

Grau en Estadística

Títol: Exploració d'una base de dades d'exercicis realitzats per estudiants de Probabilitat i Estadística

Autor: Carla Sánchez Cabanes

Director: José Antonio González Alastrué

Departament: Dep. Estadística i Investigació Operativa, UPC

Convocatòria: Juny de 2020



RESUM

Es disposa d'informació recopilada dels últims vuit anys (2012-2019) de docència de l'assignatura de Probabilitat i Estadística (PE) del grau d'Informàtica que imparteix la Universitat Politècnica de Catalunya. Bàsicament tenim les qualificacions que han obtingut aquests alumnes en els exercicis i exàmens que han fet mentre cursaven l'assignatura. L'objectiu d'aquest treball és explorar aquestes dades, mitjançant tècniques d'anàlisi multivariant, per buscar patrons de comportament dels alumnes tant dintre d'un mateix curs com entre cursos, així com trobar relacions entre els diversos processos didàctics i el rendiment de l'alumne.

Paraules clau: Anàlisi de components principals, clústering jeràrquic, alumne, treball autònom, treball a l'aula, e-status.

ABSTRACT

Over the last eight years (2012-2019) we have collected information of the subject Probability and Statistics (PE) of the degree in Computer Science taught by the Polytechnic University of Catalonia. Basically, we have the grades of each student of the exercises and exams. The aim of this work is to explore this data, using multivariate analysis techniques, to look for patterns of the student's behavior within the courses and between years, as well as trying to find relationships between the various teaching processes and the performances of the students.

Keywords: Principal Component Analysis, hierarchical clustering, student, work in the classroom, work at home and e-status.

ÍNDEX

1 - INTRODUCCIÓ.....	2
1.1 Context	2
1.2 Objectius.....	3
2 - METODOLOGIA.....	3
2.1 Clústering jeràrquic	3
2.2 Anàlisi de components principals.....	4
3 - DEPURACIÓ BASE DE DADES	7
3.1 Base de dades original.....	7
3.1.1 Notes treball autònom	7
3.1.2 Notes treball a l'aula.....	8
3.2 Base de dades resultant	9
4 - ANÀLISI UNIVARIANT I BIVARIANT	11
5 - ANÀLISI I RESULTATS	17
5.1 Clústering jeràrquic	17
5.1.1 Aplicació algoritme.....	17
5.1.2 Profiling de les classes	18
5.2 Anàlisi components principals.....	22
Alumnes avaluació continuada.....	22
Alumnes avaluació única	26
6 - CONCLUSIONS	31
7 - BIBLIOGRAFIA.....	33
ANNEX 1	34
ANNEX 2	54
ANNEX 3	66

1 - INTRODUCCIÓ

1.1 Context

Es disposa d'informació recopilada dels últims vuit anys (2012-2019), és a dir de 16 quadrimestres, de docència de l'assignatura de Probabilitat i Estadística (PE) del grau d'Informàtica que imparteix la Universitat Politècnica de Catalunya.

Des de la implantació del Pla Bologna el 2010, els estudiants de PE han hagut de treballar activament amb exercicis i proves de seguiment realitzats a l'aula així com amb problemes fets mitjançant la plataforma en línia e-status.

L'e-status és una plataforma en línia interactiva a la qual l'alumne pot accedir amb el seu usuari, en la qual el professorat habilita problemes, amb més d'un apartat cadascun, perquè els alumnes els resolguin. Es proposen una sèrie de problemes per a cada bloc, els quals l'alumne pot fer tantes vegades com cregui convenient cadascun d'ells (canvien les dades), ja sigui per a seguir practicant els continguts acadèmics o per a pujar nota, ja que la nota referent al treball autònom de cada bloc és, aproximadament, la mitjana de les notes de totes les execucions realitzades dels problemes d'e-status del bloc en qüestió.

L'assignatura de PE s'estructura en 7 blocs (6 específics i 1 transversal d'aplicació estadística): Bloc 1 de Càlcul de Probabilitat, Bloc 2 de Variable Aleatòria, Bloc 3 de Models de v.a., Bloc 4 d'Inferència, Bloc 5 de Disseny d'experiments, Bloc 6 de Previsió, i Bloc 7 d'Aplicació. Cada bloc es desenvolupa quinzenalment al llarg del quadrimestre i de cadascun es fan exercicis presencials a classe (treball a l'aula), els quals acostumen a ser en grup, i problemes a resoldre individualment fora de l'horari lectiu mitjançant l'e-status (treball autònom). A més a més es realitzen dos exàmens parcials, el primer inclou el temari dels blocs 1, 2 i 3, i el segon el dels blocs 4, 5 i 6. També es realitza un treball final d'aplicació de les competències obtingudes durant el curs.

Els primers cinc cursos (primavera 2012 – tardor 2014) es realitzaven quatre exercicis presencials per bloc, els següents cursos els van reduir a dos per bloc. De problemes d'e-status es realitzen uns deu per bloc, el nombre va canviant en funció del curs i del bloc.

La nota final de l'avaluació continuada és definida per una ponderació de totes les qualificacions obtingudes en els exercicis i exàmens comentats anteriorment, per decisió del professorat de l'assignatura es dóna més pes als últims blocs que als primers, el qual es fa saber als alumnes, amb la finalitat que aquests segueixin sent constants durant tot el curs. En cas que no s'aprovi l'avaluació continuada, o no s'hagi volgut optar per aquesta via, està l'opció de presentar-se a un únic examen final amb tot el temari del curs.

Al llarg de l'estudi que és dura a terme s'obviarà els factors de ponderació que s'apliquen a les notes de cada bloc per a determinar la nota final de l'avaluació continuada, la nota final es considerarà la mitjana dels dos exàmens parcials, ja que s'entendrà que el rendiment real de l'alumne és la nota que treu als exàmens, sense ponderacions.

Les dades de què es disposa sobre l'assignatura durant aquests 16 quadrimestres són les qualificacions dels estudiants en els exercicis, problemes i exàmens comentats, així com el temps d'execució dels problemes d'e-status i el nombre d'execucions per problema.

1.2 Objectius

La intenció del present treball és explorar la informació recopilada que s'ha esmentat anteriorment. Concretant, els objectius que es pretenen abordar són els següents:

- Exposar la situació general de l'assignatura aquests últims anys (nombre d'alumnes que la cursen, nombre de repetidors, avaluació per la qual solen optar,...)
- Detectar diferents perfils d'alumnes.
- Tractar d'observar com és el comportament dels alumnes, referent a la constància i seguiment, durant el curs i entre cursos.
- Determinar si es poden establir relacions entre els diversos processos didàctics (bàsicament, treball a l'aula, i treball autònom) i el rendiment de l'alumne (nota als exàmens).

2 - METODOLOGIA

Donat el volum de les dades que es faran servir, s'utilitzaran eines orientades a la mineria de dades. S'aplicaran tècniques d'anàlisi multivariant, les quals s'ocupen d'analitzar simultàniament conjunts de dades de les quals s'han mesurat un gran nombre de variables. També es farà ús de models estadístics més tradicionals com són els models lineals. La implementació d'aquestes eines i models es farà mitjançant el software lliure R.

La teoria que es presenta en aquest apartat s'ha obtingut d'apunts de l'assignatura Anàlisi multivariant, impartida per la Karina Gibert, i d'un treball final de grau d'un alumne del grau de matemàtiques en el qual es fa ús de mètodes multivariants.

2.1 Clústering jeràrquic

A la mineria de dades l'agrupament jeràrquic és un mètode d'anàlisi de grups puntuals, el qual busca construir una jerarquia de grups mitjançant un algoritme que agrupa les dades basant-se en la distància entre individus, buscant crear grups en el que les dades que estan dintre d'un clúster siguin similars entre si.

El mètode d'agregació que es farà servir és el de Ward. Aquest és un procediment jeràrquic en el qual, en cada etapa, s'uneixen successivament els clústers (a cada individu se'l considera inicialment com a un clúster) de manera que es tingui la menor pèrdua d'inèrcia entre classes possible. Per altra banda, cada cop que es produeix una agregació, la inèrcia dins de

les classes augmenta. La distància que s'aplicarà és la de Gower, que es calcula mitjançant la següent fórmula:

$$s(i, i') = \frac{\sum_{k=1}^K w_k(i, i') s_k(i, i')}{\sum_{k=1}^K w_k(i, i')}$$

$$w_k(i, i') = \begin{cases} 0 & \text{si } (x_{ik} = \text{missing}) \text{ o } (x_{i'k} = \text{missing}) \\ 0 & \text{si } (X_k \text{ és binària}) \text{ i } (x_{ik} = \text{missing}) \text{ o } (x_{i'k} = \text{missing}) \\ & \text{i l'absència negativa de } X_k \text{ exclosa} \\ 1 & \text{altrament} \end{cases}$$

$$s_k(i, i') = \begin{cases} 1 - \frac{|x_{ik} - x_{i'k}|}{R_k} & \text{si } X_k \text{ és numèrica} \\ 1 & \text{si } (X_k \text{ és qualitativa}) \text{ i } (x_{ik} = x_{i'k}) \\ 0 & \text{si } (X_k \text{ és qualitativa}) \text{ i } (x_{ik} \neq x_{i'k}) \end{cases}$$

S'utilitzarà la distància de Gower, ja que és una de les distàncies que s'usen principalment quan la matriu de dades està formada tant per variables qualitatives com numèriques (variables mixtes).

Es pot representar gràficament l'agrupació dels individus en diferents classes feta per l'algoritme mitjançant el dendrograma, un diagrama en forma d'arbre que organitza les dades en subcategories que es van dividint en altres fins a arribar al nivell de detall desitjat, semblant-se a les branques d'un arbre. Cada branca correspon a un individu, els individus que s'assemblen els uns als altres es combinen en aquestes branques. L'altura de les branques, proporcionada a l'eix vertical, indica la similitud entre dos individus, com més alta sigui l'altura de la fusió, menys individus són semblants. L'alçada de tall al dendrograma controla el cúmul de classes obtingudes, el qual és decisió subjectiva.

2.2 Anàlisi de components principals

Una de les qüestions que aborda l'anàlisi de dades multivariants és la de la reducció de la dimensionalitat. Reduir el nombre de variables objecte d'anàlisi sense perdre gaire informació al llarg del procés. Aquesta reducció és d'interès tant per simplificar l'anàlisi com per poder-les representar gràficament amb més facilitat.

L'anàlisi de components principals (PCA, Principal Component Analysis) és una de les tècniques utilitzades per a reduir la dimensionalitat i disminuir la redundància de les dades. S'atribueix el seu descobriment a Harold Hotelling (1895 - 1973), que inspirat per Karl Pearson (1857 - 1936) i la seva primera aproximació al problema d'obtenir un indicador que resumís un conjunt de variables, va generalitzar el 1933, amb l'ajuda de la computació, la idea de components principals introduint l'anàlisi de correlacions canòniques i permetent així resumir simultàniament dos conjunts de variables.

El que fa el PCA és realitzar una transformació lineal sobre les dades multivariants per trobar de manera seqüencial variables no correlacionades que recullin la variabilitat més gran possible, aquestes són les anomenades components principals. Són un nou sistema de coordenades ortogonals sobre el qual les dades podran representar-se en dimensió inferior. Això, i el fet que sempre es farà respectant la màxima variabilitat possible, atorguen a l'anàlisi de components principals la possibilitat tant de ser aplicat com un mètode de classificació no supervisat (és a dir, sense informació a priori de les classes a les quals pertanyen les dades), com de poder interpretar les dades a partir d'aquestes noves variables que s'han obtingut com a combinacions lineals de les originals.

Tot i que el PCA s'aplica partint de la matriu de covariàncies, com que aquesta no és invariante per canvis d'escala, es realitzarà sobre la matriu de correlacions en cas que les variables no proveniguin de la mateixa naturalesa, és a dir, que no siguin de dimensions homogènies o que l'ordre de la magnitud de les variables aleatòries mesurades no sigui el mateix.

2.2.1 Obtenció de les components principals

D'acord amb la definició donada de components principals, la manera d'obtenir-ne la primera d'elles és buscant la combinació lineal de les variables originals que tingui màxima variància.

Per tal de facilitar els càlculs sense perdre informació sobre les dades, aquestes s'estandarditzen restant-li a cada una d'elles la seva mitjana. És a dir, en comptes de treballar amb la matriu de dades multivariants, es fa ús de la matriu de dades centrades.

Matriu de dades centrades:

Els valors que tindran els individus en aquesta primera component vindran representats per un vector. On recollirà els coeficients pertinents per a aquesta transformació.

Al centrar-se les dades, les variables ara tenen mitjana zero, el qual s'ha traslladat també a les noves variables.

La variància de, que es vol que sigui màxima, serà:

$$\text{var}(Z_1) = \frac{1}{n} Z_1^t Z_1 = \frac{1}{n} a_1^t \bar{X} \bar{X} a_1 = a_1^t \left(\frac{1}{n} \bar{X} \bar{X} \right) a_1 = a_1^t S a_1$$

essent S la matriu de covariàncies de les variables X_1, \dots, X_p :

$$S = \text{cov}(X_i, X_j) = \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

Per tal de donar solució a aquest sistema indeterminat, s'imposa, sense perdre generalitat, la condició $a_1^t a_1 = 1$. Així, per maximitzar l'expressió considerada i contemplar alhora la condició esmentada es fan servir els multiplicadors de Lagrange:

$$M = a_1^t S a_1 - \lambda(a_1^t a_1 - 1).$$

Derivant respecte de les components d' a_1 i igualant l'equació a zero s'obté:

$$\frac{\partial M}{\partial a_1} = 2Sa_1 - 2\lambda a_1 = 0$$

que té per solució: $Sa_1 = \lambda a_1$.

És a dir que a_1 és vector propi de valor propi λ de la matriu de covariàncies S . Així, recuperant la variància de Z_1 calculada es té:

$$\text{var}(Z_1) = a_1^t Sa_1 = a_1^t \lambda a_1 = \lambda a_1^t a_1 = \lambda.$$

El vector propi de valor propi més gran de la matriu de covariàncies, defineix els coeficients de cada una de les variables en la primera component principal.

S'ha mostrat quin és el procés que porta a obtenir quines són les components principals. Per trobar la segona component, Z_2 , s'ha de procedir de la mateixa manera afegint-li la condició que aquesta sigui no correlacionada amb Z_1 .

2.2.2 Representació gràfica: Biplot

L'eina que farem servir per representar gràficament els resultats obtinguts pel PCA és el Biplot. Aquest és un mètode general i potent que permet visualitzar alhora tant les observacions com les variables de la matriu de dades multivariants. El Biplot consisteix a aproximar la matriu X mitjançant la descomposició en valors singulars, per una de rang dos:

$$X \approx U_2 D_2^{\frac{1}{2}} V_2^t = (U_2 D_2^{\frac{1-c}{2}}) (D_2^{\frac{c}{2}} V_2^t) = FC$$

on U_2 és $n \times 2$, $D_2^{\frac{1}{2}}$ és diagonal d'ordre 2, V_2^t és $2 \times p$ i totes tres estan definides segons la definició donada en la descomposició singular.

Prenent $0 \leq c \leq 1$ s'obtenen diferents descomposicions de la matriu X en dues matrius, i per tant diferents Biplots, on: F representa les n files d' X en un espai de dos dimensions, i C representa en el mateix espai les columnes de la matriu. Els valors més utilitzats per aquest paràmetre són $c = 0, 0.5$ i 1 . Amb cada un d'ells es prioritza la fidelització en la representació de o bé les variables, o bé les observacions. A causa de com són les dades amb les quals treballarem, a nosaltres ens interessa estudiar-ne el cas $c = 1$. Representarem per tant les files d' X per la matriu U_2 i les columnes per $D_2^{\frac{1}{2}} V_2^t$. Per distingir ambdues representacions, les observacions es dibuixen com a punts i les variables com a vectors al pla.

3 - DEPURACIÓ BASE DE DADES

Inicialment les dades es troben ubicades en diferents documents d'Excel. Es disposava de setze documents amb les qualificacions dels exercicis i proves realitzades a l'aula, un per a cada quadrimestre, i un altre amb les qualificacions de l'e-status recopilades durant tots els cursos. Aquests s'han importat a l'R i s'ha fet la depuració de dades pertinent per a corregir errors de format i poder unificar les notes de tots els anys en una sola taula. Posteriorment, s'han afegit variables noves funció de les variables originals. Els scripts d'R d'aquest procés de depuració estan inclosos a l'Annex 1.

Per qüestions de confidencialitat les dades es van proporcionar anonimitzades, no es van rebre amb els noms dels alumnes sinó amb un identificador generat de manera aleatòria.

Cal deixar clar també, que aquesta assignatura s'imparteix dos cops a l'any, en un mateix any s'hauran impartit dos cursos, quan parlem de quadrimestre ens referim a un curs acadèmic. Tardor del 2015, per exemple, equival a primer quadrimestre del 2015, i això és un curs acadèmic. Primavera del 2015 serà el mateix que segon quadrimestre del 2015.

3.1 Base de dades original

La base de dades objecte d'anàlisi de la que es parteix està formada per les diverses taules que s'expliquen a continuació.

3.1.1 Notes treball autònom

La taula amb les dades corresponents als problemes d'e-status conté 136.343 registres i 10 variables. Cada registre fa referència a l'execució d'un problema, mitjançant la plataforma e-status, realitzada per un alumne. Aquests problemes es poden realitzar més d'un cop, l'enunciat és el mateix però canvien les dades, i de cada bloc hi ha més d'un problema, de manera que hi ha més d'un registre per alumne en aquesta taula.

El data frame que emmagatzema aquestes dades, *estatus*, conté les següents variables:

Taula 3.1.1. Descripció de les variables

Nom variable	Descripció	Tipus
login	Identificador de l'alumne.	Alfanumèrica
prob	Nom del problema.	Factor
bloque	Nom del bloc de l'assignatura.	Factor
nota	Qualificació del problema. Pren valor NA quan s'ha començat a fer el problema però no s'ha acabat tramitant.	Numèrica
grupo	Grup al que pertany l'alumne.	Factor

Init	Data i hora en què l'alumne comença a realitzar el problema.	Data
Fin	Data i hora en què l'alumne acaba de realitzar el problema.	Data
cuatri	Quadrimestre.	Factor
año	Any.	Factor
blc	Identificador del bloc.	Factor
tempsmin	Temps que es triga a realitzar el problema, en minuts.	Numèrica.

3.1.2 Notes treball a l'aula

Les notes dels exercicis fets a classe i els exàmens s'han dividit en tres taules, ja que als primers quadrimestres que s'estudiaran no es feien el mateix nombre d'exercicis de classe per bloc que als últims quadrimestres. En les següents taules cada registre serà referent a un alumne. Els alumnes repetidors apareixeran en més d'un registre, però aquests, complementats del curs (any i quadrimestre) els entendrem com a diferents individus.

Els NA en les variables de qualificacions fan referència a un no presentat.

Les tres taules en què s'han dividit aquestes dades són les següents:

Notes exàmens

Cada curs es fan dos exàmens parcials i un examen final. El data frame que emmagatzema aquestes dades, notes, conté 2.641 registres i 14 variables, que són les següents:

Taula 3.1.2.1. Descripció de les variables

Nom variable	Descripció	Tipus
login	Identificador de l'alumne.	Alfanumèrica
año	Any.	Factor
cuatri	Quadrimestre.	Factor
grupo	Grup al que pertany l'alumne.	Factor
Pb1	Nota exercici referent al bloc 1. Primer parcial.	Numèrica
Pb2	Nota exercici referent al bloc 2. Primer parcial.	Numèrica
Pb3	Nota exercici referent al bloc 3. Primer parcial.	Numèrica
Pb4	Nota exercici referent al bloc 4. Segon parcial.	Numèrica
Pb5	Nota exercici referent al bloc 5. Segon parcial.	Numèrica
Pb6	Nota exercici referent al bloc 6. Segon parcial.	Numèrica

Pb1EF	Nota exercici referent als blocs 1 i 2. Exàmen final.	Numèrica
Pb2EF	Nota exercici referent als blocs 3 i 4. Exàmen final.	Numèrica
Pb2EF	Nota exercici referent als blocs 5 i 6. Exàmen final.	Numèrica
Informe	Nota treball final. Bloc 7.	Numèrica

Notes dels exercicis de classe dels primers cinc cursos (primavera 2012 – primavera 2014)

Aquests cursos es feien quatre exercicis per bloc. La taula, emmagatzemada al data frame *bda*, conté 688 registres i 27 variables, que són les següents:

Taula 3.1.2.2. Descripció de les variables

Nom variable	Descripció	Tipus
login	Identificador de l'alumne.	Alfanumèrica
año	Any.	Factor
cuatri	Quadrimestre.	Factor
iSMT	Nota exercici 1 del bloc <i>i</i> . $i \in (1, \dots, 6)$.	Numèrica
iPEM	Nota exercici 2 del bloc <i>i</i> . $i \in (1, \dots, 6)$.	Numèrica
iCoop	Nota exercici 3 del bloc <i>i</i> . $i \in (1, \dots, 6)$.	Numèrica
iSMT2	Nota exercici 4 del bloc <i>i</i> . $i \in (1, \dots, 6)$.	Numèrica

Notes dels exercicis de classe dels últims onze cursos (tardor 2014 – tardor 2019)

Aquests cursos es feien dos exercicis per bloc. La taula, emmagatzemada al data frame *bdb*, conté 1.953 registres i 15 variables, que són les següents:

Taula 3.1.2.3. Descripció de les variables

Nom variable	Descripció	Tipus
login	Identificador de l'alumne.	Alfanumèrica
año	Any.	Factor
cuatri	Quadrimestre.	Factor
iPS1	Nota exercici 1 del bloc <i>i</i> . $i \in (1, \dots, 6)$.	Numèrica
iPS2	Nota exercici 2 del bloc <i>i</i> . $i \in (1, \dots, 6)$.	Numèrica

3.2 Base de dades resultant

Tota la informació recollida a la base de dades anterior s'ha tractat per unificar-la i facilitar l'anàlisi d'aquesta. A partir de les dades de les quals es disposa s'han creat altres variables a nivell de bloc així com variables per caracteritzar la situació acadèmica de l'alumne. La base

de dades amb la qual es treballarà, emmagatzemada al data frame *pe*, conté 2.641 registres i 98 variables. Les variables amb les quals s'ha treballat principalment són les següents:

Taula 3.1.2.5. Descripció de les variables

Nom variable	Descripció	Tipus
login	Identificador de l'alumne.	Alfanumèrica
año	Any.	Factor
cuatri	Quadrimestre.	Factor
conv	Nombre de convocatòries de l'alumne. L'alumne amb més d'una convocatòria és repetidor	Factor
Pb1	Nota exercici referent al bloc 1. Primer parcial.	Numèrica
Pb2	Nota exercici referent al bloc 2. Primer parcial.	Numèrica
Pb3	Nota exercici referent al bloc 3. Primer parcial.	Numèrica
Pb4	Nota exercici referent al bloc 4. Segon parcial.	Numèrica
Pb5	Nota exercici referent al bloc 5. Segon parcial.	Numèrica
Pb6	Nota exercici referent al bloc 6. Segon parcial.	Numèrica
Pb1EF	Nota exercici referent als blocs 1 i 2. Exàmen final.	Numèrica
Pb2EF	Nota exercici referent als blocs 3 i 4. Exàmen final.	Numèrica
Pb2EF	Nota exercici referent als blocs 5 i 6. Exàmen final.	Numèrica
Informe	Nota treball final. Bloc 7.	Numèrica
MCBi	Nota mitjana dels exercicis de classe del bloc i . $i \in (1, \dots, 6)$.	Numèrica
curso	Curs. Concatenació de l'any i el quadrimestre.	Factor
CP1	Mitjana dels exercicis de classe dels blocs del primer parcial (1, 2, i 3).	Numèrica
CP2	Mitjana dels exercicis de classe dels blocs del segon parcial (4, 5, i 6).	Numèrica
np1	Nota primer parcial.	Numèrica
np2	Nota segon parcial.	Numèrica
perbi	Tant per u d'exercicis de classe fets del bloc i . $i \in (1, \dots, 6)$.	Numèrica
peri	Tant per u de problemes d'e-status fets dels bloc i . $i \in (1, \dots, 6)$.	Numèrica
n_exc_acabi	Número d'execucions acabades d'e-status de problemes del bloc i . $i \in (1, \dots, 6)$.	Numèrica
nota_mitji	Nota mitjana de tots els problemes d'e-status fets del bloc i . $i \in (1, \dots, 6)$.	Numèrica

t_mitji	Temps mig d'execució dels problemes d'estatus del bloc <i>i</i> . $i \in (1, \dots, 6)$.	Numèrica
cursfact	Identificador del curs, pren valors d'1 a 16.	Ordinal
nota_avcont	Nota final avaluació continuada.	Numèrica
nota_avunica	Nota examen final.	Numèrica
avuni	1: Avaluació única 0: Avaluació continuada.	Factor
numexacab	Nombre d'execucions de problemes d'e-status fets durant el curs.	Numèrica
notaestmitj	Nota mitjana d'e-status de tots els problemes fets durant tel curs.	Numèrica
tempsestmitj	Temps mig d'execució de tots els problemes fets durant el curs	Numèrica
mnc	Nota mitjana de tots els exercicis de classe fets durant el curs.	Numèrica
per	Suma del tant per u de problemes d'e-status fets de cada bloc. Domini entre 1 i 6.	Numèrica
perb	Suma del tant per u d'exercicis de classe fets de cada bloc. Domini entre 1 i 6.	Numèrica
apr_cont	1: Avaluació continuada aprovada 0: Avaluació continuada suspesa.	Factor
apr_uni	1: Avaluació única aprovada 0: Avaluació única suspesa.	Factor

El càlcul de les notes dels parcials és el següent:

$$\text{Nota primer parcial} = (Pb1 \times 10 + Pb2 \times 11 + Pb3 \times 12)/33$$

$$\text{Nota primer parcial} = (Pb4 \times 13 + Pb5 \times 14 + Pb6 \times 15)/42$$

La distinció entre avaluació continuada i avaluació única és definida per si els alumnes s'han presentat o no als parcials. Si no s'han presentat als dos exàmens parcials, s'entendrà que aniran a l'examen final i es dirà que fan avaluació única.

4 - ANÀLISI UNIVARIANT I BIVARIANT

S'ha fet una descriptiva bàsica per posar en context la situació de l'assignatura en aquests últims setze quadrimestres. Els scripts d'R amb la implementació d'aquestes estadístiques estan inclosos a l'Annex 2.

S'estan analitzant un total de 2.641 alumnes, en la següent taula de freqüències veiem com es distribueixen aquests per curs acadèmic.

Taula 4.1. Taula de freqüència dels cursos

	2012	2013	2014	2015	2016	2017	2018	2019	Total
1Q (tardor)	168	197	180	211	254	232	268	284	1794
2Q (prim.)	123	96	104	116	109	98	110	91	847
Total	291	293	284	327	363	330	378	375	2641

Els alumnes poden seguir l'avaluació continuada o optar per presentar-se a l'examen final, és a dir, avaluació única, a la següent taula veiem com estan distribuïts.

Taula 4.2. Taula de freqüència d'av. continuada/única

	Prim. 12	Tard. 12	Prim. 13	Tard. 13	Prim. 14	Tard. 14	Prim. 15	Tard. 15
Av. cont.	103	141	76	158	97	153	100	175
Av. única	20	27	20	39	7	27	16	36
Total	123	168	96	197	104	180	116	211

	Prim. 16	Tard. 16	Prim. 17	Tard. 17	Prim. 18	Tard. 18	Prim. 19	Tard. 19	Total
Av. cont.	103	141	76	158	97	153	100	175	2265
Av. única	20	27	20	39	7	27	16	36	376
Total	123	168	96	197	104	180	116	211	2641

Els alumnes que opten per l'avaluació continuada també poden presentar-se a l'examen final o també es pot donar els cas de què alumnes que no han seguit l'avaluació continuada tampoc es presentin al final, és a dir, alumnes que no es presenten als exàmens i que hauran de repetir. A la taula següent es mostren els percentatges d'aquestes dues casuístiques.

Taula 4.3. Taula d'alumnes que es presenten a l'examen final

	Es presenten al examen final %
Av. continuada	13,69 (310/2265)
Av. única	57,98 (218/376)

D'aquest 13,69% d'alumnes que fan l'avaluació continuada però també s'acaben presentant a l'examen final, el 84% es presenta perquè havia suspès i el 16% per intentar treure una nota millor que la que havien tret als parcials.

A continuació, la taula següent recull les notes mitjanes de l'assignatura de cada curs acadèmic, tant dels alumnes que han fet avaluació continuada (fan els exàmens parcials) com dels que han fet avaluació única (fan l'examen final).

Taula 4.4. Taula de notes finals

	Prim. 12	Tard. 12	Prim. 13	Tard. 13	Prim. 14	Tard. 14	Prim. 15	Tard. 15
Av. cont.	5,84	5,9	5,7	5,94	5,25	5,9	6,19	6,35
Av. única	1,4	2,01	2,06	3,58	0,59	2,61	1,41	3,35

	Prim. 16	Tard. 16	Prim. 17	Tard. 17	Prim. 18	Tard. 18	Prim. 19	Tard. 19
Av. cont.	5,7	5,62	5,4	5,88	5,42	5,68	4,73	6,66
Av. única	2,55	2,81	2,63	2,24	1,87	1,91	1,72	1,91

A la taula següent es mostren els percentatges d'alumnes repetidors segons els número de vegades que repeteixen.

Taula 4.5. Taula d'alumnes repetidors

	Repeteixen 1 vegades	Repeteixen 2 vegades	Repeteixen 3 vegades	Repeteixen 4 vegades	Total
Alumnes	78,35	18,04	2,6 (5/194)	1,03 (2/194)	7,35%
%	(152/194)	(35/194)			(194/2.641)

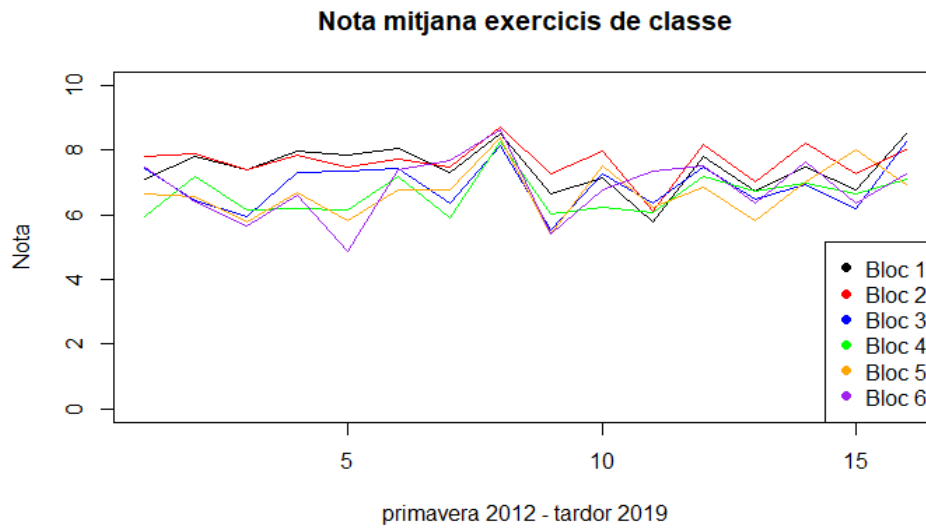
A la següent taula es té el percentatge d'alumnes repetidors (poden haver repetit un o més cops) a cada curs acadèmic.

Taula 4.6. Taula de percentatge d'alumnes repetidors

	2012	2013	2014	2015	2016	2017	2018	2019
1Q (tardor) %	17,86	33,33	9,95	21,10	8,21	26,37	22,84	26,92
2Q (prim.) %	13,01	15,00	23,28	7,33	17,27	16,67	30,21	24,17

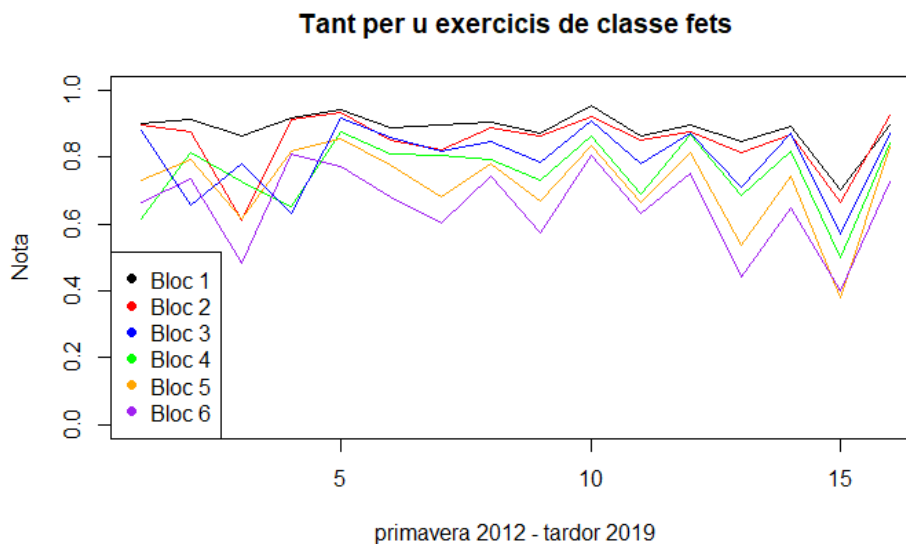
El següent gràfic mostra les notes mitjanes dels exercicis de classe de cada bloc per cada curs. Hi ha variació entre cursos i també entre blocs (dins del curs). En general als primers blocs es treuen més bones notes que als últims, aquesta tendència però, no és tan clara en els últims anys. Els pics de les notes més altes són quadrimestres de tardor (primer quadrimestre).

Figura 4.1. Notes mitjances dels exercicis de classe per bloc



El següent gràfic mostra les mitjanes dels tants per u d'exercicis de classe fets de cada bloc per cada curs. Per a aquesta variable també hi ha variació entre cursos i entre blocs. A partir del bloc 2, el percentatge d' exercicis que es fan disminueix considerablement cada per a cada bloc següent.

Figura 4.2. Tant per u exercicis de classe fets per bloc



La següent taula recull el nombre de problemes d'e-status que es proposen per a cada bloc. Recordem que l'alumne pot fer tantes execucions d'un mateix problema com cregui convenient.

Taula 4.5. Taula problemes per blocs

Bloc 1	Bloc 2	Bloc 3	Bloc 4	Bloc 5	Bloc 6
12	13	15	17	11	12

La següent és una taula de freqüències de les execucions acabades dels problemes d'e-status que s'han fet de cada bloc, així com la variació d'un bloc per l'altre.

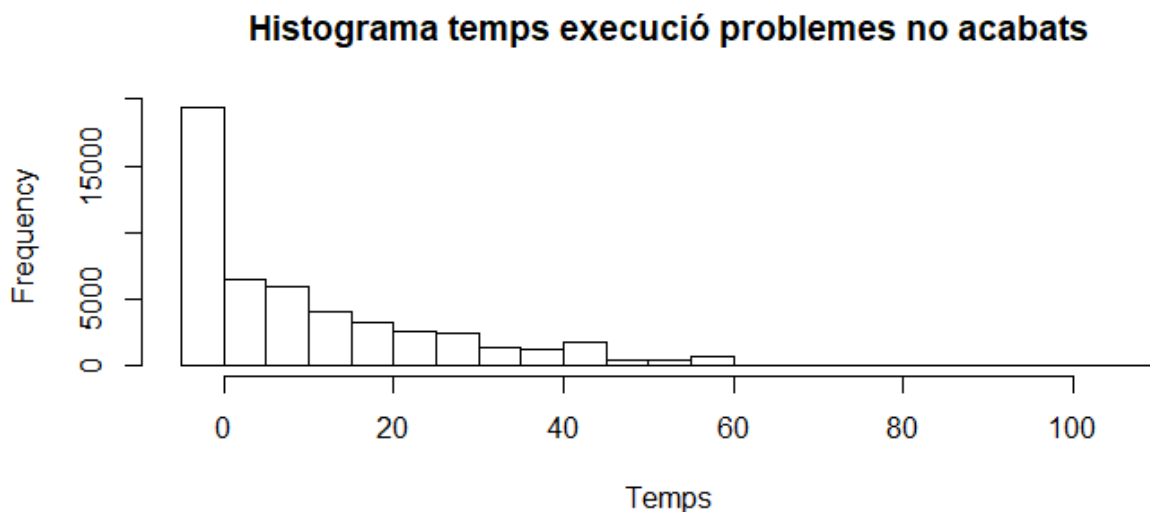
Taula 4.5. Taula execucions finalitzades

	Bloc 1	Bloc 2	Bloc 3	Bloc 4	Bloc 5	Bloc 6	Total
n	35.536	25.422	27.035	21.093	14.031	12.460	135.577
%	26,21	18,75	19,94	15,56	10,35	9,19	
variació %		-28,46	6,34	-21,98	-33,48	-11,2	

Quan diem que una execució ha estat acabada ens referim al fet que l'alumne l'ha acabat tramitant, si s'ha iniciat un problema però aquest no es tramita no tindrà qualificació. Es pot donar el cas de què es comenci un problema i que es finalitzi a l'instant, sense tramitar-lo, o que es comenci i estigui una estona obert però que al final no es tramiti (en aquest cas no tenim la informació referent a si l'alumne ha contestat algun apartat del problema i ha estat interactuant, o si no havia fet res).

D'execucions començades i no acabades hi ha un 36,46% (49.716/136.343), el següent histograma mostra la distribució del temps que ha estat obert el problema, en minuts, d'aquestes execucions no finalitzades.

Figura 4.3. Histograma temps execució

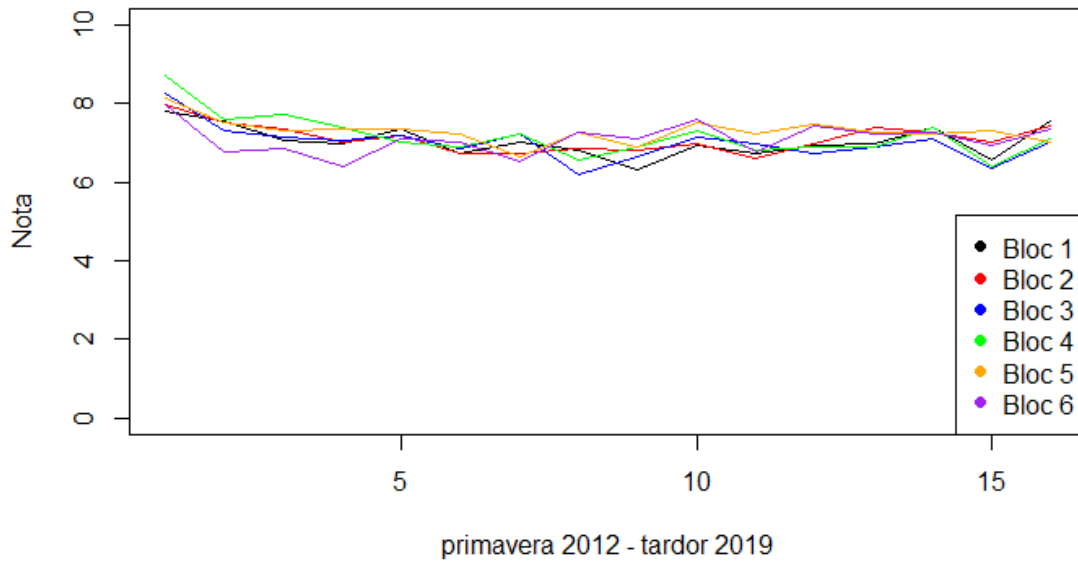


El 38,28% (19.006/49.716) tenen temps zero, és a dir, són els casos en què s'obre el problema i automàticament es tanca.

El següent gràfic mostra la nota mitjana dels problemes d'e-status per bloc de cada curs. Pràcticament no hi ha variació entre cursos ni entre blocs.

Figura 4.4. Nota mitjana prob. e-status

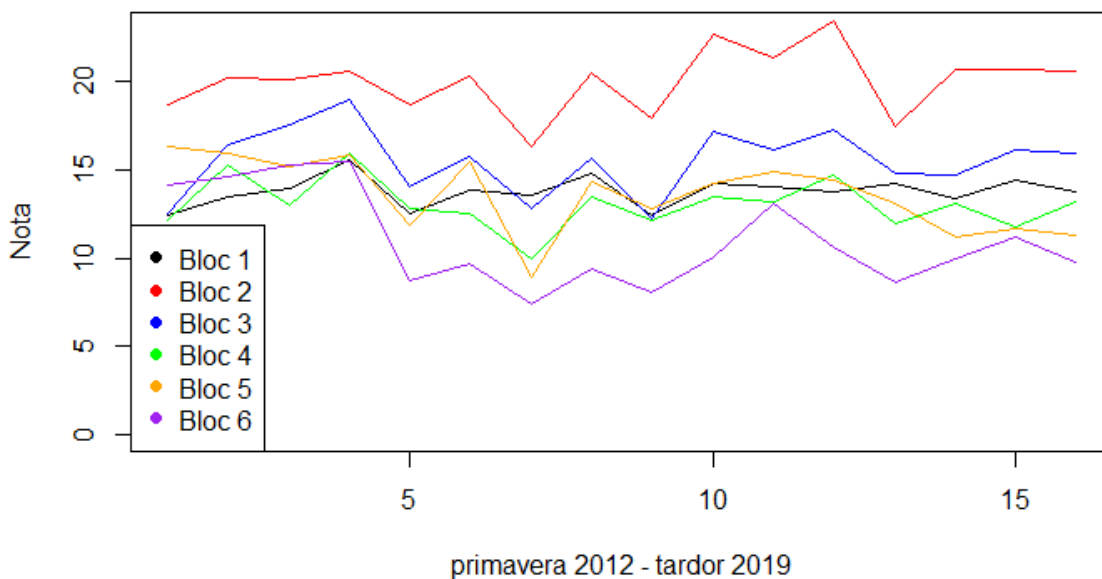
Nota mitjana problemes e-status



Al següent gràfic estan representats els temps mitjos d'execució per problema d'e-status per bloc i per cada curs. Hi ha molta variació entre blocs.

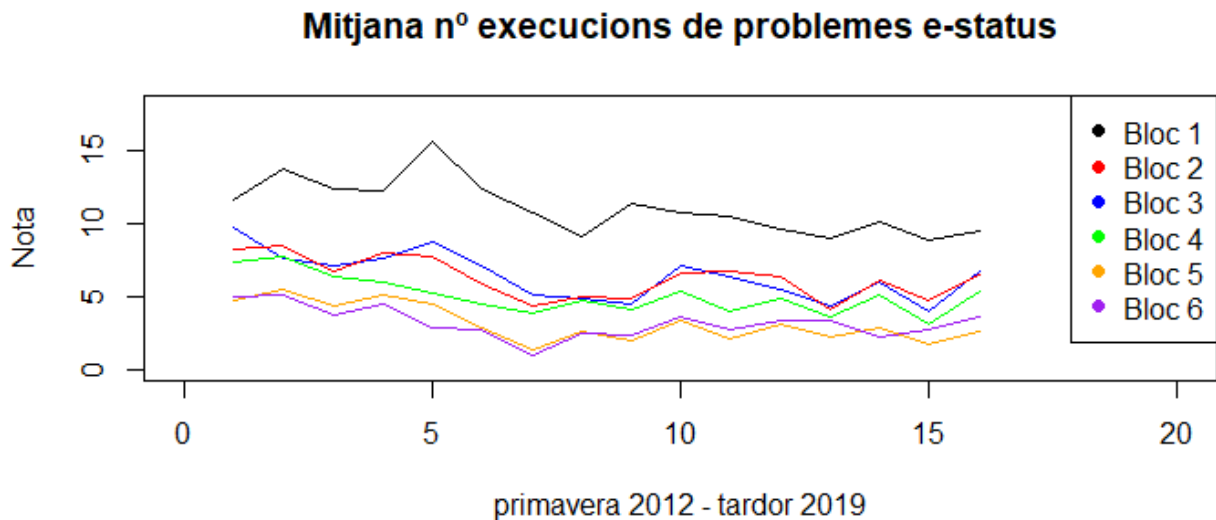
Figura 4.5. Temps mig d'execució problemes e-status.

Temps mig d'execució problemes e-status



Per últim, el gràfic següent mostra la mitjana per bloc, i de cada curs, del nombre d'execucions de problemes d'e-status fets. Entre cursos no sembla gaire diferent el patró, però entre blocs hi ha molta variació, del primer bloc es fan moltes execucions, en el segon aquestes es redueixen considerablement i en els següents blocs segueixen disminuint.

Figura 4.6. Mitjana execucions de prob. e-status



5 - ANÀLISI I RESULTATS

En aquest apartat s'exposen els resultats obtinguts a l'aplicar a les dades que ens ocupen les tècniques vistes a l'apartat 2. Els *scripts* d'R amb la implementació d'aquest anàlisi estan inclosos a l'Annex 2.

5.1 Clústering jeràrquic

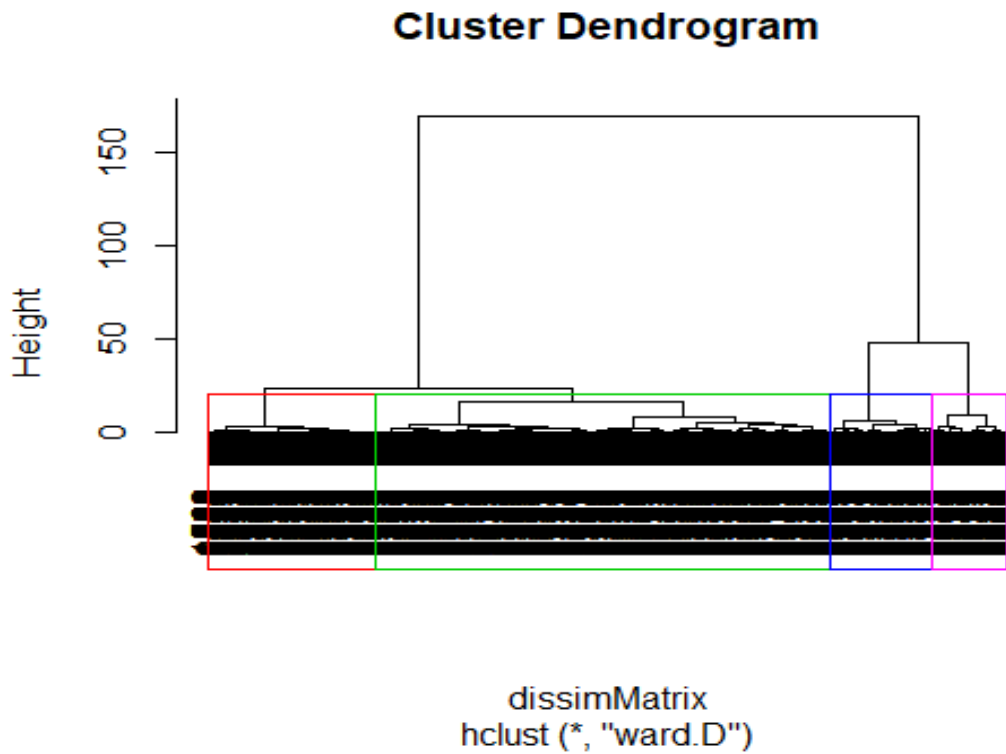
5.1.1 Aplicació algoritme

Es pretenia classificar d'alguna manera els alumnes segons patrons de comportament acadèmic similars, determinar quins perfils d'alumnes es poden diferenciar pel que fa al treball realitzat i notes obtingudes durant el curs. Per això s'ha realitzat un anàlisi de clústering jeràrquic a les dades, la població considerada ha estat la totalitat dels alumnes, és a dir els 2.641 estudiants dels últims setze cursos de l'assignatura.

L'algoritme s'ha aplicat a una matriu de dades amb les variables referents a les notes mitjanes dels exercicis de classe de cada bloc, les notes mitjanes dels problemes d'e-status de cada bloc, els percentatges d'exercicis realitzats de classe, i d'estatus, de cada bloc, els temps mitjans d'execució dels problemes d'e-status de cada bloc, el nombre d'execucions dels problemes d'estatus de cada bloc i la nota del treball final de l'assignatura.

El dendrograma de la matriu de dades obtingut després d'haver realitzat el clstering jeràrquic, amb el mètode d'agregació de Ward i la distància de Gower és el següent:

Figura 5.1. Dendrograma



S'ha decidit tallar en quatre classes, la taula de freqüències del nombre d'alumnes que pertanyen a cadascuna d'elles és la següent:

Taula 5.1. Classes resultants del clúster

Classe 1	Classe 2	Classe 3	Classe 4	Total
1504	554	338	245	2641
56,95 %	20,98 %	12,80 %	9,28 %	100%

Pràcticament el 60% pertanyen a una mateixa classe, un 20% a una altra i el 20% restant a les altres dues classes.

5.1.2 Profiling de les classes

Un cop distingits els alumnes en quatre classes, es perfilarà les característiques pròpies de cadascuna. Per això, s'han utilitzat principalment una sèrie d'eines gràfiques que ajuden a identificar els comportaments distintius dels alumnes de cada classe.

Figura 5.1.2.1 Boxplot temps

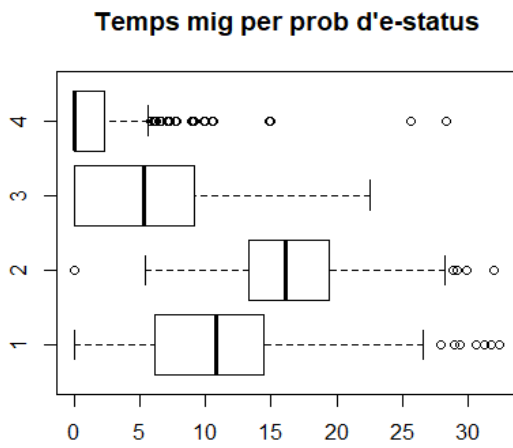


Figura 5.1.2.2. Boxplot execucions

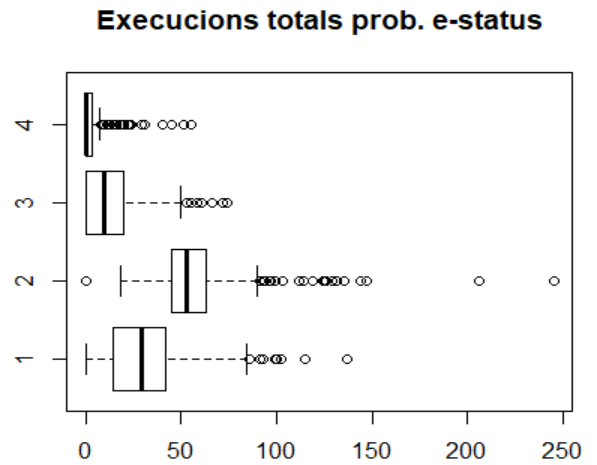


Figura 5.1.2.3. Boxplot perc. prob e-status

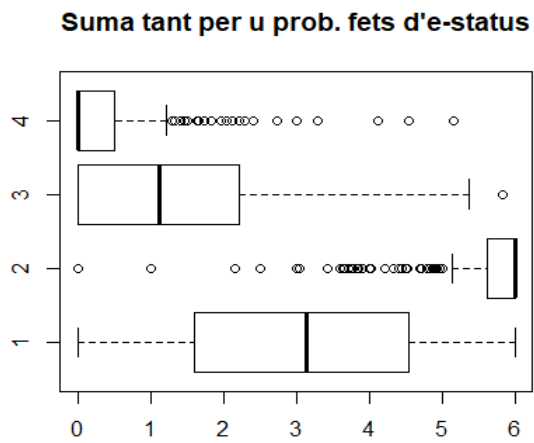


Figura 5.1.2.4. Boxplot nota e-status

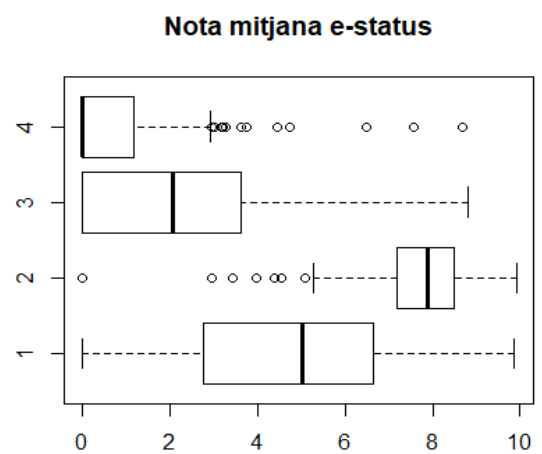


Figura 5.1.2.5. Boxplot perc ex casse

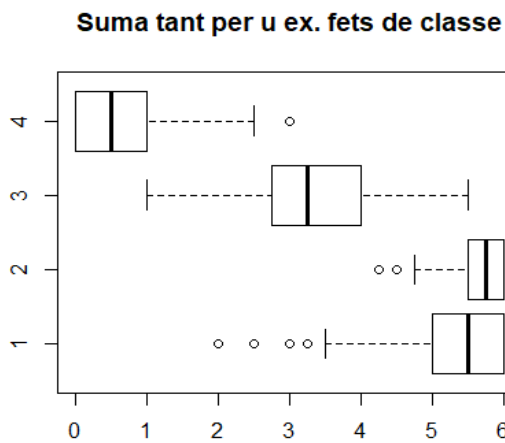


Figura 5.1.2.6. Boxplot notes classe

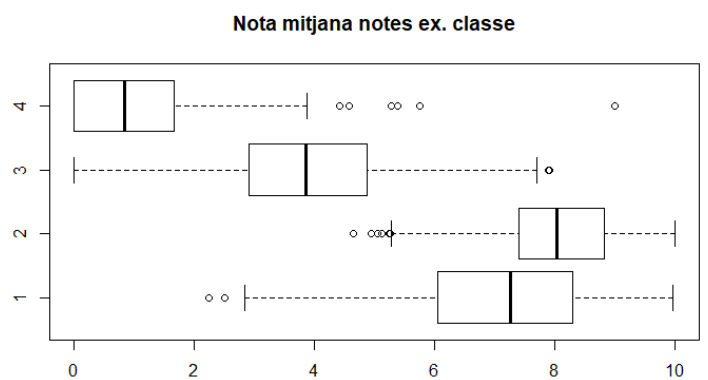
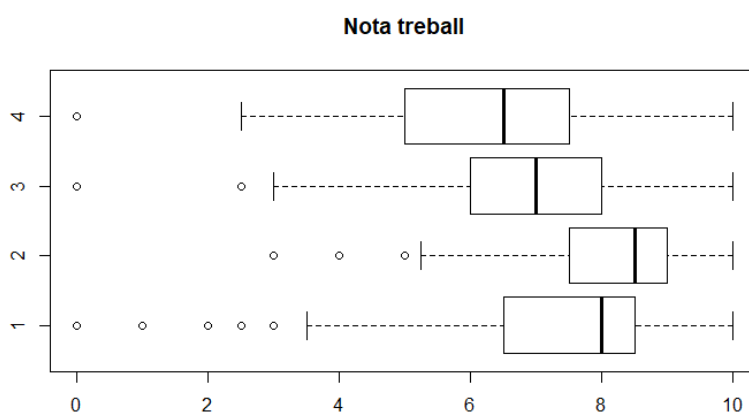


Figura 5.1.2.7. Boxplot nota treball



Taula 5.1.2.1. Tipus avaluació segons classes

	% av. cont.	% av. única
Classe 1	93,48 (1406/1504)	6,52 (98/1504)
Classe 2	98,56 (546/554)	1,44 (8/554)
Classe 3	70,41 (238/338)	29,59 (100/338)
Classe 4	30,61 (75/245)	69,39 (170/245)

Taula 5.1.2.2. Tipus avaluació segons quadrimestre

	% 1er Quadrimestre	% 2on Quadrimestre
Classe 1	68,82 (1035/1504)	31,18 (469/1504)
Classe 2	80,87 (448/554)	19,13 (106/554)
Classe 3	52,37 (177/338)	47,63 (161/338)
Classe 4	54,69 (134/245)	45,31 (111/245)

Així doncs, ja es pot obtenir un perfil prou clar del tipus d'alumne que pertany a cada classe:

Classe 1 – És el perfil de la majoria dels alumnes. En referència al seguiment del treball a l'aula es determina que són constants, fan tots els exercicis i la nota mediana d'aquests és un 7. Pel que fa al treball autònom ja no són tan constants, de mediana, de tots els problemes disponibles per fer d'e-status en fan la meitat, i la nota d'aquests de mediana és 5. De mediana dediquen 10 minuts a realitzar els problemes d'e-status i a l'acabar el curs han fet, de mediana, 30 execucions de problemes. El 90% dels alumnes d'aquest grup opten per l'avaluació continuada, i el 70% estan matriculats al primer quadrimestre.

Classe 2 – Aquests són un 20% dels alumnes. Són el grup més constant i treballador tant pel que es refereix al treball a l'aula com al treball autònom. Fan tots els exercicis, les notes medianes dels quals són 8 en els exercicis de classe i 7,9 en els problemes d'e-status. De mediana dediquen 15 minuts a realitzar els problemes d'e-status i a l'acabar el curs han fet, de mediana, 53 execucions d'aquests problemes. El 99% dels alumnes d'aquest grup fan l'avaluació continuada, i el 80% estan matriculats al primer quadrimestre.

Classe 3 – Aquests són un 10% dels alumnes. Aquests no són constants ni en el treball a l'aula ni en el treball autònom. De mediana, de tots els exercicis de classe en fan la meitat, i de tots els problemes disponibles per fer d'e-status en fan el 20%. Les notes medianes que obtenen són 3,9 en els exercicis de classe i 2 en els problemes d'e-status. De mediana dediquen 5 minuts a realitzar els problemes d'e-status i al acabar el curs han fet, de mediana, 10 execucions de problemes. El 70% d'aquests alumnes opten per l'avaluació continuada, i el 50% estan matriculats al primer quadrimestre.

Classe 4 – Aquests són un 10% dels alumnes, és el perfil d'alumne que no segueix gens el curs, no fa ni els exercicis de classe ni els problemes d'e-status. El 70% d'aquests alumnes opten per l'avaluació única, i el 55% estan matriculats al primer quadrimestre.

D'una manera molt resumida, és podria determinar que els alumnes de la classe 2 són els alumnes exemplars i excel·lents. Els de la classe 1 són els alumnes que segueixen el curs correctament, tot i que podrien dedicar-s'hi una mica més en el treball autònom. Els de la classe 3 es despengen i pràcticament no segueixen el curs i per últim els de la classe 4 són els alumnes que no segueixen el curs en cap moment.

En les següents taules es veu reflectit com els alumnes que segueixen el curs, les classes 1 i 2, acaben aprovant el curs en més proporció que els alumnes que no el segueixen, classes 3 i 4.

Taula 5.1.2.3. Avaluació continuada

	% suspesos	% aprovats
Classe 1	27,6 (388/1406)	72,4 (1018/1406)
Classe 2	12,09 (66/546)	87,91 (480/546)
Classe 3	50 (119/238)	50 (119/238)
Classe 4	44 (33/75)	56 (42/75)

Taula 5.1.2.4. Avaluació única

	% suspesos	% aprovats
Classe 1	73,47 (72/98)	26,53 (26/98)
Classe 2	37,5 (3/8)	62,5 (5/8)
Classe 3	79 (79/100)	21 (21/100)
Classe 4	88,24 (150/170)	11,76 (20/170)

5.2 Anàlisi components principals

Amb l'anàlisi de components principals (ACP) es vol reduir el nombre de variables, minimitzant la pèrdua "d'informació" de manera que es guanyi en significació. És a dir, es pretén obtenir noves variables (components) no correlacionades que agruparan a les variables originals, relacionades entre si, de forma que aquestes noves variables expliquin el comportament de les originals i, per tant, el màxim de la seva variabilitat. Té sentit doncs, si existeixen altes correlacions entre les variables, és a dir que existeix informació redundant i, per tant, pocs factors expliquen gran part de la variabilitat total.

Els alumnes que opten per l'avaluació continuada (2265 alumnes) fan els exàmens parcials, mentre que els que opten per la l'avaluació única (376 alumnes) no fan aquests i fan l'examen final, de manera que aplicarem l'ACP a aquestes dues poblacions per separat.

Alumnes avaluació continuada

Mitjançant la taula de valors propis veiem que les primeres tres components acumulen un 60% de la variabilitat de les dades, les següents components per si soles no aporten més d'un 10% d'inèrcia, així que com que no aporten gaire ens centrarem a interpretar les tres primeres que són les més explicatives, encara que l'ideal seria cobrir més variabilitat.

Taula 5.2.1. Valors propis i variància

	Valors propis	Variància %	Variància acumulada %
comp 1	4,74	36,461	36,461
comp 2	1,683	12,944	49,406
comp 3	1,319	10,147	59,553
comp 4	1,064	8,183	67,736
comp 5	0,926	7,12	74,856
comp 6	0,728	5,599	80,455
comp 7	0,625	4,805	85,26
comp 8	0,586	4,51	89,77
comp 9	0,512	3,939	93,71
comp 10	0,356	2,738	96,448
comp 11	0,245	1,885	98,333
comp 12	0,156	1,203	99,536
comp 13	0,06	0,464	100

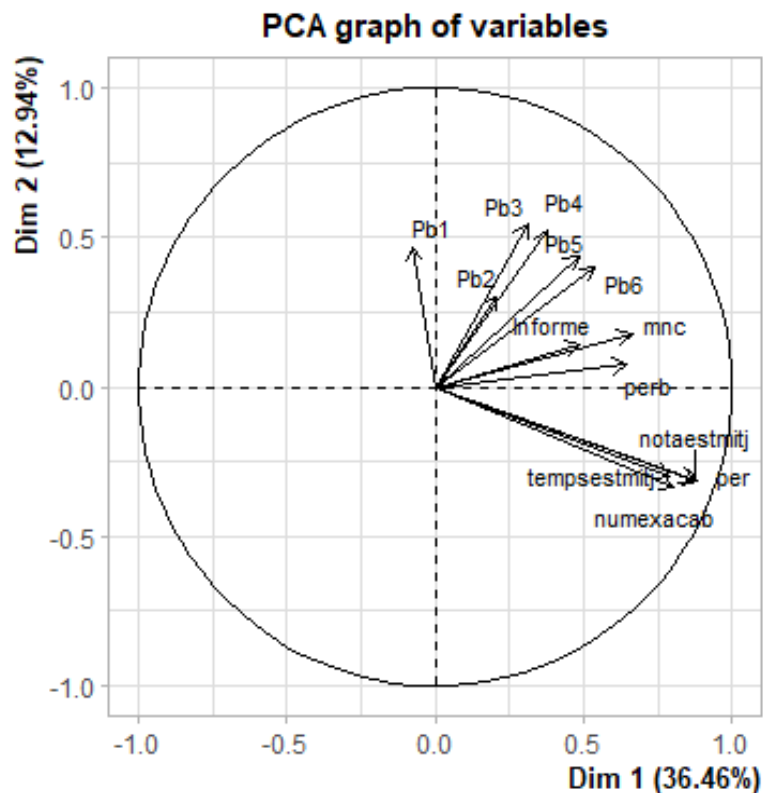
A la taula següent es mostren les correlacions entre les tres primeres components principals i les variables originals.

Taula 5.2.2 Matriu de correlacions

	Dim.1	Dim.2	Dim.3
Pb1	-0,076	0,468	0,053
Pb2	0,208	0,303	0,29
Pb3	0,312	0,542	0,112
Pb4	0,381	0,52	0,224
Pb5	0,483	0,44	0,177
Pb6	0,54	0,399	0,22
Informe	0,487	0,144	-0,545
numexacab	0,804	-0,332	0,184
notaestmitj	0,875	-0,313	0,188
tempsestmitj	0,793	-0,299	0,191
per	0,886	-0,316	0,19
perb	0,647	0,074	-0,594
mnc	0,665	0,177	-0,546

Per a interpretar el significat de cada component ho farem a partir de la matriu de correlacions vista anteriorment així com mitjançant el gràfic de les projeccions de les variables en els nous eixos.

Figura 5.2.1. Projeccions variables, eixos 1 i 2

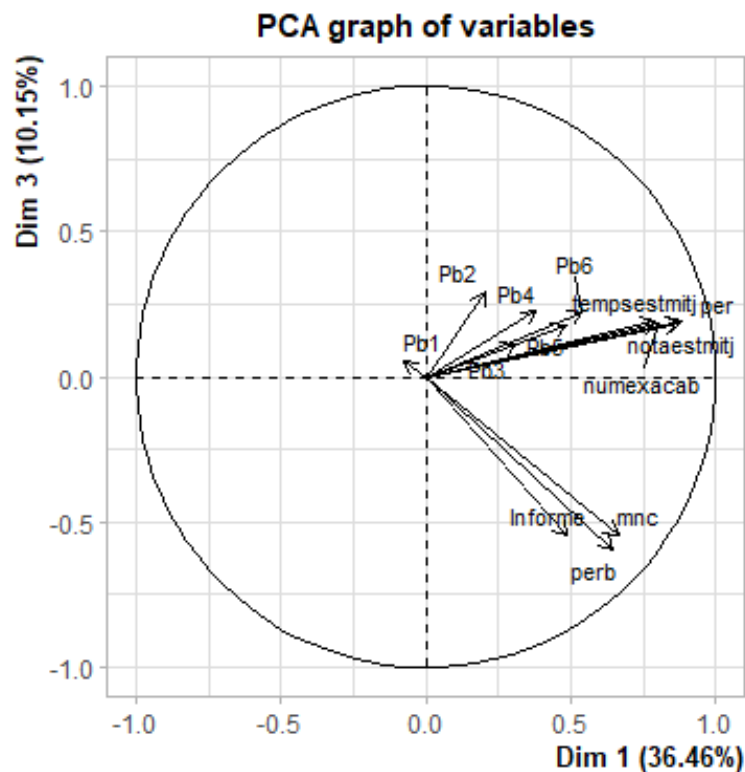


En aquest Biplot estan representades les variables sobre el primer eix (primera component) i segon eix (segona component), els quals acumulen un 50% d'inèrcia. Podríem destacar tres clústers de variables, un referent a les notes dels problemes dels exàmens, un altre referent als exercicis que es fan a l'aula i el treball final i un últim referent al treball autònom d'estatus. Tant les variables que es refereixen al treball a l'aula com les referents al treball autònom estan situades a prop del primer eix, és a dir que són les que més contribueixen en la seva formació, de manera que entendrem aquest primer eix com un eix de mida, de magnitud d'activitat acadèmica durant el curs, i com que es troben en el costat positiu, la primera component serà aproximadament funció creixent d'aquestes variables. És a dir, ordenarà l'eix de menor a major i, per tant, els alumnes a la dreta del gràfic tindran valors alts d'aquestes variables. D'altra banda, si es troben situats a l'esquerra, tindran valors baixos.

Les variables referents a les notes dels dos parcials, estan situades properes al segon eix, és a dir, són les que contribueixen a la formació d'aquest. Les variables es troben al costat positiu així que els alumnes que és situïn a la part de dalt del gràfic tindran més bones notes en els exercicis dels parcials que els que és situïn a la part de baix del gràfic.

A l'Annex 3 hi ha inclosos gràfics amb les projeccions dels alumnes sobre aquests dos eixos en els quals aquests estan distingits per colors segons els valors que prenen en diferents variables, es pot visualitzar prou clar com la primera component (primer eix/dimensió) plasma el grau d'activitat acadèmica durant el curs, i el segon eix les notes dels parcials.

Figura 5.2.2. Projeccions variables, eixos 1 i 3

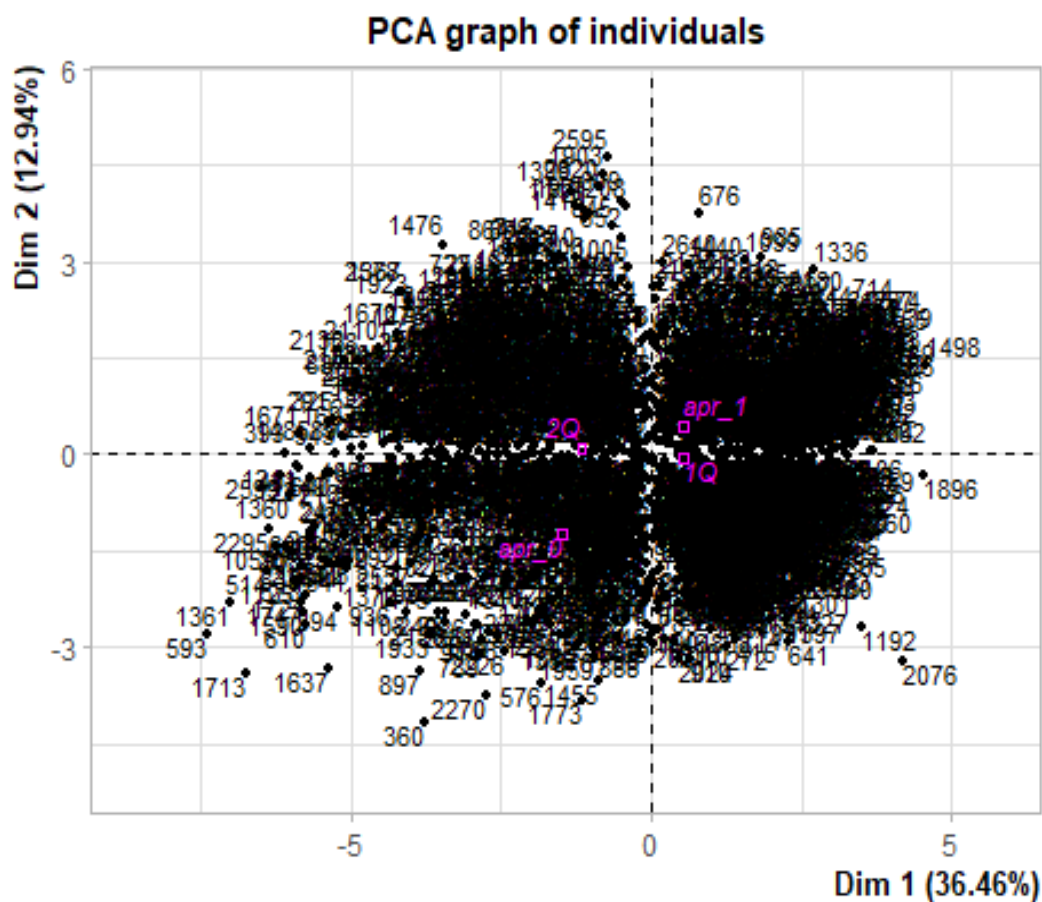


Ara s'han representat el primer i tercer eix, aquest últim recull un 10% d'inèrcia, per a poder definir la tercera component. Les úniques variables properes a aquest eix són el triplet referent al treball a l'aula i el treball final, que es troben a la meitat entre el primer i tercer eix. Com hem vist anteriorment aquestes variables formen la primera component, en funció creixent, ara veiem, ajudant-nos també de la matriu de correlacions, que aquestes variables també formen, quasi exclusivament, el tercer eix, aquest en funció decreixent, és a dir que els alumnes que és situïn a la part de dalt del gràfic tindran un rendiment més baix en aquestes variables que els que és situïn a la part de baix del gràfic.

La tercera component doncs, està explicada per unes variables que també expliquen la primera component, de manera que podríem explicar les dades només amb les dues primeres components. El primer eix representa el seguiment que es fa del curs i el segon eix les qualificacions dels parcials.

En el gràfic a continuació es projecten els alumnes sobre el primer i segon eix, així com les coordenades de dues variables categòriques suplementaries, el quadrimestre i si han aprovat o no l'assignatura. El núvol dels alumnes es reparteix bastant homogèniament per tot el gràfic, no hi ha cap patró que ressalti, cauen alumnes en tots els quadrants. Fent la interpretació que hem fet anteriorment de les dues primeres components principals, podem llegir que els alumnes situats en el primer quadrant són els que segueixen activament el curs i treuen bones notes als parcials, els alumnes que cauen al segon quadrant són els que no segueixen molt activament el curs però treuen bones notes als parcials, els que es situen al tercer quadrant són els que no segueixen el curs i no treuen bones notes als parcials, i els del quart quadrant els que segueixen el curs però no treuen bones notes als parcials. També veiem al gràfic que els alumnes que no aproven l'assignatura se situaran al tercer quadrant, ja que la coordenada d'aquest valor del factor se situa en aquest i allunyat del centre, els que aproven estan distribuïts entre el primer quadrant i part del segon i quart. Pel que fa al quadrimestre, no amb molta claredat, ja que les coordenades referents a cada quadrimestre estan bastant properes, però es podria dir que els alumnes que cursen l'assignatura al segon quadrimestre cauran més en el segon i tercer quadrant, i els alumnes del primer quadrimestre en el primer i quart quadrant.

Figura 5.2.3. Projeccions individus, eixos 1 i 2



A l'Annex 3 s'adjunta el mateix gràfic però amb els alumnes distingits segons el quadrimestre, i segons si han aprovat o no l'assignatura.

Alumnes avaluació única

Les primeres tres components acumulen un 80% de la variabilitat de les dades, de manera que són aquestes amb les que ens quedem i interpretarem.

Taula 5.2.3. Valors propis i variància

	Valors propis	Variància %	Variància acumulada %
comp 1	5,38	53,804	53,804
comp 2	1,741	17,413	71,217
comp 3	1,184	11,838	83,054
comp 4	0,471	4,705	87,759
comp 5	0,444	4,44	92,199
comp 6	0,281	2,814	95,013

comp 7	0,234	2,341	97,355
comp 8	0,122	1,22	98,574
comp 9	0,092	0,919	99,493
comp 10	0,051	0,507	100

La matriu de correlacions entre les tres primeres components principals i les variables originals és la següent.

Taula 5.2.4 Matriu de correlacions

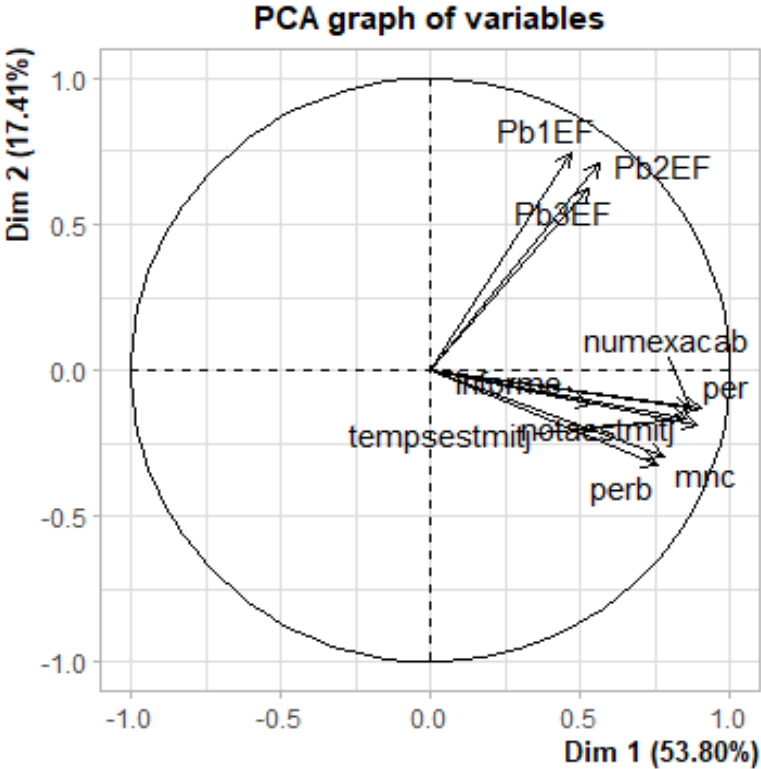
	Dim.1	Dim.2	Dim.3
Pb1EF	0,466	0,745	0,1
Pb2EF	0,562	0,707	0,03
Pb3EF	0,53	0,621	0,069
Informe	0,533	-0,112	0,674
numexacab	0,866	-0,131	-0,345
notaestmitj	0,892	-0,188	-0,316
tempsestmitj	0,855	-0,166	-0,264
Per	0,907	-0,127	-0,307
Perb	0,759	-0,323	0,411
Mnc	0,779	-0,295	0,402
numexacab	0,866	-0,131	-0,345
notaestmitj	0,892	-0,188	-0,316
tempsestmitj	0,855	-0,166	-0,264
Per	0,907	-0,127	-0,307
Perb	0,759	-0,323	0,411
Mnc	0,779	-0,295	0,402

Ara hem aplicat l'ACP sobre les mateixes variables però sobre els alumnes que fan l'avaluació única, les variables que canvien són les de les notes dels exàmens parcials, aquestes s'han substituït per les notes de l'examen final. A jutjar per aquestes correlacions les components obtingudes per als alumnes de l'avaluació única són, com era d'esperar, les mateixes que per als alumnes de l'avaluació continuada, en aquest cas però, les correlacions són més fortes, és a dir que els factors estan més ben formats, són més fidels, com hem vist les tres primeres components recullen un 80% de variabilitat, mentre que en el cas anterior recollien un 60%.

Així doncs, el primer eix també el formen les variables que expressen el seguiment del treball a l'aula i el treball autònom, és un eix de mida, els alumnes que es situïn a la dreta del gràfic seran els que tenen més activitat acadèmica, és a dir segueixen més el curs, fan més exercicis i obtenen més bones notes en aquests. El segon eix, és a dir, la segona component, representa les notes de l'examen final, també és funció creixent, alumnes que es situïn a la part de dalt del gràfic tindran qualificacions més altes i els que es situïn a la part de baix, més baixes. A continuació es mostra el biplot de les projeccions de les variables en aquests dos eixos. Les

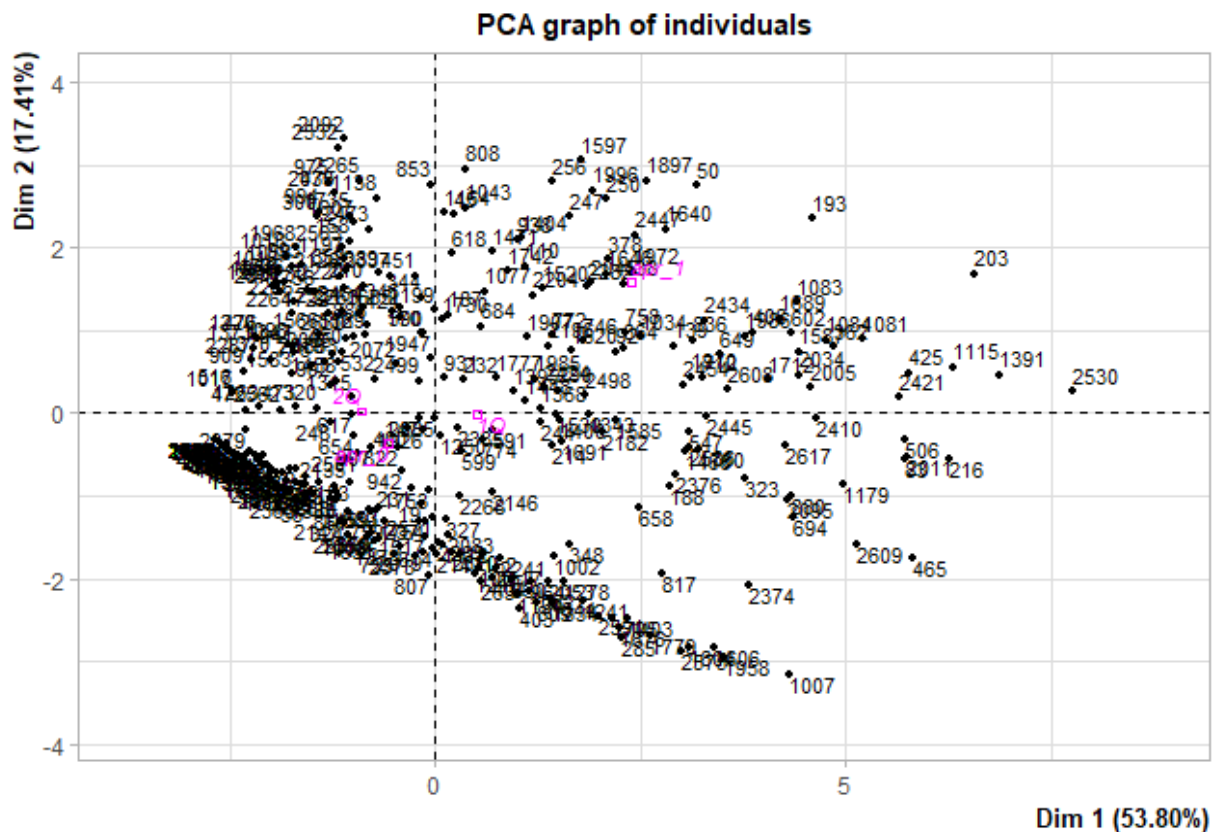
notes de l'examen final formen un angle pròxim a 90° amb les variables referents a l'activitat realitzada d'e-status, així que sembla que no hi ha relació entre el treball autònom i la nota a l'examen final.

Figura 5.2.4. Projeccions variables, eixos 1 i 2



En el següent gràfic es projecten els alumnes en el nou pla factorial, sobre la primera i segona component, així com les coordenades de dues variables categòriques suplementàries, el quadrimestre i si han aprovat o no l'assignatura.

Figura 5.2.5. Projeccions individus, eixos 1 i 2



Al gràfic no es distingeixen gaire bé les coordenades de les dues variables suplementàries, així que les afegim a la següent taula per a poder fer una millor lectura de les dades.

Taula 5.2.4 Coordenades variables suplementàries.

	Dim.1	Dim.2
1Q	0,51	-0,01
2Q	-0,9	0,02
apr_0	-0,57	-0,37
apr_1	2,40	1,58

Al segon i tercer quadrant els alumnes s'acumulen a prop del centre dels eixos, mentre que al primer i quart quadrant els alumnes estan més dispersos. Mitjançant les coordenades de les dues variables suplementàries determinem que l'eix d'abscisses discrimina el quadrimestre, els alumnes situats a l'esquerra del gràfic seran matriculats al segon semestre i els situats a la

dreta al primer semestre. També veiem que els alumnes situats al primer quadrant i allunyats de l'origen de coordenades són els que aproven, el qual és coherent tenint en compte que el primer eix es funció creixent de la quantitat i qualitat del treball que es fa durant el curs i que el segon eix es també funció creixent i referent a la nota de l'examen final.

6 - CONCLUSIONS

La intenció era explotar les dades de les que disposàvem per buscar patrons de comportament dels alumnes tan dintre d'un mateix curs com entre cursos, així com trobar relacions entre els diversos processos didàctics i el rendiment de l'alumne.

A l'utilitzar dades anònimes ens hem trobat amb què potser teníem poca informació, variables, per a caracteritzar els alumnes i l'activitat acadèmica que duen a terme durant el curs, per poder així categoritzar-los millor i fer un anàlisi més precís i trobar més casuístiques. Els anàlisis els hem realitzat bàsicament a partir de les qualificacions dels alumnes en els exercicis i exàmens, hauria estat interessant disposar de més informació, com podria ser les hores que dedica l'alumne a l'estudi d'aquesta assignatura, tenir l'historial de qualificacions d'altres assignatures, saber quantes assignatures més estava cursant a part d'aquesta o saber si treballa o no a part d'estudiar.

Tot i que pràcticament només disposàvem de les notes per a poder explicar com funcionen els alumnes en aquesta assignatura, hem pogut extreure algunes conclusions.

Al llarg d'un mateix curs, la tendència dels alumnes respecte al volum de feina que fan és decreixent, segueixen seguint el curs i fent feina però en menor mesura, és progressiu en el temps, no sembla que sigui un declivi a causa de que els últims blocs presentin més dificultat, cap bloc destaca en cap sentit.

Les components principals ens han permès distingir i explicar amb més facilitat el treball autònom i el treball a l'aula, així com les notes dels exàmens. No hem observat cap relació clara entre el procés didàctic i el rendiment de l'alumne, s'ha vist que el treball autònom va lligat amb el treball a l'aula i l'alumne es comporta de manera similar en els dos processos, així que no es destaca variació en el rendiment en funció del procés didàctic.

El que si hem observat mitjançant l'ACP i algunes descriptives bàsiques, és que el quadrimestre és un factor bastant determinant, els alumnes matriculats al segon quadrimestre tenen pitjors resultats que els del primer. Sabem que els alumnes que es matriculen al segon quadrimestre són alumnes que no van al dia amb el grau, segurament en algun moment han repetit i per tant no arriben a matricular-se al primer quadrimestre com ho faria un alumne que va superant totes les assignatures. Aquesta tendència no s'observa tan clarament durant els primers cinc cursos.

Hem pogut determinar també, que efectivament hi ha diferents perfils d'alumnes pel que fa al seguiment que fan de l'assignatura, els quals s'ha vist que obtenen notes diferents als exàmens. Principalment en podem distingir tres, el més comú és el de l'alumne que segueix l'assignatura al dia, a mesura que avança el curs va disminuint una mica l'activitat però no deixa de ser actiu durant tot curs, disminueix en major mesura el treball autònom que el treball a l'aula. Un altre perfil és el de l'alumne exemplar, el més treballador, aquest es caracteritza per ser actiu durant tot el curs, fa tots els exercicis tant de classe com d'e-status, i més d'una vegada. Per últim està el perfil d'alumne que en cap moment segueix el curs, o que el comença però es despenja molt ràpid, l'activitat acadèmica d'aquests és pràcticament nul·la.

Aquests serien els resultats més destacables obtinguts de les anàlisis realitzats, podem concloure també que l'ACP ens ha resultat molt útil per a explicar aquestes dades. Tot i que no hem obtingut respostes gaire sorprenents, hem sigut capaços de donar resposta als objectius que ens havíem plantejat. A aquestes dades encara els hi queda marge per a seguir sent explotades.

7 - BIBLIOGRAFIA

<http://people.tamu.edu/~alawing/materials/ESSM689/pca.pdf>

<http://wpage.unina.it/cafiero/books/pc.pdf>

<http://factominer.free.fr/factomethods/principal-components-analysis.html>

<https://blog.bioturing.com/2018/06/18/how-to-read-pca-biplots-and-scree-plots/>

https://web.archive.org/web/20091229050509/http://www.uoc.edu/in3/emath/docs/Componentes_principales.pdf

<http://diposit.ub.edu/dspace/bitstream/2445/121591/1/memoria.pdf>

<http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema3am.pdf>

<https://www.displayr.com/what-is-dendrogram/>

<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/hclust>

<https://www.rdocumentation.org/packages/cluster/versions/2.1.0/topics/daisy>

https://uc-r.github.io/hc_clustering

<https://es.wikipedia.org/wiki/Dendrograma>

https://uc-r.github.io/hc_clustering

<https://www.cs.us.es/~fran/TIB/semana5.pdf>

http://halweb.uc3m.es/esp/Personal/personas/agrane/ficheros_docencia/MULTIVARIANT/slides_Coarp_reducido.pdf

ANNEX 1

Script : TFG depuració e-status

```
library(openxlsx)
library(dplyr)

estatus <- read.xlsx('PE estatus net.xlsx', 1)
descri <- read.xlsx('PE estatus net.xlsx', 2)

intersect(names(estatus), names(descri))
estatus <- merge(estatus,descri, all=T)

# ---Variables
colnames(estatus)

## -- login
class(estatus$login)
sum(is.na(estatus$login))

## -- prob
summary(estatus$prob)
sum(is.na(estatus$prob)) #no hi ha missings
estatus$prob <- as.factor(estatus$prob)
table(estatus$prob)

## -- bloque
summary(estatus$bloque)
sum(is.na(estatus$bloque)) # no hi ha missings

table(estatus$bloque) #diferent nom per un mateix valor/nivell

#####
# Càlcul de Probabilitats (BLOC 1)
nrow(estatus %>% filter(bloque=="Càlcul de Probabilitats"))
nrow(estatus %>% filter(bloque=="Càlcul de Probabilitats  "))

# Variable aleatòria (BLOC 2)
```

```

nrow(estatus %>% filter(bloque=="Variable aleatòria"))
nrow(estatus %>% filter(bloque=="Variable Aleatòria"))

# Models de variable aleatòria (BLOC 3)
nrow(estatus %>% filter(bloque=="Models de variable aleatòria"))
nrow(estatus %>% filter(bloque=="Models de variable aleatòria  "))
nrow(estatus %>% filter(bloque=="&#9;Models de variable aleatòria  "))

# Evidència i inferència (BLOC 4)
nrow(estatus %>% filter(bloque=="&#9;Evidència i inferència"))
nrow(estatus %>% filter(bloque=="Evidència i inferència"))

# Models estadístics i previsió (BLOC 6)
nrow(estatus %>% filter(bloque=="Models estadístics i previsió"))
nrow(estatus %>% filter(bloque=="Models estadístics i previsió  "))
#####

for(i in 1:length(estatus$bloque)){

  if(estatus$bloque[i]=="Models de variable aleatòria  " ||
     estatus$bloque[i]=="&#9;Models de variable aleatòria  "){
    estatus$bloque[i] <- "Models de variable aleatòria"}

  if(estatus$bloque[i]=="&#9;Evidència i inferència"){
    estatus$bloque[i] <- "Evidència i inferència"}

  if(estatus$bloque[i]=="Càlcul de Probabilitats  "){
    estatus$bloque[i] <- "Càlcul de Probabilitats"}

  if(estatus$bloque[i]=="Variable Aleatòria"){
    estatus$bloque[i] <- "Variable aleatòria"}

  if(estatus$bloque[i]=="Models estadístics i previsió  "){
    estatus$bloque[i] <- "Models estadístics i previsió"}

  if((estatus$bloque[i]=="Preguntes Test") || (estatus$bloque[i]=="Proves
42") || (estatus$bloque[i]=="PS") || (estatus$bloque[i]=="PS 21"))

```



```

    || (estatus$bloque[i]=="PS B5") || (estatus$bloque[i]=="Qüestions de
repàs") || (estatus$bloque[i]=="Repàs inferència ")
    || (estatus$bloque[i]=="Repàs inferència") || (esta-
tus$bloque[i]=="Repàs") || (estatus$bloque[i]=="tests 41") || (esta-
tus$bloque[i]=="Tests")){
        estatus$bloque[i] <- "Altres"
    }
}

table(estatus$bloque)
estatus$bloque <- as.factor(estatus$bloque)

## -- nota
# els NA son els exercicis que caduquen i no s'acaben
summary(estatus$nota)
estatus$nota <- as.numeric(estatus$nota)
summary(estatus$nota)

## -- Año (del curs)
summary(estatus$Año)
estatus$Año <- as.factor(estatus$Año)
levels(estatus$Año)
colnames(estatus)[10]<-"año"

## -- Cuatrimestre (del curs)
summary(estatus$Cuatrimestre)
estatus$Cuatrimestre <- as.factor(estatus$Cuatrimestre)
levels(estatus$Cuatrimestre)
colnames(estatus)[9]<-"cuatri"

## -- Grupo
estatus$Grupo <- as.factor(estatus$Grupo)
levels(estatus$Grupo)

## -- Fin i Init

class(estatus$Fin)
estatus$Fin <- as.POSIXct(estatus$Fin)

```

```

estatus$Init <- as.POSIXct(estatus$Init)
class(estatus$Fin)

tiempomin <- as.numeric(estatus$Fin-estatus$Init)/60
summary(tiempomin)

estatus <- cbind(estatus, tiempomin)
estatus <- estatus[-1]

## -- bloque B-

blc <- c()
for(i in 1:nrow(estatus)){
  if(estatus$bloque[i]=="Càlcul de Probabilitats"){
    blc<-c(blc,1)}
  if(estatus$bloque[i]=="Variable aleatòria"){
    blc<-c(blc,2)}
  if(estatus$bloque[i]=="Models de variable aleatòria"){
    blc<-c(blc,3)}
  if(estatus$bloque[i]=="Evidència i inferència"){
    blc<-c(blc,4)}
  if(estatus$bloque[i]=="Models estadístics i previsió"){
    blc<-c(blc,6)}
  if(estatus$bloque[i]=="Disseny d'experiments"){
    blc<-c(blc,5)}
  if(estatus$bloque[i]=="Altres"){
    blc<-c(blc,99)
  }
}

estatus <- cbind(estatus, blc)
estatus$blc <- as.factor(estatus$blc)
save(estatus, file="estatus.RData")

```

Script : TFG depuració notes classe

```
setwd("C:/Users/carla/Desktop/Estadística UB - UPC/TFG/definitius")
```

```

library(openxlsx)
library(dplyr)

# --- Càrrega Excels
p12 <- read.xlsx('PE_P12.xlsx', 1) ; t12 <- read.xlsx('PE_T12.xlsx', 1)
p13 <- read.xlsx('PE_P13.xlsx', 1) ; t13 <- read.xlsx('PE_T13.xlsx', 1)
p14 <- read.xlsx('PE_P14.xlsx', 1) ; t14 <- read.xlsx('PE_T14.xlsx', 1)
p15 <- read.xlsx('PE_P15.xlsx', 1) ; t15 <- read.xlsx('PE_T15.xlsx', 1)
p16 <- read.xlsx('PE_P16.xlsx', 1) ; t16 <- read.xlsx('PE_T16.xlsx', 1)
p17 <- read.xlsx('PE_P17.xlsx', 1) ; t17 <- read.xlsx('PE_T17.xlsx', 1)
p18 <- read.xlsx('PE_P18.xlsx', 1) ; t18 <- read.xlsx('PE_T18.xlsx', 1)
p19 <- read.xlsx('PE_P19.xlsx', 1) ; t19 <- read.xlsx('PE_T19.xlsx', 1)

# --- Agrupació dels diferents anys en un mateix dataframe

# Anys p12-t12-p13-t13-p14, per cada bloc: 4 exercicis de classe
# La resta d'anys, per cada bloc: 2 exercicis de classe
# anys_a:
#   SMT : exercici classe en grup
#   SMT2 : exercici classe en grup
#   Coop : exercici classe en grup
#   PEM : exercici e-status a classe
# -- SES : e-status des de casa >>> TA
#
# anys_b:
#   PS1 : exercici classe en grup
#   PS2 : exercici classe en grup
# -- TA : e-status des de casa

anys_a <- rbind(p12, t12[,1:63],p13[,1:63],t13[,1:63],p14[,1:63])
anys_b <- rbind(t14[,1:51],p15[,1:51],t15[,1:51],p16[,1:51],t16[,1:51],
               t17[,1:51],p17[,1:51],p18[,1:51],t18[,1:51],p19[,1:51],
               t19[,1:51])

str(anys_a) ; str(anys_b)

```

```

# -- Eliminem notes ponderades
# eliminem SB (mitjana exercicis blocs)
# eliminem SES i TA (mitjana de les notes màximes )
# eliminiem NB notes blocs amb ponderacions
# del bloc 7 ens quedem amb la nota del treball (informe)

colnames(anys_a) ; colnames(anys_b)

bda <- anys_a[,c(-10,-11,-16,-17,-22,-23,-28,-29,-34,-35,-40,-41,
                -43,-44,-45,-52,-53,-54,-55,-56,-57,-58)]
bdb <- anys_b[,c(-8,-9,-12,-13,-16,-17,-20,-21,-24,-25,-28,-29,-31,-32,-
33,
                -40,-41,-42,-43,-44,-45)]

colnames(bda) ; colnames(bdb)

# s'inputen de nou algunes dades (no se sabia si era 0 o NA, s'inputen da-
des
# originals en les qe sí que es fa la distinció)

load("t14n.Rdata") ; load("p12n.Rdata") ; load("p13n.Rdata")
load("t12n.Rdata") ; load("t13n.Rdata")

# Tardor 14
intersect(colnames(bdb),colnames(t14n))
w<-(which(bdb$login==t14n$login))
w<-w[-length(w)]
sum(bdb[w,1]==t14n[,1])
bdb[w,][,6:17]<-t14n[,2:13]

# Primavera 13
bda[bda$any == 2013 & bda$cuatri == "2Q",1]==p13n[,1]
p13n<-p13n[c(-1,-98,-99),]
which(bda[bda$any == 2013 & bda$cuatri ==
"2Q",1]=="lhwunjorkuuvtrqipvldmvD") # 43 (45)
which(bda[bda$any == 2013 & bda$cuatri ==
"2Q",1]=="UQ5244JUYUMF260U060WT4") # 44 (43)
which(bda[bda$any == 2013 & bda$cuatri == "2Q",1]=="ietqgdclk83g") # 45
(44)

```

```

p13n<-rbind(p13n[c(1:42,44,45,43,46:nrow(p13n)),])
bda[bda$año == 2013 & bda$cuatri == "2Q",1]==p13n[,1]
bda[bda$año == 2013 & bda$cuatri == "2Q",][,6:29]<-p13n[,2:25]

# Primavera 12
bda[bda$año == 2012 & bda$cuatri == "2Q",1]==p12n[,1]
bda[bda$año == 2012 & bda$cuatri == "2Q",][,6:29]<-p12n[,2:25]

# Tardor 12
bda[bda$año == 2012 & bda$cuatri == "1Q",1]==t12n[,1]
which(bda[bda$año == 2012 & bda$cuatri == "1Q",1]=="1ltukd9ij89kfw5bf3") #
56 (57)
which(bda[bda$año == 2012 & bda$cuatri == "1Q",1]=="kkstiiitjbaea0") # 57
(56)
t12n<-rbind(t12n[c(1:55,57,56,58:nrow(t12n)),])
bda[bda$año == 2012 & bda$cuatri == "1Q",1]==t12n[,1]
bda[bda$año == 2012 & bda$cuatri == "1Q",][,6:29]<-t12n[,2:25]

# Tardor 13
bda[bda$año == 2013 & bda$cuatri == "1Q",1]==t13n[,1]
bda[bda$año == 2013 & bda$cuatri == "1Q",][,6:29]<-t13n[,2:25]

####

str(bda) ; str(bdb)
bda$"2SMT2" <- as.numeric(bda$"2SMT2")
bda$cuatri <- as.factor(bda$cuatri)

bdb$"3PS2"<-as.numeric(bdb$"3PS2")
bdb$cuatri <- as.factor(bdb$cuatri)

colnames(bda) ; colnames(bdb)

# --- Reordenament columnes
# per poder agrupar els dos dataframes en un de sol

notes<-bda[,c("login", "año", "cuatri", "grupo", "Pb1", "Pb2", "Pb3",
              "Pb4", "Pb5", "Pb6", "Pb1EF", "Pb2EF", "Pb3EF", "Informe")]

```

```

notes<-
rbind(notes,bdb[,c("login","año","cuatri","grupo","Pb1","Pb2","Pb3",
                  Pb4","Pb5","Pb6","Pb1EF","Pb2EF","Pb3EF","Informe")])

str(notes)
notes$año<-as.factor(notes$año)
notes$grupo<-as.factor(notes$grupo)

bda <- bda[,c(1:3,6:29)]
bdb <- bdb[,c(1:3,6:17)]

str(bda) ; str(bdb)
bda$año <- as.factor(bda$año)
bdb$año <- as.factor(bdb$año)

save(bda,file="bda.Rdata") ; save(bdb,file="bdb.Rdata")
save(notes,file = "notes.Rdata")

```

Script : TFG var e-status

```

library(dataone) ; library(pivottabler)
library(dplyr) ; library(openxlsx)

load("estatus.Rdata") ; load("notes.RData")

dim(estatus)
estatus<-estatus[-which(estatus$b1c==99),]
#eliminem les execucions que no son d'un dels 6 blocs

## Execucions totals que s'han fet de cada problema
# i problemes que s'han fet de cada bloc

pb1c<-qpvt(estatus,rows="prob",columns = "b1c", calculations =
"n()",totals=NULL)
pb1c<-pb1c$asDataFrame()
dim(pb1c)
head(pb1c)

```

```

for(i in 1:nrow(pblc)){
  for(j in 1:ncol(pblc)){
    if(is.na(pblc[i,j])){
      pblc[i,j]<-0
    }
  }
}

# problemes que hi ha a cada bloc

pblc1<-rownames(pblc)[pblc[,1]!=0]
pblc2<-rownames(pblc)[pblc[,2]!=0]
pblc3<-rownames(pblc)[pblc[,3]!=0]
pblc4<-rownames(pblc)[pblc[,4]!=0]
pblc5<-rownames(pblc)[pblc[,5]!=0]
pblc6<-rownames(pblc)[pblc[,6]!=0]
save(pblc1,pblc2,pblc3,pblc4,pblc5,pblc6,pblc,file="pblc1_6.Rdata")
# num. de problemes que hi ha a cada bloc
colSums(pblc!=0)

## num. d'execucions que fa l'alumne de cada bloc

head(estatus,2)
# es fa subdataframes porque com que hi ha repetidors, s'ha de fer els re-
comptes segons el login però també segons l'any i cuatri

q12_2 <- estatus%>%filter(año==2012,cuatri=="2Q")
q12_1 <- estatus%>%filter(año==2012,cuatri=="1Q")
(...)
q19_2 <- estatus%>%filter(año==2019,cuatri=="2Q")
q19_1 <- estatus%>%filter(año==2019,cuatri=="1Q")

# pba: problemes per bloc alumne

pb12a1<-qpvt(q12_1,rows="login",columns = "blc", calculations =
  c("num ex"="n()","num ex acab"="n()-sum(is.na(nota))",
    "Nota mitj","mean(nota, na.rm=TRUE)",
    "Nota med."="median(nota, na.rm=TRUE)",

```

```

        "Nota Sd"="sd(nota, na.rm=TRUE)"),totals=NULL)
pb12a1<-pb12a1$asDataFrame()
año<-rep(2012,nrow(pb12a1))
cuatri<-as.factor(rep("1Q",nrow(pb12a1)))
pb12a1<-cbind(pb12a1,año,cuatri)
#
pb12a2<-qpvt(q12_2,rows="login",columns = "blc", calculations =
        c("num ex"="n()","num ex acab"="n()-sum(is.na(nota))",
        "Nota mitj,"="mean(nota, na.rm=TRUE)",
        "Nota med."="median(nota, na.rm=TRUE)",
        "Nota Sd"="sd(nota, na.rm=TRUE)"),totals=NULL)
pb12a2<-pb12a2$asDataFrame()
año<-rep(2012,nrow(pb12a2))
cuatri<-as.factor(rep("2Q",nrow(pb12a2)))
pb12a2<-cbind(pb12a2,año,cuatri)

(...)

pb19a1<-qpvt(q19_1,rows="login",columns = "blc", calculations =
        c("num ex"="n()","num ex acab"="n()-sum(is.na(nota))",
        "Nota mitj,"="mean(nota, na.rm=TRUE)",
        "Nota med."="median(nota, na.rm=TRUE)",
        "Nota Sd"="sd(nota, na.rm=TRUE)"),totals=NULL)
pb19a1<-pb19a1$asDataFrame()
año<-rep(2019,nrow(pb19a1))
cuatri<-as.factor(rep("1Q",nrow(pb19a1)))
pb19a1<-cbind(pb19a1,año,cuatri)
#
pb19a2<-qpvt(q19_2,rows="login",columns = "blc", calculations =
        c("num ex"="n()","num ex acab"="n()-sum(is.na(nota))",
        "Nota mitj,"="mean(nota, na.rm=TRUE)",
        "Nota med."="median(nota, na.rm=TRUE)",
        "Nota Sd"="sd(nota, na.rm=TRUE)"),totals=NULL)
pb19a2<-pb19a2$asDataFrame()
año<-rep(2019,nrow(pb19a2))
cuatri<-as.factor(rep("2Q",nrow(pb19a2)))
pb19a2<-cbind(pb19a2,año,cuatri)

```



```

pban <- rbind(pb12a1,pb12a2,pb13a1,pb13a2,pb14a1,pb14a2,pb15a1,pb15a2,
             pb16a1,pb16a2,pb17a1,pb17a2,pb18a1,pb18a2,pb19a1,pb19a2)

pban<-cbind(login=rownames(pban),pban)
save(pban,file="pban.Rdata")

### mitjana temps execucio per bloc

mt12a1<-qpvt(q12_1,rows="login",columns = "blc", calculations =
             c("Temps mitj,"="mean(tiempomin, na.rm=TRUE)",
               "Temps med."="median(tiempomin, na.rm=TRUE)",
               "Temps Sd"="sd(tiempomin, na.rm=TRUE)"),totals=NULL)
mt12a1<-mt12a1$asDataFrame()
año<-rep(2012,nrow(mt12a1))
cuatri<-as.factor(rep("1Q",nrow(mt12a1)))
mt12a1<-cbind(mt12a1,año,cuatri)
#
mt12a2<-qpvt(q12_2,rows="login",columns = "blc", calculations =
             c("Temps mitj,"="mean(tiempomin, na.rm=TRUE)",
               "Temps med."="median(tiempomin, na.rm=TRUE)",
               "Temps Sd"="sd(tiempomin, na.rm=TRUE)"),totals=NULL)
mt12a2<-mt12a2$asDataFrame()
año<-rep(2012,nrow(mt12a2))
cuatri<-as.factor(rep("2Q",nrow(mt12a2)))
mt12a2<-cbind(mt12a2,año,cuatri)

(...)

mt19a2<-qpvt(q19_2,rows="login",columns = "blc", calculations =
             c("Temps mitj,"="mean(tiempomin, na.rm=TRUE)",
               "Temps med."="median(tiempomin, na.rm=TRUE)",
               "Temps Sd"="sd(tiempomin, na.rm=TRUE)"),totals=NULL)
mt19a2<-mt19a2$asDataFrame()
año<-rep(2019,nrow(mt19a2))
cuatri<-as.factor(rep("2Q",nrow(mt19a2)))
mt19a2<-cbind(mt19a2,año,cuatri)

```

```

mt <- rbind(mt12a1,mt12a2,mt13a1,mt13a2,mt14a1,mt14a2,mt15a1,mt15a2,
           mt16a1,mt16a2,mt17a1,mt17a2,mt18a1,mt18a2,mt19a1,mt19a2)
mt<-cbind(login=rownames(mt),mt)
save(mt,file="mt.Rdata")

```

```

# execucions fetes i acabades de cada prob de cada bloc

```

```

## -- BLOC 1

```

```

###

```

```

w1<-q12_1[which(q12_1$blc==1),]
pb1a12_1<-qpvt(w1,rows="login",columns="prob",
               calculations=c("num ex acab"="n()-
sum(is.na(nota))"),totals=NULL)
pb1a12_1<-pb1a12_1$asDataFrame()
n<-length(colnames(pb1a12_1))
per1<-as.data.frame(rowSums(pb1a12_1>0,na.rm = T)/n)
colnames(per1)<-"per"
año<-rep(2012,nrow(per1))
cuatri<-as.factor(rep("1Q",nrow(per1)))
per1<-cbind(per1,año,cuatri)

```

```

###

```

```

w2<-q12_2[which(q12_2$blc==1),]
pb1a12_2<-qpvt(w2,rows="login",columns="prob",
               calculations=c("num ex acab"="n()-
sum(is.na(nota))"),totals=NULL)
pb1a12_2<-pb1a12_2$asDataFrame()
n<-length(colnames(pb1a12_2))
per2<-as.data.frame(rowSums(pb1a12_2>0,na.rm = T)/n)
colnames(per2)<-"per"
año<-rep(2012,nrow(per2))
cuatri<-as.factor(rep("2Q",nrow(per2)))
per2<-cbind(per2,año,cuatri)

```

```

(...)

```

```

w16<-q19_2[which(q19_2$blc==1),]
pb1a19_2<-qpvt(w16,rows="login",columns="prob",

```

```

        calculations=c("num ex acab"="n()-
sum(is.na(nota))"),totals=NULL)
pb1a19_2<-pb1a19_2$asDataFrame()
n<-length(colnames(pb1a19_2))
per16<-as.data.frame(rowSums(pb1a19_2>0,na.rm = T)/n)
colnames(per16)<-"per"
año<-rep(2019,nrow(per16))
cuatri<-as.factor(rep("2Q",nrow(per16)))
per16<-cbind(per16,año,cuatri)

bloc1per <- rbind(per1,per2,per3,per4,per5,per6,per7,per8,per9,per10,
                per11,per12,per13,per14,per15,per16)

(...)

bloc5per <- rbind(per1,per2,per3,per4,per5,per6,per7,per8,per9,per10,
                per11,per12,per13,per14,per15,per16)

## -- BLOC 6

w1<-q12_1[which(q12_1$blc==6),]
pb1a12_1<-qpvt(w1,rows="login",columns="prob",
               calculations=c("num ex acab"="n()-
sum(is.na(nota))"),totals=NULL)
pb1a12_1<-pb1a12_1$asDataFrame()
n<-length(colnames(pb1a12_1))
per1<-as.data.frame(rowSums(pb1a12_1>0,na.rm = T)/n)
colnames(per1)<-"per"
año<-rep(2012,nrow(per1))
cuatri<-as.factor(rep("1Q",nrow(per1)))
per1<-cbind(per1,año,cuatri)

(...)

w16<-q19_2[which(q19_2$blc==6),]
pb1a19_2<-qpvt(w16,rows="login",columns="prob",
               calculations=c("num ex acab"="n()-
sum(is.na(nota))"),totals=NULL)

```

```

pb1a19_2<-pb1a19_2$asDataFrame()
n<-length(colnames(pb1a19_2))
per16<-as.data.frame(rowSums(pb1a19_2>0,na.rm = T)/n)
colnames(per16)<-"per"
año<-rep(2019,nrow(per16))
cuatri<-as.factor(rep("2Q",nrow(per16)))
per16<-cbind(per16,año,cuatri)

bloc6per <- rbind(per1,per2,per3,per4,per5,per6,per7,per8,per9,per10,
                 per11,per12,per13,per14,per15,per16)

names(bloc1per)[1]<"per1" ; names(bloc2per)[1]<"per2"
names(bloc3per)[1]<"per3" ; names(bloc4per)[1]<"per4"
names(bloc5per)[1]<"per5" ; names(bloc6per)[1]<"per6"

login<-rownames(bloc1per)
bloc1per<-cbind(login,bloc1per)
rownames(bloc1per)=NULL

(...)

login<-rownames(bloc6per)
bloc6per<-cbind(login,bloc6per)
rownames(bloc6per)=NULL

colnames(notes)
alu <- notes[,c(1:4,28)]
colnames(alu)

intersect(names(alu),names(bloc1per))
alu <- merge(alu,bloc1per,by = inter-
sect(names(alu),names(bloc1per)),all.x=T)
alu <- merge(alu,bloc2per,by = inter-
sect(names(alu),names(bloc2per)),all.x=T)
alu <- merge(alu,bloc3per,by = inter-
sect(names(alu),names(bloc3per)),all.x=T)
alu <- merge(alu,bloc4per,by = inter-
sect(names(alu),names(bloc4per)),all.x=T)

```

```

alu <- merge(alu,bloc5per,by = inter-
sect(names(alu),names(bloc5per)),all.x=T)
alu <- merge(alu,bloc6per,by = inter-
sect(names(alu),names(bloc6per)),all.x=T)

save(alu,file = "alu.Rdata")

#####
colnames(mt)<-
c("login","t_mitj1","t_med1","t_sd1","t_mitj2","t_med2","t_sd2",
  "t_mitj3","t_med3","t_sd3","t_mitj4","t_med4","t_sd4",
  "t_mitj5","t_med5","t_sd5","t_mitj6","t_med6","t_sd6",
  "año","cuatri")
colnames(pban)<-
c("login","n_exc1","n_exc_acab1","nota_mitj1","nota_med1","nota_sd1",

"n_exc2","n_exc_acab2","nota_mitj2","nota_med2","nota_sd2",

"n_exc3","n_exc_acab3","nota_mitj3","nota_med3","nota_sd3",

"n_exc4","n_exc_acab4","nota_mitj4","nota_med4","nota_sd4",

"n_exc5","n_exc_acab5","nota_mitj5","nota_med5","nota_sd5",

"n_exc6","n_exc_acab6","nota_mitj6","nota_med6","nota_sd6",
  "año","cuatri")

save(pban,file = "pban.Rdata")
save(mt,file = "mt.Rdata")
save(alu,file = "alu.Rdata")

# BBDD pe

colnames(notes) ; colnames(pban) ; colnames(mt) ; colnames(alu)

pe<-merge(notes,alu,by=intersect(names(notes),names(alu)),all.x = T)
pe<-merge(pe,pban,by=intersect(names(pe),names(pban)),all.x = T)
pe<-merge(pe,mt,by=intersect(names(pe),names(mt)),all.x = T)

cursfact<-c()

```

```

for(i in 1:nrow(pe)){
  if(pe$curso[i]=="2Q2012"){cursfact<-c(cursfact,1)}
  if(pe$curso[i]=="1Q2012"){cursfact<-c(cursfact,2)}
  if(pe$curso[i]=="2Q2013"){cursfact<-c(cursfact,3)}
  if(pe$curso[i]=="1Q2013"){cursfact<-c(cursfact,4)}
  if(pe$curso[i]=="2Q2014"){cursfact<-c(cursfact,5)}
  if(pe$curso[i]=="1Q2014"){cursfact<-c(cursfact,6)}
  if(pe$curso[i]=="2Q2015"){cursfact<-c(cursfact,7)}
  if(pe$curso[i]=="1Q2015"){cursfact<-c(cursfact,8)}
  if(pe$curso[i]=="2Q2016"){cursfact<-c(cursfact,9)}
  if(pe$curso[i]=="1Q2016"){cursfact<-c(cursfact,10)}
  if(pe$curso[i]=="2Q2017"){cursfact<-c(cursfact,11)}
  if(pe$curso[i]=="1Q2017"){cursfact<-c(cursfact,12)}
  if(pe$curso[i]=="2Q2018"){cursfact<-c(cursfact,13)}
  if(pe$curso[i]=="1Q2018"){cursfact<-c(cursfact,14)}
  if(pe$curso[i]=="2Q2019"){cursfact<-c(cursfact,15)}
  if(pe$curso[i]=="1Q2019"){cursfact<-c(cursfact,16)}
}

```

```

pe<-cbind(pe,cursfact)
save(pe,file="pe.Rdata")

```

Script : TFG var notes classe

```

load("bda.Rdata") ;load("bdb.Rdata") ; load("notes.Rdata")
# --- Variable MCB
# MCB : mitjana de les notes dels exercicis fets a classe del bloc (Mitjana Classe Bloc)

MCB1<-MCB2<-MCB3<-MCB4<-MCB5<-MCB6<-rep(0,nrow(bda))
for(i in 1:nrow(bda)){
  MCB1[i]<-mean(as.numeric(bda[i,4:7]),na.rm = T)
  MCB2[i]<-mean(as.numeric(bda[i,8:11]),na.rm = T)
  MCB3[i]<-mean(as.numeric(bda[i,12:15]),na.rm = T)
  MCB4[i]<-mean(as.numeric(bda[i,16:19]),na.rm = T)
  MCB5[i]<-mean(as.numeric(bda[i,20:23]),na.rm = T)
  MCB6[i]<-mean(as.numeric(bda[i,24:27]),na.rm = T)
}

```

```

var1<-cbind(MCB1,MCB2,MCB3,MCB4,MCB5,MCB6)

MCB1<-MCB2<-MCB3<-MCB4<-MCB5<-MCB6<-rep(0,nrow(bdb))
for(i in 1:nrow(bdb)){
  MCB1[i]<-mean(as.numeric(bdb[i,4:5]),na.rm = T)
  MCB2[i]<-mean(as.numeric(bdb[i,6:7]),na.rm = T)
  MCB3[i]<-mean(as.numeric(bdb[i,8:9]),na.rm = T)
  MCB4[i]<-mean(as.numeric(bdb[i,10:11]),na.rm = T)
  MCB5[i]<-mean(as.numeric(bdb[i,12:13]),na.rm = T)
  MCB6[i]<-mean(as.numeric(bdb[i,14:15]),na.rm = T)
}
var2<-cbind(MCB1,MCB2,MCB3,MCB4,MCB5,MCB6)
varn<-as.data.frame(rbind(var1,var2))
notes<-cbind(notes,varn)
notes<-cbind(notes,paste(notes[,3],notes[,2],sep=""))
colnames(notes)[ncol(notes)]<-"curso"

# --- Número de vegades que repeteixen
convoc <- as.data.frame.matrix(table(notes$login, notes$curso))
convoc <- cbind(login=rownames(convoc),convoc)
sum(convoc[,-1]) ; nrow(notes)

convoc <- con-
voc%>%mutate(conv=rowSums(convoc%>%select("1Q2012","1Q2013","1Q2014","1Q20
15","1Q2016","1Q2017","1Q2018",

"1Q2019","2Q2012","2Q2013","2Q2014","2Q2015","2Q2016","2Q2017","2Q2018","2
Q2019"))))
convoc <- convoc[,c(1,ncol(convoc))]

# --- Mitjana notes blocs 1,2,3 i blocs 4,5,6
# - CP1: Exercicis classe parcial 1 (Mitjana nota MCB1, MCB2 i MCB3)
# - CP2: Exercicis classe parcial 2 (Mitjana nota MCB4, MCB5 i MCB6)
notes <-
notes%>%mutate(CP1=rowMeans(notes%>%select("MCB1","MCB2","MCB3"),na.rm =
T))
notes <-
notes%>%mutate(CP2=rowMeans(notes%>%select("MCB4","MCB5","MCB6"),na.rm =
T))
#---

```

```

intersect(names(notes), names(convoc))
notes <- merge(notes,convoc, all=T)
str(notes)
notes$conv <- as.factor(notes$conv)

# - Percentatge d'exercicis de classe fets per bloc

perb1<-perb2<-perb3<-perb4<-perb5<-perb6<-rep(0,nrow(bda))
for(i in 1:nrow(bda)){
  perb1[i]<-rowSums(is.na(bda[i,4:7])==F)/4
  perb2[i]<-rowSums(is.na(bda[i,8:11])==F)/4
  perb3[i]<-rowSums(is.na(bda[i,12:15])==F)/4
  perb4[i]<-rowSums(is.na(bda[i,16:19])==F)/4
  perb5[i]<-rowSums(is.na(bda[i,20:23])==F)/4
  perb6[i]<-rowSums(is.na(bda[i,24:27])==F)/4
}
p1<-as.data.frame(cbind(bda[,1:3],perb1,perb2,perb3,perb4,perb5,perb6))
colnames(p1)[1:3]<-c("login","año","cuatri")

colnames(bdb)
perb1<-perb2<-perb3<-perb4<-perb5<-perb6<-rep(0,nrow(bdb))
for(i in 1:nrow(bdb)){
  perb1[i]<-rowSums(is.na(bdb[i,4:5])==F)/2
  perb2[i]<-rowSums(is.na(bdb[i,6:7])==F)/2
  perb3[i]<-rowSums(is.na(bdb[i,8:9])==F)/2
  perb4[i]<-rowSums(is.na(bdb[i,10:11])==F)/2
  perb5[i]<-rowSums(is.na(bdb[i,12:13])==F)/2
  perb6[i]<-rowSums(is.na(bdb[i,14:15])==F)/2
}
p2<-as.data.frame(cbind(bdb[,1:3],perb1,perb2,perb3,perb4,perb5,perb6))
colnames(p2)[1:3]<-c("login","año","cuatri")
p<-rbind(p1,p2)
pe<-merge(pe,p,by=intersect(names(pe),names(p)),all.x = T)
perb<-rowSums(pe[,90:95])
pe<-cbind(pe,perb)
save(notes,file = "notes.Rdata")
save(convoc,file = "convoc.Rdata")

```



```

# --- np: nota parcial ponderada
np1<-rep(0,nrow(pe2))
pe <- cbind(pe,np1)
for(i in 1:nrow(pe)){
  pe$np1[i] <- sum(as.numeric(pe[i,6:8])*10:12)/33
}

np2<-rep(0,nrow(pe))
pe <- cbind(pe,np2)
for(i in 1:nrow(pe)){
  pe$np2[i] <- sum(as.numeric(pe[i,9:11])*13:15)/42
}

# -- Avaluació única
colnames(pe)

avuni<-rep(0,nrow(pe))
for(i in 1:nrow(pe)){
  if(is.na(pe$np1[i])==T||is.na(pe$np2[i])==T){
    avuni[i]<-1}
}

pe<-cbind(pe,avuni)
pe$avuni<-as.factor(pe$avuni)
table(pe$avuni,pe$curso)
save(pe,file="pe.Rdata")

nota_avcont<-rep(0,nrow(pe))
for(i in 1:nrow(pe)){
  nota_avcont[i] <- sum(c(pe[i,25],pe[i,26]),na.rm=T)/2
}

nota_avunica<-rep(0,nrow(pe))
for(i in 1:nrow(pe)){
  nota_avunica[i]<-sum(c(pe[i,12],pe[i,13],pe[i,14]),na.rm=T)/3
}

```

```
pe<-cbind(pe,nota_avcont,nota_avunica)
save(pe,file="pe.Rdata")
numexacab<-rowSums(dataclust[,30:35],na.rm = T)
notaestmitj<-rowSums(dataclust[,36:41],na.rm = T)/6
tempsestmitj<-rowSums(dataclust[,42:47],na.rm = T)/6
per<-rowSums(dataclust[,24:29],na.rm = T)
perb<-rowSums(dataclust[,18:23],na.rm = T)
mnc<-rowSums(pe[,16:21],na.rm = T)/6

pe<-cbind(pe,numexacab,notaestmitj,tempsestmitj,per,mnc,perb)
save(pe,file="pe.Rdata")
```

ANNEX 2

Script : TFG Clustering

```
load("pe.Rdata")
library("dendextend")

colnames(pe)
dataclust<-cbind(pe$Informe,pe$MCB1,pe$MCB2,pe$MCB3,pe$MCB4,
                pe$MCB5,pe$MCB6,pe$perb1,pe$perb2,pe$perb3,pe$perb4,
                pe$perb5,pe$perb6,pe$per1,pe$per2,pe$per3,pe$per4,pe$per5,pe$per6,
                pe$n_exc_acab1,pe$n_exc_acab2,pe$n_exc_acab3,pe$n_exc_acab4,
                pe$n_exc_acab5,pe$n_exc_acab6,pe$nota_mitj1,pe$nota_mitj2,pe$nota_mitj3,
                pe$nota_mitj4,pe$nota_mitj5,pe$nota_mitj6,pe$t_mitj1,pe$t_mitj2,
                pe$t_mitj3,pe$t_mitj4,pe$t_mitj5,pe$t_mitj6)

## CLUSTERING JERARQUIC

dissimMatrix <- daisy(dataclust, metric = "gower", stand=TRUE)
class(dissimMatrix)

h1 <- hclust(dissimMatrix,method="ward.D")
plot(h1)
rect.hclust(h1, k = 4, border = 2:6)
plot(color_branches(as.dendrogram(h1), k = 4))

c4<-cutree(h1,4)
table(c4)

## PROFILING

# Nota final
boxplot(pe$nota_avcont[pe$avuni==0]~c4[pe$avuni==0], horizon-
tal=F,main="Nota final")
boxplot(pe$nota_avunica[pe$avuni==1]~c4[pe$avuni==1], horizon-
tal=F,main="Nota final")
```

```

# informe
boxplot(pe$Informe~c4, horizontal=F,main="Nota treball")

# exec total estatus
boxplot(pe$numexacab~c4, horizontal=T,main="Execucions totals prob. e-
status")
with(pe, tapply(numexacab, c4, median))

# mitjana exercicis classe
boxplot(pe$mnc~c4, horizontal=F,main="Nota mitjana notes ex. classe")
round(with(pe, tapply(mnc, c4, mean)),2)

# suma de % exercicis e-status fets de cada bloc
boxplot(pe$per~c4, horizontal=T,main="Suma tant per u prob. fets d'e-
status")
round(with(pe, tapply(per, c4, median)),2)

# % exercicis e-status bloc 1
boxplot(pe$per1~c4, horizontal=F,main="Tant per u prob. fets d'e-status
B1")
# % exercicis e-status bloc 2
boxplot(pe$per2~c4, horizontal=F,main="Tant per u prob. fets d'e-status
B2")
# % exercicis e-status bloc 2
boxplot(pe$per3~c4, horizontal=F,main="Tant per u prob. fets d'e-status
B3")
# % exercicis e-status bloc 2
boxplot(pe$per4~c4, horizontal=F,main="Tant per u prob. fets d'e-status
B4")
# % exercicis e-status bloc 2
boxplot(pe$per5~c4, horizontal=F,main="Tant per u prob. fets d'e-status
B5")
# % exercicis e-status bloc 2
boxplot(pe$per6~c4, horizontal=F,main="Tant per u prob. fets d'e-status
B6")

# suma de % exercicis classe fets de cada bloc

```

```

boxplot(pe$perb~c4, horizontal=T,main="Suma tant per u ex. fets de
classe")
round(with(pe, tapply(perb, c4, mean)),2)

# temps mig per problema e-status
boxplot(pe$tempsestmitj~c4, horizontal=T,main="Temps mig per prob d'e-
status")
round(with(pe, tapply(tempsestmitj, c4, mean)),2)

# nota mitjana probleme e-status
boxplot(pe$notaestmitj~c4, horizontal=T,main="Nota mitjana e-status")
round(with(pe, tapply(notaestmitj, c4, median)),2)

table(c4,pe$conv)[,2:5]#repetidors

table(c4,pe$avuni)

d<-as.matrix.data.frame(table(c4,pe$avuni))
d<-cbind(d,total=d[,1]+d[,2])
d<-cbind(d,av_cont=d[,1]/d[,3]*100,av_uni=d[,2]/d[,3]*100)
rownames(d)←1:4 ; d

f<-table(pe$cuatri,c4)
f<-as.matrix.data.frame(table(c4,pe$cuatri))
f<-cbind(f,total=f[,1]+f[,2])
f<-cbind(f,"1Q"=f[,1]/f[,3]*100,"2Q"=f[,2]/f[,3]*100) ; f

q<-as.matrix.data.frame(table(pe$apr_cont[pe$avuni==0],c4[pe$avuni==0]))
q<-rbind(q,total=q[1,]+q[2,])
q<-rbind(q,"sus"=q[1,]/q[3,]*100,"apr"=q[2,]/q[3,]*100) ; q

y<-as.matrix.data.frame(table(pe$apr_uni[pe$avuni==1],c4[pe$avuni==1]))
y<-rbind(y,total=y[1,]+y[2,])
y<-rbind(y,"sus"=y[1,]/y[3,]*100,"apr"=y[2,]/y[3,]*100) ; y

```

Script : TFG Descriptiva

```

#----- Càrrega dades -----

setwd("C:/Users/carla/Desktop/Estadística UB - UPC/TFG/Rdata")

library(dataone)
library(pivottabler)

load("pe.RData")

###----- ALUMNES

# Alumnes per curs
table(pe$cuatri,pe$año)
t1<-table(pe$cuatri,pe$año)
write.table(t1, file = "t1.txt", sep = ",")
# % d'alumnes que hi ha més al 1Q que al 2Q
t2<-(table(notes$cuatri,notes$año)[1,]/table(notes$cuatri,notes$año)[2,]-
1)*100
write.table(t2, file = "t2.txt", sep = ",")

# av continuada vs av unica
t3<-table(pe$avuni,pe$cursfact)
write.table(t3, file = "t3.txt", sep = ",")

# Es presenten a l'examen final, segons tipus av.
t4<-table(pe$avuni[pe$nota_avunica>0],pe$cursfact[pe$nota_avunica>0])
write.table(t4, file = "t4.txt", sep = ",")
# percentatge qe es presenten a l'examen final
#t5<-
round(table(pe$avuni[pe$nota_avunica>0],pe$cursfact[pe$nota_avunica>0])/ta
ble(pe$avuni,pe$cursfact)*100,2)
#write.table(t5, file = "t5.txt", sep = ",")

# % av unica
t5<-round(t4[2,]/(t4[2,]+t4[1,])*100,2)
write.table(t5, file = "t5.txt", sep = ",")
# % av cont
t6<-round(t4[1,]/(t4[2,]+t4[1,])*100,2)

```

```

write.table(t6, file = "t6.txt", sep = ",")

###

avcont<-pe[which(pe$avuni==0),]
length(which(avcont$nota_avunica>0))#av cont es presenten al final

avuni<-pe[which(pe$avuni==1),]
length(which(avuni$nota_avunica>0))#av uni es presenten al final

avcontuni<-avcont[which(avcont$nota_avunica>0),]

length(which(avcontuni$nota_avcont<5))/
  length(which(avcont$nota_avunica>0))
#av cont es presenten al final pq han suspes

# nota av continua
round(tapply(pe$nota_avcont[pe$avuni==0],pe$cursfact[pe$avuni==0],mean),2)
round(mean(pe$nota_avcont[pe$avuni==0]),2)

# nota av cont i unica
round(tapply(pe$nota_avunica[pe$avuni==1],pe$cursfact[pe$avuni==1],mean),2)
)
round(mean(pe$nota_avunica[pe$avuni==1]),2)

# Taula freq repetidors
t11<-table(pe$conv)[-1]/(2:5)
write.table(t11, file = "t11.txt", sep = ",")

sum(table(notes$conv)[-1]/(2:5))# total
length(which(convoc$conv>1))# repetidors (poden haver repetit un o més dun
cop)

## Repetidors per curs

```

```

repc<-qpvt(notes %>% filter(conv!=1), "cuatri", "año", "n()")
t12 <- repc$asDataFrame()
write.table(t12, file = "t12.txt", sep = ",")
# aquests 439 individus en realitat son 194 persones

t12/table(pe$cuatri,pe$año)

###----- CLASSE

colnames(pe)

plot(tapply(pe$MCB1,pe$cursfact,mean,na.rm=T),type="l",ylim=c(0,10),
      ylab = "Nota", xlab="primavera 2012 - tardor 2019", main="Nota mitja-
na exercicis de classe")
lines(tapply(pe$MCB2,pe$cursfact,mean,na.rm=T),type="l",col="red")
lines(tapply(pe$MCB3,pe$cursfact,mean,na.rm=T),type="l",col="blue")
lines(tapply(pe$MCB4,pe$cursfact,mean,na.rm=T),type="l",col="green")
lines(tapply(pe$MCB5,pe$cursfact,mean,na.rm=T),type="l",col="orange")
lines(tapply(pe$MCB6,pe$cursfact,mean,na.rm=T),type="l",col="purple")
legend("bottomright",legend=c("Bloc 1","Bloc 2","Bloc 3","Bloc 4",
                              "Bloc 5","Bloc 6"),
      pch=16,col=c("black","red","blue","green","orange","purple"))

plot(tapply(pe$perb1,pe$cursfact,mean,na.rm=T),type="l",ylim=c(0,1),
      ylab = "Nota", xlab="primavera 2012 - tardor 2019", main="Tant per u
exercicis de classe fets")
lines(tapply(pe$perb2,pe$cursfact,mean,na.rm=T),type="l",col="red")
lines(tapply(pe$perb3,pe$cursfact,mean,na.rm=T),type="l",col="blue")
lines(tapply(pe$perb4,pe$cursfact,mean,na.rm=T),type="l",col="green")
lines(tapply(pe$perb5,pe$cursfact,mean,na.rm=T),type="l",col="orange")
lines(tapply(pe$perb6,pe$cursfact,mean,na.rm=T),type="l",col="purple")
legend("bottomleft",legend=c("Bloc 1","Bloc 2","Bloc 3","Bloc 4",
                              "Bloc 5","Bloc 6"),
      pch=16,col=c("black","red","blue","green","orange","purple"))

```



```
plot(tapply(pe$Informe,pe$cursfact,mean,na.rm=T),type="l",ylim=c(0,10),
      ylab = "Nota", xlab="primavera 2012 - tardor 2019", main="Nota tre-
ball curs")
```

```
plot(tapply(pe$nota_mitj1,pe$cursfact,mean,na.rm=T),type="l",ylim=c(0,10),
      ylab = "Nota", xlab="primavera 2012 - tardor 2019", main="Nota mitja-
na problemes e-status")
lines(tapply(pe$nota_mitj2,pe$cursfact,mean,na.rm=T),type="l",col="red")
lines(tapply(pe$nota_mitj3,pe$cursfact,mean,na.rm=T),type="l",col="blue")
lines(tapply(pe$nota_mitj4,pe$cursfact,mean,na.rm=T),type="l",col="green")
lines(tapply(pe$nota_mitj5,pe$cursfact,mean,na.rm=T),type="l",col="orange"
)
lines(tapply(pe$nota_mitj6,pe$cursfact,mean,na.rm=T),type="l",col="purple"
)
legend("bottomright",legend=c("Bloc 1","Bloc 2","Bloc 3","Bloc 4",
                              "Bloc 5","Bloc 6"),
      pch=16,col=c("black","red","blue","green","orange","purple"))
```

```
plot(tapply(pe$t_mitj1,pe$cursfact,mean,na.rm=T),type="l",ylim=c(0,23),
      ylab = "Nota", xlab="primavera 2012 - tardor 2019", main="Temps mig
d'execució problemes e-status")
lines(tapply(pe$t_mitj2,pe$cursfact,mean,na.rm=T),type="l",col="red")
lines(tapply(pe$t_mitj3,pe$cursfact,mean,na.rm=T),type="l",col="blue")
lines(tapply(pe$t_mitj4,pe$cursfact,mean,na.rm=T),type="l",col="green")
lines(tapply(pe$t_mitj5,pe$cursfact,mean,na.rm=T),type="l",col="orange")
lines(tapply(pe$t_mitj6,pe$cursfact,mean,na.rm=T),type="l",col="purple")
legend("bottomleft",legend=c("Bloc 1","Bloc 2","Bloc 3","Bloc 4",
                              "Bloc 5","Bloc 6"),
      pch=16,col=c("black","red","blue","green","orange","purple"))
```

```
plot(tapply(pe$n_exc_acab1,pe$cursfact,mean,na.rm=T),type="l",ylim=c(0,18)
,xlim=c(0,20),
      ylab = "Nota", xlab="primavera 2012 - tardor 2019", main="Mitjana nº
execucions de problemes e-status")
```

```

lines(tapply(pe$n_exc_acab2,pe$cursfact,mean,na.rm=T),type="l",col="red")
lines(tapply(pe$n_exc_acab3,pe$cursfact,mean,na.rm=T),type="l",col="blue")
lines(tapply(pe$n_exc_acab4,pe$cursfact,mean,na.rm=T),type="l",col="green"
)
lines(tapply(pe$n_exc_acab5,pe$cursfact,mean,na.rm=T),type="l",col="orange
")
lines(tapply(pe$n_exc_acab6,pe$cursfact,mean,na.rm=T),type="l",col="purple
")
legend("topright",legend=c("Bloc 1","Bloc 2","Bloc 3","Bloc 4",
                           "Bloc 5","Bloc 6"),
pch=16,col=c("black","red","blue","green","orange","purple"))

```

```
###----- e-status
```

```
## Execucions totals que s'han fet de cada problema
# i problemes que s'han fet de cada bloc
```

```
dim(pblc)
head(pblc) # execucions per problema
```

```
colSums(pblc) # execucions per bloc
```

```
# num. de problemes que hi ha a cada bloc
colSums(pblc!=0)
```

```
#### notes d'estatus NULL (execucions obertes però no acabades)
```

```
notanull <- estatus %>% filter(is.na(nota)==T)
```

```
nrow(notanull)/nrow(estatus)# percentatge exec no acabades
sum(notanull$tiempomin==0)/nrow(notanull)# % temps=0
```

```
hist(notanull$tiempomin)
```

Script : TFG ACP

```

setwd("C:/Users/carla/Desktop/Estadística UB - UPC/TFG/Rdata")

install.packages("FactoMineR")
library(FactoMineR)

# -- Càrrega dades

load("pe.Rdata")
## var aprovat/suspes

#apr_cont<-rep(0,nrow(pe))
#for(i in 1:nrow(pe)){
#  if(pe$nota_avcont[i] >= 5 && pe$avuni[i]==0 ){
#    apr_cont[i]<-1
#  }
#}

#apr_uni<-rep(0,nrow(pe))
#for(i in 1:nrow(pe)){
#  if(pe$nota_avunica[i] >= 5 && pe$avuni[i]==1 ){
#    apr_uni[i]<-1
#  }
#}

#table(apr_uni,pe$avuni)
#table(apr_cont,pe$avuni)

#pe<-cbind(pe,apr_cont,apr_uni)

###

zeroNA <- function(df){
  for(i in 1:nrow(df)){
    for(j in 1:ncol(df)){
      if(is.na(df[i,j])==T){
        df[i,j]<-0
      }
    }
  }
}

```

```

    }
  }
}
return(df)
}
# -----

# --- ALUMNES AV. CONTINUADA
#acpdata<-pe[,c(6:11,15:21,27:32,34,35,39,40,44,45,49,50,54,55,59,60,
#           63,66,69,72,75,78,89:93)][pe$avuni==0,]
#dim(acpdata)
#acpdata<-zeroNA(acpdata)

#PCA(cbind(acpdata,pe$cuatri[pe$avuni==0]), graph=TRUE, quali.sup = 43)

# --- ALUMNES AV. UNICA
#acpdata0<-pe[,c(12:21,27:32,34,35,39,40,44,45,49,50,54,55,59,60,
#           63,66,69,72,75,78,89:93)][pe$avuni==1,]
#dim(acpdata0)
#acpdata0<-zeroNA(acpdata0)

#PCA(cbind(acpdata0,pe$cuatri[pe$avuni==1]), graph=TRUE, quali.sup = 40)

# --- ALUMNES AV. CONTINUADA
acpdata1<-pe[,c(6:11,15,85:88,95,96)][pe$avuni==0,]
dim(acpdata1)
colnames(acpdata1)
acpdata1 <- zeroNA(acpdata1)

b<-
PCA(cbind(acpdata1,cuatri=pe$cuatri[pe$avuni==0],apr=pe$apr_cont[pe$avuni=
=0]), graph=T, quali.sup = c(14,15))

plotellipses(b,axes = c(1,2))

b$eig # valors propis , variancia explicada
b$var$coord # correlació variables amb comp prin

```

```

plot.PCA(b,c(1,2),choix = "var",cex=0.7)
plot.PCA(b,c(1,3),choix = "var",cex=0.7)
plot.PCA(b,c(2,3),choix = "var",cex=0.7)

plot.PCA(b,c(1,2),choix = "ind",cex=0.7)
plot.PCA(b,c(1,3),choix = "ind",cex=0.7)
plot.PCA(b,c(2,3),choix = "ind",cex=0.7)

plot.PCA(b,c(1,2),choix = "ind",cex=0.7,habillage = 14,graph.type =
"ggplot")
plot.PCA(b,c(1,2),choix = "ind",cex=0.7,habillage = 15,graph.type =
"ggplot")
plot.PCA(b,c(1,2),choix = "ind",cex=0.7,habillage = 4,graph.type =
"ggplot")
plot.PCA(b,c(1,2),choix = "ind",cex=0.7,habillage = 3,graph.type =
"ggplot")
plot.PCA(b,c(1,2),choix = "ind",cex=0.7,habillage = 10,graph.type =
"ggplot")
plot.PCA(b,c(1,2),choix = "ind",cex=0.7,habillage = 13,graph.type =
"ggplot")

# --- ALUMNES AV. UNICA
acpdata2<-pe[,c(12:15,85:88,95,96)][pe$avuni==1,]
dim(acpdata2)
acpdata2 <- zeroNA(acpdata2)
dim(acpdata2)

a<-
PCA(cbind(acpdata2,cuatri=pe$cuatri[pe$avuni==1],apr=pe$apr_uni[pe$avuni==
1]), graph=T, quali.sup = c(11,12))

a$eig # valors propis , variancia explicada
a$var$coord # correlació variables amb comp prin

plot.PCA(a,c(1,2),choix = "var",cex=0.7)
plot.PCA(a,c(1,3),choix = "var",cex=0.7)
plot.PCA(a,c(2,3),choix = "var",cex=0.7)

plot.PCA(a,c(1,2),choix = "ind",cex=0.7)

```

```
plot.PCA(a,c(1,3),choix = "ind",cex=0.7)
```

```
plot.PCA(a,c(2,3),choix = "ind",cex=0.7)
```

```
plot.PCA(a,c(1,2),choix = "ind",cex=0.7,habillage = 11,graph.type =  
"ggplot")
```

```
plot.PCA(a,c(1,2),choix = "ind",cex=0.7,habillage = 12,graph.type =  
"ggplot")
```

```
plot.PCA(a,c(1,2),choix = "ind",cex=0.7,habillage = 2,graph.type =  
"ggplot")
```

```
plot.PCA(a,c(1,2),choix = "ind",cex=0.7,habillage = 5,graph.type =  
"ggplot")
```

```
plot.PCA(a,c(1,2),choix = "ind",cex=0.7,habillage = 6,graph.type =  
"ggplot")
```

```
plot.PCA(a,c(1,2),choix = "ind",cex=0.7,habillage = 10,graph.type =  
"ggplot")
```

ANNEX 3

