[iV,'pg'] + 1
    classificacio[iV,'pts'] <- classificacio[iV,'pts'] + 3
    classificacio[iL,'pp'] <- classificacio[iL,'pp'] + 1
    classificacio[iV,'ptscomV'] <- classificacio[iV,'ptscomV'] + 3
    classificacio[iV,'guanyades'] <- classificacio[iV,'guanyades'] + 1
    classificacio[iL,'guanyades'] <- 0

```

```

    quini[partit,'guanyatL'] <- 0
    quini[partit,'empat'] <- 0

```

```

quini[partit,'guanyatV'] <- 1
}
if (quini[partit,'GolesLocal'] == quini[partit,'GolesVisitante']){
  classificacio[iV,'pe'] <- classificacio[iV,'pe'] + 1
  classificacio[iV,'pts'] <- classificacio[iV,'pts'] + 1
  classificacio[iL,'pe'] <- classificacio[iL,'pe'] + 1
  classificacio[iL,'pts'] <- classificacio[iL,'pts'] + 1
  classificacio[iL,'ptscomL'] <- classificacio[iL,'ptscomL'] + 1
  classificacio[iV,'ptscomV'] <- classificacio[iV,'ptscomV'] + 1
  classificacio[iL,'guanyades'] <- 0
  classificacio[iV,'guanyades'] <- 0

  quini[partit,'guanyatL'] <- 0
  quini[partit,'empat'] <- 1
  quini[partit,'guanyatV'] <- 0
}
classificacio <- arrange(classificacio, desc(pts)) #ordenar la classificaio per punts
classificacio[, 'posicio'] <- 1:20
classificacio[, 'jornada'] <- rep(jornada1,20)
partit <- partit + 1
}

jornada1 <- jornada1 + 1
}

anys <- anys + 1
Pequip <- Pequip + 20
Fequip <- Fequip + 20

}
classificacio

quini <- quini[3965:11564,]

i <- 1

```

```

while ( i <= nrow(quini)){

  if ( quini[i,'GolesLocal'] > quini[i,'GolesVisitante'] ) {
    quini[i,'quiniela'] <- 1

  } else if ( quini[i,'GolesLocal'] == quini[i,'GolesVisitante'] ){
    quini[i,'quiniela'] <- 0

  } else if ( quini[i,'GolesLocal'] < quini[i,'GolesVisitante'] ){
    quini[i,'quiniela'] <- 2
  }

  i <- i + 1

}

subquini1 <- subset(quini, Jornada >= 10)

i <- 1

while ( i <= nrow(subquini1)){

  if ( subquini1[i,'GolesLocal'] > subquini1[i,'GolesVisitante'] ) {
    subquini1[i,'quiniela2'] <- '1'
    subquini1[i,'quiniela3'] <- 'springgreen4'

  } else if ( subquini1[i,'GolesLocal'] == subquini1[i,'GolesVisitante'] ){
    subquini1[i,'quiniela2'] <- 'X'
    subquini1[i,'quiniela3'] <- 'yellow'

  } else if ( subquini1[i,'GolesLocal'] < subquini1[i,'GolesVisitante'] ){
    subquini1[i,'quiniela2'] <- '2'
    subquini1[i,'quiniela3'] <- 'aquamarine'
  }

  i <- i + 1

}

```

```

subquini1[, 'quiniela2'] <- factor(subquini1[, 'quiniela2'])
subquini1grafics <- subquini1[, c('pos2', 'pts2', 'ptsLV2', 'pg2', 'pe2', 'pp2', 'gf2', 'gc2', 'div22',
'guanyadesLV2', 'quiniela2', 'quiniela3')]

```

```

#####
##### DESCRIPTIVA UNIVARIANT #####
#####

```

```

summary(subquini1)
subquini1$Temporada <- factor(subquini1$Temporada)
subquini1$Jornada <- factor(subquini1$Jornada)
subquini1$Local <- factor(subquini1$Local)
subquini1$Visitante <- factor(subquini1$Visitante)
subquini1$quiniela2 <- factor(subquini1$quiniela2)
descr(subquini1[, 6:55], out="viewer")
sjt.frq(subquini1[, c(1, 2, 56)], title='quiniela2')

```

```

i <- 6
while (i <= 55) {
  hist(subquini1[, i], main = paste("Histograma de", names(subquini1)[i]), xlab=
names(subquini1)[i])
  i <- i + 1
}

```

```

pie(table(subquini1$quiniela2), col=c("springgreen4", "aquamarine", "yellow"))
title ('Resultats Quiniela')

```

```

#####
##### DESCRIPTIVA BIVARIANT #####
#####

```

```

subquini1grafics <- subquini1[, c('pos2', 'pts2', 'ptsLV2', 'pg2', 'pe2', 'pp2', 'gf2', 'gc2', 'div22',
'guanyadesLV2', 'quiniela2', 'quiniela3')]

```

```

i <- 1
c <- data.frame()

```



```

k <- 1
while (i <= (ncol (subquini1grafics)-2) ){

  j <- i + 1

  while (j <= (ncol (subquini1grafics) -1 )){

    plot(subquini1grafics[,i] ~ subquini1grafics[,j], xlab = names(subquini1grafics)[j], ylab =
names(subquini1grafics)[i],
      pch = 16, col = subquini1grafics['quiniela3'], cex = 0.5)
    legend("topright", legend=c("1","X","2"), col = c("springgreen4", "yellow","aquamarine"),
pch = 15)
    title(paste(names(subquini1grafics)[i],'i',names(subquini1grafics)[j],'segons el resultat'))

    c[k,'i'] <- i
    c[k,'j'] <- j
    j <- j + 1
    k <- k + 1
  }
  i <- 1 + i
}

#####
##### MODELS #####
#####

subquini1.1 <- subquini1 [1:4930,]

#####
##### Model Complet #####
#####

subquini1.3 <- subquini1 [1:4930,]
dades3 <- subquini1.3 [,c('posL', 'ptsL','ptsLL' , 'pgL' , 'peL' , 'ppL' , 'gfl' , 'gcl' , 'div2L'
,'guanyadesL', 'posV', 'ptsV','ptsvV' , 'pgV' , 'peV' , 'ppV' , 'gfv' , 'gcv' , 'div2V'
,'guanyadesV','quiniela')]

model1 <- gam(list(quiniela ~
  posL + ptsLL + pgL + peL + gfl + gcl + div2L + guanyadesL + posV + ptsvV
+ peV + ppV + gfv + gcv + div2V + guanyadesV,

```

```

~ posL + ptslL + pgL + peL + gfl + gcL + div2L + guanyadesL + posV + ptsvV
+ peV + ppV + gfV + gcV + div2V + guanyadesV),family=multinom(K=2), data=dades3)

```

```
summary(model1)
```

```
AIC(model1)
```

```
pred.test.3 <- predict(model1,newdata=dades3 ,type="response")
```

```
dades3['Prob1'] <- pred.test.3[,2]
```

```
dades3['ProbX'] <- pred.test.3[,1]
```

```
dades3['Prob2'] <- pred.test.3[,3]
```

```
i <- 1
```

```
while ( i <= nrow(subquini1.3)){
```

```

if ( pred.test.3[i,2] > pred.test.3[i,1] && pred.test.3[i,2] > pred.test.3[i,3] ){
  subquini1.3[i,'PrediccioQuiniela'] <- 1

```

```

} else if ( pred.test.3[i,1] > pred.test.3[i,2] && pred.test.3[i,1] > pred.test.3[i,3]){
  subquini1.3[i,'PrediccioQuiniela'] <- 0

```

```

} else if ( pred.test.3[i,3] > pred.test.3[i,1] && pred.test.3[i,3] > pred.test.3[i,2]){
  subquini1.3[i,'PrediccioQuiniela'] <- 2
}

```

```
i <- i + 1
```

```
}
```

```
head(subquini1.3)
```

```
i <- 1
```

```
while ( i <= nrow(subquini1.3)){
```

```

if (subquini1.3[i,'quiniela']==subquini1.3[i,'PrediccioQuiniela']){
  subquini1.3[i,'EncertQuiniela'] <- 1

```

```

} else {
  subquini1.3[i,'EncertQuiniela'] <- 0
}
i <- i + 1
}

mean(subquini1.3['EncertQuiniela'])

roc.multi3 <- multiclass.roc(subquini1.3$quiniela, subquini1.3$PrediccioQuiniela, levels=c(0,
1, 2))
roc.multi3

#####
##### Model Diferències #####
#####
subquini1.2 <- subquini1 [1:4930,]
dades2 <- subquini1.2 [,c('pos1', 'pts1','ptsLV' , 'pg1' , 'pe1' , 'pp1' , 'gf1' , 'gc1' , 'div21'
,'guanyadesLV1','quiniela')]
model2 <- gam(list(quiniela ~
  pos1 + pp1 + ptsLV + pe1 + gf1 + gc1 + div21 + guanyadesLV1,
  ~ pos1 + pp1 + ptsLV + pe1 + gf1 + gc1 + div21 +
guanyadesLV1),family=multinom(K=2), data=dades2)

summary(model2)
AIC(model2)

pred.test.2 <- predict(model2,newdata=dades2 ,type="response")
dades2['Prob1'] <- pred.test.2[,2]
dades2['ProbX'] <- pred.test.2[,1]
dades2['Prob2'] <- pred.test.2[,3]

i <- 1
while ( i <= nrow(subquini1.2)){

if ( pred.test.2[i,2] > pred.test.2[i,1] && pred.test.2[i,2] > pred.test.2[i,3] ){
  subquini1.2[i,'PrediccioQuiniela'] <- 1

} else if ( pred.test.2[i,1] > pred.test.2[i,2] && pred.test.2[i,1] > pred.test.2[i,3]){

```

```

subquini1.2[i,'PrediccioQuiniela'] <- 0

} else if ( pred.test.2[i,3] > pred.test.2[i,1] && pred.test.2[i,3] > pred.test.2[i,2]){
  subquini1.2[i,'PrediccioQuiniela'] <- 2
}

i <- i + 1

}

i <- 1
while ( i <= nrow(subquini1.2)){
  if (subquini1.2[i,'quiniela']==subquini1.2[i,'PrediccioQuiniela']){
    subquini1.2[i,'EncertQuiniela'] <- 1
  } else {
    subquini1.2[i,'EncertQuiniela'] <- 0
  }
  i <- i + 1
}

mean(subquini1.2[, 'EncertQuiniela'])

roc.multi2 <- multiclass.roc(subquini1.2$quiniela, subquini1.2$PrediccioQuiniela, levels=c(0,
1, 2))
roc.multi2

#####
##### Model Diferències i Dividit #####
#####
subquini1.0 <- subquini1.1
dades0 <- subquini1.1 [,c('pos2', 'pts2','ptsLV2' , 'pg2' , 'pe2' , 'pp2' , 'gf2' , 'gc2' , 'div22'
, 'guanyadesLV2', 'quiniela', 'quiniela2', 'Jornada', 'Temporada', 'guanyatL', 'empat', 'guanyatV')]

model3 <- gam(list(quiniela ~
  pos2 + pp2 + ptsLV2 + pe2 +gf2 + gc2 + div22 + guanyadesLV2,
  ~ pos2 + pp2 + ptsLV2 + pe2 +gf2 + gc2 + div22 +
guanyadesLV2),family=multinom(K=2), data=dades0)

```

```
summary(model3)
```

```
AIC(model3)
```

```
pred.test.0 <- predict(model3,newdata=dades0 ,type="response")
```

```
dades0['Prob1'] <- pred.test.0[,2]
```

```
dades0['ProbX'] <- pred.test.0[,1]
```

```
dades0['Prob2'] <- pred.test.0[,3]
```

```
i <- 1
```

```
while ( i <= nrow(subquini1.0)){
```

```
  if ( pred.test.0[i,2] > pred.test.0[i,1] && pred.test.0[i,2] > pred.test.0[i,3] ){  
    subquini1.0[i,'PrediccioQuiniela'] <- 1
```

```
  } else if ( pred.test.0[i,1] > pred.test.0[i,2] && pred.test.0[i,1] > pred.test.0[i,3]){  
    subquini1.0[i,'PrediccioQuiniela'] <- 0
```

```
  } else if ( pred.test.0[i,3] > pred.test.0[i,1] && pred.test.0[i,3] > pred.test.0[i,2]){  
    subquini1.0[i,'PrediccioQuiniela'] <- 2  
  }
```

```
  i <- i + 1
```

```
}
```

```
i <- 1
```

```
while ( i <= nrow(subquini1.0)){
```

```
  if (subquini1.0[i,'quiniela']==subquini1.0[i,'PrediccioQuiniela']){  
    subquini1.0[i,'EncertQuiniela'] <- 1
```

```
  } else {  
    subquini1.0[i,'EncertQuiniela'] <- 0
```

```
  }
```

```
  i <- i + 1
```

```
}
```

```
mean(subquini1.0['EncertQuiniela'])
```

```
roc.multi0 <- multiclass.roc(subquini1.0$quiniela, subquini1.0$PrediccioQuiniela, levels=c(0,
1, 2))
roc.multi0
```

```
#####
##### MODEL 4 #####
#####
```

```
dades1 <- subquini1.1 [,c('pos2', 'pts2','ptsLV2' , 'pg2' , 'pe2' , 'pp2' , 'gf2' , 'gc2' , 'div22'
, 'guanyadesLV2', 'quiniela', 'quiniela2', 'Jornada', 'Temporada', 'guanyatL', 'empat', 'guanyatV')]
```

```
##### Correlacio #####
round(cor(dades1[,c('pos2', 'pts2', 'ptsLV2', 'pg2', 'pp2', 'gf2', 'gc2', 'pe2')]),2)
```

```
##### Model #####
model4 <- gam(list(quiniela ~
ptsLV2+gf2 + gc2,
~ pos2+ gf2 + gc2),family=multinom(K=2), data=dades1)
summary(model4)
AIC(model4)
```

```
pred.test <- predict(model4,newdata=dades1 ,type="response")
dades1[, 'Prob1'] <- pred.test[,2]
dades1[, 'ProbX'] <- pred.test[,1]
dades1[, 'Prob2'] <- pred.test[,3]
```

```
i <- 1
while ( i <= nrow(subquini1.1)){

if ( pred.test[i,2] > pred.test[i,1] && pred.test[i,2] > pred.test[i,3] ){
subquini1.1[i,'PrediccioQuiniela'] <- 1

} else if ( pred.test[i,1] > pred.test[i,2] && pred.test[i,1] > pred.test[i,3]){
subquini1.1[i,'PrediccioQuiniela'] <- 0

} else if ( pred.test[i,3] > pred.test[i,1] && pred.test[i,3] > pred.test[i,2]){
subquini1.1[i,'PrediccioQuiniela'] <- 2
}
}
```

```

i <- i + 1

}

head(subquini1.1)

i <- 1
while ( i <= nrow(subquini1.1)){
  if (subquini1.1[i,'quiniela']==subquini1.1[i,'PrediccioQuiniela']){
    subquini1.1[i,'EncertQuiniela'] <- 1
  } else {
    subquini1.1[i,'EncertQuiniela'] <- 0
  }
  i <- i + 1
}

mean(subquini1.1[, 'EncertQuiniela'])

##### Collinealitat #####
gc2lmbo <- lm(gc2 ~ gf2 + pos2 ,data=dades1)
summary(gc2lmbo)
1/(1-summary(gc2lmbo)$r.square)

gf2lmbo <- lm(gf2 ~ gc2 + pos2 ,data=dades1)
summary(gf2lmbo)
1/(1-summary(gf2lmbo)$r.square)

pos2lmbo <- lm(pos2 ~ gf2 + gc2 ,data=dades1)
summary(pos2lmbo)
1/(1-summary(pos2lmbo)$r.square)

##### Calibratge #####
calibrat1 <- dades1[order(dades1$Prob1),]
par (mfrow=c(1,2))
c1l <- c(mean(calibrat1[1:493,'Prob1']),
        mean(calibrat1[494:986,'Prob1']),
        mean(calibrat1[987:1479,'Prob1']),

```

```
mean(calibrat1[1480:1972,'Prob1']),
mean(calibrat1[1973:2465,'Prob1']),
mean(calibrat1[2466:2958,'Prob1']),
mean(calibrat1[2959:3450,'Prob1']),
mean(calibrat1[3452:3943,'Prob1']),
mean(calibrat1[3945:4437,'Prob1']),
mean(calibrat1[4438:4930,'Prob1']))
```

```
p1l <- mean(calibrat1[1:493,'guanyatL'])
p2l <- mean(calibrat1[494:986,'guanyatL'])
p3l <- mean(calibrat1[987:1479,'guanyatL'])
p4l <- mean(calibrat1[1480:1972,'guanyatL'])
p5l <- mean(calibrat1[1973:2465,'guanyatL'])
p6l <- mean(calibrat1[2466:2958,'guanyatL'])
p7l <- mean(calibrat1[2959:3450,'guanyatL'])
p8l <- mean(calibrat1[3452:3943,'guanyatL'])
p9l <- mean(calibrat1[3945:4437,'guanyatL'])
p10l <- mean(calibrat1[4438:4930,'guanyatL'])
```

```
c2l <- c( p1l,
          p2l,
          p3l,
          p4l,
          p5l,
          p6l,
          p7l,
          p8l,
          p9l,
          p10l)
```

```
lsl <- c( p1l+1.96*sqrt( (p1l*(1-p1l)/493) ),
          p2l+1.96*sqrt( (p2l*(1-p2l)/493) ),
          p3l+1.96*sqrt( (p3l*(1-p3l)/493) ),
          p4l+1.96*sqrt( (p4l*(1-p4l)/493) ),
          p5l+1.96*sqrt( (p5l*(1-p5l)/493) ),
          p6l+1.96*sqrt( (p6l*(1-p6l)/493) ),
          p7l+1.96*sqrt( (p7l*(1-p7l)/493) ),
          p8l+1.96*sqrt( (p8l*(1-p8l)/493) ),
```



```

p9l+1.96*sqrt( (p9l*(1-p9l)/493) ),
p10l+1.96*sqrt( (p10l*(1-p10l)/493) ))

lil <- c( p1l-1.96*sqrt( (p1l*(1-p1l)/493) ),
p2l-1.96*sqrt( (p2l*(1-p2l)/493) ),
p3l-1.96*sqrt( (p3l*(1-p3l)/493) ),
p4l-1.96*sqrt( (p4l*(1-p4l)/493) ),
p5l-1.96*sqrt( (p5l*(1-p5l)/493) ),
p6l-1.96*sqrt( (p6l*(1-p6l)/493) ),
p7l-1.96*sqrt( (p7l*(1-p7l)/493) ),
p8l-1.96*sqrt( (p8l*(1-p8l)/493) ),
p9l-1.96*sqrt( (p9l*(1-p9l)/493) ),
p10l-1.96*sqrt( (p10l*(1-p10l)/493) ))

plot(c1l,c2l, type='l',ylim=c(0,1),xlim=c(0,1),ylab='Observat', xlab='Predit')
abline(0,1,lty=2)
lines(c1l, lsl,col='blue')
lines(c1l, lil,col='blue')
points(c1l, c2l, pch=16)
points(c1l, lsl,col='blue', pch=16)
points(c1l, lil,col='blue', pch=16)
title ('Calibrat Prob Local')

calibrat2 <- dades1[order(dades1$Prob2),]
c1v <- c( mean(calibrat2[1:493,'Prob2']),
mean(calibrat2[494:986,'Prob2']),
mean(calibrat2[987:1479,'Prob2']),
mean(calibrat2[1480:1972,'Prob2']),
mean(calibrat2[1973:2465,'Prob2']),
mean(calibrat2[2466:2958,'Prob2']),
mean(calibrat2[2959:3450,'Prob2']),
mean(calibrat2[3452:3943,'Prob2']),
mean(calibrat2[3945:4437,'Prob2']),
mean(calibrat2[4438:4930,'Prob2']))

p1v <- mean(calibrat2[1:493,'guanyatV'])
p2v <- mean(calibrat2[494:986,'guanyatV'])
p3v <- mean(calibrat2[987:1479,'guanyatV'])

```

```

p4v <- mean(calibrat2[1480:1972,'guanyatV'])
p5v <- mean(calibrat2[1973:2465,'guanyatV'])
p6v <- mean(calibrat2[2466:2958,'guanyatV'])
p7v <- mean(calibrat2[2959:3450,'guanyatV'])
p8v <- mean(calibrat2[3452:3943,'guanyatV'])
p9v <- mean(calibrat2[3945:4437,'guanyatV'])
p10v <- mean(calibrat2[4438:4930,'guanyatV'])

```

```

c2v <- c( p1v,
          p2v,
          p3v,
          p4v,
          p5v,
          p6v,
          p7v,
          p8v,
          p9v,
          p10v)

```

```

lsv <- c( p1v+1.96*sqrt( (p1v*(1-p1v)/493) ),
          p2v+1.96*sqrt( (p2v*(1-p2v)/493) ),
          p3v+1.96*sqrt( (p3v*(1-p3v)/493) ),
          p4v+1.96*sqrt( (p4v*(1-p4v)/493) ),
          p5v+1.96*sqrt( (p5v*(1-p5v)/493) ),
          p6v+1.96*sqrt( (p6v*(1-p6v)/493) ),
          p7v+1.96*sqrt( (p7v*(1-p7v)/493) ),
          p8v+1.96*sqrt( (p8v*(1-p8v)/493) ),
          p9v+1.96*sqrt( (p9v*(1-p9v)/493) ),
          p10v+1.96*sqrt( (p10v*(1-p10v)/493) ))

```

```

liv <- c( p1v-1.96*sqrt( (p1v*(1-p1v)/493) ),
          p2v-1.96*sqrt( (p2v*(1-p2v)/493) ),
          p3v-1.96*sqrt( (p3v*(1-p3v)/493) ),
          p4v-1.96*sqrt( (p4v*(1-p4v)/493) ),
          p5v-1.96*sqrt( (p5v*(1-p5v)/493) ),
          p6v-1.96*sqrt( (p6v*(1-p6v)/493) ),
          p7v-1.96*sqrt( (p7v*(1-p7v)/493) ),
          p8v-1.96*sqrt( (p8v*(1-p8v)/493) ),

```

```

p9v-1.96*sqrt( (p9v*(1-p9v)/493) ),
p10v-1.96*sqrt( (p10v*(1-p10v)/493) ))

plot(c1v,c2v, type='l',ylim=c(0,1),xlim=c(0,1), ylab='Observat', xlab='Predit')
abline(0,1,lty=2)
lines(c1v, lsv,col='blue')
lines(c1v, liv,col='blue')
points(c1v, c2v,pch=16)
points(c1v, lsv,col='blue',pch=16)
points(c1v, liv,col='blue',pch=16)
title ('Calibrat Prob Visitant')

##### Discriminacio #####
roc.multi1 <- multiclass.roc(subquini1.1$quiniela, subquini1.1$PrediccioQuiniela, levels=c(0,
1, 2))
roc.multi1

Quiniela1 <- subset(dades1, quiniela == 1)
QuinielaX <- subset(dades1, quiniela == 0)
Quiniela2 <- subset(dades1, quiniela == 2)

aucprob1 <- multiclass.roc(subquini1.1$quiniela, dades1[, 'Prob1'], levels=c(0, 1, 2))
aucprob1 #0.62
aucprobx <- multiclass.roc(subquini1.1$quiniela, dades1[, 'ProbX'], levels=c(0, 1, 2))
aucprobx #0.5789
aucprob2 <- multiclass.roc(subquini1.1$quiniela, dades1[, 'Prob2'], levels=c(0, 1, 2))
aucprob2 #0.6204

##### Densitats
par (mfrow=c(1,3))
plot(density(Quiniela1[, 'Prob1']), xlab = 'Probabilitat', ylab= 'Densitat', main = "", xlim= c(0,1),
ylim=c(0,6))
lines(density(Quiniela1[, 'Prob1']),col="springgreen4", lwd = 3)
lines(density(QuinielaX[, 'Prob1']),col="yellow", lwd = 3)
lines(density(Quiniela2[, 'Prob1']),col="aquamarine", lwd = 3)
title(main="Funcio de Densitat Prob Local")
legend("topright", legend=c("1","X","2"), col = c("springgreen4", "yellow","aquamarine"),
pch = 15)

```

```
legend(0.4,6, legend=c("AUC: 0.6200"), border = 13)
legend(0.4,6, legend=paste("AUC: 0.6200"), border = 13)
```

```
plot(density(QuinielaX[, 'ProbX']), xlab = 'Probabilitat', ylab= 'Densitat', main = "", xlim=
c(0,0.4), ylim=c(0,18))
lines(density(Quiniela1[, 'ProbX']),col="springgreen4", lwd = 3)
lines(density(QuinielaX[, 'ProbX']),col="yellow", lwd = 3)
lines(density(Quiniela2[, 'ProbX']),col="aquamarine", lwd = 3)
title(main="Funcio de Densitat Prob Empat")
legend("topright", legend=c("1", "X", "2"), col = c("springgreen4", "yellow", "aquamarine"),
pch = 15)
legend(0.1,18, legend=c("AUC: 0.5789"), border = 13)
```

```
plot(density(Quiniela2[, 'Prob2']), xlab = 'Probabilitat', ylab= 'Densitat', main = "", xlim= c(0,1),
ylim=c(0,6))
lines(density(Quiniela1[, 'Prob2']),col="springgreen4", lwd = 3)
lines(density(QuinielaX[, 'Prob2']),col="yellow", lwd = 3)
lines(density(Quiniela2[, 'Prob2']),col="aquamarine", lwd = 3)
title(main="Funcio de Densitat Prob Visitant")
legend("topright", legend=c("1", "X", "2"), col = c("springgreen4", "yellow", "aquamarine"),
pch = 15)
legend(0.4,6, legend=c("AUC: 0.6204"), border = 13)
```

```
##### PROB 1
```

```
summary(subset(dades1, Prob1 <= 0.1)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9)[, 'quiniela2'])
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1)[, 'quiniela2'])
# Encert
(15+69+103+283+679+755+395+134+78+2)/4930
```

PROB X

```
summary(subset(dades1, ProbX <= 0.1), 'quiniela2')
summary(subset(dades1, ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')
summary(subset(dades1, ProbX > 0.2 & ProbX <= 0.3), 'quiniela2')
summary(subset(dades1, ProbX > 0.3 & ProbX <= 0.4), 'quiniela2')
summary(subset(dades1, ProbX > 0.4 & ProbX <= 0.5), 'quiniela2')
summary(subset(dades1, ProbX > 0.5 & ProbX <= 0.6), 'quiniela2')
summary(subset(dades1, ProbX > 0.6 & ProbX <= 0.7), 'quiniela2')
summary(subset(dades1, ProbX > 0.7 & ProbX <= 0.8), 'quiniela2')
summary(subset(dades1, ProbX > 0.8 & ProbX <= 0.9), 'quiniela2')
summary(subset(dades1, ProbX > 0.9 & ProbX <= 1), 'quiniela2')
#Encert
(7+301+2055+48)/4930
```

PROB 2

```
summary(subset(dades1, Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob2 > 0.1 & Prob2 <= 0.2), 'quiniela2')
summary(subset(dades1, Prob2 > 0.2 & Prob2 <= 0.3), 'quiniela2')
summary(subset(dades1, Prob2 > 0.3 & Prob2 <= 0.4), 'quiniela2')
summary(subset(dades1, Prob2 > 0.4 & Prob2 <= 0.5), 'quiniela2')
summary(subset(dades1, Prob2 > 0.5 & Prob2 <= 0.6), 'quiniela2')
summary(subset(dades1, Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
summary(subset(dades1, Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
summary(subset(dades1, Prob2 > 0.8 & Prob2 <= 0.9), 'quiniela2')
summary(subset(dades1, Prob2 > 0.9 & Prob2 <= 1), 'quiniela2')
# Encert
(173+691+1033+408+113+50+45+10)/4930
```

PROB 1 i PROB 2

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 <= 0.1), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 <= 0.1), 'quiniela2')
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 <= 0.1), 'quiniela2')
```

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.1 & Prob2 <= 0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.1 & Prob2 <=
0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.2 & Prob2 <= 0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.2 & Prob2 <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.2 & Prob2 <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.2 & Prob2 <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.2 & Prob2 <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.2 & Prob2 <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.2 & Prob2 <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.2 & Prob2 <=
0.3), 'quiniela2')
```



```

summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.3 & Prob2 <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.3 & Prob2 <=
0.4), 'quiniela2')

summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.4 & Prob2 <= 0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.4 & Prob2 <=
0.5), 'quiniela2')

summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.5 & Prob2 <= 0.6), 'quiniela2')
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.5 & Prob2 <=
0.6), 'quiniela2')
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.5 & Prob2 <=
0.6), 'quiniela2')
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.5 & Prob2 <=
0.6), 'quiniela2')
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.5 & Prob2 <=
0.6), 'quiniela2')
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.5 & Prob2 <=
0.6), 'quiniela2')
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.5 & Prob2 <=
0.6), 'quiniela2')
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.5 & Prob2 <=
0.6), 'quiniela2')

```



```
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.5 & Prob2 <= 0.6), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.5 & Prob2 <= 0.6), 'quiniela2')
```

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.6 & Prob2 <= 0.7), 'quiniela2')
```

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.7 & Prob2 <= 0.8), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.7 & Prob2 <=
0.8),'quiniela2')
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.7 & Prob2 <=
0.8),'quiniela2')
#Encert
(173+691+1033+408+113+50+45+10)/4930
```

```
##### PROB X i PROB
```

```
summary(subset(dades1, ProbX <= 0.1 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.1 & ProbX <= 0.2 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.2 & ProbX <= 0.3 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.3 & ProbX <= 0.4 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.4 & ProbX <= 0.5 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.5 & ProbX <= 0.6 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.6 & ProbX <= 0.7 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.7 & ProbX <= 0.8 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.8 & ProbX <= 0.9 & Prob2 <= 0.1),'quiniela2')
summary(subset(dades1, ProbX > 0.9 & ProbX <= 1 & Prob2 <= 0.1),'quiniela2')

summary(subset(dades1, ProbX <= 0.1 & Prob2 > 0.1 & Prob2 <= 0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.1 & ProbX <= 0.2 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.2 & ProbX <= 0.3 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.3 & ProbX <= 0.4 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.4 & ProbX <= 0.5 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.5 & ProbX <= 0.6 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.6 & ProbX <= 0.7 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.7 & ProbX <= 0.8 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.8 & ProbX <= 0.9 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
summary(subset(dades1, ProbX > 0.9 & ProbX <= 1 & Prob2 > 0.1 & Prob2 <=
0.2),'quiniela2')
```



```

summary(subset(dades1, ProbX <= 0.1 & Prob2 > 0.6 & Prob2 <= 0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.1 & ProbX <= 0.2 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.2 & ProbX <= 0.3 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.3 & ProbX <= 0.4 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.4 & ProbX <= 0.5 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.5 & ProbX <= 0.6 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.6 & ProbX <= 0.7 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.7 & ProbX <= 0.8 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.8 & ProbX <= 0.9 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.9 & ProbX <= 1 & Prob2 > 0.6 & Prob2 <=
0.7)[,'quiniela2'])

```

```

summary(subset(dades1, ProbX <= 0.1 & Prob2 > 0.7 & Prob2 <= 0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.1 & ProbX <= 0.2 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.2 & ProbX <= 0.3 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.3 & ProbX <= 0.4 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.4 & ProbX <= 0.5 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.5 & ProbX <= 0.6 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.6 & ProbX <= 0.7 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.7 & ProbX <= 0.8 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.8 & ProbX <= 0.9 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])
summary(subset(dades1, ProbX > 0.9 & ProbX <= 1 & Prob2 > 0.7 & Prob2 <=
0.8)[,'quiniela2'])

```

#Encert

(173+691+1033+408+113+50+45+10)/4930

PROB X i PROB 1

summary(subset(dades1, Prob1 <= 0.1 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & ProbX <= 0.1), 'quiniela2')

summary(subset(dades1, Prob1 <= 0.1 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & ProbX > 0.1 & ProbX <= 0.2), 'quiniela2')

summary(subset(dades1, Prob1 <= 0.1 & ProbX > 0.2 & ProbX <= 0.3), 'quiniela2')

summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & ProbX > 0.2 & ProbX <= 0.3), 'quiniela2')

summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & ProbX > 0.2 & ProbX <= 0.3), 'quiniela2')

```

summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & ProbX > 0.2 & ProbX <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & ProbX > 0.2 & ProbX <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & ProbX > 0.2 & ProbX <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & ProbX > 0.2 & ProbX <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & ProbX > 0.2 & ProbX <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & ProbX > 0.2 & ProbX <=
0.3), 'quiniela2')
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & ProbX > 0.2 & ProbX <=
0.3), 'quiniela2')

```

```

summary(subset(dades1, Prob1 <= 0.1 & ProbX > 0.3 & ProbX <= 0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & ProbX > 0.3 & ProbX <=
0.4), 'quiniela2')

```

```

# Encert
(15+69+103+283+679+755+395+134+78+2)/4930

```

```

##### PROB 1 i PROB 2,PROBX <= 0.2
summary(subset(dades1, Prob1 <= 0.1 & Prob2 <= 0.1 & ProbX <= 0.2), 'quiniela2')

```



```
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
<= 0.2),['quiniela2'])
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
<= 0.2),['quiniela2'])
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
<= 0.2),['quiniela2'])
```

```
##### PROB 1 i PROB 2, 0.2 <= PROBX <= 0.3
```

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX <= 0.3
),['quiniela2'])
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 <= 0.1 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
```

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.1 & Prob2 <= 0.2 & ProbX > 0.2 & ProbX
<= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.1 & Prob2 <= 0.2 & ProbX
> 0.2 & ProbX <= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.1 & Prob2 <= 0.2 & ProbX
> 0.2 & ProbX <= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.1 & Prob2 <= 0.2 & ProbX
> 0.2 & ProbX <= 0.3 ),['quiniela2'])
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.1 & Prob2 <= 0.2 & ProbX
> 0.2 & ProbX <= 0.3 ),['quiniela2'])
```



```

summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.2 & ProbX <= 0.3 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.2 & ProbX <= 0.3 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.2 & ProbX <= 0.3 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.2 & ProbX <= 0.3 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.2 & ProbX <= 0.3 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.2 & ProbX <= 0.3 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.2 & ProbX <= 0.3 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX >
0.2 & ProbX <= 0.3 ),'quiniela2'))

```

```

##### PROB 1 i PROB 2,PROBX <= 0.2

```

```

summary(subset(dades1, Prob1 <= 0.1 & Prob2 <= 0.1 & ProbX <= 0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 <= 0.1 & ProbX <=
0.2),'quiniela2'))

summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.1 & Prob2 <= 0.2 & ProbX <=
0.2),'quiniela2'))

```



```
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX <= 0.2), 'quiniela2')
```

```
##### PROB 1 i PROB 2, 0.3 <= PROBX <= 0.4
```

```
summary(subset(dades1, Prob1 <= 0.1 & Prob2 <= 0.1 & ProbX > 0.3 & ProbX <= 0.4), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 <= 0.1 & ProbX > 0.3 & ProbX <= 0.4), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 <= 0.1 & ProbX > 0.3 & ProbX <= 0.4), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 <= 0.1 & ProbX > 0.3 & ProbX <= 0.4), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 <= 0.1 & ProbX > 0.3 & ProbX <= 0.4), 'quiniela2')
```

```
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 <= 0.1 & ProbX > 0.3 & ProbX <= 0.4), 'quiniela2')
```



```

summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.6 & Prob2 <= 0.7 & ProbX >
0.3 & ProbX <= 0.4 ),'quiniela2'))

```

```

summary(subset(dades1, Prob1 <= 0.1 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX > 0.3 & ProbX
<= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.1 & Prob1 <= 0.2 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.2 & Prob1 <= 0.3 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.3 & Prob1 <= 0.4 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.4 & Prob1 <= 0.5 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.5 & Prob1 <= 0.6 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.6 & Prob1 <= 0.7 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.7 & Prob1 <= 0.8 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.8 & Prob1 <= 0.9 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX
> 0.3 & ProbX <= 0.4 ),'quiniela2'))
summary(subset(dades1, Prob1 > 0.9 & Prob1 <= 1 & Prob2 > 0.7 & Prob2 <= 0.8 & ProbX >
0.3 & ProbX <= 0.4 ),'quiniela2'))

```

#Encert

(93+78+2+90+41+3+6+255+305+521+497+231+149+71+33+29+21+40+5+4+3+9+5+23+1+9)
/4930

Encert final


```

i <- 1

while ( i <= nrow(subquini1.1)){

  if ( dades1[i,'Prob2'] <= 0.4 ){
    subquini1.1[i,'PrediccióQuiniela1'] <- 'Aposta 1'
  }else if ( dades1[i,'Prob2'] > 0.4){
    subquini1.1[i,'PrediccióQuiniela1'] <- 'Aposta 2'
  }

  i <- i + 1
}

tabla11<-table(subquini1.1[, 'quiniela2'],subquini1.1[, 'PrediccióQuiniela1'])
round(tabla11,4)

tabla211<-prop.table(tabla11,1)
round(tabla211,4)

tabla311<-prop.table(tabla11,2)
round(tabla311,4)

#####
##### MOSTRA DE VALIDACIO #####
#####

mvalidacio <- subquini1.1 [4931:5800,]
predict.validacio <- predict(model4,newdata=mvalidacio ,type="response")
mvalidacio[, 'Prob1'] <- predict.validacio[,2]
mvalidacio[, 'ProbX'] <- predict.validacio[,1]
mvalidacio[, 'Prob2'] <- predict.validacio[,3]

i <- 1

while ( i <= nrow(mvalidacio)){

  if ( mvalidacio[i,'Prob2'] <= 0.4 ){

```

```

    mvalidacio[i,'PrediccionQuiniela1'] <- 'Aposta 1'
  }else if ( mvalidacio[i,'Prob2'] > 0.4){
    mvalidacio[i,'PrediccionQuiniela1'] <- 'Aposta 2'
  }

  i <- i + 1
}

tabla1v<-table(mvalidacio[, 'quiniela2'],mvalidacio[, 'PrediccionQuiniela1'])
round(tabla1v,4)

tabla2v<-prop.table(tabla1v,1)
round(tabla2v,4)

tabla3v<-prop.table(tabla1v,2)
round(tabla3v,4)

aucprob1val <- multiclass.roc(mvalidacio$quiniela, mvalidacio[, 'Prob1'], levels=c(0, 1, 2))
aucprob1val #0.7029
aucprobxval <- multiclass.roc(mvalidacio$quiniela, mvalidacio[, 'ProbX'], levels=c(0, 1, 2))
aucprobxval #0.6169
aucprob2val <- multiclass.roc(mvalidacio$quiniela, mvalidacio[, 'Prob2'], levels=c(0, 1, 2))
aucprob2val #0.701

```