

DIPLOMATURA D'ESTADÍSTICA

SUPORT NO PRESENCIAL PER L'APRENTATGE
DE REGRESSIÓ LINEAL SIMPLE

Autor: Xavier Basagaña Flores
Directora: M. Pilar Muñoz

UNIVERSITAT POLITÈCNICA DE CATALUNYA
Biblioteca



1400313157

FACULTAT DE MATEMÀTIQUES I ESTADÍSTICA

Qualifous : N.H



Trendent
Enk Coho



Structais
M. P. by puire



Vocal
jord. Riera

Barcelona, 28 de Setembre de 1999

AGRAÏMENTS

En primer lloc, agrair a la Pilar Muñoz tots els consells i comentaris tant pel desenvolupament de l'aplicació com per la redacció del projecte.

També vull agrair a l'Erik Cobo tota la seva ajuda.

Molt especialment també vull agrair als meus pares tots els sacrificis que han hagut de fer per mi.

Finalment, també voldria agrair al Dr. Jordi Sunyer la seva insistència (i sobretot PACIÈNCIA) perquè acabés aquest Projecte Final de Carrera.

ÍNDEX

1.	INTRODUCCIÓ	6
1.1	Objectius i motivació del projecte	6
1.2	Antecedents	8
1.3	Estructura de la memòria	10
2.	DISSENY I IMPLEMENTACIÓ	12
2.1	Software de base utilitzat	12
2.2	Tipus de pantalles de l'aplicació	12
2.2.1	Disseny de les pantalles teòriques	13
2.2.2	Disseny de les pantalles de preguntes a l'usuari	14
2.3	Organització de l'aplicació	16
2.3.1	Funció de les preguntes	18
2.3.2	BLOC TEÒRIC	19
2.3.2.1	Pantalles d'introducció a la regressió	21
2.3.2.2	Pantalles "Identificació"	22
2.3.2.3	Pantalles "Estimació"	23
2.3.2.4	Pantalles "Validació"	25
2.3.2.5	Pantalles "Reidentificació"	32
2.3.3	BLOC "ANÀLISI D'UN CAS"	35
2.3.3.1	Com s'avalua a l'usuari	37
2.4	Com desplaçar-se per l'aplicació	40
2.5	Temps aproximat d'ús de l'aplicació	41
3.	TREBALL FUTUR	43
4.	CÀRREGA DEL TREBALL	46
ANNEX 1	RESOLUCIÓ DE L'ANÀLISI D'UN CAS	48
ANNEX 2	VALORACIÓ DE LA VERSIÓ BETA PER ESTUDIANTS DE L'ASSIGNATURA ESTADÍSTICA 2 (ES2) DE LA FIB	57
ANNEX 3	MANUAL D'USUARI	64
ANNEX 4	MANUAL D'INSTAL·LACIÓ	69
1	Requeriments	69
2	Instal·lació	70
3	Descripció dels arxius d'instal·lació	74
4	Execució de RLS	76
5	Desinstal·lació	77
BIBLIOGRAFIA		79

SECCIÓ 1

INTRODUCCIÓ

1. Introducció

1.1. OBJECTIUS I MOTIVACIÓ DEL PROJECTE

Aconseguir una millora en el rendiment acadèmic dels estudiants passa per poder tenir complements a les sessions de teoria, pràctiques o laboratori impartides pels professors.

En els últims anys, la introducció de les noves tecnologies en la docència de diferents matèries, i en concret de l'estadística, ha experimentat un gran creixement.

La possibilitat d'utilització de l'ordinador per part dels estudiants, ja sigui des de casa o des de la Universitat, permet que una eina de suport no presencial per a la docència de diferents matèries pugui ser utilitzada per la pràctica totalitat del col·lectiu d'alumnes. Aquest tipus de material, per tant, ha de ser considerat ja com un complement més a la docència, al mateix nivell dels que existien fins ara.

Un entorn informàtic disposa de força avantatges. Permet incorporar la interactivitat amb l'usuari, acostant l'aplicació creada a la relació professor-alumne. Així, una aplicació pot avaluar a un usuari, fer-li preguntes i intentar corregir-li els errors que comet. També li pot informar de quines parts li convé repassar i en quines altres té les coses bastant clares.

Dins d'una aplicació d'aquest tipus, l'usuari pot retrocedir en les pantalles, accedir només a determinades pàgines que li interessin, veure gràfics de dades, etc.

Una aplicació d'aquest tipus, i dissenyada per un estudiant, pot tenir l'avantatge d'incorporar un punt de vista diferent i potser més proper als altres estudiants, i d'insistir en les coses que costen més d'aprendre, de mostrar maneres diferents d'aprendre les coses, etc.

Tots aquests avantatges s'han volgut aprofitar per dissenyar en aquest projecte una eina de suport no presencial per a l'aprenentatge i consolidació de coneixements de la regressió lineal simple.

Aquesta eina va dirigida a tots els estudiants d'alguna assignatura d'estadística on es tracti el tema de la regressió.

La regressió lineal és un tema bàsic en l'Estadística, aplicable a una infinitat de camps, i que proporciona una base a partir de la qual és més fàcil entrar en models més complicats (com les sèries temporals, regressió logística, etc.). Així doncs, és important tenir molt clar el model de regressió. Hi ha dubtes usuals sobre aquest tema, com poden ser les premisses del model, com es validen els models obtinguts, etc., que són les qüestions bàsiques on l'aplicació dissenyada incideix de cara a

1. Introducció

deixar-les clares. Així, a part de definicions i explicacions més formals hi ha també explicacions de caire més intuïtiu, orientades més al moment de trobar-se amb un cas pràctic.

És cert que amb la regressió lineal simple no n'hi ha prou en la majoria dels casos reals, i convé utilitzar la regressió lineal múltiple. No obstant, la regressió lineal simple és més senzilla, i un cop apresada és molt més fàcil l'extensió de la teoria cap a la regressió lineal múltiple. És recomanable, per tant, en l'ensenyament de la regressió, començar per la lineal simple, i fer després l'extensió a la múltiple.

La present aplicació, doncs, tractarà només la regressió lineal simple. La realització d'un programa que s'extengués també a la múltiple comportaria aproximadament el doble d'hores de feina. Si aquesta aplicació és usada pels estudiants i rep una bona valoració pot plantejar-se dissenyar una aplicació de regressió lineal múltiple com a extensió/continuació d'aquesta.

L' objectiu final del programa és que al finalitzar l'aplicació l'usuari sigui capaç de crear els seus propis models de regressió lineal simple, que sàpiga validar-los i que en conegui les limitacions.

L'aplicació serà de gran utilitat a les assignatures bàsiques del departament d'Estadística i Investigació Operativa (EIO) que continguin el tema de regressió lineal simple.

1. Introducció

1.2. ANTECEDENTS

La importància i interès internacional que genera l'ús de les noves tecnologies en la docència de moltes matèries, i en concret de l'estadística, queda palesa amb la gran quantitat de persones, organitzacions i entitats que estan treballant en el tema actualment.

Algunes de les eines existents actualment per la docència d'estadística són les següents :

(extretes de Morin et al (1998). Computer assisted Training in Statistics: Internet and Multimedia: a Survey of Existing Tools. NTTS'98 Conference.)

- Software no presencial:
 - Software basat en fulles de càlcul com Excel o Lotus, que en general cobreixen els objectius d'un curs bàsic d'estadística. En aquesta categoria hi ha per exemple el paquet XLStatistics (<http://www.man.deakin.edu.au/rodneyc/>) i DISCUS (http://www.mis.coventry.ac.uk/research/discus/discus_home.html). Els dos productes treballen amb Excel.
 - Productes basats en paquets estadístics professionals, com per exemple els existents per executar-se en MINITAB (<http://www.minitab.com/resources/whitepapers/>).
 - Software específic desenvolupat per algunes determinades propostes, com per exemple:
 - STEPS (Statistical Education through Problem Solving <http://www.stats.gla.ac.uk/steps>). Aquest és un projecte de nou departaments de set universitats de Gran Bretanya amb l'objectiu de desenvolupar material per l'aprenentatge i l'ensenyança a través de solucionar exemples reals en estadística. El material produït està basat en problemes específics en àrees com la Biologia, Economia, Geografia i Psicologia.
 - QUERCUS (<http://www.stams.strath.ac.uk/external/CAL/quercus>). Aquest és un conjunt de programes tutorial interactius dissenyats per ajudar a l'aprenentatge de tècniques estadístiques bàsiques a estudiants de les àrees de biologia i ciències de la salut.

1. Introducció

- Productes basats en Internet. De la gran quantitat de material d'aquest tipus se'n llisten els següents:
 - Llibres electrònics com el desenvolupat per Jan de Leew, "Statistics. The Study of Stability in Variation" (<http://www.stat.ucla.edu/textbook>). L'objectiu dels autors és posar a Internet un llibre d'introducció a l'estadística, interactiu i d'accés gratuït.
 - Laboratoris de Probabilitat i Estadística com "Virtual Laboratories in Probability and Statistics" (<http://www.math.uah.edu/~stat/>) amb la finalitat de proporcionar mòduls interactius d'acord amb la tecnologia d'Internet per estudiants i professors de probabilitat i estadística.
 - El projecte de la Universitat de Newcastle, que ha desenvolupat una Intranet docent que inclou hipertext interactiu per usar-lo com a suport no presencial a la docència presencial que s'imparteix en aquesta universitat (<http://surfstat.newcastle.edu.au/surfstat/main/surfstat.html>)

A nivell nacional, la Universidad de Málaga ofereix un hipertext de suport al curs de Bioestadística que s'imparteix en aquesta universitat (<http://ftp.bioestadistica.uma.es/libro>).

La llista d'exemples donats no és exhaustiva, i la constant evolució d'Internet fa que apareguin nous productes ràpidament.

1. Introducció

1.3. ESTRUCTURA DE LA MEMÒRIA

En la primera secció de la memòria, la Introducció, s'expliquen els objectius i la motivació del projecte, així com també es fa un breu repàs als antecedents.

En la secció 2, Disseny i Implementació, es detalla la construcció de l'aplicació. S'expliquen en primer lloc els diferents tipus de pantalles que conté l'aplicació, i quin és el disseny que s'ha implementat en cada un d'ells. (seccions 2.2, 2.2.1 i 2.2.2)

Un cop fet això es descriu l'estructura de l'aplicació. Aquesta queda dividida en dos grans blocs, el bloc teòric i el bloc d'Anàlisi d'un cas, els quals vénen detallats en les seccions 2.3.1 i 2.3.2 respectivament.

Dins el bloc teòric s'explica com queden dividits els diferents conceptes que s'expliquen de regressió lineal, i també es fa esment a la manera d'explicar-los. Es detalla en aquesta secció el contingut d'algunes pantalles.

En el bloc d'Anàlisi d'un cas, s'expliquen com funcionen les pantalles, què s'hi pretén i s'explica també de quina manera s'avalua a l'usuari (secció 2.3.2.1)

La següent secció (2.4) explica com desplaçar-se per l'aplicació.

Finalment tanca la secció de Disseny i Implementació l'apartat 2.5, en el qual s'especifica el temps aproximat que triga un usuari en usar l'aplicació RLS. També es detalla en aquesta secció el nombre total de pantalles per cada bloc.

La secció 3 tracta del treball futur, de possibles extensions de l'aplicació i de noves aplicacions relacionades.

La secció 4 mostra la càrrega del treball.

L'Annex 1 és la resolució del cas que es proposa resoldre en l'aplicació.

En l'Annex 2, es mostren els resultats d'una enquesta passada a alumnes de l'assignatura Estadística II d'Enginyeria Informàtica de la FIB després de la utilització del software aquí presentat.

L'Annex 3 és el manual d'usuari de l'aplicació RLS.

Finalment, l'Annex 4 conté el manual d'instal·lació.

SECCIÓ 2

**DISSENY I
IMPLEMENTACIÓ**

2. Disseny i Implementació

2.1. SOFTWARE DE BASE UTILITZAT

La implementació d'una aplicació de les característiques preteses ha requerit la creació d'una interfície gràfica agradable a l'usuari.

L'aplicació ha estat creada utilitzant la versió 6.0 de Visual Basic. Aquest és un llenguatge de programació per Windows, i per tant, amb ell es poden crear aplicacions seguint l'aparença normal de les aplicacions Windows (botons, menús, quadres de text, etc.)

2.2. TIPUS DE PANTALLES EN L'APLICACIÓ

El disseny de RLS conté bàsicament dos grans tipus de pantalles. Un són les pantalles teòriques. En aquestes es mostren les explicacions dels diferents temes dels què tracta l'aplicació. S'hi inclouen definicions de termes, exemples, etc. El contingut teòric de l'aplicació s'explica en la secció 2.3.1 ("Bloc teòric") d'aquesta memòria.

L'altre gran tipus de pantalles són les de preguntes a l'usuari o d'exercicis. En aquestes es demana la participació de l'usuari. Així, es demanen a l'estudiant diferents aspectes relacionats amb el què s'ha explicat a les pantalles teòriques, si bé també n'hi ha d'altres que són simplement il·lustratives de diferents situacions.

Un exemple d'aquestes últimes és la pantalla on es poden col·locar uns punts en un gràfic per tal que l'aplicació mostri quina és la recta estimada per Mínims Quadrats (Pantalla "Introducció a la Regressió (6)"). Aquesta pantalla, per exemple, no fa cap pregunta concreta, sinó que simplement pot servir a l'usuari per veure com opera la regressió lineal (i en concret el criteri de Mínims Quadrats) en diferents situacions triades per ell mateix.

Les pantalles amb preguntes també fan l'aplicació més agradable i més entretinguda, doncs deixa de ser un conjunt d'explicacions per passar a ser una eina interactiva.

2. Disseny i Implementació

2.2.1 DISSENY DE LES PANTALLES TEÒRIQUES

El disseny bàsic de les pantalles teòriques és com el que es mostra en la figura 2.1.

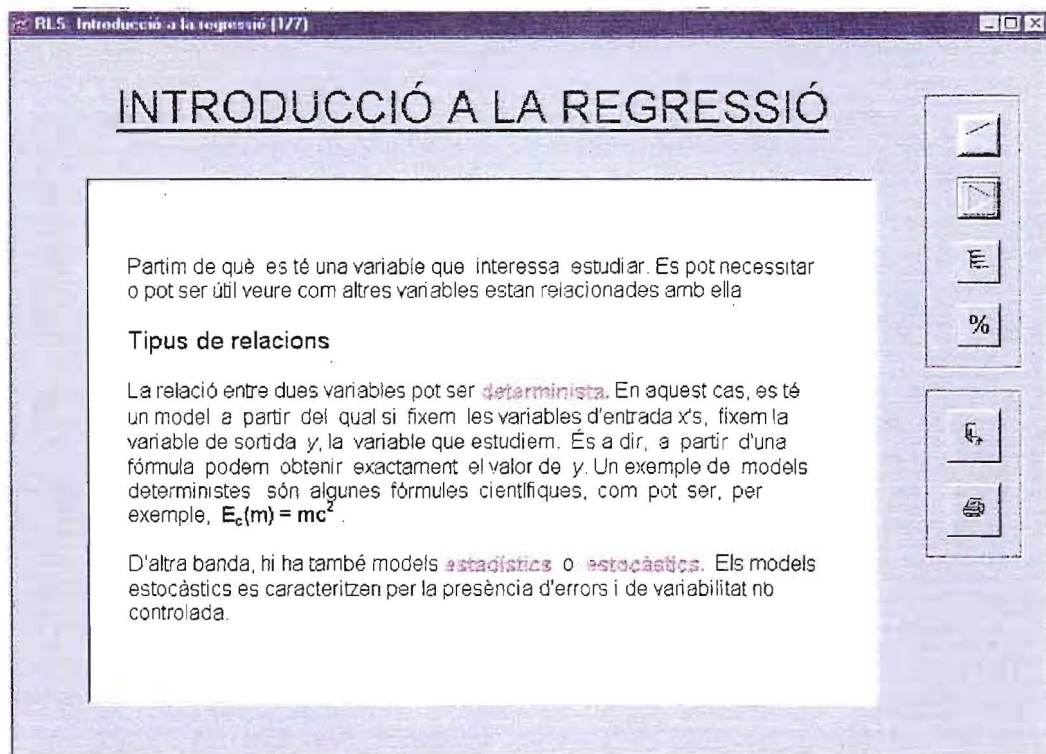


Figura 2.1. Pantalla "Introducció a la regressió (1)"

Així doncs, una pantalla teòrica es compon, generalment, d'un quadre de text i d'una barra d'eines.

En el quadre de text hi ha les explicacions pertinents segons la part que s'està tractant. El quadre de text té a sobre un títol que indica la part de l'aplicació on s'està. Les explicacions en el quadre de text s'ha intentat que siguin curtes, sense incloure "massa" informació per pantalla. A més, conté subtítols quan s'hi expliquen coses que es puguin diferenciar. D'aquesta manera l'usuari veu abans de llegir de què tracta cada paràgraf. La informació queda també d'aquesta manera més ben organitzada.

En els quadres de text també es destaquen en vermell les paraules clau. En algunes pantalles, pot haver-hi hipertext per tal d'explicar algun dels conceptes de l'explicació.

2. Disseny i Implementació

La que s'ha explicat és l'estructura més bàsica d'una pantalla teòrica. D'altres poden contenir a més del quadre de text, gràfics que l'acompanyen, i altres quadres de text que expliquen el gràfic. Els quadres de text que expliquen un gràfic tenen el fons blau, per tal de diferenciar-los dels altres on s'expliquen conceptes teòrics. Aquest tipus de pantalla queda reflexat en la figura 2.2.

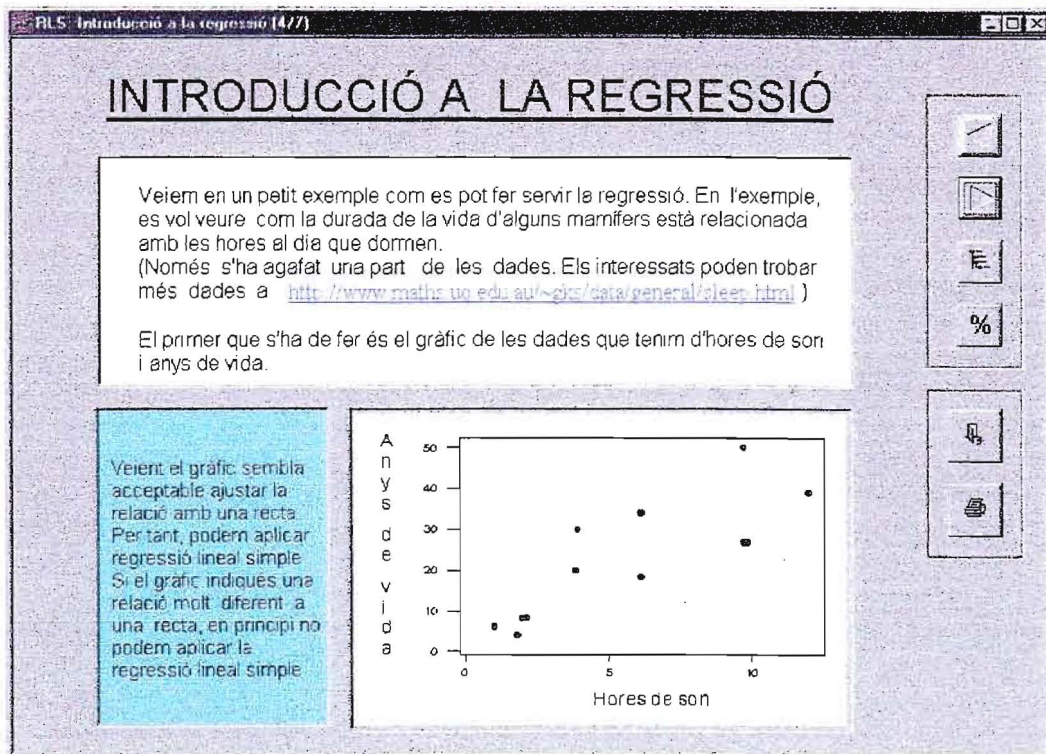


Figura 2.2. Introducció a la regressió

Les pantalles en les que es vol mostrar alguna "idea" de cara a aspectes

pràctics, tenen al començament la icona



2.2.2. DISSENY DE LES PANTALLES DE PREGUNTES A L'USUARI

En la majoria d'aquestes pantalles, l'usuari ha de contestar una pregunta que se li formula, ja sigui referent a un gràfic donat, a contingut teòric que s'ha explicat, o una pregunta que busca la intuïció de l'usuari.

2. Disseny i Implementació

Les preguntes a l'usuari són sempre preguntes tancades. En ocasions més d'una resposta és vàlida i en d'altres només una pot ser certa. En l'enunciat de la pregunta queda clar en quina situació de les dues anteriors s'està.

Així doncs, després d'una pregunta es presenten un seguit d'opcions, és a dir, un conjunt de possibles respostes. L'usuari normalment haurà de marcar les que creu que són correctes, i prémer el botó "Acceptar", per indicar a l'aplicació que ja pot avaluar la seva decisió. Un exemple de pantalla d'aquest tipus es pot veure a la figura 2.3. En aquests cas, hi ha més d'una resposta correcta.

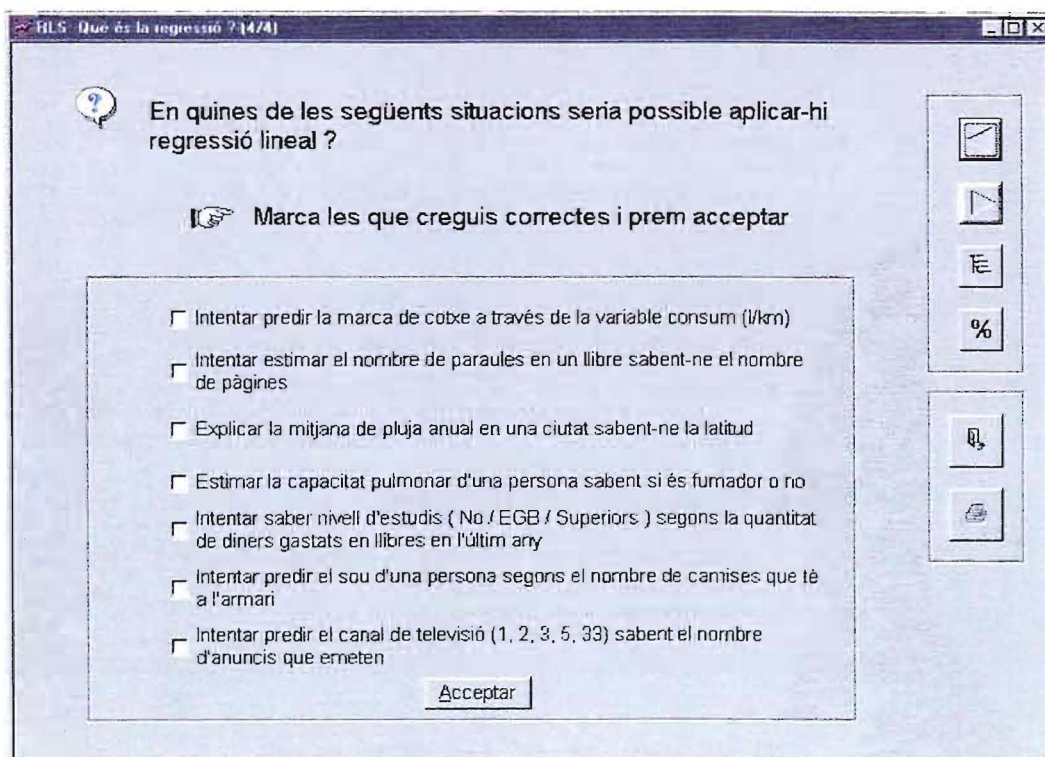


Figura 2.3. Pantalla "Què és la regressió (4)"

Les preguntes poden ser inspirades en un gràfic, poden fer referència a una sortida de Minitab quan s'ha fet un model de regressió, o simplement poden preguntar sobre un tema teòric que s'ha explicat.

Un cop es contesta una pregunta, apareix al costat de cada opció un símbol indicant si s'ha respost correctament o no. També, en un quadre de text o en un formulari emergent apareix un comentari a la solució, explicant el perquè i deixant clares les situacions que poden haver portat a confusió. En la figura 2.4 es pot veure un exemple d'una de les pantalles típiques de l'aplicació. En ella s'observa com les

2. Disseny i Implementació

preguntes fan referència als gràfics donats. En cada cas es pot contestar "Sí" o "No" a la pregunta que es formula a la part superior de la pantalla.

En la figura 2.4 es mostra la pantalla quan ja s'han contestat totes les preguntes. S'hi poden veure els símbols que indiquen que les dues primeres han estat ben contestades mentre que la última no. A més, en un quadre de text sota les opcions, s'hi pot veure l'explicació a la resposta, que, com es pot observar, s'ha donat tant per les ben contestades com per la que no ho ha estat.

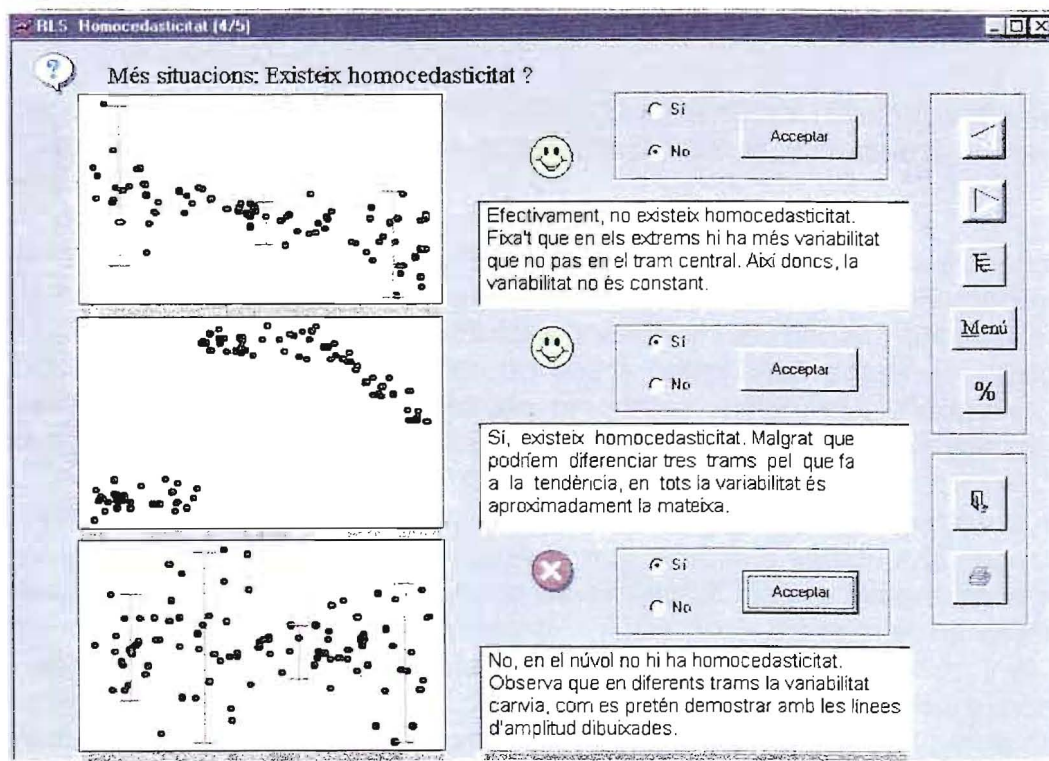


Figura 2.4. Pantalla "Homocedasticitat (4)"

2.3. ORGANITZACIÓ DE L'APLICACIÓ

L'aplicació està organitzada en dos grans blocs.

En el primer, s'expliquen tots els conceptes de la regressió lineal simple. És la part teòrica de l'aplicació o "**Bloc teòric**".

En el segon, que té per nom "**Anàlisi d'un cas**", l'usuari ha de construir un model de regressió amb unes dades concretes, i al mateix temps anar contestant les preguntes que l'aplicació li fa. Aquestes preguntes estan relacionades amb les

2. Disseny i Implementació

possibles limitacions que té el model que s'està tractant, amb quines transformacions es podria millorar-lo, etc. La part de l' "Anàlisi d'un cas" és la part on s'avalua a l'usuari, es comprova si ha adquirit bé els coneixements que pretén impartir l'aplicació, i on es recull en quins temes ha respost malament de cara a suggerir-li els temes a repassar.

Segons el tipus de pantalles que s'han descrit en l'apartat anterior, el primer bloc de l'aplicació (Bloc teòric) és clar que conté les pantalles teòriques, si bé també inclou pantalles de preguntes a l'usuari. S'han intercalat pantalles de preguntes a l'usuari en el Bloc teòric per diverses raons.

La principal raó és per motivació de l'usuari. Aquest estarà més motivat a seguir amb l'aplicació si troba, a més de tot un seguit d'explicacions, instruments per avaluar allò que aprèn.

Una altra raó és que posar tots els conceptes teòrics seguits fa que l'aplicació sigui més pesada, menys distreta. En aquest sentit, l'aplicació no es diferenciaria massa d'un llibre qualsevol que expliqués la regressió lineal, només que presentaria els continguts en format digital en lloc de sobre paper. Així doncs, la introducció de pantalles interactives com són les de preguntes a l'usuari i d'exercicis, fan que l'aplicació sigui més agradable, que no cansi a l'estudiant que l'està utilitzant.

Un altre avantatge d'incorporar pantalles interactives de preguntes a l'usuari o d'exercicis entre les pantalles teòriques és que aquestes ajuden a la consolidació de coneixements. Aspectes que poden no haver quedat clars en una explicació escrita, poden comprendre's quan, per exemple, l'usuari ha d'analitzar un gràfic donat i contesta malament a la pregunta. La solució a aquella qüestió, i el posterior comentari que mostra l'aplicació, pot ajudar a l'estudiant a veure perquè s'ha equivocat, què és el que no tenia en consideració, i en definitiva el prevé de realitzar el mateix error en una nova ocasió, la qual podria ser, per exemple, en la vida real quan està realitzant un anàlisi. Es pretén en aquest sentit que l'usuari aprengui dels seus propis errors.

Les pantalles de preguntes a l'usuari evidentment també estan presents en el segon bloc de l'aplicació, "l'Anàlisi d'un cas". De fet aquest bloc està compost exclusivament per pantalles d'aquest tipus, degut a que el seu fi és avaluar el coneixement de l'usuari i dir-li al final de la secció quines parts li convé repassar. Així, en aquest bloc es succeeixen pantalles amb sortides de Minitab i amb gràfics, amb preguntes sobre ells.

2. Disseny i Implementació

2.3.1 FUNCIO DE LES PREGUNTES

Les preguntes del bloc de teoria, fan referència a situacions on normalment poden haver-hi problemes i situacions poc clares, la solució a les quals pot deixar molt més clars els conceptes en qüestió. Així, els errors que es produeixen en el Bloc teòric no han de ser considerats avaluatius, sinó que les preguntes són només complements per assimilar la teoria. Malgrat això, els encerts a aquestes preguntes són comptabilitzats.

Les solucions i comentaris sobre les solucions a les preguntes es mostren sempre, tant si s'ha contestat bé com si no. Això és així perquè la solució pot ser molt informativa, i a l'usuari, malgrat haver encertat la resposta, pot no haver-li quedat clar del tot allò que es pretén ensenyar amb la pregunta. D'aquesta manera també s'evita que un usuari que encerti per sort una resposta no es perdi la informació útil que es transmet en el comentari de la solució. Si només es donés la solució quan s'ha respost malament, es donaria el cas que, de cara a obtenir explicacions del perquè, fos més aconsellable fallar les preguntes que encertar-les.

Les preguntes, a més, poden provocar en l'estudiant una satisfacció immediata del seu esforç, mostrant immediatament el progrés que està adquirint en el coneixement de la matèria en qüestió. Aquesta retroalimentació (feedback) facilita l'aprenentatge, i impulsa a l'estudiant a seguir treballant.

Durant tot el Bloc teòric, l'usuari té a la seva disposició el botó "puntuacions", que al prémer-lo mostra el percentatge de preguntes ben contestades en la primera ocasió que es duu en aquell moment.

En el segon bloc de l'aplicació, "l'Anàlisi d'un cas", l'objectiu de les preguntes és avaluatiu. Al final es donarà una puntuació i s'indicaran els temes a repassar segons els errors comesos. Tot i això, es continua donant un comentari a la solució, si bé és cert que és més breu que en el bloc anterior.

En la propera secció s'expliquen més a fons les estructures dels dos grans blocs de l'aplicació, el teòric i el d'anàlisi d'un cas.

2. Disseny i Implementació

2.3.2 BLOC TEÒRIC

En la "Organització de l'aplicació" (secció anterior de la memòria) ja s'ha explicat la funció del bloc teòric, que és donar els coneixements bàsics de regressió.

El què es tracta en aquesta secció és com s'estructura aquest bloc. Així, les pantalles poden agrupar-se pels diferents subtemes de la regressió que tracten.

Un esquema general de la organització dels grups de pantalles es mostra en l'esquema 1. Aquest esquema és molt semblant al que es pot trobar en l'aplicació si es fa servir el Navegador per desplaçar-se per les pantalles.

A l'hora d'implementar els conceptes teòrics en l'aplicació, aquests s'han organitzat en l'ordre que es fa servir a l'hora de crear un model de regressió, per tal que l'usuari ja quedi familiaritzat amb el procés a seguir. Així, seguint l'algoritme de creació de models que es mostra en la figura 2.5, es comença explicant la **identificació**, per continuar amb l'**estimació**, la **validació** i la **reidentificació**. Aquestes són les quatre parts més importants del Bloc teòric.

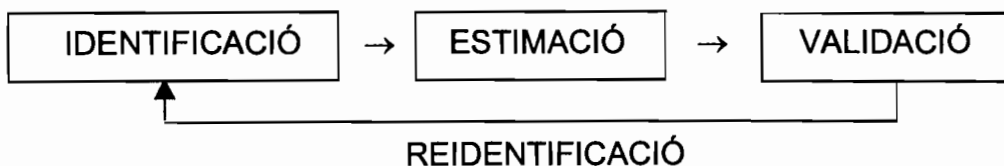
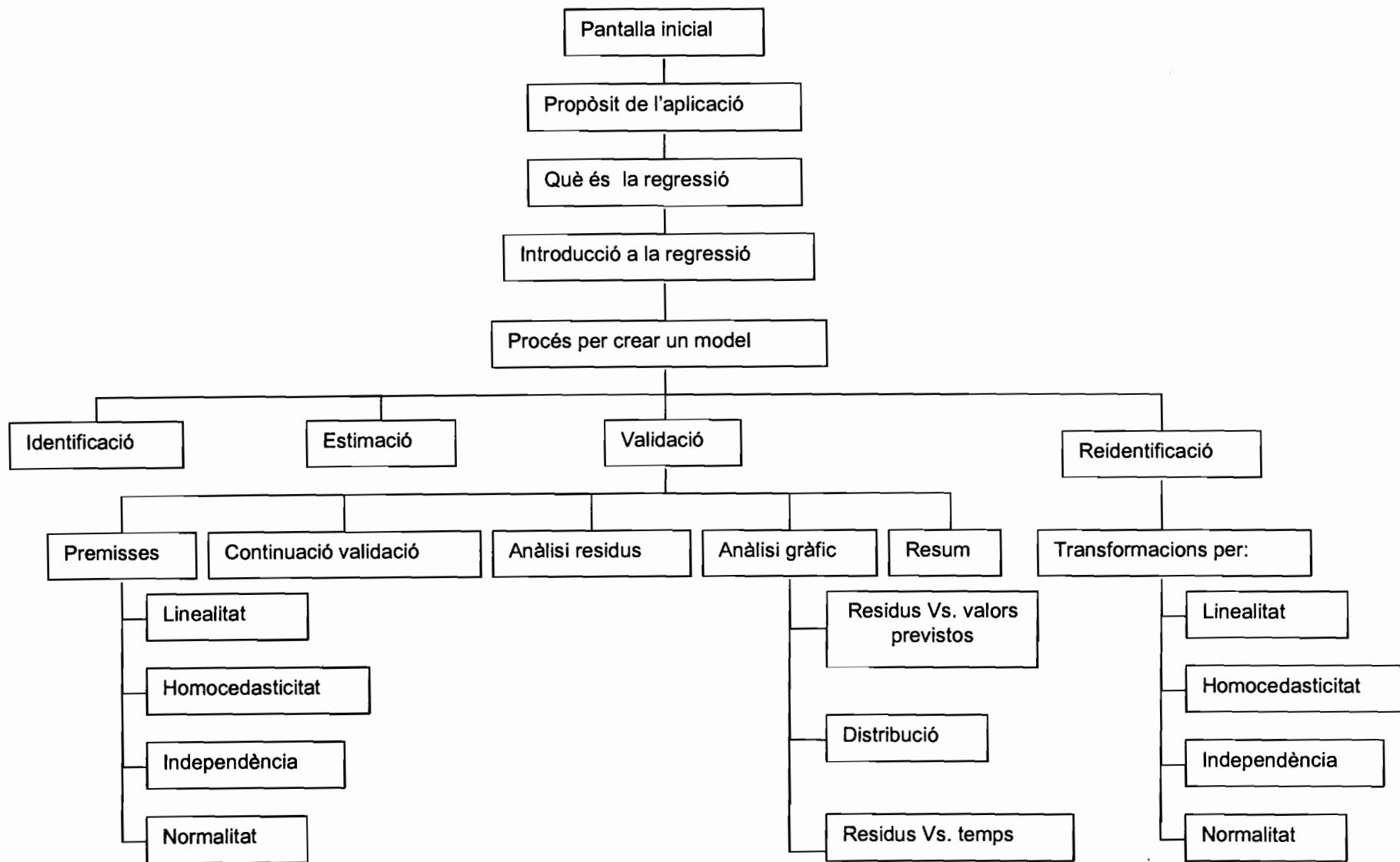


Figura 2.5. Procés de creació d'un model

Abans però d'entrar directament a l'apartat d'identificació del model, l'aplicació comença amb les pantalles d'**Introducció a la regressió**.



Esquema 1. Organització de les pantalles teòriques

2. Disseny i Implementació

2.3.2.1 PANTALLES "INTRODUCCIÓ A LA REGRESSIÓ"

Abans de procedir a explicar els passos per crear un model, s'explica què és la regressió i per què pot servir. Això és el que fan els dos primers grups de pantalles de l'aplicació, "Què és la regressió" i "Introducció a la regressió" (veure esquema 1).

En la secció "Què és la regressió" es fa una explicació intuïtiva de què és el que fa la regressió.

Per introduir l'usuari en la regressió es fa amb l'exemple de la relació entre l'alçada i el pes de les persones. Tothom pot veure que aquestes dues variables estan en general relacionades, en el sentit que les persones més altes és més probable que també pesin més que les més baixes, si bé també tothom sap que això no sempre passa.

Aquesta última és una típica relació que es pot tractar amb regressió lineal simple.

Seguint amb el mateix exemple, s'explica que es poden crear models de regressió per dues raons: per intentar explicar una relació o per predir. En l'exemple, es pot intentar explicar com es relacionen les variables pes i alçada, o bé es pot crear un model per predir el pes d'una persona si se'n coneix l'alçada.

També s'explica en aquestes pantalles quin tipus de variables es poden usar en un model de regressió. Així, s'explica que la variable resposta ha de ser no categòrica. Les variables explicatives, en canvi, poden ser contínues, dicotòmiques i fins i tot categòriques, si bé aquestes últimes necessiten un tractament especial.

Un cop acabat aquest primer grup de pantalles es passa a les pantalles "Introducció a la regressió". En aquestes es comença explicant que hi ha dos tipus de relacions entre variables, la determinista i l'estadística. La determinista és aquella mitjançant la qual donant com a entrada el valor o els valors de les variables explicatives obtenim el valor exacte de la variable resposta. Es dona un exemple de relacions d'aquest tipus, com poden ser diferents fórmules matemàtiques o científiques. En una relació no determinista (anomenada estadística o estocàstica) la relació entre dues o més variables no és exacta, hi ha error. Aquesta últim tipus de relació és la que tracta la regressió.

També s'introdueix el model de regressió lineal simple. S'explica que es pot dividir la variable resposta en una part explicada per les variables explicatives (les x's) i en una part no explicada per elles, que s'anomena "soroll".

2. Disseny i Implementació

També en aquestes pantalles s'explica la diferència entre estimador i paràmetre. En estadística freqüentista, el paràmetre és una característica intrínseca en una població, és desconegut i és el que es vol intentar estudiar. L'estimador és un estadístic mitjançant el qual obtenim informació sobre el paràmetre.

La notació que s'usa en l'aplicació per escriure els models també és presentada en aquest grup de pantalles.

Es fa la primera introducció a què és un **residu**.

En aquesta secció hi ha també una pantalla interactiva, on l'usuari pot col·locar una sèrie de punts sobre un gràfic cartesià. Un cop col·locats els punts, l'aplicació calcula la recta de regressió que donen, i la dibuixa al gràfic. L'objectiu d'aquesta pantalla és que l'usuari pugui provar diferents situacions, situacions estranyes, extremes, etc. i pugui veure de quina manera hi actua la regressió lineal simple.

Acaba la secció una pantalla on es mostren diferents avantatges i inconvenients de la regressió.

2.3.2.2. PANTALLES "IDENTIFICACIÓ"

És una de les quatre parts principals de l'aplicació. En la pantalla "Procés per construir un model" (figura 2.6) es pot entrar dins d'aquest bloc prement en el botó corresponent.

En aquest conjunt de pantalles s'explica el primer pas que es fa a l'hora de construir un model de regressió: la identificació del model. En el cas de la regressió lineal simple aquesta feina queda reduïda a trobar una variable que pugui relacionar-se linealment amb la variable resposta. Es pot comprovar si existeix una relació d'aquest tipus mitjançant un gràfic de dispersió on es creui la variable resposta amb la possible variable explicativa.

En l'aplicació s'insisteix en què s'ha de tenir en compte que si el gràfic de les dues variables no mostra un patró lineal, això no implica necessàriament que aquesta variable explicativa no sigui útil per construir un model de regressió lineal simple. Pot passar que una transformació de la variable explicativa, de la variable resposta, o d'ambdues faci que la relació sí que pugui aproximar-se per una recta.

Es mostren gràfics d'un exemple on transformant una variable aconseguim que una relació corba passi a poder-se aproximar per una recta.

2. Disseny i Implementació

També hi ha un exercici on l'usuari ha d'intentar corregir la falta de linealitat triant una transformació de les que se li proposen. Aquesta pantalla també és útil per mostrar com afectarien a les dades determinades transformacions.

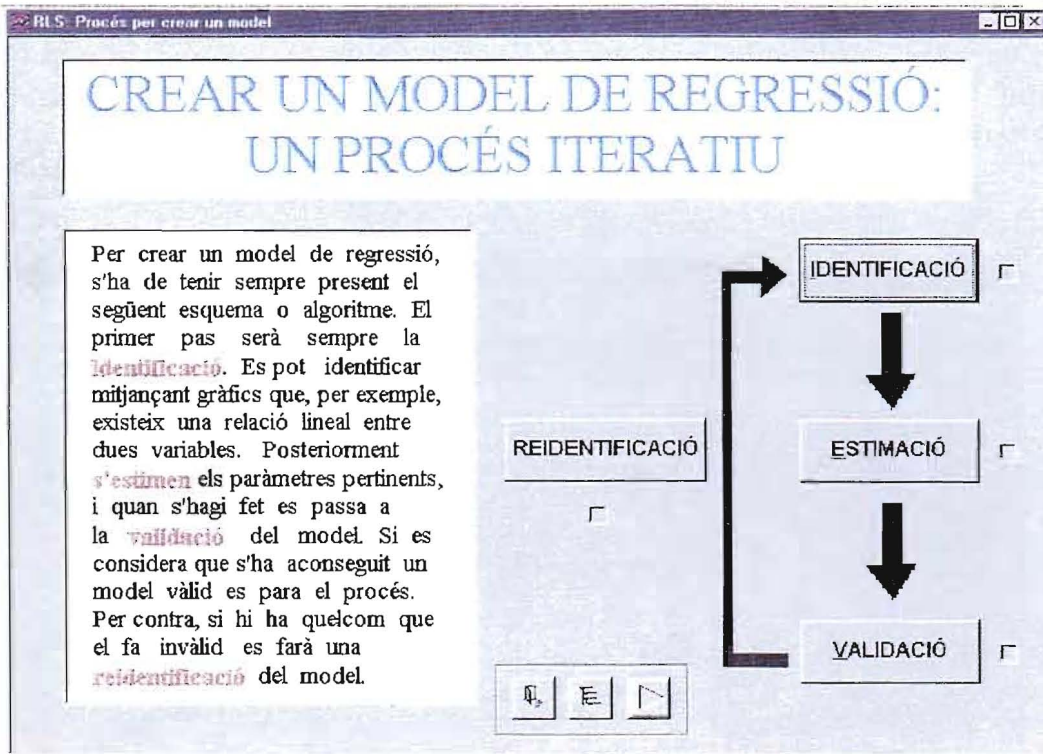


Figura 2.6. Pantalla "Procés per crear un model".

2.3.2.3. PANTALLES "ESTIMACIÓ"

El pas següent a la identificació del model és l'estimació dels seus paràmetres.

Aquestes pantalles comencen donant el model $y = \beta_0 + \beta_1 x$ amb el qual es basa la regressió lineal simple. També s'intenta fer entendre què és la pertorbació aleatòria, mostrant dos gràfics amb una relació recta entre dues variables, l'un amb pertorbació aleatòria i l'altre sense.

En les següents pantalles, sense entrar-hi massa formalment, es vol donar una idea de com s'estimen els paràmetres del model amb el mètode de mínims quadrats ordinaris.

2. Disseny i Implementació

S'explica primer què són els residus, i es fa veure que quant més petit sigui el conjunt de tots ells, millor és el model que es crea.

En un exercici, l'usuari ha de decidir amb quin o quins dels criteris que se li donen a triar les dues rectes del gràfic són igualment vàlides (figura 2.7). Amb aquest exercici es vol fer notar com el criteri de minimitzar la suma de residus és un mal criteri, ja que permet que residus positius i negatius es compensin i que, tot i sent molt grans en magnitud, la seva suma pugui ser molt propera a zero. D'aquesta manera es podria donar per bo un model que no ho fos gens, com ho és en el gràfic el model que dona la recta horitzontal.

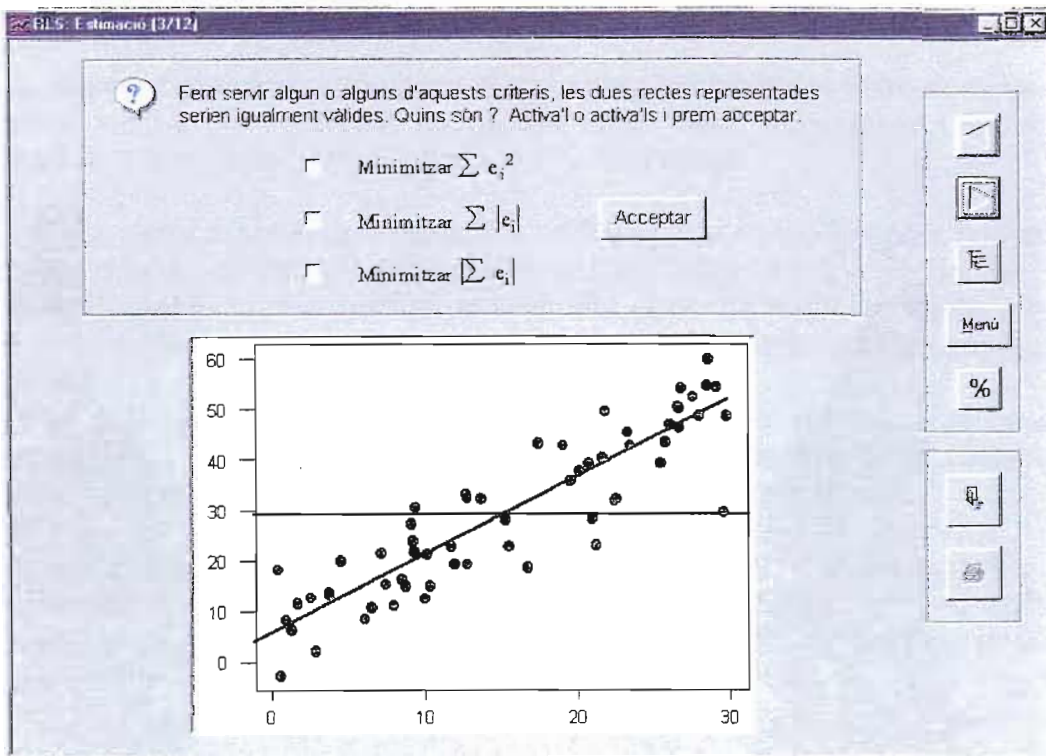


Figura 2.7. Pantalla "Estimació (3)"

En canvi, amb el criteri de mínims quadrats, o bé minimitzant la suma dels valors absoluts dels residus, aquest efecte de compensació no passa.

Es justifica fer servir el criteri de mínims quadrats en lloc del de valors absoluts per la simplicitat de les fórmules obtingudes en el primer dels dos.

En el següent pas, es donen les fórmules dels estimadors de β_0 i β_1 , i s'explica com s'ha d'interpretar cada un. β_0 és el valor de la variable resposta quan el valor de la

2. Disseny i Implementació

variable explicativa és zero. β_1 (pendent) és l'increment en la variable resposta per cada unitat que augmenta la variable explicativa.

Es posa com a exemple un model que prediu l'alçada en funció del pes, exemple que s'havia fet servir en la introducció a la regressió.

La següent pantalla proposa a l'usuari intentar ajustar una recta en un núvol de punts. Un cop ho hagi fet, podrà comprovar si el pendent que ha escollit està inclòs o no dins l'interval de confiança (al 95%). Aquesta pantalla pretén mostrar que el criteri de mínims quadrats ajusta una recta que normalment no s'allunya massa de la que hi ajustaria qualsevol persona a ull.

També en aquest apartat l'aplicació s'ocupa d'alguns resultats que s'obtenen després de l'estimació, que ajuden a fer una valoració del model.

Es fa veure en primer lloc que perquè no hi hagi relació lineal entre dues variables, el pendent ajustat per regressió ha de ser zero. Quan això passa una variable no aporta informació sobre l'altre, com a mínim linealment.

Es vol que l'usuari noti que un pendent de 0.05 pot ser suficientment llunyà respecte al zero, i que un de 100 pot ser insignificant, tot depèn de l'escala de les variables i de la seva variància. Per avaluar si realment el pendent que dona la regressió és zero o no, es fa servir el p-valor. S'explica com s'ha d'interpretar aquest estadístic.

Un altre estadístic molt important de cara a avaluar el model és el coeficient de determinació, o R^2 , que dona informació de quin percentatge de variabilitat de la variable resposta es pot explicar gràcies a la variable explicativa.

Tant pel coeficient de determinació com pel p-valor s'anomenen les diferents nomenclatures que tenen, de cara a poder-los identificar en les sortides que ofereixen els paquets estadístics. En l'aplicació es dona una part de la sortida que dona el paquet estadístic MINITAB quan es fa una regressió.

2.3.2.4. PANTALLES "VALIDACIÓ"

Un cop s'ha identificat un model i se n'han estimat els paràmetres, cal validar-lo. Per tant, aquesta és la següent secció de l'aplicació.

La part de validació és, juntament amb la part de reidentificació, la part en la qual es centra més l'aplicació. És aquesta una part crucial en la construcció d'un bon model, i moltes vegades es deixa de banda i no es duu a terme. Així doncs, és també un objectiu de l'aplicació deixar clar que aquest pas s'ha de fer sempre que es construeix un model, i és el que ens dirà si el model escollit és vàlid o no.

2. Disseny i Implementació

L'apartat de validació del model és també una de les que pot induir a més confusió i que pot quedar menys clara als estudiants d'estadística. Poden ser difícils d'interpretar, per exemple, els gràfics de residus.

El què es pretén bàsicament en aquesta secció és que l'usuari sigui capaç de validar models, mostrant-li les diferents eines que hi ha per fer-ho (bàsicament les eines gràfiques), i ensenyant-li què és el què ha d'exigir a cada gràfic, què hi ha de buscar que li indiqui que es compleix una premissa o no, com hauria de ser cada gràfic en el cas ideal, què implica trobar-hi segons quines desviacions, etc.

Abans de poder fer la validació d'un model és necessari conèixer-ne les premisses. Així doncs, en la primera pantalla de la secció de validació (figura 2.8) es mostren diferents botons que l'usuari pot triar per tal de veure les explicacions de cada una de les premisses del model de regressió (linealitat, homocedasticitat, independència i normalitat). Si ja s'han llegit les premisses o no es volen passar aquestes pantalles es pot passar a les pantalles següents, si bé és bastant recomanable veure-les.

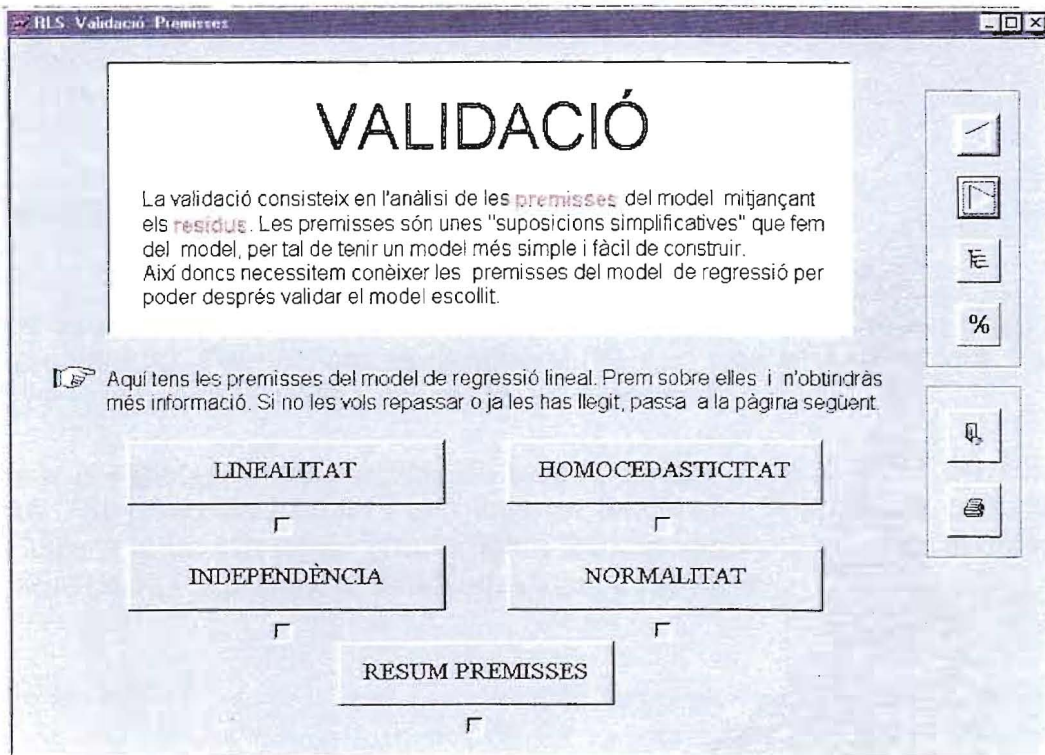


Figura 2.8. Validació: Premisses.

2. Disseny i Implementació

- **Pantalles “Linealitat”**

S'explica en aquest conjunt de pantalles què diu la premissa de linealitat, i com s'ha d'interpretar.

La premissa s'expressa d'una manera formal, $E[Y|X] = \beta_0 + \beta_1 x$ ó $E[\varepsilon_i] = 0$, és a dir, que l'esperança de la resposta depèn linealment de la variable explicativa X . També però s'intenta fer-la més intuïtiva. Un dels consells que es dóna és intentar dibuixar per diferents trams de les X el valor mig de les Y , i després unir-los. Si s'aconsegueix que aquests punts formin més o menys una recta, la premissa de linealitat es compleix.

Un cop explicats aquests punts, les dues pantalles següents mostren diferents gràfics de Y Vs. X , i l'usuari ha de marcar si hi ha linealitat o no. El gràfics que es mostren inclouen situacions que *incompleixen altres premisses* (com per exemple la d'homocedasticitat) per tal que l'usuari, encara que vegi que el gràfic no és adequat, sigui capaç d'identificar si es compleix la premissa concreta per la qual se li pregunta, en aquest cas la de linealitat. Les solucions, que es donen després d'haver contestat tant si s'ha respost bé com si no, poden ajudar a clarificar aquestes situacions a l'usuari.

- **Pantalles “Homocedasticitat”**

També s'expressa la premissa d'homocedasticitat de manera formal i d'una manera més entenedora. Formalment es diu que $V(Y) = \sigma^2$, on σ^2 és constant. També és usual dir que l'error és additiu, no depèn del valor de X .

De cara a aspectes més pràctics, es mostra un gràfic de Y Vs. X on s'hi intenta dibuixar l'amplitud dels punts i una línia de tendència. Si les línies d'amplitud són paral·leles a la de tendència (veure figura 2.9), la variància no creix ni decreix amb els valors de X , i la premissa homocedasticitat es compleix.

2. Disseny i Implementació

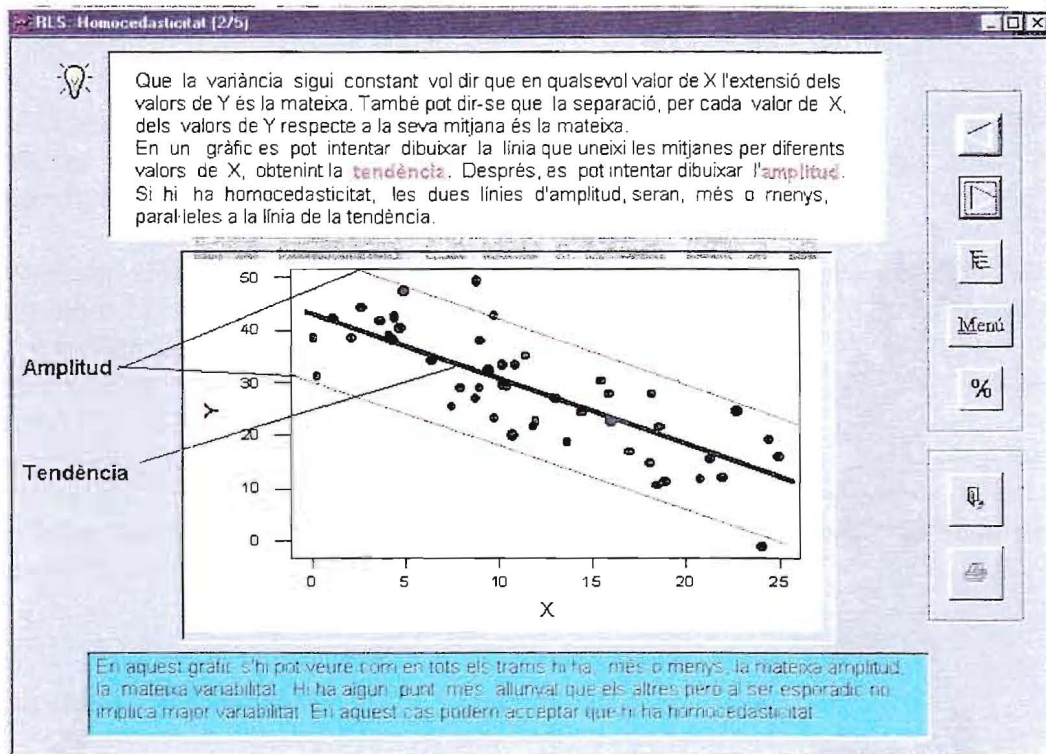


Figura 2.9. Pantalla "Homocedasticitat (2)"

Seguidament a aquestes explicacions, les tres pantalles següents són de preguntes a l'usuari.

En les dues primeres, a l'igual que ja s'ha fet a les pantalles de linealitat, es mostren diferents gràfics de Y Vs. X i l'usuari ha de contestar si es compleix la homocedasticitat o no. Com és habitual en tota l'aplicació, les solucions ajuden a clarificar el perquè de la resposta correcta.

La tercera té un format diferent. En ella es vol que l'usuari pugui veure com es pot traduir la homocedasticitat a la realitat, ja que aquest pot semblar un concepte una mica abstracte. Així, en la pantalla "Homocedasticitat (5)" es mostren quatre gràfics, on hi ha quatre alteracions diferents de la variància. A la seva dreta hi ha quatre requadres on hi ha escrites aquestes quatre alteracions, i quatre requadres més on hi ha possibles situacions que podrien explicar aquestes alteracions en la variància. El què ha de fer l'usuari és col·locar al costat de cada gràfic el requadre corresponent al tipus d'alteració que s'hi mostra i el requadre de la possible causa.

Les situacions que mostra aquesta pantalla estan relacionades amb una indústria que fabrica unes certes peces, i amb algunes situacions que la poden afectar, com lloguer de nous operaris sense experiència, canvis en la matèria prima, ...

2. Disseny i Implementació

- **Pantalles “Independència”**

La premissa expressada formalment ens diu que $E(\varepsilon_i \cdot \varepsilon_j) = 0$, quan $E(\varepsilon_i) = 0$, o el què és el mateix, que la covariància entre les pertorbacions és zero, que són independents.

Es mostren algunes situacions on si s'observa el gràfic dels residus Vs. l'ordre d'obtenció de les observacions es veu que no hi ha independència, sinó que un error porta informació de com serà el següent. Hi ha exemples de correlació positiva i negativa entre residus, i s'expliquen situacions reals on es pot donar aquest fenomen.

Aquest bloc de pantalles també n'inclou una de preguntes a l'usuari, on es mostren tres gràfics de residus Vs. temps (o ordre de recollida), i s'ha d'identificar si existeix independència o no.

- **Pantalles “Normalitat”**

En aquestes pantalles s'explica què és el concepte de Normalitat i com s'avalua. La premissa de Normalitat fa referència a la distribució dels errors. Aquests han de tenir una distribució que tingui la forma típica de la “campana de Gauss”.

En la primera pantalla es mostra una corba normal (o campana de Gauss) i se n'expliquen breument les seves característiques bàsiques (major concentració al centre i menys a les cues, funció més o menys allisada depenent de la variància,...).

La premissa de Normalitat és conseqüència del Teorema Central del Límit. El què es suposa al fer el model de regressió és que les pertorbacions no poden assignar-se a causes concretes, sinó que són l'efecte agregat de molts factors diferents, cap d'ells predominant. Suposant això, el Teorema Central del Límit ens diu que la distribució de les pertorbacions tendirà a ser Normal.

Per avaluar la Normalitat es pot fer gràficament amb dos tipus de gràfics, l'histograma i el Gràfic Probabilístic Normal.

S'expliquen com es construeixen els dos gràfics, i com han de ser per tal de poder acceptar que es compleix la Normalitat. Pels dos tipus de gràfics hi ha pantalles de preguntes a l'usuari, on donats uns gràfics l'usuari ha de contestar si creu que es compleix la Normalitat o no.

2. Disseny i Implementació

Hi ha també unes pantalles anomenades "Resum de les premisses". En aquestes s'expliquen breument les premisses, podent-les tenir totes a l'hora en poques pantalles. Primer es dona la definició de les premisses en funció de la variable resposta, i després en funció de les pertorbacions.

S'explica al final que les premisses no són l'objectiu de l'anàlisi, ja que són una condició suficient però no necessària. És per això que es poden acceptar petites desviacions de les premisses. També es recorda que les premisses del model fan referència als paràmetres (valors poblacionals) i no als valors mostrals que són dels que disposem. És per això que en el procés de validació de residus ha d'intervenir-hi una certa inferència. Aquesta es durà a terme bàsicament amb mètodes gràfics.

- **Pantalles "Anàlisi de residus"**

Havent acabat ja amb l'explicació de les premisses es continua explicant l'anàlisi de residus.

L'anàlisi de residus és el que es fa servir per tal de validar el model. En aquestes pantalles s'insisteix novament en què són els residus (la part de la resposta que no ve explicada pel model), i què pot aportar fer l'anàlisi de residus.

L'anàlisi de residus es realitza bàsicament de forma gràfica. En l'esquema 1, on es mostrava l'estructura de les pantalles teòriques, aquests mètodes queden englobats en les pantalles "Anàlisi gràfic".

- **Gràfic de residus Vs. valors previstos**

El primer gràfic que s'explica en aquesta secció és el gràfic de residus Vs. valors previstos. Aquest és el gràfic bàsic de l'anàlisi de residus. És un gràfic que sempre s'ha de fer, i que pot permetre descobrir situacions de falta de linealitat, situacions d'heterocedasticitat, valors atípics, i en algunes ocasions dependència.

S'explica què són els residus estandarditzats, els quals es fan servir generalment per construir aquest gràfic, així com d'altres tipus de gràfics de residus. Els residus estandarditzats, en cas de complir-se la hipòtesi de normalitat, seran una mostra d'una Normal (0,1).

2. Disseny i Implementació

Tenint els residus estandarditzats, s'aconsella dibuixar en aquest gràfic línies horitzontals als valors de 0, 2, 3, -2 i -3, que seran molt útils de cara a veure si tenim uns bons residus o no. Amb la línia al zero és més fàcil comprovar si es compleix la linealitat, com s'explica més endavant.

En les pantalles següents s'explica què s'ha d'exigir a aquest gràfic per tal de què es compleixin les premisses. S'introdueix el concepte de "soroll blanc".

Per veure si es compleix la linealitat, en el gràfic s'ha de veure si per tots els trams dels valors previstos la mitjana dels residus és zero. Si hi ha trams on tots els residus estan per sobre el zero i altres on estan tots per sota, la linealitat no es compleix.

Per veure si hi ha homocedasticitat s'ha de comprovar que la variabilitat, o l'amplitud dels punts és la mateixa per tots els trams de valors previstos.

Per comprovar la independència, els residus representats en el gràfic no han de seguir cap patró concret.

També en aquest gràfic es podrien detectar algunes situacions on no es compleix la normalitat. Si es compleix, els residus (estandarditzats) han de ser simètrics respecte al zero, el 95% dels punts hauran d'estar entre -2 i 2, i quasi bé cap ha d'estar fora de l'interval [-3,3]. Això és així perquè si féssim un interval de confiança al 95% d'una Normal (0,1) tindria aproximadament els límits [-2,2], i al 99.7% serien [-3,3]. També són fàcils de detectar en aquest gràfic els valors atípics.

En la pantalla següent, l'usuari ha de contestar si dona un residus com a bons en funció dels gràfics de residus Vs. valors previstos que es presenten de la forma habitual.

- **Gràfics de distribució**

En aquesta secció es recorden breument els gràfics per avaluar Normalitat, que ja s'han descrit en la secció de descripció de les premisses.

Hi ha també una pantalla de preguntes a l'usuari, on aquest ha de contestar si existeix normalitat en els gràfics que se li presenten.

2. Disseny i Implementació

- **Gràfics de residus Vs. temps**

Aquest és un gràfic que es fa servir quan l'ordre amb el qual han estat recollides les observacions té importància, i ens permet avaluar la independència dels residus. Es mostra un gràfic on no hi ha independència.

En la següent pantalla, es mostren tres gràfics de les mateixes dades: el gràfic de Y Vs. X, el gràfic de residus Vs. valors previstos i el gràfic de residus Vs. temps. Observant aquests gràfics, l'usuari ha de contestar si les afirmacions que se li presenten són certes o falses.

Per acabar amb la secció de validació, es fa un resum del què s'ha vist fins al moment referent a aquest tema, i es mostra novament el gràfic iteratiu de construcció d'un model de regressió. En aquests moments, a l'estar les pantalles ordenades tal i com si s'estigués fent un model de regressió, l'usuari pot veure que ha acabat les tres grans parts, i que a partir d'aquí, si el model no és bo (no passa la fase de validació), s'ha de passar a fer la reidentificació. Altrament, el model que s'ha provat es dona per bo, i per tant ha acabat el procés.

2.3.2.5. REIDENTIFICACIÓ

Per tal de fer que es compleixin les premisses es pot recórrer a transformacions de les variables. En la primera pantalla d'aquesta secció es mostren quatre botons corresponents a transformacions de cara a aconseguir cada una de les quatre premisses (figura 2.10).

2. Disseny i Implementació

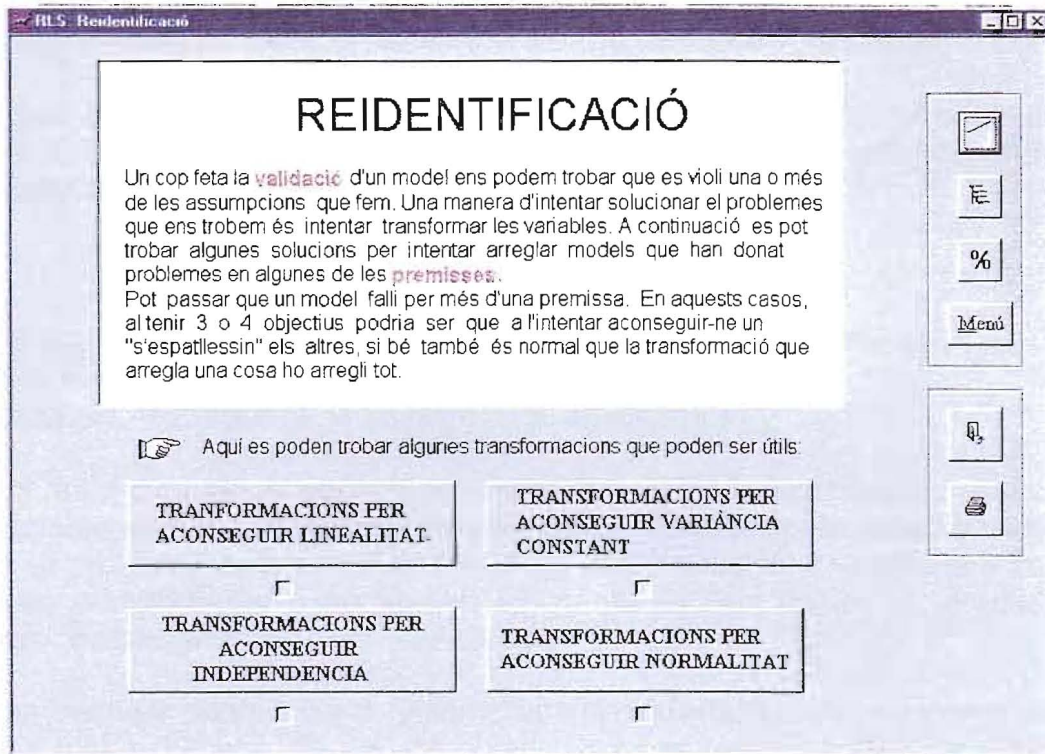


Figura 2.10. Reidentificació (1)

- Pantalles "Transformacions per aconseguir linealitat"

En la primera pantalla d'aquesta secció s'explica el concepte de "lineal". Es donen diferents exemples de funcions, de cara a deixar clar aquest concepte abans de parlar de cap transformació.

Donar una transformació que aconseguixi linealitzar les dades es pot fer tant amb criteris teòrics com examinant el gràfic de Y Vs. X. Es donen exemples dels dos casos.

En l'última pantalla, es mostra un gràfic on la linealitat no es compleix. L'usuari ha de triar una transformació de les que se li donen per tal d'aconseguir linealitzar la relació. Aquesta pantalla també és útil per veure com afecten determinades transformacions a un gràfic d'aquest tipus.

S'insisteix també en què és difícil de trobar "a la primera" una transformació que arregli les dades. Amb experiència es poden arribar a conèixer una mica les transformacions més adients. A la pràctica però, moltes vegades es proven les transformacions més usuals per veure si alguna ens soluciona el problema.

2. Disseny i Implementació

- **Pantalles "Transformacions per aconseguir variància constant"**

En aquestes pantalles es dona una família de transformacions que solucionen la falta de variabilitat constant quan la variància creix o decreix amb l'esperança de la següent manera:

$$\sigma_Y = k \cdot E(Y|X)^p$$

on $E(Y|X)$ és l'esperança de Y (variable resposta) condicionada a X (variable explicativa), i k i p són constants.

En aquests casos doncs, la transformació adequada és $y' = y^{1-p}$.

Es fa notar també que per solucionar problemes d'heterocedasticitat cal transformar la variable resposta. No serveix de res intentar transformar la variable explicativa, ja que el problema està en la distància de l'eix de les Y . Transformant l'explicativa només compactarem o separarem els punts en l'eix de les X , deixant igual la heterocedasticitat.

En la següent pantalla està desenvolupada aquesta transformació per algun dels casos més habituals. Es vol que quedi clar en aquesta pantalla que si la variància creix amb l'esperança, les transformacions $y' = y^p$ han de tenir $p < 1$. Altrament, la variància seria encara més gran en els valors més alts en relació amb la resta.

L'última pantalla mostra com en un cas particular, totes les transformacions $y' = y^p$ amb $p < 1$ redueixen l'heterocedasticitat.

- **Pantalla "Transformacions per aconseguir independència"**

Les situacions de no independència s'han de tractar amb models de sèries temporals o afegint variables noves (regressió lineal múltiple). Aquests temes no entren en el propòsit de l'aplicació, i per tant només s'informa que per corregir aquestes situacions s'han de fer servir alguna d'aquestes vies.

- **Transformacions per aconseguir normalitat**

En aquesta secció es parla de situacions on la distribució dels residus no és simètrica. En aquests casos (com en d'altres) es pot fer servir la família de transformacions de potències i arrels. Aquestes transformacions són les del tipus $y' = y^p$, on p típicament pren els valors -2 , -1 , $-\frac{1}{2}$, $\frac{1}{2}$, 2 , o 3 , encara que en pot prendre d'altres.

2. Disseny i Implementació

Les transformacions tenen diferents efectes si p és major o menor que 1 (1 representa la no transformació), i quan més allunyada estigui p de 1 més forta serà la transformació (serà més adequada per situacions amb una asimetria molt accentuada).

Les transformacions amb $p < 1$ "ajunten" més a la resta els valors més grans de Y . Són adequades quan tenim una asimetria aplanada a la dreta. Contràriament, les transformacions amb $p > 1$ "separen" més de la resta les observacions amb valors més grans de Y . És doncs una transformació adequada per situacions d'asimetria aplanada a l'esquerra.

Les dues últimes pantalles d'aquesta secció posen en pràctica aquests coneixements. L'usuari haurà de triar quina és la transformació més adient donat un gràfic de distribució concret.

2.3.3. BLOC "ANÀLISI D'UN CAS"

Després d'haver passat el bloc teòric, l'aplicació passa a la secció "Anàlisi d'un cas". Malgrat que l'usuari ja ha contestat varies preguntes durant la secció anterior, i ha rebut ja una primera puntuació segons les seves respostes, la part pròpiament avaluativa és l'anàlisi d'un cas pràctic.

Les anteriors respostes (les de l'apartat teòric) podien ajudar a assimilar els diferents conceptes que s'explicaven en cada secció. En aquests moments, l'usuari ja ha passat totes les pantalles teòriques, i per tant és ara quan es pot avaluar el seu coneixement.

Aquesta secció parteix d'un conjunt de dades referents a quaranta-set estats dels Estats Units durant l'any 1960, que són : taxa de crims, el nombre d'habitants en cada estat, una variable que mesura la riquesa de l'estat, i la despesa policial de l'estat. Les dades són extretes d'Internet, a l'adreça :

<http://www.maths.uq.edu.au/~gks/data/general/uscrime.html>

Treballar amb unes dades no simulades, com ho han estat per exemple en el bloc de teoria, fa el problema més interessant, ja que es veu com un model pot funcionar a la vida real. També cal dir que en aquestes dades les situacions de normalitat, linealitat, etc. no són tant clares com en l'apartat teòric, on s'havien dissenyat situacions on quedés molt clar si es complien o no.

2. Disseny i Implementació

En aquest bloc l'usuari haurà de ser qui construeixi un model de regressió per explicar la taxa de crims d'un estat. De fet, es demana construir un model de regressió lineal simple per cada una de les tres variables restants. L'Annex 1 d'aquesta memòria conté la resolució a aquest anàlisi.

En la pantalla inicial es pot decidir amb quina variable fer l'anàlisi. Per fer l'anàlisi amb cada una caldrà prémer el botó corresponent, com es mostra en la figura 2.11.

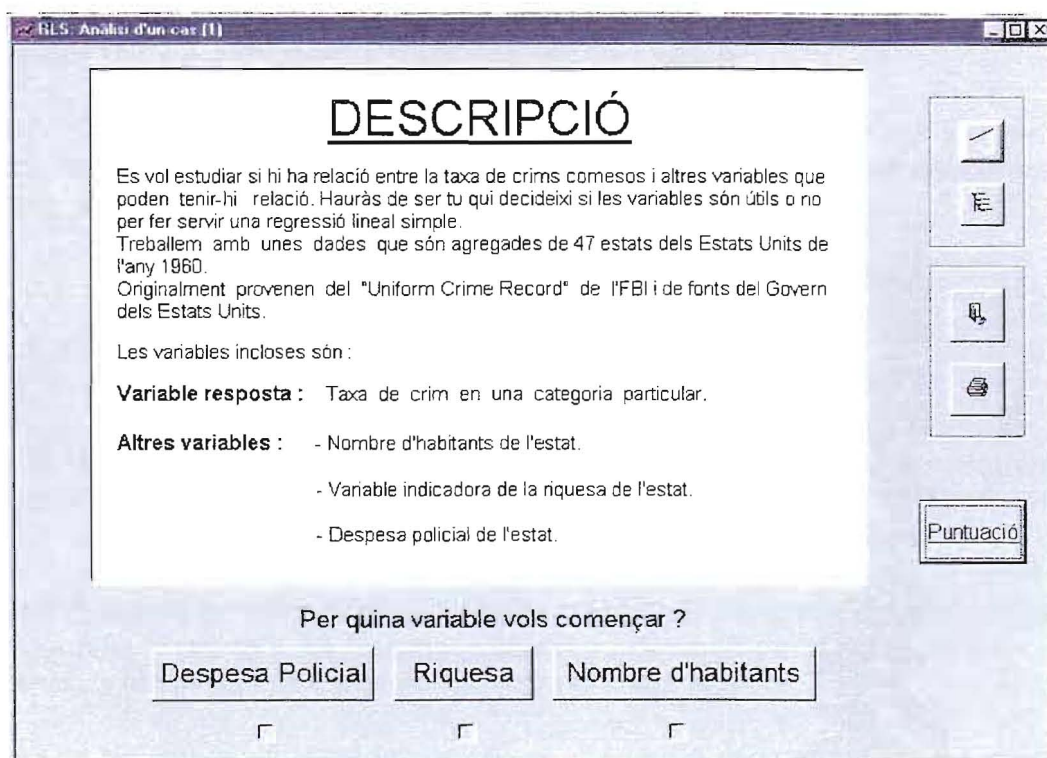


Figura 2.11. Pantalla "Anàlisi d'un cas (1)"

L'anàlisi amb una variable pot fer-se més d'una vegada, si bé de cara a la puntuació final només comptaran les preguntes ben i mal contestades en la primera ocasió.

Un cop s'ha triat una variable, les pantalles mostren sortides de regressió de MINITAB o bé diferents gràfics, i a partir d'això es formulen preguntes a l'usuari de cara a valorar el model que s'està estimant i de cara a triar el proper pas a seguir.

En les sortides de MINITAB estan destacats en vermell els valors que ha de mirar l'usuari, que són els que s'han explicat en l'aplicació. D'aquesta manera l'usuari no pot confondre's amb valors que potser no coneix, com els "t-ratio", la "s" o la "Stdev" dels coeficients. Els valors en els quals s'ha de fixar són els coeficients estimats, els p-valors, el coeficient de determinació (R^2) i el coeficient de determinació ajustat (R^2 adj.).

2. Disseny i Implementació

Les preguntes que es fan referents a les sortides incideixen bàsicament en :

- la interpretació del model
- la variància explicada pel model
- la significació dels coeficients

També hi ha preguntes que es refereixen a gràfics. Es mostren gràfics de Y Vs. X, de residus Vs. valors previstos i histogrames de residus.

A partir d'aquests gràfics l'usuari ha de contestar si es compleixen les premisses, quines fallen, quina transformació podria arreglar un model que no compleixi les hipòtesi, etc.

2.3.3.1. COM S'AVALUA A L'USUARI

Com ja s'ha dit, en el primer bloc d'explicacions teòriques s'hi inclouen també preguntes a l'usuari i exercicis. Malgrat que en aquesta part l'objectiu principal no és avaluar, es porta un recompte de les preguntes ben i mal contestades.

Degut a que les pantalles es poden visitar en més d'una ocasió (també les pantalles de preguntes a l'usuari), el recompte de preguntes ben i mal contestades només fa referència a les ben i mal contestades en la primera ocasió.

Al final de les pantalles teòriques es dona el nombre de preguntes ben contestades (en la primera oportunitat), el total de contestades i el percentatge d'encert. Aquesta informació, així i tot, és accessible quan encara no s'ha acabat la secció, mitjançant el botó "puntuació" que apareix en cada pantalla. Aquest botó mostra la informació que es pot veure en la figura 2.12.

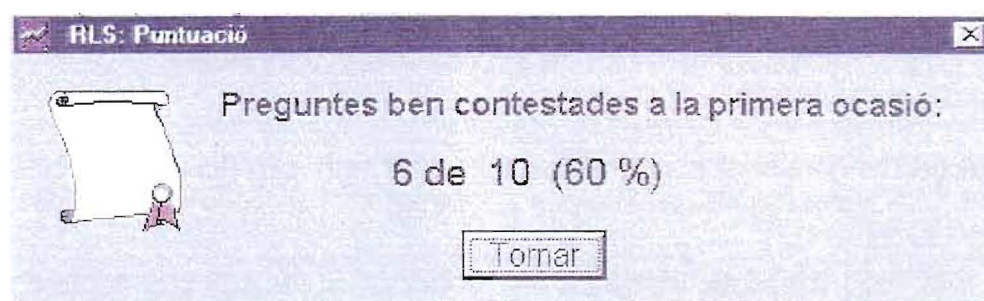


Figura 2.12. Puntuació

2. Disseny i Implementació

El percentatge es calcula només amb les preguntes contestades, i no amb el total de preguntes que conté l'aplicació. Així, si només se'n contesten dues el percentatge es calcula segons els encerts en aquestes dues.

El total de preguntes incloses en l'aplicació és de 48 en el bloc teòric i de 24 més en l'Anàlisi d'un cas, fent així un total de 72 possibles preguntes a contestar.

En la figura 2.13 es pot veure una de les pantalles. En aquesta, segons la sortida que es dona s'han de marcar les respostes correctes. Al poder-hi haver més d'una resposta correcta, de cara a la puntuació es compta un punt per ítem, de manera que si es contestessin totes bé serien quatre punts sobre quatre possibles.

Anem a estudiar la relació entre la taxa de crims i el PNB per càpita en els diferents estats.

Observa el resultat de fer la regressió lineal simple entre la taxa de crims i el PNB

Regression Analysis

The regression equation is
 $crime = -24 - 1.71 PNB$

Predictor	Coef	Stdev	t-ratio	p
Constant	-24.0	286.3	-0.08	0.933
PNB	-1.769	0.5362	3.30	0.002

s = 350.9 R-sq = 24.54 R-sq(adj) = 17.74

Tenint en compte la sortida (només es necessita allò destacat en vermell), tria les opcions correctes i prem acceptar.

- 1) La variable PNB explica una part de la resposta, i és estadísticament significativa
- 2) Aquest model no pot ser donat com a bo perquè el terme constant no és gens significatiu
- 3) La variable explica aproximadament un 15-20 % de la variabilitat de la resposta
- 4) Són els estats més pobres els que tenen més taxa de crims.

Acceptar

Figura 2.13. Anàlisi d'un cas amb la variable Riquesa.

L'avaluació pròpiament dita que fa l'aplicació es duu a terme en el segon bloc, en "l'Anàlisi d'un cas". La figura 2.13 pertany a aquest grup de pantalles.

Aquí se suposa que l'usuari ja ha passat les pantalles de teoria, i per tant és capaç de passar un "test" sobre el què ha après.

2. Disseny i Implementació

Les preguntes ben i mal contestades en la primera ocasió queden emmagatzemades, tal i com passava en l'apartat teòric. Al final, quan l'usuari ha acabat, arriba a la pantalla de puntuacions finals, on es troba el percentatge de ben i mal contestades en el cas pràctic, el percentatge per les preguntes de la part de pantalles teòriques, i un percentatge global que és la unió dels dos.

Cal recordar que les preguntes de les pantalles teòriques són més a títol il·lustratiu, i per això es mostren separatament les dues puntuacions. A més, les primeres preguntes poden haver ajudat en l'aprenentatge, i a que l'usuari no repeteixi després en el cas pràctic els mateixos errors.

No obstant, també es dóna el promig total de ben contestades a la primera oportunitat.

Durant el cas pràctic s'emmagatzemen les respostes correctes i incorrectes per tal de donar una puntuació al final, però també es recull per cada resposta errada de quina temàtica tractava la pregunta. Al final de l'anàlisi del cas i després d'obtenir les puntuacions, es suggereix a l'usuari quins temes s'ha de repassar perquè són els que ha errat en les preguntes.

Els temes amb els que queden englobades les respostes són:

- el p-valor
- l'estadístic R^2
- interpretació dels coeficients
- linealitat
- homocedasticitat
- normalitat
- independència
- outliers
- transformacions per aconseguir homocedasticitat
- transformacions per aconseguir normalitat
- temes agrupats en les consideracions finals

2. Disseny i Implementació

Haver fallat una resposta referent a un tema ja implica que l'aplicació suggerirà que es repassi. Només si no s'ha errat cap qüestió en el cas pràctic l'aplicació no en suggerirà cap.

2.4. COM DESPLAÇAR-SE PER L'APLICACIÓ

L'aplicació compta amb una barra d'eines que permet el desplaçament entre pantalles. Aquesta barra d'eines està descrita en l'Annex 2 "Manual d'usuari".

El desplaçament normal per l'aplicació és seqüencial, és a dir, avançant d'una pantalla a una altra. L'aplicació compta també amb un Navegador, que permet en qualsevol moment desplaçar-se a la pantalla desitjada. Aquest Navegador també està descrit en el Manual d'usuari.

Hi ha situacions, però, on l'aplicació no té l'estructura típica seqüencial. Situacions d'aquestes passen quan hi ha submenús. Un exemple és la pantalla "Validació: Premisses", mostrada en la figura 2.14.

En aquesta pantalla es pot triar veure cada una de les premisses. Per fer-ho s'ha de prémer el botó corresponent. Si, per exemple, es prem al botó "Linealitat" s'entra a les pantalles que fan referència a la premissa de linealitat. Un cop s'hagin passat totes aquestes pantalles, apareix el botó "Menú", que al prémer-lo tornarà a l'usuari a la pantalla on es mostren les premisses, i es podrà prosseguir a veure les restants pel mateix procediment. Al tornar a la pantalla veurem que la casella que hi ha sota el botó corresponent queda marcada, indicant que ja s'ha vist tot el contingut d'aquell botó. Això no impedeix, no obstant, que si es vol es puguin tornar a veure les pantalles associades al botó.

2. Disseny i Implementació

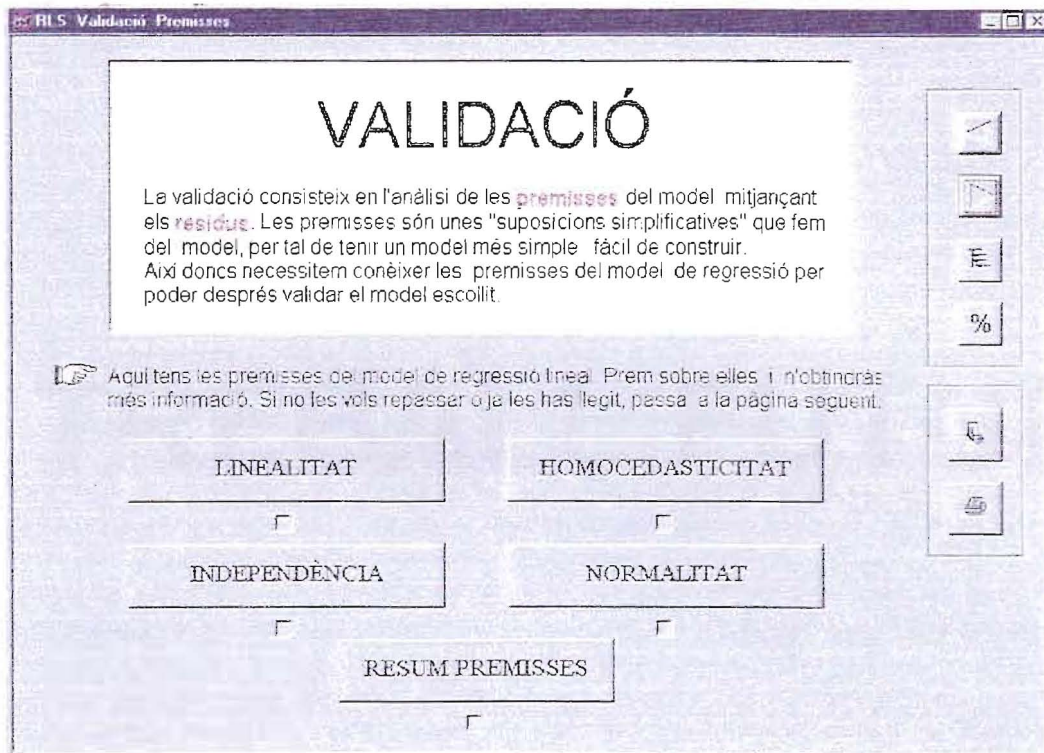


Figura 2.14. Pantalla "Validació: Premisses"

Quan s'han vist totes les pantalles associades a tots els botons, totes les caselles apareixen marcades. Llavors, es segueix l'aplicació prement el botó pantalla següent, com s'ha anat fent fins a aquell moment, seguint amb l'esquema seqüencial normal de l'aplicació.

2.5. TEMPS APROXIMAT D'ÚS DE L'APLICACIÓ

El temps aproximat que necessita un usuari per a la utilització completa de l'aplicació RLS va de 60 a 80 minuts.

L'aplicació consta de 116 pantalles, 95 del bloc teòric i 21 de l'Anàlisi d'un cas.

SECCIÓ 3

TREBALL FUTUR

3. Treball futur

3. TREBAL FUTUR

Una aplicació de suport a l'ensenyament com la que s'ha dissenyat en aquest projecte pot tenir una bona acollida pels estudiants com a complement a l'assignatura.

En una enquesta sobre l'aplicació feta a alumnes de l'assignatura d'Estadística 2 (ES2) d'Enginyeria Informàtica de la FIB, els resultats de la qual es presenten en l'Annex 2, hi ha, malgrat que l'enquesta era per avaluar la present aplicació, peticions de més programes d'aquest tipus referents a més temes (i en concret es suggereix una aplicació per tractar la comparació de poblacions).

Una valoració positiva de la utilització d'aquesta aplicació per part dels estudiants, refermaria la utilitat d'aquest tipus d'aplicacions i plantejaria la conveniència del disseny de noves aplicacions de més temes.

L'aplicació dissenyada en aquest projecte tracta el tema de la regressió lineal simple.

Malgrat ser un dels temes més bàsics de l'estadística, no s'acaba de tractar en profunditat. L'aplicació podria veure's ampliada amb diferents mòduls nous. Per exemple, no es tracten en la present aplicació mesures per trobar dades anòmales (leverage, distància de Cook, etc.).

Per aprofundir més en la regressió, és evident que no n'hi ha prou amb la regressió lineal simple, i és necessària la incursió en la regressió amb més d'una variable explicativa, la regressió lineal múltiple.

Com s'ha comentat ja en aquesta memòria, en una situació real poques vegades s'acabarà treballant amb un model de regressió lineal simple, i serà necessari fer servir un model amb més d'una variable. No obstant, aprofundir primer en la regressió lineal simple fa molt més fàcil entrar en la seva extensió natural, la regressió lineal múltiple. Moltes de les coses que s'aprenen en l'aplicació RLS són perfectament aplicables en situacions de regressió lineal múltiple, com l'anàlisi de les premisses, etc.

3. Treball futur

Així, podria estudiar-se la creació d'una nova aplicació com a extensió o continuació de RLS, en la que es tractés la regressió lineal múltiple. Podria ser una aplicació a part, o acoblar-la a l'actual, formant en aquest últim cas una aplicació amb diferents nivells de dificultat, dirigit a diferents tipus d'usuaris.

En aquesta, es podria realitzar només la part de regressió lineal simple dissenyada en aquesta aplicació, i els usuaris que volguessin aprofundir podrien realitzar també les parts de regressió lineal múltiple, d'avaluació "avançada" de dades anòmales, etc.

L'extensió natural d'aquesta aplicació passa doncs per tractar el model lineal general, és a dir, la regressió lineal múltiple, els models d'anàlisi de la variància (ANOVA) i d'anàlisi de la covariància (ANCOVA).

SECCIÓ 4

CÀRREGA DEL TREBALL

4. Càrrega del treball

4. CÀRREGA DEL TREBALL

La realització d'aquest Projecte Final de Carrera ha suposat 200 hores en el disseny i la programació de l'aplicació i de 40 hores en la redacció.

Temps de realització: un any.

Cost del projecte:

Fotocòpies:	400 ptes.
Enquadernació:	5000 ptes.
CD's:	1250 ptes.

Estimació del temps i el cost del projecte:

Cost : 6650 ptes.
Temps: 240 hores.

ANNEX 1

RESOLUCIÓ DE L'ANÀLISI D'UN CAS

INTRODUCCIÓ

En el segon bloc de l'aplicació RLS es proposa a l'usuari que resolgui un cas utilitzant un conjunt de dades reals, per tal d'avaluar el seu coneixement sobre la regressió i la seva capacitat per aplicar-lo a un cas pràctic.

Les dades que s'han usat intenten relacionar la taxa de crims en diferents estats dels Estats Units d'Amèrica amb una variable que mesura la riquesa de l'estat, amb la població de cada estat, i amb la despesa policial realitzada per l'estat .

S'han fet servir dades de gairebé tots els estats dels Estats Units. Bàsicament, doncs, l'objectiu dels models és explicatiu, ja que un model predictiu no té sentit al no quedar pràcticament més estats que no participin en la creació del model.

Per veure si existeix relació entre aquestes variables i per explicar-la s'han usat models de regressió lineal simple.

MÈTODES

Les dades que s'han usat són una part de les que es poden trobar a l'adreça següent : <http://www.maths.uq.edu.au/~gks/data/general/uscrime.html>.

En la pàgina d'Internet en qüestió hi ha més variables de les que es fan servir en aquest anàlisi, ja que està pensada per crear un bon model de regressió lineal múltiple. Com que en l'aplicació només es treballa amb models de regressió lineal simple, s'han triat, a part de la variable resposta, només tres possibles variables explicatives. Haver fet servir més variables hagués suposat una tasca massa repetitiva a l'hora de realitzar els anàlisi en l'aplicació.

La variable que s'ha estudiat és la taxa de crims en una categoria particular en la pràctica totalitat dels estats dels Estats Units d'Amèrica a l'any 1960. En concret aquesta variable és el nombre de delictes denunciats a la policia per milió d'habitants. S'ha disposat de dades de 47 estats.

Les dades originalment provenen del "Uniform Crime Record" de l'FBI i de fonts del Govern dels Estats Units. Les possibles variables explicatives incloses en l'anàlisi han estat : el nombre d'habitants en cada estat, una mesura de la riquesa de cada estat, i la despesa policial realitzada en l'estat.

Annex 1. Resolució de l'Anàlisi d'un cas

En concret, la variable població en cada estat ve expressada en centenars de milers de persones. La variable indicadora de riquesa és el valor mig dels béns transferibles i els actius o la renda familiar, expressat en desenes de dòlars. La variable despesa policial és la despesa per càpita realitzada en policia per estat i govern local, expressada en dòlars.

Com ja s'ha comentat, s'han usat models de regressió lineal simple (criteri de mínims quadrats), i per la realització dels anàlisi s'ha usat el paquet estadístic MINITAB.

Anàlisi i resultats

Quan s'ha estudiat la variable despesa policial s'ha fet el següent anàlisi:

Creant un model que relacioni directament la variable despesa policial amb la taxa de crims, s'ha observat en els residus falta d'homocedasticitat. Per solucionar aquest problema s'han provat diverses transformacions, de les quals la que ha donat lloc a uns millors residus és fer servir l'arrel quadrada de la taxa de crims. El gràfic de residus Vs. valors previstos corresponent a aquesta relació es mostra en la figura A1.1. Basant-nos en el gràfic podem acceptar les premisses i donar el model com a bo.

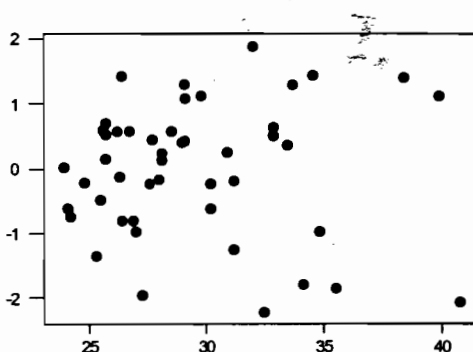


Figura A1.1. Residus Vs. valors previstos (arrel taxa crims Vs. despesa)

El model finalment escollit fa servir doncs aquesta variable. El resultat de la regressió es mostra en la taula A1.1. El model és capaç d'explicar una mica menys de la meitat de la variabilitat de la resposta (44 - 46%). Aquesta relació serà analitzada en l'apartat de discussió.

Annex 1. Resolució de l'Anàlisi d'un cas

Variable : $\sqrt{\text{Taxa de crims}}$

	Coefficient	p-valor
Constant	17.586	<0.001
Despesa policial	0.13976	<0.001
$R^2 = 45.8\%$	R^2 ajustat = 44.6%	

Taula A1.1. Regressió amb la variable despesa policial.

Quan es treballa amb la variable Riquesa de l'estat, succeeix una situació similar a l'anterior. Relacionant directament la taxa de crims amb aquesta variable, s'incompleix l'homocedasticitat, la normalitat, i a més s'observen punts bastant allunyats de la resta. Si s'aplica la mateixa transformació que amb l'anterior variable, és a dir, es passa a treballar amb l'arrel quadrada de la taxa de crims, es soluciona l'heterocedasticitat i la manca de normalitat, i els punts allunyats de la resta no hi queden tant. La figura A1.2 mostra el gràfic de residus Vs. valors previstos d'aquesta relació. La majoria dels residus estan ara entre -2 i 2, i cap està molt allunyat. Els residus són acceptables, i es pot donar el model com a bo.

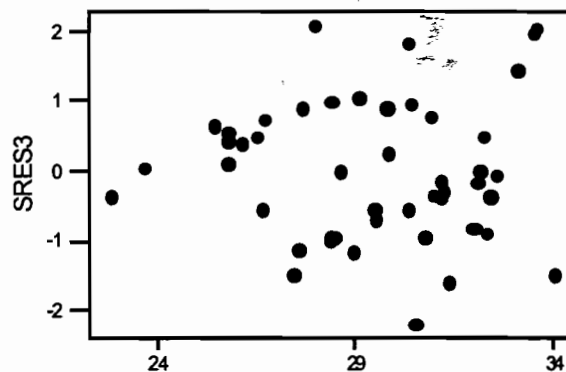


Figura A1.2. Residus Vs. valors previstos (arrel taxa crims Vs. Riquesa)

El resultat de la regressió es mostra en la taula A1.2.

Annex 1. Resolució de l'Anàlisi d'un cas

Variable : $\sqrt{\text{Taxa de crims}}$

	Coefficient	p-valor
Constant	14.838	0.002
Riquesa	0.0278	0.002
$R^2 = 19.2\%$	R^2 ajustat = 17.4%	

Taula A1.2. Regressió amb la variable Riquesa

L'última variable a analitzar és el nombre d'habitants en un estat. Creuant les dues variables directament queda clar que la relació entre ambdues, si existeix, no és recta. Treballant amb el logaritme de la població aconseguim que els punts siguin més propers en l'espai de les x, i la relació millora molt. Es veu clarament un punt molt allunyat de la resta; més endavant s'explica com s'ha tractat el punt en qüestió.

Els residus de la relació crims Vs log(població) mostren però una asimetria aplanada a la dreta, en part deguda a la observació allunyada, però no totalment, ja que excloent aquesta observació l'asimetria encara hi és present. Per tal de corregir-la es pren també el logaritme de la variable taxa de crims. El model és presentat en la taula A1.3, i en la figura A1.3 es pot veure el gràfic de residus Vs. valors previstos. En ell es pot observar una observació amb un residu estandarditzat de 2.8. Aquesta, correspon a un estat amb una població petita i una taxa de crims molt gran en comparació amb la resta, just el contrari del que indica la relació positiva entre les dues variables. Excloent aquesta observació s'obtenen els resultats mostrats en la taula A1.4 i en la figura A1.4.

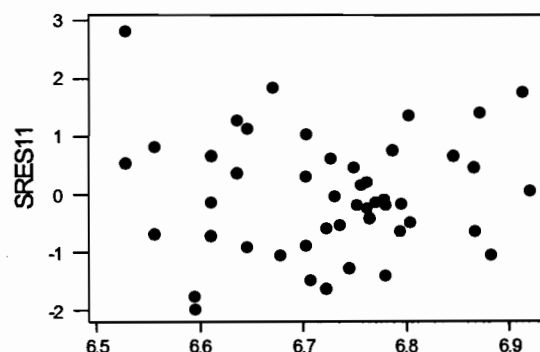


Figura A1.3. Residus Vs. valors previstos (log crims Vs. log població)

Annex 1. Resolució de l'Anàlisi d'un cas

Variable : log (Taxa crims)

	Coeficient	p-valor
Constant	6.42	<0.001
log (població)	0.097	0.099
$R^2 = 5.9\%$	R^2 ajustat = 3.8%	

Taula A1.3. Regressió amb la variable log(Població)

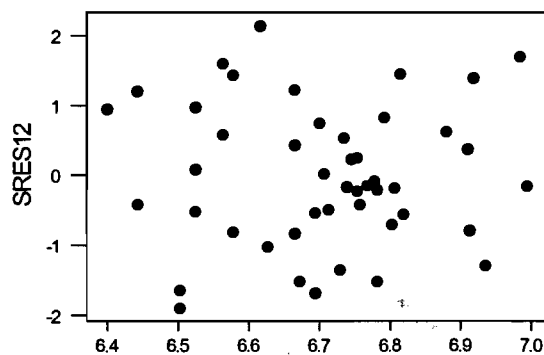


Figura A1.4. Residus Vs. valors previstos (log crims Vs. log població) excloent una observació

Variable : log (Taxa crims)

	Coeficient	p-valor
Constant	6.238	<0.001
log (població)	0.1476	0.011
$R^2 = 13.8\%$	R^2 ajustat = 11.8%	

Taula A1.4. Regressió amb la variable log(Població) excloent 1 observació

DISCUSIÓ

Els millors models de regressió lineal simple que s'han trobat relacionen l'arrel quadrada de la despesa policial amb la taxa de crims i amb la riquesa, i el logaritme de la taxa de crims amb el logaritme de la població de cada estat.

El propòsit original de les dades és explicar la taxa de crims de l'any 60 mitjançant tot un conjunt de covariables. En l'aplicació només es tracta regressió lineal simple, i per tant no es fan servir en la resolució del cas models multivariants. Cal tenir present això a l'hora d'interpretar els models resultants, doncs no es tindran en compte en una relació possibles factors de confusió. Les relacions que s'obtidran doncs, no podran interpretar-se com a causals, ja que podrien venir explicades per una altra variable que no tenim en compte en el model.

Centrant-nos primer en la despesa policial, el model estimat és :

$$\sqrt{\text{taxa de crims}} = 17.59 + 0.14 \times \text{Despesa policial}$$

la interpretació, si el model fos el que es dona en la realitat, és que per cada augment d'un dòlar en despesa policial per càpita de l'estat l'arrel quadrada de la taxa de crims de l'any següent augmenta 0.14.

Si es fa aquesta interpretació com a causal, cosa que no queda garantida degut a no controlar per possibles confusors, la interpretació pot semblar carent de lògica, doncs el model sembla suggerir que reduint la despesa policial s'aconseguiria reduir la taxa de crims. El què està passant és que potser en aquesta relació hagués estat més convenient utilitzar com a variable resposta la despesa policial, ja que aquesta variable és la que canviarà en funció de la taxa de crims que hi hagi. Construint el model d'aquesta manera, sí que és lògica la relació positiva existent entre les dues variables, doncs quan augmenta la taxa de crims augmenta també la despesa policial i a l'inrevés.

Amb aquesta variable s'aconsegueix explicar aproximadament un 45% de la variabilitat de la resposta, i la variable despesa policial és altament significativa.

Pel què fa a la variable riquesa, el model finalment escollit és :

$$\sqrt{\text{Taxa de crims}} = 14.84 + 0.028 \times \text{Riquesa}$$

Annex 1. Resolució de l'Anàlisi d'un cas

Novament, com en tots els casos, no podem parlar de relació causal, ja que no es donen les circumstàncies per a poder-ho assegurar. No obstant, si assumim el model com a bo, queda reflectit que per cada 10 dòlars que augmenta aquesta mesura de riquesa d'un estat l'arrel quadrada de la taxa de crims creix 0.028. La relació ens diu que són els països amb una riquesa més alta (segons l'índex usat) aquells que tenen una taxa de crims més elevada.

La variable riquesa és significativa, i el model final aconsegueix explicar aproximadament un 17-19% de la variabilitat de la resposta.

Per últim, usant la variable població en cada estat, el model resultant és el següent :

$$\log(\text{Taxa de crims}) = 6.42 + 0.097 \times \log(\text{població})$$

Observant els residus d'aquest model es pot veure una observació molt allunyada de la resta. Aquesta correspon a un estat amb una població molt petita i una taxa de crims molt gran. És bàsicament només aquesta observació la que s'allunya de la resta. El seu comportament diferent a la resta d'estats pot ser degut al valor d'una altra variable també relacionada amb la taxa de crims. Com que usem models univariants, aquesta possibilitat no es pot comprovar en un model ajustant per confusors, per tant, és convenient deixar el model tal i com està, doncs per la resta de dades el model és acceptable. Es pot comprovar el gran pes que té aquesta observació ajustant el mateix model sense incloure aquest estat. El model estimat llavors és :

$$\log(\text{Taxa de crims}) = 6.238 + 0.1476 \times \log(\text{Població})$$

Es pot comprovar com els coeficients varien notablement, i la variable $\log(\text{població})$ passa a ser significativa. El model, exclouent l'estat en qüestió, passa d'explicar un 3.8% de la variabilitat de la resposta a explicar-ne un 11.8%. Els residus havent exclòs aquesta observació no presenten cap anomalia.

La relació trobada, per bé que pot estar confosa per altres variables, indica que els estats amb més alta població tenen també la taxa de crims més alta. Cal recordar però que s'ha exclòs un estat amb una taxa molt alta i una població petita en comparació amb la resta. Sense usar models multivariants no es pot saber si aquest fet pot ser degut a una altra variable. No obstant això, el model que inclou l'estat que es diferencia té la mateixa direcció (positiva), si bé no acaba de ser significativa.

La relació és en aquest model més difícil d'interpretar, doncs ha calgut transformar les dues variables. El model indica que per cada unitat que creix el logaritme de la variable població, el logaritme de la taxa de crims creix 0.1476. El model és equivalent (algebraicament) a la relació:

$$\text{Taxa de crims} = e^{6.238} \cdot \text{Població}^{0.1476} = 511.8 \cdot \text{Població}^{0.1476}$$

Annex 1. Resolució de l'Anàlisi d'un cas

Per concloure, podem dir que les tres variables estudiades poden relacionar-se univariantment amb la taxa de crims de l'any 1960, fent ús de diferents transformacions. La relació és més forta amb la variable despesa policial, on s'explica aproximadament un 45% de la variabilitat. La variable riquesa quasi n'explica un 20%, i la població de l'estat només un 12%.

El proper pas en l'anàlisi seria incloure les tres variables, juntament amb d'altres que podrien actuar com a confusores (com poden ser la desigualtat en la renda, el percentatge de blancs en la població, anys d'escolarització, etc.), en un model multivariant. Podria passar que variables que hi estaven relacionades en l'anàlisi cru deixin d'estar-ho, o contràriament, que la relació sigui més forta que abans.

ANNEX 2

**AVALUACIÓ DE LA VERSIÓ
BETA PELS ALUMNES
D'ESTADÍSTICA 2 (ES2)
D'ENGINYERIA INFORMÀTICA DE
LA FIB**

ANNEX 2: AVALUACIÓ DE LA VERSIÓ BETA PELS ALUMNES D'ESTADÍSTICA 2 (ES2) D'ENGINYERIA INFORMÀTICA DE LA FIB

Una versió beta de l'aplicació RLS es va passar a dos grups de laboratori de l'assignatura Estadística 2 (ES2) d'Enginyeria Informàtica de la FIB. Es va demanar als alumnes que valoressin l'aplicació, donant-ne els punts forts, els punts febles, propostes de millora i una nota numèrica com a puntuació del programa. Cada enquesta estava contestada per dues persones que van utilitzar l'aplicació en el mateix ordinador.

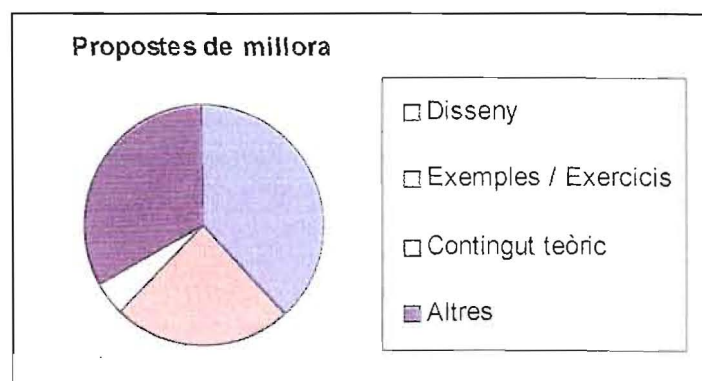
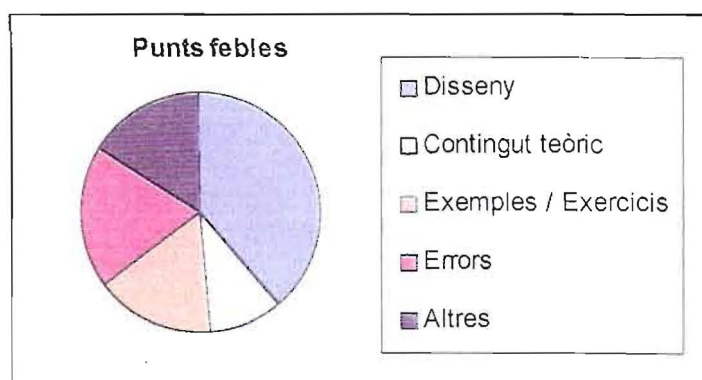
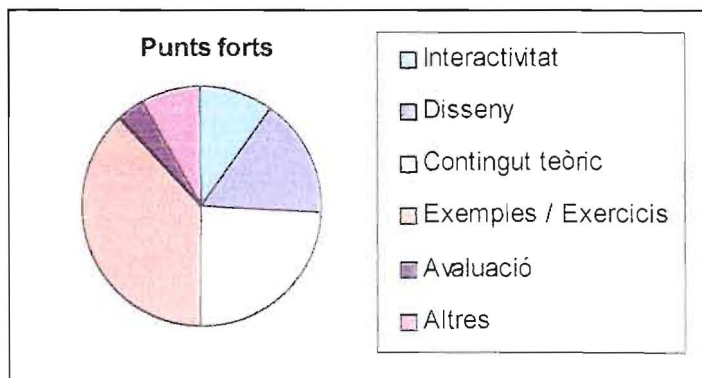
L'enquesta va ser molt positiva de cara a veure què esperaven els estudiants, els usuaris potencials del programa, d'un software d'aquestes característiques.

Els resultats es poden veure a les taules 1, 2 i 3, i en els gràfics 1, 2, 3. La puntuació mitjana que van donar els estudiants a la versió beta de RLS va ser de 7.5 (sobre 10) .

El què més va destacar en els punts forts són els exemples, els exercicis i el contingut teòric.

Pel què fa als punts febles i a les propostes de millora, la majoria es van centrar en el disseny, si bé també en els exercicis i en menor mesura en les explicacions teòriques. En els punts febles també hi figuraven errors, que òbviament van ser corregits en la versió final.

Annex 2. Avaluació de la versió beta pels alumnes d'ES2 d'Enginyeria Informàtica de la FIB



Annex 2. Avaluació de la versió beta pels alumnes d'ES2 d'Enginyeria Informàtica de la FIB

<u>PUNTS FORTS</u>	<u>freq</u>
Interactivitat	5 (10 %)
Interactiu	4
Menús interactius	1
Disseny	8 (16 %)
Facilitat d'ús	2
Disseny en pantalla	1
Estructuració del temari	1
Molt visual	1
Navegador útil	2
Senzill d'usar	1
Contingut teòric	12 (24 %)
Explicacions detallades, ben explicat	9
Explicacions intuïtives	2
Bones introduccions	1
Exemples / Exercicis	19 (38 %)
Molts exemples	4
Els comentaris de les respostes	4
Bones preguntes	4
Bons gràfics	3
Bons exemples	1
Exemples basats en dubtes més usuals	1
Tipus test és ràpid	1
Molts gràfics	1
Avaluació	2 (4 %)
Valoració dels coneixements	1
Donar una puntuació al final	1
Altres	4 (8 %)
Autodidacta	2
No massa avorrit	1
Bon reforç per aprenentatge	1

Taula 1. Punts forts

Annex 2. Avaluació de la versió beta pels alumnes d'ES2 d'Enginyeria Informàtica de la FIB**PUNTS FEBLES**

Disseny	12 (39 %)
Massa lletra explicativa	4
Icones haurien d'estar al mateix lloc	2
Complex per moure't dins el programa	2
Pantalla ajust pendent: No es pot posar directament	1
No indica quan has acabat una fase	1
No aprofita tot l'espai de pantalla per lletra + gran	1
Poca ajuda en la navegació	1
Contingut teòric	3 (10 %)
Dependència del Minitab	1
Introducció difícil d'entendre	1
Coses sense explicar	1
Exemples / Exercicis	5 (16 %)
Solucions: està malament però no diu perquè	2
Validació respostes	1
En preguntes no s'informa si és multiresposta o no	1
Exercicis llargs de fer varies vegades	1
Errors	6 (19 %)
Altres	5 (16 %)
Tamany fitxer	1
La barra d'estat no canvia	1
Es necessita 800 x 600	1
No té animacions	1
No té memòria de les pantalles fetes	1

Taula 2. Punts febles

Annex 2. Avaluació de la versió beta pels alumnes d'ES2 d'Enginyeria Informàtica de la FIB**PROPOSTES DE MILLORA**

Disseny	8 (38 %)
Més divertit	1
Entorn més intuïtiu	1
Colors menys pesants	1
Hipertext explicatiu	1
Mostrar nombre pàgines d'un tema (3/7)	1
Subratllar idees importants	1
Més dinàmic	1
Evitar la correcció de les respostes	1
Exemples / Exercicis	5 (24 %)
Més exemples	2
Respostes a millorar	1
Més exercicis a la primera part	2
Contingut teòric	1 (5 %)
Introducció més didàctica	1
Altres	7 (33 %)
Més programes de més temes, poder estudiar a casa	3
Fer càlculs amb dades externes	1
Incorporar so	1
Control per pantalla de les puntuacions	1
Més links	1

Taula 3. Propostes de millora

MODIFICACIONS A LA VERSIÓ BETA A PARTIR DE L'ENQUESTA ALS ESTUDIANTS

A la versió beta se li van fer diverses correccions, moltes d'elles basades en les suggerències dels alumnes.

A part dels errors tipogràfics i els errors en la programació, òbviament corregits, una de les principals coses que es van canviar referien al disseny. Així, es van dividir totes les pantalles teòriques en dues, a fi i efecte de reduir el text i els conceptes que es mostraven per pantalla. Les pantalles es van fer d'aquesta manera més agradables a la vista. També es van destacar en vermell termes importants, i es van afegir subtítols que donen idea de què s'està explicant en cada moment. Es va afegir també més hipertext a les explicacions.

Es va intentar fer més entretinguda la introducció, que semblava fer-se pesada, afegint pantalles interactives de les quals abans no disposava.

Una altra millora interessant va ser col·locar en la barra de títol quantes pantalles té la secció en la qual s'està (p.ex.: 1/4), així com també es va canviar part de l'estructuració del temari. La organització del programa va quedar d'aquesta manera més clara que abans.

També es va incloure un nou botó a les pantalles, que permet veure en cada moment la proporció de preguntes ben i mal contestades en la primera ocasió. En l'antiga versió, la puntuació només es podia veure al final del bloc teòric.

ANNEX 3

MANUAL D'USUARI

MANUAL D'USUARI

L'aplicació RLS és un programa d'autoaprenentatge de regressió lineal simple. L'aplicació està adreçada bàsicament a estudiants de qualsevol assignatura d'Estadística on es tracti el tema de la regressió.

En l'aplicació s'expliquen diferents conceptes teòrics referents a la regressió, i s'aconsella a l'usuari en aspectes pràctics. L'aplicació conté també una gran quantitat de gràfics per una millor il·lustració del què s'està explicant.

D'altra banda, l'aplicació conté també un gran nombre de preguntes a l'usuari, les respostes a les quals són comptabilitzades.

En la segona part del programa ("Anàlisi d'un cas") l'usuari ha de contestar nombroses preguntes referents a diferents models que es van construir amb unes dades concretes, i ha de decidir sobre com prosseguir i sobre les valoracions de cada model que se li presenta. Les preguntes van referides a gràfics o a sortides que poden donar els paquets estadístics (Minitab en aquest cas).

BARRA D'EINES



Botó "Pàgina anterior"

Per moure's a la pantalla immediatament anterior segons l'ordre seqüencial del programa. Si s'ha anat a una pantalla amb el Navegador, aquest botó no fa tornar a la pantalla de la qual es prové, sinó a l'anterior segons l'esquema del programa.



Botó "Pàgina següent"

Per passar a la pantalla següent, en l'ordre seqüencial del programa. Quan està desactivat és que no queden més pantalles dins del subtema, i cal tornar al menú del qual es prové mitjançant el botó "Menú" .



Botó "Menú"

Botó que torna al menú on s'havia entrat en una de les opcions.



Botó "Navegador"

Botó que porta al Navegador, des d'on es pot accedir directament a qualsevol pantalla del programa.



Botó "Puntuació"

Mitjançant aquest botó s'activa un formulari on l'usuari pot veure les respostes ben i mal contestades fins al moment, així com el percentatge d'encerts.



Botó "Sortir"

Botó per abandonar el programa.



Botó "Imprimir"

Imprimeix el contingut de les finestres de text.

Annex 3. Manual d'usuari

EL NAVEGADOR

El programa pot ser seguit seqüencialment mitjançant els botons "Pàgina següent", "Pàgina anterior" i "Menú", però també poden fer-se salts de pantalla i accedir directament a una pantalla concreta. Això és possible mitjançant el Navegador.

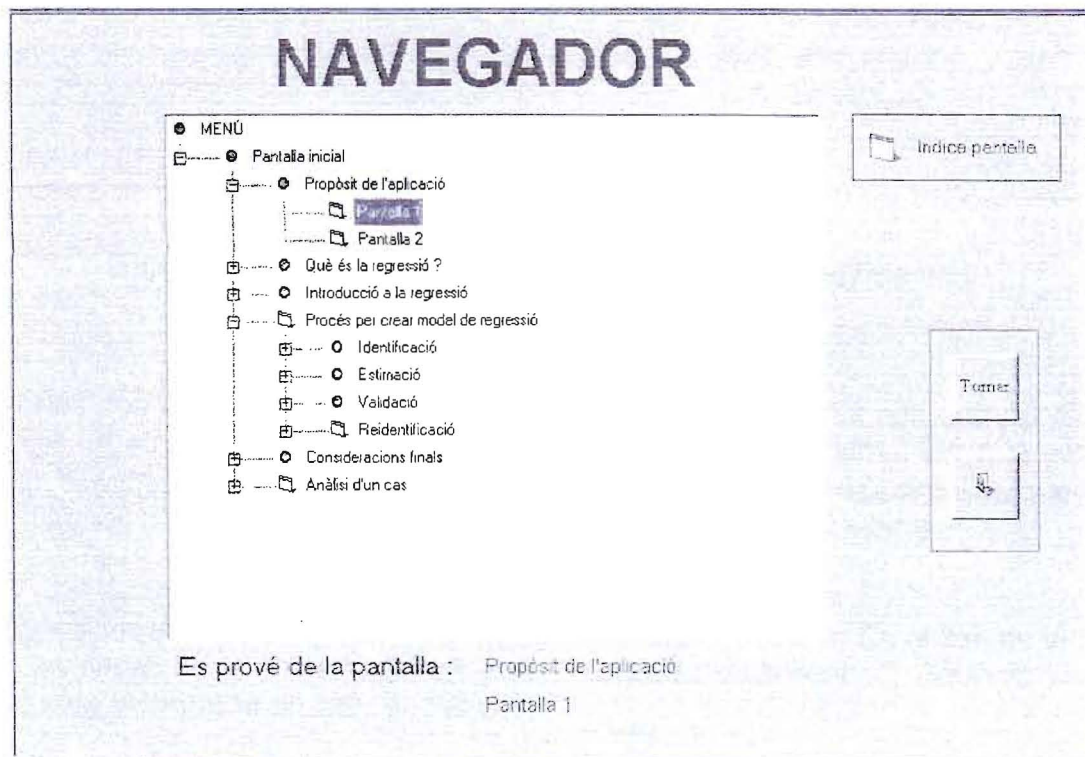


Figura A3.1. Navegador

En la pantalla "Navegador" (figura A3.1), hi ha tota l'estructura de les pantalles del programa en forma d'arbre. A partir dels diferents nodes es poden anar desagregant les parts del programa.

Si en el node apareix el signe "+" vol dir que hi ha més d'una pantalla dins aquell node, i cliquejant sobre el "+" podrem veure-les. Aquesta situació queda reflectida en la figura A3.2. Quan apareix el signe "-" ja no queden més pantalles dins el node.

Annex 3. Manual d'usuari

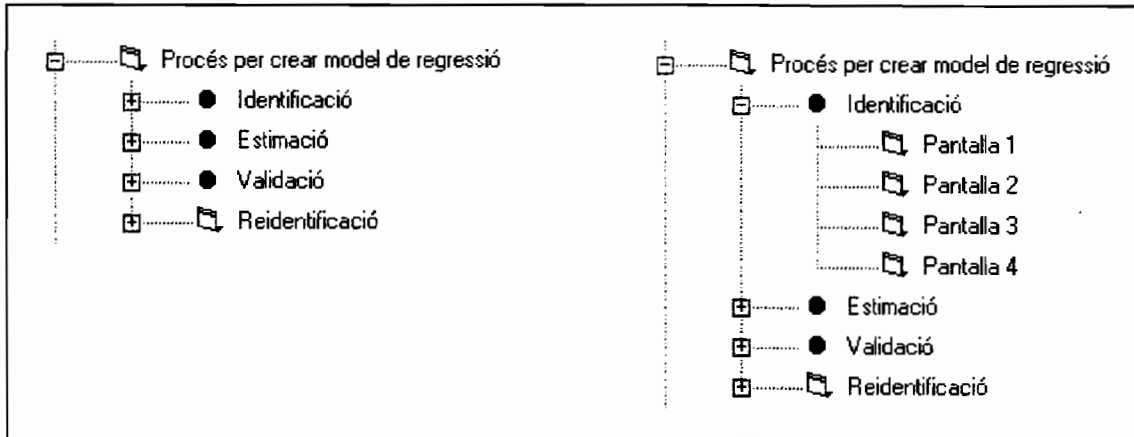



Figura A3.2. Pantalles "Identificació agregades (+) i desagregades (-)

Els nodes que duen incorporat el símbol  són branques que són una pantalla. Al cliquejar un cop sobre ells accedirem a la pantalla corresponent. Els que tenen el símbol ● són només indicatius del tema de què tracta el grup de pantalles que estan dins d'aquell node. Si cliquegem sobre ells no es produirà cap efecte.

El Navegador també informa de la pantalla de la qual es prové. En el cas de la figura A3.1, es prové de la primera pantalla de "Propòsit de l'aplicació". Aquesta pantalla també està destacada en blau en l'arbre.

El botó "Tornar" que apareix en aquesta pantalla torna a l'usuari a la pantalla d'on ha cridat al Navegador.

ANNEX 4

**MANUAL
D'INSTAL·LACIÓ**

1. Requisites

Per l'aplicació es requereix el següent hardware i software:

- Microprocessador Pentium ® 90 MHz o superior.
- Pantalla VGA 640x480 o de resolució superior (és molt recomanable resolució 800x600 per una bona visualització de l'aplicació) compatible amb Microsoft Windows.
- 24 MB de RAM per Windows 95, 32 MB per Windows NT.
- Microsoft Windows 95 o posterior, o Microsoft Windows NT 3.51 o posterior.

Requisits d'espai al disc dur:

En el pitjor dels casos (quan no es té cap dels arxius DLL que es descriuen més endavant):

6.34 MB

Tamany de l'executable: 2.2 MB

2. INSTAL·LACIÓ

Per la distribució del programa es fan servir 3 arxius per la distribució en CD-ROM i 4 arxius per la distribució en disquets (2 disquets).

Per iniciar la instal·lació s'ha d'entrar al primer disquet o al CD i obrir l'arxiu SETUP.EXE . A partir d'aquí comença el programa d'instal·lació de RLS.

El primer que apareix és un missatge com el de la figura A4.1:

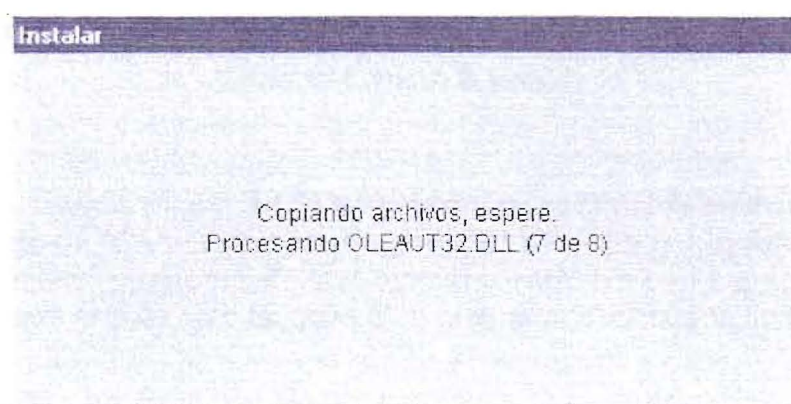


Figura A4.1. Còpia dels arxius necessaris

El programa d'instal·lació copia 8 arxius que li són necessaris per l'execució. Si el procés d'instal·lació no s'acaba perquè l'usuari ha decidit interrompre'l, el programa d'instal·lació automàticament borrarà aquests arxius que ha copiat.

Seguidament, apareix la pantalla d'instal·lació de RLS, on apareix el missatge de la figura A4.2.

És recomanable no tenir cap aplicació oberta. Si es vol sortir del programa per tancar altres aplicacions, s'ha de prémer a "Salir". El programa d'instal·lació borrarà els arxius que ha copiat, i la instal·lació podrà realitzar-se més tard. Per continuar amb la instal·lació de RLS, cal prémer "Aceptar"

Annex 4. Manual d'instal·lació

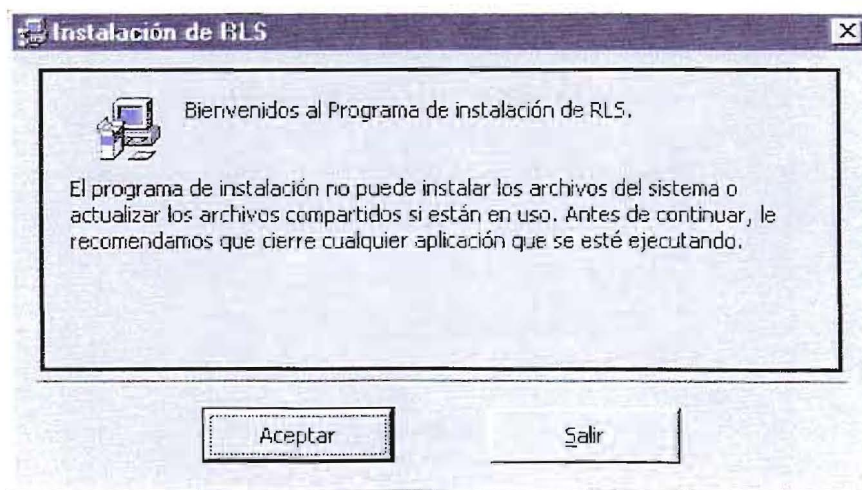


Figura A4.2. Iniciar la instal·lació

En la següent pantalla (figura A4.3) es pot canviar el directori on instal·lar l'aplicació, que per defecte és "C:\Archivos de programa\RLS\" . Es pot indicar un directori no existent i el mateix programa d'instal·lació el crearà. Un cop especificat el directori de destí, prement el botó amb la icona d'un ordinador s'iniciarà la instal·lació de RLS.

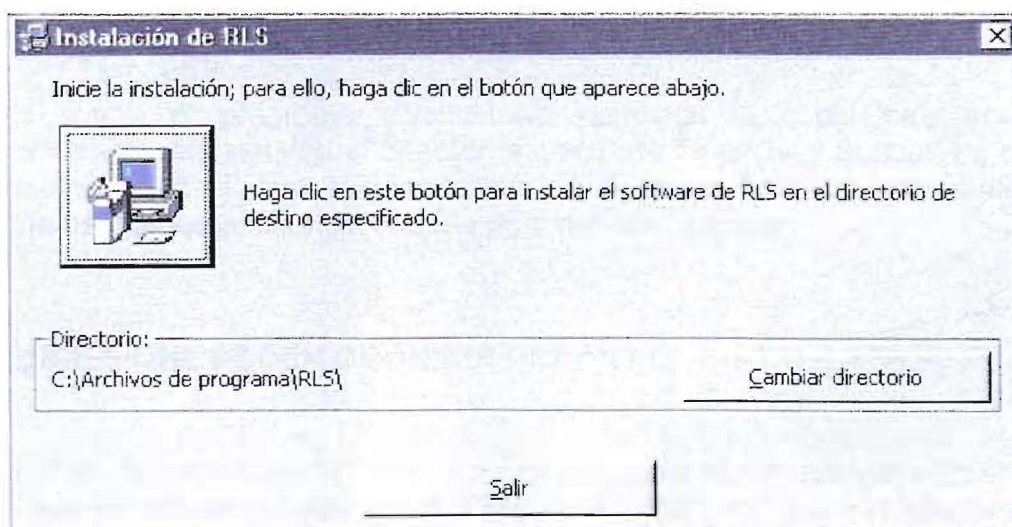
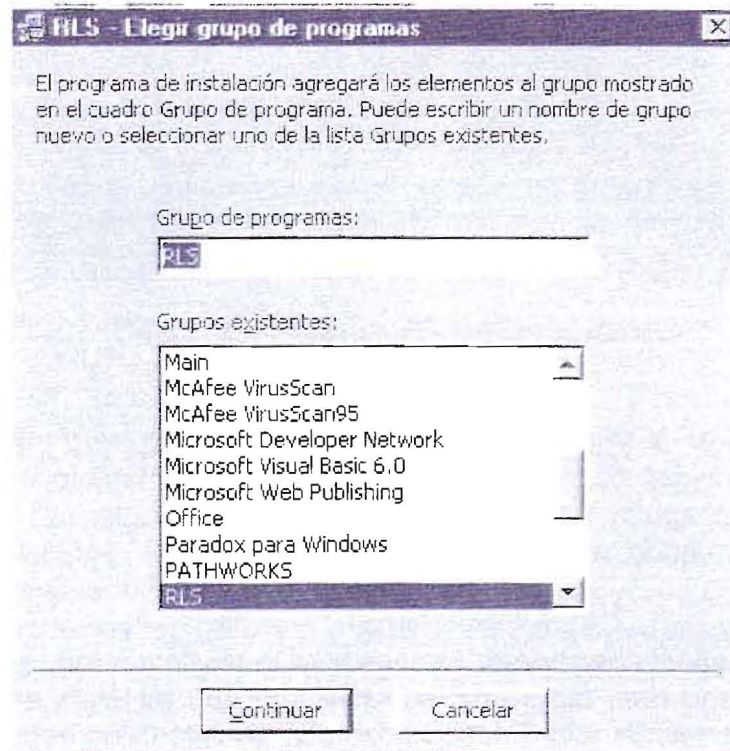


Figura A3.3. Directori on instal·lar el programa

Annex 4. Manual d'instal·lació**Figura A4.4. Triar grup de programes**

L'última pantalla abans de dur a terme el propi procés d'instal·lació és la que es mostra en la figura A4.4. En aquesta pantalla es deixa "RLS" com ja ve per defecte (encara que si l'usuari vol pot canviar-ho).

A partir d'aquí el programa d'instal·lació comença la còpia dels arxius, els descomprimeix i els instal·la al directori especificat. Quan hagi acabat, es mostrarà un missatge indicant que la instal·lació de RLS ha finalitzat correctament. El programa ha quedat ja instal·lat i està a punt per ser executat.

INCIDÈNCIES QUE PODEN OCÓRRER DURANT LA INSTAL·LACIÓ

Pot ser que al començar la instal·lació el programa digui que cal actualitzar uns arxius i que cal reiniciar (figura A4.5). Quan això passa cal reiniciar i després tornar a executar el programa d'instal·lació. Els arxius ja hauran estat actualitzats i la instal·lació es durà a terme normalment. Els arxius que s'actualitza vénen detallats en la propera secció.

Annex 4. Manual d'instal·lació

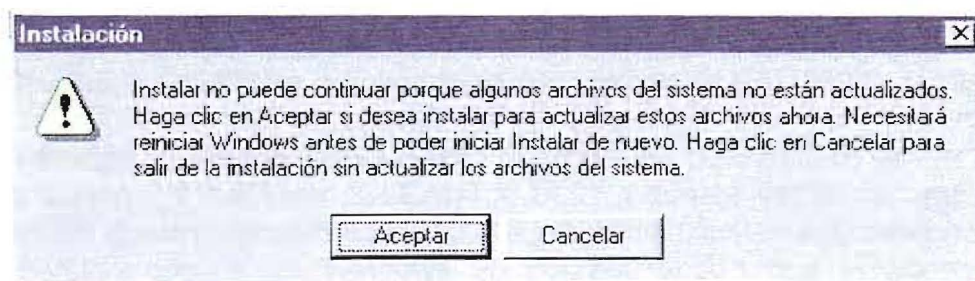


Figura A4.5. Missatge d'actualització d'arxius

Un altre missatge que pot aparèixer és el que es mostra a la figura A4.6. Aquest missatge apareix al final de la instal·lació, quan ja s'estan copiant tots els arxius. Si apareix aquest missatge és millor conservar la versió antiga de l'arxiu, tal i com recomana el missatge. Per tant s'ha de prémer "Sí" per conservar-lo. El programa funcionarà bé igual amb la versió antiga d'aquest arxiu.

Aquest missatge pot aparèixer si es disposa d'una versió de Visual Basic anterior a la 6.0, o si s'ha instal·lat un programa que s'ha creat amb una versió anterior de Visual Basic. Altres productes de Microsoft poden haver instal·lat la versió antiga de l'arxiu. Si s'actualitza l'arxiu amb la nova versió, a l'executar l'aplicació que usava la versió antiga poden produir-se problemes. És per això que es recomana conservar-lo.

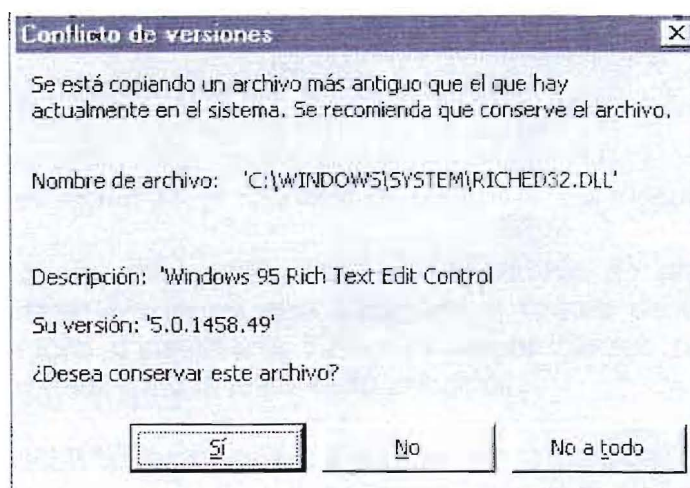


Figura A4.6. Conflicte de versions

Annex 4. Manual d'instal·lació

Un altre error que pot ocórrer és el que fa aparèixer el següent missatge:

"The SETUP.EXE file is linked to missing export OLEAUT32.DLL:185."

Aquest missatge apareix quan s'executa el programa d'instal·lació en un ordinador que té la versió 2.1 del fitxer "OLEAUT32.DLL". Aquesta versió ve amb la versió original de Windows 95 (versió 4.00.950). El programa d'instal·lació creat no funciona amb la versió original de Windows 95 pel poc suport que té aquest en API (Application Program Interface).

Aquest error no es produirà en ordinadors que tinguin versions posteriors de Windows 95, Windows 98 o Windows NT 4.0 o superior. Tampoc es produirà si un altre programa d'instal·lació, com el de Microsoft Office 97 o Internet Explorer 3.0 o superior, ja ha actualitzat l'arxiu OLEAUT32.DLL.

Per resoldre aquest problema, cal instal·lar les últimes versions d'arxius que necessita Visual Basic 6.0 per funcionar. Per aquesta incidència, en el CD d'instal·lació s'ha inclòs el fitxer VBRUN60.EXE, que facilita gratuïtament Microsoft per aquests casos. Només cal usar aquest arxiu si es produeix l'error que aquí es descriu. Executant primer aquest arxiu, es podrà utilitzar normalment el programa d'instal·lació.

L'arxiu VBRUN60.EXE es troba en el directori AUX1 del CD de distribució de l'aplicació.

3. DESCRIPCIÓ DELS ARXIS D'INSTAL·LACIÓ

Els arxius que es veuen en el CD de distribució o en els disquets són els següents:

- **Setup.exe** : Aquest arxiu actua com un executable de preinstal·lació. És el primer programa que s'executa a l'equip de l'usuari durant el procés d'instal·lació i realitza els processos necessaris abans de procedir amb la instal·lació principal.
- Arxiu o arxius **RLS*.cab**: L'aplicació queda empaquetada en un o dos arxius *.cab, depenent del suport per la instal·lació. Aquest arxiu conté els diferents arxius necessaris pel programa en format comprimit.

Annex 4. Manual d'instal·lació

En la instal·lació del programa, primer s'executa una preinstal·lació. En aquesta es graven uns arxius que són necessaris per l'execució. Aquests arxius són els següents:

- Msvbvm60.dll
- VB6STKIT.dll
- VB6ES.dll
- *Stdole2.tlb*
- *Oleaut32.dll*
- *Olepro32.dll*
- *Comcat.dll*
- *Asyncfilt.dll*

Aquests arxius poden estar ja presents en l'ordinador de l'usuari, ja que poden venir amb algunes versions de Windows, d'Explorer o de Office. En cas que hi hagi versions més antigues, el programa d'instal·lació les actualitzarà i s'haurà de reiniciar l'ordinador. Aquesta incidència ve explicada en l'apartat anterior ("Incidències que poden ocórrer durant la instal·lació").

Posteriorment, ja en el procés d'instal·lació del programa s'instal·len els següents arxius:

- CmCtlES.dll
- CmDlgES.dll
- RchTxES.dll
- RICHED32.dll
- RICHTX32.OCX
- COMCTL32.OCX
- COMDLG32.OCX

Algun d'aquests últims arxius poden existir a l'ordinador de l'usuari. Aquests arxius no s'executen durant la preinstal·lació. Probablement en el procés d'instal·lació apareixerà un missatge demanant si es vol actualitzar l'arxiu. Aquesta incidència també està descrita en l'apartat anterior.

Tots els arxius esmentats fins ara s'instal·laran al directori "C:\Windows\System\".

A més d'aquests arxius s'instal·laran dues subcarpetes en el directori on s'ha instal·lat el programa. Una té per nom "Doc", i conté els textos que es mostren en el programa. L'altre té per nom "Img", i conté algun dels gràfics que mostra el programa.

Annex 4. Manual d'instal·lació

L'eliminació de qualsevol d'aquests arxius comportarà el mal funcionament del programa.

Per últim, s'instal·len els arxius St6unst.log i St6unst.exe, necessaris per la desinstal·lació del programa (veure apartat de desinstal·lació). L'arxiu Setup.lst conté les instruccions que s'han usat per la instal·lació.

4. EXECUCIÓ DE RLS

El programa afegeix un accés a RLS a "...Menu Inicio\Programas\"", de manera que per executar RLS n'hi ha prou amb buscar-lo a "Inicio" - "Programas" i fer un clic sobre ell, com mostra la figura A4.7.

No obstant, també es pot executar obrint l'executable al directori on l'hem instal·lat.

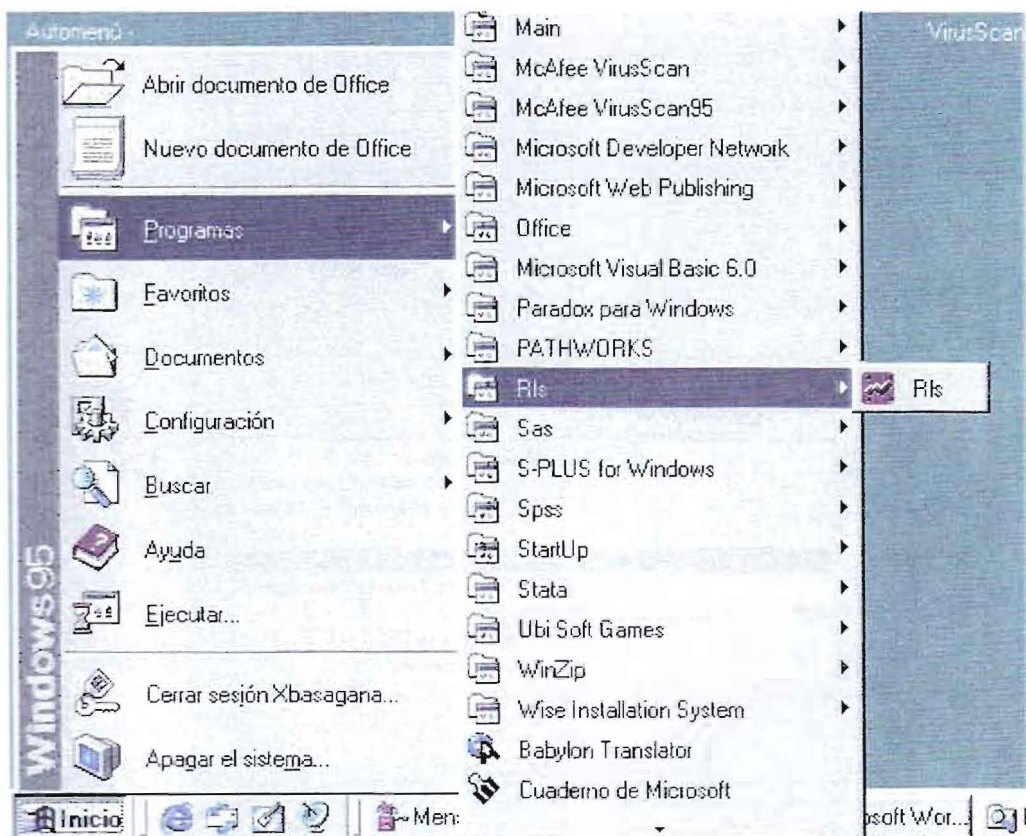


Figura A4.7. Execució de RLS mitjançant el botó "Inicio"

Annex 4. Manual d'instal·lació

5. DESINSTAL·LACIÓ DE RLS

El programa d'instal·lació de RLS copia un arxiu al directori que s'escull per instal·lar el programa (per defecte és "c:\Archivos de programa\RLS\") que es fa servir en la desinstal·lació de RLS. Aquest arxiu té per nom "St6unst.log". És molt important no borrar ni modificar aquest arxiu, ja que conté informació sobre la instal·lació de l'aplicació. L'arxiu es fa servir per eliminar automàticament totes les components de l'aplicació quan s'escull fer-ho.

Per desinstal·lar el programa pot fer-se amb el programa que proporciona Windows, el qual farà servir la informació de St6unst.log per desinstal·lar-lo.

Per fer servir el desinstal·lador de Windows, en Windows 95, cal anar al botó "Inicio", un cop allí anar a "Configuración" i allí a "Panel de control". En aquesta carpeta s'hi trobaran diferents icones. Cal triar "Agregar o quitar programas". Un cop dins (figura A4.8), es tria "RLS" de la llista de programes que ofereix el programa i es prem "Agregar o quitar...". El programa demanarà confirmació i desinstal·larà l'aplicació.

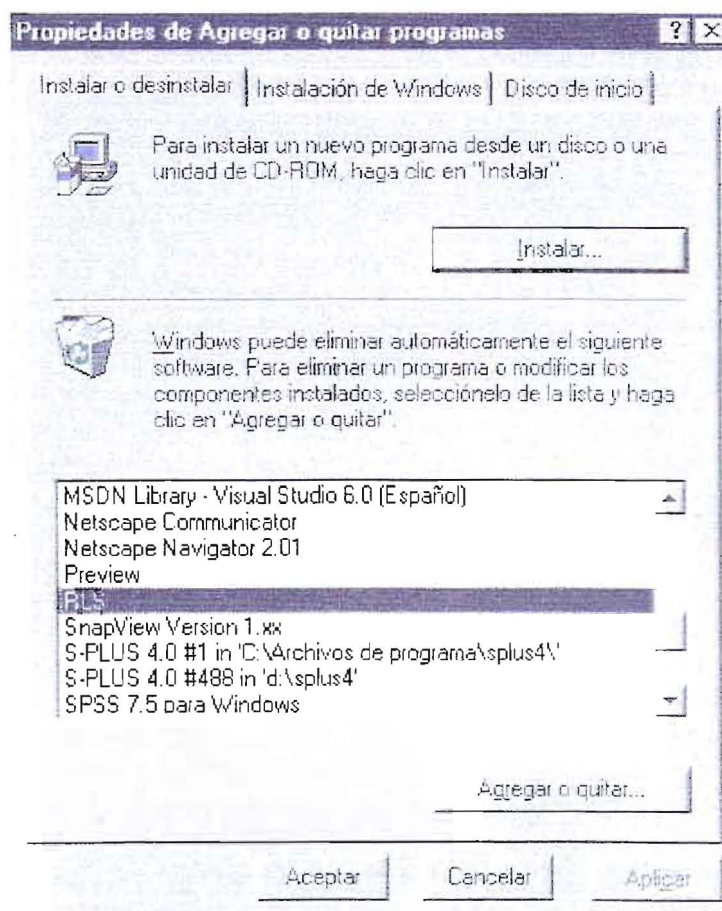


Figura A4.8. Agregar o quitar programas.

Bibliografia

BIBLIOGRAFIA

1. Wonnacott, T.H ; Wonnacott, R.J. (1997). Introducción a la estadística. 5ª ed. Ed. LIMUSA.
2. Peña, D (1989). Estadística. Modelos y métodos. 2. Modelos lineales y series temporales. 2ª ed. Ed. ALIANZA
3. Chatterjee, S; Price, B. (1991). Regression analysis by example. 2ª ed. Ed. Wiley-Interscience.
4. Marriott, F.H.C. (1990). A Dictionary of statistical terms. 5ª ed. Ed. Wiley
5. Microsoft Corporation.(1998). Microsoft Visual Basic 6.0. Manual del programador. Ed. McGraw-Hill.
6. Apunts de la diplomatura.