

**Anàlisi dels factors associats a  
l'assoliment d'algunes competències  
bàsiques en l'àrea de Matemàtiques**

**Una proposta de millora en la connexió  
curricular en la transició Primària–Secundària**

**Autor: Jose Barrera Gómez  
Supervisor: Josep Gascón Pérez (UAB)  
Curs acadèmic 2008–2009**

La realització d'aquest treball ha estat possible gràcies a una llicència retribuída concedida pel Departament d'Educació de la Generalitat de Catalunya. Resolució EDU/2276/2008, de 9 de juliol (DOGC núm. 5176 18/07/2008).

## Agraïments

Per a la realització d'aquest treball he gaudit de la col·laboració de diverses persones i organismes que ha resultat una ajuda molt important i, en molts casos, imprescindible. Per això, vull donar les gràcies

als departaments de matemàtiques i als equips directius dels 9 centres analitzats per facilitar-me les dades acadèmiques del seu alumnat, sense les quals és evident que aquest treball no hagués estat possible, i, molt especialment, a l'equip directiu de l'IES Thos i Codina que va avalar aquest projecte, fins i tot sabent que jo en seria l'autor,

al Josep Gascón Pérez, del Departament de Matemàtiques de la UAB, per acceptar la supervisió del projecte i orientar-me en la interpretació dels resultats obtinguts i per fer-me pensar sovint en els problemes de la didàctica de les matemàtiques,

al Pere Puig, del mateix departament, per haver encertat en suggerir-me el Josep com a supervisor i haver-me posat en contacte amb ell,

al Josep A. Sánchez Espigares, per introduir-me en el concepte del Model Lineal Mixt, ensenyar-me a modelitzar bases de dades amb mesures repetides i aclarir tots els meus dubtes (que no han estat pocs) sobre modelització estadística, sempre amb disponibilitat, paciència i interès,

al Servei d'Estadística de la UAB, per aclarir els meus dubtes de manera instantània, amable i desinteressada,

a la Mari Teruel, experta i valuosíssima professora de matemàtiques, companya en el Departament de Matemàtiques de l'IES Thos i Codina i amiga, per les seves aportacions, idees i crítiques (sempre constructives) que han fet guanyar qualitat a aquest treball,

al Pep Timoneda que, a més d'haver-me tutelat en els meus inicis com a professor de matemàtiques a secundària i d'haver-me suportat després com a company i com amic, m'ha facilitat la informació necessària des del Centre de Recursos Pedagògics del Maresme I,

a la Ruth Pomés, amb qui vaig tenir la sort de compartir interessants reunions d'Equip Docent, per la correcció lingüística d'aquest document, al molt raonable preu d'"una birra per cada cinquanta paraules",

a la Judith, per la seva gran ajuda en la depuració de la base de dades (ella en sap, és de poble) i,

sobretot, a la mare que em va parir, em va criar i, des de petit, em va inculcar que l'educació és un dels valors més importants pels quals val la pena lluitar.

# Índex

## Agraïments

<b>1</b>	<b>Introducció</b>	<b>1</b>
<b>2</b>	<b>Les dades</b>	<b>2</b>
2.1	Individus . . . . .	2
2.2	Variables . . . . .	2
2.3	Recollida de dades . . . . .	3
2.4	Validació de la base de dades . . . . .	4
<b>3</b>	<b>Metodologia</b>	<b>4</b>
<b>4</b>	<b>Anàlisi del flux de l'alumnat des dels centres de primària als centres de secundària considerats en l'estudi</b>	<b>7</b>
4.1	Anàlisi del flux brut de l'alumnat . . . . .	7
4.1.1	Anàlisi descriptiva del flux d'alumnat . . . . .	9
4.1.2	Anàlisi de Correspondències Simple . . . . .	9
4.2	Avaluació de l'aleatorietat en la pèrdua d'informació deguda al flux de sortida i al flux d'entrada . . . . .	11
4.2.1	Alumnat que surt al juny . . . . .	11
4.2.2	Alumnat que entra al setembre . . . . .	16
4.3	Individus finalment inclosos en l'estudi segons els casos complets en el flux net intern . . . . .	16
<b>5</b>	<b>Anàlisi descriptiva</b>	<b>19</b>
<b>6</b>	<b>Modelització dels resultats en la prova de juny</b>	<b>25</b>
6.1	Model ajustat . . . . .	27
6.2	Interpretació del model ajustat i dels seus paràmetres . . . . .	27
6.3	Prediccions del model ajustat . . . . .	29
<b>7</b>	<b>Estimació del Risc Relatiu com a mesura d'assoliment de continguts a curt termini (juny–setembre)</b>	<b>33</b>
<b>8</b>	<b>Modelització de la diferència entre el resultat en la prova de setembre i el resultat de la prova de juny</b>	<b>34</b>
8.1	Resposta quantitativa. Diferència numèrica entre juny i setembre	35
8.1.1	Model ajustat . . . . .	35
8.1.2	Interpretació del model ajustat i dels seus paràmetres . .	37
8.1.3	Prediccions del model ajustat . . . . .	38
8.2	Resposta qualitativa. Pèrdua de qualificació entre juny i setembre	38
8.2.1	Model ajustat . . . . .	38
8.2.2	Interpretació del model ajustat i dels seus paràmetres . .	41
8.2.3	Prediccions del model ajustat . . . . .	44

<b>9 Modelització dels resultats en la primera avaluació de 1r d'ESO (desembre)</b>	<b>44</b>
9.1 Resposta quantitativa. Qualificació numèrica . . . . .	44
9.1.1 Model ajustat . . . . .	44
9.1.2 Interpretació del model ajustat i dels seus paràmetres . .	47
9.1.3 Prediccions del model ajustat . . . . .	47
9.2 Resposta qualitativa. Obtenció de l'aprovat . . . . .	49
9.2.1 Model ajustat . . . . .	49
9.2.2 Interpretació del model ajustat i dels seus paràmetres . .	51
9.2.3 Prediccions del model ajustat . . . . .	51
<b>10 Resultats</b>	<b>51</b>
<b>11 Limitacions</b>	<b>57</b>
<b>A Enunciat de la prova escrita</b>	<b>59</b>
<b>B Instruccions per a la realització de la prova</b>	<b>65</b>
<b>C Codi R</b>	<b>66</b>
<b>Referències</b>	<b>113</b>



# 1 Introducció

L'elevat fracàs escolar, el risc d'exclusió social en l'entorn de gran part dels centres analitzats i la necessitat d'avaluar la tendència a l'atomització dels continguts curriculars en la pràctica docent d'aquesta àrea, amb la conseqüent falta d'articulació prevista en el currículum oficial de primària i en el de secundària, fan necessària la recerca de noves eines per a la millora del rendiment acadèmic.

Amb aquest treball, basat en el disseny i la realització d'un estudi estadístic, pretenc poder detectar els factors que s'associen a l'assoliment d'algunes competències bàsiques a curt termini d'una manera objectiva (mesurable) que permeti proporcionar arguments per a ajudar a prendre decisions sobre adaptacions curriculars en la transició Primària–Secundària. En aquest sentit, espero poder aportar arguments a favor del fet que un ensenyament “constructiu” en l'àrea de Matemàtiques, és a dir, amb la conjugació de diferents conceptes i procediments tal com preveuen els currículums oficials de primària i secundària, en contraposició a la tendència a “atomitzar” continguts, es tradueix en un increment de les possibilitats d'evolucionar favorablement en aquesta àrea, especialment en la transició Primària–Secundària i, en tot cas, poder proporcionar resultats útils per a la proposta de millores en la connexió curricular entre aquests dos nivells educatius.

La base de l'estudi es troba en l'anàlisi estadística de les dades aparellades que s'obtenen dels resultats d'una prova escrita que va fer l'alumnat seleccionat al final de Primària (juny de 2008) i a l'inici de Secundària (setembre de 2008), i també de la consideració dels resultats en la primera avaluació del 1r curs d'ESO (desembre 2008).

El marc del projecte és el *Seminari de Coordinació Curricular Primària–Secundària Mataró–Nord* (SCPS), integrat per sis CEIPs i tres IES de titularitat pública de Mataró. Aquest seminari, dins del Pla de Formació de Zona del Centre de Recursos Pedagògics del Maresme I<sup>1</sup>, té com a un dels objectius fomentar la col·laboració entre els centres integrants per tal de cercar eines que permetin introduir propostes per a la millora dels resultats acadèmics de l'alumnat així com de la connexió curricular Primària–Secundària. El fet que els centres integrants es trobin geogràficament pròxims entre ells implica un important trasvàs intern d'alumnat entre aquests dos nivells acadèmics i és aquest el fet principal que dóna sentit a la constitució del SCPS.

No obstant això, el seguiment de l'alumnat en el procés de transició Primària–Secundària resulta complicat donat que habitualment implica un canvi de centre educatiu per part de l'alumnat (és el cas de la totalitat dels centres de primària integrants del SCPS) i, en general, la informació de què els IES disposen sobre el nou alumnat que hi arriba per a començar el 1r curs de l'ESO sol, molt

---

<sup>1</sup>Aquest seminari, juntament amb un altre similar a la zona sud de Mataró, es va crear el curs 2002–2003 com una iniciativa dels propis centres integrants, degut a la necessitat de millorar la coordinació entre els centres de primària i de secundària de la zona. Aquest seminari va ser reconegut oficialment a partir del curs 2006–2007, dins del Pla de Formació de Zona del Centre de Recursos Pedagògics del Maresme.

sovint, reduir-se als informes dels CEIPs i a les proves d'avaluació inicial a Secundària. Aquesta situació pot presentar dos problemes: els informes dels CEIPs solen ser qualitatius i poc exhaustius pel gran volum de feina que implica sintetitzar la informació. A més, la prova d'avaluació inicial a Secundària pot mostrar l'estat actual de l'alumnat però no informa de la seva evolució fins al moment de la prova, ni tan sols a curt termini. En aquest sentit, aquest treball pretén proporcionar una eina per tal de millorar el traspàs d'informació de Primària a Secundària, avaluant l'associació entre els resultats en la prova de juny i els de la de setembre. Tanmateix, el fet de modelitzar les qualificacions en la primera avaluació del 1r curs de Secundària, pot aportar una eina predictora del fracàs escolar a curt termini que pot ser útil al professorat per a la detecció, al començament del curs, de l'alumnat amb risc de dificultats d'aprenentatge.

## 2 Les dades

### 2.1 Individus

Els individus, a partir dels quals es va obtenir la base de dades que s'analitza en aquest treball, es corresponen amb l'alumnat que va cursar 6è curs de Primària en 6 CEIPs (curs acadèmic 2007–2008) i va continuar cursant 1r curs d'ESO en 3 IES (curs acadèmic 2008–2009) de titularitat pública de Mataró. La proximitat geogràfica entre aquests 9 centres fa que la major part de l'alumnat que acaba els estudis de Primària en algun dels centres analitzats continui els estudis de Secundària també en un dels centres analitzats. Aquest fet és important atès que la part essencial d'aquest estudi va consistir a fer un seguiment de l'alumnat des de juny fins a desembre de 2008.

Per motius de preservació de l'anonimat, tant de persones com de centres, es van etiquetar aquests centres com CEIP 1, CEIP 2, CEIP 3, CEIP 4, CEIP 5, CEIP 6 (primària); IES 1, IES 2, IES 3 (secundària).

### 2.2 Variables

La Taula 1 mostra la descripció de les variables mesurades per l'estudi.

Algunes de les variables estan relacionades amb les característiques de l'alumnat: sexe, edat, mes de naixement, repetició de 6è curs de Primària, incorporació recent a Catalunya (menys de 2 anys), alumnat tipificat NEE (Necessitats Educatives Especials), CEIP i IES.

També es van considerar les variables corresponents als resultats obtinguts en una prova escrita. Aquesta prova escrita (veure Annex A) va ser dissenyada pels departaments de matemàtiques dels 9 centres que van participar en l'estudi tot i que el Bloc 6 (Problemes) es va afegir expressament amb motiu d'aquest treball. L'alumnat va realitzar la prova el juny de 2008, just després d'acabar el 6è curs de Primària i també el setembre, just abans de començar el 1r curs de Secundària. La prova es va portar a terme sota la supervisió del professorat



Variable	Etiqueta	Tipus	Codificació/Valors
Sexe	"Sexe"	Nominal	"Noia", "Noi"
Edat	"Edat"	Contínua	Amb una xifra decimal
MesNa	"Mes de naixement"	Ordinal	"Gener", ..., "Desembre"
Rep	"Ha repetit 6è?"	Nominal	"No", "Sí"
Cat2	"Vingut a Catalunya fa menys de 2 anys?"	Nominal	"No", "Sí"
NEE	"NEE"	Nominal	"No", "Sí"
Cprim	"Centre de primària"	Nominal	"CEIP 1", ..., "CEIP 6", "Altres CEIP"
B1P	"+ - · : amb naturals (juny)"	Discreta	{0, ..., 10}
B2P	"Jerarquia d'operacions (juny)"	Discreta	{0, ..., 10}
B3P	"Càlcul amb decimals (juny)"	Discreta	{0, ..., 10}
B4P	"Fraccions i canvis d'unitats (juny)"	Discreta	{0, ..., 10}
B5P	"Geometria (juny)"	Discreta	{0, ..., 10}
B61P	"Problema 1 (juny)"	Discreta	{0, 1}
B62P	"Problema 2 (juny)"	Discreta	{0, 1}
B63P	"Problema 3 (juny)"	Discreta	{0, 1}
B64P	"Problema 4 (juny)"	Discreta	{0, 1}
B65P	"Problema 5 (juny)"	Discreta	{0, 1}
Csec	"Centre de secundària"	Nominal	"IES 1", ..., "IES 3", "Altres IES"
B1S	"+ - · : amb naturals (setembre)"	Discreta	{0, ..., 10}
B2S	"Jerarquia d'operacions (setembre)"	Discreta	{0, ..., 10}
B3S	"Càlcul amb decimals (setembre)"	Discreta	{0, ..., 10}
B4S	"Fraccions i canvis d'unitats (setembre)"	Discreta	{0, ..., 10}
B5S	"Geometria (setembre)"	Discreta	{0, ..., 10}
B61S	"Problema 1 (setembre)"	Discreta	{0, 1}
B62S	"Problema 2 (setembre)"	Discreta	{0, 1}
B63S	"Problema 3 (setembre)"	Discreta	{0, 1}
B64S	"Problema 4 (setembre)"	Discreta	{0, 1}
B65S	"Problema 5 (setembre)"	Discreta	{0, 1}
Flux	"Flux primària-secundària"	Nominal	"Prim→Sec", "Prim→Altres", "Altres→Sec"
Aval1	"Qualificació 1a Avaluació (desembre)"	Discreta	{0, ..., 10}

**Taula 1:** Descripció de les variables mesurades en l'estudi.

responsable de l'alumnat en cadascuna de les dues realitzacions (juny i setembre) i tenint en compte unes instruccions determinades per tal de garantir la igualtat de condicions en els diferents centres participants (veure Annex B).

Finalment, també es va incorporar la qualificació de l'alumnat en l'àrea de Matemàtiques en la primera avaluació de 1r d'ESO (desembre de 2008).

## 2.3 Recollida de dades

Per a recollir les dades es va dissenyar un full de càlcul Excel que es va distribuir a cadascun dels 6 CEIPs. Cadascun d'aquests centres va buidar en el full tant les dades del seu alumnat com els resultats de la prova escrita realitzada en juny.

Quan es van rebre els 6 fulls de càlcul corresponents als 6 centres de primària, es van fusionar en un únic full que es va validar després d'haver-ne fet una anàlisi de possibles errors i de dades mancants.

A continuació es va fer arribar aquest full als 3 IES per tal que cadascun d'ells hi buidés la informació relativa a la prova escrita realitzada al setembre.

Una altra vegada, un cop rebuts els 3 fulls corresponents als 3 IES, es van fusionar en un únic full que es va tornar a validar.

El pas següent va consistir a enviar còpia d'aquest full als 3 IES per tal d'incorporar la qualificació en la primera avaluació de 1r d'ESO.

Per garantir l'anonimat de l'alumnat participant, es va codificar convenientment la variable identificadora de l'individu.

Finalment, després d'una nova fusió i validació de la base de dades es va tancar.

## 2.4 Validació de la base de dades

Per validar la base de dades es va contactar amb els 9 centres per tal de corregir possibles errades, incoherències o dades faltants. Aquest procediment es va realitzar en diferents moments del procés de recollida de dades descrit en la secció 2.3.

Totes les incoherències i errades es van corregir però va resultar inevitable la presència de dades faltants (essencialment alumnat que no va fer la prova de juny o la de setembre per absència, i alumnat del qual no es va disposar de la seva qualificació en la primera avaluació de 1r d'ESO perquè s'havia traslladat anteriorment a un altre centre).

Es va optar per no fer cap tipus d'imputació a les dades faltants i simplement es van considerar els casos complets quan ho van requerir les característiques tècniques de l'eina estadística utilitzada en cada moment.

## 3 Metodologia

En la secció 2 s'ha exposat el procediment que es va seguir per a l'obtenció de la base de dades objecte d'aquest estudi. D'altra banda, els mètodes i criteris que s'apliquen en el tractament estadístic de les dades han estat els següents:

- **Software estadístic:** Tots els procediments estadístics dels quals deriven els resultats que es mostren en aquest treball van ser realitzats amb el paquet estadístic R en la versió R 2.9.0 GUI 1.28 Tiger build 32-bit (5395)<sup>2</sup>.
- **Significació estadística:** Totes les decisions estadístiques es van prendre fixant un llindar de significació del 5%.
- **Anàlisi descriptiva:** Les variables qualitatives es van descriure amb percentatges, nombre total de casos, diagrames de sectors i diagrames de barres. Les variables quantitatives es van descriure amb el nombre de casos, estadístics de posició (mitjana, mediana, mínim, màxim), la desviació estàndard com a estadístic de dispersió i gràfics com l'histograma i el diagrama de caixa. (Secció 5).
- **Anàlisi bivariant de variables qualitatives:** Per a aquest tipus d'anàlisi es van fer servir taules de contingència, diagrames de barres apilades, diagrames de Venn, l'Anàlisi de Correspondències Simple i l'estimació puntual i per interval de confiança del Risc Relatiu (RR). (Seccions 4 i 7).
- **Anàlisi bivariant de variables quantitatives:** Per a aquest tipus d'anàlisi es van fer servir el diagrama de dispersió i el coeficient de correlació parcial. (Secció 5).

---

<sup>2</sup><http://www.r-project.org/>

- **Comparació de mitjanes:** Per a comparar les mitjanes en diferents grups es va fer servir el  $t$ -test. (Secció 4.2).
- **Modelització:** Per a modelitzar els resultats de la prova de juny, de la de setembre i els resultats de la primera avaluació de 1r d'ESO en l'àrea de Matemàtiques es van fer servir models lineals prenent com a variables explicatives aquelles descrites en la secció 2. En funció del tipus de variable resposta, es van fer servir els models lineals següents:
  - Per a modelitzar una variable resposta numèrica amb mesures repetides es va fer servir el **Model Lineal Mixt**. Aquest model considera un efecte fix de les variables explicatives comú a tot l'alumnat analitzat i, a més, un efecte aleatori degut a cada individu en concret. Les eines estadístiques utilitzades permeten aïllar i interpretar els efectes fixes del model ajustat. Aquest model va ser utilitzat en la modelització de les variables següents:
    - \* Qualificació numèrica (de 0 a 10) en la prova de juny, en cadascun dels 6 Blocs de la prova. (Secció 6).
    - \* Diferència (de -10 a 10) entre la qualificació numèrica en la prova de juny i la de setembre, en cadascun dels 6 Blocs de la prova. (Secció 8.1).
  - Per a modelitzar una variable resposta dicotòmica amb mesures repetides es va fer servir el **Model Lineal Mixt amb funció d'enllaç Logit**. Aquest model considera un efecte fix de les variables explicatives comú a tot l'alumnat analitzat i, a més, un efecte aleatori degut a cada individu en concret. La variable resposta es transforma prèviament segons la transformació Logit<sup>3</sup> per tal de poder ajustar el model a una variable resposta no numèrica. Aquest model va ser utilitzat en la modelització de la variable següent:
    - \* Pèrdua de qualificació (no/sí) entre la prova de juny i la de setembre, en cadascun dels 6 Blocs de la prova. (Secció 8.2).
  - Per a modelitzar una variable resposta numèrica sense mesures repetides es va fer servir el **Model Lineal General**. Aquest model

---

<sup>3</sup>La transformació Logit és

$$\text{Logit}(x) = \ln\left(\frac{x}{1-x}\right).$$

Així, si la variable resposta,  $Y$ , és dicotòmica amb possibles valors  $Y = 0$ ,  $Y = 1$  (per exemple, 0 = "no" i 1 = "sí"), es pot considerar la modelització de la probabilitat de resposta "sí",  $P(Y = 1) \in (0, 1)$  i, amb la transformació Logit, es modelitza la variable

$$\tilde{Y} = \text{Logit}(P(Y = 1)) = \ln\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right) \in (-\infty, +\infty)$$

que té un recorregut coherent amb un predictor lineal,  $X\beta$  que pugui prendre qualsevol valor real, de manera que el model  $\tilde{Y} = X\beta + e$  resulta més adient que no pas el model  $Y = X\beta + e$ .

només considera un efecte fix de les variables explicatives comú a tot l'alumnat analitzat, i va es ser utilitzar en la modelització de la variable següent:

- \* Qualificació numèrica (de 0 a 10) en l'àrea de Matemàtiques en la primera avaluació de 1r d'ESO. (Secció 9.1).

- Per a modelitzar una variable resposta dicotòmica sense mesures repetides es va fer servir el **Model Lineal Generalitzat amb enllaç Logit**. Aquest model només considera un efecte fix de les variables explicatives comú a tot l'alumnat analitzat i requereix una transformació prèvia de la variable resposta per tal de poder ajustar el model a una variable resposta no numèrica (transformació Logit). Es va utilitzar en la modelització de la variable següent:

- \* Aprovat (no/sí) en l'àrea de Matemàtiques en la primera avaluació de 1r d'ESO. (Secció 9.2).

- **Selecció del millor model:** El procediment seguit per a la selecció del millor model va ser el que es descriu en els passos següents:

- Inicialment es va prendre el model saturat d'efectes principals; és a dir, es van incloure en el model totes les variables explicatives de la base de dades però no les interaccions entre elles. El motiu de la no inclusió d'interaccions és que, des d'un punt de vista pedagògic, són molt difícils de justificar i això resta interpretabilitat al model ajustat.
- Una vegada fixat el model inicial, es van anant treient el termes estadísticament no significatius, sempre d'un en un, i justificant la seva exclusió o inclusió en funció de la prova de variància (o deviància, segons el cas) incremental.
- Finalment, una vegada obtingut el model definitiu després del pas anterior, es va procedir a la seva validació. Aquesta validació es va fer exigint un comportament correcte dels residus del model –en el cas de resposta numèrica– o una avaluació de la matriu de confusió, la sensibilitat i l'especificitat –en el cas de resposta política.
- Per a cada model final ajustat es mostra l'estimació dels paràmetres i se'n va fer la seva interpretació.

Tot i que en aquest treball només es mostra el model final seleccionat en cada cas, en l'Annex C es mostra el procés de selecció realitzat en cada cas.

- **Prediccions:** Per a cada model ajustat, es van obtenir les prediccions puntuals i per interval de confiança al 95%. Aquestes prediccions es van

obtenir tant per a tractaments<sup>4</sup> concrets com per a perfils de risc (combinacions de valors de les variables explicatives per a les quals la predicció és la pitjor possible) i perfils protectors (combinacions de valors de les variables explicatives per a les quals la predicció és òptima).

## 4 Anàlisi del flux de l'alumnat des dels centres de primària als centres de secundària considerats en l'estudi

Abans de seleccionar l'alumnat que va fer les dues proves per a realitzar-ne un estudi longitudinal, es va analitzar el flux d'alumnat entre juny i setembre, des del punt de vista tant de canvi de centre com de realització de les dues proves, atès que l'alumnat que va realitzar la prova de juny i el que va realitzar la prova de setembre no és exactament el mateix.

### 4.1 Anàlisi del flux brut de l'alumnat

La Figura 1 (pàgina 8) mostra la classificació de l'alumnat segons la realització de les proves. En aquest sentit podem definir tres conjunts d'alumnat:

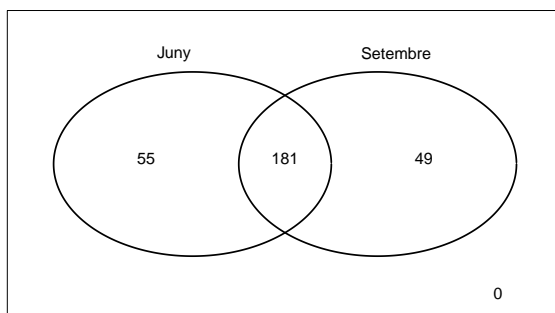
- “Alumnat que surt”. És l'alumnat que va realitzar la prova de juny però no va realitzar la de setembre perquè no es va matricular en cap dels 3 IES en estudi. La mida d'aquest conjunt va ser de 55 individus, i representa un 19,3% del total.
- “Alumnat que entra”. És l'alumnat que va realitzar la prova de setembre però no va realitzar la de juny perquè es tracta d'alumnat procedent d'altres CEIPs diferents als estudiats. La mida d'aquest conjunt va ser de 49 individus, i representa un 17,2% del total.
- “Alumnat intern”. És l'alumnat que va realitzar la prova tant de juny com de setembre; és a dir, és l'alumnat que va cursar 6è de Primària en un dels 6 CEIPs en estudi i que ha cursat 1r d'ESO en un dels 3 IES de l'estudi. La mida d'aquest conjunt va ser de 181 individus, representant un 63,5% del total.

L'estudi longitudinal es va realitzar sobre l'“Alumnat intern” però abans es va comparar aquest conjunt amb l'“Alumnat que surt” i amb l'“Alumnat que entra” en funció dels resultats en la prova de juny i en la prova de setembre respectivament. Aquestes comparacions s'analitzen detalladament en la secció 4.2.

---

<sup>4</sup>Un tractament és una combinació de valors de les variables explicatives.

Distribució de l'alumnat segons la seva participació en les proves



**Figura 1:** Distribució de l'alumnat corresponent als 6 CEIPs i als 3 IES segons si van sortir de la mostra després de fer la prova del juny (55; 19,3%), si van romandre en la mostra fent les dues proves (181; 63,5%) o si es van incorporar a la mostra al setembre des d'altres CEIPs (49; 17,2%).

		<b>IES 1</b>	<b>IES 2</b>	<b>IES 3</b>	<b>Altres</b>	<b>Total</b>
	<i>N</i>	3	5	3	13	24
<b>CEIP 1</b>	% f.	12,5	20,8	12,5	54,2	100
	% c.	3,5	5,6	5,5	23,6	8,4
	<i>N</i>	16	25	3	7	51
<b>CEIP 2</b>	% f.	31,4	49,0	5,9	13,7	100
	% c.	18,8	27,8	5,5	12,7	17,9
	<i>N</i>	13	3	24	3	43
<b>CEIP 3</b>	% f.	30,2	7,0	55,8	7,0	100
	% c.	15,3	3,3	43,6	5,5	15,1
	<i>N</i>	13	17	3	8	41
<b>CEIP 4</b>	% f.	31,7	41,5	7,3	19,5	100
	% c.	15,3	18,9	5,5	14,5	14,4
	<i>N</i>	19	15	1	15	50
<b>CEIP 5</b>	% f.	38,0	30,0	2,0	30,0	100
	% c.	22,4	16,7	1,8	27,3	17,5
	<i>N</i>	9	2	7	9	27
<b>CEIP 6</b>	% f.	33,3	7,4	25,9	33,3	100
	% c.	10,6	2,2	12,7	16,4	9,5
	<i>N</i>	12	23	14	0	49
<b>Altres</b>	% f.	24,5	46,9	28,6	0,0	100
	% c.	14,1	25,6	25,5	0,0	17,2
	<i>N</i>	85	90	55	55	285
<b>Total</b>	% f.	29,8	31,6	19,3	19,3	100
	% c.	100	100	100	100	100

**Taula 2:** Distribució de l'alumnat segons el CEIP de procedència i l'IES de destinació. Es mostren el volum d'alumnat (*N*) i els percentatges per fila (% f.) i per columna (% c.).

#### 4.1.1 Anàlisi descriptiva del flux d'alumnat

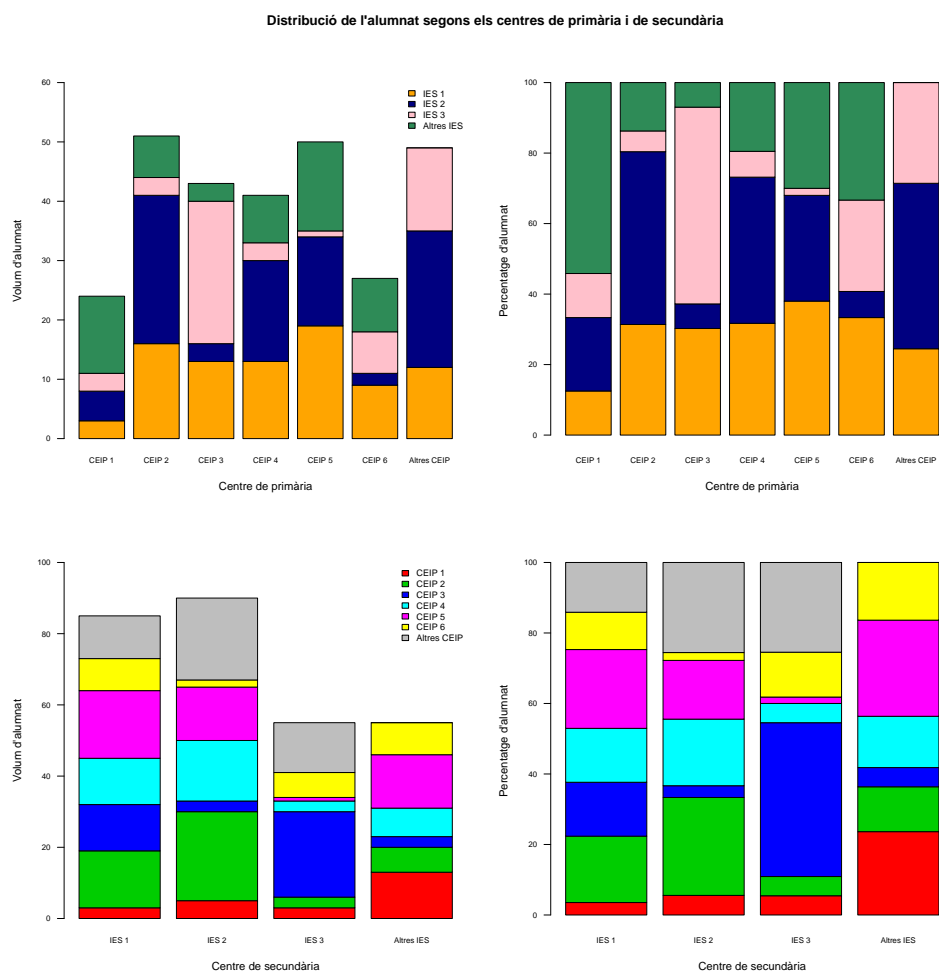
La Taula 2 (pàgina 9) mostra la distribució de l'alumnat segons el seu flux entre el CEIP i l'IES; s'hi mostren les freqüències absolutes (*N*), els percentatges per files (% f.) i els percentatges per columna (% c.). La no homogeneïtat dels percentatges fila indica que alguns CEIPs tendeixen a relacionar-se amb determinats IES. Aquesta heterogeneïtat es pot observar clarament en la Figura 2 (pàgina 10).

#### 4.1.2 Anàlisi de Correspondències Simple

Per tal de caracteritzar l'associació entre els CEIPs i els IES, es va realitzar una Anàlisi de Correspondències Simple que es va representar en dues dimensions<sup>5</sup> en la Figura 3 (pàgina 3) on es pot observar que:

- El CEIP 1 es caracteritza pel seu flux d'alumnat cap a IES diferents dels 3 estudiats.

<sup>5</sup>Reduint les tres dimensions inicials (3 IES) a dues, es va conservar un 94,8% d'explicabilitat.



**Figura 2:** Distribució de l'alumnat segons el CEIP de procedència i l'IES de destinació. En les dues figures superiors es mostren els “perfils d'emissió” dels CEIPs (absoluts a l'esquerra i relatius a la dreta). Anàlogament, en les dues figures inferiors es mostren els “perfils de recepció” dels IES.



- El CEIP 2 s'associa a l'IES 2.
- El CEIP 3 s'associa a l'IES 3.
- El CEIP 4 s'associa als IES 1 i 2.
- El CEIP 5 s'associa als IES 1 i 2 i a altres IES diferents als estudiats.
- El CEIP 6 s'associa als IES 1 i 3 i a altres IES diferents als estudiats.
- Quasi la meitat de l'alumnat provinent d'altres CEIPs diferents als estudiats a un dels 3 IES de l'estudi es matricula en l'IES 2 i la resta es reparteix gairebé en parts iguals entre els IES 1 i 3.

Es va poder determinar que aquesta caracterització de flux d'alumnat és coherent amb la situació geogràfica dels 9 centres.

## 4.2 Avaluació de l'aleatorietat en la pèrdua d'informació deguda al flux de sortida i al flux d'entrada

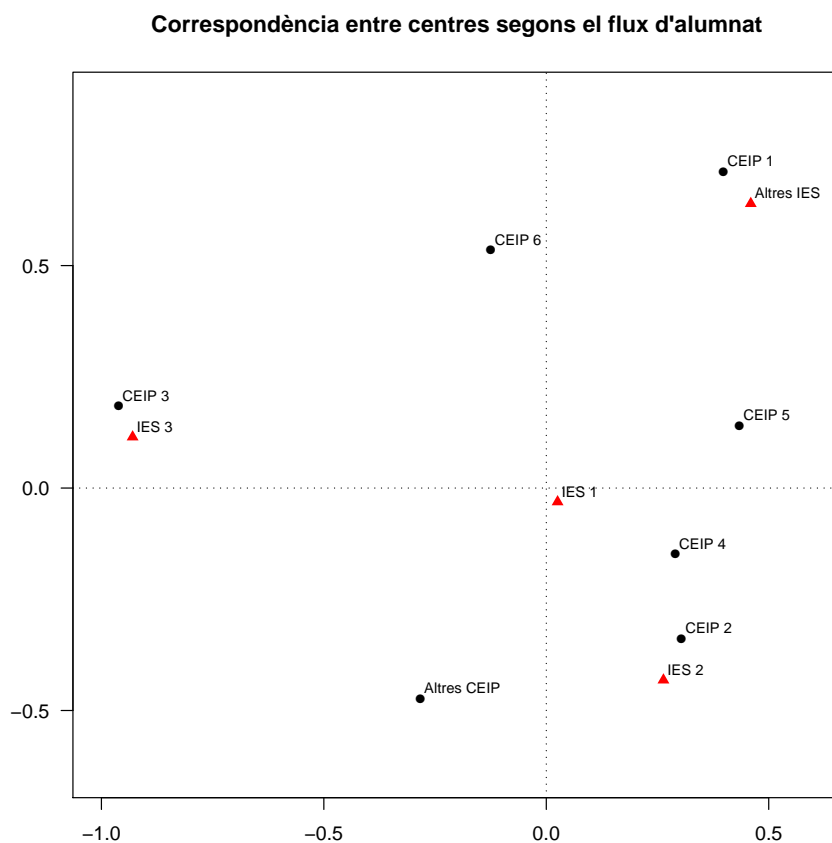
Tal i com s'ha indicat en la secció 4.1, abans d'analitzar el seguiment de l'"Alumnat intern", es va comparar aquest conjunt amb l'"Alumnat que surt" i amb l'"Alumnat que entra", en funció dels resultats en la prova de juny i en la prova de setembre respectivament. Aquesta comparació es va fer amb la intenció de comprovar si hi va haver diferències significatives entre els resultats pel fet que l'alumnat pertanyés a un dels tres grups o si, pel contrari, no van haver tals diferències i, per tant, el flux d'alumnat entre centres va ser aleatori en aquest sentit. Per a analitzar aquestes diferències es va fer servir el  $t$ -test per a la comparació de mitjanes.

### 4.2.1 Alumnat que surt al juny

Respecte la comparació entre l'"Alumnat intern" i l'"Alumnat que surt", la Taula 3 (pàgina 13) mostra un resum descriptiu dels resultats en cadascun dels 6 Blocs de la prova de juny, per a cadascun dels dos grups d'alumnat. S'hi observa una variabilitat similar en els resultats (Desviació estàndar) però una mitjana més gran en tots els Blocs en l'"Alumnat intern".

D'altra banda, la Figura 4 (pàgina 14) i la Figura 5 (pàgina 15) mostren una comparació gràfica del resultat entre els dos grups d'alumnat mitjançant histogrames i diagrames de caixa, respectivament.

Es va realitzar el  $t$ -test per a la comparació de mitjanes per a decidir si aquestes diferències van ser significatives. Els resultats es mostren en la Taula 4 (pàgina 13). Prenent un nivell de significació del 5%, es pot observar (columna  $p$ -valor) que les diferències entre els dos grups d'alumnat són significatives únicament en el Bloc 1 (+ - · : amb naturals), a favor de "Alumnat intern". Així, es pot suposar que, des del punt de vista dels resultats en la prova, els resultats entre els dos grups no van ser significativament diferents.



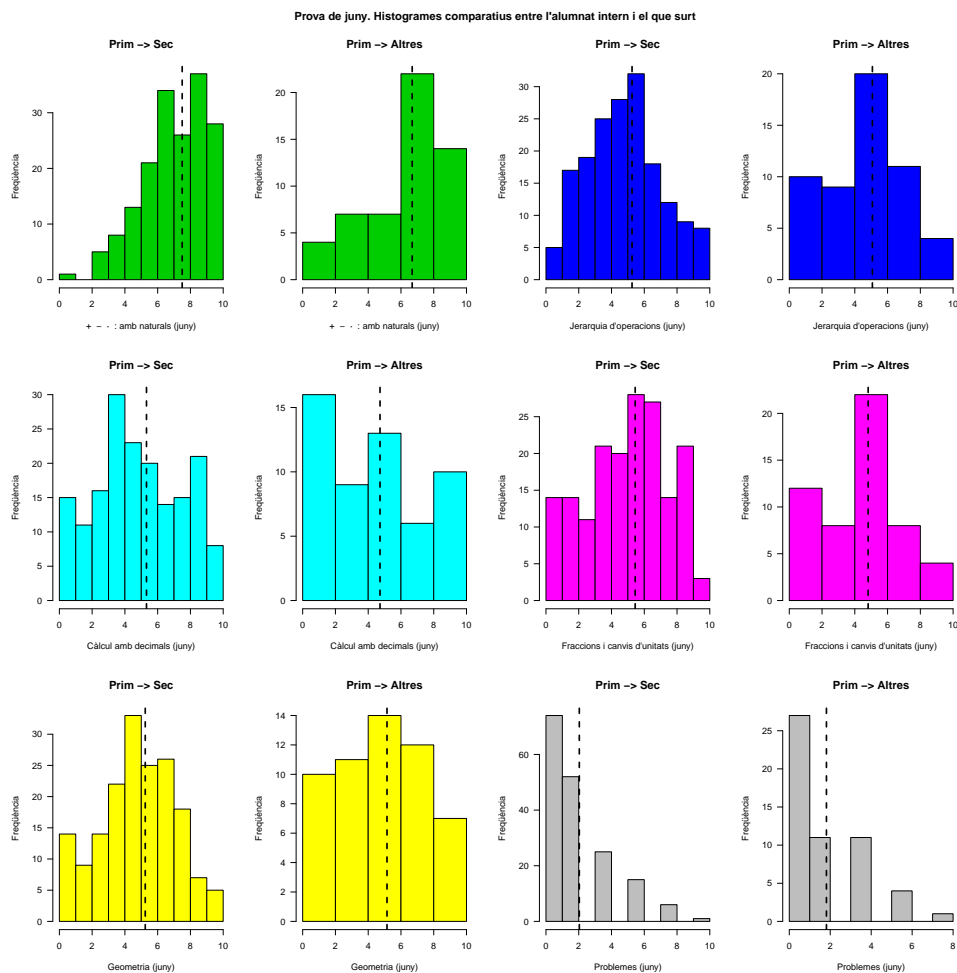
**Figura 3:** Representació gràfica corresponent a l'Anàlisi de Correspondències Simple entre els CEIPs i els IES segons el flux d'alumnat.

Bloc	Alumnat	N	Mitjana	Desv. Est.	Mediana	Mínim	Màxim
1	Intern	173	7,497	1,955	8	0	10
	Surt	54	6,685	2,502	7	0	10
2	Intern	173	5,254	2,283	5	0	10
	Surt	54	5,074	2,655	5	0	10
3	Intern	173	5,318	2,669	5	0	10
	Surt	54	4,722	3,062	5	0	10
4	Intern	173	5,445	2,527	6	0	10
	Surt	54	4,815	2,578	5	0	10
5	Intern	173	5,243	2,389	5	0	10
	Surt	54	5,148	2,771	6	0	10
6	Intern	173	2,035	2,307	2	0	10
	Surt	54	1,815	2,173	1	0	8

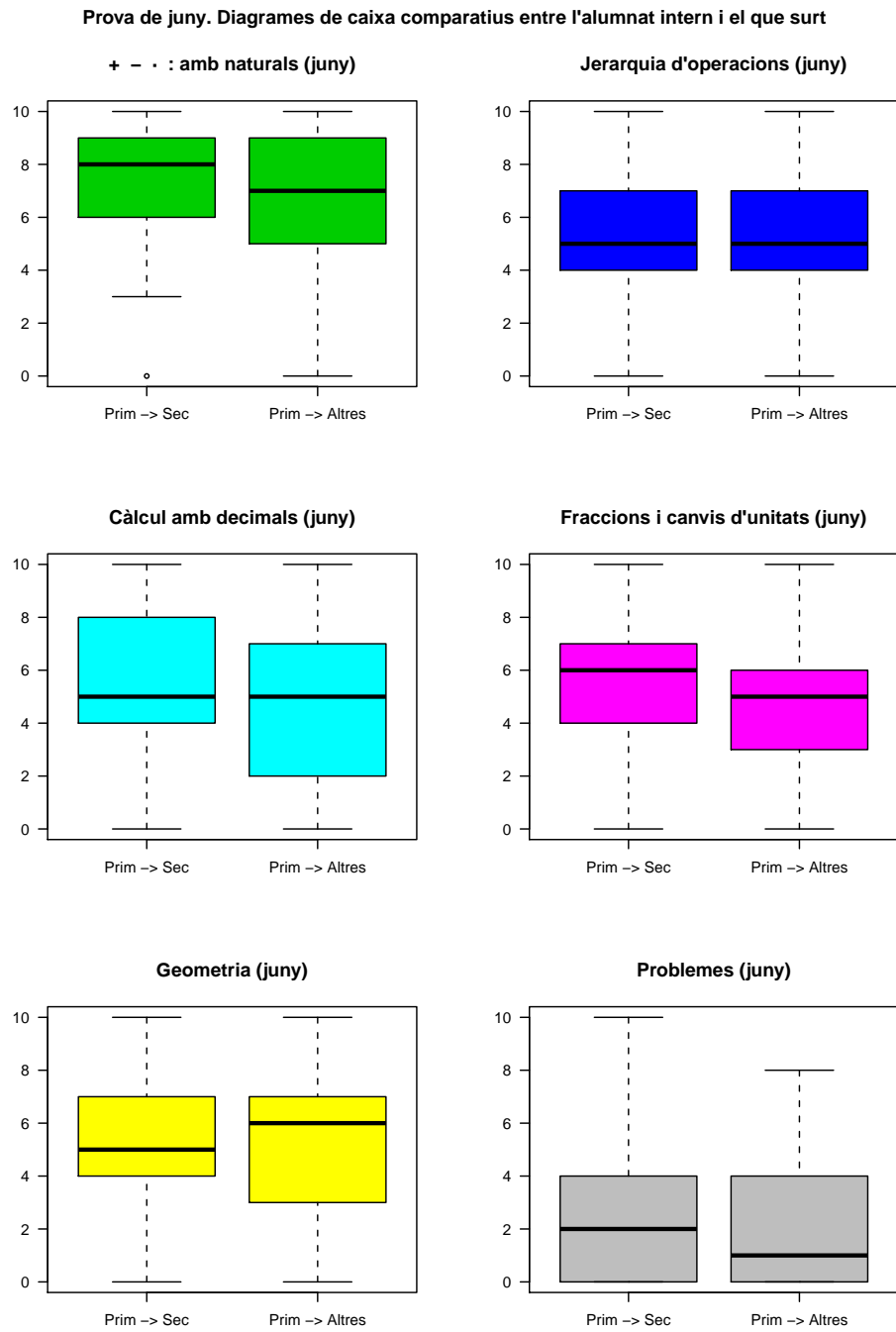
**Taula 3:** Descriptiva de les puntuacions en els 6 blocs de la prova del juny segons si l'alumnat es va matricular al setembre a un dels 3 IES de l'estudi ("Alumnat intern") o no ("Alumnat que surt"). Només s'han considerat els casos complets.

Bloc	Estimació	Error estàndar	$t$	$p$ -valor
+ - · : amb naturals (juny)	-0,8119	0,3268	-2,4845	0,0137
Jerarquia d'operacions (juny)	-0,1803	0,3704	-0,4866	0,6270
Càlcul amb decimals (juny)	-0,5957	0,4313	-1,3812	0,1686
Fraccions i canvis d'unitats (juny)	-0,6303	0,3958	-1,5922	0,1127
Geometria (juny)	-0,0946	0,3873	-0,2444	0,8072
Problemes (juny)	-0,2199	0,3548	-0,6196	0,5361

**Taula 4:** Resultats del  $t$ -test per a la comparació de mitjanes en els resultats dels diferents blocs de la prova del juny. L'estimació és per a la diferència corresponent a la mitjana en l'"Alumnat que surt" menys la mitjana en l'"Alumnat intern".



**Figura 4:** Histogrames de les puntuacions en els 6 blocs de la prova del juny segons si l'alumnat es va matricular al setembre a un dels 4 IES de l'estudi ("Alumnat intern") o no ("Alumnat que surt"). La línia discontinua en cada histograma indica la mitjana corresponent. Només s'han considerat els casos complets.



**Figura 5:** Diagrames de caixa de les puntuacions en els 6 blocs de la prova del juny segons si l'alumnat es va matricular al setembre a un dels 4 IES de l'estudi ("Alumnat intern") o no ("Alumnat que surt").

Bloc	Alumnat	N	Mitjana	Desv. Est.	Mediana	Mínim	Màxim
1	Intern	176	6,767	2,332	7	0	10
	Entra	48	6,354	2,037	7	2	10
2	Intern	176	4,807	2,460	5	0	10
	Entra	48	3,854	2,306	4	0	10
3	Intern	176	4,273	2,624	4	0	10
	Entra	48	3,646	2,419	3,5	0	9
4	Intern	176	4,875	2,506	5	0	10
	Entra	48	3,583	2,112	3	0	9
5	Intern	176	4,830	2,326	5	0	9
	Entra	48	4,396	2,541	4	0	9
3	Intern	176	1,636	2,063	0	0	8
	Entra	48	0,625	1,248	0	0	4

**Taula 5:** Descriptiva de les puntuacions en els sis blocs de la prova del setembre segons si l'alumnat provenia d'un dels 6 CEIPs de l'estudi ("Alumnat intern") o no ("Alumnat que entra"). Només s'han considerat els casos complets.

#### 4.2.2 Alumnat que entra al setembre

De manera anàloga a la secció 4.2.1, es va comparar l'"Alumnat intern" i l'"Alumnat que entra", en funció dels resultats en la prova de setembre.

La Taula 5 (pàgina 16) mostra un resum descriptiu dels resultats en cadascun dels 6 Blocs de la prova de setembre, per a cadascun dels dos grups d'alumnat. Novament s'hi observa una variabilitat similar en els resultats (desviació estàndar) però l'"Alumnat que entra" va mostrar mitjanes força pitjors en els Blocs 2, 4 i 6.

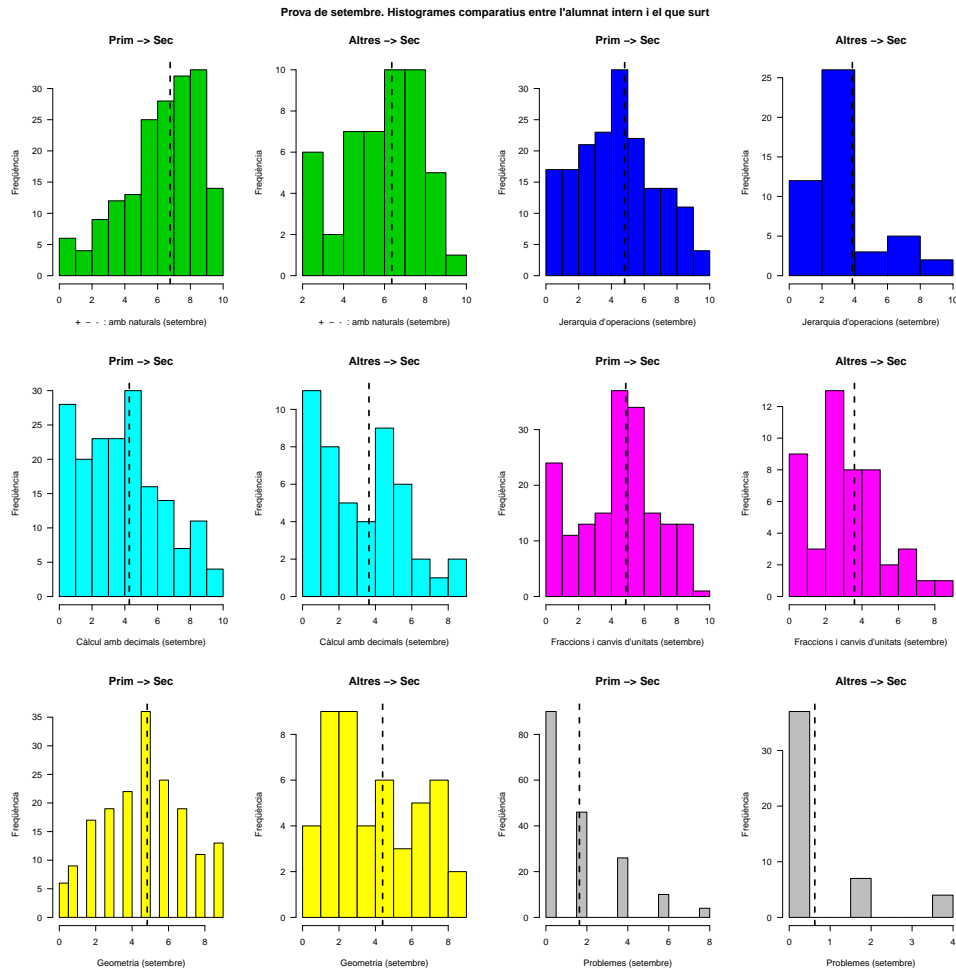
Aquestes diferències també es poden apreciar en la descriptiva gràfica mitjançant histogrames (Figura 6, pàgina 17) i diagramens de caixa (Figura 7, pàgina 18).

Els resultats del corresponent  $t$ -test mostrats en la Taula 6 (pàgina 17) confirmen que són significatives (columna  $p$ -valor) les mitjanes més baixes entre l'"Alumnat que entra" en els Blocs 2 (Jerarquia d'operacions), 4 (Fraccions i canvis d'unitats) i 6 (Problemes).

Tres possibles motius d'aquestes diferències són, en primer lloc, que l'"Alumnat intern" ja havia fet la prova el mes de juny; en segon lloc, que la prova es va dissenyar en consonància amb el currículum treballat en els 6 CEIPs estudiats; en tercer lloc, la possibilitat que part de l'alumnat nou incorporat al setembre sigui nouvingut i tingui més dificultats acadèmiques.

### 4.3 Individus finalment inclosos en l'estudi segons els casos complets en el flux net intern

Els resultats obtinguts en la secció 4.2 es poden interpretar com que, des del punt de vista dels resultats en la prova de juny, no va haver diferències significatives

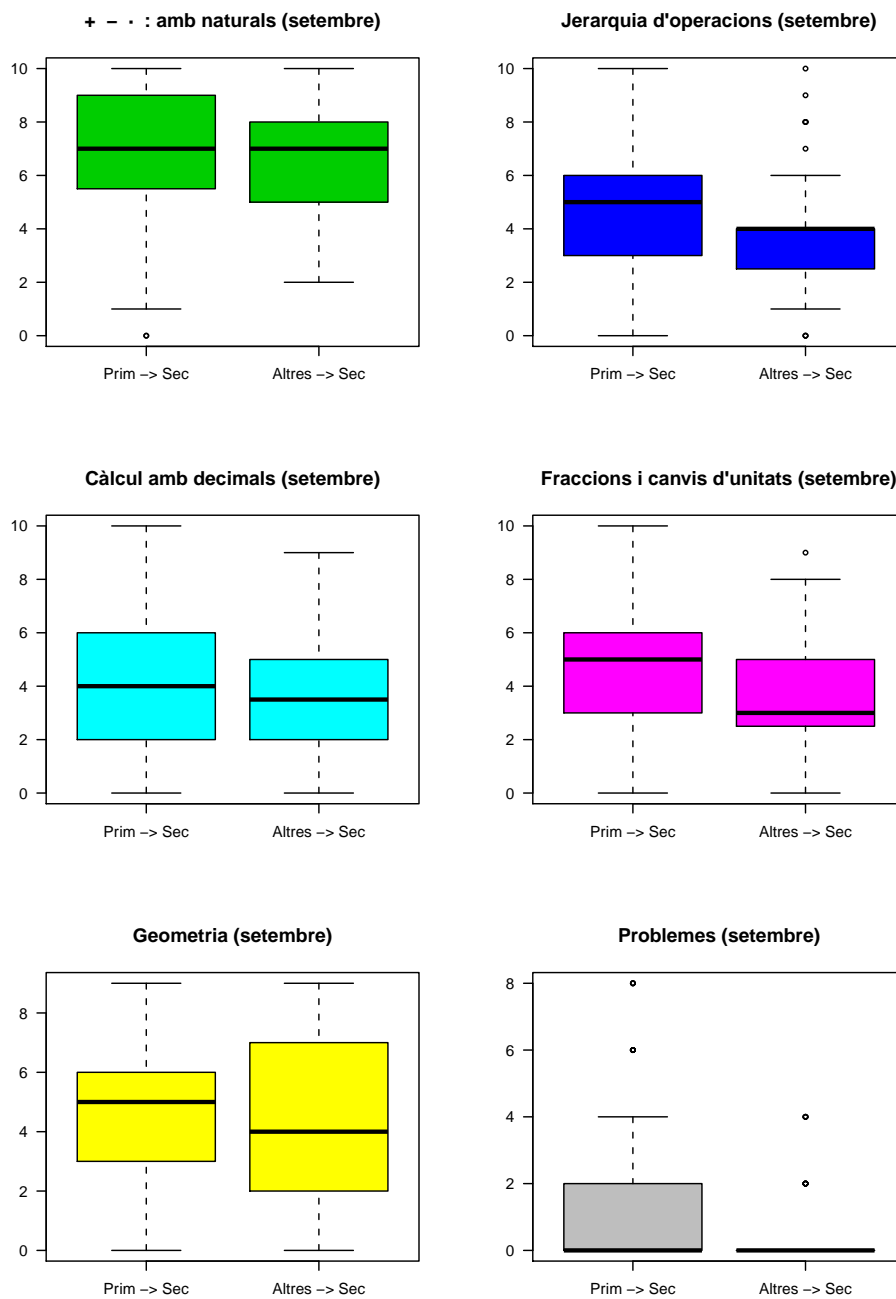


**Figura 6:** Histogrames de les puntuacions en els 6 blocs de la prova del setembre segons si l'alumnat provenia d'un dels 6 CEIPs de l'estudi ("Alumnat intern") o no ("Alumnat que entra"). La línia discontinua en cada histograma indica la mitjana corresponent. Només s'han considerat els casos complets.

Bloc	Estimació	Error estàndar	$t$	$p$ -valor
+ - · : amb naturals (setembre)	-0,4129	0,3700	-1,1158	0,2657
Jerarquia d'operacions (setembre)	-0,9527	0,3955	-2,4089	0,0168
Càlcul amb decimals (setembre)	-0,6269	0,4204	-1,4911	0,1374
Fraccions i canvis d'unitats (setembre)	-1,2917	0,3954	-3,2668	0,0013
Geometria (setembre)	-0,4337	0,3864	-1,1223	0,2629
Problemes (setembre)	-1,0114	0,3126	-3,2357	0,0014

**Taula 6:** Resultats del  $t$ -test per a la comparació de mitjanes en els resultats dels diferents blocs de la prova del setembre. L'estimació és per a la diferència corresponent a la mitjana en l'"Alumnat que entra" menys la mitjana en l'"Alumnat intern".

**Prova de setembre. Diagrames de caixa comparatius entre l'alumnat intern i el que surt**



**Figura 7:** Diagrames de caixa de les puntuacions en els 6 blocs de la prova del setembre segons si l'alumnat provenia d'un dels 6 CEIPs de l'estudi ("Alumnat intern") o no ("Alumnat que entra").



entre l'“Alumnat intern” i l'“Alumnat que surt” i, per tant, es pot acceptar que la “pèrdua” d'alumnat va ser aleatòria.

En canvi es va comprovar que l'“Alumnat que entra” va obtenir resultats significativament pitjors en la prova de setembre en 3 dels 6 Blocs.

En tot cas, per a realitzar un estudi longitudinal entre juny i desembre de l'alumnat, es va seleccionar l'“Alumnat intern” perquè va ser l'alumnat que va realitzar la prova tant al juny com al setembre.

D'aquest grup d'alumnat, però, només es va seleccionar el subconjunt que va rebre qualificació en l'àrea de Matemàtiques en la primera avaluació de 1r d'ESO (desembre) i mostrava tota la informació referent a la resta de variables analitzades. D'aquesta manera, el grup “Alumnat intern”, de mida 181 (Figura 1, pàgina 8) va derivar en la mostra final formada 168 individus.

La resta de l'estudi es va realitzar sobre aquest grup d'alumnat.

## 5 Anàlisi descriptiva

Es va procedir a realitzar una anàlisi descriptiva de les variables en estudi, després de seleccionar l'alumnat segons s'ha exposat en la secció 4.3.

La descripció dels factors (variables explicatives categòriques), es mostra en forma de percentatges i de diagrames de sectors en la Figura 8 (pàgina 20).

S'hi observa que només l'1% de l'alumnat analitzat va repetir l'últim curs de Primària. Aquest fet va implicar que aquest factor no fos significatiu en cap dels models ajustats; és a dir, la no repetició sistemàtica impossibilita que es pugui analitzar el mode que afecta la repetició a resultats acadèmics posteriors.

Tanmateix, s'observa el mateix amb els factors corresponents a l'alumnat nouvingut (factor “Vingut a Catalunya fa menys de 2 anys”) i a l'alumnat amb Necessitats Educatives Especials (factor “NEE”). Tot i això, el percentatge d'alumnat nouvingut i d'alumnat NEE va ser suficient per tal que aquests factors fossin significatius en alguns dels models ajustats<sup>6</sup>.

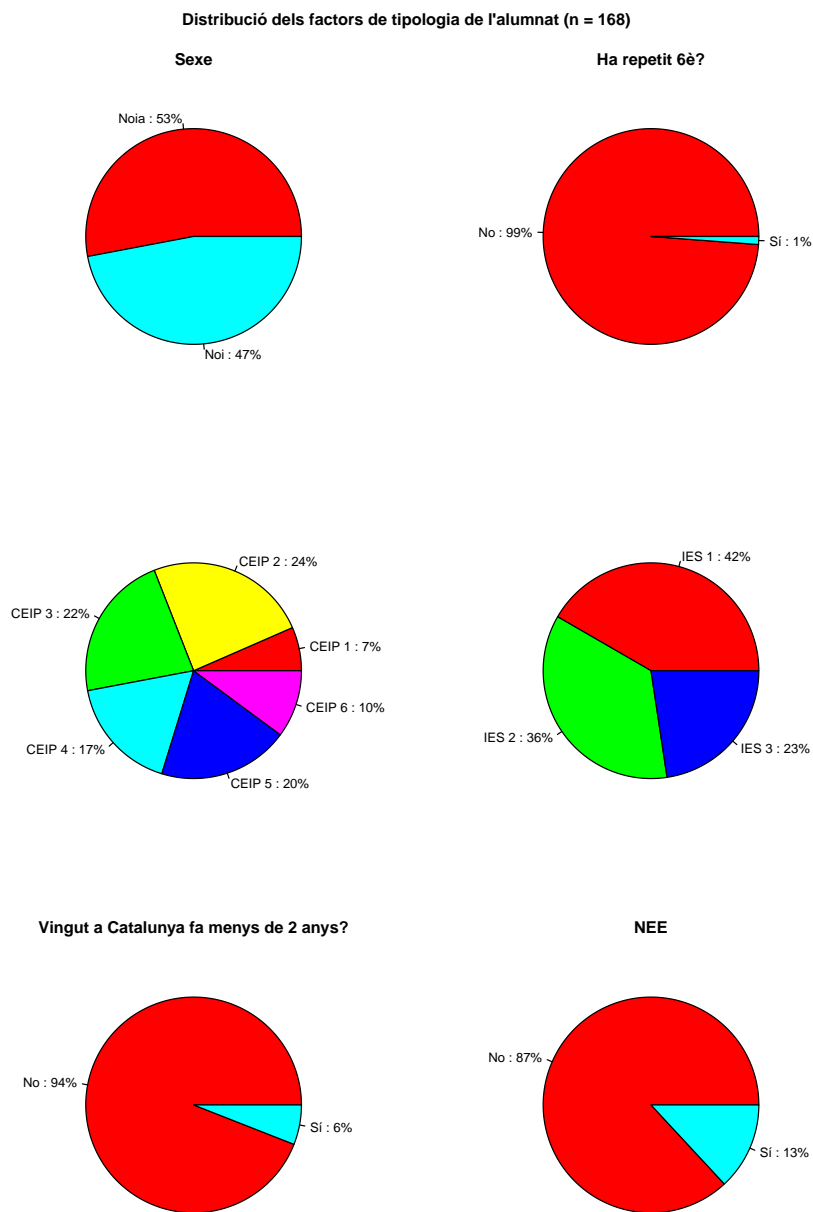
La resta de factors van representar equilibradament totes les categories.

La Figura 9 (pàgina 21) mostra la distribució de l'alumnat segons el mes de naixement. Aquest factor es va incloure en l'estudi per la possibilitat d'existència de diferències significatives entre l'alumnat no repetidor més jove i el menys jove d'un mateix curs.

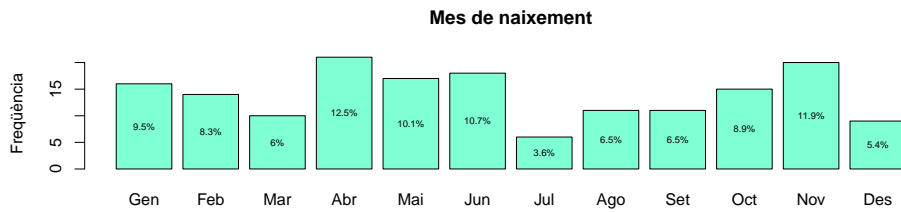
L'edat es va considerar tant com factor com a covariable (variable explicativa numèrica). El gràfic de l'esquerra de la Figura 10 mostra la distribució de l'edat entera (factor) de l'alumnat on s'observa la petita representació d'edats diferents als 12 anys. Aquest fet va motivar la conveniència de considerar també l'edat com a covariable (gràfic de la dreta). L'aspecte d'aquest últim histograma suggereix la categorització de l'edat per quantils (en categories de freqüències

---

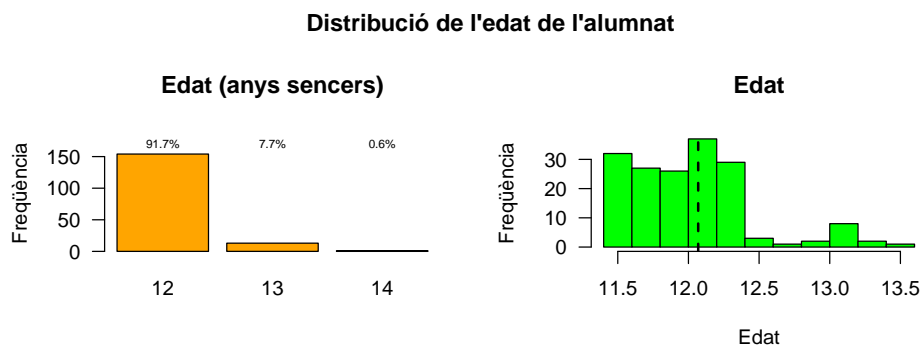
<sup>6</sup>L'anàlisi de la taula de contingència que classifica l'alumnat en funció d'aquests dos factors va permetre acceptar que el percentatge d'alumnat NEE va ser independent del fet de tractar-se o no d'alumnat nouvingut.



**Figura 8:** Distribució de freqüències dels factors corresponents a la mostra final considerada en l'estudi.



**Figura 9:** Distribució del mes de naixement corresponent a la mostra final considerada en l'estudi.



**Figura 10:** Distribució de l'edat corresponent a la mostra final considerada en l'estudi.

Variable	Mitjana	Desv. Est.	Mediana	Mínim	Màxim
Edat	12,1	0,4	12	11,5	13,5

**Taula 7:** Resum descriptiu de la covariable *Edat*.

Variable	Mitjana	Desv. Est.	Mediana	Mínim	Màxim
B1P	7,5	1,9	8,0	0	10,0
B2P	5,2	2,3	5,0	0	10,0
B3P	5,3	2,7	5,0	0	10,0
B4P	5,5	2,5	6,0	0	10,0
B5P	5,2	2,4	5,0	0	10,0
B6P	2,0	2,3	2,0	0	10,0
GP	5,1	1,9	5,2	0	9,8
B1S	6,8	2,3	7,0	0	10,0
B2S	4,8	2,5	5,0	0	10,0
B3S	4,3	2,6	4,0	0	10,0
B4S	5,0	2,5	5,0	0	10,0
B5S	4,9	2,3	5,0	0	9,0
B6S	1,7	2,1	2,0	0	8,0
GS	4,6	1,9	4,6	0	8,8
dif.B1	-0,7	2,0	-1,0	-6	4,0
dif.B2	-0,4	1,9	0,0	-5	5,0
dif.B3	-1,0	2,0	-1,0	-9	4,0
dif.B4	-0,5	1,6	0,0	-6	3,0
dif.B5	-0,4	1,9	0,0	-6	4,0
dif.B6	-0,3	1,9	0,0	-4	6,0
dif.G	-0,6	1,0	-0,5	-5	2,2
Aval1	5,4	1,9	5,0	2	10,0

**Taula 8:** Resum descriptiu de les covariables resposta. Resultats en la prova de juny, en la prova de setembre, en la diferència entre les dues proves i en l'àrea de Matemàtiques en la primera avaluació de 1r curs d'ESO.

iguals). De fet, en els models on l'edat va resultar significativa, la millor manera d'introduir-la (quant a significació) va ser segons una categorització per quantils<sup>7</sup>. La descripció numèrica de l'edat es mostra en la Taula 7 (pàgina 22).

Pel que fa a la resta de covariables, la Taula 8 (pàgina 22) mostra el corresponent resum numèric on es pot observar la pèrdua de qualificació ( $\text{dif.Bi} = \text{BiS} - \text{BiP}$ ,  $i = 1, \dots, 6$ ) entre juny ( $\text{BiP}$ ) i setembre ( $\text{BiS}$ ), en cadascun dels 6 Blocs. També es mostren les qualificacions globals en la prova de juny ( $\text{GP} \times$ ) i de setembre ( $\text{GS}$ ) i en la primera avaluació de 1r curs d'ESO en l'àrea de Matemàtiques ( $\text{Aval1}$ ).

Resulta interessant destacar que la qualificació mitjana en la primera avaluació de 1r curs d'ESO en l'àrea de Matemàtiques va ser superior a la qualificació mitjana de la prova al juny i al setembre.

Es va realitzar una anàlisi descriptiva també de les variables resposta. La

<sup>7</sup>Es van provar categoritzacions de 2, 3 i 4 categories.

Histogrammes de les qualificacions segons bloc i prova (juny, setembre i diferència)

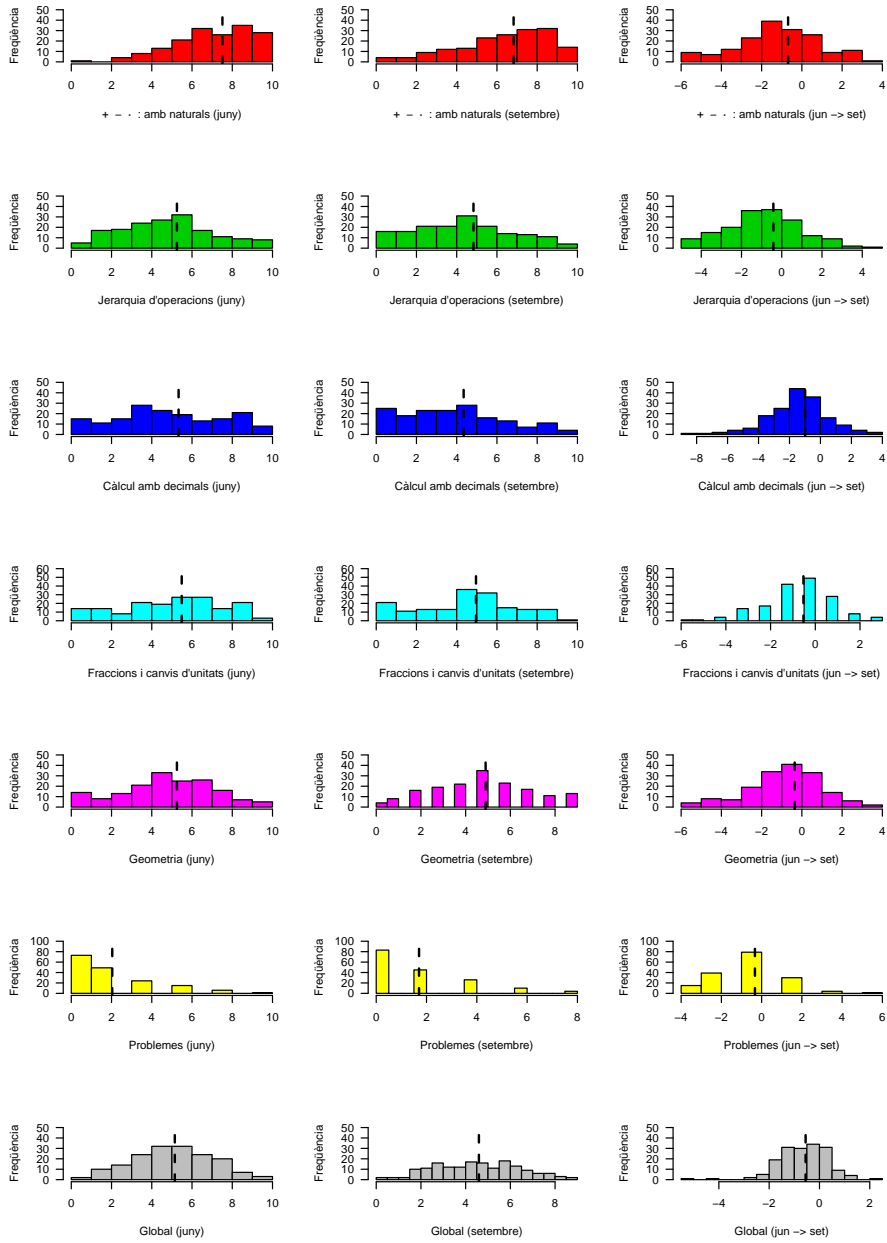
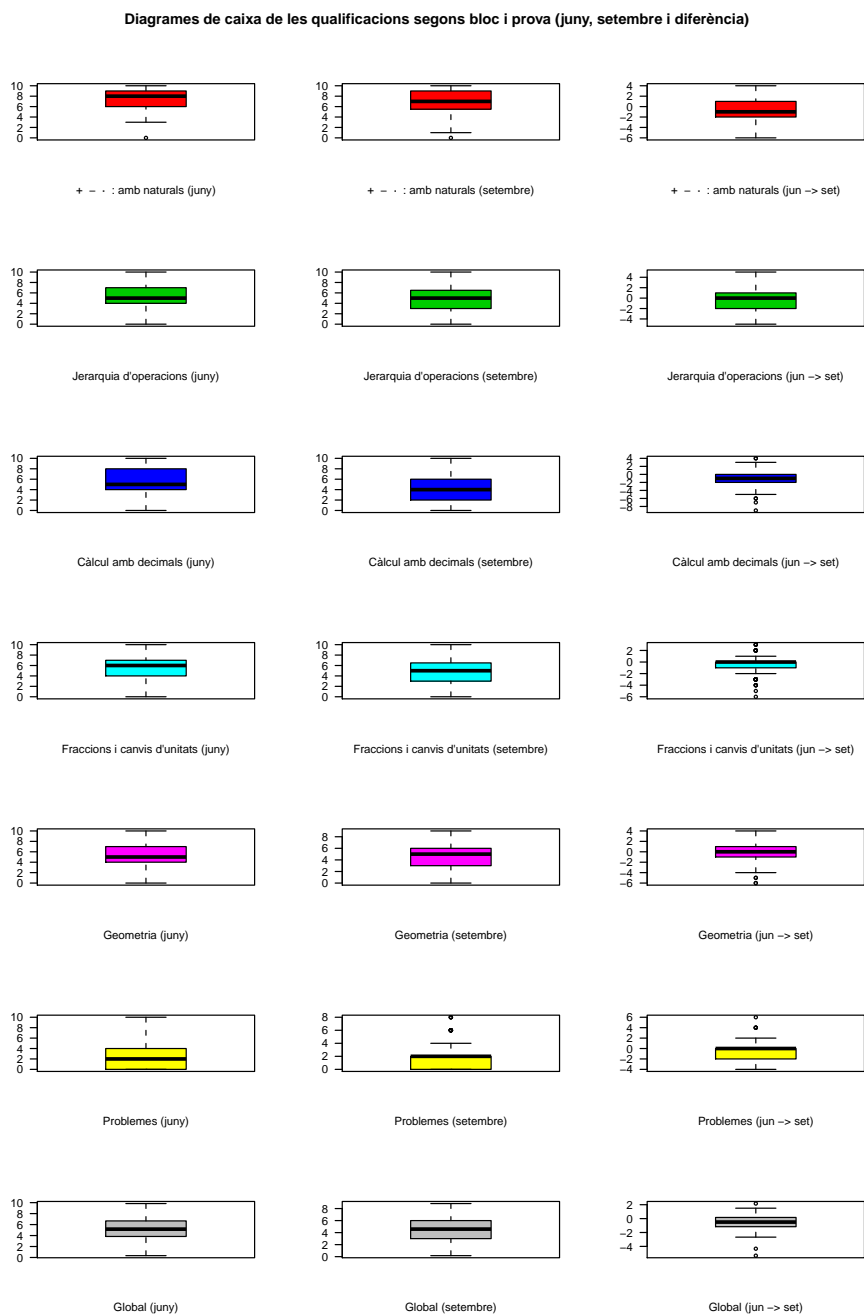
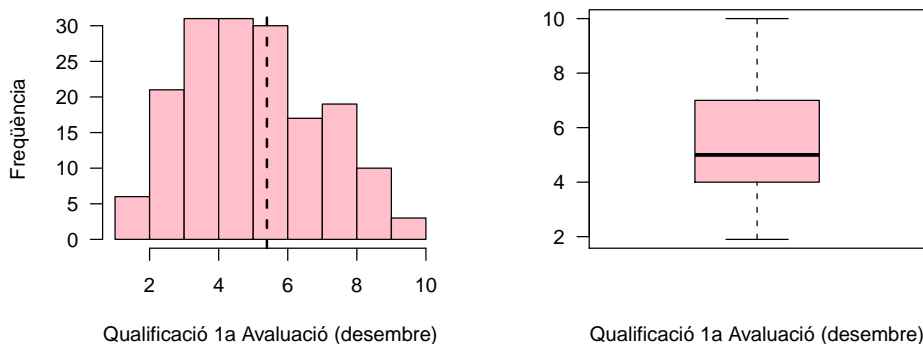


Figura 11: Distribució de les qualificacions de la prova al juny, al setembre i de la diferència entre el juny i el setembre.



**Figura 12:** Diagrames de caixa de les qualificacions de la prova al juny, al setembre i de la diferència entre el juny i el setembre.

**Distribució de les qualificacions en la primera avaluació (1r d'ESO)**



**Figura 13:** Distribució de les qualificacions en la primera avaluació de 1r d'ESO (desembre).

Figura 11 (pàgina 23) i la Figura 12 (pàgina 24) mostren la representació gràfica (histogrames i diagrames de caixa, respectivament) dels resultats de la prova al juny i al setembre –i també de la diferència entre juny i setembre– per a cadascun dels 6 Blocs. Una descripció gràfica similar es presenta en la Figura 13 (pàgina 25).

Finalment, es va analitzar la correlació entre la qualificació en la prova de juny i la de la prova de setembre per a cadascun dels Blocs. Les correlacions parcials es mostren en la Taula 9 (pàgina 26) on s'observa que el valor més gran implica una correlació parcial inferior al 47%<sup>8</sup>. Aquestes correlacions tenen sentit positiu, com es pot observar en els diagrames de dispersió correponents (Figura 14, pàgina 26); és a dir, com més elevades són les qualificacions en la prova de juny, més elevades tendeixen a ser també en la de setembre.

## 6 Modelització dels resultats en la prova de juny

En aquesta secció es presenta la modelització dels resultats en la prova de juny. Es va procedir a l'ajustament d'un Model Lineal Mixt, considerant com a variable resposta la qualificació numèrica (entre 0 i 10) corresponent a un Bloc de la prova de juny. Com a efectes fixos es van considerar tots els factors i covariables possibles que es descriuen en la Taula 1 (pàgina 1), incloent-hi el factor Bloc. Com a efecte aleatori es va considerar l'individu.

Així, el model ajustat és de la forma

$$J = X\beta + Iv + e \tag{1}$$

<sup>8</sup> $100 \cdot 0,683^2 \approx 46,6\%$ .

	B1P	B2P	B3P	B4P	B5P	B6P
B1P	1,000	0,583	0,627	0,536	0,594	0,442
B2P		1,000	0,683	0,645	0,581	0,589
B3P			1,000	0,644	0,636	0,534
B4P				1,000	0,636	0,627
B5P					1,000	0,535
B6P						1,000

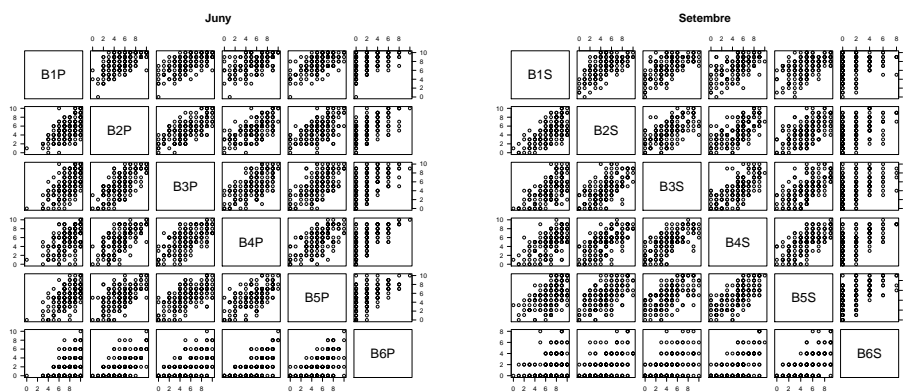
  

	B1S	B2S	B3S	B4S	B5S	B6S
B1S	1,000	0,640	0,628	0,601	0,532	0,338
B2S		1,000	0,639	0,658	0,599	0,454
B3S			1,000	0,678	0,622	0,470
B4S				1,000	0,682	0,546
B5S					1,000	0,581
B6S						1,000

B1PS	B2PS	B3PS	B4PS	B5PS	B6PS
0,541	0,677	0,720	0,805	0,683	0,638

**Taula 9:** Correlacions parcials 2 a 2 entre les qualificacions obtingudes per l'alumnat que va fer la prova tant al juny (P) com al setembre (S), i correlacions, per a cada Bloc, entre juny i setembre.



**Figura 14:** Diagrames de dispersió 2 a 2 de les qualificacions de la prova de juny i de la prova de setembre.



	Estimació ( $\beta$ )	Error est.	Wald	p-valor
B1P: + - : amb naturals	7,36	0,59	155,93	0,0000
B2P: Jerarquia d'operacions	5,16	0,59	75,70	0,0000
B3P: Càlcul amb decimals	5,27	0,61	75,41	0,0000
B4P: Fraccions i canvis d'unitats	5,38	0,59	83,80	0,0000
B5P: Geometria	5,17	0,59	76,00	0,0000
B6P: Problemes	1,94	0,58	11,38	0,0007
Cat2	-1,39	0,53	6,88	0,0087
NEE	-2,47	0,31	62,52	0,0000
CEIP 2	1,48	0,63	5,62	0,0177
CEIP 3	-0,38	0,63	0,37	0,5408
CEIP 4	0,19	0,68	0,08	0,7797
CEIP 5	0,69	0,63	1,20	0,2736
CEIP 6	0,21	0,66	0,10	0,7519

**Taula 10:** Estimació dels coeficients en el model per als resultats de la prova de juny.

on  $J$  és la qualificació (de 0 a 10),  $X$  és el vector format per tots els factors i covariables d'efectes fixos,  $\beta$  és el vector de coeficients corresponent als efectes fixos,  $I$  és la variable identificadora de l'individu,  $v$  és el coeficient corresponent a l'efecte aleatori introduït per l'individu i  $e$  és el terme d'error del model.

Tant el coeficient corresponent a l'efecte aleatori de l'individu,  $v$ , com el terme d'error,  $e$ , se suposen amb distribució normal, esperança zero, variància constant i independència entre sí. Així, les prediccions segons el model (1) són

$$\hat{J} = X\hat{\beta}$$

i, per tant, cal estimar els paràmetres  $\beta$ . Aquesta estimació es fa per màxima versemblança.

## 6.1 Model ajustat

Després del procés de modelització, el millor model ajustat va ser el següent:

$$J = \beta_1 \text{Bloc} + \beta_2 \text{Cat2} + \beta_3 \text{NEE} + \beta_4 \text{CEIP} + Iv + e$$

i les prediccions

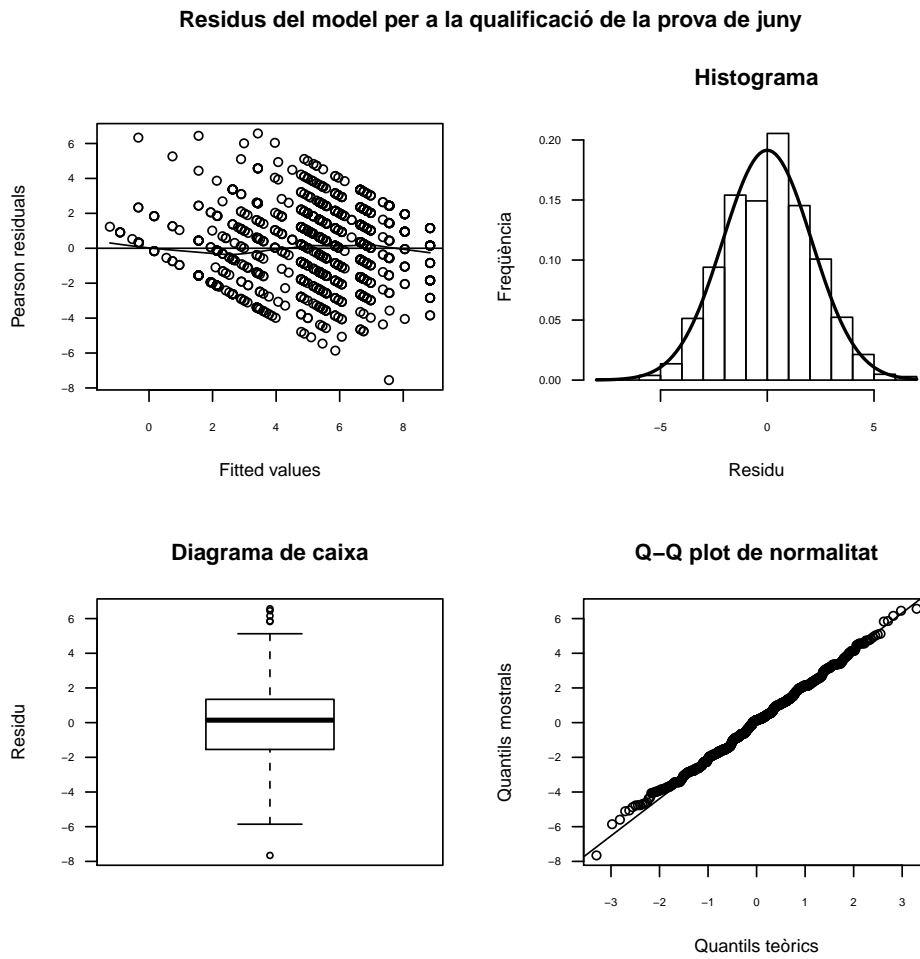
$$\hat{J} = \hat{\beta}_1 \text{Bloc} + \hat{\beta}_2 \text{Cat2} + \hat{\beta}_3 \text{NEE} + \hat{\beta}_4 \text{CEIP}, \quad (2)$$

de manera que la qualificació de la prova de juny depenia del Bloc considerat, si l'individu s'havia incorporat a Catalunya feia menys de dos anys, si era tipificat NEE i del CEIP. Les estimacions dels coeficients del model es mostren a la Taula 10 (pàgina 27).

Tanmateix, la Figura 15 (pàgina 28) mostra un comportament acceptable dels residus ( $e$ ) del model i, per tant, es va donar el model per validat.

## 6.2 Interpretació del model ajustat i dels seus paràmetres

La interpretació del model es fa combinant l'equació del model (2) i la Taula 10 i es fa de la manera següent.



**Figura 15:** Comportament dels residus del model per als resultats de la prova de juny.

Fixant els nivells de referència en tots els factors –no incorporat a Catalunya fa menys de dos anys i no NEE i pertanyent al CEIP 1– el model prediu una qualificació en cada Bloc igual a l'estimació corresponent ( $\hat{\beta}$ ). Per exemple, en el Bloc 1 la predicció per a aquest perfil d'alumnat és 7,36.

A partir d'aquí, si varia el nivell en algun factor, s'hi ha de sumar l'estimació que correspon a l'efecte considerat. Així, si, per exemple, l'individu sí que es va incorporar a Catalunya fa menys de 2 anys, s'hi ha de restar 1,39; si l'individu és tipificat NEE, s'hi ha de restar 2,47; i si l'individu pertanyia al CEIP 5 en lloc de al CEIP 1, s'hi ha de sumar 0,69.

No obstant això, observant el  $p$ -valor, és palès que cap CEIP no és significativament diferent del CEIP 1 quant a resultats, excepte el CEIP 2, que es va mostrar millor (coeficient positiu).

D'aquesta manera, el model ajustat prediu que:

- El Bloc 1 (+ - · : amb naturals) presenta una qualificació acceptable (7,36), si es fixen els nivells de referència en la resta de factors.
- El Bloc 6 (Problemes) presenta una qualificació molt baixa (1,94), si es fixen els nivells de referència en la resta de factors.
- La resta de Blocs presenten una qualificació lleugerament per sobre del 5, si es fixen els nivells de referència en la resta de factors.
- Un individu que es va incorporar a Catalunya fa menys de 2 anys perd 1,39 punts en qualsevol Bloc, a diferència de si no ho va fer.
- Un individu tipificat NEE perd 2,47 punts en qualsevol Bloc, a diferència de si no ho és.

### 6.3 Prediccions del model ajustat

Les prediccions –puntuals i també per interval de confiança al 95%– que el model ajustat fa en la prova de juny, per a tots els tractaments possibles obtinguts en la combinació de les variables **Cat2**, **NEE** i **Cprim**, es mostren en la Taula 11 (pàgina 30), i corresponen als Blocs 1, 2 i 3, i també en la Taula 12 (pàgina 31), que corresponen als Blocs 4, 5 i 6.

Tanmateix, la Taula 13 (pàgina 32) mostra, per a cadascun dels 6 Blocs, els 5 pitjors i els 5 millors tractaments on es pot observar clarament que l'alumnat incorporat a Catalunya fa menys de 2 anys i la tipificació NEE representen factors de risc.

Bloc	Cat2	NEE	Cprim	Predicció	L.I. (95%)	L.S. (95%)
+ - · : amb naturals	No		CEIP 1	7,4	6,8	7,9
			CEIP 2	8,8	8,5	9,2
			CEIP 3	7,0	6,6	7,4
			CEIP 4	7,6	7,1	8,0
			CEIP 5	8,0	7,6	8,5
			CEIP 6	7,6	7,1	8,1
		Sí	CEIP 1	4,9	4,2	5,6
			CEIP 2	6,4	5,9	6,9
			CEIP 3	4,5	4,0	5,0
			CEIP 4	5,1	4,5	5,6
			CEIP 5	5,6	5,1	6,1
			CEIP 6	5,1	4,5	5,7
	Sí	No	CEIP 1	6,0	5,2	6,7
			CEIP 2	7,4	6,8	8,1
			CEIP 3	5,6	5,0	6,2
			CEIP 4	6,2	5,6	6,8
			CEIP 5	6,7	6,0	7,3
			CEIP 6	6,2	5,5	6,9
		Sí	CEIP 1	3,5	2,7	4,3
			CEIP 2	5,0	4,2	5,7
			CEIP 3	3,1	2,4	3,8
			CEIP 4	3,7	3,0	4,4
			CEIP 5	4,2	3,5	4,9
			CEIP 6	3,7	2,9	4,5
Jerarquia d'operacions	No		CEIP 1	5,2	4,6	5,7
			CEIP 2	6,6	6,3	7,0
			CEIP 3	4,8	4,4	5,2
			CEIP 4	5,4	4,9	5,8
			CEIP 5	5,9	5,4	6,3
			CEIP 6	5,4	4,9	5,9
		Sí	CEIP 1	2,7	2,0	3,4
			CEIP 2	4,2	3,7	4,7
			CEIP 3	2,3	1,8	2,8
			CEIP 4	2,9	2,3	3,4
			CEIP 5	3,4	2,9	3,9
			CEIP 6	2,9	2,3	3,5
	Sí	No	CEIP 1	3,8	3,0	4,5
			CEIP 2	5,3	4,6	5,9
			CEIP 3	3,4	2,8	4,0
			CEIP 4	4,0	3,4	4,6
			CEIP 5	4,5	3,8	5,1
			CEIP 6	4,0	3,3	4,7
		Sí	CEIP 1	1,3	0,5	2,1
			CEIP 2	2,8	2,0	3,5
			CEIP 3	0,9	0,2	1,6
			CEIP 4	1,5	0,8	2,2
			CEIP 5	2,0	1,3	2,7
			CEIP 6	1,5	0,7	2,3
Càlcul amb decimals	No		CEIP 1	5,3	4,7	5,8
			CEIP 2	6,8	6,4	7,1
			CEIP 3	4,9	4,5	5,3
			CEIP 4	5,5	5,0	5,9
			CEIP 5	6,0	5,5	6,4
			CEIP 6	5,5	5,0	6,0
		Sí	CEIP 1	2,8	2,1	3,5
			CEIP 2	4,3	3,8	4,8
			CEIP 3	2,4	1,9	2,9
			CEIP 4	3,0	2,5	3,5
			CEIP 5	3,5	3,0	4,0
			CEIP 6	3,0	2,4	3,6
	Sí	No	CEIP 1	3,9	3,1	4,6
			CEIP 2	5,4	4,7	6,0
			CEIP 3	3,5	2,9	4,1
			CEIP 4	4,1	3,5	4,7
			CEIP 5	4,6	3,9	5,2
			CEIP 6	4,1	3,4	4,8
		Sí	CEIP 1	1,4	0,6	2,2
			CEIP 2	2,9	2,1	3,6
			CEIP 3	1,0	0,3	1,7
			CEIP 4	1,6	0,9	2,3
			CEIP 5	2,1	1,4	2,8
			CEIP 6	1,6	0,8	2,4

**Taula 11:** Prediccions (puntuals i per interval de confiança al 95%) segons el model ajustat per a les qualificacions en la prova de juny. Blocs 1, 2 i 3.

Bloc	Cat2	NEE	Cprim	Predicció	L.I. (95%)	L.S. (95%)
Fraccions i canvis d'unitats	No		CEIP 1	5,4	4,8	6,0
			CEIP 2	6,9	6,5	7,2
			CEIP 3	5,0	4,6	5,4
			CEIP 4	5,6	5,1	6,0
			CEIP 5	6,1	5,7	6,5
			CEIP 6	5,6	5,1	6,1
		Sí	CEIP 1	2,9	2,2	3,6
			CEIP 2	4,4	3,9	4,9
			CEIP 3	2,5	2,0	3,0
			CEIP 4	3,1	2,6	3,6
			CEIP 5	3,6	3,1	4,1
			CEIP 6	3,1	2,5	3,7
	Sí	No	CEIP 1	4,0	3,2	4,7
			CEIP 2	5,5	4,8	6,1
			CEIP 3	3,6	3,0	4,2
			CEIP 4	4,2	3,6	4,8
			CEIP 5	4,7	4,0	5,3
			CEIP 6	4,2	3,5	4,9
		Sí	CEIP 1	1,5	0,7	2,3
			CEIP 2	3,0	2,3	3,7
			CEIP 3	1,1	0,4	1,8
			CEIP 4	1,7	1,0	2,4
			CEIP 5	2,2	1,5	2,9
			CEIP 6	1,7	0,9	2,5
Geometria	No		CEIP 1	5,2	4,6	5,8
			CEIP 2	6,7	6,3	7,0
			CEIP 3	4,8	4,4	5,2
			CEIP 4	5,4	4,9	5,8
			CEIP 5	5,9	5,5	6,3
			CEIP 6	5,4	4,9	5,9
		Sí	CEIP 1	2,7	2,0	3,4
			CEIP 2	4,2	3,7	4,7
			CEIP 3	2,3	1,8	2,8
			CEIP 4	2,9	2,4	3,4
			CEIP 5	3,4	2,9	3,9
			CEIP 6	2,9	2,3	3,5
	Sí	No	CEIP 1	3,8	3,0	4,5
			CEIP 2	5,3	4,6	5,9
			CEIP 3	3,4	2,8	4,0
			CEIP 4	4,0	3,4	4,6
			CEIP 5	4,5	3,8	5,1
			CEIP 6	4,0	3,3	4,7
		Sí	CEIP 1	1,3	0,5	2,1
			CEIP 2	2,8	2,0	3,5
			CEIP 3	0,9	0,2	1,6
			CEIP 4	1,5	0,8	2,2
			CEIP 5	2,0	1,3	2,7
			CEIP 6	1,5	0,7	2,3
Problemes	No		CEIP 1	1,9	1,4	2,5
			CEIP 2	3,4	3,0	3,8
			CEIP 3	1,6	1,2	2,0
			CEIP 4	2,1	1,7	2,6
			CEIP 5	2,6	2,2	3,0
			CEIP 6	2,2	1,6	2,7
		Sí	CEIP 1	0,0	0,0	0,2
			CEIP 2	1,0	0,4	1,5
			CEIP 3	0,0	0,0	0,0
			CEIP 4	0,0	0,0	0,2
			CEIP 5	0,2	0,0	0,7
			CEIP 6	0,0	0,0	0,3
	Sí	No	CEIP 1	0,5	0,0	1,3
			CEIP 2	2,0	1,4	2,7
			CEIP 3	0,2	0,0	0,8
			CEIP 4	0,7	0,1	1,3
			CEIP 5	1,2	0,6	1,9
			CEIP 6	0,8	0,0	1,5
		Sí	CEIP 1	0,0	0,0	0,0
			CEIP 2	0,0	0,0	0,3
			CEIP 3	0,0	0,0	0,0
			CEIP 4	0,0	0,0	0,0
			CEIP 5	0,0	0,0	0,0
			CEIP 6	0,0	0,0	0,0

**Taula 12:** Prediccions (puntuals i per interval de confiança al 95%) segons el model ajustat per a les qualificacions en la prova de juny. Blocs 4, 5 i 6.

Bloc	Cat2	NEE	Cprim	Predicció	L.I. (95%)	L.S. (95%)
+ - : amb naturals	Sí	Sí	CEIP 3	3,1	2,4	3,8
	Sí	Sí	CEIP 1	3,5	2,7	4,3
	Sí	Sí	CEIP 4	3,7	3,0	4,4
	Sí	Sí	CEIP 6	3,7	2,9	4,5
	Sí	Sí	CEIP 5	4,2	3,5	4,9
	Sí	No	CEIP 2	7,4	6,8	8,1
	No	No	CEIP 4	7,6	7,1	8,0
	No	No	CEIP 6	7,6	7,1	8,1
	No	No	CEIP 5	8,0	7,6	8,5
	No	No	CEIP 2	8,8	8,5	9,2
Jerarquia d'operacions	Sí	Sí	CEIP 3	0,9	0,2	1,6
	Sí	Sí	CEIP 1	1,3	0,5	2,1
	Sí	Sí	CEIP 4	1,5	0,8	2,2
	Sí	Sí	CEIP 6	1,5	0,7	2,3
	Sí	Sí	CEIP 5	2,0	1,3	2,7
	Sí	No	CEIP 2	5,3	4,6	5,9
	No	No	CEIP 4	5,4	4,9	5,8
	No	No	CEIP 6	5,4	4,9	5,9
	No	No	CEIP 5	5,9	5,4	6,3
	No	No	CEIP 2	6,6	6,3	7,0
Càlcul amb decimals	Sí	Sí	CEIP 3	1,0	0,3	1,7
	Sí	Sí	CEIP 1	1,4	0,6	2,2
	Sí	Sí	CEIP 4	1,6	0,9	2,3
	Sí	Sí	CEIP 6	1,6	0,8	2,4
	Sí	Sí	CEIP 5	2,1	1,4	2,8
	Sí	No	CEIP 2	5,4	4,7	6,0
	No	No	CEIP 4	5,5	5,0	5,9
	No	No	CEIP 6	5,5	5,0	6,0
	No	No	CEIP 5	6,0	5,5	6,4
	No	No	CEIP 2	6,8	6,4	7,1
Fraccions i canvis d'unitats	Sí	Sí	CEIP 3	1,1	0,4	1,8
	Sí	Sí	CEIP 1	1,5	0,7	2,3
	Sí	Sí	CEIP 4	1,7	1,0	2,4
	Sí	Sí	CEIP 6	1,7	0,9	2,5
	Sí	Sí	CEIP 5	2,2	1,5	2,9
	Sí	No	CEIP 2	5,5	4,8	6,1
	No	No	CEIP 4	5,6	5,1	6,0
	No	No	CEIP 6	5,6	5,1	6,1
	No	No	CEIP 5	6,1	5,7	6,5
	No	No	CEIP 2	6,9	6,5	7,2
Geometria	Sí	Sí	CEIP 3	0,9	0,2	1,6
	Sí	Sí	CEIP 1	1,3	0,5	2,1
	Sí	Sí	CEIP 4	1,5	0,8	2,2
	Sí	Sí	CEIP 6	1,5	0,7	2,3
	Sí	Sí	CEIP 5	2,0	1,3	2,7
	Sí	No	CEIP 2	5,3	4,6	5,9
	No	No	CEIP 4	5,4	4,9	5,8
	No	No	CEIP 6	5,4	4,9	5,9
	No	No	CEIP 5	5,9	5,5	6,3
	No	No	CEIP 2	6,7	6,3	7,0
Problemes	Sí	Sí	CEIP 3	0,0	0,0	0,0
	Sí	Sí	CEIP 1	0,0	0,0	0,0
	Sí	Sí	CEIP 4	0,0	0,0	0,0
	Sí	Sí	CEIP 6	0,0	0,0	0,0
	Sí	Sí	CEIP 5	0,0	0,0	0,0
	Sí	No	CEIP 2	2,0	1,4	2,7
	No	No	CEIP 4	2,1	1,7	2,6
	No	No	CEIP 6	2,2	1,6	2,7
	No	No	CEIP 5	2,6	2,2	3,0
	No	No	CEIP 2	3,4	3,0	3,8

**Taula 13:** Prediccions (puntuals i per interval de confiança al 95%) per als 5 pitjors perfils i per als 5 millors, segons el model ajustat per a les qualificacions en la prova de juny.

## 7 Estimació del Risc Relatiu com a mesura d'assoliment de continguts a curt termini (juny–setembre)

Una primera aproximació per avaluar l'assoliment a curt termini de continguts relacionats amb la prova va ser mitjançant el concepte de Risc Relatiu (RR).

L'RR és una mesura comparativa de dues probabilitats. Per exemple, si la probabilitat de patir una determinada malaltia és  $\frac{1}{100}$  entre els homes i  $\frac{1}{250}$  entre les dones, l'RR de patir la malaltia si un individu és home en relació a si és dona és  $RR = \frac{P(\text{malaltia}|\text{♂})}{P(\text{malaltia}|\text{♀})} = \frac{1/100}{1/250} = 2,5$ , i direm que la probabilitat de patir la malaltia és 2,5 vegades més gran entre els homes que entre les dones.

La Taula 14 (pàgina 34) mostra taules de contingència classificant l'alumnat, en cadascun dels 6 Blocs, segons si el Bloc va ser aprovat en juny i/o en setembre. Per exemple, observant la taula de contingència referent al Bloc 1, es pot observar que 155 persones van aprovar-lo<sup>9</sup> en la prova de juny (de les quals 132 també el van aprovar en la prova de setembre) i només 13 persones no el van aprovar (de les quals 7 sí van aprovar-lo en la prova de setembre). Si s'observen els percentatges fila (% f.), s'obté que el 53,8% de l'alumnat que no va aprovar el Bloc en la prova de juny sí que el va aprovar en la prova de setembre i el 85,2% de l'alumnat que va aprovar el Bloc en la prova de juny també el va aprovar en la prova de setembre. Sent així, es pot fer l'estimació següent

$$\frac{P(\text{Bloc 1 } \checkmark \text{ setembre} | \text{Bloc 1 } \checkmark \text{ juny})}{P(\text{Bloc 1 } \checkmark \text{ setembre} | \text{Bloc 1 } \times \text{ juny})} = \frac{0,852}{0,538} \approx 1,58;$$

és a dir, la probabilitat d'aprovar el Bloc 1 de la prova de setembre és 1,58 vegades superior si es va aprovar també el de la prova de juny en relació a si no es va aprovar. Per tant, diríem que aprovar el Bloc 1 de la prova de juny és un “factor de protecció”<sup>10</sup> per a aprovar el mateix Bloc en la prova de setembre.

Es pot obtenir la Taula 15, procedint de manera anàloga en la resta de Blocs, en la qual, a més d'estimar l'RR puntualment, es mostra el seu interval de confiança al 95%. S'hi pot observar que:

- En el Bloc 1, tot i que l'estimació puntual de l'RR és 1,58, l'interval de confiança al 95% conté la unitat i, per tant, es pot acceptar  $RR = 1$ , cosa la qual s'interpretaria com que la probabilitat d'aprovar el Bloc 1 en la prova de setembre és independent d'haver-lo aprovat o no en la prova de juny.

<sup>9</sup>Es va considerar aprovat un Bloc si la qualificació no era inferior a 5 sobre 10.

<sup>10</sup>El concepte d'RR es defineix en l'àmbit de l'epidemiologia on la variable resposta sol ser una malaltia, una infecció, una mort o, en general, un esdeveniment advers. És per això que es parla de risc. En el context d'aquest estudi, s'ha d'entendre el concepte de risc com quelcom positiu, ja que es tracta de risc d'aprovar; per això, resulta més adient parlar de factor de “protecció”.

Bloc 1		Setembre		
+ - : amb naturals		X	✓	Total
	N	6	7	13
Juny	X	% f. 46,2	53,8	100
		% c. 20,7	5,0	7,7
	N	23	132	155
Juny	✓	% f. 14,8	85,2	100
		% c. 79,3	95,0	92,3
	N	29	139	168
Total	% f.	17,3	82,7	100
		% c. 100	100	100
	% c.	100	100	100

Bloc 2		Setembre		
Jerarquia d'operacions		X	✓	Total
	N	48	16	64
Juny	X	% f. 75,0	25,0	100
		% c. 64,9	17,0	38,1
	N	26	78	104
Juny	✓	% f. 25,0	75,0	100
		% c. 35,1	83,0	61,9
	N	74	94	168
Total	% f.	44,0	56,0	100
		% c. 100	100	100
	% c.	100	100	100

Bloc 3		Setembre		
Càlcul amb decimals		X	✓	Total
	N	58	11	69
Juny	X	% f. 84,1	15,9	100
		% c. 65,2	13,9	41,1
	N	31	68	99
Juny	✓	% f. 31,3	68,7	100
		% c. 34,8	86,1	58,9
	N	89	79	168
Total	% f.	53,0	47,0	100
		% c. 100	100	100
	% c.	100	100	100

Bloc 4		Setembre		
Fraccions i canvis d'unitats		X	✓	Total
	N	45	12	57
Juny	X	% f. 78,9	21,1	100
		% c. 77,6	10,9	33,9
	N	13	98	111
Juny	✓	% f. 11,7	88,3	100
		% c. 22,4	89,1	66,1
	N	58	110	168
Total	% f.	34,5	65,5	100
		% c. 100	100	100
	% c.	100	100	100

Bloc 5		Setembre		
Geometria		X	✓	Total
	N	43	13	56
Juny	X	% f. 76,8	23,2	100
		% c. 62,3	13,1	33,3
	N	26	86	112
Juny	✓	% f. 23,2	76,8	100
		% c. 37,7	86,9	66,7
	N	69	99	168
Total	% f.	41,1	58,9	100
		% c. 100	100	100
	% c.	100	100	100

Bloc 6		Setembre		
Problemes		X	✓	Total
	N	141	5	146
Juny	X	% f. 96,6	3,4	100
		% c. 91,6	35,7	86,9
	N	13	9	22
Juny	✓	% f. 59,1	40,9	100
		% c. 8,4	64,3	13,1
	N	154	14	168
Total	% f.	91,7	8,3	100
		% c. 100	100	100
	% c.	100	100	100

**Taula 14:** Taules de contingència, per a cadascun dels sis Blocs, classificant l'alumnat segons va aprovar o no (✓: qualificació igual o superior a 5 sobre 10; X: qualificació inferior a 5). Es mostren les freqüències absolutes i els perfils percentuals fila i columna.

- En els Blocs 2, 3, 4 i 5, l'RR s'estima entre 3 i 4, de manera que es pot dir que la probabilitat d'aprovar un d'aquests Blocs en la prova de setembre és entre 3 i 4 vegades més gran si es va aprovar en la prova de juny en relació a si no es va aprovar.
- En el Bloc 6, l'RR s'estima en 12, de manera que es pot dir que la probabilitat d'aprovar aquest Bloc en la prova de setembre és de l'ordre de 12 vegades més gran si es va aprovar en la prova de juny en relació a si no es va aprovar.

## 8 Modelització de la diferència entre el resultat en la prova de setembre i el resultat de la prova de juny

En aquesta secció es presenten els resultats de la modelització de la diferència entre el resultat de la prova de juny i de setembre.



Bloc	Estimació de RR = $\frac{P(\checkmark \text{ setembre}   \checkmark \text{ juny})}{P(\checkmark \text{ setembre}   \times \text{ juny})}$		
	Puntual	Inf. 95%	Sup. 95%
1	1,58	0,95	2,63
2	3,00	1,93	4,65
3	4,31	2,47	7,53
4	4,19	2,53	6,96
5	3,31	2,03	5,38
6	11,95	4,41	32,38

**Taula 15:** Estimació del risc relatiu (RR) corresponent a la probabilitat d'aprovar el Bloc en setembre si el mateix Bloc es va aprovar al juny, respecte la mateixa probabilitat si el Bloc no es va aprovar al juny.

Aquesta modelització es va realitzar considerant la variable resposta contínua resultant de restar la qualificació en la prova de juny a la qualificació en la prova de setembre (secció 8.1), i també considerant la variable resposta qualitativa consistent a observar si la qualificació en la prova de setembre va ser inferior que la corresponent a la de juny (secció 8.2).

## 8.1 Resposta quantitativa. Diferència numèrica entre juny i setembre

Es va considerar, per a cada individu i per a cada Bloc, el resultat d'haver restat la qualificació de la prova de juny a la qualificació de la prova de setembre i aleshores es va modelitzar aquesta variable,  $D$ .

El model ajustat va ser un model lineal mixt, exactament igual que en la modelització dels resultats en la prova de juny (veure secció 6).

### 8.1.1 Model ajustat

Una vegada realitzat l'ajustament del model, es va arribar al model final

$$D = \beta_1 \text{Bloc} + \beta_2 (J - 5) + \beta_3 \text{NEE} + \beta_4 \text{CEIP} + Iv + e$$

i les prediccions

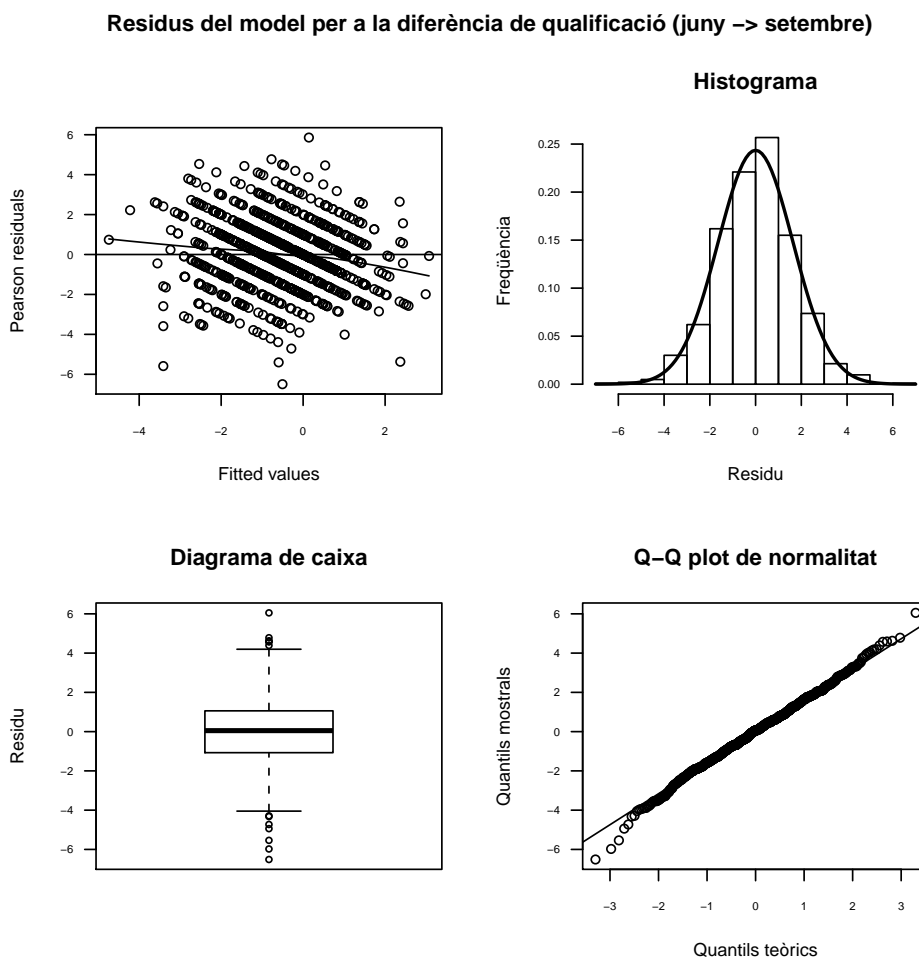
$$\hat{D} = \hat{\beta}_1 \text{Bloc} + \hat{\beta}_2 (J - 5) + \hat{\beta}_3 \text{NEE} + \hat{\beta}_4 \text{CEIP}, \quad (3)$$

de manera que la diferència entre la qualificació de la prova de juny i la de la prova de setembre ( $D = S - J$ ) es podia explicar segons el Bloc, la qualificació basal (juny), la tipificació NEE i el centre de primària. Les estimacions dels coeficients del model es mostren en la Taula 16 (pàgina 36).

La Figura 16 (pàgina 36) mostra un comportament acceptable dels residus ( $e$ ) del model i, per tant, es va donar el model per validat.

	Estimació ( $\beta$ )	Error est.	Wald	p-valor
B1: + - · : amb naturals	1,36	0,37	13,50	0,0002
B2: Jerarquia d'operacions	0,49	0,36	1,87	0,1720
B3: Càlcul amb decimals	-0,07	0,35	0,04	0,8344
B4: Fraccions i canvis d'unitats	0,46	0,35	1,71	0,1908
B5: Geometria	0,54	0,35	2,43	0,1191
B6: Problemes	-1,07	0,35	9,35	0,0022
Qual. juny - 5	-0,51	0,03	306,27	0,0000
NEE	-1,33	0,24	30,01	0,0000
CEIP 2	0,04	0,35	0,02	0,8999
CEIP 3	-0,64	0,36	3,05	0,0806
CEIP 4	-1,32	0,40	10,81	0,0010
CEIP 5	-0,96	0,37	6,84	0,0089
CEIP 6	-0,53	0,41	1,70	0,1916

**Taula 16:** Estimació dels coeficients en el model per a la diferència resultant de restar la qualificació en la prova de juny a la qualificació corresponent al mateix Bloc en la prova de setembre.



**Figura 16:** Comportament dels residus del model per a la diferència de resultats entre la prova de juny i la de setembre.

### 8.1.2 Interpretació del model ajustat i dels seus paràmetres

Per a interpretar el model ajustat, cal incorporar al model (3) les estimacions dels seus paràmetres mostrats en la Taula 16.

El perfil de referència d'un individu s'obté fixant una qualificació de 5 sobre 10 en el Bloc considerat en un individu no tipificat NEE i pertanyent al CEIP 1. Per a l'alumnat amb aquest perfil, el model prediu una qualificació en cada Bloc igual a l'estimació corresponent ( $\hat{\beta}$ ). Per exemple, en el Bloc 1, la predicció per a aquest perfil d'alumnat és 1,42 (increment respecte la prova de juny), i, en el Bloc 6, és -1,04 (decrement respecte la prova de juny).

El coeficient corresponent a la variable Nota de juny - 5, que té el valor -0,51, s'ha d'interpretar com un pendent: per cada punt per sobre (sota) de 5 en la prova de juny, el model prediu una pèrdua (un guany) de 0,51 punts en el mateix Bloc.

A partir d'aquí, si varia el nivell en algun factor, s'hi ha de sumar l'estimació corresponent a l'efecte considerat de manera que si, per exemple, l'individu és tipificat NEE, s'hi ha de restar 1,34 o si l'individu pertanyia al CEIP 5 en lloc del CEIP 1, s'hi ha de restar 0,99.

Els  $p$ -valors indiquen que, fixant els nivells de referència, només es prediuen diferències significatives entre juny i setembre en els Blocs 1 i 6.

Així, el model ajustat prediu que:

- El Bloc 1 (+ - · : amb naturals) presenta, de juny a setembre, una millora en la qualificació de 1,36 punts (fixant els nivells de referència en la resta de factors).
- El Bloc 6 (Problemes) presenta, de juny a setembre, un empitjorament en la qualificació d'aproximadament un punt (fixant els nivells de referència en la resta de factors).
- Els Blocs 2 (Jerarquia d'operacions), 4 (Fraccions i canvis d'unitats) i 5 (Geometria) presenten, de juny a setembre, una millora en la qualificació d'aproximadament mig punt (fixant els nivells de referència en la resta de factors) tot i que no és significativa.
- El Bloc 3 (Càlcul amb decimals) no presenta canvis en la qualificació de juny a setembre.
- Un individu tipificat NEE perd 1,33 punts en qualsevol Bloc a diferència de si no ho és.
- En funció del centre de primària, la diferència entre la qualificació de juny i la de setembre pot arribar a empitjorar respecte el CEIP 1 fins a 1,32 punts en qualsevol Bloc (CEIP 4).

### 8.1.3 Prediccions del model ajustat

En la Taula 17 (pàgina 39) es mostren les prediccions –puntuals i també per interval de confiança al 95%– que el model ajustat fa en la diferència de qualificacions entre juny i setembre, per a tots els tractaments possibles obtinguts en la combinació de les variables “NEE” i “Cprim”.

La Taula 18 (pàgina 40) mostra, per a cadascun dels 6 Blocs, els 3 pitjors i els 3 millors tractaments on es pot observar que la tipificació NEE representa un factor de risc. Respecte els centres de primària, els pitjors resultats s'associen als CEIPs 1 i 2, i els millors, als CEIPs 4 i 5.

## 8.2 Resposta qualitativa. Pèrdua de qualificació entre juny i setembre

En aquest cas es va mesurar, per a cada individu i per a cada Bloc, si va haver pèrdua de qualificació entre juny i setembre; és a dir, si la qualificació en la prova de setembre va ser inferior a la de juny. Després, es va procedir a modelitzar aquesta nova variable dicotòmica, mitjançant un model lineal mixt similar al model 1 (secció 6), però sota l'enllaç Logit:

$$\text{Logit}(P) = \ln\left(\frac{P}{1-P}\right) = X\beta + Iv + e, \quad (4)$$

on  $P \equiv$  Probabilitat de pèrdua  $= P(\text{Qual}(\text{set}) < \text{Qual}(\text{jun})) = P(S < J)$ ,  $X$  és el vector format per tots els factors i covariables d'efectes fixos (incloent les qualificacions en la prova de juny);  $\beta$  és el vector de coeficients corresponent als efectes fixos;  $I$  és la variable identificadora de l'individu,  $v$  és el coeficient corresponent a l'efecte aleatori introduït per l'individu, i  $e$  és el terme d'error del model.

Tant el coeficient corresponent a l'efecte aleatori de l'individu,  $v$ , com el terme d'error,  $e$ , se suposen amb distribució normal, esperança zero, variància constant i independència entre sí.

Les prediccions segons el model (4) són

$$\ln\left(\frac{\hat{P}}{1-\hat{P}}\right) = X\hat{\beta}$$

o, equivalentment

$$\hat{P} = \frac{1}{1 + e^{-X\hat{\beta}}}.$$

### 8.2.1 Model ajustat

El millor model ajustat va ser el següent:

$$\ln\left(\frac{P}{1-P}\right) = \hat{\beta}_1 \text{Bloc} + \hat{\beta}_2 (J - 5) + \hat{\beta}_3 \text{NEE} + \hat{\beta}_4 \text{CEIP} Iv + e$$

Bloc	NEE	Cprim	Predicció	L.I. (95%)	L.S. (95%)
<b>+ - · :</b> <b>amb naturals</b>	No	CEIP 1	1,4	0,9	1,8
		CEIP 2	1,4	1,0	1,8
		CEIP 3	0,7	0,4	1,0
		CEIP 4	0,0	-0,3	0,4
		CEIP 5	0,4	0,0	0,8
		CEIP 6	0,8	0,4	1,2
	Sí	CEIP 1	0,0	-0,5	0,6
		CEIP 2	0,1	-0,3	0,5
		CEIP 3	-0,6	-1,0	-0,2
		CEIP 4	-1,3	-1,7	-0,9
		CEIP 5	-0,9	-1,3	-0,5
		CEIP 6	-0,5	-1,0	0,0
<b>Jerarquia</b> <b>d'operacions</b>	No	CEIP 1	0,5	0,0	1,0
		CEIP 2	0,5	0,2	0,9
		CEIP 3	-0,1	-0,5	0,2
		CEIP 4	-0,8	-1,2	-0,5
		CEIP 5	-0,5	-0,8	-0,1
		CEIP 6	0,0	-0,4	0,4
	Sí	CEIP 1	-0,8	-1,4	-0,3
		CEIP 2	-0,8	-1,2	-0,4
		CEIP 3	-1,5	-1,9	-1,0
		CEIP 4	-2,2	-2,6	-1,7
		CEIP 5	-1,8	-2,2	-1,4
		CEIP 6	-1,4	-1,8	-0,9
<b>Càlculs amb</b> <b>decimals</b>	No	CEIP 1	-0,1	-0,5	0,4
		CEIP 2	0,0	-0,3	0,3
		CEIP 3	-0,7	-1,0	-0,4
		CEIP 4	-1,4	-1,7	-1,1
		CEIP 5	-1,0	-1,4	-0,7
		CEIP 6	-0,6	-1,0	-0,2
	Sí	CEIP 1	-1,4	-1,9	-0,9
		CEIP 2	-1,4	-1,8	-0,9
		CEIP 3	-2,0	-2,5	-1,6
		CEIP 4	-2,7	-3,2	-2,3
		CEIP 5	-2,4	-2,8	-1,9
		CEIP 6	-1,9	-2,4	-1,5
<b>Fracions i</b> <b>canvis d'unitats</b>	No	CEIP 1	0,5	0,0	0,9
		CEIP 2	0,5	0,2	0,8
		CEIP 3	-0,2	-0,5	0,1
		CEIP 4	-0,9	-1,2	-0,5
		CEIP 5	-0,5	-0,8	-0,2
		CEIP 6	-0,1	-0,5	0,3
	Sí	CEIP 1	-0,9	-1,4	-0,3
		CEIP 2	-0,8	-1,2	-0,4
		CEIP 3	-1,5	-1,9	-1,1
		CEIP 4	-2,2	-2,6	-1,7
		CEIP 5	-1,8	-2,2	-1,4
		CEIP 6	-1,4	-1,9	-0,9
<b>Geometria</b>	No	CEIP 1	0,5	0,1	1,0
		CEIP 2	0,6	0,3	0,9
		CEIP 3	-0,1	-0,4	0,2
		CEIP 4	-0,8	-1,1	-0,4
		CEIP 5	-0,4	-0,7	-0,1
		CEIP 6	0,0	-0,4	0,4
	Sí	CEIP 1	-0,8	-1,3	-0,2
		CEIP 2	-0,7	-1,2	-0,3
		CEIP 3	-1,4	-1,9	-1,0
		CEIP 4	-2,1	-2,5	-1,7
		CEIP 5	-1,7	-2,2	-1,3
		CEIP 6	-1,3	-1,8	-0,8
<b>Problemes</b>	No	CEIP 1	-1,1	-1,6	-0,6
		CEIP 2	-1,0	-1,3	-0,7
		CEIP 3	-1,7	-2,1	-1,3
		CEIP 4	-2,4	-2,8	-2,0
		CEIP 5	-2,0	-2,4	-1,7
		CEIP 6	-1,6	-2,0	-1,2
	Sí	CEIP 1	-2,4	-3,0	-1,8
		CEIP 2	-2,4	-2,8	-1,9
		CEIP 3	-3,0	-3,5	-2,5
		CEIP 4	-3,7	-4,2	-3,2
		CEIP 5	-3,4	-3,8	-2,9
		CEIP 6	-2,9	-3,4	-2,4

**Taula 17:** Prediccions (puntuals i per interval de confiança al 95%) per a l'alumnat que obté una qualificació de 5 sobre 10 en cadascun dels 6 Blocs en la prova de juny, segons el model ajustat per a la diferència de qualificació entre juny i setembre.

Bloc	NEE	Cprim	Predicció	L.I. (95%)	L.S. (95%)
<b>+ - · :</b> <b>amb naturals</b>	Sí	CEIP 4	-1.3	-1.7	-0.9
	Sí	CEIP 5	-0.9	-1.3	-0.5
	Sí	CEIP 3	-0.6	-1.0	-0.2
	No	CEIP 6	0.8	0.4	1.2
	No	CEIP 1	1.4	0.9	1.8
	No	CEIP 2	1.4	1.0	1.8
<b>Jerarquia</b> <b>d'operacions</b>	Sí	CEIP 4	-2.2	-2.6	-1.7
	Sí	CEIP 5	-1.8	-2.2	-1.4
	Sí	CEIP 3	-1.5	-1.9	-1.0
	No	CEIP 6	0.0	-0.4	0.4
	No	CEIP 1	0.5	0.0	1.0
	No	CEIP 2	0.5	0.2	0.9
<b>Càlcul amb</b> <b>decimals</b>	Sí	CEIP 4	-2.7	-3.2	-2.3
	Sí	CEIP 5	-2.4	-2.8	-1.9
	Sí	CEIP 3	-2.0	-2.5	-1.6
	No	CEIP 6	-0.6	-1.0	-0.2
	No	CEIP 1	-0.1	-0.5	0.4
	No	CEIP 2	0.0	-0.3	0.3
<b>Fraccions i</b> <b>canvis d'unitats</b>	Sí	CEIP 4	-2.2	-2.6	-1.7
	Sí	CEIP 5	-1.8	-2.2	-1.4
	Sí	CEIP 3	-1.5	-1.9	-1.1
	No	CEIP 6	-0.1	-0.5	0.3
	No	CEIP 1	0.5	0.0	0.9
	No	CEIP 2	0.5	0.2	0.8
<b>Geometria</b>	Sí	CEIP 4	-2.1	-2.5	-1.7
	Sí	CEIP 5	-1.7	-2.2	-1.3
	Sí	CEIP 3	-1.4	-1.9	-1.0
	No	CEIP 6	0.0	-0.4	0.4
	No	CEIP 1	0.5	0.1	1.0
	No	CEIP 2	0.6	0.3	0.9
<b>Problemes</b>	Sí	CEIP 4	-3.7	-4.2	-3.2
	Sí	CEIP 5	-3.4	-3.8	-2.9
	Sí	CEIP 3	-3.0	-3.5	-2.5
	No	CEIP 6	-1.6	-2.0	-1.2
	No	CEIP 1	-1.1	-1.6	-0.6
	No	CEIP 2	-1.0	-1.3	-0.7

**Taula 18:** Prediccions (puntuals i per interval de confiança al 95%) per als 3 pitjors perfils i per als 3 millors, segons el model ajustat per a la diferència de qualificació entre juny i setembre, amb qualificació de 5 sobre 10 en la prova de juny.

	Estimació ( $\hat{\beta}$ )	Error est.	Wald	p-valor
B1: + - : amb naturals	-1,85	0,46	16,49	0,0000
B2: Jerarquia d'operacions	-1,05	0,42	6,15	0,0131
B3: Càlcul amb decimals	-0,49	0,43	1,30	0,2550
B4: Fraccions i canvis d'unitats	-1,16	0,44	7,00	0,0082
B5: Geometria	-1,29	0,45	8,07	0,0045
B6: Problemes	-0,30	0,43	0,49	0,4860
Nota juny-5	0,47	0,04	123,85	0,0000
NEE	1,12	0,26	18,31	0,0000
CEIP 2	0,05	0,43	0,02	0,9004
CEIP 3	0,81	0,44	3,38	0,0659
CEIP 4	1,20	0,46	6,68	0,0098
CEIP 5	1,20	0,44	7,34	0,0068
CEIP 6	0,45	0,49	0,82	0,3640

**Taula 19:** Estimació dels coeficients en el model per a la probabilitat d'obtenir una qualificació en la prova de setembre inferior a la de juny.

i les prediccions

$$\hat{P} = \frac{1}{1 + e^{-(\hat{\beta}_1 \text{Bloc} + \hat{\beta}_2 (J-5) + \hat{\beta}_3 \text{NEE} + \hat{\beta}_4 \text{CEIP})}} \quad (5)$$

de manera que la probabilitat que la qualificació en la prova de setembre sigui inferior a la de la prova de juny depenia del Bloc en consideració, de la qualificació basal en el Bloc (prova de juny), del fet que l'individu fos tipificat NEE, i del centre de primària. Les estimacions dels coeficients del model es mostren en la Taula 19 (pàgina 41).

Es va crear la matriu de confusió, mostrada en la Taula 20 (pàgina 42), per tal de validar el model. Aquesta matriu és una taula de contingència que classifica tots els individus simultàniament segons si va haver pèrdua de qualificació entre juny i setembre i segons la predicció del model. Tres mesures possibles de bondat d'ajustament de les dades al model, relacionades amb la matriu de confusió són:

- Sensibilitat: és la probabilitat de detectar correctament un individu que perd qualificació. La seva estimació és  $\frac{313}{174+313} \approx 0,64 = 64\%$ .
- Especificitat: és la probabilitat de detectar correctament un individu que no perd qualificació. La seva estimació és  $\frac{392}{392+153} \approx 0,72 = 72\%$ .
- Classificació correcta: és la probabilitat de fer una predicció correcta sobre un individu. La seva estimació és  $\frac{392+313}{392+153+174+313} \approx 0,68 = 68\%$ .

Aquests valors, no massa elevats, fan acceptar el model amb reserves sobre la seva capacitat predictiva.

## 8.2.2 Interpretació del model ajustat i dels seus paràmetres

La interpretació del model ajustat es fa incorporant al model (5) les estimacions dels seus paràmetres que es mostren en la Taula 19.

L'enllaç Logit fa que la interpretació d'aquest tipus de model no resulti trivial. Per començar, es pot fer una interpretació qualitativa. Es pot veure fàcilment que l'expressió

$$\hat{P} = \frac{1}{1 + e^{-X\hat{\beta}}} \quad (6)$$

Observat	Predicció	
	No	Sí
No	392	153
Sí	174	313

**Taula 20:** Matriu de confusió corresponent al model per a la probabilitat d'obtenir en la prova de setembre una qualificació inferior a la de juny. Aquesta matriu classifica l'alumnat en funció de si va perdre qualificació i de la predicció del model.

implica que  $\hat{p}$  és monòtona amb  $\hat{\beta}$ , de manera que valors elevats del paràmetre  $\beta_i$  impliquen valors elevats de  $p$ . Per tant, si la variable explicativa  $X_i$  té associat un valor positiu de  $\beta_i$ , llavors  $X_i$  resulta un factor de risc en la pèrdua de qualificació entre juny i setembre. Contràriament, si la variable explicativa  $X_i$  té associat un valor negatiu de  $\beta_i$ , llavors  $X_i$  resulta un factor de protecció en aquesta pèrdua de qualificació. Un valor nul de  $\beta_i$  implicaria que  $p$  és independent de  $X_i$ .

Per a una interpretació quantitativa, resulta convenient introduir els conceptes d'*Odds*<sup>11</sup> i d'*Odds Ratio* (OR).

L'*Odds* d'un esdeveniment  $A$  es defineix com

$$\text{Odds}(A) = \frac{P(A)}{P(\bar{A})} = \frac{P(A)}{1 - P(A)},$$

és a dir, com el quocient entre la probabilitat de l'esdeveniment  $A$  i la del seu complementari,  $\bar{A}$ .

Per exemple, si la probabilitat que demà faci sol és 0,80, llavors

$$\text{Odds}(\text{Demà farà sol}) = \frac{P(\text{Demà farà sol})}{1 - P(\text{Demà farà sol})} = \frac{0,80}{1 - 0,80} = \frac{0,80}{0,20} = 4,$$

i es pot dir que demà és 4 vegades més probable que faci sol que que no en faci, o bé, que l'avantatge de probabilitats que demà faci sol és de 4 contra 1.

D'altra banda, l'*Odds Ratio* entre dos esdeveniments  $A$  i  $B$  es defineix com

$$\text{OR}(A, B) = \frac{\text{Odds}(A)}{\text{Odds}(B)} = \frac{\frac{P(A)}{1-P(A)}}{\frac{P(B)}{1-P(B)}}.$$

Tenint en compte aquests dos conceptes, l'expressió 4 permet expressar les prediccions del model com

$$\ln(\widehat{\text{Odds}}(P)) = X\hat{\beta};$$

de manera que

$$\left. \begin{aligned} \ln(\widehat{\text{Odds}}(P); X_i = 1, X_j = x_j, j \neq i) &= \hat{\beta}_1 x_1 + \dots + \hat{\beta}_i \cdot 1 + \dots + \hat{\beta}_n x_n \\ \ln(\widehat{\text{Odds}}(P); X_i = 0, X_j = x_j, j \neq i) &= \hat{\beta}_1 x_1 + \dots + \hat{\beta}_i \cdot 0 + \dots + \hat{\beta}_n x_n \end{aligned} \right\}$$

<sup>11</sup>Es pot traduir *Odds* com *avantatge*.



i restant les dues expressions s'obté

$$\hat{\beta}_i = \frac{\ln(\widehat{\text{Odds}}(P); X_i = 1, X_j = x_j, j \neq i)}{\ln(\widehat{\text{Odds}}(P); X_i = 0, X_j = x_j, j \neq i)} = \ln(\widehat{\text{OR}}(X_i))$$

o, equivalentment,

$$\widehat{\text{OR}}(X_i) = e^{\hat{\beta}_i}.$$

Així docs, l'exponencial del paràmetre  $\beta_i$  s'interpreta com el quocient d'avantatges que s'obté si el factor  $X_i$  és present en relació a si és absent, mantenint constants els valors de la resta de variables explicatives.

Per exemple, en la Taula 19 (pàgina 41), s'observa que  $\beta_{\text{NEE}} = 1,12$ . Així,

$$\widehat{\text{OR}}(\text{NEE}) = e^{\hat{\beta}_{\text{NEE}}} = e^{1,12} \approx 3,1,$$

és a dir,

$$\frac{\frac{P(\text{pèdre qualificació, sent NEE})}{P(\text{no pèdre qualificació, sent NEE})}}{\frac{P(\text{pèdre qualificació, no sent NEE})}{P(\text{no pèdre qualificació, no sent NEE})}} \approx 3,1.$$

D'aquesta manera, l'avantatge –en aquest context, desavantatge– de perdre qualificació entre juny i setembre en relació a no perdre-la augmenta al triple pel fet de ser alumnat tipificat NEE.

Atès que la interpretació en termes d'OR pot ser complicada, resulta més convenient analitzar les prediccions en termes de probabilitat. En aquest sentit, considerant novament l'expressió 6, s'obté que, fixant el factor  $X_i = 1$  i la resta de factors i covariables en els seus valors de referència (de manera que  $X_j = 0$  per a tot  $j \neq i$ ), resulta

$$\hat{P}(\text{pèrdua de qualificació} | X_i = 1, X_j = 0, j \neq i) = \frac{1}{1 + e^{-\hat{\beta}_i}}.$$

Així, per exemple, segons la Taula 19, la predicció que fa el model per a la probabilitat de pèrdua de qualificació entre juny i setembre en el Bloc 1, per a un individu amb qualificació de 5 en aquest Bloc, no tipificat NEE, i pertanyent al CEIP 1 (valors de referència), és  $\frac{1}{1 + e^{-\hat{\beta}_{\text{Bloc 1}}}} = \frac{1}{1 + e^{1,85}} \approx 0,136 = 13,6\%$ .

Un altre exemple: predicció per a un individu en el Bloc 6 (Problemes) amb qualificació de 6 sobre 10 en juny en aquest Bloc, no tipificat NEE, i pertanyent al CEIP 6:

$$\frac{1}{1 + e^{-(\hat{\beta}_{\text{Bloc 6}} + (J-5) \cdot \hat{\beta}_{J-5} + \hat{\beta}_{\text{CEIP 6}})}} = \frac{1}{1 + e^{0,30 - 1 \cdot 0,47 - 0,45}} \approx 0,650 = 65,0\%.$$

De manera anàloga, es pot calcular la predicció per a qualsevol altre tractament (combinació de valors de les variables explicatives).

### 8.2.3 Prediccions del model ajustat

La Taula 21 (pàgina 45) mostra les prediccions –puntuals i també per interval de confiança al 95%– que el model ajustat fa per a la probabilitat de pèrdua de qualificació entre juny i setembre, per a tots els tractaments possibles obtinguts en la combinació de les variables “NEE” i “Cprim”.

La Taula 22 (pàgina 46) mostra, per a cadascun dels 6 Blocs, els 3 pitjors i els 3 millors tractaments on es pot observar (per a una qualificació de 5 sobre 10 en la prova de juny) que la tipificació NEE representa un factor de risc, en coherència amb els resultats de la secció 8.1. També s’observa que els pitjors resultats s’associen als CEIPs 1 i 2, i els millors, als CEIPs 4 i 5 (com també en la secció 8.1).

## 9 Modelització dels resultats en la primera avaluació de 1r d’ESO (desembre)

Finalment, es va modelitzar la qualificació en l’àrea de Matemàtiques corresponent a la primera avaluació del primer curs d’ESO.

Es va modelitzar tant la qualificació numèrica –de 0 a 10– (secció 9.1) com la qualificació qualitativa –aprovat/no aprovat– considerant l’aprovat amb una qualificació mínima de 5 sobre 10 (secció 9.2).

### 9.1 Resposta quantitativa. Qualificació numèrica

En aquest cas, cada individu va obtenir una única qualificació en l’àrea de Matemàtiques al final de la primera avaluació de 1r d’ESO i, per tant, va ser suficient ajustar un Model Lineal General similar al model (1) (secció 6) però sense l’efecte aleatori que introduïa l’individu ja que no hi havia mesures repetides (com era el cas de la prova escrita de juny i setembre, degut als diferents Blocs). Així, es tracta del model

$$A_1 = X\beta + e, \quad (7)$$

on  $A_1$  és la qualificació (de 0 a 10),  $X$  és el vector format per tots els factors i covariables d’efectes fixos (incloent les qualificacions en la prova de juny i en la de setembre),  $\beta$  és el vector de coeficients corresponent als efectes fixos i  $e$  és el terme d’error del model, que se suposa amb distribució normal, esperança zero i variància constant.

Les prediccions segons el model (7) són

$$\widehat{A}_1 = X\hat{\beta}.$$

#### 9.1.1 Model ajustat

En aquesta ocasió, el millor model ajustat va ser el següent:

$$A_1 = \beta_0 + \beta_1\text{NEE} + \beta_2\text{Bloc 2 (set)} + \beta_3\text{Bloc 3 (set)} + \beta_4\text{Bloc 6 (set)} + e$$

Bloc	NEE	Cprim	Predicció	L.I. (95%)	L.S. (95%)
<b>+ - · ÷ : amb naturals</b>	No	CEIP 1	13,6	7,3	24,0
		CEIP 2	14,3	9,2	21,4
		CEIP 3	26,1	18,5	35,3
		CEIP 4	34,4	24,6	45,7
		CEIP 5	34,3	24,4	45,8
		CEIP 6	19,8	12,2	30,4
	Sí	CEIP 1	32,5	18,4	50,6
		CEIP 2	33,7	22,5	47,0
		CEIP 3	51,8	38,3	65,1
		CEIP 4	61,5	47,6	73,8
		CEIP 5	61,5	48,1	73,3
		CEIP 6	43,0	29,0	58,2
<b>Jerarquia d'operacions</b>	No	CEIP 1	25,8	15,2	40,3
		CEIP 2	26,9	19,4	36,0
		CEIP 3	43,8	34,0	54,1
		CEIP 4	53,7	42,6	64,4
		CEIP 5	53,6	42,8	64,1
		CEIP 6	35,3	24,2	48,3
	Sí	CEIP 1	51,5	33,1	69,5
		CEIP 2	52,9	39,3	66,1
		CEIP 3	70,4	57,1	80,9
		CEIP 4	78,0	66,2	86,5
		CEIP 5	77,9	67,0	86,0
		CEIP 6	62,5	47,1	75,7
<b>Càlcul amb decimals</b>	No	CEIP 1	38,1	24,0	54,6
		CEIP 2	39,4	29,8	49,9
		CEIP 3	57,9	47,4	67,8
		CEIP 4	67,2	56,3	76,5
		CEIP 5	67,1	56,6	76,2
		CEIP 6	49,1	35,9	62,4
	Sí	CEIP 1	65,3	46,4	80,3
		CEIP 2	66,5	53,1	77,6
		CEIP 3	80,8	69,9	88,4
		CEIP 4	86,2	77,2	92,0
		CEIP 5	86,2	77,9	91,7
		CEIP 6	74,6	60,9	84,8
<b>Fracions i canvis d'unitats</b>	No	CEIP 1	23,8	13,8	37,7
		CEIP 2	24,8	17,7	33,6
		CEIP 3	41,1	31,6	51,3
		CEIP 4	50,9	39,9	61,8
		CEIP 5	50,9	40,0	61,6
		CEIP 6	32,8	22,2	45,5
	Sí	CEIP 1	48,8	30,8	67,1
		CEIP 2	50,1	36,7	63,5
		CEIP 3	68,0	54,5	79,1
		CEIP 4	76,0	63,7	85,1
		CEIP 5	76,0	64,5	84,6
		CEIP 6	59,9	44,4	73,6
<b>Geometria</b>	No	CEIP 1	21,7	12,4	35,1
		CEIP 2	22,6	15,9	31,1
		CEIP 3	38,2	28,8	48,6
		CEIP 4	47,9	37,0	59,1
		CEIP 5	47,8	37,0	58,9
		CEIP 6	30,2	20,0	42,8
	Sí	CEIP 1	45,8	28,1	64,6
		CEIP 2	47,1	33,8	60,8
		CEIP 3	65,4	51,2	77,2
		CEIP 4	73,7	60,8	83,6
		CEIP 5	73,7	61,5	83,1
		CEIP 6	56,9	41,2	71,4
<b>Problemes</b>	No	CEIP 1	42,6	26,9	60,0
		CEIP 2	43,9	33,6	54,9
		CEIP 3	62,4	50,2	73,3
		CEIP 4	71,2	59,6	80,5
		CEIP 5	71,1	60,3	80,0
		CEIP 6	53,8	39,5	67,5
	Sí	CEIP 1	69,4	49,3	84,1
		CEIP 2	70,5	56,1	81,7
		CEIP 3	83,5	71,6	91,1
		CEIP 4	88,3	78,9	93,8
		CEIP 5	88,3	79,6	93,5
		CEIP 6	78,0	63,4	88,0

**Taula 21:** Prediccions (puntuals i per interval de confiança al 95%) per a l'alumnat amb qualificació de 5 sobre 10 en cadascun dels Blocs de la prova de juny, segons el model ajustat per a la probabilitat de pèrdua de qualificació entre juny i setembre.

Bloc	NEE	Cprim	Predicció	L.I. (95%)	L.S. (95%)
<b>+ - · :</b> <b>amb naturals</b>	No	CEIP 1	13,6	7,3	24,0
	No	CEIP 2	14,3	9,2	21,4
	No	CEIP 6	19,8	12,2	30,4
	Sí	CEIP 3	51,8	38,3	65,1
	Sí	CEIP 5	61,5	48,1	73,3
	Sí	CEIP 4	61,5	47,6	73,8
<b>Jerarquia</b> <b>d'operacions</b>	No	CEIP 1	25,8	15,2	40,3
	No	CEIP 2	26,9	19,4	36,0
	No	CEIP 6	35,3	24,2	48,3
	Sí	CEIP 3	70,4	57,1	80,9
	Sí	CEIP 5	77,9	67,0	86,0
	Sí	CEIP 4	78,0	66,2	86,5
<b>Càlcul amb</b> <b>decimals</b>	No	CEIP 1	38,1	24,0	54,6
	No	CEIP 2	39,4	29,8	49,9
	No	CEIP 6	49,1	35,9	62,4
	Sí	CEIP 3	80,8	69,9	88,4
	Sí	CEIP 5	86,2	77,9	91,7
	Sí	CEIP 4	86,2	77,2	92,0
<b>Fraccions i</b> <b>canvis d'unitats</b>	No	CEIP 1	23,8	13,8	37,7
	No	CEIP 2	24,8	17,7	33,6
	No	CEIP 6	32,8	22,2	45,5
	Sí	CEIP 3	68,0	54,5	79,1
	Sí	CEIP 5	76,0	64,5	84,6
	Sí	CEIP 4	76,0	63,7	85,1
<b>Geometria</b>	No	CEIP 1	21,7	12,4	35,1
	No	CEIP 2	22,6	15,9	31,1
	No	CEIP 6	30,2	20,0	42,8
	Sí	CEIP 3	65,4	51,2	77,2
	Sí	CEIP 5	73,7	61,5	83,1
	Sí	CEIP 4	73,7	60,8	83,6
<b>Problemes</b>	No	CEIP 1	42,6	26,9	60,0
	No	CEIP 2	43,9	33,6	54,9
	No	CEIP 6	53,8	39,5	67,5
	Sí	CEIP 3	83,5	71,6	91,1
	Sí	CEIP 5	88,3	79,6	93,5
	Sí	CEIP 4	88,3	78,9	93,8

**Taula 22:** Prediccions (puntuals i per interval de confiança al 95%) per als 3 pitjors perfils i per als 3 millors, segons el model ajustat per a la probabilitat de pèrdua de qualificació entre juny i setembre, amb qualificació de 5 sobre 10 en la prova de juny.

	Estimació ( $\hat{\beta}$ )	Error est.	Wald	$p$ -valor
Constant	2,64	0,26	10,15	0,0000
NEE	1,20	0,32	3,77	0,0002
B2 (Set.): Jerarquia d'operacions	0,16	0,05	2,99	0,0032
B3 (Set.): Càlcul amb decimals	0,27	0,05	5,35	0,0000
B6 (Set.): Problemes	0,37	0,05	6,85	0,0000

**Taula 23:** Estimació dels coeficients en el model per a la qualificació en la primera avaluació de 1r d'ESO (desembre).

i les prediccions

$$\widehat{A}_1 = \hat{\beta}_0 + \hat{\beta}_1 \text{NEE} + \hat{\beta}_2 \text{Bloc 2 (set)} + \hat{\beta}_3 \text{Bloc 3 (set)} + \hat{\beta}_4 \text{Bloc 6 (set)}, \quad (8)$$

de manera que la qualificació en la primera avaluació de 1r d'ESO es va poder explicar segons si l'individu era tipificat NEE i segons les qualificacions en els Blocs 2 (Jerarquia d'operacions), 3 (Càlcul amb decimals) i 6 (Problemes) corresponents a la prova de juny. Les estimacions dels coeficients del model es mostren en la Taula 23 (pàgina 47).

Es va poder validar el model ajustat atès el comportament correcte dels residus (Figura 17, pàgina 28).

### 9.1.2 Interpretació del model ajustat i dels seus paràmetres

La interpretació del model ajustat es fa considerant l'equació del model (9.2.1) i les estimacions dels seus paràmetres, que es mostren en la Taula 23.

El perfil de referència correspon a un individu no tipificat NEE i amb qualificació de 0 sobre 10 en els Blocs 2, 3 i 6; de manera que, per a l'alumnat amb aquest perfil, el model prediu una qualificació en la primera avaluació de 1r curs d'ESO de  $\hat{\beta}_0 = 2,64$  sobre 10. Si l'individu és tipificat NEE, s'hi han de sumar 1,20 ( $\hat{\beta}_{\text{NEE}}$ ) punts més. Per cada punt obtingut en el Bloc 2, en el Bloc 3 o en el Bloc 6, s'estima un increment en la qualificació de la primera avaluació de 0,16 ( $\hat{\beta}_{\text{B2}}$ ), 0,27 ( $\hat{\beta}_{\text{B3}}$ ) o 0,37 ( $\hat{\beta}_{\text{B6}}$ ) punt, respectivament.

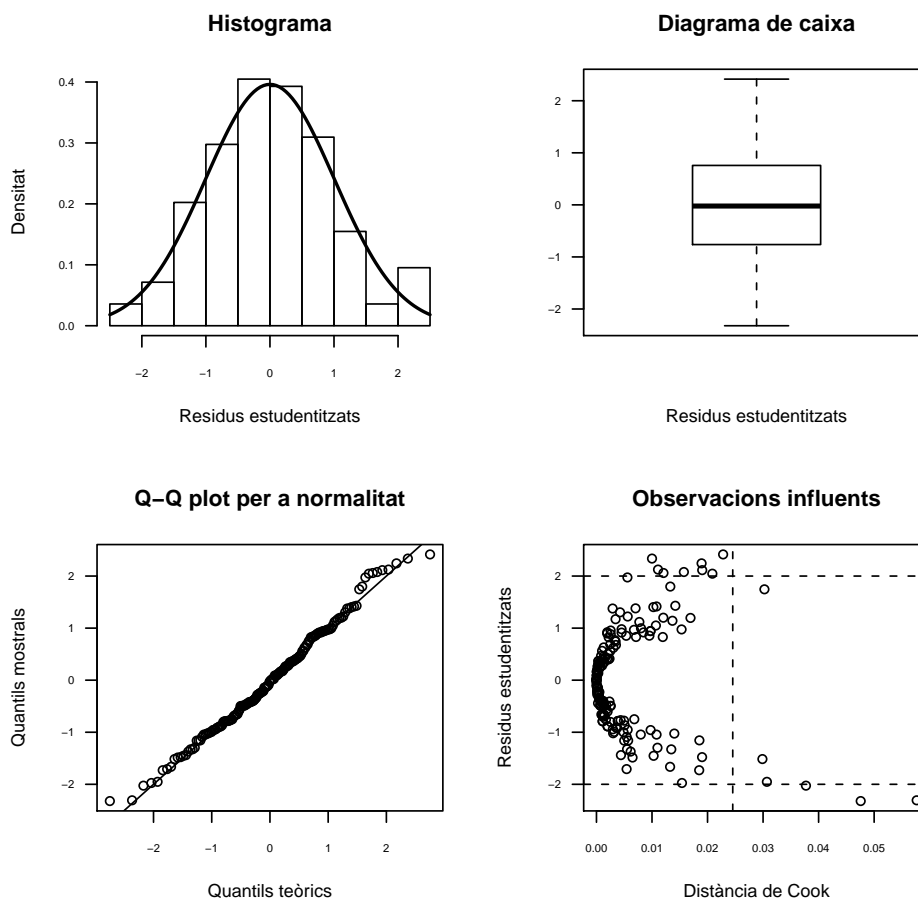
Per exemple, en el cas d'un individu que va obtenir una qualificació de 5 sobre 10 en cadascun dels Blocs 2, 3 i 6 en la prova de setembre, el model prediu una qualificació en la primera avaluació del 1r curs d'ESO de  $2,64 + 5 \cdot (0,16 + 0,27 + 0,37) = 6,7$  si no és tipificat NEE, i de  $6,64 + 1,20 = 7,9$  si sí ho és. Aquest exemple es completa amb els corresponents intervals de confiança en la Taula 24 (pàgina 48).

### 9.1.3 Prediccions del model ajustat

La Taula 25 (pàgina 49) mostra les prediccions del model per a individus amb qualificació entre 4 i 6 en els Blocs 2, 3 i 6 de la prova de setembre<sup>12</sup>. Es pot observar com una qualificació de 4 sobre 10 en els Blocs 2, 3 i 6 de la prova de

<sup>12</sup>Considerar totes les possibles combinacions de qualificacions entre 0 i 10 provocaria una taula de prediccions de 2662 files.

**Residus del model per a la 1a Avaluació**



**Figura 17:** Comportament dels residus del model per a la qualificació en la primera avaluació de 1r d'ESO (desembre).

NEE	B2 (set.)	B3 (set.)	B6 (set.)	Predicció	L.I. (95%)	L.S. (95%)
No	5	5	5	6,7	6,3	7,1
Sí	5	5	5	7,9	7,2	8,5

**Taula 24:** Prediccions (puntuals i per interval de confiança al 95%), per a l'alumnat que obté qualificació de 5 sobre 10 en els Blocs 2, 3 i 6 en setembre, segons el model ajustat per a qualificació en la primera avaluació de 1r d'ESO (desembre).

NEE	B2 (set.)	B3 (set.)	B6 (set.)	Predicció	L.I. (95%)	L.S. (95%)
	4	4	4	5,9	5,5	6,2
	5	4	4	6,0	5,7	6,4
	4	5	4	6,1	5,8	6,5
	6	4	4	6,2	5,9	6,5
	4	4	5	6,2	5,8	6,7
	5	5	4	6,3	6,0	6,6
	4	6	4	6,4	6,0	6,8
	5	4	5	6,4	6,0	6,8
	6	5	4	6,5	6,2	6,8
	4	5	5	6,5	6,1	6,9
	5	6	4	6,6	6,3	6,9
	6	4	5	6,6	6,1	7,0
	4	4	6	6,6	6,1	7,1
No	6	6	4	6,7	6,4	7,0
	5	5	5	6,7	6,3	7,1
	6	5	5	6,8	6,5	7,2
	4	6	5	6,8	6,3	7,2
	5	4	6	6,8	6,3	7,3
	5	6	5	6,9	6,5	7,3
	6	4	6	6,9	6,4	7,4
	4	5	6	6,9	6,4	7,4
	5	5	6	7,0	6,6	7,5
	6	6	5	7,1	6,7	7,5
	4	6	6	7,1	6,6	7,7
	6	5	6	7,2	6,7	7,7
	5	6	6	7,3	6,8	7,8
	6	6	6	7,5	7,0	7,9
	4	4	4	7,1	6,5	7,7
	5	4	4	7,2	6,6	7,9
	4	5	4	7,3	6,7	8,0
	6	4	4	7,4	6,7	8,0
	4	4	5	7,4	6,8	8,1
	5	5	4	7,5	6,9	8,1
	4	6	4	7,6	7,0	8,3
	5	4	5	7,6	6,9	8,3
	6	5	4	7,7	7,0	8,3
	4	5	5	7,7	7,0	8,4
	5	6	4	7,8	7,1	8,4
	6	4	5	7,8	7,1	8,5
	4	4	6	7,8	7,1	8,5
Sí	6	6	4	7,9	7,3	8,6
	5	5	5	7,9	7,2	8,5
	6	5	5	8,0	7,3	8,7
	4	6	5	8,0	7,3	8,7
	5	4	6	8,0	7,2	8,7
	5	6	5	8,1	7,5	8,8
	6	4	6	8,1	7,4	8,9
	4	5	6	8,1	7,3	8,8
	5	5	6	8,2	7,5	9,0
	6	6	5	8,3	7,6	9,0
	6	5	6	8,4	7,7	9,1
	4	6	6	8,4	7,6	9,1
	5	6	6	8,5	7,8	9,2
	6	6	6	8,7	7,9	9,4

**Taula 25:** Prediccions del model per a individus amb qualificació entre 4 i 6 en els Blocs 2, 3 i 6 de la prova de setembre, segons el model ajustat per a la qualificació en la primera avaluació de 1r d'ESO (desembre).

setembre provoca una predicció de 5,9 punts en alumnat no NEE, i de 7,1 en alumnat NEE.

## 9.2 Resposta qualitativa. Obtenció de l'aprovat

### 9.2.1 Model ajustat

En aquest cas es va ajustar un Model Lineal Generalitzat amb transformació Logit. El millor model ajustat va ser

$$\ln\left(\frac{P}{1-P}\right) = \text{Constant} + \hat{\beta}_{\text{NEE}} + \hat{\beta}_{\text{B1J}}\text{B1J} + \hat{\beta}_{\text{B2S}}\text{B2S} + \hat{\beta}_{\text{B3S}}\text{B3S} + \hat{\beta}_{\text{B6S}}\text{B6S} + e$$

	Estimació	Error est.	Wald	p-valor
Constant	-6,15	1,25	-4,93	0,0000
NEE	2,87	0,75	3,83	0,0001
B1 (Juny): + - · : amb naturals	0,30	0,13	2,24	0,0254
B2 (Set.): Jerarquia d'operacions	0,40	0,14	2,95	0,0031
B3 (Set.): Càlcul amb decimals	0,48	0,14	3,45	0,0006
B6 (Set.): Problemes	0,46	0,17	2,73	0,0064

**Taula 26:** Estimació dels coeficients en el model per a la probabilitat d'aprovar la primera avaluació de 1r d'ESO (desembre).

Observat	Predicció	
	No	Sí
No	48	13
Sí	11	96

**Taula 27:** Matriu de confusió corresponent al model per a la qualificació en la primera avaluació de 1r d'ESO (desembre). Aquesta matriu classifica l'alumnat simultàniament segons si va aprovar l'avaluació i segons la predicció del model.

i les prediccions

$$\hat{P} = \frac{1}{1 + e^{-(\text{Constant} + \hat{\beta}_{\text{NEE}} + \hat{\beta}_{\text{B1J}}\text{B1J} + \hat{\beta}_{\text{B2S}}\text{B2S} + \hat{\beta}_{\text{B3S}}\text{B3S} + \hat{\beta}_{\text{B6S}}\text{B6S})}}, \quad (9)$$

on  $P$  = Probabilitat d'aprovar la primera avaluació del 1r curs d'ESO i les estimacions dels coeficients del model es mostren en la Taula 26 (pàgina 50).

Així, segons el millor model ajustat, la probabilitat d'aprovar la primera avaluació del 1r curs d'ESO depèn de les mateixes covariables que el model quantitatiu (equació ) i de la qualificació en el Bloc 1 en la prova de juny.

Per a validar el model ajustat es va considerar que la matriu de confusió (Taula 27, pàgina 50) reflectia una molt bona qualitat d'ajustament:

- Sensibilitat: és la probabilitat de detectar correctament un individu que suspèn la primera avaluació (detecció de fracàs). La seva estimació és  $\frac{48}{48+13} \approx 0,79 = 79\%$ .
- Especificitat: és la probabilitat de detectar correctament un individu que aprova la primera avaluació. La seva estimació és  $\frac{96}{96+11} \approx 0,90 = 90\%$ .
- Classificació correcta: és la probabilitat de fer una predicció correcta sobre un individu. La seva estimació és  $\frac{48+96}{48+96+13+11} \approx 0,86 = 86\%$ .



## 9.2.2 Interpretació del model ajustat i dels seus paràmetres

La interpretació del model ajustat s'ha de fer en termes d'*Odds Ratio* (OR), segons s'exposa en la secció 8.2.2 on es mostra que l'exponencial del paràmetre  $\beta_i$  s'interpreta com el quocient d'avantatges que s'obté si el factor  $X_i$  és present en relació a si és absent, mantenint constants els valors de la resta de variables explicatives. Si es tracta d'una covariable en lloc d'un factor, l'exponencial del paràmetre  $\beta_i$  s'interpreta com el quocient d'avantatges que s'obté si la covariable  $X_i$  augmenta el seu valor en una unitat.

Per exemple, en la Taula 26 (pàgina 50), s'observa que  $\beta_{\text{NEE}} = 2,87$ . Així,

$$\widehat{\text{OR}}(\text{NEE}) = e^{\hat{\beta}_{\text{NEE}}} = e^{2,87} \approx 17,6,$$

és a dir,

$$\frac{\frac{P(\text{aprovar 1a avaluació, sent NEE})}{P(\text{no aprovar 1a avaluació, sent NEE})}}{\frac{P(\text{aprovar 1a avaluació, no sent NEE})}{P(\text{no aprovar 1a avaluació, no sent NEE})}} \approx 17,6.$$

Resulta més entenedor presentar les prediccions en termes de probabilitat, per a facilitar la interpretació del model, tal com es fa en la secció següent.

## 9.2.3 Prediccions del model ajustat

La Taula 29 (pàgina 52) mostra les prediccions per a un individu amb qualificació de 5 sobre 10 en el Bloc 1 (juny) i en els Blocs 2, 3 i 6 (setembre) on es pot observar que el model prediu una probabilitat d'aprovar la primera avaluació de 1r curs d'ESO tant si l'individu és tipificat NEE (99,3%) com si no ho és (88,8%).

Tanmateix, la Taula 28 (pàgina 52) mostra les prediccions per a diferents combinacions de les qualificacions en els Blocs involucrats en el model. S'han considerat únicament els casos amb la mateixa qualificació –i només entre 3 i 7–, atenenet al gran nombre de combinacions de qualificacions possible<sup>13</sup> i tenint en compte que els quatre Blocs presents en el model tenem pesos similars (Taula 26). Per a obtenir prediccions en altres casos no presents en la Taula 28, es pot procedir al seu càlcul manual segons s'explica en l'exemple que es mostra al final de la secció 8.2.2.

# 10 Resultats

En aquest estudi es va fer un seguiment a un grup d'alumnat des del final de l'últim curs de Primària (juny de 2008; 6 CEIPs) fins al final de la primera avaluació del 1r curs de Secundària (desembre de 2008; 3 IES). Aquest seguiment va consistir en una prova escrita que avaluava continguts curriculars de Primària de l'àrea de Matemàtiques –que es va passar al juny i també al setembre–, i en

<sup>13</sup>El nombre total de combinacions possibles considerant el factor NEE i les qualificacions enteres entre 0 i 10 proporcionaria una taula de prediccions de  $2 \cdot 11^4 = 29282$  files.

NEE	B1 (juny)	B2 (set.)	B3 (set.)	B6 (set.)	Pred.	L.I. (95%)	L.S. (95%)
No	3	3	3	3	22,8	6,2	56,8
	4	4	4	4	60,5	26,5	86,7
	5	5	5	5	88,8	61,8	97,5
	6	6	6	6	97,6	86,3	99,6
	7	7	7	7	99,5	95,7	100,0
Sí	3	3	3	3	83,9	50,7	96,3
	4	4	4	4	96,4	81,2	99,4
	5	5	5	5	99,3	94,2	99,9
	6	6	6	6	99,9	98,3	100,0
	7	7	7	7	100,0	99,5	100,0

**Taula 28:** Prediccions (puntuals i per interval de confiança al 95%) per a l'alumnat amb qualificació entre 3 i 7 sobre 10, igual en els Blocs 1 (juny) i 2, 3 i 6 (setembre), segons el model ajustat per a la probabilitat d'aprovar la primera avaluació de 1r d'ESO (desembre).

NEE	B1 (juny)	B2 (set.)	B3 (set.)	B6 (set.)	Pred.	L.I. (95%)	L.S. (95%)
No	5	5	5	5	88,8	61,8	97,5
Sí	5	5	5	5	99,3	94,2	99,9

**Taula 29:** Prediccions (puntuals i per interval de confiança al 95%) per a l'alumnat que va obté qualificació de 5 sobre 10 en el Bloc 1 en juny i en els Blocs 2, 3 i 6 en setembre, segons el model ajustat per a la probabilitat d'aprovar la primera avaluació de 1r d'ESO (desembre).

l'obtenció de la qualificació en l'àrea de matemàtiques corresponent a la primera avaluació del 1r curs d'ESO (desembre). La modelització de les dades obtingudes va permetre obtenir els resultats següents:

- Del grup inicial de 255 individus sobre el qual es va començar el seguiment en acabar l'últim curs de Primària del curs 2008–2009:
  - 55 individus només van realitzar la prova de juny perquè al setembre no es van matricular en cap dels 3 IES analitzats. Els resultats d'aquest alumnat en la prova de juny no van ser significativament diferents dels de la resta d'alumnat.
  - 49 individus només van realitzar la prova de setembre perquè no van cursar l'últim curs de la Primària en cap dels 6 CEIPs analitzats. Els resultats d'aquest alumnat en la prova de setembre van ser significativament inferiors als corresponents a l'alumnat que va fer la prova al juny i al setembre, en els Blocs 2 (Jerarquia d'operacions; 1 punt sobre 10 menys en mitjana), 4 (Fraccions i canvis d'unitats, 1,3 punts menys) i 6 (Problemes; 1 punt menys). Una possible explicació d'aquest resultat és la suma de dues causes: d'una banda, aquest alumnat no havia fet la prova al juny i, per tant, no en podien recordar res i, d'altra banda, el contingut d'aquesta prova es va dissenyar en coherència amb els continguts treballats en els 6 CEIPs analitzats.
  - La resta d'alumnat, 181 individus, va realitzar la prova al juny i al setembre encara que només 168 dels quals presentaven també la qualificació en la primera avaluació de 1r d'ESO. Sobre aquest conjunt

de 168 individus (65,9% de la mostra inicial) es va fer el seguiment del qual deriven la resta de resultats que s'exposen a continuació.

- L'anàlisi descriptiva va indicar que els factors "Sexe", "CEIP" i "IES" mostraven les seves categories ben representades. Això no va passar en el cas dels factors "Ha repetit 6è?", amb només un 1% d'alumnat repetidor; "Cat2" (vingut a Catalunya fa menys de 2 anys), amb només un 6% de casos afirmatius, i "NEE (Alumnat tipificat amb Necessitats Educatives Especials)", amb un 13% de casos afirmatius. De fet, la gairebé nul·la presència d'alumnat repetidor va implicar que aquest factor no fos significatiu en cap dels models ajustats i, per tant, no va ser possible mesurar la seva associació amb els resultats acadèmics considerats en l'estudi. A més, els factors "Mes de naixement" i "Sexe" tampoc no van ser significatius en cap dels models ajustats i, per tant, es pot acceptar que no hi ha diferències significatives entre nois i noies<sup>14</sup>. La covariable "Edat" tampoc no va resultar significativa en cap model ajustat, ni considerant-la com a factor després d'haver-la categoritzat per intervals.

Un fet descatable és que la qualificació mitjana en la primera avaluació de 1r curs d'ESO (5,4) va ser superior tant a la qualificació mitjana en la prova de setembre (4,6) com a la de la prova de juny (5,1). Això es pot interpretar com una possible relaxació en el nivell dels continguts mínims necessaris per aprovar la primera avaluació de 1r curs d'ESO, i un possible motiu d'això podria ser la tendència, per part del professorat de 1r curs d'ESO, a "partir de zero", repassant continguts que ja es van treballar en profunditat a Primària. També es palesa que, en tots els Blocs de la prova, la mitjana de la diferència entre juny i setembre va resultar negativa, indicant una devallada generalitzada dels resultats de la prova entre juny i setembre (Bloc 1: 0,7; Bloc 2: 0,4; Bloc 3: 1,0; Bloc 4: 0,5; Bloc 5: 0,4; Bloc 6: 0,3).

- Les correlacions parcials 2 a 2 entre les qualificacions d'un mateix Bloc en la prova de juny i en la prova de setembre van donar explicabilitats del 29%, 46%, 52%, 65%, 47% i 41% en els 6 Blocs, respectivament. Així, el Bloc menys correlacionat entre juny i setembre va ser el Bloc 1 (+ - · : amb naturals), i el més correlacionat, el Bloc 4 (Fraccions i canvis d'unitats).

Quant a la correlació parcial entre els diferents Blocs d'una mateixa prova, en la prova de juny la correlació més baixa es va donar entre els Bloc 3 –Càlcul amb decimals– i 6 –Problemes– (29%), i la és elevada entre els Blocs 2 –Jerarquia d'operacions– i 3 (47%). En la prova de setembre, la correlació més baixa es va donar entre els Bloc 1, + - · : amb naturals, i

<sup>14</sup>La no significació del factor "Sexe" també es va donar en la modelització dels resultats obtinguts en les proves Cangur 2007 de l'àmbit de Mataró, reflectida en un treball realitzat per mi mateix.

6 –Problemes– (11%), i la és elevada entre els Blocs 4 –Fraccions i canvis d'unitats– i 5 –Geometria– (47%).

Aquestes correlacions tan baixes es podrien interpretar com una conseqüència d'una metodologia d'ensenyament–aprenentatge “atomitzada” en la que es tendeix a ensenyar (i, per tant, a aprendre) els diferents continguts matemàtics en base a procediments senzills, independents i repetitius que, tot i que poden afavorir-ne la memotització, s'allunyen d'un ensenyament constructivista de les matemàtiques que permeti assolir els objectius curriculars i conservar-los a curt termini.

- En la modelització dels resultats de la prova de juny es va ajustar un Model Lineal Mixt amb resposta contínua per a explicar la qualificació en cadascun dels 6 Blocs de la prova. El millor model ajustat, una vegada validat, conté, com a variables explicatives, els factors “Cat2”, “NEE” i “CEIP”. El model preveu que la incorporació recent a Catalunya (“Cat2”) implica una reducció mitjana d'1,4 punts en qualsevol Bloc i la tipificació NEE implica també una reducció de 2,5 punts. El factor “CEIP” tot i que significatiu al model, implica diferències menys importants entre centres. Prenent com a referència un individu no incorporat a Catalunya fa menys de 2 anys, no tipificat NEE i pertanyent al CEIP 1, el model prediu les qualificacions següents: Bloc 1 (+ – · : amb naturals), 7,36; Bloc 2 (Jerarquia d'operacions), 5,16; Bloc 3 (Càlcul amb decimals), 5,27; Bloc 4 (Fraccions i canvis d'unitats), 5,38; Bloc 5 (Geometria), 5,17; Bloc 6 (Problemes), 1,94.

Així, el model va reflectir que el Bloc 1 va presentar una bona qualificació, mentre que les qualificacions en els Blocs 2, 3, 4 i 5 van ser baixes. La molt baixa qualificació en el Bloc 6 es pot interpretar com l'existència d'una dificultat molt important per part de l'alumnat per a resoldre els problemes plantejats en la prova.

- Una primera aproximació per a comparar els resultats de la prova de juny i els de la de setembre va ser la utilització del Risc Relatiu (RR).

Per a un determinat Bloc, l'RR es va definir com

$$RR = \frac{P(\checkmark \text{ setembre} | \checkmark \text{ juny})}{P(\checkmark \text{ setembre} | \times \text{ juny})}$$

on  $\checkmark$  significa que en el Bloc considerat es va obtenir una qualificació no inferior a 5 sobre 10 i  $\times$  el contrari. Així, en aquest context, la paraula “Risc” s'ha d'entendre com un efecte positiu (risc d'aprovar). Es va obtenir  $RR_{B1} \approx 1$ ,  $RR_{B2} \approx 3$ ,  $RR_{B3} \approx 4$ ,  $RR_{B4} \approx 4$ ,  $RR_{B5} \approx 3$  i  $RR_{B6} \approx 12$ . Es pot observar el comportament especial del Bloc 6 (Problemes) que s'interpreta com que la probabilitat de superar aquest Bloc en la prova de setembre, si es va superar també en la prova de juny, és 12 vegades la mateixa probabilitat si el Bloc no es va superar en la prova de juny.

Això es pot interpretar com que l'alumnat que aprèn a resoldre problemes tendeix a assimilar-ho almenys a curt termini (juny–setembre).

- En la modelització dels resultats comparatius entre la prova de juny i la de setembre es va ajustar un Model Lineal General Mixt amb resposta contínua per a modelitzar la diferència entre la qualificació de juny i la de setembre i un Model Lineal Generalitzat Mixt amb resposta binària per a modelitzar la probabilitat de pèrdua de qualificació; és a dir, la probabilitat que la qualificació en la prova de setembre fos inferior a la del mateix Bloc en la prova de juny.

Quant a la modelització de la diferència entre la qualificació en la prova de juny i la prova de setembre, el millor model ajustat i validat va indicar que aquesta diferència, en un determinat Bloc, es pot explicar pels factors “NEE” i “CEIP”, i la qualificació basal (és a dir, la qualificació en la prova de juny). Així, el model prediu que un individu no tipificat NEE, pertanyent al CEIP 1 i amb qualificació de 5 sobre 10 en cadascun dels Blocs de la prova de juny té un guany igual a 1,36 punts entre juny i setembre en el Bloc 1 (que va ser el millor assimilat en juny); de 0,5 punts en els Blocs 2, 4 i 5, i una pèrdua igual a 1 punt en el Bloc 6. En el Bloc 3 no hi ha variació.

De nou s'observa un comportament diferenciat en el Bloc dels problemes on la pèrdua esperada és més accentuada. Aquest fet podria semblar contradir els resultats que es mostren anteriorment respecte l'RR però això es pot explicar tenint en compte que la qualificació basal també va resultar significativa. En aquest sentit, el model prediu una pèrdua de 0,5 punts per cada punt per sobre de 5 en la qualificació basal. Per exemple, si la qualificació basal d'un determinat individu va ser de 9 punts, s'espera una diferència de  $-(9 - 5) \cdot 0,5 = -2$  punts (pèrdua) i, en canvi, si la qualificació basal va ser 4 s'espera una diferència de  $-(4 - 5) \cdot 0,5 = +0,5$  punts (guany). Així, les qualificacions més elevades mostren tendència, en mitjana, a una pèrdua més important que les qualificacions més baixes. Una possible explicació d'aquest fet és que l'alumnat que assimila deficientment alguns continguts tendeix a conservar la petita part assimilada i, en canvi, l'alumnat amb qualificació més elevada en la prova de juny té més marge per a cometre errades i/o oblidar alguns conceptes i procediments que es pot interpretar com una manca d'assimilació a curt termini.

El model prediu una pèrdua extra igual a 1,3 punts pel fet que l'individu sigui tipificat NEE. Aquest fet és esperable atenent les diferències curriculars de l'alumnat NEE, tal com passa en el model per a explicar les qualificacions de la prova de juny.

Es van obtenir resultats molt similars respecte al model anterior en la modelització de la probabilitat de pèrdua de qualificació entre juny i setembre: les variables explicatives van resultar les mateixes (“NEE”, “CEIP” i qualificació basal).

- En la modelització dels resultats en la primera avaluació de 1r curs d'ESO, es va ajustar un Model Lineal General amb resposta contínua per a modelitzar la qualificació numèrica i, a més, un Model Lineal Generalitzat amb resposta binària per a modelitzar la probabilitat d'aprovar l'avaluació.

El millor model ajustat i validat per a la qualificació numèrica en la primera avaluació del 1r curs d'ESO va incloure com a variables explicatives el factor "NEE" i les covariables corresponents a les qualificacions en els Blocs 2 (Jerarquia d'operacions), 3 (Càlcul amb decimals) i 6 (Problemes) de la prova de setembre. El model prediu una qualificació d'1,2 punts més si l'individu és tipificat NEE a diferència de si no ho és. Això es pot explicar tenint en compte l'atenció en petit grup que rep l'alumnat NEE i, sobretot, l'adaptació curricular que afecta aquest alumnat i suposa menys dificultat per assolir l'aprovat.

Quant a les covariables, el model prediu un augment de 0,16 punts, de 0,27 punts i de 0,37 punts en la qualificació de la primera avaluació de 1r curs d'ESO per cada punt que hagués augmentat la qualificació en el Bloc 2, en el 3 i en el 6, respectivament, de la prova de setembre. Així, el Bloc que més pes presenta és el 6 (Problemes) de manera que una bona qualificació en aquest Bloc fa millorar la predicció en la qualificació de la primera avaluació de 1r curs d'ESO en relació a la resta de Blocs. Cal destacar que cap dels 3 IES analitzats va treballar específicament continguts de resolució de problemes durant la primera avaluació del 1r curs d'ESO.

Segons aquest model, per a un individu no tipificat NEE i amb qualificació de 5 sobre 10 en cadascun dels Blocs 2, 3 i 6 de la prova de setembre s'espera una qualificació en la primera avaluació del 1r curs d'ESO de 6,7 punts, mentre que per a un individu tipificat NEE i amb una qualificació de només 4 sobre 10 en els mateixos Blocs 2 s'espera una qualificació de 7,0 punts.

Resulta interessant observar que ni el CEIP de procedència de l'alumnat ni l'IES de destinació no van resultar factors significatius. És a dir, s'ha donat una homogeneïtzació de l'alumnat de manera que s'ha dissolt l'efecte que cada CEIP va poder causar sobre el seu alumnat i, a més, no s'han detectat diferències significatives entre IES quant a la qualificació en la primera avaluació del 1r curs d'ESO. Això últim no vol dir que les diferents característiques en la metodologia d'ensenyament que poden haver, de manera específica, en els diferents IES no tinguin efectes diferents sobre l'alumnat considerent-les aïlladament una per una; però, en tot cas, considerant-les en conjunt, centre per centre, no es pot dir que aquestes metodologies diferents hagin afectat els resultats acadèmics de l'alumnat.

Un avantatge d'aquest model és que es pot fer servir en l'IES com a predictor, al setembre, dels resultats en la primera avaluació de 1r curs d'ESO, amb tres mesos d'antelació. Això és possible perquè només cal conèixer les qualificacions en els Blocs 2, 3 i 6 de la prova de setembre –que es fa

en el mateix IES i és avaluada pel mateix professorat que impartirà classe a aquest alumnat en el 1r curs d'ESO, i la tipificació NEE de l'individu d'interès. Si aquestes prediccions fossin bones –comprovant-ho una vegada conegudes les qualificacions reals al desembre–, es podria acceptar l'estabilitat<sup>15</sup> del model i fer-lo servir com a predictor en promocions futures dels IES analitzats, de manera que aquestes prediccions fetes al setembre es podrien considerar una eina de detecció d'alumnat en risc de fracàs escolar (en l'àrea de Matemàtiques).

Finalment, es va ajustar i validar un model per a la probabilitat d'aprovar la primera avaluació del 1r curs d'ESO. El millor model va incloure les mateixes variables explicatives que el model quantitatiu comentat anteriorment i, a més, la covariable corresponent al Bloc 1 (+ - · : amb naturals) en la prova de juny. Aquest Bloc és el més elemental i va ser el que va mostrar millors resultats respecte la resta de Blocs, amb diferència (mitjanes de 7,5 i de 6,8 en juny i setembre, respectivament). Així, la seva inclusió en el model per a explicar la probabilitat d'aprovar la primera avaluació de 1r curs d'ESO es pot interpretar com que calen els coneixements mínims avaluats en aquest Bloc com a requisit mínim per a superar l'avaluació. Aquest model va presentar una capacitat predictiva elevada, amb una sensibilitat del 79% –probabilitat de predicció correcta de fracàs–, una especificitat del 90% –probabilitat de predicció correcta d'un aprovat– i una probabilitat de classificació correcta suspès/aprovat del 86%.

## 11 Limitacions

Els resultats que s'han presentat en la secció anterior han de ser considerats amb les reserves derivades de les inevitables limitacions que sorgeixen en un treball com aquest, essencialment experimental.

En aquest sentit, cal tenir en compte els aspectes següents:

- Tots els resultats presentats en aquest document es deriven d'una prova escrita que va ser dissenyada entre les persones representants dels departaments de matemàtiques dels 9 centres analitzats, amb l'objectiu de millorar la continuïtat curricular en l'àrea de Matemàtiques en la transició Primària–Secundària entre aquests centres. Tot i així, resulta inevitable la component subjectiva per part d'aquestes persones (entre les que m'incloc) pel que fa a la tria del contingut, l'estructura i el format de la prova. És per això que s'ha de considerar la possibilitat que els resultats obtinguts siguin sensibles a petites variacions en el disseny de la prova.

---

<sup>15</sup>Més correctament, per acceptar l'estabilitat d'un model, caldria ajustar-lo de nou sobre un conjunt diferent de dades i comprovar que les variables explicatives en el nou model són les mateixes i que els seus coeficients en el model són molt similars.

- Pel que fa a l'avaluació de la prova, tant al juny com al setembre, s'ha de tenir en compte que va ser realitzada per diferents persones (les responsables en cada centre) i, tot i que es va dissenyar un protocol d'avaluació (Annex B) molt precís i amb poc marge per a la variabilitat, no es poden descartar petites variacions en els criteris d'avaluació.
- Un altre aspecte a tenir en compte és que durant el període de temps entre la prova de juny i la de setembre (vacances estivals) no es va fer cap tipus de seguiment de l'alumnat analitzat i, per tant, no es pot garantir la igualtat de tractament en aquest sentit (realització d'algun tipus d'activitat acadèmica durant les vacances...).
- Per tal que un model ajustat i validat, com tots els que es presenten en aquest treball, sigui predictiu a més d'explicatiu, cal analitzar la seva estabilitat. És per això que idealment caldria tornar a ajustar tots els models sobre la següent promoció d'alumnat dels nivells acadèmics i dels centres analitzats i comprovar si les variables explicatives que els models inclouen són les mateixes i si els seus coeficients són similars. Si això no es fa, no es pot garantir la predictibilitat.
- Només un 1% de l'alumnat analitzat és repetidor de 6è curs de Primària. Aquesta menyspreable fracció implica que el factor "Repetició" no hagi estat significatiu en cap model i, per tant, no s'hagi pogut avaluar el seu efecte en les variables resposta modelitzades.
- El fet que el factor "Csec" no hagi estat significatiu en la modelització de la qualificació en la primera avaluació del 1r curs d'ESO implica que no van haver diferències significatives entre els tres IES considerats. Això s'ha d'entendre a nivell global; és a dir, no es pot avaluar l'efecte de les diferents estratègies didàctiques aplicades en els tres centres per separat, sinó globalment. Per exemple, si en un centre es fa un ús del projector a classe més intens que en un altre, no es pot dir que això no afecta els resultats perquè l'ús del projector és una eina aïllada dins del conjunt d'estratègies d'ensenyament que pot aplicar cada centre (ús de la pissarra, treball en petit grup, activitats de lectura comprensiva...), i en el model ajustat no es consideren aquestes estratègies per separat sinó totes juntes integrades en el factor "Csec".



## A Enunciat de la prova escrita

Nom i cognoms: \_\_\_\_\_ Data: \_\_\_\_\_

Centre: \_\_\_\_\_

**No escriguis dins del quadres**

**BLOC 1**

**Puntuació:**

1. Escriu amb paraules la quantitat 8.037
2. Escriu amb xifres la quantitat "vint mil quatre-cents cinc"
3. Escriu la xifra de les centenes del nombre 1.698
4. Calcula  $495 + 1.737 =$
5. Calcula  $147 - 59 =$
6. Calcula  $26 - 15 + 40 - 34 =$
7. Calcula  $52 \cdot 6 =$
8. Calcula  $3 \cdot 307 =$
9. Calcula  $6.000 \cdot 50.000 =$
10. Escriu una multiplicació que tingui per resultat 56

**BLOC 2**

**Puntuació:**

1. Calcula  $10 + 3 \cdot 4 =$

2. Calcula  $(10+3) \cdot 4 =$

3. Calcula  $4 \cdot 7 - 6 \cdot 3 =$

4. Calcula  $3 + (2 + 5 - 7) =$

5. Escriu què et contestarà una calculadora si hi tecleges  $8 \times 2 - 4 \times 3 =$

6. Escriu què et contestarà una calculadora si hi tecleges  $8 \times 2 \div 4 - 3 =$

7. Escriu i calcula el quadrat de 6

8. Escriu i calcula el cub de 10

9. Calcula  $(8 - 5)^2 =$

10. Calcula  $8^2 - 5^2 =$

**BLOC 3**

Puntuació:

1. Fes la divisió de 48 entre 3
  
2. Calcula el quocient de la divisió sense decimals de 92 entre 7
  
3. Calcula el residu de la divisió sense decimals de 137 entre 9
  
4. Fes la divisió amb un decimal de 75 entre 4
  
5. Fes la divisió amb un decimal de 6 entre 7
  
6. Escribeu amb xifres el decimal "quatre centèsims"
  
7. Calcula  $5,32 + 17,8 =$
  
8. Calcula  $17,8 - 5,32 =$
  
9. Calcula  $43 : 100 =$
  
10. Calcula  $5,6 \cdot 1.000 =$

**BLOC 4**

**Puntuació:**

1. Escribe amb xifres “tres setens”

2. Calcula el valor decimal de  $\frac{2}{5}$

3. Calcula la meitat de 192

4. Calcula la quarta part de 84

5. Marca la tercera part de la quadrícula següent



6. Quants centímetres té un metre?

7. Quants metres té mig quilòmetre?

8. Quants minuts hi ha a la tercera part d'una hora?

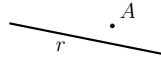
9. Quines són, de manera aproximada i indicant les unitats, la llargada i l'amplada d'aquest full de paper?

10. Un refresc val 90 cèntims. Cinc amics posen 75 cèntims cadascun per comprar-ne tres. Calcula quant canvi li correspon a cada amic.

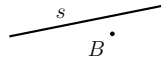
**BLOC 5**

Puntuació:

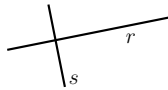
1. Dibuixa la paral·lela a la recta  $r$  que passa pel punt  $A$



2. Dibuixa la perpendicular a la recta  $s$  que passa pel punt  $B$



3. Com són les rectes  $r$  i  $s$  del dibuix següent?



4. Dibuixa un angle obtús

5. Aquest angle té per mesura un dels tres valors següents:  $70^\circ$ ,  $100^\circ$ ,  $30^\circ$ . Quin és el correcte?



6. Dibuixa un triangle rectangle

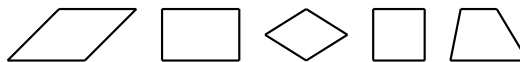
7. Com s'anomena la forma d'aquest full de paper?

8. Com es calcula l'àrea d'un cercle?

9. Dibuixa un diàmetre de la circumferència següent



10. Encercla la figura que **no** és un paral·lelogram



**BLOC 6**

**Puntuació:**

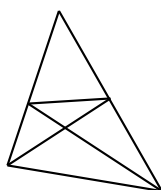
1. Escribe tres multiplicacions de dos nombres de manera que totes elles donin per resultat 462.

2. Quin nombre hem de multiplicar per 7 per tal que, si al resultat hi sumem 8, ens dóna el mateix resultat que si elevem 13 al quadrat?

3. Quants quadrats de 3 cm de costat caben en un rectangle de 20 cm de llargada i 13 cm d'amplada?

4. L'Alba, el Blai i la Carla tenen monedes de 5 cèntims. L'Alba i el Blai tenen el mateix nombre de monedes. La Carla en té quatre més que l'Alba i el Blai junts. Si entre tots tres tenen 3€, esbrina quantes monedes té cadascun.

5. Quants triangles hi ha en la figura següent?



## B Instruccions per a la realització de la prova

### 1. Condicions de realització de les proves

- **Sense calculadora.**
- **En “fred”, sense cap preparació prèvia.** És molt important que sigui així ja que estem interessats a analitzar el nivell d'assoliment dels continguts per part de l'alumnat.
- **Atenent l'elevat nombre de preguntes, la prova es fa en dues parts.** La primera part conté els blocs 1, 2, 3 i 4; la segona part conté els blocs 5 i 6.
- **Convé passar les dues parts en dies diferents** per tal d'evitar fatigar massa l'alumnat.
- La manca de temps no ha de ser un obstacle per a la realització de la prova. En aquest sentit, hauria de ser suficient **una hora per a cada part.**
- Si un alumne no assisteix el dia que es realitza alguna de les parts de la prova, pot fer-la en un altre moment però evitant que la prepari (que no accedeixi a l'enunciat).

### 2. Quan es passa la prova?

- A Primària, al final del curs, després que l'alumnat hagi estat avaluat i no tingui activitats acadèmiques simultànies que el puguin fer arribar fatigat a la realització de la prova. No ha d'haver fet cap altra prova de matemàtiques pocs dies abans. Pot ser interessant que el professorat miri de motivar l'alumnat per tal que es prenguin la prova amb interès, ja que si saben que no s'avalua, es poden relaxar.
- A Secundària, a l'inici del curs, idealment en la primera classe de matemàtiques. Aquesta prova es pot fer servir com a avaluació inicial.

### 3. Puntuació de les proves

Es puntuarà cada pregunta amb un 1 (resultat correcte) o un 0 (qualsevol altre cas).

## C Codi R

```
#####
#####
# Preparació de directoris:
#####
#####

# Fixem el directori actual com a directori de treball
# (on és el arxiu amb el codi):
workingDir <- getwd()
# Fixem la carpeta d'on llegir les dades:
dataDir <- file.path(workingDir, "Dades")
# Fixem la carpeta on desar els resultats:
resultatsDir <- file.path(workingDir,"Resultats")
# Fixem el directori de treball:
setwd(workingDir)

#####
#####
# Càrrega de llibreries:
#####
#####

library(doBy)
library(MASS)
library(limma)

#####
#####
# Opcions predeterminades:
#####
#####

a.latex <- TRUE
a.pdf <- TRUE
pdf.options(paper="a4", family="Helvetica", width=10, height=18)

#####
#####
# Funcions:
#####
#####

#####
# Resum descriptiu d'una covariable:
#####
#####

sumfun<-function(x,r=1,...)
{
  c(
    N=sum(complete.cases(x)),
    m=round(mean(x, ...),r),
```



```

    sd=round(sd(x, ...),r),
    Med=round(median(x, ...),r),
    Min=round(min(x, ...),r),
    Max=round(max(x, ...),r)
  )
}

#####
# Diagrama de sectors personalitzat:
#####

diagrama.sectors <- function(factor)
{
  library(Hmisc)
  x <- summary(factor)
  l <- length(x)
  perc <- round(100*x/sum(x))
  info <- paste(levels(factor),":", perc)
  info <- paste(info,"%", sep="")
  pie(x, labels=info, col=rainbow(l), main=label(factor))
  detach(package:Hmisc)
}

#####
# Diagrama de barres personalitzat:
#####

diagrama.barres <- function(factor, color)
{
  library(Hmisc)
  x <- summary(factor)
  l <- length(x)
  perc <- round(100*x/sum(x), 1)
  info <- paste(levels(factor),":", perc)
  info <- paste(info,"%", sep="")
  barplot(x, main=label(factor), ylab="Freqüència", col=color)
  text(1.2*(0:(l-1))+0.7, x/2, paste(perc,"%",sep=""), cex=0.6)
  detach(package:Hmisc)
}

#####
# Anàlisi de residus d'un lm:
#####

analisi.residus.lm <- function(model, titol, titol.pdf=NULL, a.pdf=FALSE)
{
  library(Hmisc)
  if (a.pdf==TRUE)
  {
    setwd(resultsDir)
    pdf(titol.pdf, family="Helvetica", width=6, height=6)
  }
  par(mfrow=c(2,2), oma=c(0,0,1,0), las=1, cex=0.7, cex.axis=0.6)
}

```

```

hist(rstudent(model), main="Histograma", xlab="Residus estudentitzats",
     ylab="Densitat",
     freq=FALSE)
mr <- mean(rstudent(model))
sdr <- sd(rstudent(model))
curve(dnorm(x, mean=mr, sd=sdr), lwd=2, add=TRUE)
boxplot(rstudent(model), main="Diagrama de caixa", xlab="Residus estudentitzats")
qqnorm(rstudent(model), main="Q-Q plot per a normalitat", xlab="Quantils teòrics",
       ylab="Quantils mostrals")
abline(0, 1)
plot(cooks.distance(model), rstudent(model), main="Observacions influents",
     xlab="Distància de Cook", ylab="Residus estudentitzats")
abline(h=-2, lty=2)
abline(h=2, lty=2)
abline(v=4/(dim(dades)[1]-model$rank), lty=2)
title(paste("Residus del model per ", titol, sep=""), outer=T)
if (a.pdf)
  {
    dev.off()
    setwd(workingDir)
  }
detach(package:Hmisc)
}

#####
# Anòmals d'un lm:
#####

anomals <- function(model)
{
  anomals <- NULL
  anomals$extrems <- which(abs(rstudent(model))>2)
  anomals$cooks <-
    which(cooks.distance(model)>4/(dim(mostra)[1]-model$rank)|rstudent(model)>2)
  anomals
}

#####
#####
# Manipulació de dades:
#####
#####

#####
# Dades originals, mesures no repetides
# (dadesoriginals):
#####

library(Hmisc)
dades <- csv.get("dades.csv", sep=";", dec=",", datevars="dbay")
detach(package:Hmisc)

# Qualificació global en la prova (mitjana dels 6 blocs):

```

```
dades$GP <- (dades$B1P + dades$B2P + dades$B3P +
            dades$B4P + dades$B5P + dades$B6P)/6
dades$GS <- (dades$B1S + dades$B2S + dades$B3S +
            dades$B4S + dades$B5S + dades$B6S)/6

# Diferències entre juny i setembre:
dades$dif.B1 <- dades$B1S - dades$B1P
dades$dif.B2 <- dades$B2S - dades$B2P
dades$dif.B3 <- dades$B3S - dades$B3P
dades$dif.B4 <- dades$B4S - dades$B4P
dades$dif.B5 <- dades$B5S - dades$B5P
dades$dif.B61 <- dades$B61S - dades$B61P
dades$dif.B62 <- dades$B62S - dades$B62P
dades$dif.B63 <- dades$B63S - dades$B63P
dades$dif.B64 <- dades$B64S - dades$B64P
dades$dif.B65 <- dades$B65S - dades$B65P
dades$dif.B6 <- dades$B6S - dades$B6P
dades$dif.G <- dades$GS - dades$GP

# Etiquetes i definició de factors:
library(Hmisc)
label(dades$Id) <- "Id"

label(dades$Edat) <- "Edat"

dades$Sexe <- as.factor(dades$Sexe)
label(dades$Sexe) <- "Sexe"
levels(dades$Sexe) <- c("Noia", "Noi")

dades$Edat.ent <- as.factor(round(dades$Edat))
label(dades$Edat.ent) <- "Edat (anys sencers)"

dades$MesNa <- as.factor(dades$MesNa)
label(dades$MesNa) <- "Mes de naixement"
levels(dades$MesNa) <- c("Gen", "Feb", "Mar", "Abr", "Mai", "Jun",
                        "Jul", "Ago", "Set", "Oct", "Nov", "Des")

dades$Rep <- as.factor(dades$Rep)
label(dades$Rep) <- "Ha repetit 6è?"
levels(dades$Rep) <- c("No", "Sí")

dades$Cat2 <- as.factor(dades$Cat2)
label(dades$Cat2) <- "Vingut a Catalunya fa menys de 2 anys?"
levels(dades$Cat2) <- c("No", "Sí")

dades$Cprim <- as.factor(dades$Cprim)
label(dades$Cprim) <- "Centre de primària"
levels(dades$Cprim) <- c("CEIP 1", "CEIP 2", "CEIP 3", "CEIP 4",
                        "CEIP 5", "CEIP 6", "Altres CEIP")

dades$Csec <- as.factor(dades$Csec)
label(dades$Csec) <- "Centre de secundària"
levels(dades$Csec) <- c("IES 1", "IES 2", "IES 3", "Altres IES")
```

```

label(dades$B1P) <- "+ - . : amb naturals (juny)"
label(dades$B2P) <- "Jerarquia d'operacions (juny)"
label(dades$B3P) <- "Càlcul amb decimals (juny)"
label(dades$B4P) <- "Fraccions i canvis d'unitats (juny)"
label(dades$B5P) <- "Geometria (juny)"
label(dades$B61P) <- "Problema 1 (juny)"
label(dades$B62P) <- "Problema 2 (juny)"
label(dades$B63P) <- "Problema 3 (juny)"
label(dades$B64P) <- "Problema 4 (juny)"
label(dades$B65P) <- "Problema 5 (juny)"
label(dades$B6P) <- "Problemes (juny)"
label(dades$GP) <- "Global (juny)"
label(dades$B1S) <- "+ - . : amb naturals (setembre)"
label(dades$B2S) <- "Jerarquia d'operacions (setembre)"
label(dades$B3S) <- "Càlcul amb decimals (setembre)"
label(dades$B4S) <- "Fraccions i canvis d'unitats (setembre)"
label(dades$B5S) <- "Geometria (setembre)"
label(dades$B61S) <- "Problema 1 (setembre)"
label(dades$B62S) <- "Problema 2 (setembre)"
label(dades$B63S) <- "Problema 3 (setembre)"
label(dades$B64S) <- "Problema 4 (setembre)"
label(dades$B65S) <- "Problema 5 (setembre)"
label(dades$B6S) <- "Problemes (setembre)"
label(dades$GS) <- "Global (setembre)"
label(dades$dif.B1) <- "+ - . : amb naturals (jun -> set)"
label(dades$dif.B2) <- "Jerarquia d'operacions (jun -> set)"
label(dades$dif.B3) <- "Càlcul amb decimals (jun -> set)"
label(dades$dif.B4) <- "Fraccions i canvis d'unitats (jun -> set)"
label(dades$dif.B5) <- "Geometria (jun -> set)"
label(dades$dif.B61) <- "Problema 1 (jun -> set)"
label(dades$dif.B62) <- "Problema 2 (jun -> set)"
label(dades$dif.B63) <- "Problema 3 (jun -> set)"
label(dades$dif.B64) <- "Problema 4 (jun -> set)"
label(dades$dif.B65) <- "Problema 5 (jun -> set)"
label(dades$dif.B6) <- "Problemes (jun -> set)"
label(dades$dif.G) <- "Global (jun -> set)"
label(dades$Aval1) <- "Qualificació 1a Avaluació (desembre)"

dades$NEE <- as.factor(dades$NEE)
label(dades$NEE) <- "NEE"
levels(dades$NEE) <- c("No", "Sí")

dades$Aprova <- as.factor(dades$Aprova)
label(dades$Aprova) <- "Aprova 1a Aval."
levels(dades$Aprova) <- c("No", "Sí")

dades$Flux <- as.factor(dades$Flux)
label(dades$Flux) <- "Flux primària-secundària"
levels(dades$Flux) <- c("Prim -> Sec", "Prim -> Altres", "Altres -> Sec")
detach(package:Hmisc)

dadesoriginals <- dades

```

```
save(dadesoriginals, file=file.path(dataDir, "dadesoriginals.RData"))
dades <- NULL
dadesoriginals <- NULL

#####
# Dades originals, mesures repetides
# (dadesorigexpand):
#####

library(Hmisc)
dades <- csv.get("dadesexpand.csv", sep=";", dec=",", datevars="dbay")
detach(package:Hmisc)

# Etiquetes i definició de factors:

library(Hmisc)
dades$Sexe <- as.factor(dades$Sexe)
label(dades$Sexe) <- "Sexe"
levels(dades$Sexe) <- c("Noia", "Noi")

label(dades$Edat) <- "Edat (anys)"

dades$Edat.ent <- as.factor(round(dades$Edat))
label(dades$Edat.ent) <- "Edat (anys sencers)"

dades$MesNa <- as.factor(dades$MesNa)
label(dades$MesNa) <- "Mes de naixement"
levels(dades$MesNa) <- c("Gen", "Feb", "Mar", "Abr", "Mai", "Jun",
                        "Jul", "Ago", "Set", "Oct", "Nov", "Des")

dades$Rep <- as.factor(dades$Rep)
label(dades$Rep) <- "Ha repetit 6è?"
levels(dades$Rep) <- c("No", "Sí")

dades$Cat2 <- as.factor(dades$Cat2)
label(dades$Cat2) <- "Vingut a Catalunya fa menys de 2 anys?"
levels(dades$Cat2) <- c("No", "Sí")

dades$NEE <- as.factor(dades$NEE)
label(dades$NEE) <- "NEE"
levels(dades$NEE) <- c("No", "Sí")

dades$Cprim <- as.factor(dades$Cprim)
label(dades$Cprim) <- "Centre de primària"
levels(dades$Cprim) <- c("CEIP 1", "CEIP 2", "CEIP 3", "CEIP 4",
                        "CEIP 5", "CEIP 6", "Altres CEIP")

dades$Csec <- as.factor(dades$Csec)
label(dades$Csec) <- "Centre de secundària"
levels(dades$Csec) <- c("IES 1", "IES 2", "IES 3", "Altres IES")

dades$Flux <- as.factor(dades$Flux)
label(dades$Flux) <- "Flux primària-secundària"
```

```

levels(dades$Flux) <- c("Prim -> Sec","Prim -> Altres", "Altres -> Sec")

dades$Bloc <- as.factor(dades$Bloc)
label(dades$Bloc) <- "Bloc"
levels(dades$Bloc) <- c("+ - . : amb naturals",
                        "Jerarquia d'operacions",
                        "Càlcul amb decimals",
                        "Fraccions i canvis d'unitats",
                        "Geometria",
                        "Problemes")

label(dades$Np) <- "Qualificació prova de juny"
label(dades$Ns) <- "Qualificació prova de setembre"
label(dades$Dif) <- "Increment qualificació (juny -> setembre)"

dades$Perd <- as.factor(dades$Perd)
label(dades$Perd) <- "Empitjora de juny a setembre?"
levels(dades$Perd) <- c("No", "Sí")
detach(package:Hmisc)

dadesoriginrepet <- dades
save(dadesoriginrepet, file=file.path(dataDir, "dadesoriginrepet.RData"))
dades <- NULL
dadesoriginrepet <- NULL

#####
# Dades completes, alumnat intern, mesures no repetides
# (dadesnetes):
#####

load(file=file.path(dataDir, "dadesoriginals.RData"))
dades <- dadesoriginals
dades <- subset(dadesoriginals, Flux=="Prim -> Sec")
dades$Flux <- NULL
dades <- subset(dades, complete.cases(dades)==TRUE)
library(Hmisc)
new.levels.Edat.ent <- levels(dades$Edat.ent)[2:4]
dades$Edat.ent <- as.factor(as.numeric(dades$Edat.ent))
levels(dades$Edat.ent) <- new.levels.Edat.ent
label(dades$Edat.ent) <- "Edat (anys sencers)"
new.levels.Cprim <- levels(dades$Cprim)[1:6]
new.levels.Csec <- levels(dades$Csec)[1:3]
dades$Cprim <- as.factor(as.numeric(dades$Cprim))
levels(dades$Cprim) <- new.levels.Cprim
label(dades$Cprim) <- "Centre de primària"
dades$Csec <- as.factor(as.numeric(dades$Csec))
levels(dades$Csec) <- new.levels.Csec
label(dades$Csec) <- "Centre de secundària"

# Categorització de l'edat segons quantils:

Edat.quant2.mesos.mitjana <- quantile(dades$Edat, probs = c(0, 2/3, 1))

```

```

Edat.quant2.mesos.mitjana[1] <- Edat.quant2.mesos.mitjana[1]-0.001
Edat.quant2.mesos.mitjana[3] <- Edat.quant2.mesos.mitjana[3]+0.001

dades$Edat2 <- cut(dades$Edat, breaks=Edat.quant2.mesos.mitjana,
                  labels=c("(11,4;12,2]", "(12,2;13,5]"))

Edat.quant3.mesos.mitjana <- quantile(dades$Edat, probs = seq(0, 1, 1/3))
Edat.quant3.mesos.mitjana[1] <- Edat.quant3.mesos.mitjana[1]-0.001
Edat.quant3.mesos.mitjana[4] <- Edat.quant3.mesos.mitjana[4]+0.001

dades$Edat3 <- cut(dades$Edat, breaks=Edat.quant3.mesos.mitjana,
                  labels=c("(11,4;11,8]", "(11,8;12,2]", "(12,2;13,5]"))

Edat.quant4.mesos.mitjana <- quantile(dades$Edat, probs = seq(0, 1, 1/4))
Edat.quant4.mesos.mitjana[1] <- Edat.quant4.mesos.mitjana[1]-0.001
Edat.quant4.mesos.mitjana[5] <- Edat.quant4.mesos.mitjana[5]+0.001

dades$Edat4 <- cut(dades$Edat, breaks=Edat.quant4.mesos.mitjana,
                  labels=c("(11,4;11,7]", "(11,7;12,0]",
                           "(12,0;12,3]", "(12,3;13,5]"))

detach(package:Hmisc)
dadesnetes <- dades
save(dadesnetes, file=file.path(dataDir, "dadesnetes.RData"))
dades <- NULL
dadesnetes <- NULL

#####
# Dades completes, alumnat intern, mesures repetides
# (dadesnetesrep):
#####

load(file=file.path(dataDir, "dadesoriginrepet.RData"))
dades <- subset(dadesoriginrepet, Flux=="Prim -> Sec")
dades$Flux <- NULL
dades <- subset(dades, complete.cases(dades)=="TRUE")
library(Hmisc)
new.levels.Cprim <- levels(dades$Cprim)[1:6]
new.levels.Csec <- levels(dades$Csec)[1:3]
dades$Cprim <- as.factor(as.numeric(dades$Cprim))
levels(dades$Cprim) <- new.levels.Cprim
dades$Csec <- as.factor(as.numeric(dades$Csec))
levels(dades$Csec) <- new.levels.Csec

# Categorització de l'edat segons quantils:

Edat.quant2.mesos.mitjana <- quantile(dades$Edat, probs = c(0, 2/3, 1))
Edat.quant2.mesos.mitjana[1] <- Edat.quant2.mesos.mitjana[1]-0.001
Edat.quant2.mesos.mitjana[3] <- Edat.quant2.mesos.mitjana[3]+0.001

dades$Edat2 <- cut(dades$Edat, breaks=Edat.quant2.mesos.mitjana,
                  labels=c("(11,4;12,2]", "(12,2;13,5]"))

```

```

Edat.quant3.mesos.mitjana <- quantile(dades$Edat, probs = seq(0, 1, 1/3))
Edat.quant3.mesos.mitjana[1] <- Edat.quant3.mesos.mitjana[1]-0.001
Edat.quant3.mesos.mitjana[4] <- Edat.quant3.mesos.mitjana[4]+0.001

dades$Edat3 <- cut(dades$Edat, breaks=Edat.quant3.mesos.mitjana,
                  labels=c("(11,4;11,8]", "(11,8;12,2]", "(12,2;13,5]"))

Edat.quant4.mesos.mitjana <- quantile(dades$Edat, probs = seq(0, 1, 1/4))
Edat.quant4.mesos.mitjana[1] <- Edat.quant4.mesos.mitjana[1]-0.001
Edat.quant4.mesos.mitjana[5] <- Edat.quant4.mesos.mitjana[5]+0.001

dades$Edat4 <- cut(dades$Edat, breaks=Edat.quant4.mesos.mitjana,
                  labels=c("(11,4;11,7]", "(11,7;12,0]",
                          "(12,0;12,3]", "(12,3;13,5]"))

detach(package:Hmisc)
dadesnetesrep <- dades
save(dadesnetesrep, file=file.path(dataDir, "dadesnetesrep.RData"))
dades <- NULL
dadesnetesrep <- NULL

#####
#####
# Caracterització de la distribució bruta de
# l'alumnat segons "interns"- "surten"- "entren":
#####
#####

load(file=file.path(dataDir, "dadesoriginals.RData"))
dades <- dadesoriginals

# Alumnat intern a primària:
cpr <- as.numeric(dades$Cprim) < 7

# Alumnat intern a secundària:
cse <- as.numeric(dades$Csec) < 4

#####
# Diagrama de Venn:
#####

if(a.pdf)
{
  setwd(resultsDir)
  pdf("01vennalumnat.pdf", paper="a4r", family="Helvetica", height=10)
}
par(oma=c(0,0,5,0))
vennDiagram(vennCounts(cbind(cpr, cse)),
            main="Distribució de l'alumnat segons la seva participació en les proves",
            names=c("Juny","Setembre"), lwd=2, counts.col="black", cex=1.2)
if (a.pdf)
{
  dev.off()
}

```



```

    setwd(workingDir)
  }

#####
# Taules i gràfics de la distribució bruta d'alumnat:
#####

library(Epi)

# Taules de contingència:

flux.prim.sec <- stat.table(list("C. primària"=Cprim, "C. secundària"=Csec),
                           list(count(), percent(Csec), percent(Cprim)),
                           data=dades, margins=TRUE)

detach(package:Epi)

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(flux.prim.sec[1,,], title="02fluxprimsecabs")
  latex(round(flux.prim.sec[2,,], 1), title="03fluxprimsecfila")
  latex(round(flux.prim.sec[3,,], 1), title="04fluxprimseccol")
  detach(package:Hmisc)
  setwd(workingDir)
}

# Diagrames de barres:

if(a.pdf)
{
  setwd(resultsDir)
  pdf("05distribruta.pdf", family="Helvetica", width=12, height=12)
}
par(mfrow=c(2,2), oma=c(0,0,1,0), las=1, cex.axis=0.7)
barplot(t(flux.prim.sec[1,1:7,1:4]), xlab="Centre de primària",
        ylab="Volum d'alumnat", ylim=c(0,60),
        col=c("orange", "navy", "pink", "seagreen"))
legend("topright", levels(dades$Csec),
      fill=c("orange", "navy", "pink", "seagreen"), cex=0.8, bty="n")
barplot(t(flux.prim.sec[2,1:7,1:4]), xlab="Centre de primària",
        ylab="Percentatge d'alumnat",
        col=c("orange", "navy", "pink", "seagreen"))
barplot(flux.prim.sec[1,1:7,1:4], xlab="Centre de secundària",
        ylab="Volum d'alumnat", ylim=c(0,100), col=2:8)
legend("topright", levels(dades$Cprim), fill=2:8, cex=0.8, bty="n")
barplot(flux.prim.sec[3,1:7,1:4], xlab="Centre de secundària",
        ylab="Percentatge d'alumnat", col=2:8)
title("Distribució de l'alumnat segons els centres de primària i de secundària",
      outer=T)
if (a.pdf)
{

```

```

dev.off()
setwd(workingDir)
}

#####
#####
# Anàlisi de correspondències simple:
#####
#####

conting <- flux.prim.sec[1,1:7,1:4]
library(ca)
corresp <- ca(conting)
corresp
summary(corresp)
names(corresp)

# Valors propis:

corresp.lambda <- rbind((corresp$sv)^2, 100*(corresp$sv)^2/(sum((corresp$sv)^2)))
rownames(corresp.lambda) <- c("Valor", "%")
colnames(corresp.lambda) <- 1:3
corresp.lambda

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(corresp.lambda, title="05acorresplambda")
  detach(package:Hmisc)
  setwd(workingDir)
}

# Gràfic:

if(a.pdf)
{
  setwd(resultsDir)
  pdf("05bcorresp.pdf", family="Helvetica", width=12, height=8)
}
par(las=1)
plot(corresp, main="Correspondència entre centres segons el flux d'alumnat",
      xlim=c(-1,0.6))
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

dades <- NULL

#####
#####

```

```

# Comparació de l'alumnat "interns"- "surten":
#####
#####

#####
# Subconjunt per a comparar els grups:
#####

load(file=file.path(dataDir,"dadesoriginals.RData"))
dades <- dadesoriginals

interns.surten <- subset(dades[,10:16], as.numeric(dades$Flux)<3)
interns.surten <- subset(interns.surten, complete.cases(interns.surten)==TRUE)
interns.surten$Flux <- as.factor(as.numeric(interns.surten$Flux))

library(Hmisc)
label(interns.surten$Flux) <- label(dades$Flux)
levels(interns.surten$Flux) <- levels(dades$Flux)[1:2]
detach(package:Hmisc)

#####
# Comparació descriptiva de cadascun dels 6 blocs entre "surten" i "interns":
#####

library(Hmisc)
sumby <- NULL
for(i in 2:7)
{
  sumby <- summaryBy(interns.surten[,i]~Flux, interns.surten, FUN=sumfun, na.rm=T)
  names(sumby)<-c('Alumnat','N', 'Mitjana','SD','Mediana','Mín.','Màx.')
  if(a.latex)
  {
    setwd(resultsDir)
    latex(sumby,title=paste("06resumintsurt",i))
    setwd(workingDir)
  }
  print(sumby)
}
detach(package:Hmisc)

#####
# Histogrames:
#####

interns.surten.int <- subset(interns.surten, Flux=="Prim -> Sec")
interns.surten.surt <- subset(interns.surten, Flux=="Prim -> Altres")

if(a.pdf)
{
  setwd(resultsDir)
  pdf("07histointsurten.pdf", family="Helvetica", width=12, height=12)
}
par(mfrow=c(3,4), oma=c(0,0,1,0), las=1)

```

```

library(Hmisc)
for (i in 2:7)
{
  hist(interns.surten.int[,i], main=interns.surten.int[1,1],
       xlab=label(dades[, (9+i)]),
       ylab="Frequència",
       nclass=round(sqrt(dim(interns.surten.int)[1]), 0), col=i+1)
  abline(v=mean(interns.surten.int)[i], lty=2, lwd=2)
  hist(interns.surten.surt[,i], main=interns.surten.surt[1,1],
       xlab=label(dades[, (9+i)]),
       ylab="Frequència",
       nclass=round(sqrt(dim(interns.surten.surt)[1]), 0), col=i+1)
  abline(v=mean(interns.surten.surt)[i], lty=2, lwd=2)
}
detach(package:Hmisc)
title("Prova de juny. Histogrames comparatius entre l'alumnat intern i el que surt",
      outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
# Boxplots:
#####

if (a.pdf)
{
  setwd(resultsDir)
  pdf("08boxplotintsurten.pdf", family="Helvetica", height=10)
}
par(mfrow=c(3,2), oma=c(0,0,1,0), las=1)
library(Hmisc)
for (i in 2:7)
boxplot(interns.surten[,i]~Flux, data=interns.surten, main=label(dades[, (9+i)]),
        col=i+1)
detach(package:Hmisc)
title("Prova de juny.
Diagrames de caixa comparatius entre l'alumnat intern i el que surt", outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
# Comparació per t-test:
#####

coef.juny <- summary(glm(interns.surten[,2]~Flux,
                        data=interns.surten))$coefficients[2,]
for (i in 3:7)

```

```

coef.juny <- rbind(coef.juny,
                  summary(glm(interns.surten[,i]~Flux,
                             data=interns.surten))$coefficients[2,])

coef.juny <- round(coef.juny,4)
rownames(coef.juny) <- label(interns.surten)[1:6]
colnames(coef.juny) <- c("Estimació", "Error estàndar", "t", "p-valor")

coef.juny

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(coef.juny, title="09coef_juny_int_surten")
  detach(package:Hmisc)
  setwd(workingDir)
}

dades <- NULL

#####
#####
# Comparació de l'alumnat "interns"- "entren":
#####
#####

#####
# Subconjunt per a comparar els grups:
#####

load(file=file.path(dataDir, "dadesoriginals.RData"))
dades <- dadesoriginals

interns.entren <- subset(dades[,c(10,22:27)], as.numeric(dades$Flux)!=2)
interns.entren <- subset(interns.entren, complete.cases(interns.entren)==TRUE)
interns.entren$Flux <- as.factor(as.numeric(interns.entren$Flux))

library(Hmisc)
label(interns.entren$Flux) <- label(dades$Flux)
levels(interns.entren$Flux) <- levels(dades$Flux)[c(1,3)]
detach(package:Hmisc)

#####
# Comparació descriptiva de cadascun dels 6 blocs entre "interns" i "entren":
#####

sumby <- NULL
for(i in 2:7)
{
  sumby <- summaryBy(interns.entren[,i]~Flux, interns.entren,
                    FUN=sumfun, na.rm=T)
  names(sumby)<-c('Alumnat', 'N', 'Mitjana', 'SD', 'Mediana', 'Mín.', 'Màx.')
}

```

```

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(sumbly,title=paste("10resumintentra",i))
  setwd(workingDir)
  detach(package:Hmisc)
}
print(sumbly)
}

#####
# Histogrames:
#####

interns.entren.int <- subset(interns.entren, Flux=="Prim -> Sec")
interns.entren.entra <- subset(interns.entren, Flux=="Altres -> Sec")

if(a.pdf)
{
  setwd(resultsDir)
  pdf("11histointentren.pdf", family="Helvetica", width=12, height=12)
}
par(mfrow=c(3,4), oma=c(0,0,1,0), las=1)
library(Hmisc)
for (i in 2:7)
{
  hist(interns.entren.int[,i], main=interns.entren.int[1,1],
        xlab=label(dades[, (20+i)]),
        ylab="Freqüència",
        nclass=round(sqrt(dim(interns.entren.int)[1]), 0), col=i+1)
  abline(v=mean(interns.entren.int)[i], lty=2, lwd=2)
  hist(interns.entren.entra[,i], main=interns.entren.entra[1,1],
        xlab=label(dades[, (20+i)]),
        ylab="Freqüència",
        nclass=round(sqrt(dim(interns.entren.entra)[1]), 0), col=i+1)
  abline(v=mean(interns.entren.entra)[i], lty=2, lwd=2)
}
detach(package:Hmisc)
title("Prova de setembre.
Histogrames comparatius entre l'alumnat intern i el que surt", outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
# Boxplots:
#####

if (a.pdf)
{

```

```

    setwd(resultsDir)
    pdf("12boxplotintentren.pdf", family="Helvetica", height=10)
  }
par(mfrow=c(3,2), oma=c(0,0,1,0), las=1)
library(Hmisc)
for (i in 2:7)
boxplot(interns.entren[,i]~Flux, data=interns.entren,
        main=label(dades[, (20+i)]), col=i+1)
detach(package:Hmisc)
title("Prova de setembre.
      Diagrames de caixa comparatius entre l'alumnat intern i el que surt",
      outer=T)
if (a.pdf)
  {
    dev.off()
    setwd(workingDir)
  }

#####
# Comparació per t-test:
#####

coef.setembre <- summary(glm(interns.entren[,2]~Flux,
                             data=interns.entren))$coefficients[2,]

for (i in 3:7)
coef.setembre <- rbind(coef.setembre,
                      summary(glm(interns.entren[,i]~Flux,
                                   data=interns.entren))$coefficients[2,])

coef.setembre <- round(coef.setembre,4)
rownames(coef.setembre) <- label(interns.entren)[1:6]
colnames(coef.setembre) <- c("Estimació", "Error estàndar", "t", "p-valor")

coef.setembre

if(a.latex)
  {
    library(Hmisc)
    setwd(resultsDir)
    latex(coef.setembre, title="13coef_setem_int_entren")
    detach(package:Hmisc)
    setwd(workingDir)
  }

dades <- NULL

#####
#####
# Anàlisi descriptiva (dadesnetes):
#####
#####

load(file=file.path(dataDir, "dadesnetes.RData"))

```

```

dades <- dadesnetes

#####
# Gràfics dels factors:
#####

if (a.pdf)
{
  setwd(resultsDir)
  pdf("14graffactors.pdf", family="Helvetica", height=15)
}
par(mfrow=c(3,2), oma=c(0,0,1,0), las=1)
diagrama.sectors(dades$Sexe)
diagrama.sectors(dades$Rep)
diagrama.sectors(dades$Cprim)
diagrama.sectors(dades$Csec)
diagrama.sectors(dades$Cat2)
diagrama.sectors(dades$NEE)
title(paste("Distribució dels factors de tipologia de l'alumnat (n = ",
            dim(dades)[1],")", sep=""), outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
# Gràfics de l'edat:
#####

  if (a.pdf)
  {
    setwd(resultsDir)
pdf("15agrafmesnaix.pdf", family="Helvetica", width=10, height=3)
  }
diagrama.barres(dades$MesNa, "aquamarine")
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

  if (a.pdf)
  {
    setwd(resultsDir)
pdf("15bgrafedat.pdf", family="Helvetica", width=8, height=3)
  }
par(mfrow=c(1,2), oma=c(0,0,1,0), las=1)
diagrama.barres(dades$Edat.ent, "orange")
hist(dades$Edat, main="Edat", xlab="Edat", ylab="Freqüència",
      nclass=round(sqrt(length(dades$Edat))), col="green")

```



```

    abline(v=mean(dades$Edat), lty=2, lwd=2)
title(paste("Distribució de l'edat de l'alumnat", sep=""), outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
# Descriptiva de l'edat:
#####

t(as.matrix(sumfun(Edat,r=1)[-1]))
#      m sd Med Min Max
#    12.1 0.4 12 11.5 13.5

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(t(as.matrix(sumfun(Edat,r=1)[-1])), title="16descripedat")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Descriptiva de covariables resposta:
#####

descrip.covar.resp <- rbind(sumfun(B1P,r=1)[-1], sumfun(B2P,r=1)[-1],
  sumfun(B3P,r=1)[-1], sumfun(B4P,r=1)[-1],
  sumfun(B5P,r=1)[-1], sumfun(B6P,r=1)[-1],
  sumfun(GP,r=1)[-1], sumfun(B1S,r=1)[-1],
  sumfun(B2S,r=1)[-1], sumfun(B3S,r=1)[-1],
  sumfun(B4S,r=1)[-1], sumfun(B5S,r=1)[-1],
  sumfun(B6S,r=1)[-1], sumfun(GS,r=1)[-1],
  sumfun(dif.B1,r=1)[-1], sumfun(dif.B2,r=1)[-1],
  sumfun(dif.B3,r=1)[-1], sumfun(dif.B4,r=1)[-1],
  sumfun(dif.B5,r=1)[-1], sumfun(dif.B6,r=1)[-1],
  sumfun(dif.G,r=1)[-1], sumfun(Aval1,r=1)[-1])
rownames(descrip.covar.resp) <- labels(dades)[[2]][c(10:15, 34, 21:26,
  35:40, 46:47, 32)]
descrip.covar.resp[,4] <- round(descrip.covar.resp[,4])
descrip.covar.resp

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(descrip.covar.resp, title="17descripcovarresp")
  detach(package:Hmisc)
  setwd(workingDir)
}

```

```

# Histogrames:

indica <- c(10:15, 34, 21:26, 35:40, 46:47, 32)
indica <- indica[c(rep(1+7*(0:2),7) + rep(0:6, each=3))]
ymax <- c(rep(50,9), rep(60,3), rep(50,3), rep(100,3), rep(50,3))

if (a.pdf)
{
  setwd(resultsDir)
  pdf("18histresp.pdf", family="Helvetica", height=16)
}
par(mfrow=c(7,3), oma=c(0,0,1,0), las=1, cex=0.6)
library(Hmisc)
for (i in 1:length(indica))
{
  hist(dades[,indica[i]], main="", xlab=label(dades)[[indica[i]]],
       ylab="Freqüència", nclass=round(sqrt(dim(dades)[1]), 0),
       ylim=c(0,ymax[i]), col=(i-0.1)%/%3+2)
  abline(v=mean(dades[,indica[i]]), lty=2, lwd=2)
}
detach(package:Hmisc)
title("Histogrames de les qualificacions segons bloc
i prova (juny, setembre i diferència)",
      outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

# Boxplots:

if (a.pdf)
{
  setwd(resultsDir)
  pdf("19boxplotresp.pdf", family="Helvetica", height=16)
}
par(mfrow=c(7,3), oma=c(0,0,1,0), las=1, cex=0.6)
library(Hmisc)
for (i in 1:length(indica))
boxplot(dades[,indica[i]], main="", xlab=label(dades)[[indica[i]]], ylab="",
        col=(i-0.1)%/%3+2)
detach(package:Hmisc)
title("Diagrames de caixa de les qualificacions segons bloc i prova
(juny, setembre i diferència)",
      outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

```

```
#####
# Histograma i boxplot per a "Aval1":
#####

if (a.pdf)
{
  setwd(resultsDir)
  pdf("20grafAval1.pdf", family="Helvetica", width=8, height=4)
}
library(Hmisc)
par(mfrow=c(1,2), oma=c(0,0,1,0), las=1)
hist(Aval1, main="", xlab=label(Aval1), ylab="Frequència",
      nclass=round(sqrt(length(Aval1))-2), col="pink")
abline(v=mean(Aval1), lty=2, lwd=2)
boxplot(Aval1, main="", xlab=label(Aval1), ylab="", col="pink")
detach(package:Hmisc)
title("Distribució de les qualificacions en la primera avaluació (1r d'ESO)",
      outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
#####
# Risc Relatiu dels 6 Blocs (Juny - Setembre):
#####
#####

#####
# Taules de contingències per als problemes com a factors:
#####

detach(dades)
library(Hmisc)

dades$P1J <- as.factor(dades$B61P)
label(dades$P1J) <- label(dades$B61P)
levels(dades$P1J) <- c("Incorrecte", "Correcte")

dades$P2J <- as.factor(dades$B62P)
label(dades$P2J) <- label(dades$B62P)
levels(dades$P2J) <- c("Incorrecte", "Correcte")
dades$P2J

dades$P3J <- as.factor(dades$B63P)
label(dades$P3J) <- label(dades$B63P)
levels(dades$P3J) <- c("Incorrecte", "Correcte")

dades$P4J <- as.factor(dades$B64P)
label(dades$P4J) <- label(dades$B64P)
levels(dades$P4J) <- c("Incorrecte", "Correcte")
```

```
dades$P5J <- as.factor(dades$B65P)
label(dades$P5J) <- label(dades$B65P)
levels(dades$P5J) <- c("Incorrecte", "Correcte")

dades$P1S <- as.factor(dades$B61S)
label(dades$P1S) <- label(dades$B61S)
levels(dades$P1S) <- c("Incorrecte", "Correcte")

dades$P2S <- as.factor(dades$B62S)
label(dades$P2S) <- label(dades$B62S)
levels(dades$P2S) <- c("Incorrecte", "Correcte")

dades$P3S <- as.factor(dades$B63S)
label(dades$P3S) <- label(dades$B63S)
levels(dades$P3S) <- c("Incorrecte", "Correcte")

dades$P4S <- as.factor(dades$B64S)
label(dades$P4S) <- label(dades$B64S)
levels(dades$P4S) <- c("Incorrecte", "Correcte")

dades$P5S <- as.factor(dades$B65S)
label(dades$P5S) <- label(dades$B65S)
levels(dades$P5S) <- c("Incorrecte", "Correcte")

detach(package:Hmisc)

# Taules de contingència per als problemes:

library(Epi)
Pr1 <- stat.table(list("Problema 1 (juny)"=P1J, "Problema 1 (setembre)"=P1S),
  list(count(), percent(P1S), percent(P1J)),
  data=dades, margins=TRUE)

Pr2 <- stat.table(list("Problema 2 (juny)"=P2J, "Problema 2 (setembre)"=P2S),
  list(count(), percent(P2S), percent(P2J)),
  data=dades, margins=TRUE)

Pr3 <- stat.table(list("Problema 3 (juny)"=P3J, "Problema 3 (setembre)"=P3S),
  list(count(), percent(P3S), percent(P3J)),
  data=dades, margins=TRUE)

Pr4 <- stat.table(list("Problema 4 (juny)"=P4J, "Problema 4 (setembre)"=P4S),
  list(count(), percent(P4S), percent(P4J)),
  data=dades, margins=TRUE)

Pr5 <- stat.table(list("Problema 5 (juny)"=P5J, "Problema 5 (setembre)"=P5S),
  list(count(), percent(P5S), percent(P5J)),
  data=dades, margins=TRUE)

detach(package:Epi)

if(a.latex)
{
```

```

library(Hmisc)
setwd(resultsDir)
latex(Pr1[1,,], title="21pr11")
latex(round(Pr1[2,,], 1), title="21pr12")
latex(round(Pr1[3,,], 1), title="21pr13")

latex(Pr2[1,,], title="21pr21")
latex(round(Pr2[2,,], 1), title="21pr22")
latex(round(Pr2[3,,], 1), title="21pr23")

latex(Pr3[1,,], title="21pr31")
latex(round(Pr3[2,,], 1), title="21pr32")
latex(round(Pr3[3,,], 1), title="21pr33")

latex(Pr4[1,,], title="21pr41")
latex(round(Pr4[2,,], 1), title="21pr42")
latex(round(Pr4[3,,], 1), title="21pr43")

latex(Pr5[1,,], title="21pr51")
latex(round(Pr5[2,,], 1), title="21pr52")
latex(round(Pr5[3,,], 1), title="21pr53")
detach(package:Hmisc)
setwd(workingDir)
}

# Estimació dels RR en els problemes:

library(epitools)
RR.p1 <- riskratio(Pr1[1,1:2,1:2], rev="neither")
RR.p2 <- riskratio(Pr2[1,1:2,1:2], rev="neither")
RR.p3 <- riskratio(Pr3[1,1:2,1:2], rev="neither")
RR.p4 <- riskratio(Pr4[1,1:2,1:2], rev="neither")
RR.p5 <- riskratio(Pr5[1,1:2,1:2], rev="neither")
detach(package:epitools)

RR.p <- round(rbind(RR.p1$measure[2,], RR.p2$measure[2,], RR.p3$measure[2,],
                    RR.p4$measure[2,], RR.p5$measure[2,]), 2)
rownames(RR.p) <- 1:5
colnames(RR.p) <- c("Puntual", "Inf. 95%", "Sup. 95%")

RR.p

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(RR.p, title="22RRprob")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Dicotomització de les notes dels Blocs:

```

```
#####
```

```
library(Hmisc)

dades$B11J <- as.factor(dades$B1P>=5)
label(dades$B11J) <- label(dades$B1P)
levels(dades$B11J) <- c("Suspès", "Aprovat")

dades$B12J <- as.factor(dades$B2P>=5)
label(dades$B12J) <- label(dades$B2P)
levels(dades$B12J) <- c("Suspès", "Aprovat")

dades$B13J <- as.factor(dades$B3P>=5)
label(dades$B13J) <- label(dades$B3P)
levels(dades$B13J) <- c("Suspès", "Aprovat")

dades$B14J <- as.factor(dades$B4P>=5)
label(dades$B14J) <- label(dades$B4P)
levels(dades$B14J) <- c("Suspès", "Aprovat")

dades$B15J <- as.factor(dades$B5P>=5)
label(dades$B15J) <- label(dades$B5P)
levels(dades$B15J) <- c("Suspès", "Aprovat")

dades$B16J <- as.factor(dades$B6P>=5)
label(dades$B16J) <- label(dades$B6P)
levels(dades$B16J) <- c("Suspès", "Aprovat")

dades$B11S <- as.factor(dades$B1S>=5)
label(dades$B11S) <- label(dades$B1S)
levels(dades$B11S) <- c("Suspès", "Aprovat")

dades$B12S <- as.factor(dades$B2S>=5)
label(dades$B12S) <- label(dades$B2S)
levels(dades$B12S) <- c("Suspès", "Aprovat")

dades$B13S <- as.factor(dades$B3S>=5)
label(dades$B13S) <- label(dades$B3S)
levels(dades$B13S) <- c("Suspès", "Aprovat")

dades$B14S <- as.factor(dades$B4S>=5)
label(dades$B14S) <- label(dades$B4S)
levels(dades$B14S) <- c("Suspès", "Aprovat")

dades$B15S <- as.factor(dades$B5S>=5)
label(dades$B15S) <- label(dades$B5S)
levels(dades$B15S) <- c("Suspès", "Aprovat")

dades$B16S <- as.factor(dades$B6S>=5)
label(dades$B16S) <- label(dades$B6S)
levels(dades$B16S) <- c("Suspès", "Aprovat")

detach(package:Hmisc)
```

```
# Taules de contingència per als Blocs dicotomitats:

library(Epi)
B11 <- stat.table(list("Bloc 1 (juny)"=B11J, "Bloc 1 (setembre)"=B11S),
                  list(count(), percent(B11S), percent(B11J)),
                  data=dades, margins=TRUE)

B12 <- stat.table(list("Bloc 2 (juny)"=B12J, "Bloc 2 (setembre)"=B12S),
                  list(count(), percent(B12S), percent(B12J)),
                  data=dades, margins=TRUE)

B13 <- stat.table(list("Bloc 3 (juny)"=B13J, "Bloc 3 (setembre)"=B13S),
                  list(count(), percent(B13S), percent(B13J)),
                  data=dades, margins=TRUE)

B14 <- stat.table(list("Bloc 4 (juny)"=B14J, "Bloc 4 (setembre)"=B14S),
                  list(count(), percent(B14S), percent(B14J)),
                  data=dades, margins=TRUE)

B15 <- stat.table(list("Bloc 5 (juny)"=B15J, "Bloc 5 (setembre)"=B15S),
                  list(count(), percent(B15S), percent(B15J)),
                  data=dades, margins=TRUE)

B16 <- stat.table(list("Bloc 6 (juny)"=B16J, "Bloc 6 (setembre)"=B16S),
                  list(count(), percent(B16S), percent(B16J)),
                  data=dades, margins=TRUE)

detach(package:Epi)

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(B11[1,,], title="21B111")
  latex(round(B11[2,,], 1), title="21B112")
  latex(round(B11[3,,], 1), title="21B113")

  latex(B12[1,,], title="21B121")
  latex(round(B12[2,,], 1), title="21B122")
  latex(round(B12[3,,], 1), title="21B123")

  latex(B13[1,,], title="21B131")
  latex(round(B13[2,,], 1), title="21B132")
  latex(round(B13[3,,], 1), title="21B133")

  latex(B14[1,,], title="21B141")
  latex(round(B14[2,,], 1), title="21B142")
  latex(round(B14[3,,], 1), title="21B143")

  latex(B15[1,,], title="21B151")
  latex(round(B15[2,,], 1), title="21B152")
  latex(round(B15[3,,], 1), title="21B153")
}
```

```

    latex(B16[1,,], title="21B161")
    latex(round(B16[2,,], 1), title="21B162")
    latex(round(B16[3,,], 1), title="21B163")
    detach(package:Hmisc)
    setwd(workingDir)
}

# Estimació dels RR en els Blocs dicotomitats:

library(epitools)
RR.b1 <- riskratio(B11[1,1:2,1:2], rev="neither")
RR.b2 <- riskratio(B12[1,1:2,1:2], rev="neither")
RR.b3 <- riskratio(B13[1,1:2,1:2], rev="neither")
RR.b4 <- riskratio(B14[1,1:2,1:2], rev="neither")
RR.b5 <- riskratio(B15[1,1:2,1:2], rev="neither")
RR.b6 <- riskratio(B16[1,1:2,1:2], rev="neither")
detach(package:epitools)

RR.b <- round(rbind(RR.b1$measure[2,], RR.b2$measure[2,], RR.b3$measure[2,],
                    RR.b4$measure[2,], RR.b5$measure[2,], RR.b6$measure[2,]), 2)
rownames(RR.b) <- 1:6
colnames(RR.b) <- c("Puntual", "Inf. 95%", "Sup. 95%")

RR.b

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(RR.b, title="22RRblocs")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Correlació entre Blocs (JUNY):
#####

cor.juny <- round(cor(dades[,10:15]), 3)
cor.juny
if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(cor.juny, title="23corjuny")
  detach(package:Hmisc)
  setwd(workingDir)
}

if (a.pdf)
{
  setwd(resultsDir)

```



```
pdf("24scatterjuny.pdf", family="Helvetica", height=7)
}
par(las=1)
plot(dades[,10:15], main="Juny")
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
# Transformacions Box-Cox (JUNY):
#####

attach(dades)
par(mfrow=c(2,3), las=1)
boxcox(B1P+0.01~1) # 1.5
boxcox(B2P+0.01~1) # 1
boxcox(B3P+0.01~1) # 0.75
boxcox(B4P+0.01~1) # 1
boxcox(B5P+0.01~1) # 1
boxcox(B6P+0.01~1) # 0
detach(dades)

#####
# Correlació entre Blocs (SETEMBRE):
#####

cor.setembre <- round(cor(dades[,21:26]), 3)
cor.setembre
if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(cor.setembre, title="23acorsetembre")
  detach(package:Hmisc)
  setwd(workingDir)
}

if (a.pdf)
{
  setwd(resultsDir)
  pdf("24ascattersetembre.pdf", family="Helvetica", height=7)
}
par(las=1)
plot(dades[,21:26], main="Setembre")
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
```

```

# Correlació entre Blocs (JUNY-SETEMBRE):
#####

cor.j.s <- c(round(cor(dades[,c(10,21)]), 3)[1,2],
round(cor(dades[,c(11,22)]), 3)[1,2],
round(cor(dades[,c(12,23)]), 3)[1,2],
round(cor(dades[,c(13,24)]), 3)[1,2],
round(cor(dades[,c(14,25)]), 3)[1,2],
round(cor(dades[,c(15,26)]), 3)[1,2])

cor.j.s # 0.541 0.677 0.720 0.805 0.683 0.638

dades <- NULL

#####
#####
# Model per a la prova del Juny (Np):
#####
#####

load(file=file.path(dataDir,"dadesnetesrep.RData"))
dades <- dadesnetesrep

library(geepack)

modNp0 <- geeglm(Np ~ Bloc-1 + Sexe + Edat + Rep + Cat2 + NEE + Cprim + Csec,
family = gaussian, data=dades, id = Id,
corstr = "exchangeable")

modNp1 <- geeglm(Np ~ Bloc-1 + Sexe + Edat4 + Rep + Cat2 + NEE + Cprim + Csec,
family = gaussian, data=dades, id = Id,
corstr = "exchangeable")

modNp2 <- geeglm(Np ~ Bloc-1 + Sexe + Edat3 + Rep + Cat2 + NEE + Cprim + Csec,
family = gaussian, data=dades, id = Id,
corstr = "exchangeable")

modNp3 <- geeglm(Np ~ Bloc-1 + Sexe + Edat2 + Rep + Cat2 + NEE + Cprim + Csec,
family = gaussian, data=dades, id = Id,
corstr = "exchangeable")

anova(modNp2,modNp3)

modNp <- modNp3
summary(modNp)

# Deixem l'edat com a factor Edat2 i tornem a començar:

modNp0 <- geeglm(Np ~ Bloc-1 + Sexe + Edat2 + Rep + Cat2 + NEE + Cprim + Csec,
family = gaussian, data=dades, id = Id, corstr = "exchangeable")
modNp <- modNp0
summary(modNp)

modNp1 <- geeglm(Np ~ Bloc-1 + Edat2 + Rep + Cat2 + NEE + Cprim + Csec,

```

```

        family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modNp,modNp1)
modNp <- modNp1
summary(modNp)

modNp2 <- geeglm(Np ~ Bloc-1 + Edat2 + Cat2 + NEE + Cprim + Csec,
                 family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modNp,modNp2)
modNp <- modNp2
summary(modNp)

modNp3 <- geeglm(Np ~ Bloc-1 + Edat2 + Cat2 + NEE + Cprim,
                 family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modNp,modNp3)
modNp <- modNp3
summary(modNp)

modNp4 <- geeglm(Np ~ Bloc-1 + Cat2 + NEE + Cprim,
                 family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modNp,modNp4)
modNp <- modNp4
summary(modNp)

#####
# Recategorització de Cprim (pooled):
#####

dades$CprimPooled <- as.factor(as.numeric(dades$Cprim) == 2)
library(Hmisc)
label(dades$CprimPooled) <- "Centre de primària"
levels(dades$CprimPooled) <- c("CEIPs 1,3,4,5,6", "CEIP 2")
detach(package:Hmisc)

modNp5 <- geeglm(Np ~ Bloc-1 + Cat2 + NEE + CprimPooled,
                 family = gaussian, data=dades, id = Id,
                 corstr = "exchangeable")
anova(modNp,modNp5)

#####
# Model ajustat:
#####

model.juny <- geeglm(Np ~ Bloc-1 + Cat2 + NEE + Cprim,
                    family = gaussian, data=dades, id = Id,
                    corstr = "exchangeable")
summary(model.juny)

#####
# Taula dels coeficients:
#####

coef.model.juny <- summary(model.juny)$coefficients
coef.model.juny[,1:3] <- round(coef.model.juny[,1:3], 2)

```

```

coef.model.juny[,4] <- round(coef.model.juny[,4], 4)
coef.model.juny

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(coef.model.juny, title="25taula_mod_juny")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Plots de diagnosi:
#####

if (a.pdf)
{
  setwd(resultsDir)
  pdf("26grafresjuny.pdf", family="Helvetica", width=6, height=6)
}
par(mfrow=c(2,2), oma=c(0,0,1,0), las=1, cex=0.7, cex.axis=0.6)
plot(model.juny, main="Prediccions")
hist(model.juny$residuals, freq=F,
      main="Histograma", xlab="Residu", ylab="Freqüència")
mr <- mean(model.juny$residuals)
sdr <- sd(model.juny$residuals)
curve(dnorm(x, mean=mr, sd=sdr), lwd=2, add=TRUE)
boxplot(model.juny$residuals, main="Diagrama de caixa", ylab="Residu")

qqnorm(model.juny$residuals, main="Q-Q plot de normalitat",
        xlab="Quantils teòrics", ylab="Quantils mostrals")
qqline(model.juny$residuals)
title("Residus del model per a la qualificació de la prova de juny", outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}

#####
# Prediccions:
#####

tractaments.juny <- expand.grid(Bloc=levels(dades$Bloc),
                               Cat2=levels(dades$Cat2),
                               NEE=levels(dades$NEE),
                               Cprim=levels(dades$Cprim))

prediccions.juny <- predict(model.juny, newdata=tractaments.juny,
                           se.fit=T, dispersion=model.juny$geese$gamma)

pred.juny <- cbind(tractaments.juny,

```

```

        Pred=prediccions.juny$fit,
        Lim.Inf=prediccions.juny$fit-1.96*prediccions.juny$se.fit,
        Lim.Sup=prediccions.juny$fit+1.96*prediccions.juny$se.fit)

pred.juny <- with(pred.juny, pred.juny[order(Bloc, Cat2, NEE, Cprim,
                                             decreasing=FALSE),])
pred.juny[,5:7] <- (pred.juny[,5:7] >=0)*round(pred.juny[,5:7], 1)
pred.juny[,1] <- rep(c(rep("",11), rep("Bloc",2), rep("",11)), 6)
pred.juny[,2] <- rep(c(rep("",5), "No", rep("",6), rep("",5), "Sí", rep("",6)), 6)
pred.juny[,3] <- rep(c(rep("",2), "No", rep("",5), "Sí", rep("",3)), 12)
rownames(pred.juny) <- NULL
pred.juny

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.juny[1:72,], title="27predjunyestandar1")
  latex(pred.juny[73:144,], title="27predjunyestandar2")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Taula de prediccions top 5 per blocs:
#####

pred.juny.top5 <- cbind(tractaments.juny,
                       Pred=prediccions.juny$fit,
                       Lim.Inf=prediccions.juny$fit-1.96*prediccions.juny$se.fit,
                       Lim.Sup=prediccions.juny$fit+1.96*prediccions.juny$se.fit)
pred.juny.top5 <- with(pred.juny.top5,
                       pred.juny.top5[order(Bloc, Pred, decreasing=FALSE),])
pred.juny.top5[,5:7] <- (pred.juny.top5[,5:7] >=0)*round(pred.juny.top5[,5:7], 1)
pred.juny.top5$index.top5 <-rep(c(rep(1,5), rep(0,14), rep(1,5)), 6)
pred.juny.top5 <- subset(pred.juny.top5, index.top5==1)[,1:7]
pred.juny.top5[,1] <- rep(c(rep("",4), rep("Bloc",2), rep("",4)), 6)
rownames(pred.juny.top5) <- NULL

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.juny.top5, title="27top5_juny")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
#####
### Diferència Juny-Setembre (Dif)
#####
#####

```

```

dades$Np5 <- dades$Np-5

modDif0 <- geeglm(Dif ~ Bloc-1 + Np5 + Sexe + Edat + MesNa +
                  Rep + Cat2 + NEE + Cprim + Csec,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
modDif <- modDif0
summary(modDif)

modDif1 <- geeglm(Dif ~ Bloc-1 + Np5 + Sexe + Edat3 + MesNa +
                  Rep + Cat2 + NEE + Cprim + Csec,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
summary(modDif1)

modDif2 <- geeglm(Dif ~ Bloc-1 + Np5 + Sexe + Edat2 + MesNa +
                  Rep + Cat2 + NEE + Cprim + Csec,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
summary(modDif2)
modDif <- modDif2

modDif3 <- geeglm(Dif ~ Bloc-1 + Np5 + Edat2 + MesNa + Rep + Cat2 + NEE + Cprim + Csec,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modDif,modDif3)
modDif <- modDif3
summary(modDif)

modDif4 <- geeglm(Dif ~ Bloc-1 + Np5 + Edat2 + MesNa + Cat2 + NEE + Cprim + Csec,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modDif,modDif4)
modDif <- modDif4
summary(modDif)

modDif5 <- geeglm(Dif ~ Bloc-1 + Np5 + MesNa + Cat2 + NEE + Cprim + Csec,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modDif,modDif5)
modDif <- modDif5
summary(modDif)

modDif6 <- geeglm(Dif ~ Bloc-1 + Np5 + MesNa + Cat2 + NEE + Cprim,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modDif,modDif6)
modDif <- modDif6
summary(modDif)

modDif7 <- geeglm(Dif ~ Bloc-1 + Np5 + Cat2 + NEE + Cprim,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modDif,modDif7)
modDif <- modDif7

modDif8 <- geeglm(Dif ~ Bloc-1 + Np5 + NEE + Cprim,
                  family = gaussian, data=dades, id = Id, corstr = "exchangeable")
anova(modDif,modDif8)
modDif <- modDif8

```

```
#####
# Model ajustat:
#####

model.dif <- geeglm(Dif ~ Bloc-1 + Np5 + NEE + Cprim,
                    family = gaussian, data=dades, id = Id, corstr = "exchangeable")
summary(model.dif)

#####
# Taula dels coeficients:
#####

coef.model.dif <- summary(model.dif)$coefficients
coef.model.dif[,1:3] <- round(coef.model.dif[,1:3], 2)
coef.model.dif[,4] <- round(coef.model.dif[,4], 4)
coef.model.dif

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(coef.model.dif, title="28taula_mod_dif")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Plots de diagnosi:
#####

if (a.pdf)
{
  setwd(resultsDir)
  pdf("29grafresdif.pdf", family="Helvetica", width=6, height=6)
}
par(mfrow=c(2,2), oma=c(0,0,1,0), las=1, cex=0.7, cex.axis=0.6)
plot(model.dif, main="Prediccions")
hist(model.dif$residuals, freq=F, main="Histograma", xlab="Residu",
      ylab="Freqüència")
mr <- mean(model.dif$residuals)
sdr <- sd(model.dif$residuals)
curve(dnorm(x, mean=mr, sd=sdr), lwd=2, add=TRUE)
boxplot(model.dif$residuals, main="Diagrama de caixa", ylab="Residu")
qqnorm(model.dif$residuals, main="Q-Q plot de normalitat", xlab="Quantils teòrics",
       ylab="Quantils mostrals")
qqline(model.dif$residuals)
title("Residus del model per a la diferència de qualificació (juny -> setembre)",
      outer=T)
if (a.pdf)
{
  dev.off()
  setwd(workingDir)
}
```

```

}

#####
# Prediccions:
#####

tractaments.dif <- expand.grid(Bloc=levels(dades$Bloc),
                              Np5=0,
                              NEE=levels(dades$NEE),
                              Cprim=levels(dades$Cprim))
prediccions.dif <- predict(model.dif, newdata=tractaments.dif,
                          se.fit=T, dispersion=model.dif$geese$gamma)
pred.dif <- cbind(tractaments.dif,
                 Pred=prediccions.dif$fit,
                 Lim.Inf=prediccions.dif$fit-1.96*prediccions.dif$se.fit,
                 Lim.Sup=prediccions.dif$fit+1.96*prediccions.dif$se.fit)
pred.dif[,2] <- NULL
pred.dif <- with(pred.dif, pred.dif[order(Bloc, NEE, Cprim, decreasing=FALSE),])

pred.dif.estandar <- pred.dif
pred.dif.estandar[,4:6] <- round(pred.dif.estandar[,4:6], 1)
rownames(pred.dif.estandar) <- NULL
pred.dif.estandar[,1] <- rep(c(rep("",5), rep("Bloc",2), rep("",5)), 6)
pred.dif.estandar[,2] <- rep(c(rep("",2), "No", rep("",5), "Sí", rep("",3)), 6)
pred.dif.estandar

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.dif.estandar, title="30preddifest")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Taula de prediccions top 3 per blocs:
#####

pred.dif.top3 <- with(pred.dif, pred.dif[order(Bloc, Pred, decreasing=FALSE),])
rownames(pred.dif.top3) <- NULL
pred.dif.top3$index.top3 <-rep(c(rep(1,3), rep(0,6), rep(1,3)), 6)
pred.dif.top3 <- subset(pred.dif.top3, index.top3==1)[,1:6]
pred.dif.top3[4:6] <- round(pred.dif.top3[4:6], 1)
pred.dif.top3[,1] <- rep(c(rep("",2), rep("Bloc",2), rep("",2)), 6)
pred.dif.top3

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.dif.top3, title="30top3_dif")
  detach(package:Hmisc)
}

```



```

    setwd(workingDir)
}

#####
#####
### Model per a la diferència Juny-Setembre (Perd)
#####
#####

dades$PerdNum <- as.numeric(dades$Perd) -1
summary(dades$PerdNum)
summary(dades$Perd)

modPerd0 <- geeglm(PerdNum ~ Bloc-1 + Np5 + Sexe + Edat + MesNa +
                    Rep + Cat2 + NEE + Cprim + Csec,
                    family = binomial, data=dades, id = Id, corstr = "exchangeable")
modPerd <- modPerd0
summary(modPerd)

modPerd1 <- geeglm(PerdNum ~ Bloc-1 + Np5 + Sexe + Edat2 + MesNa +
                    Rep + Cat2 + NEE + Cprim + Csec,
                    family = binomial, data=dades, id = Id, corstr = "exchangeable")
summary(modPerd1)
modPerd <- modPerd1

modPerd2 <- geeglm(PerdNum ~ Bloc-1 + Np5 + Sexe + Edat2 + MesNa +
                    Cat2 + NEE + Cprim + Csec,
                    family = binomial, data=dades, id = Id, corstr = "exchangeable")
anova(modPerd,modPerd2)
modPerd <- modPerd2
summary(modPerd)

modPerd3 <- geeglm(PerdNum ~ Bloc-1 + Np5 + Edat2 + MesNa + Cat2 + NEE + Cprim + Csec,
                    family = binomial, data=dades, id = Id, corstr = "exchangeable")
anova(modPerd,modPerd3)
modPerd <- modPerd3
summary(modPerd)

modPerd4 <- geeglm(PerdNum ~ Bloc-1 + Np5 + MesNa + Cat2 + NEE + Cprim + Csec,
                    family = binomial, data=dades, id = Id, corstr = "exchangeable")
anova(modPerd,modPerd4)
modPerd <- modPerd4
summary(modPerd)

modPerd5 <- geeglm(PerdNum ~ Bloc-1 + Np5 + MesNa + Cat2 + NEE + Cprim,
                    family = binomial, data=dades, id = Id, corstr = "exchangeable")
anova(modPerd,modPerd5)
modPerd <- modPerd5
summary(modPerd)

modPerd6 <- geeglm(PerdNum ~ Bloc-1 + Np5 + MesNa + NEE + Cprim,
                    family = binomial, data=dades, id = Id, corstr = "exchangeable")
anova(modPerd,modPerd6)

```

```

modPerd <- modPerd6
summary(modPerd)

modPerd7 <- geeglm(PerdNum ~ Bloc-1 + Np5 + NEE + Cprim,
                  family = binomial, data=dades, id = Id, corstr = "exchangeable")
anova(modPerd,modPerd7)
modPerd <- modPerd7

modPerd8 <- geeglm(PerdNum ~ Bloc-1 + Np5 + NEE,
                  family = binomial, data=dades, id = Id, corstr = "exchangeable")
anova(modPerd,modPerd8)

#####
# Model ajustat:
#####

model.perd <- geeglm(PerdNum ~ Bloc-1 + Np5 + NEE + Cprim,
                   family = binomial, data=dades, id = Id, corstr = "exchangeable")
summary(model.perd)

#####
# Taula dels coeficients:
#####

coef.model.perd <- summary(model.perd)$coefficients
coef.model.perd[,1:3] <- round(coef.model.perd[,1:3], 2)
coef.model.perd[,4] <- round(coef.model.perd[,4], 4)
coef.model.perd

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(coef.model.perd, title="31taula_mod_perd")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Prediccions:
#####

tractaments.perd <- expand.grid(Bloc=levels(dades$Bloc),
                              Np5=-4:5,
                              NEE=levels(dades$NEE),
                              Cprim=levels(dades$Cprim))
prediccions.perd <- predict(model.perd, newdata=tractaments.perd,
                          se.fit=T, dispersion=1)
pred.perd <- cbind(tractaments.perd,
                  Pred=100/(1+exp(-prediccions.perd$fit)),
                  Lim.Inf=100/(1+exp(-prediccions.perd$fit+
                                     1.96*prediccions.perd$se.fit)),
                  Lim.Sup=100/(1+exp(-prediccions.perd$fit-

```

```

1.96*prediccions.perd$se.fit)))
pred.perd <- with(pred.perd, pred.perd[order(Bloc, Pred, decreasing=FALSE),])
pred.perd

#####
# Taula de prediccions top 3 per blocs:
#####

pred.perd.top3 <- subset(pred.perd, pred.perd$Np5==0)
pred.perd.top3 <- pred.perd.top3[,-2]
pred.perd$index.top3 <- rep(c(rep(1,3), rep(0,6), rep(1,3)), 6)
pred.perd.top3 <- subset(pred.perd.top3, index.top3==1)[,1:6]
rownames(pred.perd.top3) <- NULL
pred.perd.top3[,1] <- rep(c(rep("",2), rep("Bloc",2), rep("",2)), 6)
pred.perd.top3[,4:6] <- round(pred.perd.top3[,4:6], 1)
pred.perd.top3

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.perd.top3, title="32top3_perd")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Prediccions estàndar:
#####

tractaments.perd.estandar <- expand.grid(Bloc=levels(dades$Bloc),
                                         Np5=0,
                                         NEE=levels(dades$NEE),
                                         Cprim=levels(dades$Cprim))
prediccions.perd.estandar <- predict(model.perd, newdata=tractaments.perd.estandar,
                                     se.fit=T, dispersion=1)
pred.perd.estandar <- cbind(tractaments.perd.estandar,
                           Pred=round(100/(1+exp(-prediccions.perd.estandar$fit)),1),
                           Lim.Inf=round(100/(1+exp(-prediccions.perd.estandar$fit +
                                                    1.96*prediccions.perd.estandar$se.fit)),1),
                           Lim.Sup=round(100/(1+exp(-prediccions.perd.estandar$fit -
                                                    1.96*prediccions.perd.estandar$se.fit)),1))
pred.perd.estandar <- with(pred.perd.estandar,
                           pred.perd.estandar[order(Bloc, NEE, Cprim, decreasing=FALSE),])
pred.perd.estandar[,2] <- NULL
rownames(pred.perd.estandar) <- NULL
pred.perd.estandar

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.perd.estandar, title="32pred_perd_estandar")
}

```

```

detach(package:Hmisc)
setwd(workingDir)
}

#####
# Capacitat predictiva del model (matriu de confusió):
#####

observat <- as.numeric(dades$Perd)-1
predit <- (1+sign(predict(model.perd, dispersion=1)))/2
mat.conf <- with(dades,table(observat,predit))
rownames(mat.conf) <- c("No", "Sí")
colnames(mat.conf) <- c("No", "Sí")

especificitat <- 100*mat.conf[1,1]/sum(mat.conf[1,])
sensibilitat <- 100*mat.conf[2,2]/sum(mat.conf[2,])
clasif.correctes <- 100*sum(diag(mat.conf))/sum(mat.conf)

mat.conf
especificitat # 72,2%
sensibilitat # 64,3%
clasif.correctes # 68,5%

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(mat.conf, title="33confusio_perd")
  detach(package:Hmisc)
  setwd(workingDir)
}

dades <- NULL

#####
#####
### Model per a la 1a avaluació desembre (Aval1)
#####
#####

load(file=file.path(dataDir,"dadesnetes.RData"))
dades <- dadesnetes

modAval0 <- lm(Aval1 ~ Sexe + Edat + MesNa + Rep + Cat2 + NEE + Cprim + Csec +
  B1P + B2P + B3P + B4P + B5P + B6P +
  B1S + B2S + B3S + B4S + B5S + B6S,
  data=dades)

summary(modAval0)

modAval1 <- lm(Aval1 ~ Sexe + Edat + Rep + Cat2 + NEE + Cprim + Csec +
  B1P + B2P + B3P + B4P + B5P + B6P +
  B1S + B2S + B3S + B4S + B5S + B6S,
  data=dades)

```

```

anova(modAval0,modAval1)
modAval <- modAval1
summary(modAval)

modAval2 <- lm(Aval1 ~ Sexe + Edat + Cat2 + NEE + Cprim + Csec +
               B1P + B2P + B3P + B4P + B5P + B6P +
               B1S + B2S + B3S + B4S + B5S + B6S,
               data=dades)
anova(modAval,modAval2)
modAval <- modAval2
summary(modAval)

modAval3 <- lm(Aval1 ~ Sexe + Cat2 + NEE + Cprim + Csec +
               B1P + B2P + B3P + B4P + B5P + B6P +
               B1S + B2S + B3S + B4S + B5S + B6S,
               data=dades)
anova(modAval,modAval3)
modAval <- modAval3
summary(modAval)

modAval4 <- lm(Aval1 ~ Sexe + Cat2 + NEE + Csec +
               B1P + B2P + B3P + B4P + B5P + B6P +
               B1S + B2S + B3S + B4S + B5S + B6S,
               data=dades)
anova(modAval,modAval4)
modAval <- modAval4
summary(modAval)

modAval5 <- lm(Aval1 ~ Sexe + Cat2 + NEE +
               B1P + B2P + B3P + B4P + B5P + B6P +
               B1S + B2S + B3S + B4S + B5S + B6S,
               data=dades)
anova(modAval,modAval5)
modAval <- modAval5
summary(modAval)

modAval6 <- lm(Aval1 ~ Cat2 + NEE +
               B1P + B2P + B3P + B4P + B5P + B6P +
               B1S + B2S + B3S + B4S + B5S + B6S,
               data=dades)
anova(modAval,modAval6)
modAval <- modAval6
summary(modAval)

modAval7 <- lm(Aval1 ~ Cat2 + NEE +
               B1P + B2P + B3P + B5P + B6P +
               B1S + B2S + B3S + B4S + B5S + B6S,
               data=dades)
anova(modAval,modAval7)
modAval <- modAval7
summary(modAval)

modAval8 <- lm(Aval1 ~ Cat2 + NEE +

```

```

        B1P + B2P + B3P + B5P +
        B1S + B2S + B3S + B4S + B5S + B6S,
        data=dades)
anova(modAval,modAval8)
modAval <- modAval8
summary(modAval)

modAval9 <- lm(Aval1 ~ Cat2 + NEE +
        B1P + B2P + B3P +
        B1S + B2S + B3S + B4S + B5S + B6S,
        data=dades)
anova(modAval,modAval9)
modAval <- modAval9
summary(modAval)

modAval10 <- lm(Aval1 ~ Cat2 + NEE +
        B1P + B2P +
        B1S + B2S + B3S + B4S + B5S + B6S,
        data=dades)
anova(modAval,modAval10)
modAval <- modAval10
summary(modAval)

modAval11 <- lm(Aval1 ~ Cat2 + NEE + B1P + B1S + B2S + B3S + B4S + B5S + B6S,
        data=dades)
anova(modAval,modAval11)
modAval <- modAval11
summary(modAval)

modAval12 <- lm(Aval1 ~ Cat2 + NEE + B1S + B2S + B3S + B4S + B5S + B6S, data=dades)
anova(modAval,modAval12)
modAval <- modAval12
summary(modAval)

modAval13 <- lm(Aval1 ~ Cat2 + NEE + B1S + B2S + B3S + B4S + B6S, data=dades)
anova(modAval,modAval13)
modAval <- modAval13
summary(modAval)

modAval14 <- lm(Aval1 ~ Cat2 + NEE + B1S + B2S + B3S + B6S, data=dades)
anova(modAval,modAval14)
modAval <- modAval14
summary(modAval)

modAval15 <- lm(Aval1 ~ Cat2 + NEE + B2S + B3S + B6S, data=dades)
anova(modAval,modAval15)
modAval <- modAval15
summary(modAval)

modAval16 <- lm(Aval1 ~ NEE + B2S + B3S + B6S, data=dades)
anova(modAval,modAval16)
modAval <- modAval16
summary(modAval)

```

```
#####
# Model ajustat:
#####

model.Aval1 <- lm(Aval1 ~ NEE + B2S + B3S + B6S, data=dades)
summary(model.Aval1)

#####
# Taula de coeficients del model:
#####

coef.model.aval1 <- summary(model.Aval1)$coefficients
coef.model.aval1[,1:3] <- round(coef.model.aval1[,1:3], 2)
coef.model.aval1[,4] <- round(coef.model.aval1[,4], 4)
coef.model.aval1

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(coef.model.aval1, title="34taula_mod_aval1")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Plots de diagnosi:
#####

analisi.residus.lm(model.Aval1, "a la 1a Avaluació", "35grafresaval1", a.pdf=TRUE)

#####
# Prediccions:
#####

tractaments.aval1 <- expand.grid(NEE=levels(dades$NEE),
                                B2S=seq(4,6,by=1),
                                B3S=seq(4,6,by=1),
                                B6S=seq(4,6,by=1))

prediccions.aval1 <- predict(model.Aval1, newdata=tractaments.aval1, se.fit=T)
pred.aval1 <- cbind(tractaments.aval1,
                   Pred=round(prediccions.aval1$fit,1),
                   Lim.Inf=round(prediccions.aval1$fit-
                                 1.96*prediccions.aval1$se.fit,1),
                   Lim.Sup=round(prediccions.aval1$fit+
                                 1.96*prediccions.aval1$se.fit,1))
pred.aval1 <- with(pred.aval1, pred.aval1[order(NEE, Pred, decreasing=FALSE),])
rownames(pred.aval1) <- NULL
pred.aval1[,1] <- c(rep("", 13), "No", rep("", 26), "Sí", rep("", 13))
pred.aval1

if(a.latex)
```

```

{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.aval1, title="36predaval1")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Predicció estàndar:
#####

tractaments.estandar <- expand.grid(NEE=levels(dades$NEE),
                                   B2S=5,
                                   B3S=5,
                                   B6S=5)
prediccions.aprova.estandar <- predict(model.Aval1, newdata=tractaments.estandar,
                                       se.fit=T)

pred.aval1.estandar <- cbind(tractaments.estandar,
                             Pred=round(prediccions.aprova.estandar$fit,1),
                             Lim.Inf=round(prediccions.aprova.estandar$fit -
                                           1.96*prediccions.aprova.estandar$se.fit,1),
                             Lim.Sup=round(prediccions.aprova.estandar$fit +
                                           1.96*prediccions.aprova.estandar$se.fit,1))
pred.aval1.estandar

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.aval1.estandar, title="36a_aval1_estandar")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
#####
### Model per a la 1a avaluació desembre (Aprova)
#####
#####

modAprova0 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + Cat2 + NEE + Cprim + Csec +
                  B1P + B2P + B3P + B4P + B5P + B6P +
                  B1S + B2S + B3S + B4S + B5S + B6S,
                  family = binomial, data=dades)

modAprova <- modAprova0
summary(modAprova)

modAprova1 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + Cat2 + NEE + Csec +
                  B1P + B2P + B3P + B4P + B5P + B6P +
                  B1S + B2S + B3S + B4S + B5S + B6S,
                  family = binomial, data=dades)

```



```

modcomp <- modAprova1
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova1
summary(modAprova)

modAprova2 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + Cat2 + NEE +
                 B1P + B2P + B3P + B4P + B5P + B6P +
                 B1S + B2S + B3S + B4S + B5S + B6S,
                 family = binomial, data=dades)

modcomp <- modAprova2
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova2
summary(modAprova)

modAprova3 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + NEE +
                 B1P + B2P + B3P + B4P + B5P + B6P +
                 B1S + B2S + B3S + B4S + B5S + B6S,
                 family = binomial, data=dades)

modcomp <- modAprova3
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova3
summary(modAprova)

modAprova4 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + NEE +
                 B1P + B3P + B4P + B5P + B6P +
                 B1S + B2S + B3S + B4S + B5S + B6S,
                 family = binomial, data=dades)

modcomp <- modAprova4
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova4
summary(modAprova)

modAprova5 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + NEE +
                 B1P + B3P + B4P + B6P +
                 B1S + B2S + B3S + B4S + B5S + B6S,
                 family = binomial, data=dades)

modcomp <- modAprova5
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova5
summary(modAprova)

modAprova6 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + NEE +
                 B1P + B3P + B6P +

```

```

                                B1S + B2S + B3S + B4S + B5S + B6S,
                                family = binomial, data=dades)

modcomp <- modAprova6
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df [2]
1-pchisq(x,df)
modAprova <- modAprova6
summary(modAprova)

modAprova7 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + NEE +
                  B1P + B3P + B6P +
                  B2S + B3S + B4S + B5S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova7
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df [2]
1-pchisq(x,df)
modAprova <- modAprova7
summary(modAprova)

modAprova8 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + NEE +
                  B1P + B3P + B6P +
                  B2S + B3S + B5S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova8
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df [2]
1-pchisq(x,df)
modAprova <- modAprova8
summary(modAprova)

modAprova9 <- glm(Aprova ~ Sexe + Edat + MesNa + Rep + NEE +
                  B1P + B3P + B6P +
                  B2S + B3S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova9
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df [2]
1-pchisq(x,df)
modAprova <- modAprova9
summary(modAprova)

dades$MesNaJuny <- as.factor(dades$MesNa == "Jun")
library(Hmisc)
label(dades$MesNaJuny) <- "Naixement en juny"
levels(dades$MesNaJuny) <- c("No", "Sí")
detach(package:Hmisc)

modAprova10 <- glm(Aprova ~ Sexe + Edat + MesNaJuny + Rep + NEE +
                  B1P + B3P + B6P +
                  B2S + B3S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova10

```

```

x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova10
summary(modAprova)

modAprova11 <- glm(Aprova ~ Sexe + Edat + MesNaJuny + Rep + NEE +
                  B1P + B3P + B2S + B3S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova11
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova11
summary(modAprova)

modAprova12 <- glm(Aprova ~ Sexe + Edat + MesNaJuny + NEE +
                  B1P + B3P + B2S + B3S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova12
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova12
summary(modAprova)

modAprova13 <- glm(Aprova ~ Sexe + MesNaJuny + NEE +
                  B1P + B3P + B2S + B3S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova13
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova13
summary(modAprova)

modAprova14 <- glm(Aprova ~ MesNaJuny + NEE +
                  B1P + B3P + B2S + B3S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova14
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)
modAprova <- modAprova14
summary(modAprova)

modAprova15 <- glm(Aprova ~ MesNaJuny + NEE +
                  B1P + B2S + B3S + B6S,
                  family = binomial, data=dades)

modcomp <- modAprova15
x <- anova(modcomp,modAprova)$Deviance[2]
df <- anova(modcomp,modAprova)$Df[2]
1-pchisq(x,df)

```

```

modAprova <- modAprova15
summary(modAprova)

#####
# Model ajustat:
#####

# Treiem MesNaJuny del model perquè no és interpretable
# (l'AIC augmenta poc respecte "step")

model.aprova <- glm(Aprova ~ NEE + B1P + B2S + B3S + B6S,
                    family = binomial, data=dades)
summary(model.aprova)

#####
# Taula de coeficients del model:
#####

coef.model.aprova <- summary(model.aprova)$coefficients
coef.model.aprova[,1:3] <- round(coef.model.aprova[,1:3], 2)
coef.model.aprova[,4] <- round(coef.model.aprova[,4], 4)
coef.model.aprova

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(coef.model.aprova, title="37taula_mod_aprova")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Prediccions:
#####

tractaments.aprova <- expand.grid(NEE=levels(dades$NEE),
                                B1P=3:7,
                                B2S=3:7,
                                B3S=3:7,
                                B6S=3:7)

prediccions.aprova <- predict(model.aprova, newdata=tractaments.aprova, se.fit=T)
pred.aprova <-
  cbind(tractaments.aprova,
        Pred=round(100/(1+exp(-prediccions.aprova$fit)),1),
        Lim.Inf=round(100/(1+exp(-prediccions.aprova$fit +
                                1.96*prediccions.aprova$se.fit)),1),
        Lim.Sup=round(100/(1+exp(-prediccions.aprova$fit -
                                1.96*prediccions.aprova$se.fit)),1))
pred.aprova <- with(pred.aprova, pred.aprova[order(NEE, B1P, B2S, B3S, B6S,
                                                  decreasing=FALSE),])

```

```

pred.aprova <- subset(pred.aprova, (B2S==B1P)&(B3S==B1P)&(B6S==B1P))
rownames(pred.aprova) <- NULL
pred.aprova[,1] <- c(rep("", 2), "No", rep("", 4), "Sí", rep("", 2))
pred.aprova

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.aprova, title="38predaprova")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Predicció estàndar:
#####

tractaments.estandar <- expand.grid(NEE=levels(dades$NEE),
                                   B1P=5, B2S=5, B3S=5, B6S=5)
prediccions.aprova.estandar <- predict(model.aprova, newdata=tractaments.estandar,
                                       se.fit=T)

pred.aprova.estandar <-
  cbind(tractaments.estandar,
        Pred=round(100/(1+exp(-prediccions.aprova.estandar$fit)),1),
        Lim.Inf=round(100/(1+exp(-prediccions.aprova.estandar$fit +
                                1.96*prediccions.aprova.estandar$se.fit)),1),
        Lim.Sup=round(100/(1+exp(-prediccions.aprova.estandar$fit -
                                1.96*prediccions.aprova.estandar$se.fit)),1))
pred.aprova.estandar

if(a.latex)
{
  library(Hmisc)
  setwd(resultsDir)
  latex(pred.aprova.estandar, title="38a_pred_estandar")
  detach(package:Hmisc)
  setwd(workingDir)
}

#####
# Capacitat predictiva del model (matriu de confusió):
#####

observat <- as.numeric(dades$Aprova)
predit <- (1+sign(predict(model.aprova)))/2
mat.conf <- with(dades,table(observat,predit))
rownames(mat.conf) <- c("No", "Sí")
colnames(mat.conf) <- c("No", "Sí")

especificitat <- 100*mat.conf[1,1]/sum(mat.conf[1,])
sensibilitat <- 100*mat.conf[2,2]/sum(mat.conf[2,])

```

```
clasif.correctes <- 100*sum(diag(mat.conf))/sum(mat.conf)
```

```
mat.conf  
especificitat # 78,7%  
sensibilitat # 89,7%  
clasif.correctes # 85,7%
```

```
if(a.latex)  
{  
  library(Hmisc)  
  setwd(resultsDir)  
  latex(mat.conf, title="39confusio_aprova")  
  detach(package:Hmisc)  
  setwd(workingDir)  
}
```

## Referències

- [1] Agresti, Alan. *Categorical Data Analysis*. Wiley, 2nd Edition, 2002.
- [2] Greenacre, M.J. *Correspondence Analysis in Practice*. Academic Press, London, 1993.
- [3] Nenadić, O., Greenacre, M. (2007). Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. *Journal of Statistical Software*, May 2007, **20**, Issue 3.
- [4] Halekoh, U., Højsgaard, S., Yan, J. (2006). The R Package geePack for Generalized Estimating Equations. *Journal of Statistical Software*, January 2006, **15**, Issue 2.
- [5] Härdle, W., Simar, L. *Greenacre, M.J, Correspondence Analysis in Practice*. Academic Press, London. *Applied Multivariate Statistical Analysis*. Springer, 2003.
- [6] Hosmer, D.W., S. Lemeshow. *Applied Logistic Regression*. Wiley, 1989.
- [7] Johnson, D.E. *Métodos Multivariados Aplicados al Análisis de Datos*. International Thomson Editores, 2000.
- [8] McCullagh, P., Nelder, J.A. *Generalized Linear Models*. Chapman and Hall, 1983.
- [9] Lee, Y., Nelder, J.A., Pawitan, Y. *Generalized Linear Models with Random Effects. Unified Analysis via H-likelihood*. Chapman and Hall, 2006.
- [10] Peña, Daniel. *Análisis de Datos Multivariantes*. McGraw-Hill, 2002.
- [11] Peña, Daniel. *Regresión y Diseño de Experimentos*. Alianza Editorial (Manuales de Ciencias Sociales), 2002.
- [12] Rao, C.R., Toutenburg, H. *Linear Models. Least Squares and Alternatives*. Springer, 1999.
- [13] Rencher, A.C. *Methods of Multivariate Analysis*. Wiley Series in Probability and Mathematical Statistics, 1995.