

R Commander eta Datuen Analisia

Paula Elosua Oliden

Juan Etxeberria Murgiondo



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea



© Udako Euskal Unibertsitatea

© Paula Elosua Oliden, Juan Etxeberria Murgiondo

ISBN: 978-84-8438-447-2

Lege-gordailua: BI-170-2013

Inprimategia: PRINTHAUS S.L., Bilbo

Azalaren diseinua: Igor Markaida

Hizkuntza-zuzenketen arduraduna: Ander Altuna Gabiola

Banatzaileak: UEU. Erribera 14, 1. D BILBO telf. 946790546 Faxe. 944793039

Helbide elektronikoa: argitalpenak@ueu.org

www.ueu.org

Elkar Banaketa: Igerabide, 88 DONOSTIA

Galarazita dago liburu honen kopia egitea, osoa nahiz zatikakoa, edozein modutara delarik ere, edizio honen Copyright-jabeen baimenik gabe.

Aurkibidea

HITZAURREA.....	7
------------------------	----------

I. ATALA

1. R?.....	11
-------------------	-----------

2. R ESKURATZEA ETA INSTALATZEA.....	13
---	-----------

3. NOLA EGITEN DA LANA R-REKIN?	17
--	-----------

3.1. Interfaze grafikoak. R-rekin ingurune intuitiboetan lan egitea	18
---	----

3.2. Programaziorako interfazeak	18
--	----

4. PAKETEAK	21
--------------------------	-----------

4.1. Paketeak instalatzea	21
---------------------------------	----

4.2. Nola kargatu paketeak.....	22
---------------------------------	----

4.3. Paketeak biltegi birtualetan bilatzea	23
--	----

4.4. Paketeak sailkatzea	23
--------------------------------	----

5. R COMMANDER.....	27
----------------------------	-----------

5.1. R Commander eskuratzea eta instalatzea.....	27
--	----

5.2. R Commander-en deskribapena	29
--	----

5.3. Menu-barra.....	30
----------------------	----

5.4. R Commander-en konfigurazioa	31
---	----

5.5. Laneko direktorioa	32
-------------------------------	----

5.6. Objektuak izendatzea	33
---------------------------------	----

5.7. R Commander zuzenean exekutatzea	33
---	----

6. NOLA SARTU DITZAKET DATUAK R COMMANDER-EN?.....	35
---	-----------

6.1. Zuzenean datuak idaztea	35
------------------------------------	----

6.1.1. Nire datuak editatu eta aldatu al ditzaket?.....	37
---	----

6.2. Datu multzoak kargatzea.....	38
-----------------------------------	----

6.2.1. Geuk sortutako datu multzoak kargatzea	38
6.2.2. Euren jatorria estatistikako paketeetan duten datu multzoak irekitzea	39
6.3. Datuak inportatzea	41
6.3.1. SPSSko datu-artxiboak inportatzea	41
6.3.2. ASCII formatua duten datuak irakurtzea	43
6.3.3. Excel formatua duten datuak irakurtzea	45
6.4. Egin al dezaket lana hainbat datu multzorekin aldi berean?	45
6.5. Nola gorde ditzaket datuak?	46
6.6. Adibideko datuen aurkezpena	47
7. DATU MULTZOA MANIPULATZEA.....	51
7.1. Datu multzoa hautatzea	51
7.2. Datu multzoa eguneratzea	52
7.3. Datu multzoaren inguruko laguntza	52
7.4. Aktibo dagoen datu multzoko aldagaiak	53
7.5. Kasuei izenak esleitzea	53
7.6. Kasuak iragaztea edo aukeratzea	53
7.7. Lerroak ezabatzea	57
7.8. Balio galduak dituzten kasuak ezabatzea	57
8. ALDAGAIK SORTZEA ETA ALDATZEA	59
8.1. Aldagaiak birkodetzea	59
8.2. Aldagai berriak sortzea	64
8.3. Lerroei zenbakiak esleitzea	65
8.4. Aldagaiak tipifikatzea	65
8.5. Zenbakizko aldagaiak faktore bihurtzea	66
8.6. Zenbakizko aldagaiak segmentutan banatzea	68
8.7. Faktoreen aldagaiak berrantolatzea	69
8.8. Aldagaien izenak aldatzea	69
8.9. Aldagaiak ezabatzea	69
9. R COMMANDER-EN LUZAPENAK.....	71
9.1. RcmdrPlugin.Export	72
9.2. RcmdrPlugin.TeachingDemos	75

II. ATALA

10. AZTERKETA ESTATISTIKOAK R COMMANDER-EN	79
10.1. Laburpenak	79
10.1.1. Aktibatutako datu multzoa	80
10.1.2. Zenbakizko laburpenak	81
10.1.3. Maiztasunen banaketa	82
10.1.4. Balio galduak	83
10.1.5. Estatistikoen taulak	84
10.1.6. Korrelazio-matrizea	85
10.1.7. Esanguratsutasunaren arazoa	88
10.1.8. Korrelazio-testa	91
10.1.9. Shapiro-Wilken normaltasunaren testa	92
10.2. Kontingentzia-etaulak	93
10.2.1. Sarrera bikoitzeko taulak	93
10.2.2. Sarrera anitzeko taula	95
10.2.3. Sarrera bikoitzeko taula bat eraikitzea eta aztertzea	96
10.3. Batezbestekoak	98
10.3.1. Talde batentzako t testa	98
10.3.2. Lagin independenteentzako t testa	100
10.3.3. Erlazionatutako datuentzako t proba	102
10.3.4. Faktore bakarreko ANOVA	105
10.3.5. Faktore anitzeko ANOVA	111
10.4. Proportzioak	116
10.5. Bariantzak	119
10.6. Test ez-parametrikokoak	121
10.6.1. Wilcoxonon testa (bi talde independente)	122
10.6.2. Parekatutako taldeentzako Wilcoxonon testa	123
10.6.3. Kruskal-Wallis test (k talde independente)	124
10.6.4. Friedmanen testa (parekatutako k talde)	125
10.7. Dimentsio-analisisa	126
10.7.1. Eskalaren fidagarritasuna	127
10.7.2. Faktore-analisisa	129
10.7.3. Osagai nagusien analisisa	134
10.7.4. Taldekatze-analisisa	136
10.8. Ereduen doiketa	144
10.8.1. Erregresio anizkoitza	144

11. EREDUAK – EREDU LINEAL OROKORRA.....	149
11.1. Ereduak. Erregresioaren analisia	151
11.2. Ereduak. Bariantzaren analisia.....	157
12. GRAFIKOAK	159
12.1. Grafiko sekuentziala.....	160
12.2. Histograma.....	162
12.3. Adar eta hostoen grafikoa	165
12.4. Kaxen diagramak	166
12.5. Kuantilak konparatzeko grafikoa	168
12.6. Banaketaren diagrama. Banaketen diagramen matrizea	169
12.7. XY grafikoa.....	170
12.8. Batezbestekoen grafikoa	172
12.9. Barren eta sektoreen grafikoa.....	174
12.10. 3D grafikoa.....	175
13. BANAKETAK.....	179
13.1. Banaketa normala.....	179
13.1.1. Kuantil normalak	179
13.1.2. Probabilitateen kalkulua.....	180
13.1.3. Banaketa normalaren grafikoa (dentsitatearen funtzioa)	181
13.1.4. Banaketa baten zorizko lagina	182
13.2. Banaketa binomiala.....	184
13.2.1. Kuantil binomialak.....	184
13.2.2. Metatutako probabilitate binomialak	185
13.2.3. Probabilitate binomialak	186
13.2.4. Banaketa binomialaren grafikoa	186
13.2.5. Banaketa binomial baten lagina	187
14. TEACHINGDEMOS. ESTATISTIKA IRAKASTEKO ERREMINTA ..	191
14.1. Limite zentralaren teorema	192
14.2. Populazioaren batezbestekoentzako konfiantza-tarteak.....	195
14.3. Testaren potentzia	196
14.4. Erregresio lineal bakuna	199
14.5. Korrelazio-koefizienteak.....	200
15. ERREFERENTZIA BIBLIOGRAFIKOAK.....	203

Hitzaurrea

Gizarte eta Osasun Zientzietan datu estatistikoak IBM-SPSS, SAS, STATISTICA, SYSTAT... pakete informatikoen laguntzarekin aztertu eta maneiatu ohi dira. Erabiltzaile berriarentzako grafikoak sortzea eta manipulatzeko errazagoa izan dadin, menu eta leiho hedagarriek osatzen dituzten sistemez baliatzen dira. Maneiatzeko errazak direla, interfaze grafiko intuitiboak dituztela eta erreminta eraginkorrak direla argi dago. Baina zoritxarrez, are eta garbiago dago pakete estatistiko-informatiko hauek ikasle eta erabiltzaile gehienok ordaindu dezakeguna baino gehiago balio duten lizentzia pribatiboekin saltzen dituztela.

Dena den, ez dira datuak aztertzeko eta maneiatzeko aukera bakarrak, ezta egokienak ere. 90eko hamarkadaren amaieran R agertu eta garatzen hasi zenetik, goitik behera aldatzen ari dira eredu estatistikoak eta grafikoak eraikitzeke, datuak aztertzeko eta estatistikan simulatzeko eta programatzeko bideak. R, datuak aztertzeko eta grafikoak sortzeko, lengoaia irekia duen programa da. Baina horretaz gain, softwarea erabiltzeko askatasuna eta dohakotasuna bermatzen dituen GNU lizentziarekin banatzen da. Alegia, R ASKEA eta DOAKOA da.

Izan al liteke tresna bat doakoa eta eraginkorra aldi berean? Bai, noski. Ezagutzaren alor desberdinetako ikertzaileek egunero egiten dituzten ekarpenei esker ari da R garatzen. Eskaintzen dituen eredu eta analisisien barietatea eta kopurua izugarria da, ezin da ezta alderatu ere software pribatiboaren eskaintzarekin. Zentzu horretan, ez dago R-rekin lehiatu dezakeen pakete komertzialik. Hain da baliotsua eta hain dago gaurkotua, ezen pakete komertzial garrantzitsuenak R-rekin konektatzeko moduluak sortzen baitaibiltza.

Baina R-rekin lan egiteko modu naturala ez da interfaze grafikoetan oinarritzen. Hau erlazionatuta egon liteke R-k bere hedapenean aurkitu duen erresistentziarekin. Izan ere, R maneiatzen ikasteko beharrezko denbora beste pakete komertzialekin behar dena baino gehiago luzatzen da. Baina haren hedapena geldotu duen eta batzuentzat eragozpena izan litekeen ezaugarri hori, R `Commander` interfaze grafikoa erabiliz gainditu daiteke.

R `Commander` menu eta leiho hedagarrietan oinarritzen den paketea da. Programazio-ingurune intuitibo baten bidez, erabiltzaileari datuak kudeatzea eta eredu estatistiko eta grafikoak sortzea errazten dio. R `Commander`-ek barneratzen dituen edukiak etengabe zabaltzen, garatzen eta hobetzen ari dira. Ez dago

zalantzarik, Gizarte eta Osasun Zientzietan datuen analisia irakasteko nahiz ikerketa aplikaturako tresna moduan, R Commander-en potentziala bikaina da.

Aurkezten dugun testuaren helburua R Commander-entzako gidaliburu bat eskeintzea da. R Commander nola erabiltzen eta aplikatzen den azaltzeaz gain, datuak aztertzeko eta emaitzak interpretatzeko oinarri teorikoak ere aurkeztuko ditugu. Kontuan izan murgilduta gauden informazioaren gizarte honetan, gutxienez, emaitza estatistikoen «irakurleak» garela. Idazki hau zuei zuzenduta dago. Liburuaren ikuspegia erabat praktikoa da. Gure asmoa estatistikaren inguruko oinarritzko ezagutzak dituzten eta R Commander-ekin lan egin nahi duten erabiltzaileentzat gida probetxugarria aurkeztea da.

Liburuak bi atal ditu. Lehenengoak R Commander interfaze grafikoari sarrera ematen dio. Instalatzeko prozesua azaltzen du eta erabiltzeko funtsezko jarraibideak ematen ditu. Interfazearen funtzionamenduaren oinarriak eta logika azaltzen ditu, baita datuak lantzeko eta kudeatzeko integratzen dituen aukerak erakutsi ere.

Bigarren atalaren bidez, R Commander-en analisi estatistikoak egiteko eta grafikoak eskuratzeko bideak zeintzuk diren azaltzen dugu. Azterketa deskribatzaile sinpleenak erakusten ditugu; baina baita konplexuagoak diren eredu batzuk ere. Gure asmoa ez da eredu bakoitzaren oinarri teorikoetan sakontzea, ezta emaitzak interpretatzeko gakoak zehatz-mehatz azaltzea ere. Horretarako hainbat liburuki beharko genituzke. Erakusten ditugunak aurkitu daitezkeen analisi guztien artean gutxi batzuen selekzioa besterik ez da.

Liburu honetako gaia, edukia eta prozedura nazioartean probatua izan den eta etengabe zabaltzen eta hobetzen ari den tresna batekin lan egin nahi duten Gizarte eta Osasun Zientzietako alorretako unibertsitateko ikasle eta irakasleei zuzentzen zaizkie.

Gidaliburu honen garapenean zehar kontuan izan dugu Confuciori esleitu izan ohi zaion esaera txinatar hau: «Entzun eta ahaztu egiten dut, ikusi eta buruan gordetzen dut, egin eta ulertzen dut». Testu hau bertan aurkezten diren azterketak burutzeko eta ulertzeko prozesuan erabilgarria izatea gustatuko litzaiguke.

Lan hau Euskal Herriko Unibertsitateak finantzaturako ikerketa baten testuinguruan dago (GIU12-32), Ekonomia eta Lehiakortasun Ministerioaren laguntzez gauzatu da (PSI2011-30256) eta Donostia Research on Education and Multilingualism (DREAM) IT362-10 eta UFI 11/54 barnean kokatzen da.

I. ATALA

1. R?

Noski. Doakoa den, kode librea duen eta beti eguneratuta dagoen software bat erabiliz datuak aztertu, grafikoak edo kodeak sortu edota estatistika ikasi eta irakatsi nahi al dugun galdetzen badigute, nola esan ezetz? Eskaera horiek betetzen dituen ingurune bakarra R da.

R (Ihaka eta Gentleman, 1996; R Development Core Team, 2011), formalki, S programatzeko lengoaiatik (Becker, Chambers eta Wilks, 1988; Chambers, 1998; Chambers eta Hastie, 1992; Venables eta Ripley, 2000) eratorritako programaziorako eta azterketa estatistikorako eta grafikorako ingurunea da. Auckland (Zeelanda Berria) Unibertsitateko Estatistika Saileko **Ross Ihaka** eta **Robert Gentleman**ek garatu zuten. R izena bere autoreen inisialetatik eta software askearen filosofiarekin erlazionatzen duen R-ren *our* ('gure' ingelesez) fonetikatik eratortzen da. R-ren lehen bertsioa oso azkar hedatu zen, eta gaur egun bere hedapena geldiezin da (Elosua, 2009). R ezagutzaren alor guztietako ikertzaileek egindako irabazi-askorik gabeko ekarpenekin hazten da. 1997. urtean osatu zen *R Development Core Team* taldea arduratzen da R eguneratzeaz.

R GNU *General Public Licence* ('Lizentzia Publiko Orokorra' ingelesez) proiektuan sartuta dago. GNU proiektuaren xedea bi ardatzetan banatzen da. Batetik softwarea aske erabiltzea, aldatzea eta banatzea aldarrikatzen du. Eta bestetik, bere erabilera mugatu lezaketen pribatizazio-saiekeretatik babestu nahi du (<http://www.gnu.org/copyleft/gpl.html>). Lizentzia horrekin, beste batzuen artean, *OpenOffice* suite ofimatikoa, *Mozilla* nabigatzailea, *wikipediako* artikuluak, *GNU/Linux* sistema eragilea eta *Emacs* testu-editorea banatzen dira.

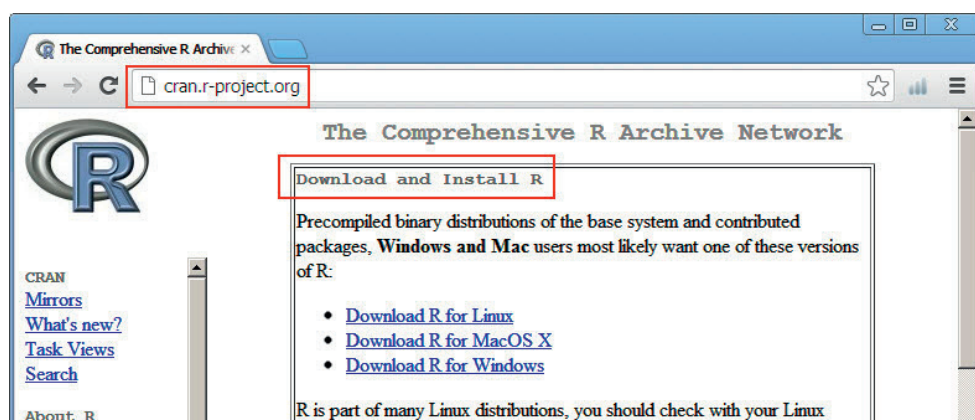
R-ren ezaugarriak:

- Askea da.
- Doakoa da.
- Zabala (2.500 pakete baino gehiago!) da.
- S lengoaian oinarritua dago.
- Konputazioaren lengoia irekia du.
- Pakete komertzialek ez dituzten funtzioak ditu.
- Ezagutza-arlo desberdinetako ikertzaileen arteko lankidetzaz errazten du.
- Tresna grafiko ahaltsua da.

2. R eskuratzea eta instalatzea

R-rekin lan egiteko lehen pausoa programa lortzea eta instalatzea da.

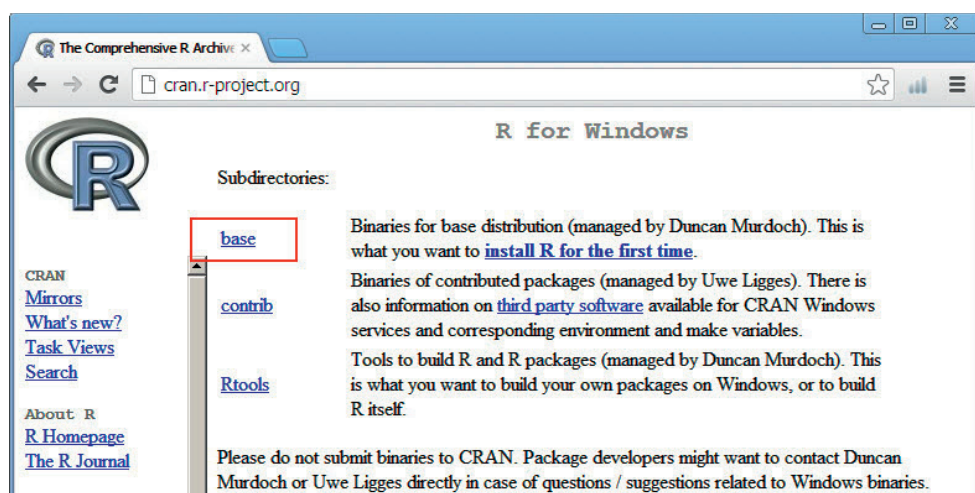
- R instalatzeko artxiboa R-ren webgune ofizialetik (<http://cran.r-project.org/>) jaitsi dezakezu. *Download and Install R* izeneko koadroan dago.



1. irudia.

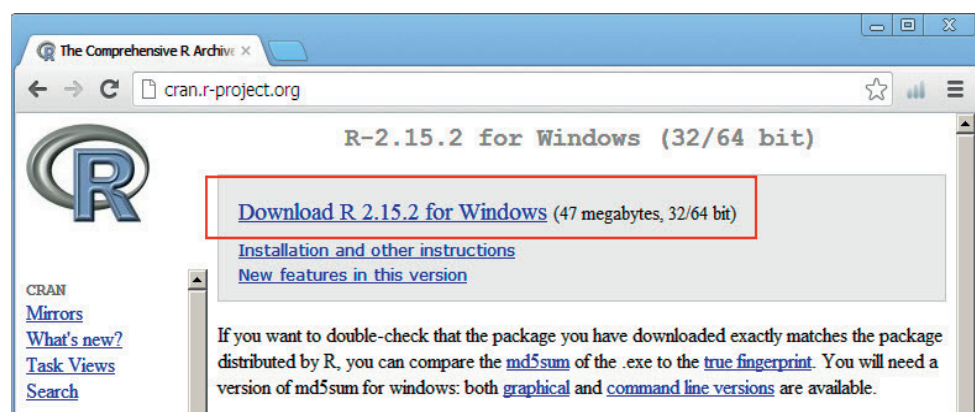
<http://cran.r-project.org/mirrors.html> webgunearen bidez aurkitu daitezkeen CRANeko ispiluetatik ere eskuratu daiteke R instalatzeko artxiboa.

- Sakatu zure sistema eragileari dagokion plataforma (Linux, MacOSX, Windows); eta hurrengo leihoan sakatu *base*.



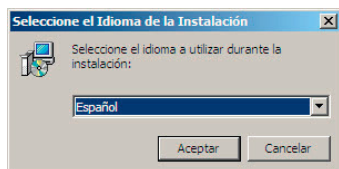
2. irudia.

- Pantaila berri batetik R-ren azken bertsioa jaitsi ahal izango duzu (gidali-buru hau idazten ari garela, *R-2.15.2*).

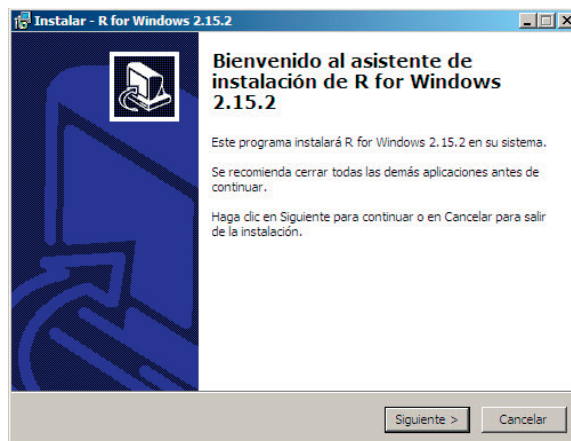


3. irudia.

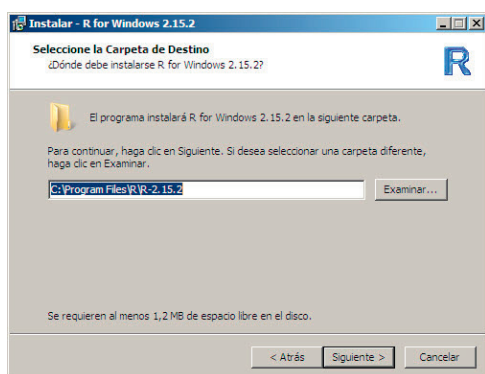
- Instalazio-prozesua automatikoa da. Agertzen doazen leiho desberdinetan berezko balioak onartu (Aceptar, Siguiente>... sakatuz) Opciones de Configuración izeneko pantaila agertu arte. Bertan SI botoia sakatu. Gero, Opciones de Display izeneko pantailan, SDI aukeratu. Jarraitu berezko balioak onartzen instalazioa amaitu arte.



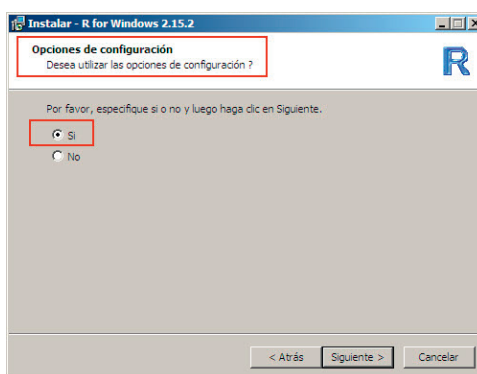
4. irudia.



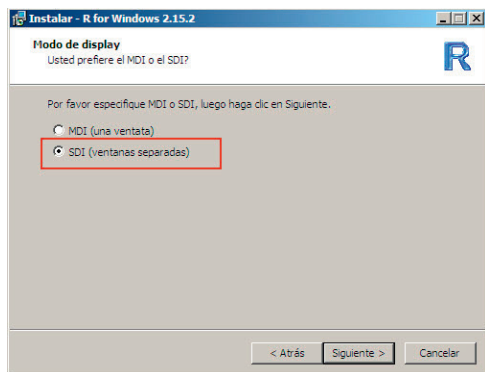
5. irudia.



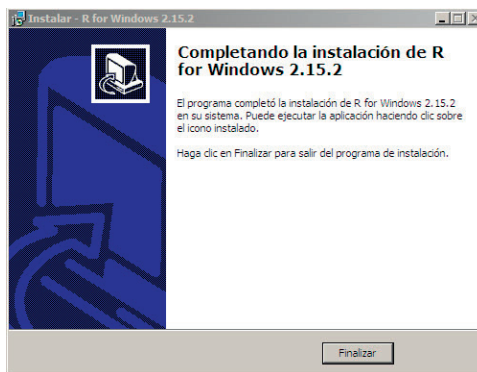
6. irudia.



7. irudia.



8. irudia.

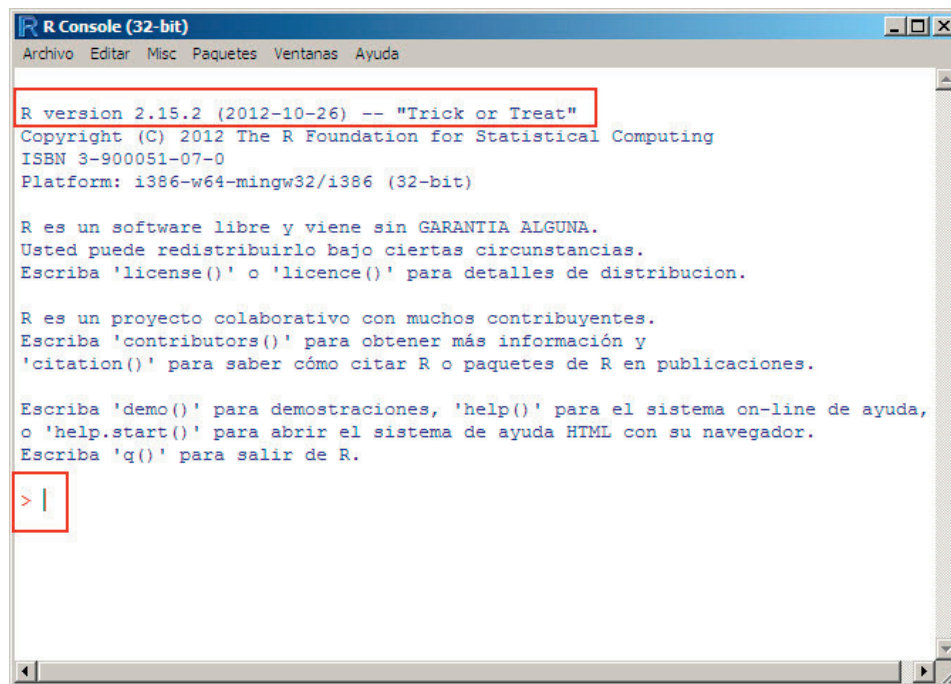


9. irudia.

- Behin R instalatuta dagoela, bere ikonoa mahaigainean agertuko da.

3. Nola egiten da lana R-rekin?

R exekutatzean agertzen den lehenengo pantailari R kontsola (R console) deituko diogu. Bertan instalatutako bertsioa zein den ikusi daiteke (R version 2.15.2). R Kontsola-n sistemako ikurra edo *prompta* kolore gorritz irudikatua (>) dago. R prest dagoela komandoak jasotzeko adierazten du.



```
R Console (32-bit)
Archivo  Editar  Misc  Paquetes  Ventanas  Ayuda

R version 2.15.2 (2012-10-26) -- "Trick or Treat"
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-w64-mingw32/i386 (32-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |
```

10. irudia.

R lehen aldiz irekitzean, harrituko zaitu hain aukera gutxi dituen interfazea aurkitzeak. Menu-barrak dituen aukerak oinarritzkoak dira. Ez dago datuak aztertzeke edo grafikoak sortzeko menurik. Hau al da nahi genuena? Argi dago ezetz.

Berez, R kodeak sortzeko diseinatuta dago. Baina erabiltzaile kopuru handi batentzat egokiagoak izan daitezkeen interfaze intuitiboak erabil daitezke. Azken finean, haiei zuzentzen zaie gidaliburu hau.

3.1. INTERFAZE GRAFIKOAK. R-REKIN INGURUNE INTUITIBOETAN LAN EGITEA

Kodeen interfaze grafikoekin harremanik izan ez dutenek eta menu klasikoetan oinarritzen diren interfaze grafikoetan erosoago sentitzen direnek, hainbat lan-aukera dituzte. Izan ere, hainbat GUI (*Graphical User Interfaz*) daude eskuragarri:

- *R.NET* (<http://www.u.arizona.edu/~ryckman/RNet.php>).
- *Poor Man's GUI* (<http://www.math.csi.cuny.edu/pmg>).
- *Rkward* (<http://es.wikipedia.org/wiki/RKward>). Oraindik ez dago Windowsen erabiltzeko prest.
- *R Commander* (Fox, 2005).
- *RExcel* (<http://www.statconn.com>). Lizentzia pribatiboarekin saltzen da.

Horien guztien artean hedatuen dagoena *R Commander* da. Haren kalitate aipagarrienen artean, Gizarte Zientzietan ohikoenak diren eduki metodologikoen egokitasuna eta erabilerraztasuna daude. Pakete komertzialen erabiltzailea *R* programazio-ingurunera gerturatzeko bitartekari bikaina da, bien arteko bidea izugarri errazten baitu.

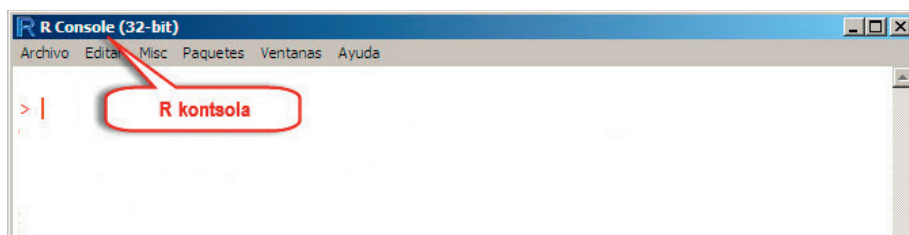
R Commander-en abantailak:

- Askea da.
- Doakoa da.
- Interfaz intuitiboa (leihoen sistema) du.
- Aldagai bakarreko ereduak azertu daitezke.
- Aldagai anitzeko ereduak azertu daitezke.
- Grafikoekin erlazionatutako aukera ugari ditu.
- Estatistika ikasteko edo irakasteko luzapenak ditu.
- Analisi konplexuagoak gauzatzeko luzapenak ditu.

3.2. PROGRAMAZIORAKO INTERFAZEAK

Kodeak sortzeko hainbat aukera daude:

- Zuzenean *R* kontsola-n (*R console*) lan egitea sistemako ikurraren (>) alboan komandoak idatziz, ekintza sinpleak burutzeko metodo baliagarria da; izan ere, aldiko komando bakarra exekutatu daiteke. Edozein testu-editoretan egin daitekeen moduan, hemen ere komandoak kopiatu eta itsatsi daitezke *Ctrl+c* eta *Ctrl+v* funtzioekin.



11. irudia.

- Idazketaren leihoetan (*Script window*) lan egitea. Horretarako menu-barran Archivo > Nuevo Script sakatuko dugu. Aukera hau komando konplexuagoak edota komando-sortak idazteko eta exekutatzeko erabilgarria da. Kodeak zuzenean leiho honetan teklatu daitezke. Beste aukera bat ASCII formatua duen artxibo bateko kodeak kopiatzea eta idazketaren leihoan itsastea izango litzateke. Intereseko komando-sorta aukeratu genuke, eta jarraian *Ctrl+r* funtzioa erabili edo  irudia sakatu exekutatzeko. Idazketaren leihoa erabiltzen denean, irteerak R-ko kotsolan agertzen dira. Hortaz, gomendagarria da aldi berean bi leihoak zabalduta edukitzea: R-ko kotsola eta komandoen leihoa.



12. irudia.

- Kodeak eraikitzeako interfazeak. Kodeak eraikitzeako eta editatzeko programa bereziak daude. R-ren erabiltzaile aurreratuek gogokoen duten aukera da. Erabilienetakoen artean hurrengoak aurkitzen dira:
 - Tinn-R, (<http://www.sciviews.org/Tinn-R/>).
 - WinEdit (<http://www.winedt.com/>).
 - Emacs (<http://www.gnu.org/software/emacs/>). Hasiera batean GNU filosofian oinarritzen den Unix plataforman erabiltzeko sortu zen.

4. Paketeak

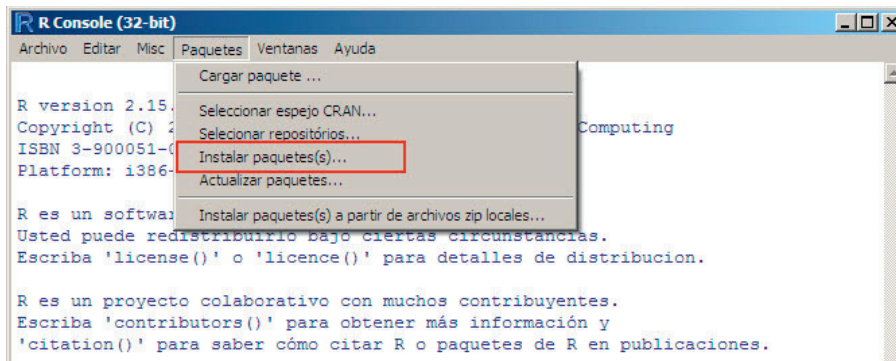
R sistema integratua da, *package* moduan ezagutzen ditugun osagarriak gehitu diezazkiokegu. *Package* batean erlazionaturik dauden hainbat funtzio, laguntza-artxibo bat, eta datuen artxibo bat edo gehiago daude. R-n integratu daitezkeen *package* kopurua 2.500 baino handiagoa da (2011ko martxoan 2.871 *package* zeuden eskuragarri). *Package*en dibertsitatea ezagutza-alorren, aplikazioen eta eredu estatistikoen dibertsitatearen parekoa da. Mikroarrayak aztertzeko, kredituen arriskuen ereduak sortzeko, Gizarte Zientzietarako, sudokuak egiteko eta ebazteko... *package*ak daude. Horietako batzuk R-ko lehen instalazioan integratzen dira; besteak lortzeko publikoak diren biltegi birtualetara jo behar dugu. *Package*ak erabiltzeko, ordea, beharrezkoa da lehenik eta behin instalatzea eta kargatzea.

4.1. PAKETEAK INSTALATZEA

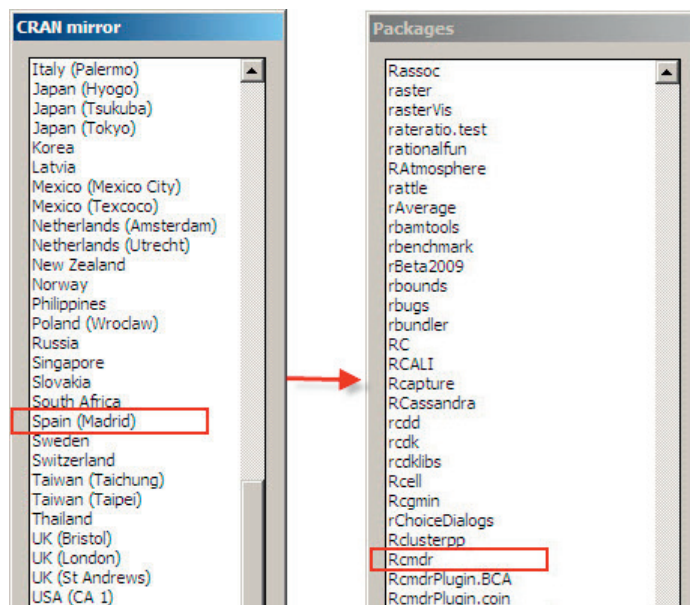
*Package*ak R-n instalatzea oso erraza da. Menu-barraren bidez edo kontsolaren bidez instalatu daitezke:

- Menu-barran `Paquetes > Instalar Paquetes...` sakatuz.
- R kontsola-n zuzenean `install.packages(paketearen izena)` idatziz.

Bai aukera batek, bai besteak, ispiluen edo *mirroren* zerrenda ageri den leiho batera eramango gaitu. Ohikoena geografikoki hurbilen dagoen lekua hautatzea da. Jarraian eskuragarri dauden *package*en zerrenda agertuko da. Hautatu instalatu nahi duzun *package*a eta sakatu OK. R-k jaitsi eta instalatuko du *package*a.



13. irudia.



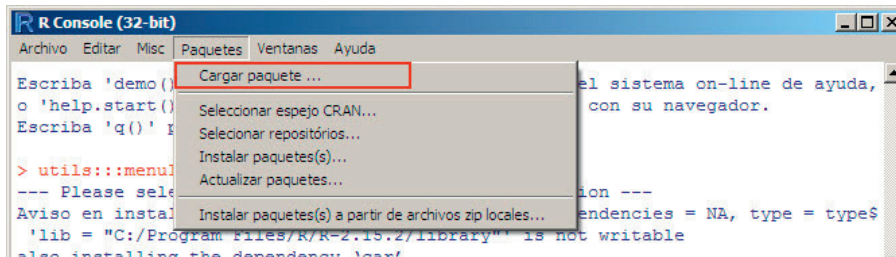
14. irudia.

4.2. NOLA KARGATU PAKETEAK

Paketeak erabiltzeko, normalean beharrezkoa da instalatzeaz gain kargatzea. Instalazioa behin bakarrik gauzatzea nahikoa den arren, pakete bat erabili nahi dugun bakoitzean kargatu beharko dugu. Paketea menu-barraren bidez edo kontsolaren bidez kargatu daiteke:

- Menu-barran Paquetes > Cargar Paquete > paketearen izena sakatuz.

- `library(paketearen izena)` komandoa idatziz.
- `require(paketearen izena)` komandoa idatziz.



15. irudia.

Kargatuta ez dagoen pakete bateko funtzio bat exekutatzen saiatzen bagara, errore baten berri emango digu R-k.

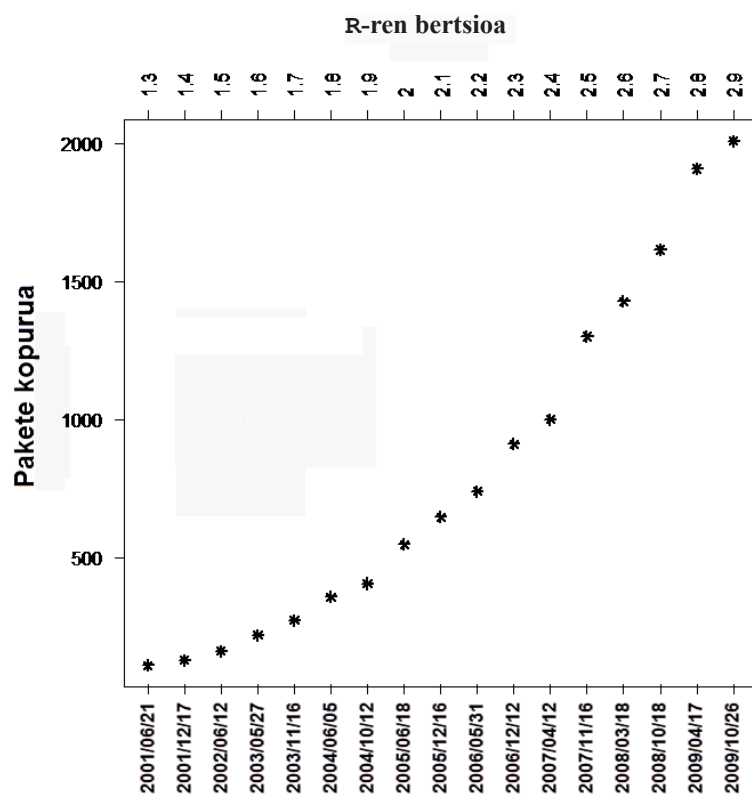
4.3. PAKETEA BILTEGI BIRTUALETAN BILATZEA

R-rekin bateragarriak diren pakete gehienak R-ren webgune ofizialean (CRAN, Comprehensive R Archive Network) eta Bioconductor-en daude. Azken hau, genomikako datuak aztertzeke erremintak sortzeko helburuarekin eratutako proiektu bat da.

Webgunearen helbideak	
CRAN	http://cran.Rproject.org/web/packages
Bioconductor	http://www.bioconductor.org/packages/release/Software.html

4.4. PAKETEA SAILKATZEA

R-ren arrakastaren indikatzailenetako bat eskuragarri dauden pakete kopuruaren etengabeko gehikuntza da. Foxek (2009) eskaintako informazioan oinarritzen den hurrengo grafikoa, 2001. urteko R-1.3 eta 2009. urteko R-2.9 bertsio eta urteen artean pakete kopuruak pairatu duen hazkuntza islatzen du.



16. irudia.

R-n integra daitekeen pakete kopuru handia dela-eta, erabiltzailea orientatzeko sailkapen-sistema sortu da. Paketeak ezagutza-alor bakoitzean duten erabilgarritasunaren arabera sailkatu dira. Horiei dagokien informazioa *Task Views* izena duen CRANeko webgunean (<http://www.cran.r-project.org/web/views/>) aurkitu daiteke.

Informazioaren egitura kategoria hauetan laburtu da:

Paketeen sailkapena	
Bayesian	Estatistika bayesiarra
ChemPhys	Kimimetria eta fisika konputazionala
ClinicalTrials	Entsegu klinikoaren diseinua, monitorizazioa eta azterketa
Cluster	Konglomeratuen eta eredu mixtoen azterketa
Distributions	Banaketak
Econometrics	Ekonometria
Environmetrics	Ekologiarako ereduak
ExperimentalDesign	Diseinu esperimentalak
Finance	Finantzako azterketak
Genetics	Genetikarako ereduak
Graphics	Grafikoak
gR	<i>gRaf</i> oseko ereduak
High Performance Computing	Konputazio intentsiboa
MachineLearning	Makinen ikaskuntza
MedicalImaging	Medikuntzako irudien azterketa
Multivariate	Aldagai anitzeko ereduaren azterketa
NaturalLanguageProcessing	Lengoaia naturalaren prozesamendua
Optimization	Estatistika operatiboa
Pharmacokinetics	Datu farmakinetikoak
Psychometrics	Eredu psikometrikoak
Robust	Metodo sendoak
SocialSciences	Portaera Zientzietara aplikaturiko estatistika
Spatial	Datu espazialen azterketa
Survival	Biziraupenaren azterketa
TimeSeries	Denborazko serieen azterketa

Kategoria bakoitzean eremu zehatz batentzat diseinaturiko pakete multzoak deskribatzen dira. Gizarte Zientzien alorrean, adibidez, Eredu Lineal Orokorrean eta Orokortutako Eredu Linealean, datu kategorikoen azterketan eta beste erregresio ereduetan (erregresio ez-linealak) oinarritzen diren funtzioak barne hartzen dituzten paketeak laburki deskribatuta daude.

5. R Commander

5.1. R Commander ESKURATZEA ETA INSTALATZEA

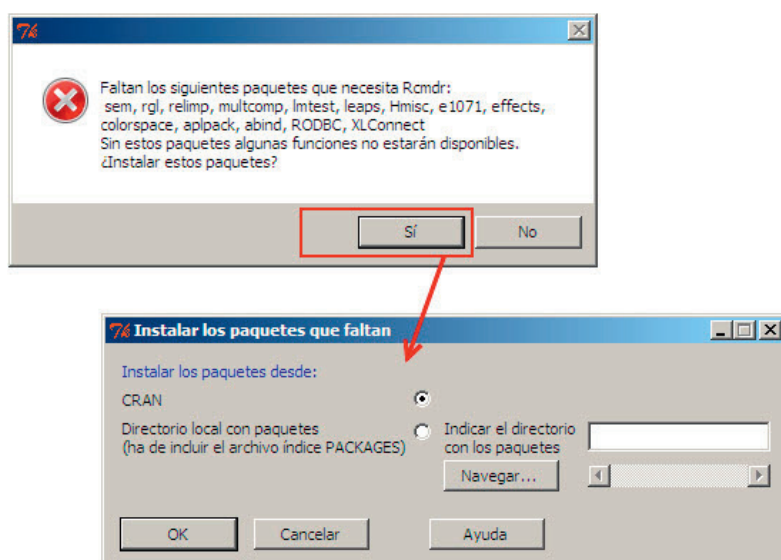
R Commander R-ko paketea da (Fox, 2005, 2007). Zehazkiago, azterketa estatistikorako eta grafikoak sortzeko funtzioak dituen interfaze grafikoa da (*Graphical User Interface*, GUI). Leihuetan oinarritzen den sistema batez baliatzen da R ingurune intuitibo bilakatzeko. R Commander-en laguntzarekin lan egiteak beste edozein pakete estatistikorekin (PASW, SAS, Splus...) lortu daitekeen esperientziara gerturatzen gaitu. R Commander-en menuek artxiboak inportatzeko eta esportatzeko, aldagaiak manipulatzeko (birkodetzeko, kalkulatzeko...), kasuak hautatzeko, datuak deskribatzeko, grafikoak sortzeko eta ereduak doitzeko aukerak dituzte. Aldi berean, R Commander-ekin oinarritzko komandoak aldatu edota komando berriak erantsi daitezke behar zehatzei erantzuna emateko.

R Commander-en gaitasuna (R-renarekin batera) etengabe zabaltzen ari da. Guk erabiliko dugun bertsioa datuen azterketa oinarritzkoenak gauzatzeko eraiki zen. Alegia, R Commander bidez R-ren benetako potentzial estatistikoa eta kudeaketa-potentziala urratu besterik ez dugu egingo. R Commander-i atxikitzen doazkion luzapenek eta barneratzen dituen funtzioek Gizarte Zientzietako ikerketa aplikatuan erabiltzen diren analisi eta ereduak erantzuten diete.

R Commander-ekin lanean hasi aurretik (paketearen izena `Rcmdr` da) beharrezkoa da `Rcmdr` instalatzea (4.1) eta kargatzea (4.2).

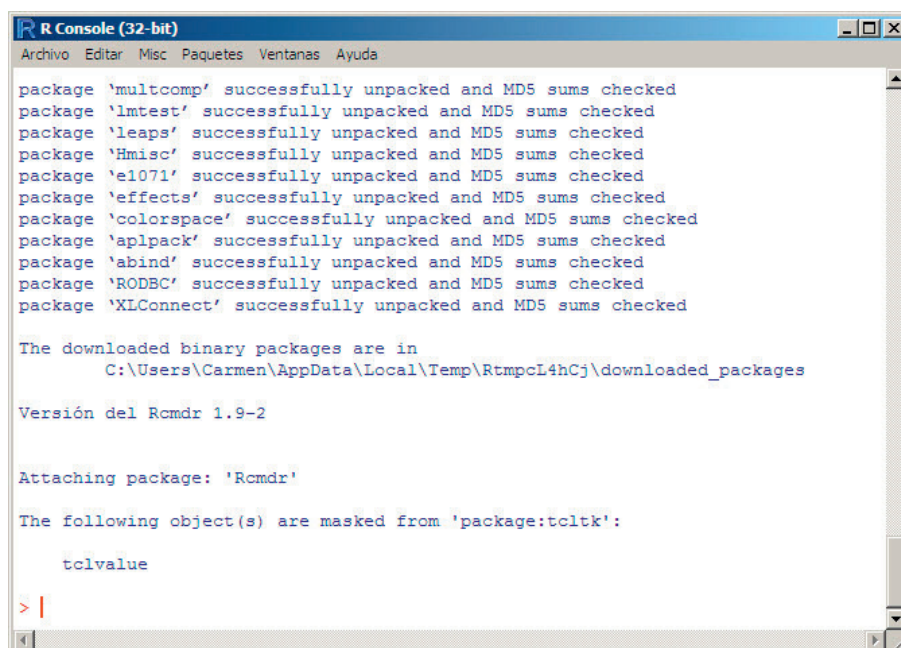
- `Rcmdr` instalatzea. Eskuragarri dauden paketeen artean bila ezazu `Rcmdr`. Hautatu ostean instalazio-prozesua automatikoki hasiko da.
- `Rcmdr` kargatzea. `Paquetes > Cargar` sakatu eta aukeratu `Rcmdr` eskuragarri dauden paketeen zerrendan.

`Rcmdr` kargatzen saiatuko garen lehenengo aldian (lehenengoan soilik), `Rcmdr`-en funtzionamendua egokia izan dadin, beste pakete batzuk instalatzea beharrezkoa dela adieraziko digu. Instalazio-prozesuak beharrezko paketeak automatikoki bilatuko ditu eta erabiltzaileari hauek instalatzeko baimena eskatuko dio. Prozesuaren eskaerak onartu behar ditugu.



17. irudia.

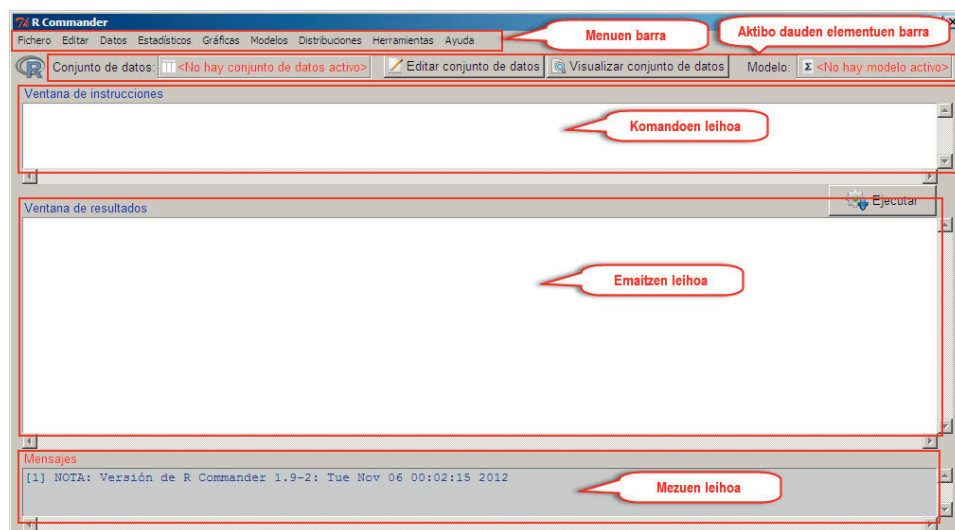
Rcmdr instalatu eta kargatu ostean (instalazioak behar duen denbora ez da beti berdina, hainbat minutu iraun ditzake) instalatzea beharrezkoa izan den paketeen zerrenda eta R Commander-en bertsioa agertuko dira. Kasu honetan, 1.6-3. R-ko kotsolak itxura hau izango du.



18. irudia.

5.2. R Commander-EN DESKRIBAPENA

R Commander-en instalazioa amaitu eta gero, bost ataletan zatitu daitekeen leihoa agertuko da pantailan:



19. irudia.

Menu-barra. Artxiboak manipulatzeko eta deskribatzeko, eredu estatistikoak doitzeko, grafikoak sortzeko eta R Commander konfiguratzeko hainbat aukera biltzen ditu.

Aktibo dauden elementuen barra. R Commander-eko leiho nagusiko bigarren lerroak aktibo dauden elementuen (datuak eta ereduak) berri ematen du. Komandoak eta funtzioak aktibo dagoen datu multzo bakarrari aplika diezazkiokegun arren, lan-saio berean hainbat datu multzorekin egin dezakegu lana gure interesen arabera bata edo bestea aktibatuz. Horretarako, nahikoa da aktibo dagoen datu multzoa zein den adierazten duen leihoan klikatzea eta interesatzen zaiguna aukeratzea. Gainera, R Commander-ek datu multzoak editatzeko eta ikusteko aukerak eskaintzen ditu.

Komandoen leihoa. R Commander-en menuen bidez exekutatutako ekintza guztiek automatikoki komandoak sortzen dituzte. Hauek komandoen leihoan (*script window*) agertuko dira. Gainera, leiho honek edizio-kontsola baten modura funtzionatzen du. Izan ere, zuzenean komandoak idatzi edo aldatu daitezke. Funtzio bat idatzi ondoren nahikoa da dagokion lerroa aukeratzea eta **Ejecutar** sakatzea edo **Ctrl+r** zapaltzea. **Ctrl+a** botoiak konbinatzen baditugu, komandoen leihoko lerro guztiak hautatuta agertuko dira, eta **Ctrl+s** sakatzen badugu, komandoak artxibo batean gordetzeko aukera emango digun leiho bat zabalduko da.

R Commander-en exekututako ekintza guztien ondorioz, hauei dagozkien komandoak agertzen dira komandoen leihoan. Honi esker, R-ko programazio-hizkuntza barneratzen hasi zaitezke; eta horrela, interes eta dedikazio apur batekin, leiho honetako oinarritzko komandoak aldatzen eta manipulatzeko ikasiko duzu.

Emaitzen leihoa. Edizioako leiho honetan, bai exekutututako komandoak (kolore gorritz), baita hauek eragindako emaitzak ere (kolore urdin ilunez) agertuko dira.

Mezuen leihoa. Mezuen leihoak eragindako erroreari (gorritz irudikatutako mezuak), oharrei (kolore berdez irudikatutako mezuak) eta erabiltzen ari garen datu multzoari (kolore urdin ilunez irudikatutako mezuak) buruzko informazioa eskaintzen du.

5.3. MENU-BARRA

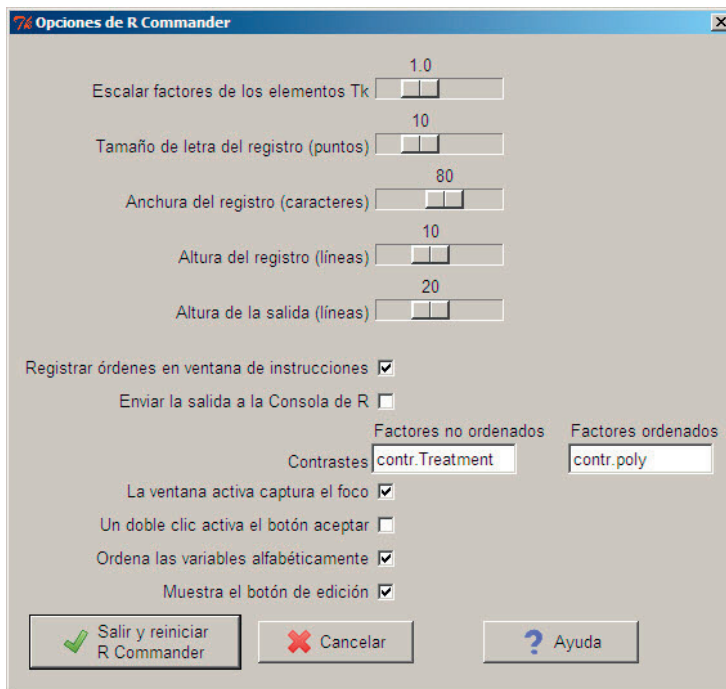
R Commander-eko menu-barrak hurrengo informazio orokorra eskaintzen du:

- **Fichero.** Menu honek barneratzen dituen aukerak R Commander-eko lan-saioak irakurtzearekin eta gordetzearekin erlazionatuta daude. Besterik esan ezean, komandoen artxiboak «R» luzapenarekin gordeko dira, eta emaitzen irteerak «.txt» formatuan. Lan-ingurunearekin erlazionatutako aukerak lan saio batean sortutako objektu guztiak (bektoreak, matrizeak, datu-markoak...) gordetzeko aukera ematen dute hurrengo lan-saio batean berreskuratzea posible izan dadin. Objektu horiek *Datos > Cargar conjunto de datos* sakatuz berreskuratu ahal izango ditugu. Dena den, lan-saioa edozein luzapenekin gorde dezakegu. Kontuan izan, besterik esan ezean, datuak kargatzeko aukerak «.Rda» luzapena erabiltzen duela.
- **Editar.** Menu honek beste edozein editorek jasotzen dituen aukerak ditu (moztu, itsatsi, kopiatu, bilatu...). Aukera horiek komandoen leihoan eta emaitzen leihoan erabili ahal izango ditugu.
- **Datos.** Menu honek beste azpimenu batzuetara eramango gaitu. Azpimenu horiek guztiek, funtsean, datu multzoak inportatzeko eta irakurtzeko eta bertan aurkitzen diren aldagaiak manipulatzeko aukerak eskaintzen dituzte.
- **Estadísticos.** Datos menuaren antzera, Estadísticos menuan azpimenuen zerrenda bat aurkituko dugu. Azpimenu horiek datuak deskribatzeko, oinarritzko azterketa estatistikoak erabiltzeko, parametrikokoak ez diren testak aplikatzeko, fidagarritasuna zenbatesteko, baliagarritasuna aztertzeko eta eredu linealak doitzeko aukerak biltzen dituzte.

- **Gráficas.** Menu honetan sakatuz gero, oinarrizko grafikoak sortzeko aukera-sorta batekin topo egingo dugu. R Commander-ek eskaintzen duen grafiko-aukera oso zabala izateaz gain, beste edozein funtzioekin bezala, behar eta egoera zehatzetara egokitu daitezke grafikoentzako funtzioak ere.
- **Modelos.** Eredu bat zenbatetsi ondoren eredu horretan sakontzeko aukera ematen du R Commander-ek. Horretarako, ereduaren doiketari dagokion informazio kuantitatiboa eta grafikoa eskuratzeko aukera ematen du.
- **Distribuciones.** Menu honek sakabanatze-estatistiko erabilienekin, jarraituekin nahiz jarraituak ez direnekin, jarduteko aukerak biltzen ditu. Besteak beste, datuak sortzeko, grafikoak irudikatzeko eta kuantilak eta probabilitateak zenbatesteko aukerak daude.
- **Herramientas.** R-ko pakete desberdinak kargatzea eta R Commander-ekin erlazionatutako hainbat parametro finkatzea ahalbidetzen duen aukera-sorta dauka menu honek.
- **Ayuda.** Menu honek R Commander erabiltzeko moduari buruzko informazioa duen artxibo batera iristeko aukerak ematen dizkigu.

5.4. R COMMANDER-EN KONFIGURAZIOA

R Commander gure lan-ohituretara egokitzeko, Herramientas > Opciones sakatuz haren itxura nahiz zehaztasun tekniko batzuetan aldaketak egin ditzakegu.

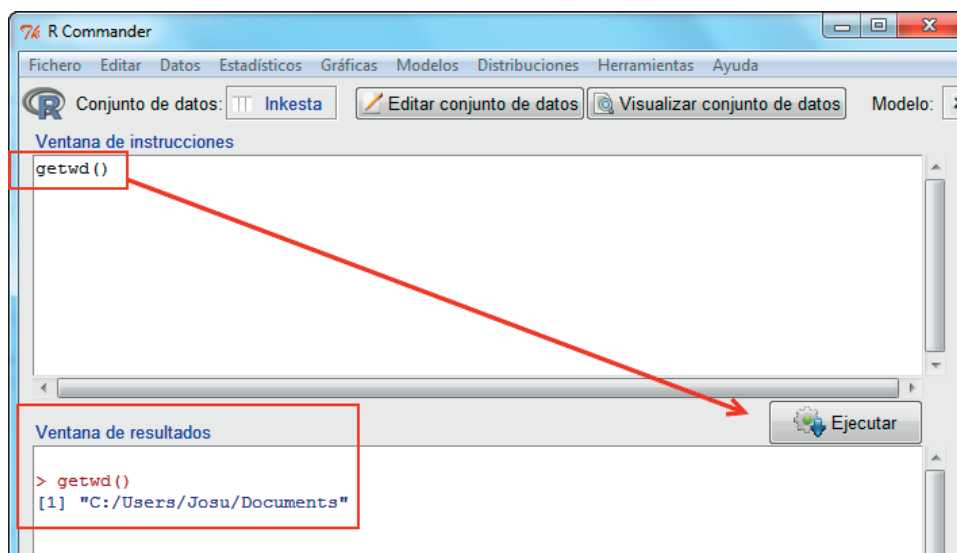


20. irudia.

Ziur asko aldaketa hauen artean praktikoena datu multzoa osatzen duten aldagaien irudikatzearekin erlazionatzen da. Normalean, ikertzaileak aldagaiak sortzeko orduan aurrez finkatutako ordena bati jarraitzen dio. Baina R Commander-ek bere leihoetan, besterik esan ezean, aldagaiak alfabetikoki ordenatzen ditu. Aldaketa hau finkatzeak analisi estatistikoen emaitzak aldatuko ez dituen arren, komenigarria izan liteke interfaze grafikoa norbere lan-ohituretara egokitu dadin.

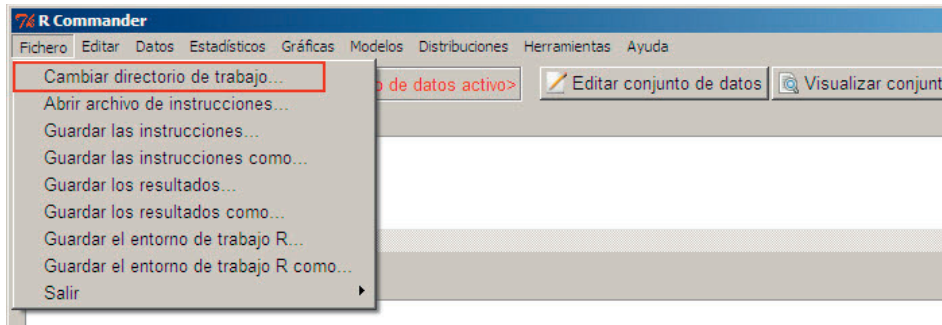
5.5. LANEKO DIREKTORIOA

Datuak, emaitzak eta komandoak gordeak izango diren karpetaen helbidea zein izango den erabiltzaileak zehaztea komenigarria da. Modu horretan denbora aurreztu liteke, eta bide batez, ezustekoren bat saihestu bila gabiltzan artxiboa ez dugulako aurkitzen. R-k artxiboak gordetzeko automatikoki hautatu duen karpeta zein den jakiteko, komandoen leihoan `getwd()` idatzi behar dugu. Komandoa exekutatzean emaitzen leihoan agertuko da laneko direktorioa.



21. irudia.

Laneko direktorioa aldatzeko eta gure beharretara egokitzeko Fichero > Cambiar directorio de trabajo sakatuko dugu.



22. irudia.

Aukera honekin guri interesatzen zaigun laneko direktorioa zein izango den zehazteko aukera izango dugu. Alegia, besterik esan ezean, gorde nahi ditugun objektuak bilduko diren lekua finkatuko dugu.

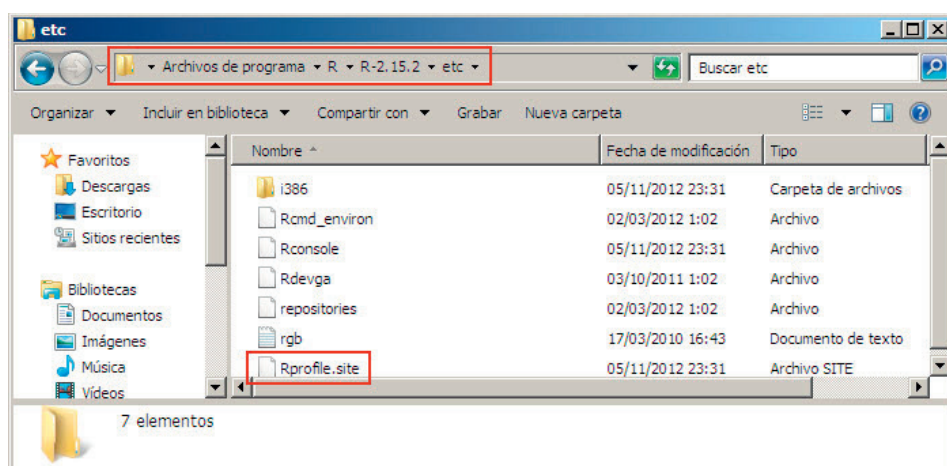
5.6. OBJEKTUAK IZENDATZEA

R-ko objektuei izena emateko hizkiak, zenbakiak eta «.» karakterea erabil daitezke. Kontuan izan beharreko bi ohar: 1) objektu baten izena ezin da zenbaki batekin hasi, eta 2) R-k letra larri eta xeheak desberdintzen ditu.

5.7. R Commander ZUZENEAN EXEKUTATZEA

R Commander exekutatzeke ohiko bideak bi pauso egiten ditu: R kargatzea, eta R Commander kargatzea. Baina ez da beharrezkoa bi pauso horiek egitea. R kargatzean R Commander automatikoki ireki daiteke. Horretarako nahikoa da R instalatu den karpetan (bestarik esan ezean, normalean *Archivos de programa* edo *Program files* karpetan) etc > Rprofile.site artxiboa irekitzea eta jarraian azalduko dugun komandoa idaztea.

Rprofile.site irekitzeko nahikoa da sagua artxiboaren gainean ipintzea eta eskuineko botoia sakatzeko *Bloc de Notas* edo bestelako testu-editore bat aukeratzeko.

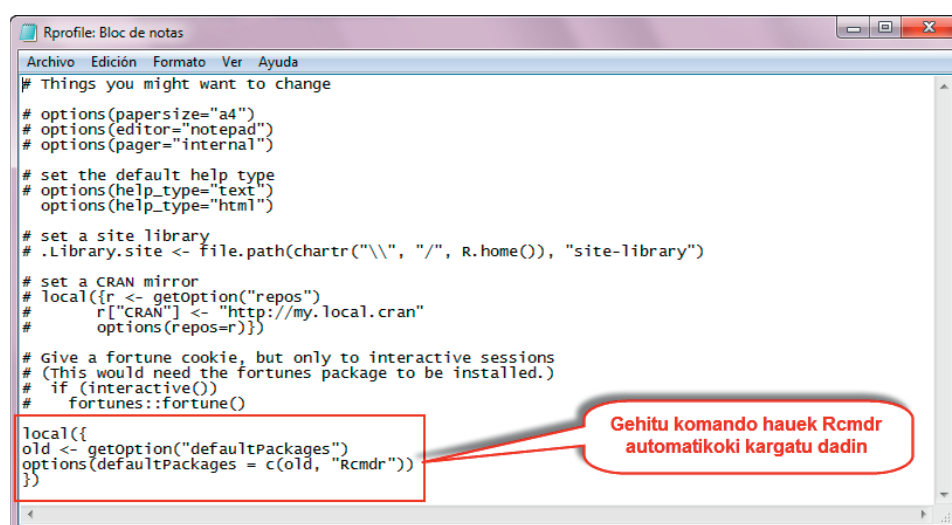


23. irudia.

Hau izango da artxiboan idatzi beharreko komandoa:

```
local({
  old <- getOption("defaultPackages")
  options(defaultPackages = c(old, "Rcmdr"))
})
```

R exekutatzen dugun hurrengo aldian R Commander ere automatikoki ireki dadin, aldaketa hauek gordeko ditugu.



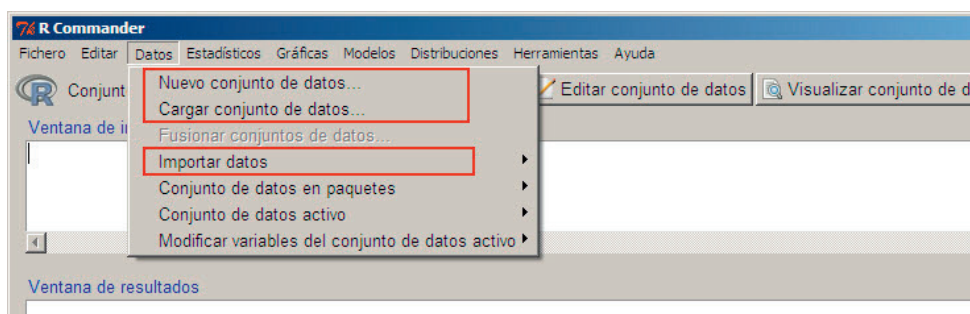
24. irudia.

6. Nola sartu ditzaket datuak R Commander-en?

Datuak aztertzeko edozein prozesuren lehen pausoa datuetara heltzea da. Aztertu nahi ditugun datuak aktibatu egin behar ditugu. R-k ez du erabiltzen datuen artxibo kontzeptua. Haren lekuan, datu multzo (edo datu-marko) kontzeptua erabiltzen du. Datu multzoa mota desberdineko aldagaiek osatzen duten matrizea da.

R Commander-ek datu multzo aktiboa zehazteko hiru aukera ditu: zuzenean datuak idaztea, aurreko lan-saioetan sortutako datuak irakurtzea, eta R-k erabiltzen ez dituen beste formatu batzuetako datuak inportatzea. Aipatutako hiru aukerak Datos menuan aurkituko ditugu.

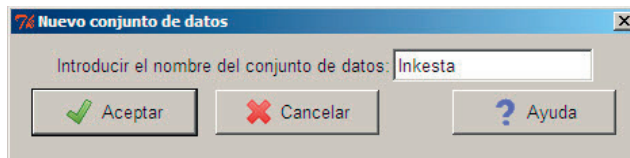
R-k ez du erabiltzen datuen artxibo kontzeptua, datu multzo kontzeptua baizik.



25. irudia.

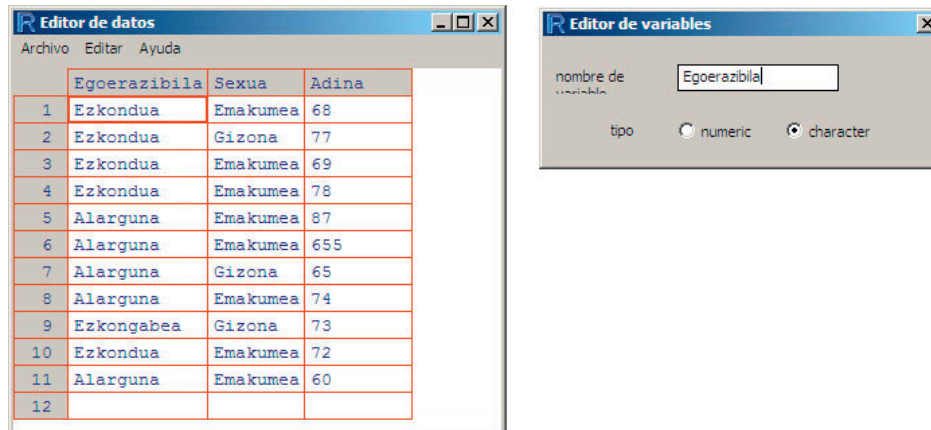
6.1. ZUZENEAN DATUAK IDAZTEA

Zuzenean datuak idazteko Datos > Nuevo Conjunto de Datos klikatu behar da. Orduan leiho bat irekiko da eta bertan datu multzo berriaren izena idatzi beharko dugu.



26. irudia.

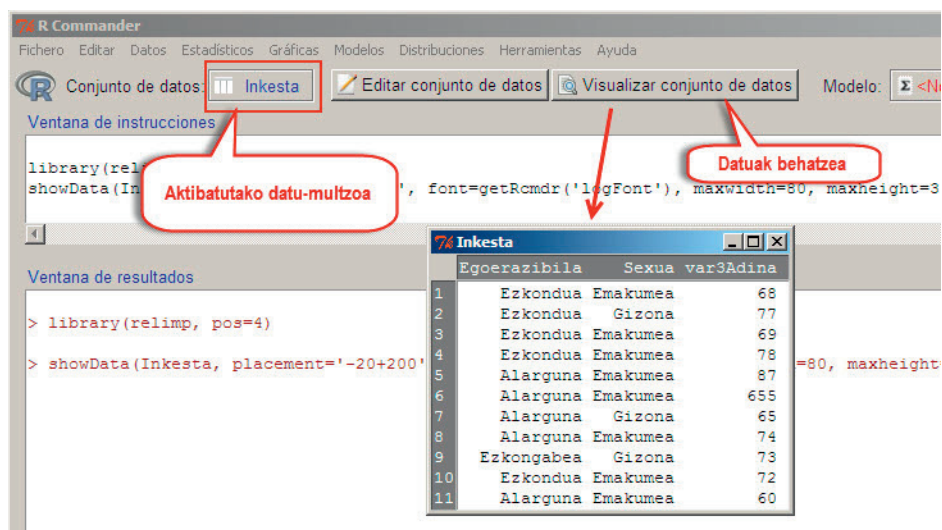
Adibide honetan Inkesta datu multzoa asmatu dugu.



27. irudia.

Datuak editatzeko balio duen leiho honetan aldagai guztiak izendatu ahal izango ditugu, baita zein motatakoak izango diren zehaztu ere (zenbakizkoak edo karakterezkoak). Horretarako editore lehenengo lerroko edozein zutaberen gainean klikatuko dugu (`var1`, `var2`, `var3...`).

Orria aurreratzeko eta atzeratzeko teklek (*AvPag* eta *RePag*) datu multzoan gora eta behera eramango gaituzte, ikusten diren kasuen ondorengoak (*AvPag*) eta aurrekoak (*RePag*) ikusteko. *Inicio* teklak datu-matrizeko lehenengo laukian ipiniko du kurtsorea. Hots, lehenengo lerroa eta zutabea gurutzatzen diren laukian. *Fin* teklak kurtsorea matrizearen azkeneko laukira eramaten du. Zutabeen tamaina dagokien aldagaien balioen tamainari egokitu dakion, zutabe horretako lauki bat aukeratu ostean saguaren eskuineko botoia sakatu eta *Autoajustar columna* sakatuko dugu. Datuen idazketa amaitutakoan Archivo menuko Cerrar aukera sakatuko dugu (Archivo > Cerrar) datu-editoretik irteteko. Orduan, komandoen leihoan Inkesta izena duen datu multzo bat (datuen artxibo bat) sortu dela adierazten duen komandoa agertuko da. Orain, Conjunto de Datos leihoan Inkesta ikusiko dugu. Guk adierazitako funtzioak datu multzo horri aplikatuko zaizkio.



28. irudia.

Visualizar conjunto de datos aukerak aurreko irudian ageri den pantaila erakutsiko du. Bertan, aldagaien izenak (Egoerazibila, Sexua, Adina) eta lehen idatzi ditugun datuak agertuko dira.

R Commander-eko datu-editorea mugatua da. Hortaz, datu multzoan sartu behar dugun datu kopurua txikia denean soilik izango da egokia haren erabilera.

6.1.1. Nire datuak editatu eta aldatu al ditzaket?

Datu-editoreak, datuak idazteko aukeraz gain, hauek aldatzeko aukera ere ematen du. Adibidez, aurreko adibidean balio oker bat idatzi dugu Adina aldagaian (6 lerroa). Beraz, datu-editorean sartuko gara (Editar conjunto de datos), balio okerra duen laukia aukeratuko dugu eta balio zuzena idatziko dugu. Ekintza sinple horrekin gure datuak zuzenduta egongo dira.

	Egoerazibila	Sexua	Adina
1	Ezkondua	Emakumea	68
2	Ezkondua	Gizona	77
3	Ezkondua	Emakumea	69
4	Ezkondua	Emakumea	78
5	Alarguna	Emakumea	87
6	Alarguna	Emakumea	655
7	Alarguna	Gizona	65
8	Alarguna	Emakumea	74
9	Ezkongabea	Gizona	73
10	Ezkondua	Emakumea	72
11	Alarguna	Emakumea	60
12			

29. irudia.

OHARRA

Datos > Editar
Conjunto de Datos
sakatuz datuak aldatu
daitezke.

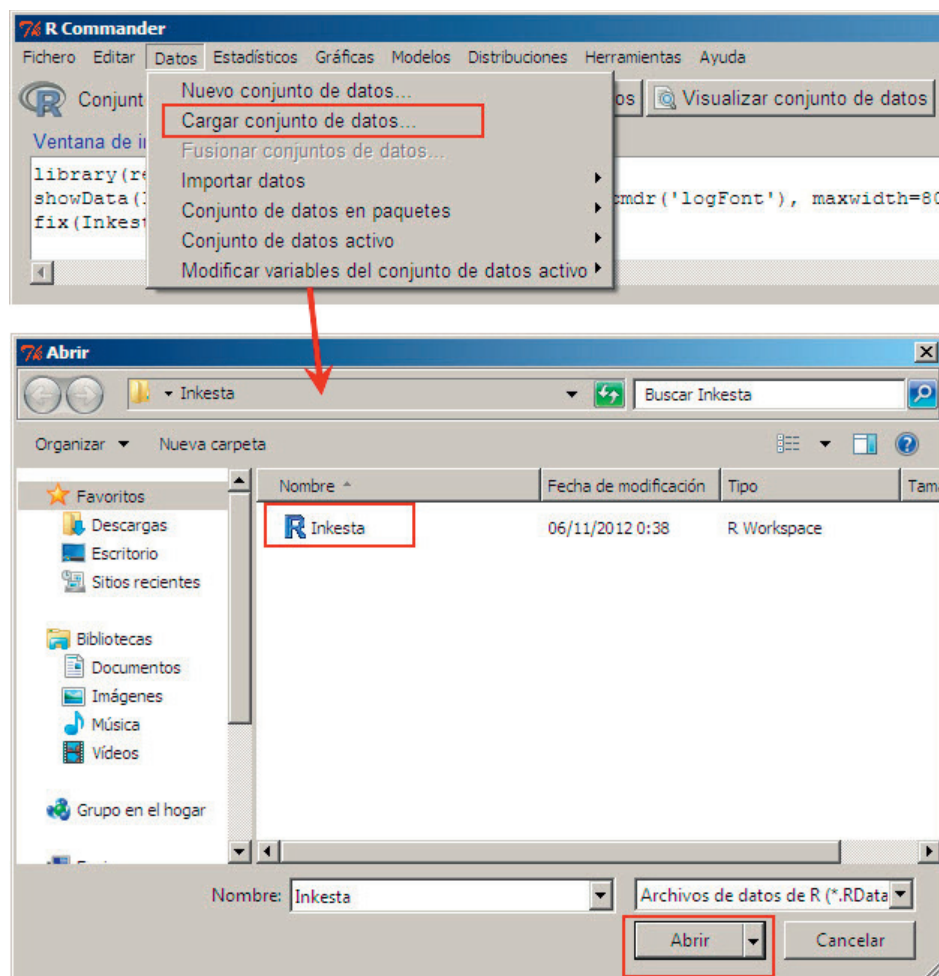
Datuak gordetzeko Datos
> Conjunto de datos
activo > Guardar
conjunto de datos
activo sakatuko dugu.

Kontuan izan behar da datu multzoa sortzeak eta aldatzeak ez duela esan nahi Inkesta izena ipini diogun datu multzoa gordeta dagoenik. Hurrengo lan-saioetan datu multzo honekin berriro lan egin nahi badugu, aurrez adierazi dugun moduan, gordetzea beharrezkoa da.

6.2. DATU MULTZOAK KARGATZEA

6.2.1. Geuk sortutako datu multzoak kargatzea

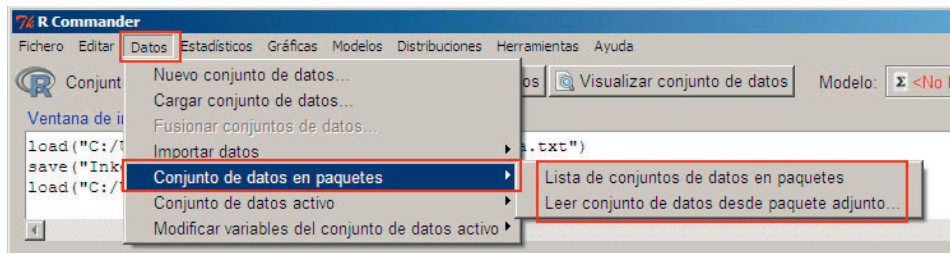
Aurrez datu multzo bat «.Rda» luzapenarekin gorde badugu, aukera honen bidez berreskuratu dezakegu. Aurreko adibidean gordetako datuak berreskuratzeko eman beharreko pausoak hurrengo irudian aurkituko dituzu.



30. irudia.

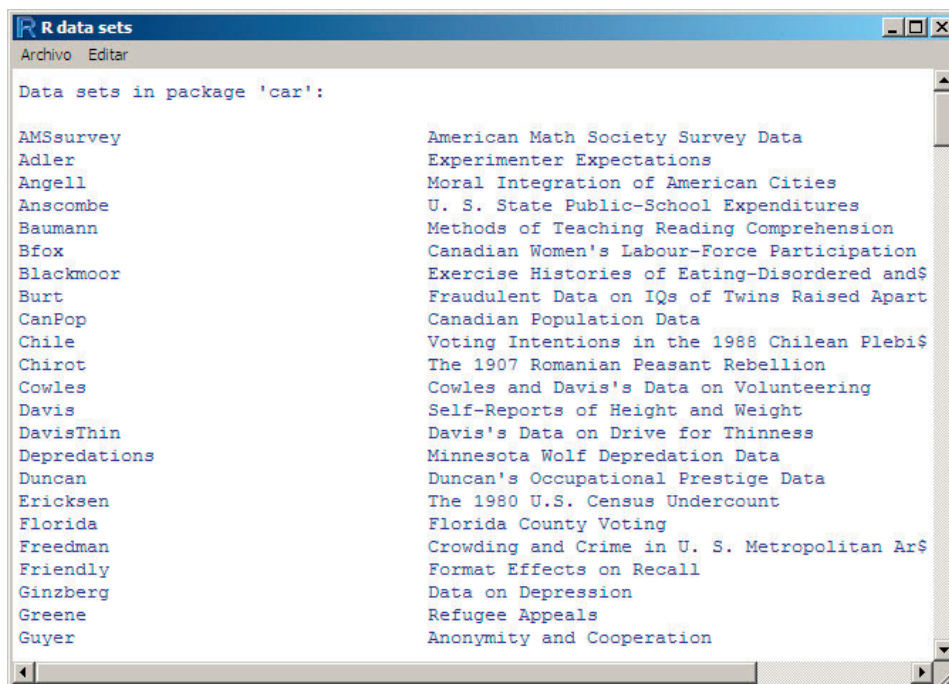
6.2.2. Euren jatorria Estatistikako paketeetan duten datu multzoak irekitzea

Normalean R-ko paketeek adibide gisa erabiltzeko datu multzoak dituzte. Horieta hartzeko R Commander-ek funtzio bereziak eskuragarri ditu hemen: Datos > Conjunto de datos en paquetes.



31. irudia.

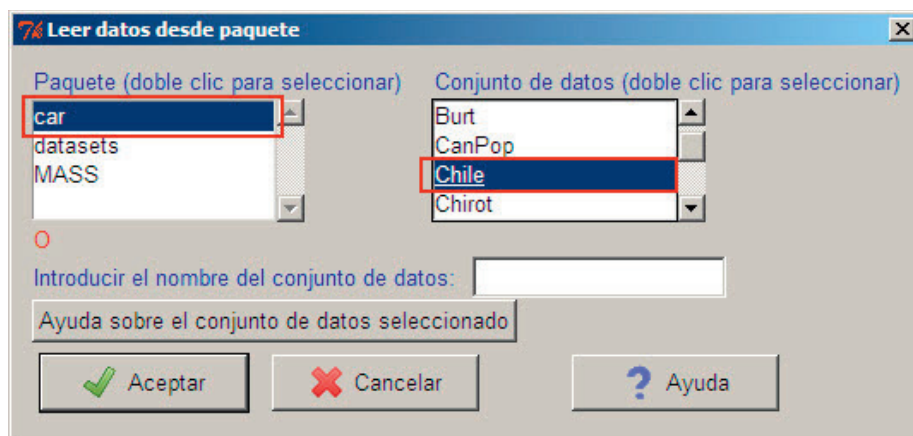
Lehenengo aukerak, Listado de conjuntos de datos en paquetes, gure ordenagailuan instalatuta dauden datu multzo guztiei buruzko informazioa ematen du. Informazioa paketeen arabera sailkatuta dago.



32. irudia.

Aurreko irudian ikus daitekeen moduan, car paketeak hainbat datu multzo ditu eta horiei buruzko informazioa dago.

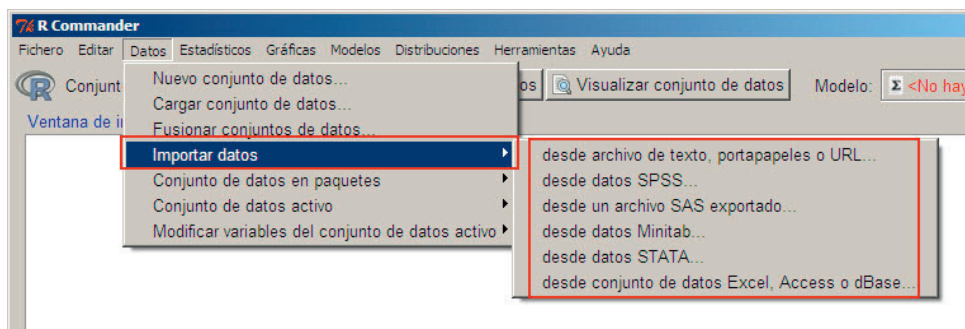
Leer datos desde paquete leihoan edozein paketetako edozein datu multzo hautatu ahal izango dugu.



33. irudia.

6.3. DATUAK INPORTATZEA

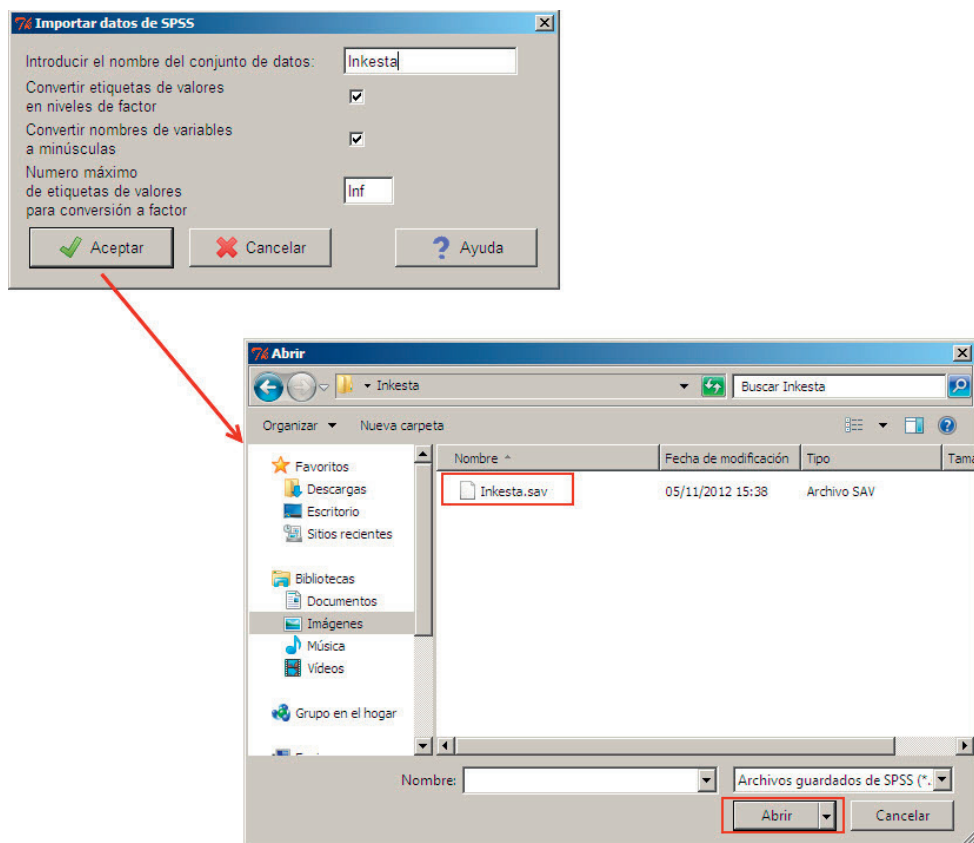
R Commander-ek formatu desberdinetan gordeta dauden datuak irakur ditzake. Baldin eta aztertu nahi ditugun datuak ASCII formatuan badaude, software estatistiko batean sortutako datuak badira (STATA, Minitab, SPSS...) edo Excel edo Acces programetan badute euren jatorria, R Commander-ek Datos > Importar datos aukeraren bidez inportatuko ditu. Aukera honen bidez irekiko zaigun azpimenuan agertzen dira inportatu daitezkeen formatuak.



34. irudia.

6.3.1. SPSSko datu-artxiboak inportatzea

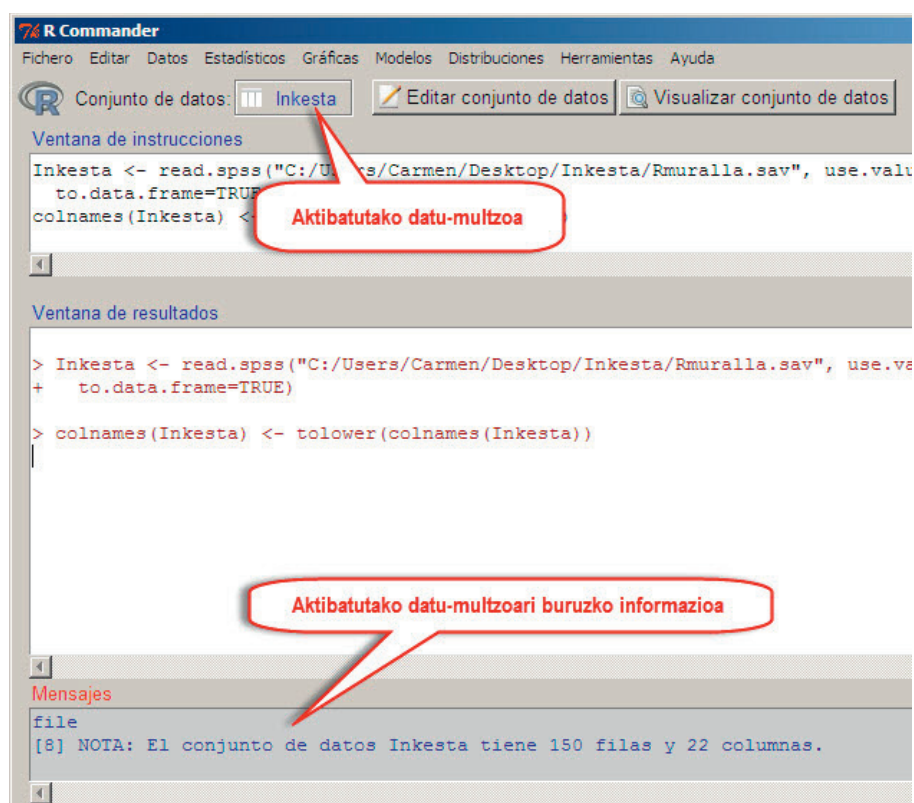
Datos > Importar Datos > desde datos SPSS... bidez irekitzen den leihoan agertuko zaizkigu datu multzoa inportatzeko eskainitako aukerak.



35. irudia.

Hasteko eta behin, datu multzoari izen bat esleituko diogu. Horrek ez du zertan bat etorri inportatu behar dugun datuen artxiboaren izenarekin. Convertir etiquetas de valores en niveles de factor aukeratuko bagenu, balioak esleituta dituzten aldagaiak faktore bailiran inportatuko lirateke. R-k zenbakizko aldagaiak eta faktoreak desberdintzen ditu, objektu desberdinak dira eta tratamendu desberdina jasotzen dute.

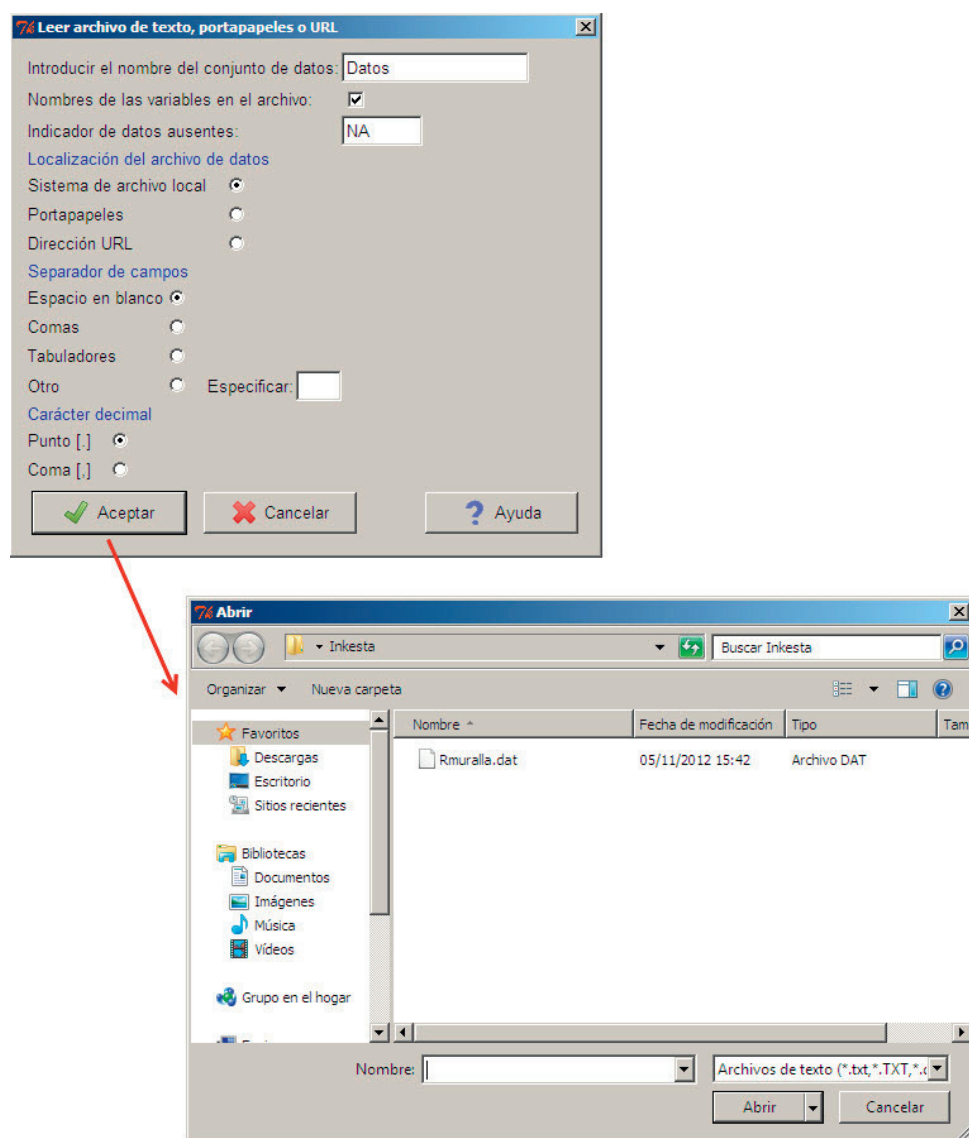
Aceptar sakatu ostean, R Commander-era inportatu nahi dugun artxiboa aukeratu behar dugu. Prozesua burutu denean, mezuen leihoan hau agertuko da: NOTA: El conjunto de datos Inkesta tiene 150 filas y 22 columnas. Informazio hori garrantzitsua da datuak zuzen inportatu ote diren jakiteko. Aldi berean, aktibatutako datu multzoaren leihoan inportatu berri dugun datu multzoaren izena ikusi ahal izango dugu.



36. irudia.

6.3.2 ASCII formatua duten datuak irakurtzea

ASCII formatua (programen artean artxiboak trukatzeko oinarritzko formatua) duen artxibo bat irakurtzeko Datos > Importar Datos > Desde archivo de texto o portapapeles sakatuko dugu. Irekiko den leihoa inportazio-prozesuaren egokitasuna mugatuko duen aukera-sorta aurkeztuko du.



37. irudia.

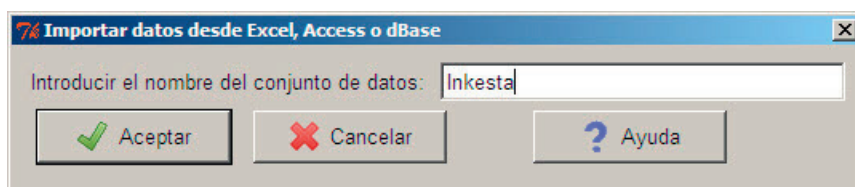
Balio galduen eta dezimalen karaktereei dagozkien aukerak oso garrantzitsuak dira. Balio galduak NA (*Non Avaliable*) karaktereekin identifikatzen dira. Jatorrizko artxiboan balio galduak nola zeuden kodifikatuta adierazten ez bada, NA izan ezik, balio guztiak normaltasunez tratatuko dira. Ondorioz, emaitza okerrak edo arraroak lortuko genituzke. ASCII formatuko artxiboan karaktere dezimala zein den adierazi behar da. Erabili dena koma («,») bada eta ez badugu horrela adierazten, karaktere dezimalak dituzten balio guztiak faktoreak bailiran irakurriko lituzke R-k.

Besterik esan ezean, balio galduak definitzeko R Commander-ek erabiltzen dituen karaktereak NA dira.

Datu-artxibo bat inportatzen denean, oso garrantzitsua da balio galdu guztiak modu berean kodifikatuta egotea.

6.3.3. Excel formatua duten datuak irakurtzea

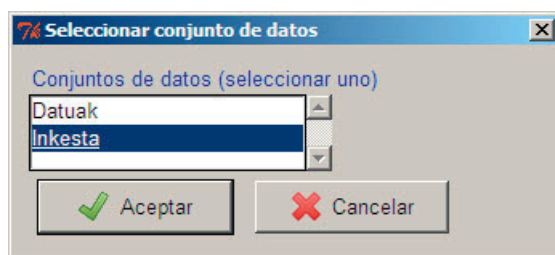
Datos > Importar Datos > Desde conjunto de datos Excel sakatuz, Exceleko kalkulu-orri bateko datuak inporta ditzakegu. Aurreko prozesuen antzera, hemen ere datu multzoari izen bat esleitzeko leiho bat zabalduko zaigu. Jarraian inportatu nahi dugun kalkulu-orria hautatuko dugu. Kalkulu-orrietako datuen irakurketa automatikoa da eta ez da beharrezkoa ezer adieraztea.



38. irudia.

6.4. EGIN AL DEZAKET LANA HAINBAT DATU MULTZOREKIN ALDI BEREAN?

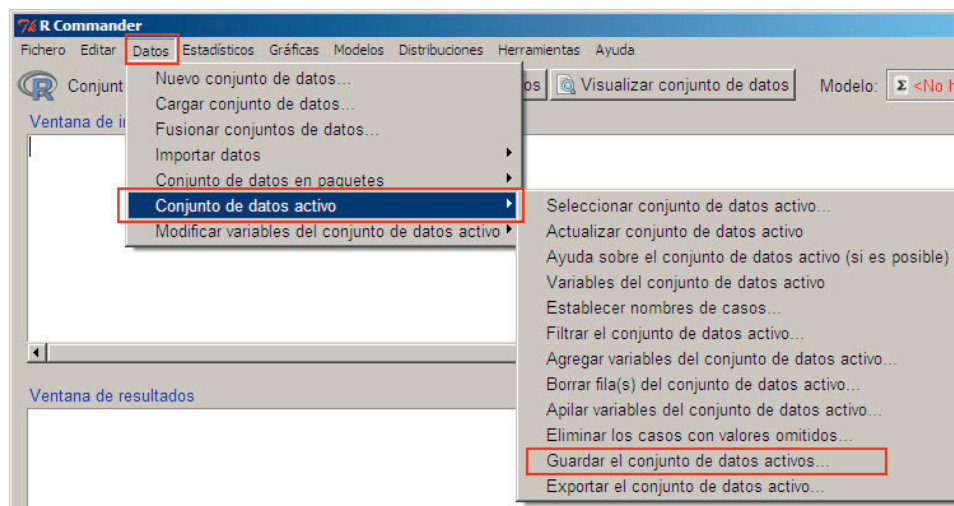
Aldi berean hainbat datu multzo irekita eduki ditzakezu, nahiz eta bakarra egon daitekeen aktibatuta. Aktibo dagoen datu multzoaren izena aktibo dauden elementuen barran adierazten da. Beste datu multzo bat aktibatzeko nahikoa da aktibo dagoen datu multzoaren gainean klikatzea eta interesatzen zaigun datu multzoa aukeratzea.



39. irudia.

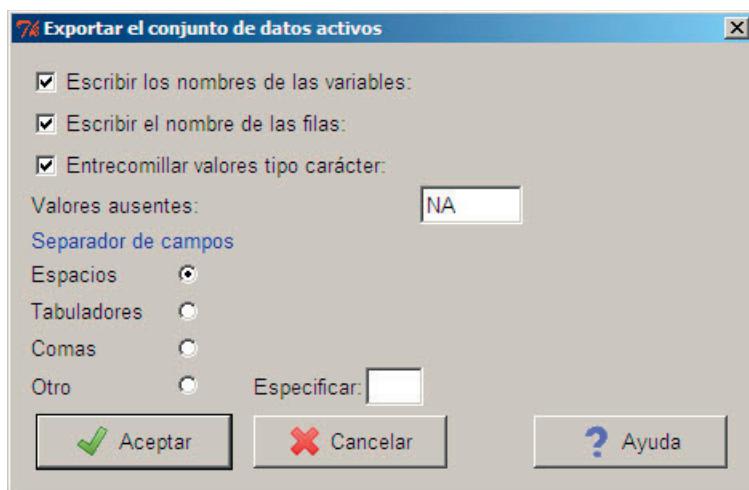
6.5. NOLA GORDE DITZAKET DATUAK?

Datos > Conjunto de datos activo > Guardar conjunto de datos activo bidez «.Rda» (R-ko datuak) luzapenarekin gorde ahal izango dut aktibatuta dagoen datu multzoa. Kontuan izan behar da datu multzoa gordetzean artxiboari edozein izen ematen diogula ere, objektuaren izena mantendu egingo dela eta Conjunto de datos leihoan agertzen dena izango dela.



40. irudia.

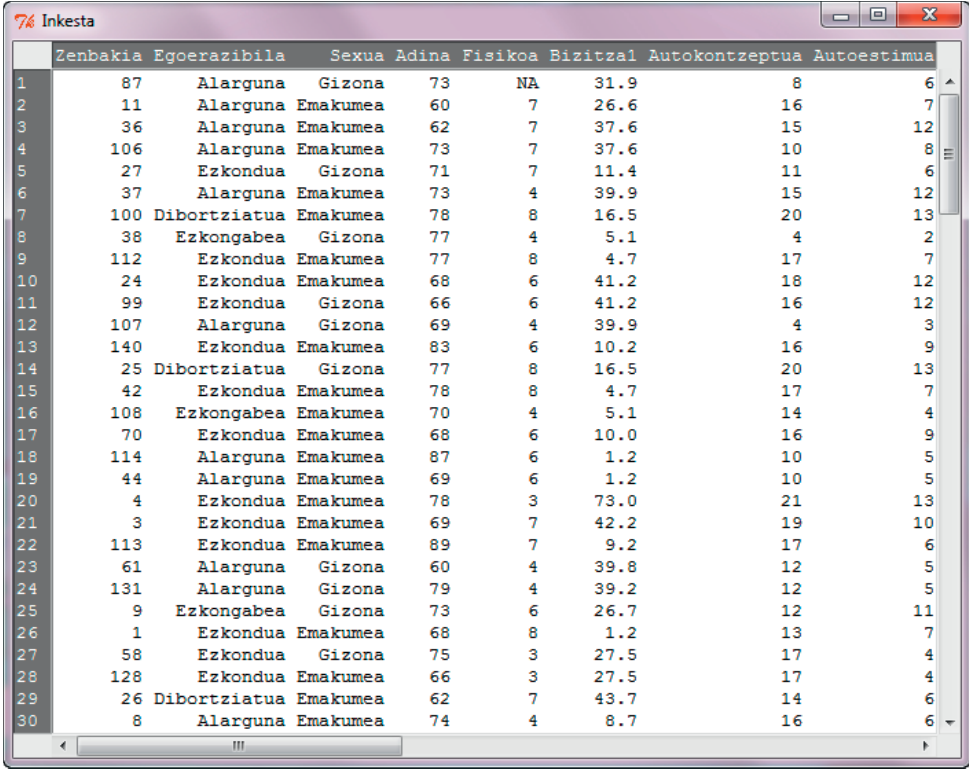
Beste aukera bat datuak testu-artxibo moduan esportatzea da («.txt» eta «.dat» luzapenak). Horretarako Datos > Conjunto de datos activo > Exportar el conjunto de datos activo bidez irekiko den leihoan interesatzen zaizkigun aukerak markatu beharko ditugu.



41. irudia.

6.6. ADIBIDEKO DATUEN AURKEZPENA

Aurreko ataletan Inkesta izeneko datu multzoa inportatu dugu. Objektu horren datuak gidaliburu honetan aurkeztuko ditugun adibide eta ariketen oinarria dira. Visualizar conjunto de datos sakatzen badugu, honek dituen aldagai eta datu guztiak ezagutu ahal izango ditugu.



	Zenbakia	Egoerazibila	Sexua	Adina	Fisikoa	Bizitzal	Autokontzeptua	Autoestimua
1	87	Alarguna	Gizona	73	NA	31.9	8	6
2	11	Alarguna	Emakumea	60	7	26.6	16	7
3	36	Alarguna	Emakumea	62	7	37.6	15	12
4	106	Alarguna	Emakumea	73	7	37.6	10	8
5	27	Ezkondua	Gizona	71	7	11.4	11	6
6	37	Alarguna	Emakumea	73	4	39.9	15	12
7	100	Dibortziatua	Emakumea	78	8	16.5	20	13
8	38	Ezkongabea	Gizona	77	4	5.1	4	2
9	112	Ezkondua	Emakumea	77	8	4.7	17	7
10	24	Ezkondua	Emakumea	68	6	41.2	18	12
11	99	Ezkondua	Gizona	66	6	41.2	16	12
12	107	Alarguna	Gizona	69	4	39.9	4	3
13	140	Ezkondua	Emakumea	83	6	10.2	16	9
14	25	Dibortziatua	Gizona	77	8	16.5	20	13
15	42	Ezkondua	Emakumea	78	8	4.7	17	7
16	108	Ezkongabea	Emakumea	70	4	5.1	14	4
17	70	Ezkondua	Emakumea	68	6	10.0	16	9
18	114	Alarguna	Emakumea	87	6	1.2	10	5
19	44	Alarguna	Emakumea	69	6	1.2	10	5
20	4	Ezkondua	Emakumea	78	3	73.0	21	13
21	3	Ezkondua	Emakumea	69	7	42.2	19	10
22	113	Ezkondua	Emakumea	89	7	9.2	17	6
23	61	Alarguna	Gizona	60	4	39.8	12	5
24	131	Alarguna	Gizona	79	4	39.2	12	5
25	9	Ezkongabea	Gizona	73	6	26.7	12	11
26	1	Ezkondua	Emakumea	68	8	1.2	13	7
27	58	Ezkondua	Gizona	75	3	27.5	17	4
28	128	Ezkondua	Emakumea	66	3	27.5	17	4
29	26	Dibortziatua	Emakumea	62	7	43.7	14	6
30	8	Alarguna	Emakumea	74	4	8.7	16	6

42. irudia.

«RInkesta.Rdata» datu multzoan dauden aldagaiak hirugarren adinean eginiko fikziozko ikerketa batean dute euren jatorria. 150 pertsonen hartu zuten parte ikerketan eta 22 aldagairi dagokien informazioa jaso zen.

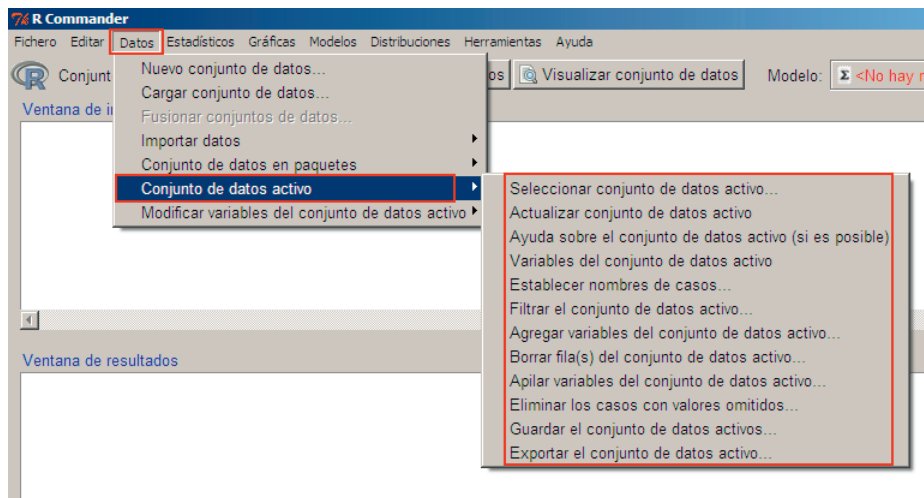
- Zenbakia. Ikerketako parte-hartzailearen identifikazioa.
- Egoerazibila. 4 kategoria dituen (Ezkondua, Alarguna, Ezkongabea eta Dibortziatua) faktore motako aldagaia.
- Sexua. 2 kategoria dituen (Emakumea eta Gizona) faktore motako aldagaia.
- Adina. Zenbakizko aldagaia.
- Fisikoa. Sendotasun fisikoa. Zenbakizko aldagaia.
- Bizitzal. Bizitzarekiko gogobetetze-maila. Zenbakizko aldagaia.
- Sostengua. Hautemandako sostengu sozialaren neurria. Zenbakizko aldagaia.
- Menpekotasuna. Mendekotasun-mailaren neurria. Zenbakizko aldagaia.
- Antsietatea. Antsietate-mailaren neurria. Zenbakizko aldagaia.

- Autokontzeptua. Autokontzeptuaren neurria. Zenbakizko aldagaia.
- Autoestima. Autoestimuaaren neurria. Zenbakizko aldagaia.
- Adimena. Adimen orokorraren neurria. Zenbakizko aldagaia.
- Ekonomia. Maila sozioekonomikoaren neurria. Zenbakizko aldagaia.
- Elikadura. Elikaduraren neurria. Zenbakizko aldagaia.
- Higienea. Higienearen neurria. Zenbakizko aldagaia.
- Tratamendua. 2 kategoria dituen (Klasikoa eta Berritzailea) faktore motako aldagaia.
- Bizitza2. Tratamendu osteko bizitzarekiko gogobetze-maila. Zenbakizko aldagaia.
- item1. Bizitzarekiko gogobetetzeari buruzko test baten lehen galdera. Zenbakizko aldagaia.
- item2. Bizitzarekiko gogobetetzeari buruzko test baten bigarren galdera. Zenbakizko aldagaia.
- item3. Bizitzarekiko gogobetetzeari buruzko test baten hirugarren galdera. Zenbakizko aldagaia.
- item4. Bizitzarekiko gogobetetzeari buruzko test baten laugarren galdera. Zenbakizko aldagaia.
- item5. Bizitzarekiko gogobetetzeari buruzko test baten bosgarren galdera. Zenbakizko aldagaia.

http://www.ehu.es/gip/esperientzia_argitarapenak/ eta http://www.ueu.org/liburuak/r_commander/inkesta.rda web orrietan aurkituko duzu «Inkesta.Rda». Inportatu artxiboa dagozkion komando eta aukerak erabiliz, eta datu multzo aktiboari esleitu iezaiozu Inkesta izena.

7. Datu multzoa manipulatzeari

Datu multzoa aktibatu ostean R Commander-eko menu hedagarriak erabiliz manipulatu daiteke. Alegia, aktibatuta dagoen datu multzoa definitzeko eta kudeatzeko ekintzak burutu daitezke; adibidez: kasuak hautatzea, lerroak ezabatzea, balio galduak aldatzea edo datuak gordetzea. Ekintza horietako edozein gauzatzeko Datos menuko Conjunto de datos activo azpimenua zabaldu behar dugu.



43. irudia.

7.1. DATU MULTZOA HAUTATZEA

Datos > Conjunto de datos activo > Seleccionar conjunto de datos activo... aukeraren bidez irekiko den leihoari esker, posible izango da datu multzo bat aktibatzea. Ekintza hau leiho nagusian Conjunto de datos-en ondoan dagoen botoia sakatzearen parekoa da.

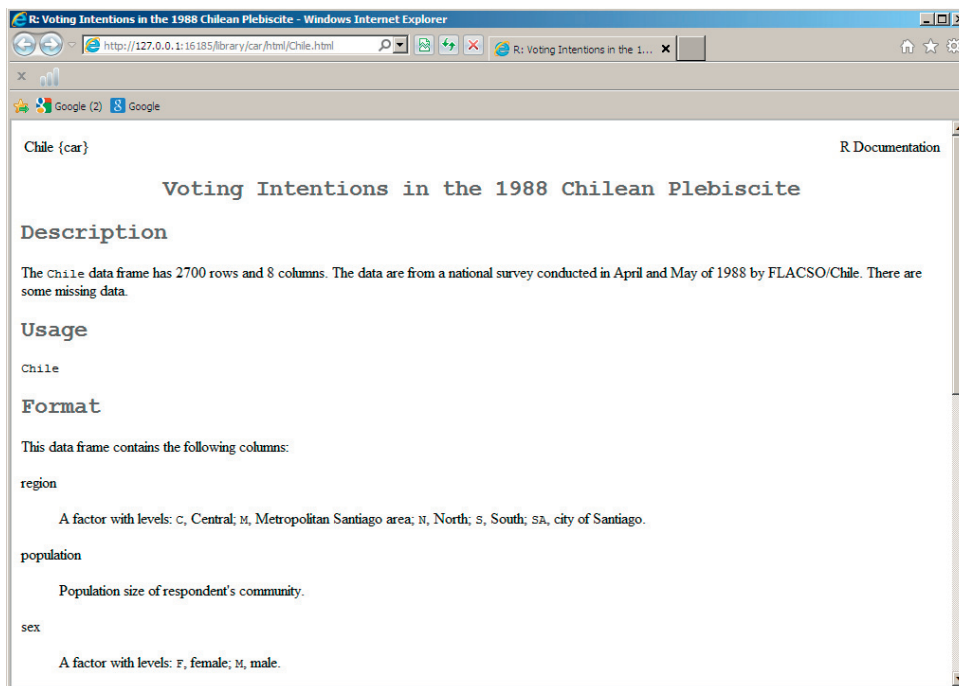
7.2. DATU MULTZOA EGUNERATZEA

Datuekin lan egiteko R Commander edota R-ko kontsola erabil daitezke. Bigarren kasuan, lanean hasi aurretik komenigarria izango litzateke R Commander-en datu multzoa eguneratzea egindako aldaketak finkatu daitezen. Datos > Conjunto de datos activo > Actualizar conjunto de datos activo sakatzuz datu multzoa eguneratuko da.

7.3. DATU MULTZOAREN INGURUKO LAGUNTZA

Datos > Conjunto de datos activo > Ayuda sobre el conjunto de datos activo sakatzean instalatuta ditugun paketeetan dauden datu multzoen inguruko informazioa eskainiko du.

Adibidez, car paketeak, R Commander-ekin batera instalatu zen paketeetako bat, Chile izeneko datu multzoa du. Datuak irakurri ostean (ikusi 6.2.2. atala) Datos > Conjunto de datos activo > Ayuda sobre el conjunto de datos activo (si es posible) aukera exekutatzan badugu, hurrengo informazioa agertuko da:



44. irudia.

7.4. AKTIBO DAGOEN DATU MULTZOKO ALDAGAIAK

Variables del conjunto de datos activo aukerak emaitzen lehoan agertuko den zerrenda bat sortuko du. Bertan datu multzoak barneratzen dituen aldagaien izenak agertuko dira. Modu horretan, Inkesta datu multzoaren aldagaiak zeintzuk diren ikusi ahal izango dugu. [1] zenbakiak "Zenbakia" aldagaiak datu multzoan duen lekua adierazten du. Adina aldagaiak [4] lekua betetzen du. item2 aldagaiak [19] lekua betetzen du.

```
> names(Inkesta)
[1] "Zenbakia"      "Egoerazibila"  "Sexua"         "Adina"         "Fisikoa"
[6] "Bizitza1"      "Sostengua"     "Menpekotasuna" "Antsietatea"   "Autokontzeptua"
[11] "Autoestima"    "Adimena"       "Ekonomia"      "Elikadura"     "Higienea"
[16] "Tratamendua"   "Bizitza2"      "item1"         "item2"         "item3"
[21] "item4"         "item5"
```

45. irudia.

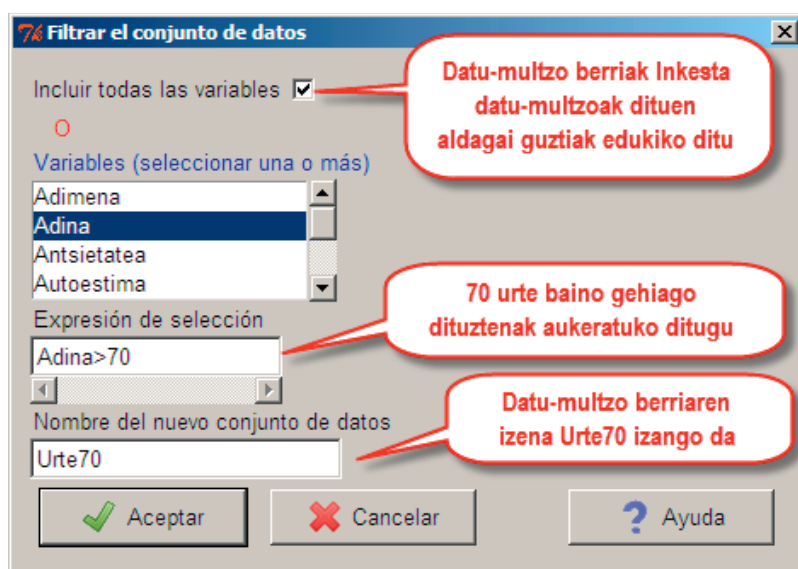
7.5. KASUEI IZENAK ESLEITZEA

Datu multzoko aldagaiei izenak jartzea posible den moduan, lerroei ere izenak jarri diezazkiekegu. Horretarako beharrezkoa da lerroen izenak jasoko dituen aldagai bat (*rowname*) edukitzea.

Inkesta datu multzoko kasuei izenak emateak zentzu handirik ez duen arren, aukera hau interesgarria izan liteke jasotako hainbat emaitzari aplikatzeko. Gure adibidean Zenbakia da parte-hartzaile bakoitza desberdintzeko edo identifikatzeko erabiliko dugun aldagaia.

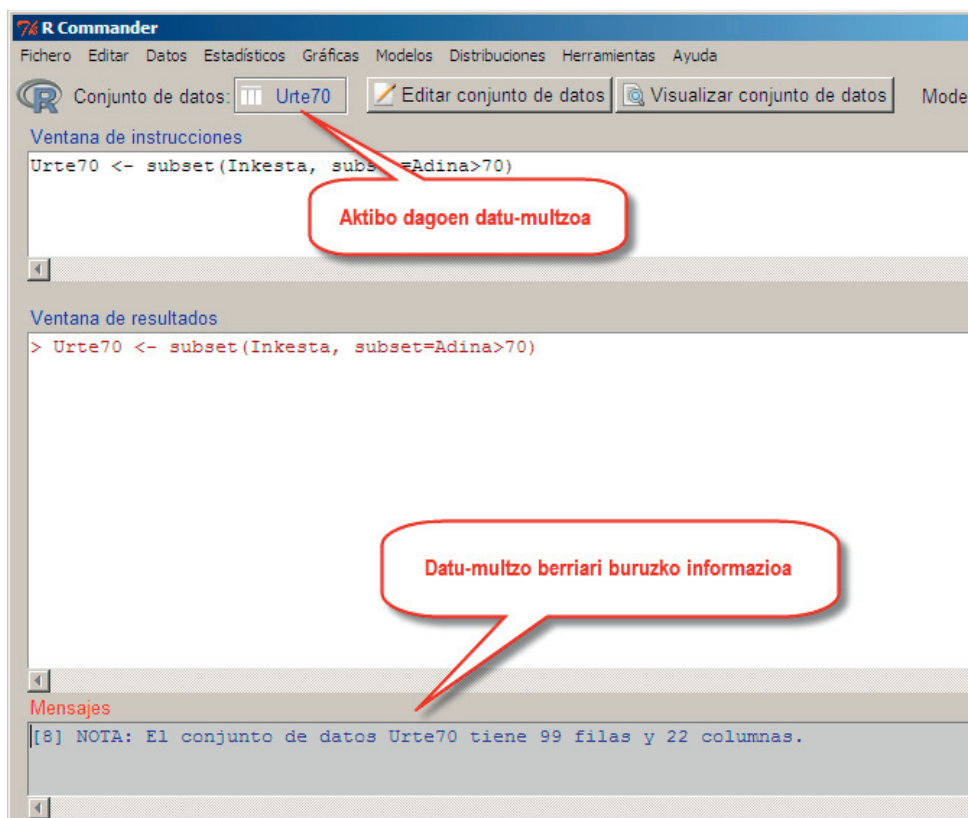
7.6. KASUAK IRAGAZTEA EDO AUKERATZEA

Datos > Conjunto de datos activo > Filtrar el conjunto de datos activo sakatzuz, baldintza konkretu bat (edo batzuk) betetzen duen kasu multzoa aukera dezakegu. Adibide honetan 70 urte baino gehiago dituzten parte-hartzaileak hautatu nahi ditugu. Irekiko den leihoan, datu multzo berria osatuko duten aldagaiak aukeratu ahal izango ditugu.



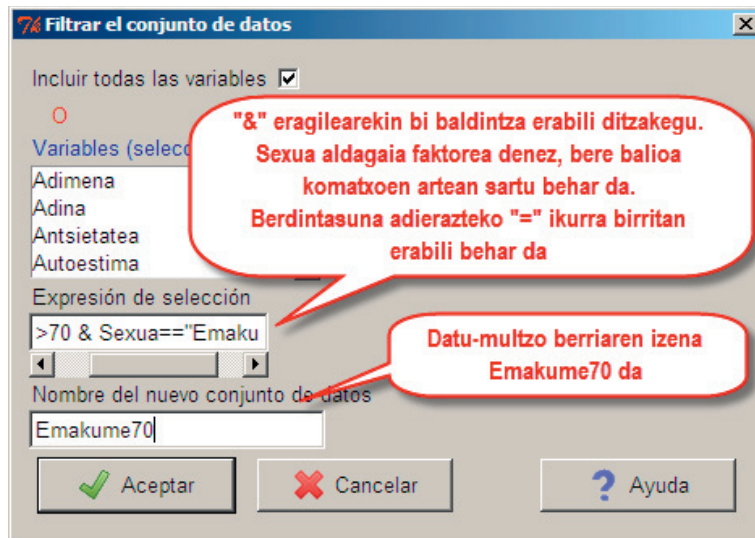
46. irudia.

Expresión de selección azpian dagoen espazio zurian iragazi nahi ditugun datuek bete beharreko baldintzak adierazi behar ditugu. Adibide honetan 70 urte baino gehiago dituzten kasuak hautatu ditugu. Datu multzo berriari Urte70 izena eman diogu eta automatikoki aktibatutako datu multzo bihurtuko da. Mezuen leihoan mezu hau agertuko da: **NOTA:** El conjunto de datos Urte70 tiene 99 filas y 22 columnas.



47. irudia.

Kasuak iragazteko aldi berean baldintza bat baino gehiago erabili daiteke. Adibidez, Inkesta datu multzotik 70 urte baino gehiago dituzten emakumeak iragaziko ditugu eta datu multzo berriari Emakume70 izena esleituko diogu. Horretarako, berriro ere Inkesta datu multzoa aktibatu dugu eta Filtrar el conjunto de datos leihoko Expresión de selección-i dagokion espazioan `Adina>70 & Sexua=="Emakumea"` idatzi dugu. Mezuen leihoan agertzen den moduan, datu multzo berriaren izena Emakume70 da eta 58 lerro (parte-hartzaile) eta 22 zutabe (aldagai) ditu.



48. irudia.

Jarraian agertzen den taulako eragile logikoak erabiliz ezartzen dira elementuak hautatzeko baldintzak.

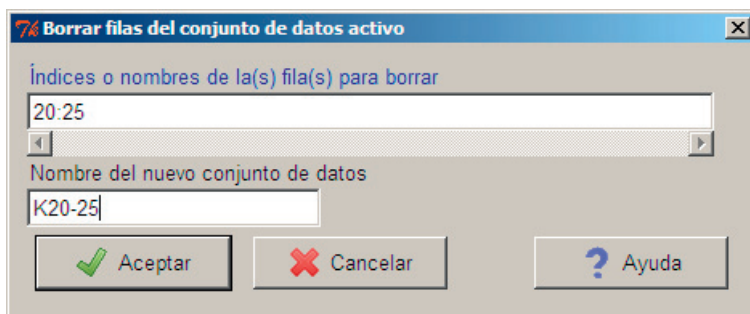
Eragile logikoa		Adibidea
<	Baino txikiagoa	$x < y$; $x < 4$
<=	Baino txikiagoa edo berdina	$x \leq y$; $x \leq 4$
>	Baino handiagoa	$x > y$; $x > 4$
>=	Baino handiagoa edo berdina	$x \geq y$; $x \geq 4$
==	Berdina	$x == y$; $x == 4$
!=	Desberdina	$x != y$; $x != 4$
&	Juntagailua (eta)	$x \& y$
	Disjuntzioa (edo)	$x y$
!	Ezeztapena	!x
xor	Ezeztapen baztertzaila	xor(x,y)

Aldagai bandintzatzailea kategorikoa bada, haren balioa komatxoaren artean idatzi behar da. Adibidez: SEXUA=="GIZONA".

Oso garrantzitsua da ezer iragazi baino lehen, interesatzen zaigun datu multzoa aktibatuta dagoela ziurtatzea.

7.7. LERROAK EZABATZEA

Datos > Conjunto de datos activo > Borrar fila(s) del conjunto de datos activo... bidez, datu multzoko lerroak ezabatu ahal izango ditugu eta datu multzo berriari izen berri bat esleitu. Adibide honetan Inkesta datu multzoko 20, 21, 22, 23, 24 eta 25 (20:25) lerroak ezabatu ditugu eta K20-25 izeneko datu multzoa sortu dugu.



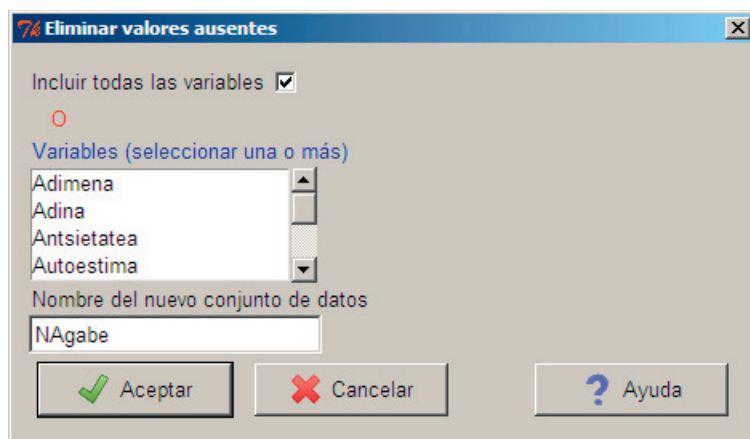
49. irudia.

Ezabatu nahi ditugun lerroak bata bestearen jarraian badaude, ezabatu nahi dugun lehen eta azken lerroak bi puntu bertikalekin («:») banatuta idatzi behar ditugu.

Ekintza burutu ostean K20-25 datu multzoa sortuko da. Horrek 144 lerro eta 22 zutabe izango ditu (ikusitako mezuen lehiogan).

7.8. BALIO GALDUAK DITUZTEN KASUAK EZABATZEA

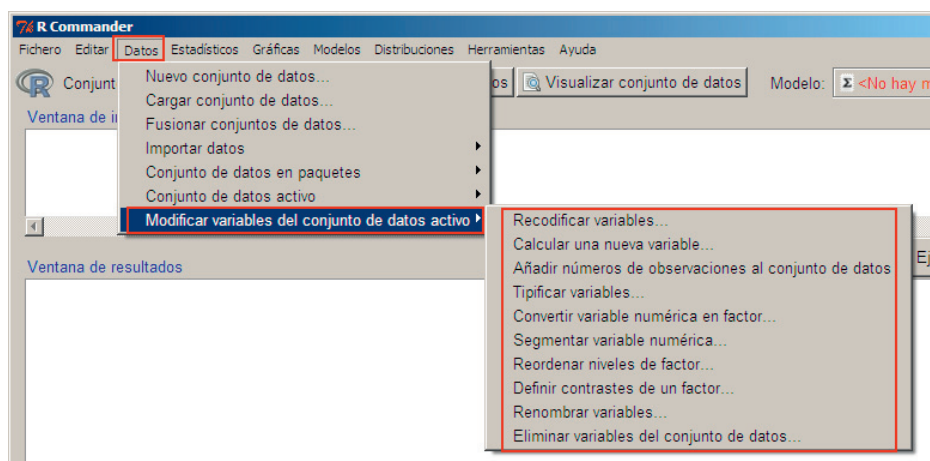
Datu multzoko aldagaiaren batean balio galdurik (NA) duen kasu oro ezabatu nahi badugu, Datos > Conjunto de datos activo > Eliminar los casos con valores omitidos aukera erabili behar dugu. Balio galduak NA (*Non available*) karaktereekin adierazten dira besterik esan ezean. Irekiko den lehiogan datu multzo berriari izen bat eman behar diogu; gure adibidean datu multzo berriaren izena NAgabe izango da.



50. irudia.

8. Aldagaiak sortzea eta aldatzea

Datu multzo aktiboko aldagaiak sortzeko eta aldatzeko aukerak hemen aurkituko ditugu: Datos > Modificar variables del conjunto de datos activo.



51. irudia.

8.1. ALDAGAIK BIRKODETZEA

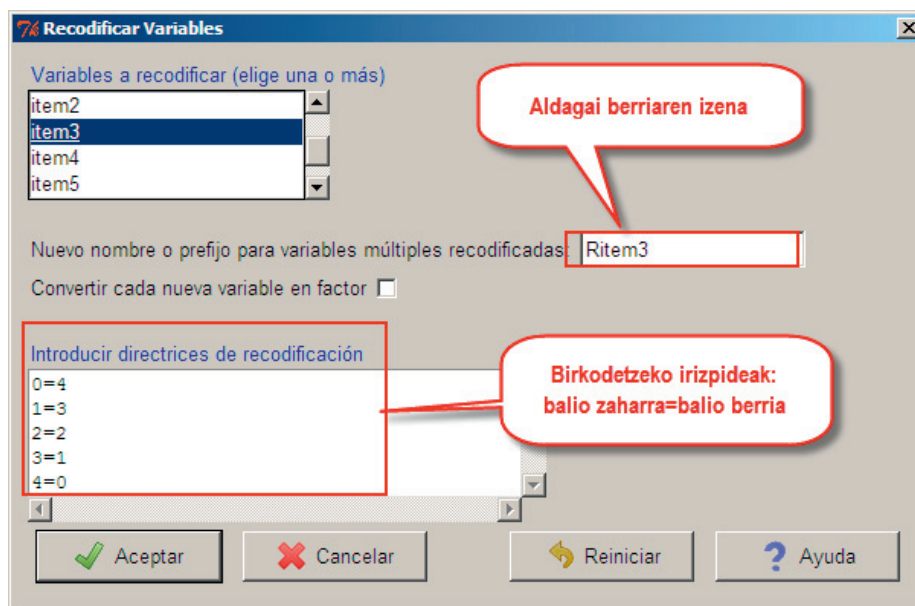
Recodificar variables... bidez datu multzo aktiboko aldagaiei balio berriak esleitu diezazkiekegu, eta horrela aldagai zaharretatik abiatuz aldagai berriak sortu. Adibidez, Inkesta datu multzoan item1, item2, item3, item4 eta item5 aldagaiak bizitzarekiko gogobetetzeari buruzko galde-sorta batetik hartu dira. Galdera (item) hauen edukia hurrengo taulan irakur daiteke:

Erantzun itzazu hurrengo esaldiak, aurkezten zaizkizun erantzunetako bat aukeratuz: Inoiz, Ia inoiz, Batzutan, Ia beti edo Beti.

Inoiz ez	Ia inoiz ez	Batzuetan		Ia beti	Beti	
0	1	2		3	4	
1- Nire bizi-kalitatea ona da		0	1	2	3	4
2- Sostengu soziala dut		0	1	2	3	4
3- Bakarrik sentitzen naiz		0	1	2	3	4
4- Nire osasuna ona da		0	1	2	3	4
5- Jarduera sozialak dauzkat		0	1	2	3	4

Eskala honetan puntuazio handi batek puntuazio txiki batek baino bizitzarekiko gogobetetze-maila handiagoa adierazten du. Bost itemen artean, `item3` aldagaiak (*Bakarrik sentitzen naiz*) parte-hartzaileen bizitzarekiko gogobetetze-maila alderantzizko galdera baten bidez neurtzen du. Alegia, puntuazio txiki batek handi batek baino bizitzarekiko gogobetetze-maila handiagoa adierazten du.

Galdera-sortetan oso arruntak diren alderantzizko itemak aurkitzen ditugunean, beharrezkoa da haien balioak birkodetzea. `Recodificar variables...` sakatuz agertuko den leihoan aldatu nahi dugun aldagaia hautatuko dugu. Jarraian sortu nahi den aldagaiari izen bat ipiniko diogu eta bukatzeko esleipen-arauak idatziko ditugu.



52. irudia.

Adibide honetan aldagai bakarra birkodetu dugu. Baina posible da ekintza bera aldi berean hainbat aldagaitan aplikatzea (pentsa galde-sorta bateko item azpimultzo batean). Kontuan izan aldagai berria faktorea (bariantza-analisiak egiteko erabilgarria) edo zenbakizko aldagaia izatea interesatzen zaigula. Aldagai berria izendatzeko oso erabilgarria da aldatutako aldagaiei aurrizki bat jartzea. Modu horretan, aldagai berriak erraz identifikatuko ditugu.

- Esleipen-arauetan balio zaharra = balio berria adierazi behar da.
- Elkarren segidako balioak birkodetzeko bi aukera daude: komekin tartekatuz balio guztiak banan-banan idaztea eta bi puntu bertikalekin («:») tartekatuz segidako lehen eta azken zenbakiak idaztea.
- Hainbat balio birkodetzeko nahikoa da soilik horietako batzuk adieraztea eta falta diren balio guztiei beste balio bat esleitzeko `else` funtzioa erabiltzea:

```
18:23, 25, 27:30=1
```

```
ELSE = 2
```

- Aldatu nahi diren aldagaiak faktoreak badira, haien balioak komatxoaren artean adierazi behar dira.
- Faktore motako aldagai bat zenbakizko aldagai bihurtzeko, balio berriak komatxo gabe adierazi behar dira.

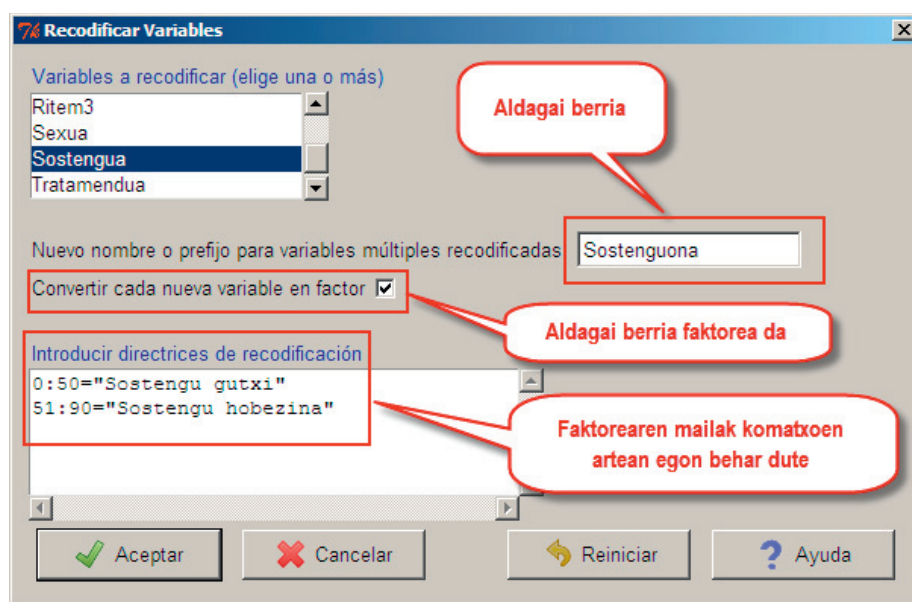


Inkesta

	Bizitza2	item1	item2	item3	item4	item5	Ritem3
1	41	1	0	1	1	1	3
2	51					0	3
3	44					0	3
4	34					1	4
5	59	0	0	1	0	0	3
6	59	0	1	1	0	0	3
7	46	1	1	1	1	1	3
8	56	1	1	1	1	0	3
9	54	1	1	1	0	1	3
10	44	0	0	1	0	0	3
11	23	1	1	0	1	1	4
12	30	1	1	0	1	1	4
13	70	1	0	1	1	1	3
14	51	1	1	0	1	1	4
15	61	1	1	0	1	1	4
16	21	1	1	1	1	1	3
17	52	1	1	1	1	1	3
18	64	1	0	1	1	1	3

53. irudia.

Gure inkestako datuetatik abiatuz hautemandako sostengu sozialaren arabera bi talde sortzeko, Sostengua aldagaia birkodetu genezake. Pentsa dezagun aldagai horretan 50 baino balio handiagoa dutenek sostengu sozial hobeazina dutela. Birkodetzeko baldintzak ondorengoak lirateke:



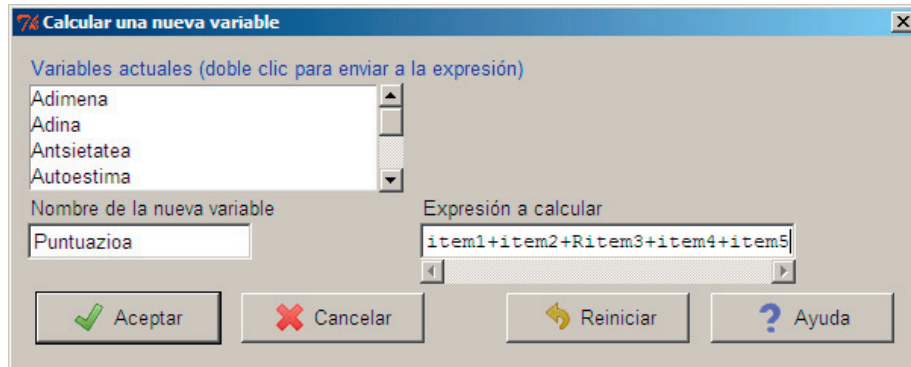
The screenshot shows the 'Inketa' data view in SPSS. The data is organized into columns: 'item2', 'item3', 'item4', 'item5', 'Ritem3', and 'Sostenguona'. The 'Sostenguona' column shows the result of the recoding process, with values 'Sostengu gutxi' and 'Sostengu hobeza' appearing in the list.

	item2	item3	item4	item5	Ritem3	Sostenguona
1	0	1	1	1	3	Sostengu gutxi
2	1	1	0	0	3	Sostengu gutxi
3	1	1	1	0	3	Sostengu gutxi
4	1	0	1	1	4	Sostengu gutxi
5	0	1	0	0	3	Sostengu gutxi
6	1	1	0	0	3	Sostengu gutxi
7	1	1	1	1	3	Sostengu gutxi
8	1	1	1	0	3	Sostengu gutxi
9	1	1	0	1	3	Sostengu hobeza
10	0	1	0	0	3	Sostengu hobeza
11	1	0	1	1	4	Sostengu gutxi
12	1	0	1	1	4	Sostengu gutxi
13	0	1	1	1	3	Sostengu gutxi
14	1	0	1	1	4	Sostengu gutxi
15	1	0	1	1	4	Sostengu gutxi
16	1	1	1	1	3	Sostengu gutxi

54. irudia.

8.2. ALDAGAI BERRIAK SORTZEA

Calcular una nueva variable bidez aldagai berriak sortzeko aukera izango dugu, horretarako jatorrizko aldagaiak eraldatuz edota operazio aljebraikoak gehituz. Aldagai berria datu multzoko zutabe berri batean agertuko da. Hurrengo adibidean, item1, item2, Ritem3, item4 eta item5 aldagaietako balioak batuz aldagai berri bat sortu dugu.



55. irudia.

OHARRA. Funtzio aljebraikoak

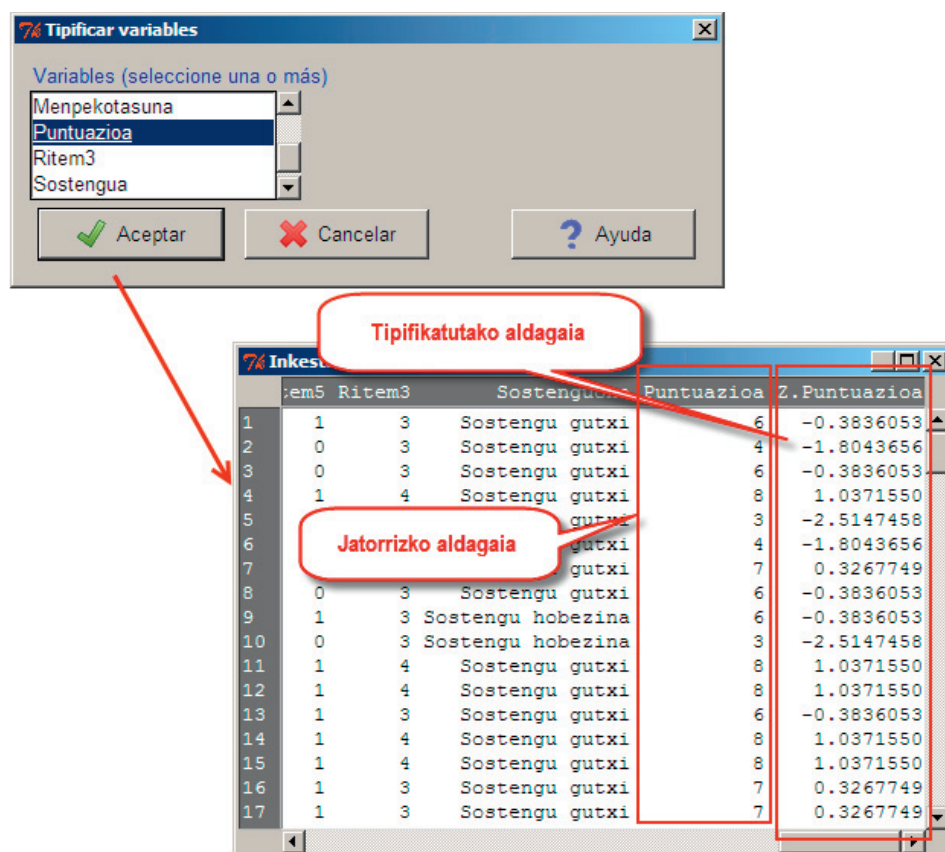
Funtzioa	Eragiketa
+, -, *, /	Batuketa, kenketa, biderketa, zatiketa
abs	Balio absolutua
asin, acos, atan	Funtzio trigonometrikoen alderantzizkoak
exp, log	Esponentzialak eta logaritmo naturalak
round	Biribiltzea
sin, con, tan	Funtzio trigonometrikoak
sqrt(), ^	Erro karratua, berreketa
%/%	Zatiketa osoa
%%	Zatiketaren hondarra

8.3. LERROEI ZENBAKIAK ESLEITZEA

Datos > Modificar variables del conjunto de datos activo > Añadir números de observaciones al conjunto de datos aukerak datu multzo aktiboko behaketa (kasu edo lerro) bakoitzari zenbaki bat esleitu dio. Besterik esan ezean aldagai berriaren izena ObsNumber izango da.

8.4. ALDAGAIK TIPIFIKATZEA

Tipificar Variables-ek datu multzoan aldagai berri bat bailitzan erantsiko den aldagai tipifikatua sortzen du. Aldagai berri horren izena besterik esan ezean jatorrizko aldagaiaren izenari aurretik «Z.» aurrizkia erantsiz sortuko da.

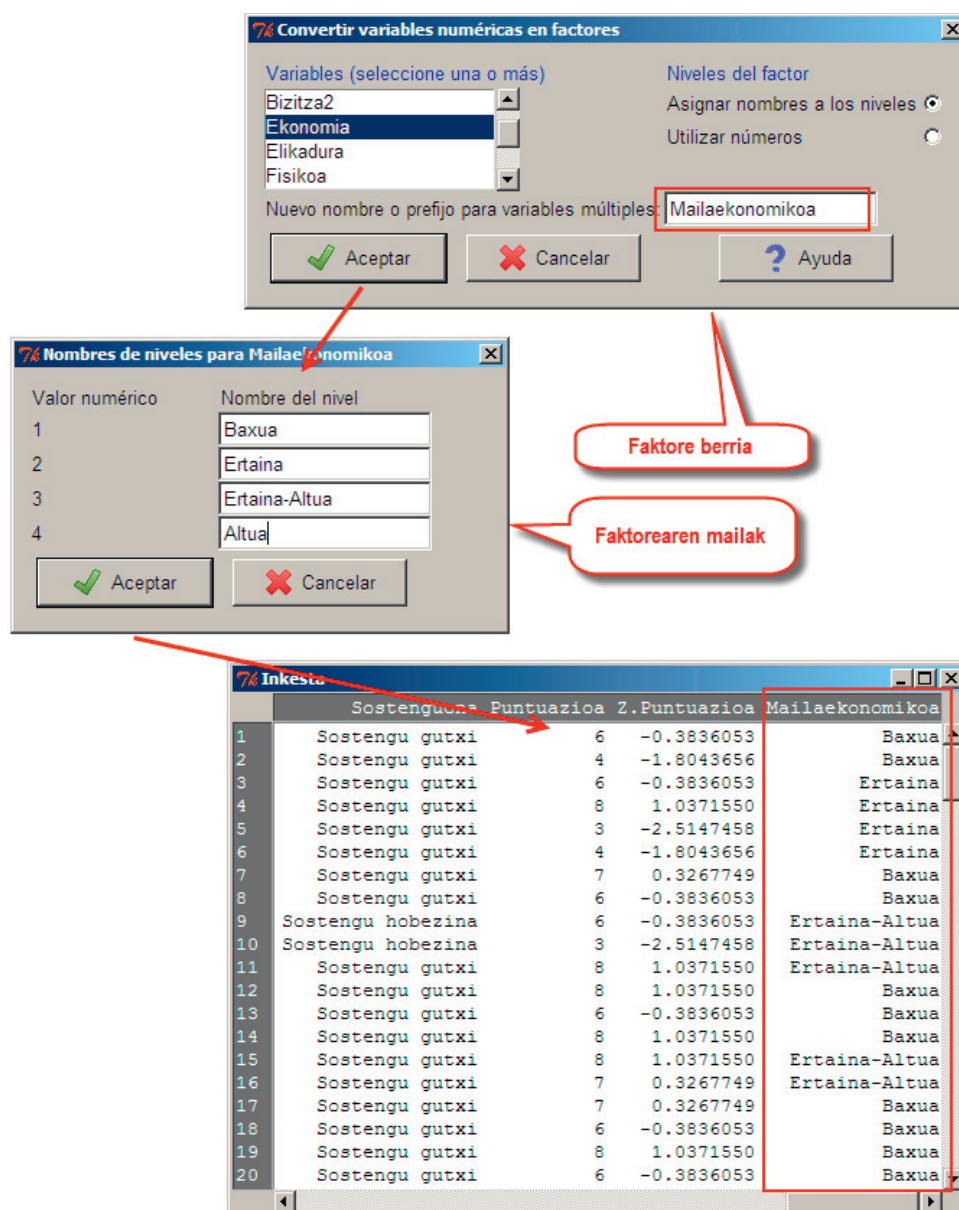


56. irudia.

8.5.ZENBAKIZKO ALDAGAIK FAKTORE BIHURTzea

Faktoreak datuak sailkatzeko eta antolatzeko erabiltzen diren eta maila kopuru finitua duten aldagai kategorikoak dira. Adibidez, Egoerazibila, Sexua, Tratamendua eta Sostengua aldagaiak faktoreak dira; eta haien mailak, hurrenez hurren, hauek dira: Ezkongabea, Alarguna, Ezkondua eta Dibortziatua; Gizona eta Emakumea; Klasikoa eta Berria; Sostengu asko eta Sostengu gutxi. R-k hainbat analisi burutzeko aldagaiak faktore motakoak izatea eskatzen du.

Convertir una variable numérica en factor... sakatuz irekitzen den leihoan aldatu nahi den aldagaia aukeratu behar da eta jarraian sortuko diren kategoria edo maila izen bat esleitu behar zaie. Adibide honetan, Ekonomia (maila sozioekonomikoa) aldagaitik abiatuko gara faktore bat sortzeko. Bere jatorrian zenbakizkoa den aldagaiak 1 eta 4 arteko balioak jasotzen ditu. Aldagai berriaren izena Mailaekonomikoa izango da eta lau maila izango ditu: Baxua, Ertaina, Ertaina-Altua, Altua.

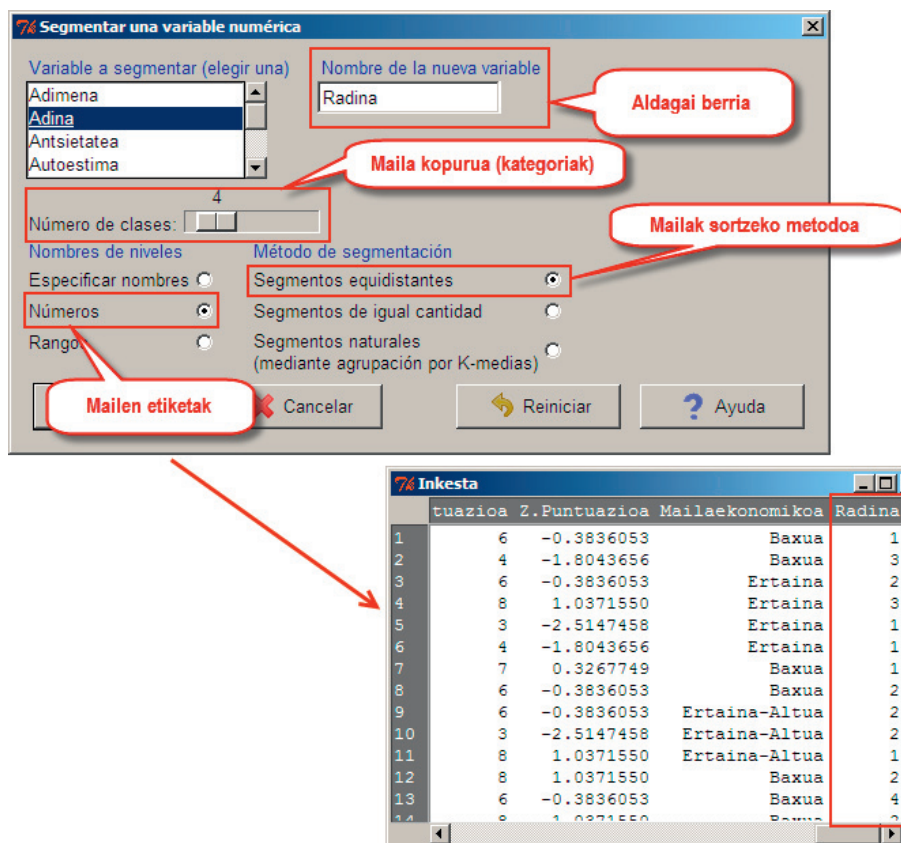


57. irudia.

Leihoan aldatu nahi dugun aldagaia hautatuko dugu, izen berri bat esleituko diogu (zaharra mantendu liteke) eta maila bakoitzari etiketa bat emango diogu. Etiketa moduan zenbaki arruntak ere erabil daitezke. Aldagai berria datu multzoari erantsiko zaio.

8.6. ZENBAKIZKO ALDAGAIK SEGMENTUTAN BANATZEA

Segmentar una variable numérica aukerak zenbakizko aldagaientzako kategoriak sortzen ditu. Aukera hau sakatuz irekiko den leihoan jatorrizko aldagaia hautatuko dugu, aldagai berriaren izena idatziko dugu eta segmentuak sortzeko baldintzak definituko ditugu. Ikertzaileak erabaki behar du zenbat segmentu interesatzen zaizkion. Aldagai berria faktorea izango da. Hurrengo adibidean Adina aldagaia kategorizatu dugu. Aldagai berria Radina izena du, lau maila distantziakide ditu eta bakoitza zenbaki batek identifikatzen du.



58. irudia.

Rcommander-ek aldagaiak segmentutan hiru irizpideren arabera banatu ditzake:

- Segmentos equidistantes: tarte osoa adierazitako segmentu kopuruetan banatzen da, baina segmentu guztien tarteen tamaina mantenduz.

- Segmentos de igual cantidad: segmentu bakoitzak elementu kopuru berdina biltzen du.
- Segmentos naturales (mediante agrupación por K-medias): segmentuak definitzeko k-batezbestekoen algoritmoa erabiltzen da.

8.7. FAKTOREEN ALDAGAIAK BERRANTOLATZEA

Besterik esan ezean, faktoreen mailak ordena alfanumeriko baten arabera biltzen dira. Hortaz, Emakumea kategoria Gizona kategoriaren aurretik egongo litzateke. Modu berean, 1 kategoria 2 kategoriaren aurretik kokatuko litzateke. Faktoreen kategorien ordena garrantzitsua izan liteke. Izan ere, taulak eta grafikoak sortzean etiketak zein ordenatan agertuko diren erabakiko du. Ordena zehatz bat ezartzeko Reordenar niveles de factor sakatuko dugu. Irekiko den leihoan jatorrizko aldagaiaren izena mantendu dezakegu edo aldagai berri bat sortu dezakegu izena aldatuz. Oso garrantzitsua da Factor de tipo ordenado aukera, faktoreak aldagai kategoriko ordenatutzat edo ez-ordenatutzat hartzea ahalbidetzen baitu.

8.8. ALDAGAIEN IZENAK ALDATZEA

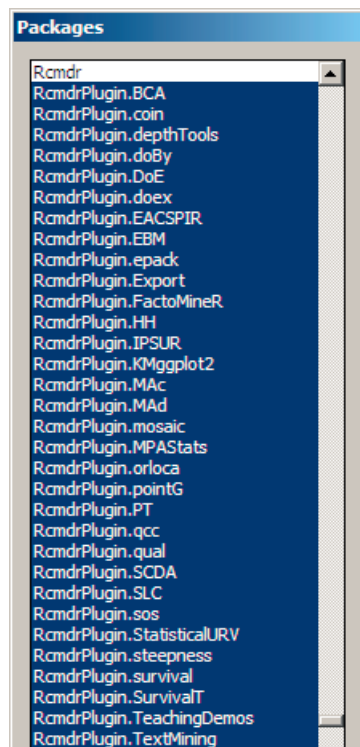
Renombrar variables... aukeraren bidez, datu multzoko aldagaien izenak alda ditzakegu.

8.9. ALDAGAIAK EZABATZEA

Eliminar variables... aukerarekin datu multzoko aldagaiak ezabatu ditzakegu. Horretarako nahikoa da irekiko den leihoan ezabatu nahi ditugun aldagaiak hautatzea.

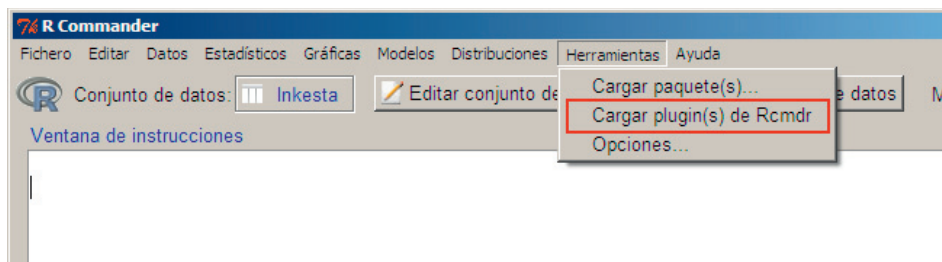
9. R Commander-en luzapenak

R-ko hainbat pakete instalatu ostean, luzapenak (plug-in) R Commander-en menuen sisteman integratzen dira. Gidaliburu hau idatzi zen unean 23 plug-in (RcmdrPlugin) zeuden. Gidaliburu honekiko duten harremana dela-eta, horietako bi deskribatuko ditugu: Export eta TeachingDemos.



59. irudia.

Plug-in luzapenak instalatzeko prozesua beste edozein paketeren instalazioaren modukoa da (ikusi 4.1 atala). Plug-in bat instalatzean Herramientas menuari Cargar plugin(s) de Rcmdr aukera gehitzen zaio automatikoki.

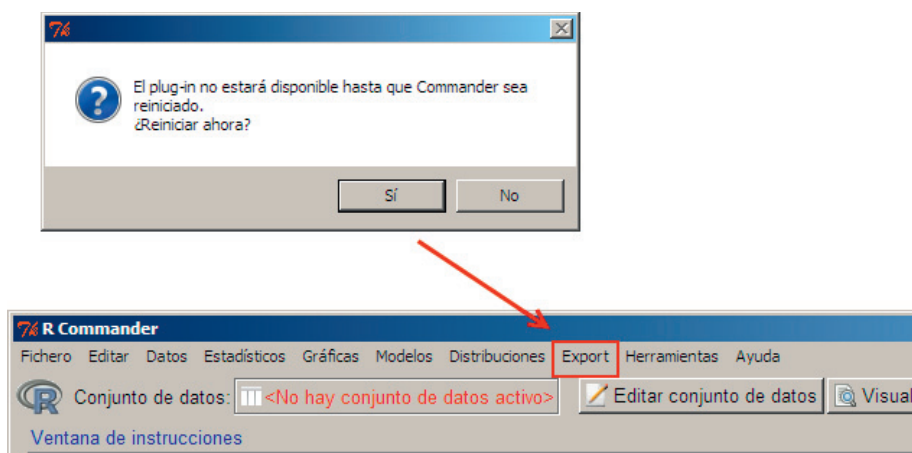


60. irudia.

9.1. RcdcmdrPlugin.Export

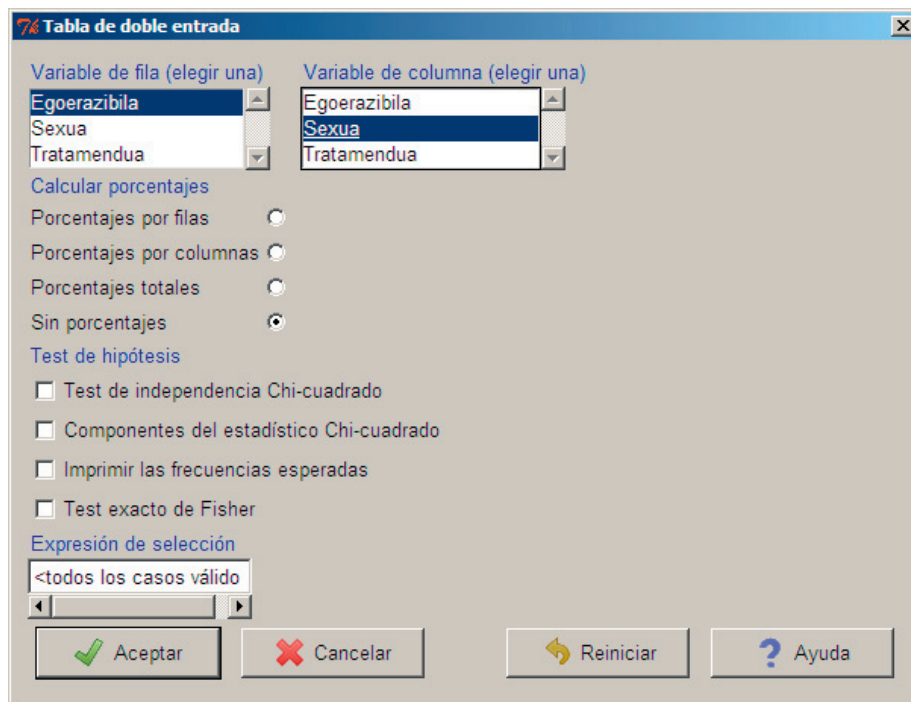
Beste software pribatibo batzuen alboan, R Commander-en alderdi kritikagarrienetako bat bere irteeren zakartasuna da; emaitzen formatuaren irteera estetikoki ez da batere dotorea. Erabiltzaile batzuek oztopotzat har dezaketen ezaugarri hau, RcmdrPlugin.Export-ekin gainditu daiteke. Plug-in honek R Commander-eko irteerak «LaTeX» eta «Html» formatuetara bihurtzen ditu. «LaTeX» eta «Html» formatuetara esportatu daitezke kontingentzia-etaulak, datu-markoak eta ereduak aplikazioetatik eratorritako informazioa.

Paketea instalatu eta gero, kargatzeko Herramientas > Cargar Plugin(s) > RcmdrPlugin.Export sakatu behar da. Aukera sakatzean R Commander-ek berrabiatzeko baimena eskatuko du. Ekintza horrek ez dio eragingo inola ere datu multzoari. Baimena ematean R Commander-eko menu-barran menu berri bat agertuko da, Export.

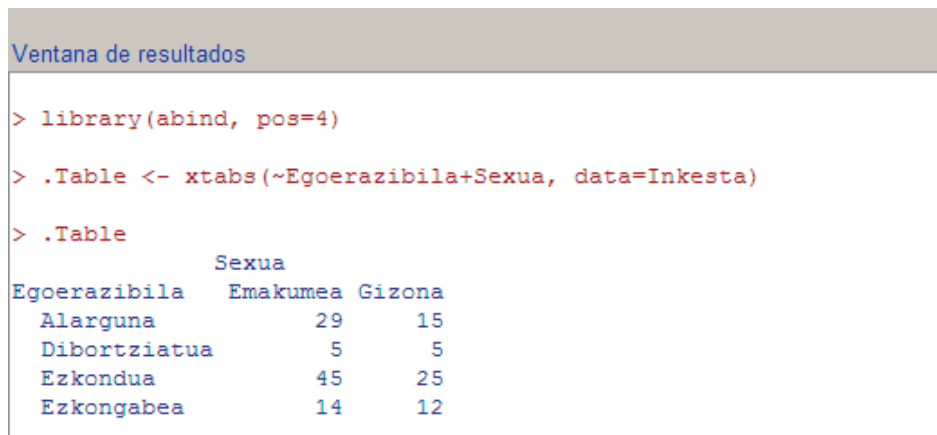


61. irudia.

Hurrengo adibidean, Egoerazibila eta Sexua aldagaien bitartez sortutako kontingentzia-taula nola esportatzen den azalduko dugu. Jarraitu beharreko lehen pausoa kontingentzia-taula sortzea izango da Estadísticos > Tablas de contingencia > Tablas de doble entrada... sakatuz.

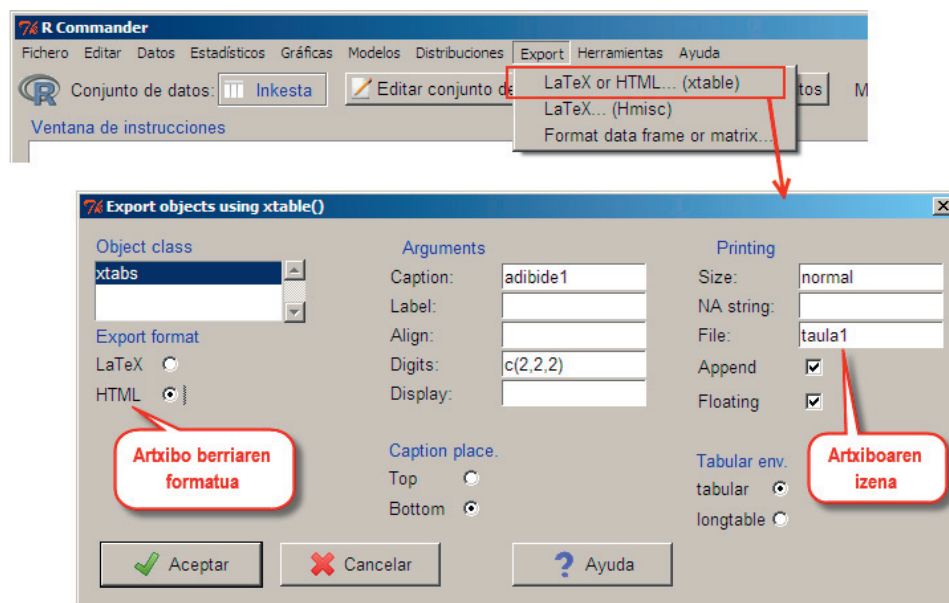


62. irudia.



63. irudia.

Behin taula sortu dela esportatu dezakegu. Export > LaTeX or HTML... (xtable) sakatuz irekiko den leihoan irteerako artxiboaren izena adieraziko dugu. Artxibo hori aktibo dagoen laneko direktorioan gordeko da. Gure kasuan «Html» formatuan sortu den taula `taula1.html` izenarekin gorde da laneko direktorioan.



64. irudia.

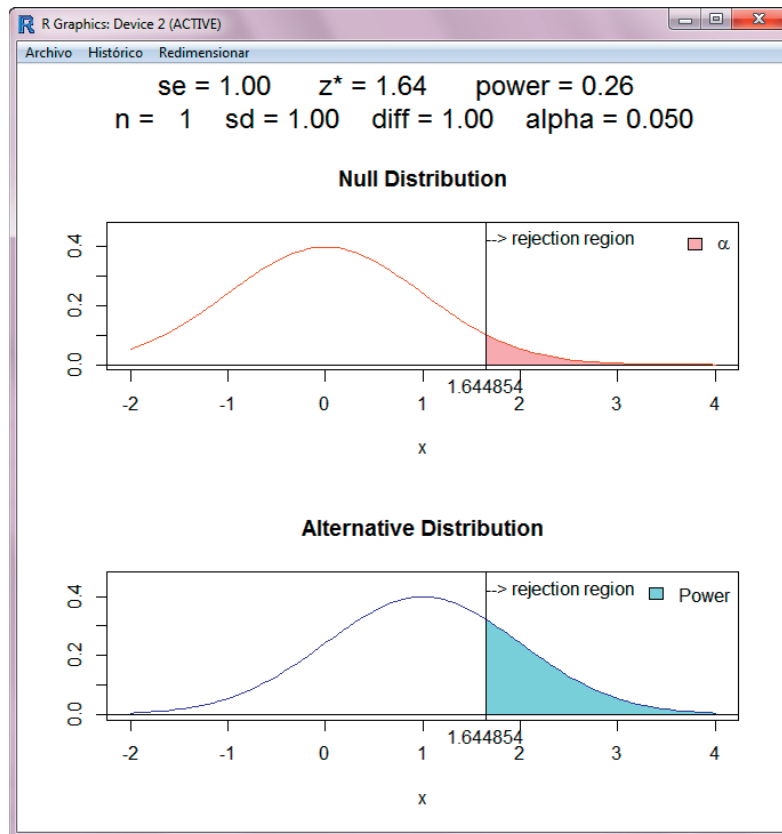
	Emakumea	Gizona
Alarguna	29.00	15.00
Dibortziatua	5.00	5.00
Ezkondua	45.00	25.00
Ezkongabea	14.00	12.00

adibide1

65. irudia.

9.2. RcmdrPlugin.TeachingDemos

R Commander-eko luzapen moduan integratzen den pakete hau J. Fox-ek diseinatu zuen. Irakaskuntzara eta ikaskuntzara zuzentzen diren funtzioak barneratzen ditu. Besteak beste, limite zentralaren teorema, probabilitate-tartearen zenbatespena, potentziaren kalkulua, erregresio bakuna eta korrelazioa kontzeptuak jasotzen ditu (Elosua eta Bully, 2011). Plug-in honen edukiak gidaliburuko azken kapituluaren aurkeztuko dira.



66. irudia.

II. ATALA

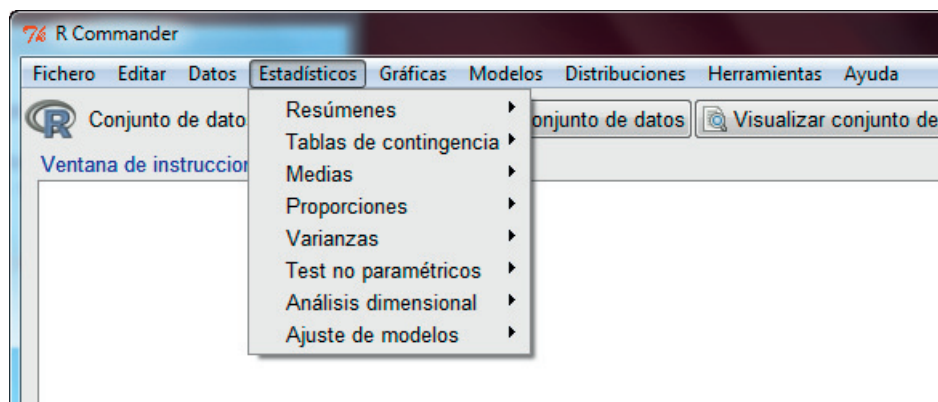
R Commander-ekin datuak aztertzea eta grafikoak sortzea

Bigarren atal hau, R Commander-eko menu-barrako aukeren arabera egituratu dugu. Estadísticos aukeran agertzen diren analisiekin hasiko da, ondoren Modelos-ek eskaintzen dituen aukeren laburpenari ekingo dio, eta grafikoen eraikuntzarekin eta probabilitate-banaketen azterketarekin jarraituko du. Atala amaitzeko, potentzia edo korrelazioa moduko kontzeptuen irakaskuntza- eta ikaskuntza-prozesuetan bereziki erabilgarria den Teaching Demos izeneko erremintaren nondik norakoez arituko gara.

10. Azterketa estatistikoak R Commander-en

Egilea: Josu Mujika Lizaso

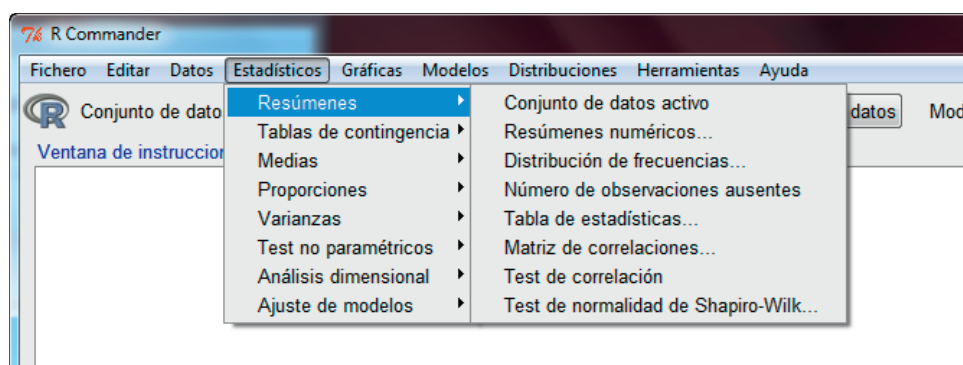
Datu multzoa eraiki ostean, «Inkesta.Rda», datuen azterketari ekingo diogu. R Commander-en azterketa estatistikoen aukerak menu-barrako Estadísticos-en kokatzen dira. R Commander-ek eskaintzen duen ordenari jarraituko diogu aukera bakoitza azaltzeko.



67. irudia.

10.1. LABURPENAK

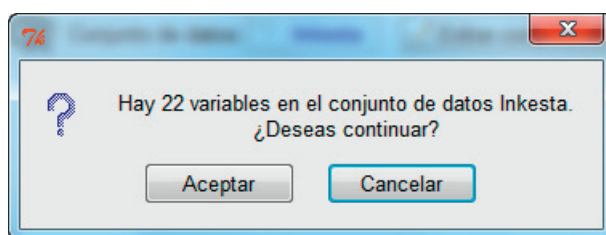
Resúmenes azpimenuak aldagaien laburpen deskribatzaileak, batezbestekoen arteko konparaketak, korrelazio-koefizienteen kalkuluak eta datuek banaketa normalarekiko duten doiketaren azterketa gauzatzeko aukerak biltzen ditu.



68. irudia.

10.1.1. Aktibatutako datu multzoa

Datuen azterketan eman beharreko lehen urratsa datuen portaera modu orokorrean ezagutzea da. Horretarako Estadísticos > Resúmenes > Conjunto de datos activo sakatuko dugu. Aukera horrekin irekiko den leihoak datu multzoko aldagai kopurua adieraziko digu. Hain zuzen ere, R Commander-ek eskainiko dizkigun laburpenen kopurua.



69. irudia.

Emaitzen aurkezpena onartu ostean, aldagai bakoitzaren laburpena eskuratuko dugu.

Ventana de resultados

```
> summary(Inkesta)
```

Zenbakia	Egoerazibila	Sexua	Adina	Fisikoa	Bi:
Min. : 1.00	Alarguna :44	Emakumea:93	Min. :60.00	Min. :0.000	Min.
1st Qu.: 38.25	Dibortziatua:10	Gizona :57	1st Qu.:68.00	1st Qu.:4.500	1st Qu.
Median : 75.50	Ezkondua :70		Median :74.00	Median :6.000	Median
Mean : 75.50	Ezkongabea :26		Mean :73.57	Mean :5.687	Mean
3rd Qu.:112.75			3rd Qu.:78.00	3rd Qu.:7.000	3rd Qu.
Max. :150.00			Max. :98.00	Max. :8.000	Max.

Sostengua	Menpekotasuna	Antsietatea	Autokontzeptua	Autoestima	Adim:
Min. : 5.00	Min. :0.000	Min. : 4.00	Min. : 1.00	Min. : 2.00	Min.
1st Qu.:29.00	1st Qu.:2.000	1st Qu.:55.25	1st Qu.:10.00	1st Qu.: 4.00	1st Qu.
Median :43.50	Median :4.000	Median :72.00	Median :14.00	Median : 7.00	Median
Mean :43.78	Mean :3.826	Mean :65.99	Mean :13.36	Mean : 6.16	Mean
3rd Qu.:56.75	3rd Qu.:5.000	3rd Qu.:81.00	3rd Qu.:17.00	3rd Qu.: 8.00	3rd Qu.
Max. :80.00	Max. :5.000	Max. :99.00	Max. :21.00	Max. :10.00	Max.

Ekondua	Hi	amendua	Bizitza2	i
Min. : 2.000	Min. : 2.000	Min. : 75	Min. : 9.00	Min.
1st Qu.:2.000	1st Qu.:2.000	1st Qu.:75	1st Qu.:39.00	1st Qu.

Faktoreak: aldagai kualitatiboak

Zenbakizko aldagaia

Balio galduak

70. irudia.

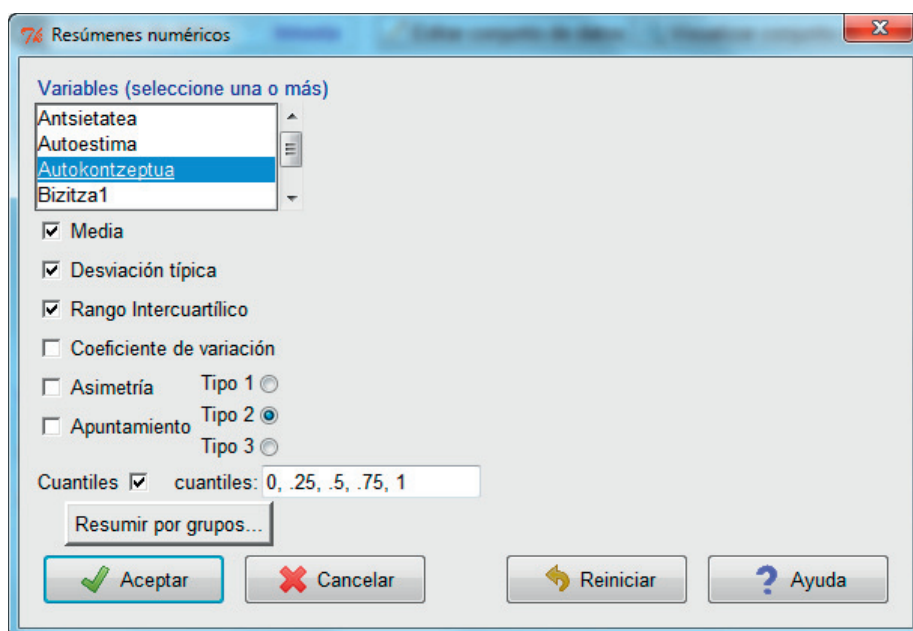
Aipatzekoa da R Commander-ek aldagai kualitatiboei (faktoreak) eta kuantitatiboei tratamendu desberdina ematen diela. Kuantitatiboak diren aldagaiarentzat ematen duen informazioak balio minimoa eta maximoa, lehen, bigarren, hirugarren eta laugarren kuartilak eta banaketaren batezbestekoa jasotzen ditu.

Aldiz, faktoreen inguruan eskaintzen duen informazioa maiztasunen banaketak jasotzen dituen taula batean eskaintzen du. Faktoreak mailatan antolatzeko interfazeak esleitzen dizkiogun izenak edo balioak erabiltzen ditu. Ezkongabea eta Ezkongabea kategoria edo maila desberdintzat hartzen ditu. Hortaz, datuak kodetzeko prozesuan hanka-sartzeak gerta daitezke. Ondorioz, oso garrantzitsua da beste analisiekin hasi aurretik datu multzoko aldagai guztiei buruzko informazioa eskuratzea, eta posible den heinean, gaizki kodetutako daturik ez dagoela ziurtatzea. Soilik, datu multzoa garbi dagoela ziurtatzean ekingo diogu azterketa-prozesuari bere zentsu hertsian.

10.1.2. Zenbakizko laburpenak

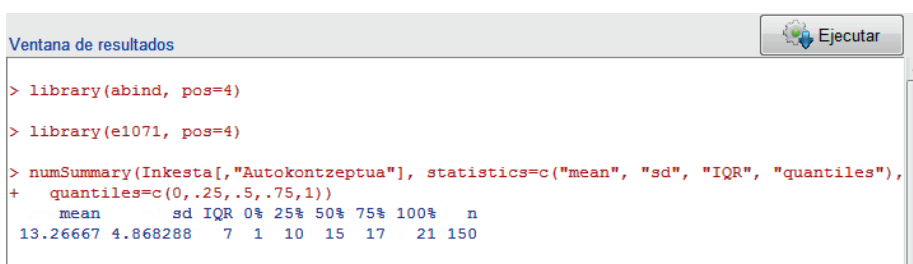
Zenbakizko aldagai batzuen (kuantitatiboak) laburpen zehatzago bat lortzeko Estadísticos > Resúmenes > Resúmenes numéricos... sakatuko dugu.

Hurrengo leihoan Autokontzeptua aldagaiaren laburpena eskatu dugu eta besterik esan ezean aktibatuta dauden aukerak mantendu ditugu: Media, Desviación típica eta Cuartiles. Lagina faktore baten mailen arabera zatitu nahi izango bagenu, Resumir por grupos sakatuko genuke, eta jarraian, interesatzen zaigun faktorea hautatu.



71. irudia.

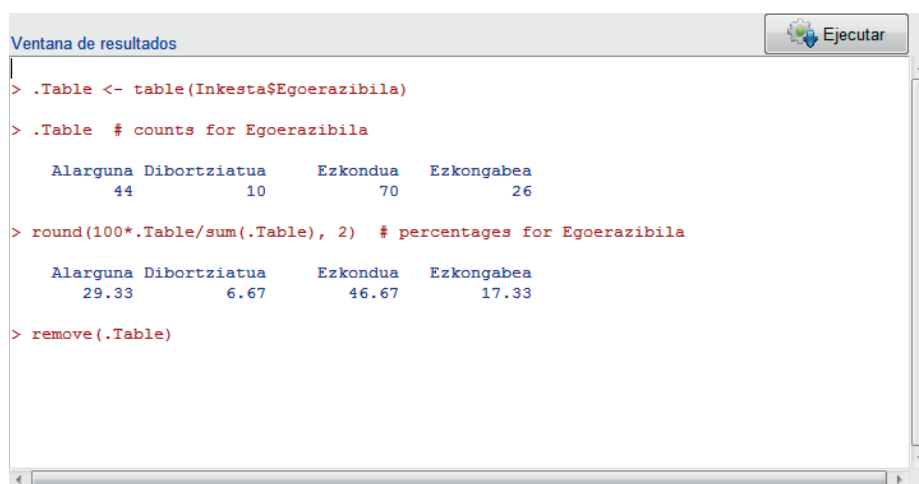
Emaitzen leihoak batezbestekoari, desbideratze tipikoari, kuartilei eta laginaren tamainari dagozkien balioak aurkezten ditu.



72. irudia.

10.1.3. Maiztasunen banaketa

Laburtu eta deskribatu nahi dugun aldagaia kualitatiboa bada, Estadísticos > Resúmenes > Distribución de frecuencias... sakatuko dugu. Hurrengo emaitzek Egoerazibila aldagaiaren kategoria bakoitzeko eskuratutako maiztasunak eta dagozkien probabilitateak aurkezten dituzte.



```

> .Table <- table(Inkesta$Egoerazibila)

> .Table # counts for Egoerazibila

  Alarguna Dibortziatua   Ezkondua   Ezkongabea
      44           10        70        26

> round(100*.Table/sum(.Table), 2) # percentages for Egoerazibila

  Alarguna Dibortziatua   Ezkondua   Ezkongabea
   29.33      6.67      46.67      17.33

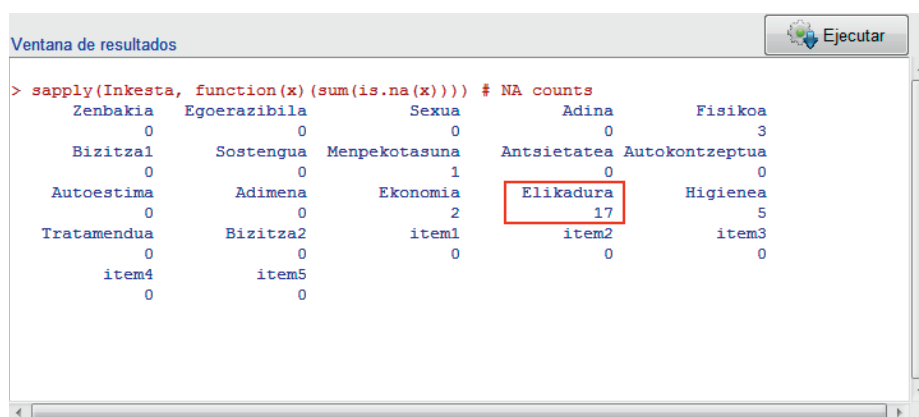
> remove(.Table)

```

73. irudia.

10.1.4. Balio galduak

Nahiz eta aukera hau ez den azterketa estatistiko bat bere adiera hertsian, datu multzoko aldagai bakoitzean zenbat kasutan informaziorik ez dugun jakiteko erabilgarria da. Estadísticos > Resúmenes > Número de observaciones ausentes sakatzen badugu, hau izango da R Commander-ek itzuliko diguna:



```

> sapply(Inkesta, function(x) (sum(is.na(x)))) # NA counts

  Zenbakia   Egoerazibila    Sexua    Adina    Fisikoa
      0           0          0          0          3
  Bizitza1   Sostengua  Menpekotasuna  Antsietatea Autokontzeptua
      0           0            1            0            0
  Autoestima   Adimena    Ekonomia   Elikadura   Higienea
      0           0            2          17            5
  Tratamendua  Bizitza2    item1    item2    item3
      0           0            0            0            0
    item4    item5
      0          0

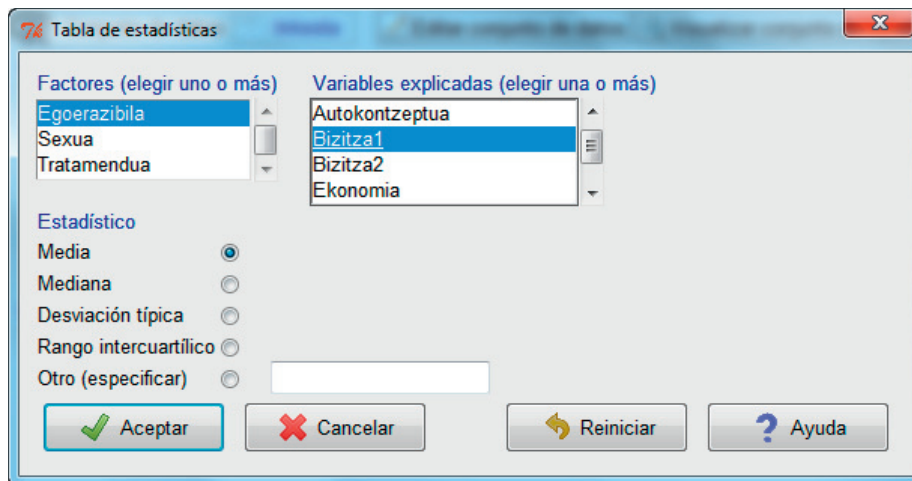
```

74. irudia.

Emaitzetan ikus daitekeen moduan, Elikadura aldagaian 17 kasutan ez daukagu baliorik edo informaziorik. Ekonomia aldagaian bi kasutan...

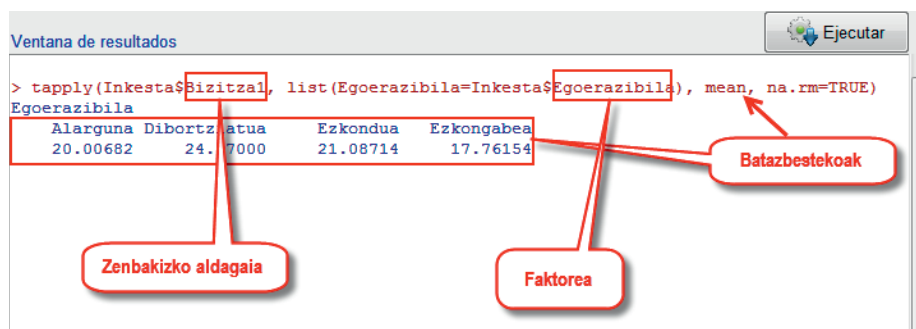
10.1.5. Estatistikoen taulak

Aukera honek faktorea baten edo gehiagoren kategoria bati edo gehiagori dagokion informazioa eskaintzen du. Hurrengo adibidean egoera zibilaren arabera bizitzarekiko gogobetetze-mailaren batezbestekoak eskatu ditugu. Alegia, zein den parte-hartzaileek Egoerazibila faktoreko maila bakoitzeko eskuratu duten batez besteko puntuazioa Bizitza1 aldagaian. Horretarako jarraitu beharreko aukeren sekuentzia hurrengoa da: Estadísticos > Resúmenes > Tabla de estadísticas...



75. irudia.

R Commander-ek batezbestekoaren balioa Egoerazibila aldagaiaren maila bakoitzarentzat aurkeztuko du.



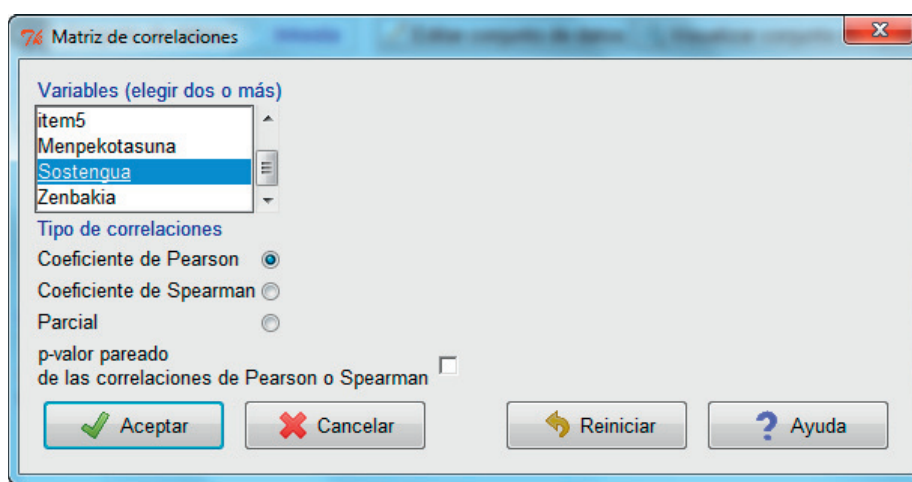
76. irudia.

Datuen arabera, bizitzarekiko gogobetetze-maila handiena dibortzietuek dute (24.07). Aldiz, balio baxuena ezkongabeek aurkeztu dute (17.76).

Modu berean, kategoria bakoitzeko medianaren, desbideratze tipikoaren... balioak eskura ditzakegu. Baina estatistiko bakar bat aldi bakoitzeko.

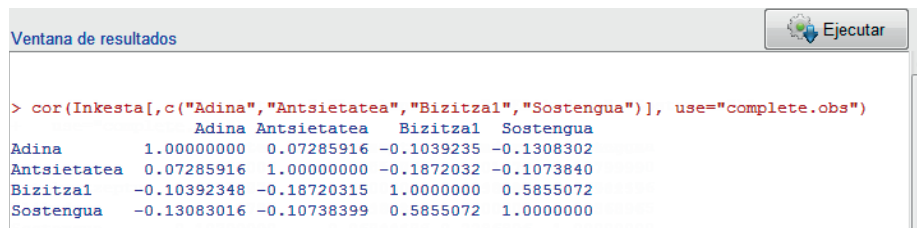
10.1.6. Korrelazio-matrizea

Aldagai kuantitatibo multzo batean binakako korrelazioak kalkulatzeko ditugu `Matriz de correlaciones...` aukerak. Prozesua oso sinplea da; nahikoa da kuantifikatu nahi ditugun erlazioetan parte hartzen duten aldagaiak aukeratzea. Korrelazio-matrizea eskuratzeko `Estadísticos > Resúmenes > Matriz de correlaciones...` sakatuko dugu.



77. irudia.

OHARRA. Hautatu nahi ditugun aldagaiak elkarren segidan badaude, lehenengoa markatu dezakegu eta `Shift` (ezkerrean `Block` `Mayús` teklaren azpian dago eta eskuinean `Intro` teklaren azpian) tekla sakatuta daukagun bitartean azken aldagaiaren gainean klikatu. Aldiz, hautatu nahi ditugun aldagaiak ez badaude elkarren segidan, aldagaiak banan-banan markatuko ditugu `Ctrl` tekla zapalduta mantenduz.



```
> cor(Inkesta[,c("Adina", "Antsietatea", "Bizitzal", "Sostengua")], use="complete.obs")
```

	Adina	Antsietatea	Bizitzal	Sostengua
Adina	1.00000000	0.07285916	-0.1039235	-0.1308302
Antsietatea	0.07285916	1.00000000	-0.1872032	-0.1073840
Bizitzal	-0.10392348	-0.18720315	1.00000000	0.5855072
Sostengua	-0.13083016	-0.10738399	0.5855072	1.00000000

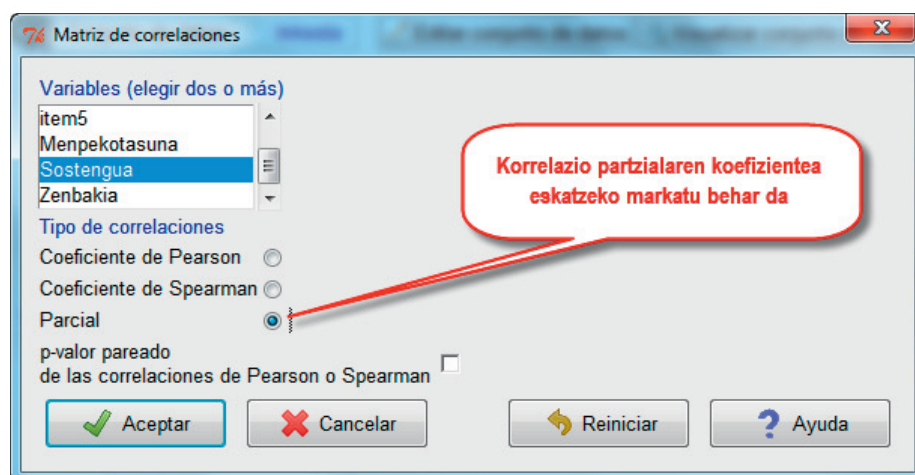
78. irudia.

Korrelazio-koefizientearen balioa interpretatzeko bi dira kontuan izan beharreko alderdiak, zeinua eta koefizientearen balioa. Zeinu positibo batek aldagaien arteko erlazio zuzena adierazten duen bitartean, zeinu negatibo batek aldagaien arteko erlazioa alderantzizkoa dela esan nahi du. Eskuratu ditugun emaitzetan Sostengua eta Bizitzal arteko erlazioa zuzena dela ikus daiteke. Alegia, sostengu soziala handiagoa den heinean bizitzarekiko gogobetetze-maila ere handiagoa da. Ostean, adinaren eta bizitzarekiko gogobetetze-mailaren arteko erlazioa negatiboa da; bata handitzen den heinean, bestea txikitzen doa. Koefizientearen balioa 0tik zenbat eta urrunago egon (-1 edo +1 balioetatik zenbat eta gertuago), are eta handiagoa izango da erlazioaren intentsitatea. Aztertutako lau aldagaien artean hautemandako sostengu sozialaren eta bizitzarekiko gogobetetze-mailaren arteko erlazioak du indar handiena ($r = 0.5855072$).

10.1.6.a. Korrelazio partzialaren koefizienteak

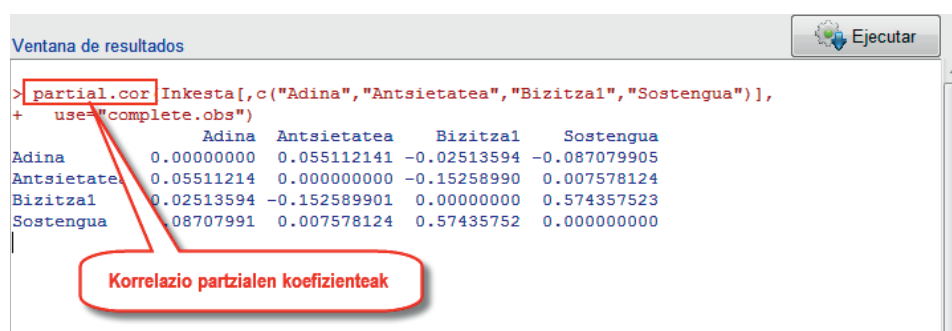
Kasu batzuetan, aldagaien arteko erlazioak beste aldagai batzuen eraginpean egon daitezke. Korrelazio partzialaren koefizienteak bi aldagaien arteko erlazioa kuantifikatzen du beste aldagai baten edo batzuen eragina kontrolatuz. Hurrengo taulan Sostengua, Antsietatea, Adina eta Bizitzal aldagaien arteko korrelazio-koefizienteen taula irudikatu dugu.

Korrelazio partzialaren koefizienteak kalkulatzeko, Matriz de correlaciones leihoan Parcial aukera markatu behar da.



79. irudia.

R Commander-ek aurkezten dituen emaitzetan korrelazio-koefizienteek pairatu dituzten aldaketak antzeman daitezke. Antsietatea eta Bizitzal aldagaien arteko erlazioak intentsitate baxuagoa du Adina eta Sostengua aldagaiek honetan duten eragina kontrolatzen denean.

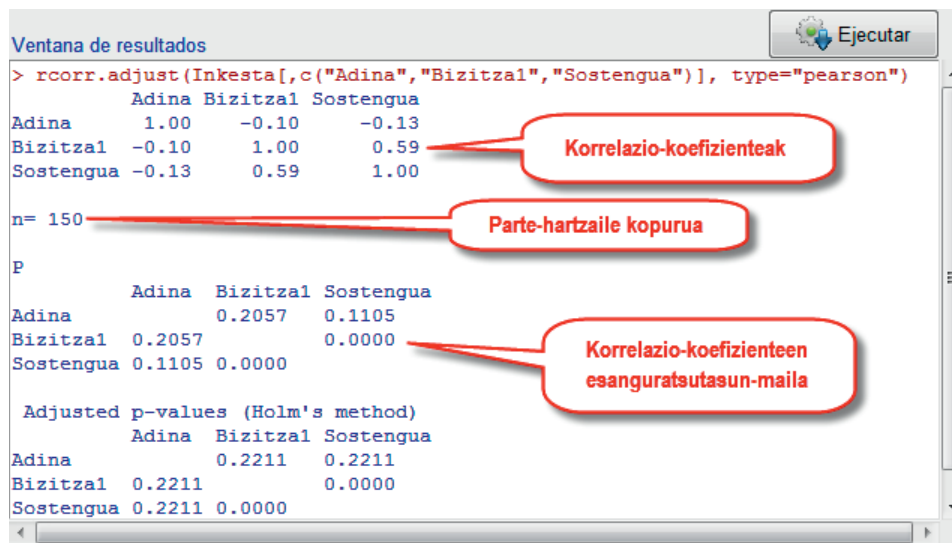


80. irudia.

Bi aldagai ordinalen arteko erlazioa kuantifikatzeko Spearmanen korrelazio-koefizientea kalkulatzeko aukera ematen du R Commander-ek. Kalkulatzeko prozesuak aldagai bakoitzeko balioak txikienetik handienera ordenatzen ditu eta dagokien ordenaren zenbakizko balioagatik ordezkatu. Datuekin «jolastea» eta koefiziente desberdinak erabilia lortutako emaitzen arteko desberdintasunak aztertzea gomendatzen dizuegu.

Matriz de correlaciones leihoko azken aukerari esker korrelazio-koefizienteak zoriz eskuratuak izateko probabilitatea kalkulatu daiteke. Hots, emaitzen esanguratsutasuna zenbatetsi daiteke. Horretarako p-valor pareado

de las correlaciones de Pearson o Spearman aukera aktibatuko dugu. Emaitzetan ikus daitekeen moduan, korrelazio-koefizienteek gain, taula batean horiei dagozkien probabilitateak aurkezten dira.



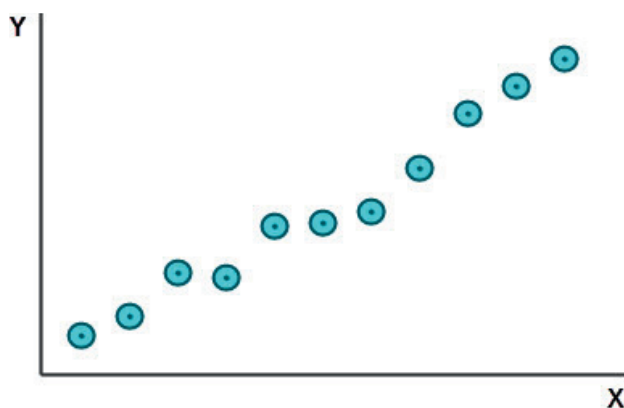
81. irudia.

Aurreko irudia behatuz, 150 parte-hartzailek osatzen duten laginean ($n = 150$) Sostengua eta Bizitza1 aldagaien artean 0.59 balioa duen korrelazio-koefizientea eskuratzeko probabilitatea, bi aldagaiak independentetzat joz gero, 0.0000 dela ikus daiteke. Ondorioz, bi aldagaien arteko erlazioa esanguratsua dela (aldagaiak ez direla independenteak) eta errore-maila 0.0000 balioaren azpitik dagoela esan genezake.

Adina eta Bizitza1 aldagaien arteko erlazioan arreta jartzen badugu, korrelazio-koefizientearen balioa -0.10 dela eta bi aldagaiak independenteak balira balio hori zoriz eskuratzeko probabilitatea 0.2057 izango litzatekeela ikusiko dugu. Ikertzen ari garen populazioan bi aldagai hauen artean erlazorik ez balego ere, 150 pertsonako lagin batean emaitza hau 100 saiakeratan 20,57 aldiz lor genezakeela adierazten du emaitza honek.

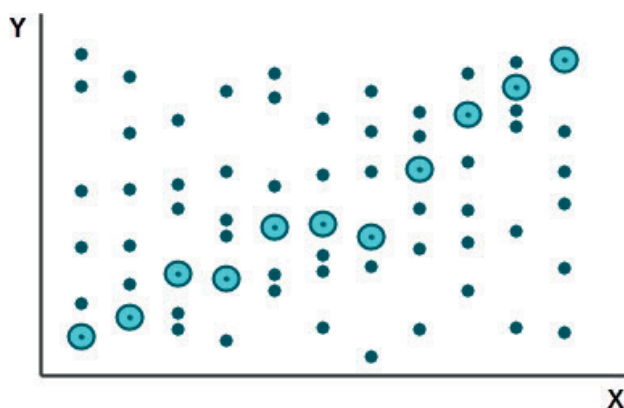
10.1.7. Esanguratsutasunaren arazoa

Esanguratsutasunaren analisiak emaitza estatistikoen interpretazioan duen garrantzia dela-eta, alderdi honi buruz hitz egiteko geldituko gara. Imajina dezagun 11 parte-hartzailek hurrengo irudiko puntuazioak eskuratu dituztela. Grafikoa ikusita aldagaien arteko erlazioa positiboa dela esan genezake.



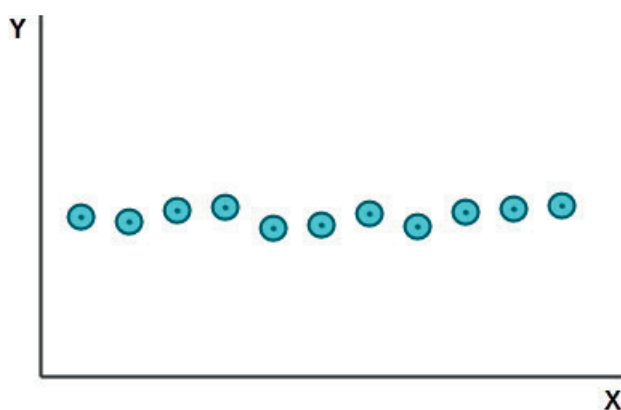
1. grafikoa. $n = 11$ parte-hartzailek osatutako laginak lortutako emaitzak.

Populazioaren errealitatea, ordea, hurrengo grafikoan irudikatzen dena izan liteke. Populazioko parte-hartzaile guztiak irudikatu dira eta argi dago aldagaien arteko ERLAZIOA NULUA dela. Horri 1. MOTAKO ERROREA deitzen zaio. Erlazio positibo hori zoriaren ondorioa izan da. Azken matrizean eskuratu ditugun p (esanguratsutasun) balioak egoera hau errepikatzeko probabilitatea zenbatekoa den adierazten du.



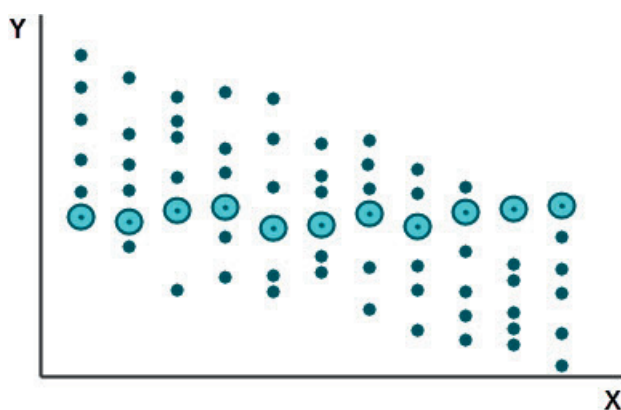
2. grafikoa. $n = 11$ laginaren jatorrizko populazioaren datuak.

Baina kontrako egoera ematea ere posible da. Laginaren emaitzak hurrengoak izan litezke.



3. grafikoa. $n = 11$ parte-hartzailek osatutako laginak lortutako emaitzak.

Aurreko grafikoa behatuz, emaitzek aldagaien arteko erlazioa nulua dela pentsatzera garamatzate. Baina gure laginaren jatorrizko populazioaren datuak hurrengo hauek balira?



4. grafikoa. $n = 11$ laginaren jatorrizko populazioaren datuak.

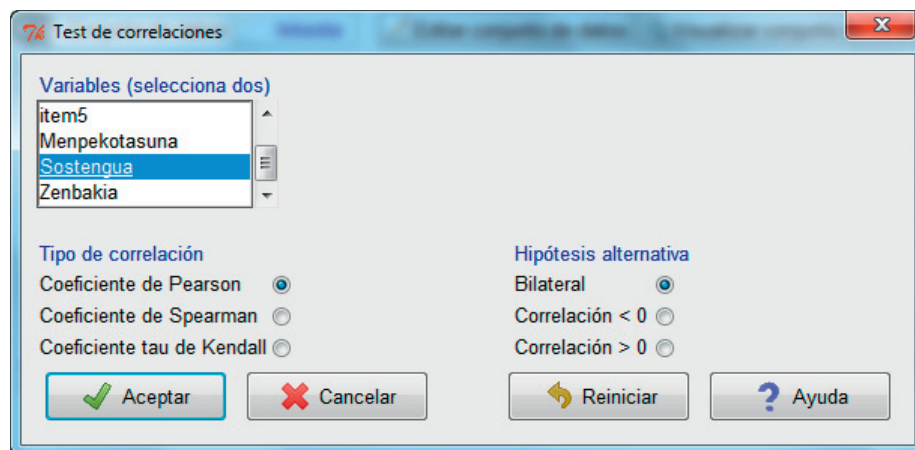
Hau 2. MOTAKO ERROREA da. Kasu honetan, lagineko datuetatik abiatuz erlazioirik ez dagoela erabakiko genuke nahiz eta populazioaren errealitatea kontrakoa izan. Baina errore mota honen azterketa beste baterako utziko dugu.

Jarraitu aurretik, edozein proba estatistikorekin hipotesien testa burutzean egin daitezkeen erroreen eta har daitezkeen erabakien laburpena aurkezten da.

Proba estatistikoetako erabaki eta erroreak		
Lagina	Populazioa	
	Desberdintasunak BAI	Desberdintasunak EZ
Esanguratsutasuna BAI (Hasierako hipotesia baztertu)	Erabaki zuzena <i>Hau gertatzeko probabilitatea:</i> Probaren POTENTZIA $= 1 - \beta$	Erabaki okerra I. Motako Errorrea <i>Hau gertatzeko probabilitatea:</i> α
Esanguratsutasuna EZ (Hasierako hipotesia onartu)	Erabaki okerra II. Motako Errorrea <i>Hau gertatzeko probabilitatea:</i> β	Erabaki zuzena <i>Hau gertatzeko probabilitatea:</i> Probaren POTENTZIA $= 1 - \alpha$

10.1.8. Korrelazio-testa

Resúmenes menuko aukera honen bidez, R Commander-ek korrelazio-koefizienteen balio zehatzak zoriz eskuratzeko probabilitatea zein den kalkulatu du. Hots, emaitzen esanguratsutasuna kalkulatu du. Horretarako Estadísticos > Resúmenes > Test de correlación sakatu behar da.



82. irudia.

```

Ventana de resultados
Ejecutar

> cor.test(Inkesta$Autoestima, Inkesta$Sostengua, alternative="two.sided",
+ method="pearson")

Pearson's product-moment correlation

data: Inkesta$Autoestima and Inkesta$Sostengua
t = 4.3416, df = 148, p-value = 2.61e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1858637 0.4710039
sample estimates:
cor
0.3361135
  
```

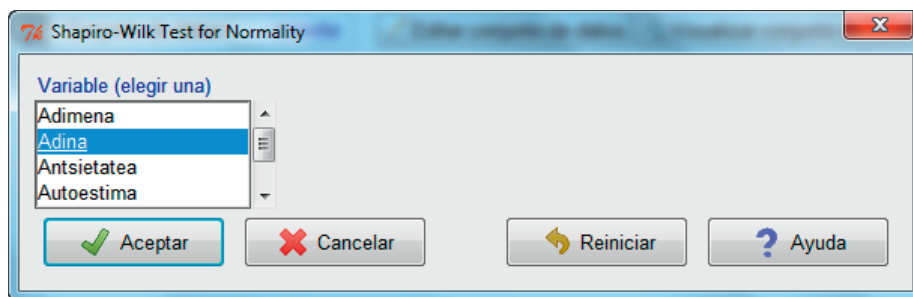
83. irudia.

Adibide honetan Sostengua eta Autoestima aldagai independenteak direla egiaztatu nahi dugu. Alegia, populazioan korrelazio-koefizientearen balioa 0 izan al litekeen jakin nahi dugu. Horretarako, laginean korrelazio-koefizientearen balioa kalkulatu dugu, $r = 0.3361135$. Balio hori zorizkoa eta populazioarekiko adierazgarria delakoan gaude. R Commander-ek % 95eko konfiantza-tartea (KT) aurkezten du. Zero balioa ez dago KTaren barnean. Emaiza hauekin independentzia hipotesia baztertuko dugu.

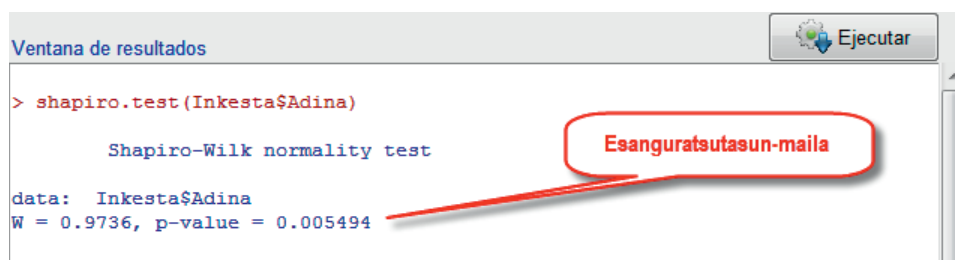
Emaitzak irakurtzeko beste modu bat Studenten t testa aztertzea da. Emaitzen leihoak laginaren t balioa, askatasun-graduak ($df = 148$; kasu kopurua $- 2$) eta t balioa zoriz eskuratzeko probabilitatea ($p = 0.0000261$) formatu esponontzialean aurkezten ditu. Probabilitate hau α errore-mailaren azpitik dago ($\alpha = 0,05$). Hortaz, hasierako hipotesia baztertuko dugu. Noski, bi ikuspegieta emaitzek, Studenten t testak eta KTak, ondorio berera garamatzate.

10.1.9. Shapiro-Wilken normaltasunaren testa

Edozein azterketa estatistiko burutu aurretik egiaztatzea komeni da datuek eredu aplikatua izan dadin beharrezko baldintza guztiak betetzen dituztela. Ziur asko ereduaren baldintzen artean gehien errepikatzen dena azterketaren objektu izango den aldagaia normaltasunaren legearen arabera banatzea da. Shapiro-Wilken normaltasunaren testak baldintza hori betetzen den egiaztatzea ahalbidetzen du. Mota honetako galderak erantzuten laguntzen digu: laginean Adina aldagaian lortu ditugun emaitzak ikusita, ondorioztatu al genezake aldagaia kurba normalaren arabera banatzen dela? (Betiere, lagina zoriz atera dela pentsatuz). Estadísticos > Resúmenes > Test de normalidad de Shapiro-Wilk... bidez irekiko den leihoan Adina aldagaia markatuko dugu:



84. irudia.



85. irudia.

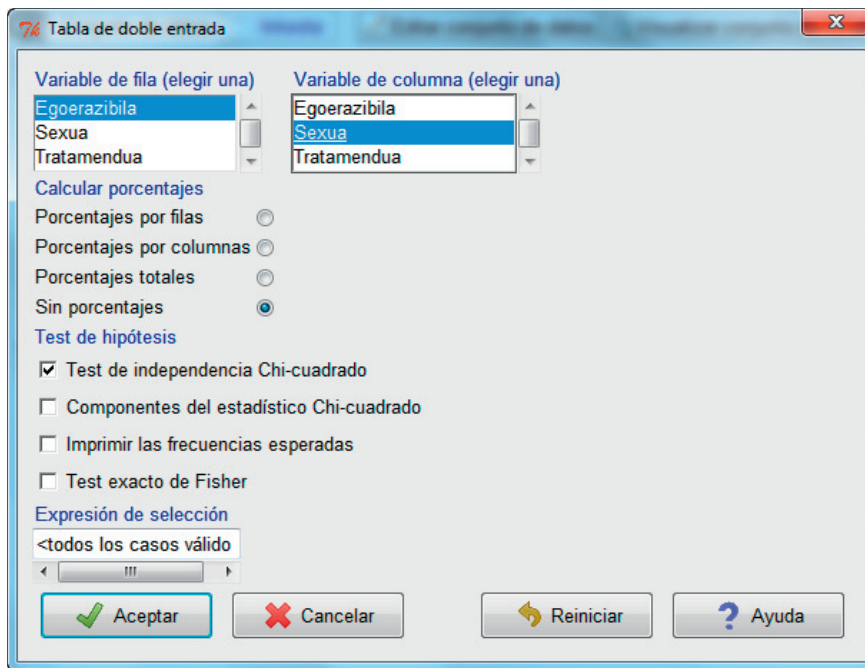
Laginarentzat Shapiro-Wilk estatistikoaren balioa $W = 0.9736$ da. Estatistiko hau banaketa teorikoaren eta lagineko datuen arteko desadostasunaren indize bat da. Bi banaketen arteko desberdintasunak zoriaren menpe egoteko probabilitatea 0.005494 da. Ondorioz, ez daukagu Adina aldagaia kurba normalaren arabera banatzen dela adierazten duen ebidentziarik ($0,05$ eko errore-mailarekin). Hortaz, aldagaia normaltasunaren legearen arabera banatzen dela jasotzen duen hasierako hipotesia baztertuko genuke.

10.2. KONTINGENTZIA-TAULAK

Kontingentzia-taulek eta haiei lotutako hipotesien testek aldagai kualitatiboen (faktoreen) arteko erlazioak aztertzeko informazioa eskaintzen dute.

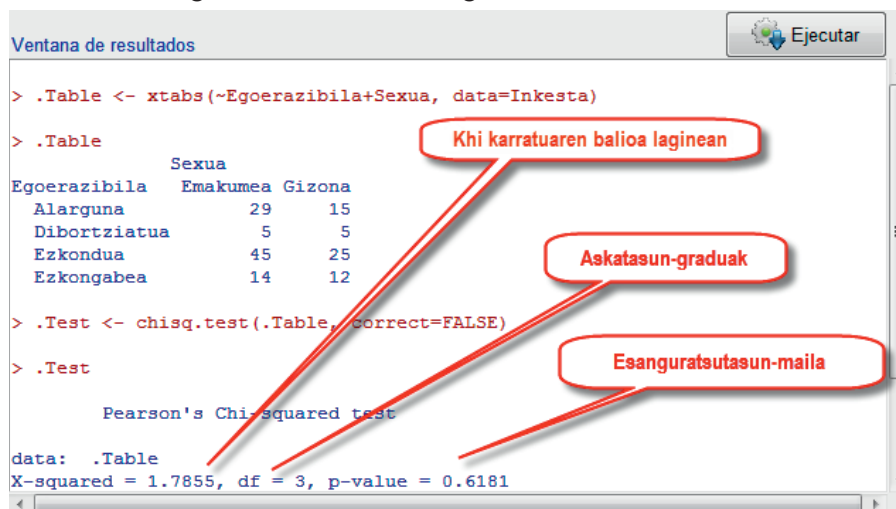
10.2.1. Sarrera bikoitzeko taulak

Estadísticos > Tablas de contingencia > Tabla de doble entrada... bidez, hurrengo irudiko leihoa eskuratuko dugu. Bertan, Egoerazibila eta Sexua aldagaiekin sarrera bikoitzeko (kontingentzia) taula eraikitzeko eta khi karratuaren independentzia-testa exekutatzeko eskatuko dugu.



86. irudia.

Emaitzen leihoan sarrera bikoitzeko taula eta khi karratuaren testaren emaitzak agertuko dira. Khi karratu estatistikoaren balioa lagingarentzat 1.7855 da, askatasun-graduak kopurua 3 ((lerro kopurua - 1) x (zutabe kopurua - 1)) da eta bi faktoreak independenteak balira zoriz emaitza hauekin topo egiteko probabilitatea 0.6181 da. Beraz, bi aldagaiak independenteak direla baztertzeko beharrezko ebidentziak ez ditugula ondorioztatuko dugu.

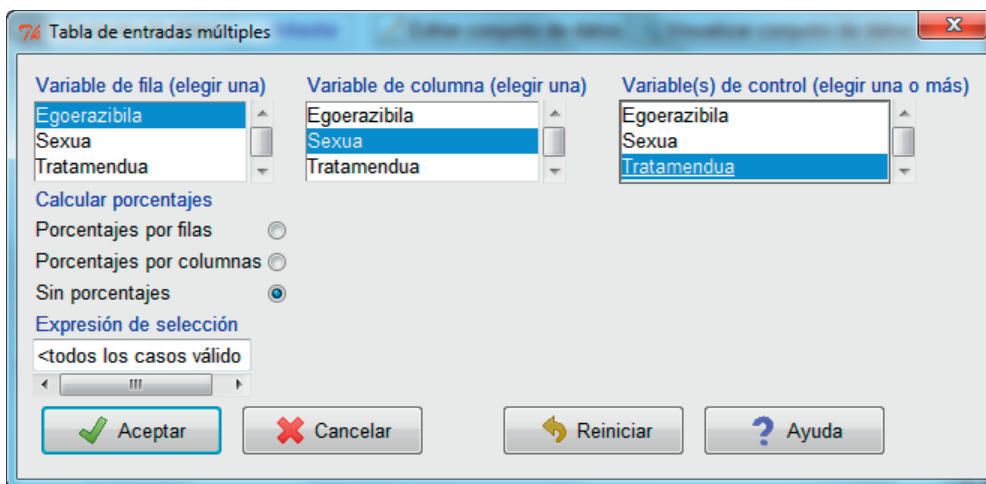


87. irudia.

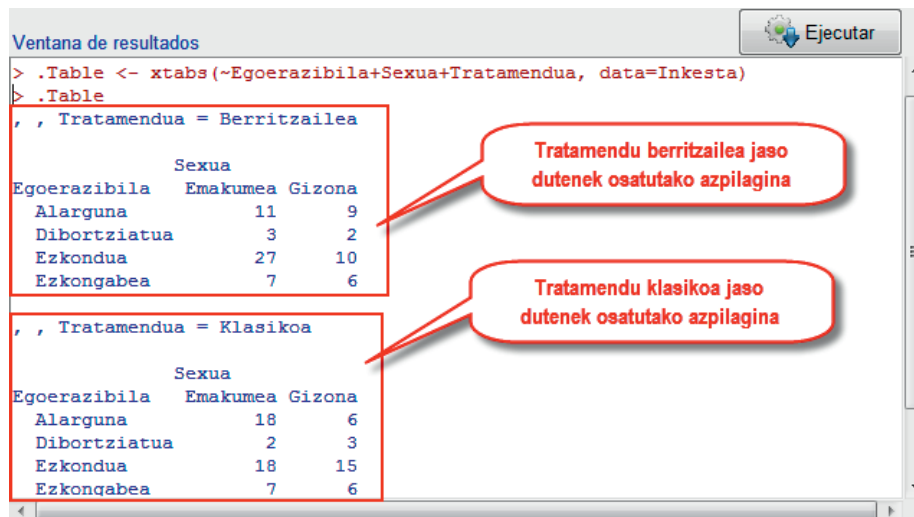
Emandako informazioa osatzeko portzentajeak lerroka edo zutabeka, portzentaje osoak, espero diren maiztasunak eta khi karratuaren osagarri estatistikoak kalkulatzeko aukera dago R Commander-en. Gainera Fisherren testa eskaintzen du, bereziki egokia faktore dikotomikoak ditugunean eta laukiren batean maiztasuna 5 baino txikiagoa denean.

10.2.2. Sarrera anitzeko taula

Tabla de entradas múltiples... aukerari esker dimentsio anitzeko taulak sor daitezke. Taula hauek bi aldagai baino gehiago gurutzatuz sortzen dira. Adibide honetan Egoerazibila, Sexua eta Tratamendua aldagaiak gurutzatu ditugu.



88. irudia.



89. irudia.

10.2.3. Sarrera bikoitzeko taula bat eraikitzea eta aztertzea

Aukera honekin maiztasunen taula zuzenean sor dezakegu. Pentsa dezagun Araba, Bizkai eta Gipuzkoako ikastetxe publiko eta pribatuetako ikasleen banaketa ezagutzen dugula eta hurrengoa dela:

	Araba	Bizkaia	Gipuzkoa
Publikoa	70	120	60
Pribatua	80	100	40

Probintzia eta ikastetxe motaren artean independentziaren hipotesia aztertu nahi dugu. Hau da, ea probintzia bakoitzeko ikastetxe publiko eta pribatuen portzentajeen arteko desberdintasunak zoriaren ondorio diren ala ez. Informazioa sartu ostean khi karratu estatistikoarekin batera, emaitza zoriaren ondorio izateko probabilitatea kalkula daiteke. Horretarako sakatu beharrekoa Estadísticos > Tablas de Contingencia > Introducir y analizar una tabla de doble entrada... da.

Introducir una tabla de doble entrada

Número de filas: 2
 Número de columnas: 3

Introducir las frecuencias:

	Araba	Bizkaia	Gipuzkoa
Publikoa	70	120	60
Privatua	80	100	40

Calcular porcentajes

☐ Porcentajes por filas
☐ Porcentajes por columnas
☐ Porcentajes totales
☒ Sin porcentajes

Test de hipótesis

☒ Test de independencia Chi-cuadrado
☐ Componentes del estadístico Chi-cuadrado
☐ Imprimir las frecuencias esperadas
☐ Test exacto de Fisher

90. irudia.

Emaitzen leihoan sarrera bikoitzeko taula agertuko da, eta jarraian, khi karratuaren testaren informazioa agertuko da. Adibide honetan zenbatetsitako balioek ($p = 0.1008$) ez dute adierazten desberdintasun esanguratsurik dagoenik hiru probintzien artean.

Ventana de resultados

```
> .Table # Counts
      Araba Bizkaia Gipuzkoa
Publikoa   70    120     60
Privatua   80    100     40

> .Test <- chisq.test(.Table, correct=FALSE)

> .Test

      Pearson's Chi-squared test

data:  .Table
X-squared = 4.5887, df = 2, p-value = 0.1008
```

91. irudia.

10.3. BATEZBESTEKOAK

Resúmenes azpimenuan aurkitzen den Medias aukerak, parte-hartzaileak hautatutako faktorearen edo faktoreen arabera taldekatzen ditu eta faktore-maila bakoitzeko batezbestekoak kalkulatu. Medias menuan aipatutako informazioa laginaren batezbestekoekin erlazionatutako hipotesi-testekin osatzeko aukerak aurkituko ditugu. Atal honetako proba guztien logika berdina da. Lehenengo laginaren estatistikoa eskuratzen da eta jarraian balio hori zoriz eskuratzeko probabilitatea zenbatekoa den zenbatesten da.

10.3.1. Talde batentzako t testa

Talde bateko t testa moduan ezagutzen den hipotesi-probaren hasierako hipotesiak aldagai kuantitatibo baten batezbestekoa populazioan (μ) ikertzaileak ezartzen duen balio baten berdina dela jasotzen du. Talde batentzako t testak laginaren batezbestekoa ($\bar{x}$) populazioaren batezbestekotik (μ) eratortzeko probabilitatea kalkulatzeko du. t estatistikoak bere formulazioan kontuan hartzen ditu laginaren tamaina eta balioen banaketa:

$$t = \frac{\bar{x} - \mu}{s_x / \sqrt{(n - 1)}}$$

1. formula.

Non:

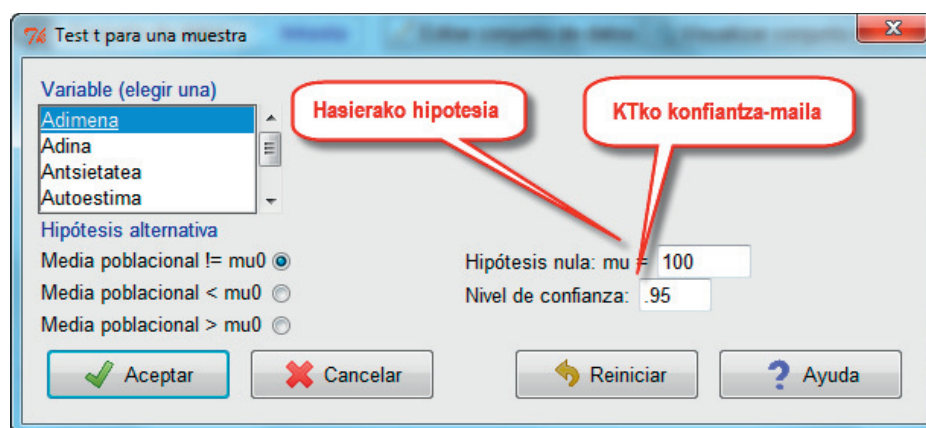
$\bar{x}$: laginaren batezbestekoa den.

n : parte-hartzaile kopurua den.

μ : populazioaren batezbestekoa den.

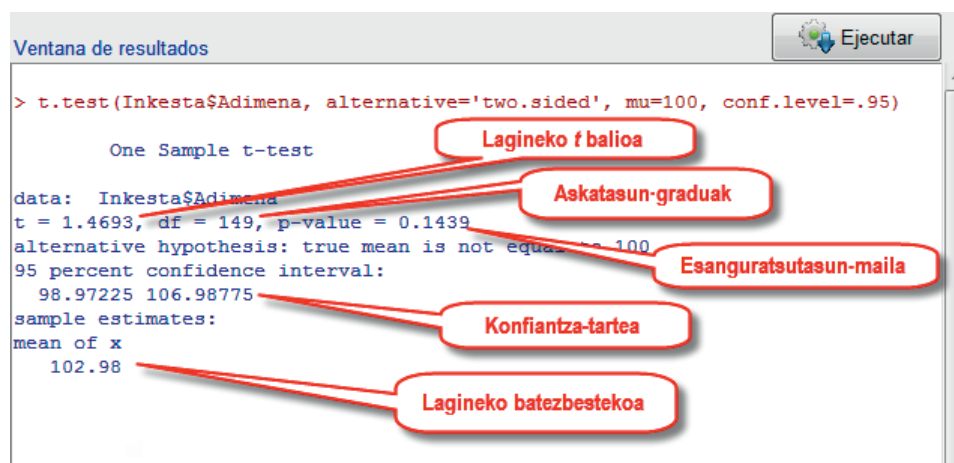
s_x : lagineko desbideratze tipikoa den.

Aurkezten dugun adibidean, Adimena aldagaiaren batezbestekoa populazioan 100 izatea onargarria den jakin nahi dugu. Estadísticos > Medias > Test t para una muestra... aukeren sekuentziari jarraituko diogu.



92. irudia.

Ireki den leihoan aldebiko test batentzako % 95eko konfiantza-tartea eraikitzeke eskatu diogu. Alegia, gehienez ere % 5eko errore-maila onartuko dugu. Kontuan izanik aldebikoa dela, errorearen erdia (% 2,5) banaketaren ezkerreko buztanean geratuko da eta beste hainbeste eskuinekoan.



93. irudia.

R Commander-ek banaketaren t balioa ($t = 1.4693$), askatasun-graduak (parte-hartzaile kopurua $- 1$) eta populazioaren batezbestekoa 100 izanik t balio hori eskuratzeko probabilitatea ($p = 0.1439$) zenbatekoa den erakutsiko digu. t balioa zoriz eskuratzeko probabilitatea % 5 baino handiagoa denez, laginean eskuratutako batezbestekoa (102.98) bat dator populazioaren batezbestekoa 100 dela jasotzen duen hasierako hipotesiarekin.

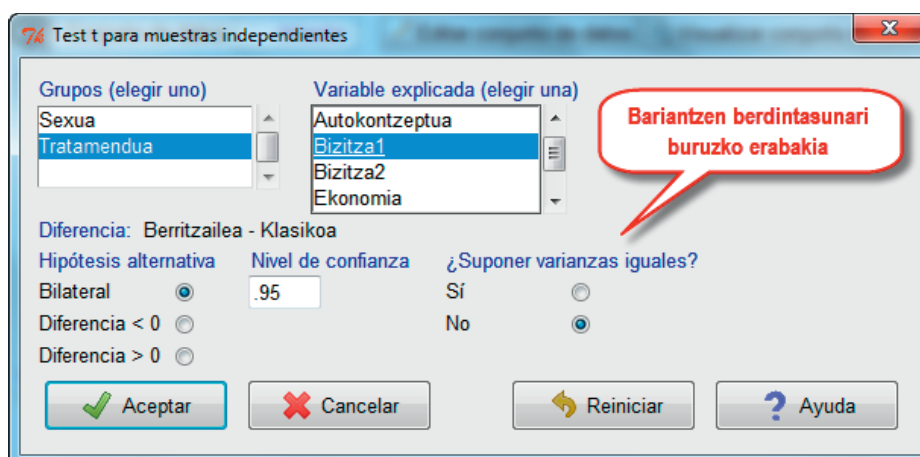
Emaizak interpretatzeko beste modua konfiantza-tartea (KT) erabiltzea da. Horretarako, kontuan izanik laginaren batezbestekoa 102.98 dela, % 95eko konfiantza-tartea eraikitzen da. KTa behetik eta gohitik mugatzen duten balioak 98.97225 eta 106.98775 dira. Alegia, populazioaren batezbestekoa 98.97225 eta 106.98775 balioen artean egoteko % 95eko probabilitatea dago. Hasierako hipotesiak populazioaren batezbestekoa 100 dela jasotzen duenez eta balio hori KT barruan dagoenez, ez dugu baztertuko hasierako hipotesia.

10.3.2. *Lagin independenteentzako t testa*

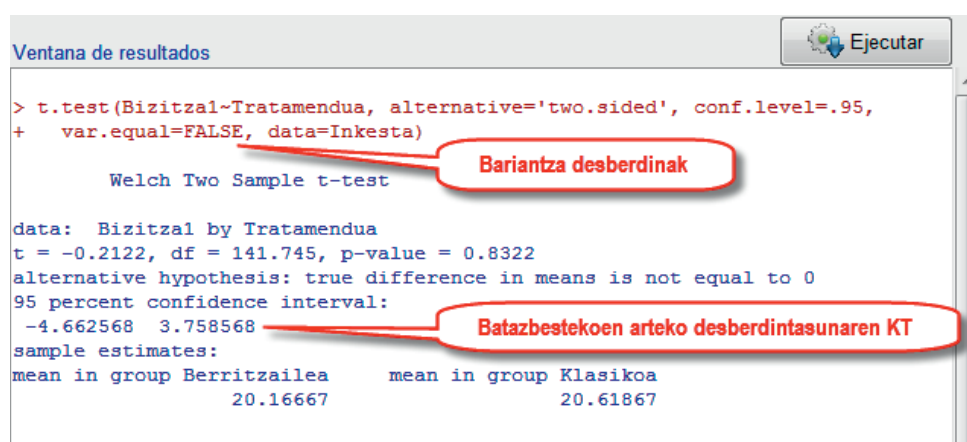
Imajina dezagun bizitzarekiko gogobetetze-maila hobetzen duten bi tratamendu desberdin ikertzen ari garela (liburu hau irakurtzea horietako bat izan liteke) eta bien emaitzak konparatu nahi ditugula. Hasteko, tratamenduak hasi aurretik bi taldeen bizitzarekiko gogobetetze-maila parekoa dela baieztatu behar dugu. Alegia, tratamendu Klasikoa eta Berritzailea jaso dutenen batezbestekoa Bizitzal aldagaian berdina ote den ikusi behar dugu. Bi talde desberdin dira eta lehenengo taldea definitzen duten elementuen hautapenak ez du bigarren taldea definitzen duten elementuen hautapena baldintzatzen. Horretarako, Estadísticos > Medias > Test t para muestras independientes... erabiliko dugu.

Azpiko irudian ikusi daitekeen moduan, taldeak Tratamendua aldagaiaren arabera definitu ditugu eta Bizitzal aldagaian talde bakoitzak jasotako balioak konparatuko ditugu. Bariantzak berdinak direla pentsatuz¹ Klasikoa eta Berritzailea taldeen arteko desberdintasunak aztertu nahi ditugu aldebiko proba batean. Konfiantza-maila moduan Gizarte Zientzietan ohikoa den balioa mantenduko dugu, % 95.

1. Bi formula daude Studenten proban laginaren t estatistikoa kalkulatzeko. Bata, bariantzen berdintasunerako; eta bestea, bariantzak berdinak ez direla pentsatzen dugunerako. Hortaz, analisisa gauzatu baino lehen bariantzen berdintasunaren hipotesia aztertu beharko genuke eta horren inguruan erabaki bat hartu. Ikusi 10.5. atalean analisi hau aurrera eramateko modua.



94. irudia.



95. irudia.

Emaitzen leihoan lagingaren t estatistikoaren balioa ($t = -0.2122$) agertuko da. Estatistikoa kalkulatzeko formula hurrengoa da:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

2. formula.

Jarraian, R Commander-ek t banaketaren askatasun-graduen balioa ($df = 141.745$) eta lagingaren t balioa zoriz eskuratzeko probabilitatearen zenbatekoa ($p = 0.8322$) emango dizkigu.

Hurrengo lerroetan populazioaren batezbestekoen desberdintasunaren konfiantza-tartea aurkezten da. Kasu honetan hasierako hipotesiak populazioaren batezbestekoak berdinak direla dioenez eta aipatutako balioa konfiantza-tartean mugaturik dagoenez ($KT = (-4.662568, 3.758568)$), ez dugu baztertuko hasierako hipotesia. Ez baitauekagu populazioaren batezbestekoak desberdinak direla esateko ebidentziarik.

Irakurleak baieztatu dezakeen moduan, bariantzen berdintasunaren proba burutu ostean (Varianzas azpimenutik abiatuz egin dezakegu) bariantza berdinen analisia aukeratu gendake eta erabakiaren norabidea berdina izango litzateke. Hurrengo leihoan emaitzak aurkezten dira.

```

Ventana de resultados
Ejecutar

> t.test(Bizitzal~Tratamendua, alternative='two.sided', conf.level=.95,
+ var.equal=TRUE, data=Inkesta)

Two Sample t-test

data: Bizitzal by Tratamendua
t = -0.2122, df = 148, p-value = 0.8322
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.661036  3.757036
sample estimates:
mean in group Berritzailea      mean in group Klasikoa
      20.16667                  20.61867
  
```

96. irudia.

Bi kasuetan analisi-prozesua berbera den arren, bigarrenean lagainen t estatistikoa eskuratzeko formulak itxura hau hartuko luke:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 \times n_1 + s_2^2 \times n_2}{n_1 + n_2 - 2} \times \frac{n_1 + n_2}{n_1 \times n_2}}}$$

3. formula.

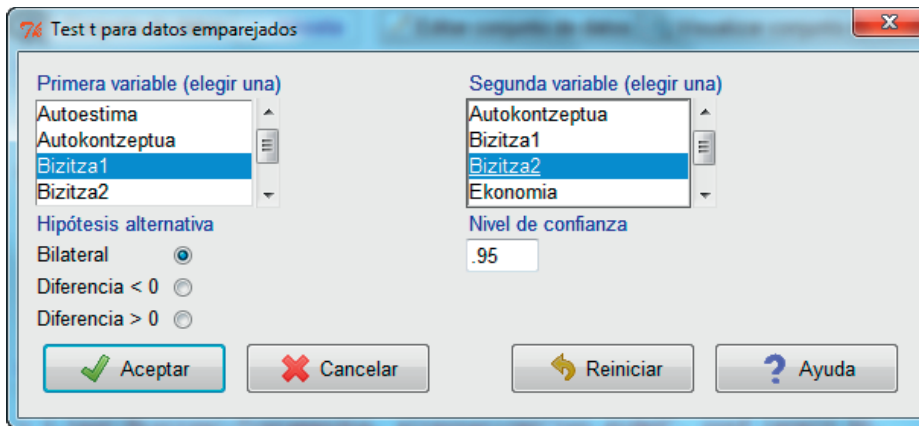
10.3.3. Erlazionatutako datuentzako t proba

Erlazionatutako lagainen eredu hiru kasutan aplikatu daiteke:

- Talde bereko parte-hartzaileen emaitzak konparatzeko esperimendu bat egin aurretik eta ostean. Liburu honen adibidean bizitzarekiko gogobetetze-maila hobekuntza-programa (tratamendua) egin aurretik eta ostean alderatu nahi dugu.

- Talde berean eskala berdinean neurtutako bi aldagai konparatzeko. Adibidez, bi ikasgai desberdinetan eskuratutako emaitzak.
- «Parekatuz» eraiki diren bi taldeetako parte-hartzaileen emaitzak konparatzeko. Hots, lagineko parte-hartzaileak «pare baliokidetan» banatu dira, eta ondoren, pareetako parte-hartzaile bakoitzari zoriz talde bat esleitzen zaio.

Gure adibidearekin jarraituz, tratamenduaren aurretik eta ostean gogobetetze-mailaren emaitzak konparatu nahi ditugu. Horretarako Estadísticos > Medias > Test t para datos relacionados... sakatuko dugu.



97. irudia.

Aldebiko testa aplikatuko dugu eta Bizitza1 eta Bizitza2 aldagaien batezbestekoen arteko diferentziarako % 95eko konfiantza-tartea zenbatetsiko dugu.

Emaitzetan ikus daitekeen moduan, gure laginean bi aldagaien batezbestekoen arteko desberdintasuna -28.814 da. Laginean Bizitza2 aldagaiaren batezbestekoa Bizitza1 aldagaiarena baino 28.814 unitate handiagoa da.

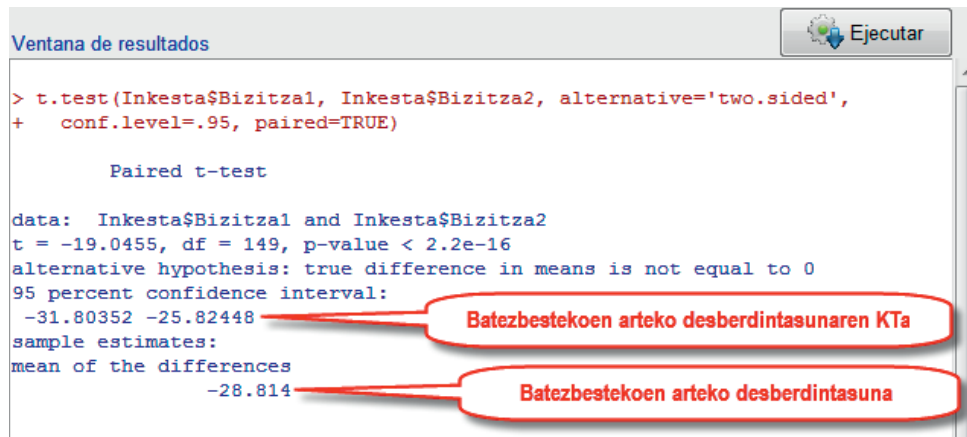
Desberdintasun horren esanguratsutasuna aztertzeke eta % 95eko konfiantza-tartea eraikitzeke, lehenengo t estatistikoaren kalkuluari ekingo diogu:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 + s_2^2 - 2 \times r_{12} \times s_1 \times s_2}}$$

4. formula.

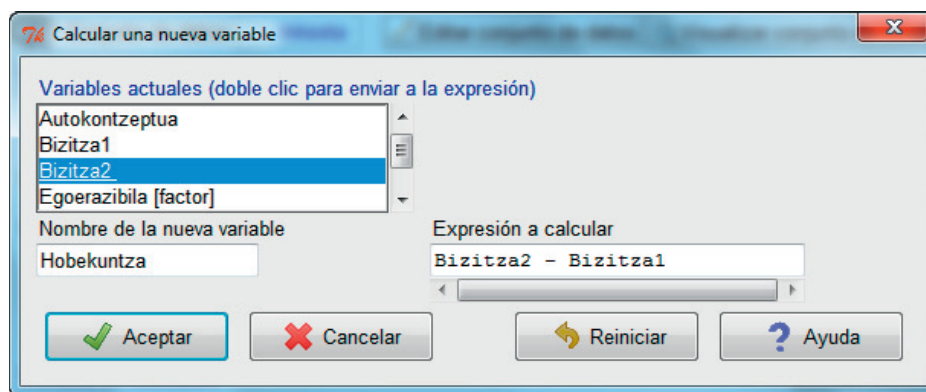
Askatasun-graduen kopurua (parte-hartzaile kopurua $- 1$) 149 da. Laginean t balioa ($t = -19.0455$) eskuratzeko probabilitatea hain da txikia, ezen esponentzialki adierazi behar baitu R Commander-ek. Hortaz, hasierako hipotesia

baztertuko dugu. Alegia, esperimendua egin aurretik eta ostean eskuratutako batezbestekoak ez direla berdinak adierazten du ebidentziak. Momentu honetan desberdintasunen azterketan pauso bat haratago joateko eta desberdintasun horien tamaina zenbatesteko beharrak azpimarratzea komeni da. Horretarako % 95eko konfiantza-tartearen (KT) mugak aurkezten dira, -31.80352 eta -25.82448 .



98. irudia.

Adibidearekin amaitzeko, kontuan izanik bizitzarekiko gogobetetze-mailan aldaketa esanguratsua egon dela ondorioztatu dugula, Hobekuntza aldagaia kalkulatzeko nahiko arrazoi daukagu. Aldagai hau Bizitza2 eta Bizitza1 aldagaien diferentziak definituko du. Kalkulatzeko Datos > Modificar variables del conjunto de datos activo > Calcular una nueva variable... erabiliko dugu.



99. irudia.

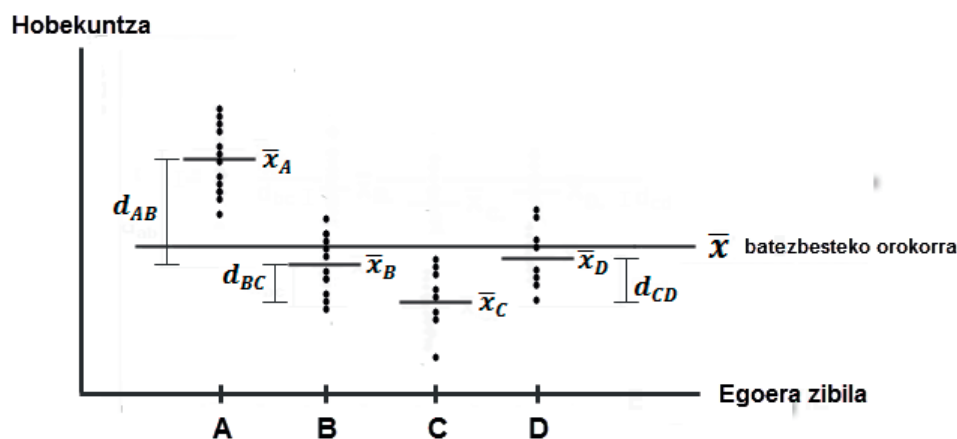
Hobekuntza aldagai berri gisa gehituko zaio datu multzoari. Ez ahaztu aldagai berria hurrengo lan-saioetan eskuragarri izan dezazun aldatu berri duzun

datu multzoa gorde egin behar duzula. Horretarako `Datos > Conjunto de datos activo > Guardar el conjunto de datos activos...` sakatu dezakezu.

10.3.4. Faktore bakarreko ANOVA

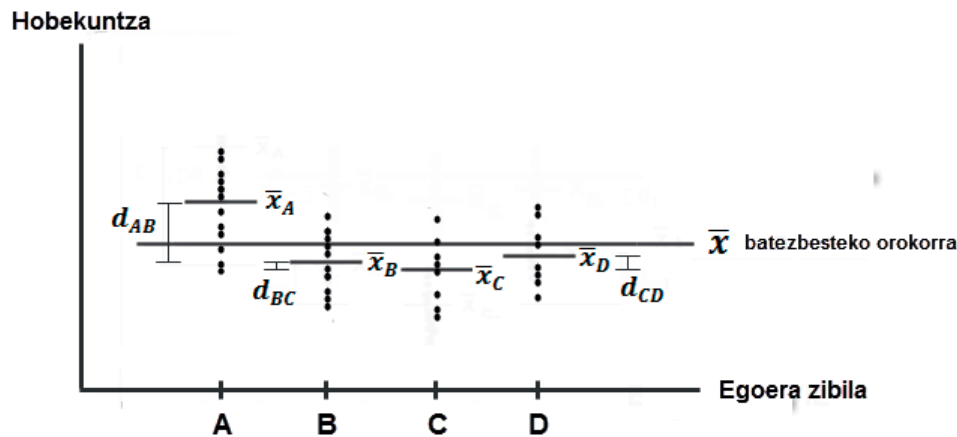
Studenten t probak talde independenteen nahiz erlazionatutako taldeen diseinuetan bi batezbestekoren arteko desberdintasunaren esanguratsutasuna aztertzeko aukera ematen du. Baina bi talde baino gehiago konparatu behar baditugu, bi tratamendu baino gehiago baditugu edo, besterik gabe, bi neurri baino gehiago konparatzeko desberdintasun horien azterketa ahalbidetzen duen proba estatistikoa bariantzaren analisisa (ANOVA) da.

ANOVA mota desberdineko bi distantzian oinarritzen da, aldagaiaren kategoria bakoitzaren balioen (TALDE BARNEKOAK) arteko distantzietan eta kategoria desberdinen batezbestekoen arteko distantzietan (TALDE ARTEKOAK). Talde bakoitza osatzen duten parte-hartzaileen arteko distantziak talde horren batezbestekoarekin konparatu ditzakegu. Adibidez, A taldean, parte-hartzaile bakoitzetik batezbestekora ($\bar{x}_A$) dagoen distantzia neur dezakegu; B taldean parte-hartzaile bakoitzetik batezbestekora ($\bar{x}_B$) dagoen distantzia neur dezakegu... Gainera, talde desberdinen batezbestekoen arteko desberdintasunak kalkula ditzakegu (d_{AB} , d_{BC} , d_{CD}). Hurrengo grafikoan taldeen batezbestekoak desberdinak direla eta parte-hartzaileen arteko distantziak baino handiagoak direla ikus daiteke.



5 grafikoa. Batezbesteko desberdinak dituzten taldeak.

Antzeko batezbestekoak dituzten taldeak irudikatzen dituen grafiko honetan (hurrengoan), taldeen arteko distantziak (d_{AB} , d_{BC} , d_{CD}) txikiak dira talde barruan dauden distantziekin konparatzen baditugu.

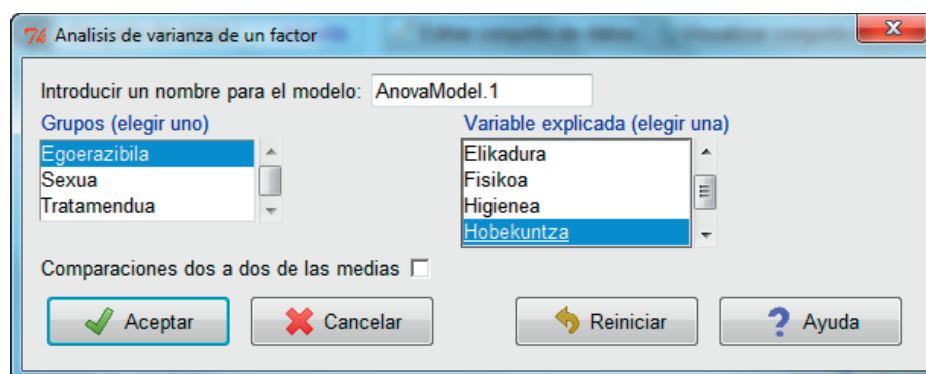


6. grafikoa. ANOVA. Antzeko batezbestekoak dituzten taldeak.

Talde arteko distantzien batezbestekoaren eta talde barneko distantzien batezbestekoaren arteko desberdintasunaren indizeari Snedecorren F deritzo. Taldeetako batezbestekoak berdinak direnean, F estatistikoaren balioa 0 da. ANOVAREN hasierako hipotesiak populazio arteko batezbestekoak berdinak direla adierazten du. Alegia, laginean aipatutako bi distantzien (talde artekoa eta talde barnekoa) arteko kenketak balio baxua ematen duela. Laginean F balio zehatz bat zoriz topatzeko probabilitatearen balioak hasierako hipotesia onartzera edo baztertzeraz bultzatuko gaitu.

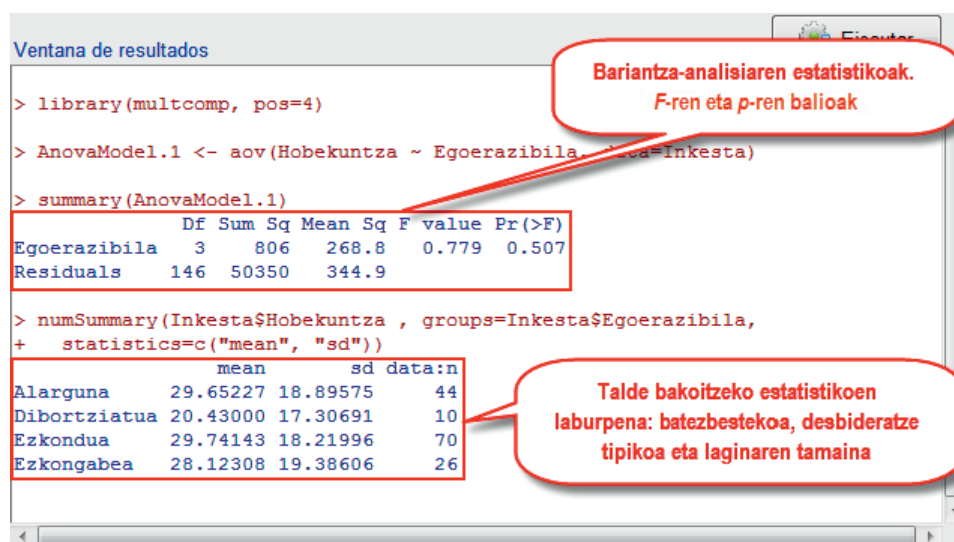
Adibide moduan Hobekuntza aldagaiko (aurreko atalean sortu dugu) emaitzak aztertuko ditugu. Egoerazibila-ren kategoriek Hobekuntza aldagaian dituzten batezbestekoak alderatzea eta kategorien arteko desberdintasunak zoriaren ondorio direla baieztatzea izango da gure helburua. Hau da, populazioan Hobekuntza aldagaiaren batezbestekoan ez dagoela Egoerazibila-ren araberako desberdintasunik. `Estadísticos > Medias > ANOVA de un factor...` sakatuko dugu analisiari ekiteko.

Ezkerreko aldean taldeak sortzeko aldagaia aukeratu dugu, Egoerazibila, eta eskuinaldean neurtu dugun aldagaia markatu dugu, Hobekuntza.



100. irudia.

Emaitzen leihoan bariantzaren analisiaren taula eta Hobekuntza aldagaien laburpena alderatutako lau taldeen arabera aurkezten dira. Ikus daitekeen moduan, dibortziatuen batez besteko hobekuntza (20.43) beste taldeena baino apur bat txikiagoa da. Baina kontuan izan behar da laginean daukagun dibortziatu kopurua ($n = 10$) oso txikia dela eta haren barne-aldakortasuna handia dela; haren desbideratze tipikoa 17.30691 da.



101. irudia.

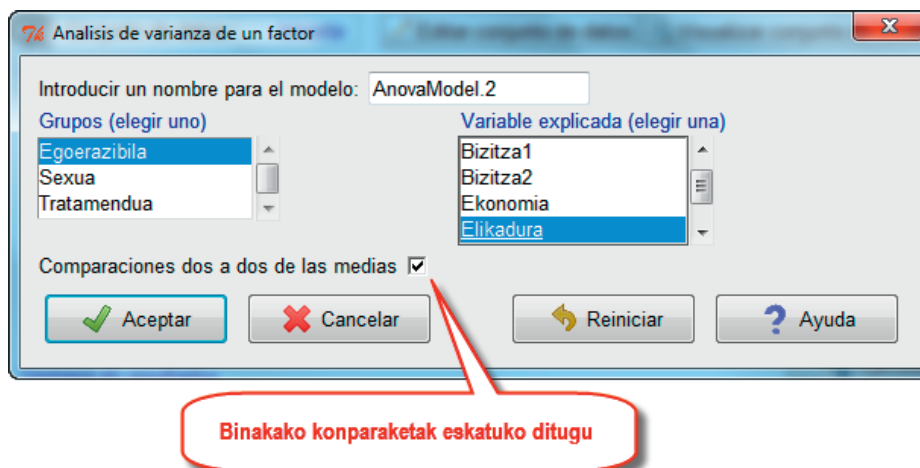
Bariantzaren analisiaren taulak (emaitzen lehen zatia) lagineko F balioa ($F = 0.779$) eta hasierako hipotesia onartuz (Hobekuntza aldagaiko batezbestekoak populazioko lau kategorietan berdina da) balio hori eskuratzeko probabilitatea ($p = 0.507$) ematen digu. Laburtuz, emaitzek laginetako balioek hasierako

hipotesia onartzeraz gero. Irakurleak baieztatu dezakeen moduan F balioa talde arteko distantzien karratuen batezbestekoaren (talde arteko batezbesteko koadratikoa; $\text{Mean Sq}_{\text{Egoerazibila}} = 268.8$) eta parte-hartzaileen arteko distantzien karratuen batezbestekoaren (talde barneko batezbesteko koadratikoa; $\text{Mean Sq}_{\text{Residuals}} = 344.9$) arteko zatidura da. Egoerazibila aldagaiaren askatasun-gradu kopurua kategoria kopuruari 1 kenduta kalkulatzen da eta hondakinen (*Residuals*) kasuan parte-hartzaile kopuruari 1 kenduta.

Lau taldeen arteko konparaketa modu orokorrean egiten da. Emaitzek batezbestekoen artean aurkitutako desberdintasunak hasierako hipotesia baztertzeko behar bezain handiak ez direla adierazten dute.

Emaitzek hasierako hipotesia baztertzea eskatuko balute; hots, populazioaren batezbestekoak desberdinak izan litezkeela adieraziko balute, galdera hau egin beharko genuke: zeintzuk dira batezbesteko desberdinak dituzten talde pareak? Galdera horri erantzuteko *post hoc* alborapenak egin behar dira. Hau da, batezbestekoak binaka konparatu behar dira.

Hurrengo adibidean, Elikadura aldagaia Egoerazibila aldagaiko lau kategorietan konparatuko dugu eta binakako konparaketak egiteko aukia aktibatuko dugu.



102. irudia.

Irteerako lehen atalean bariantzaren analisisiko oinarritzko emaitzak aurkituko ditugu.


```

> AnovaModel.2 <- aov(Elikadura ~ Egoerazibila, d
> summary(AnovaModel.2)
              Df Sum Sq Mean Sq F value    Pr(>F)
Egoerazibila   3  13.08   4.359    3.938 0.00999 **
Residuals    129 142.80   1.107
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
17 observations deleted due to missingness

> numSummary(Inkesta$Elikadura , groups=Inkesta$Egoerazibila,
+   statistics=c("mean", "sd"))
      mean      sd data:n data:NA
Alarguna  2.658537 1.1315864    41      3
Dibortziatua 3.750000 0.7071068     8      2
Ezkondua   2.524590 1.0584038    61      9
Ezkongabea  2.304348 0.9739695    23      3

```

103. irudia.

F -ren balioak ($F = 3.938$) adierazten du taldeen arteko distantzia taldeen barneko distantzia baino ia lau aldiz handiagoa dela. Katetoria bakoitzeko batezbestekoa populazioan berdina izango balitz, emaitza hau zoriz eskuratzeko probabilitatea 0.00999 izango litzateke. Ondorioz, % 5eko errore-mailarekin batezbestekoen arteko berdintasuna jasotzen duen hasierako hipotesia baztertuko dugu. Orain erantzun beharreko galdera hurrengoa da: batezbestekoak desberdinak dira, baina zein kategoriaren artean? Batezbestekoen arteko diferentzia handienak dituztenez susmatu genezake (Ezkongabea eta Dibortziatua). Baina batezbestekoen arteko desberdintasunak baloratzeko beharrezkoa da desbideratze tipikoak eta taldeen tamainak ere kontuan hartzea.

R Commander-ek binakako konparaketei eusteko edo *post hoc* probak gauzatzeko Tukeyren proba erabiltzen du.

```

Ventana de resultados
Ejecutar

Quantile = 2.5741
95% family-wise confidence level

Linear Hypotheses:
              Estimate lwr      upr
Dibortziatua - Alarguna == 0    1.09146  0.04467  2.13826
Ezkondua - Alarguna == 0      -0.13395 -0.68089  0.41300
Ezkongabea - Alarguna == 0     -0.35419 -1.05975  0.35137
Ezkondua - Dibortziatua == 0   -1.22541 -2.24380 -0.20702
Ezkongabea - Dibortziatua == 0 -1.44565 -2.55731 -0.33399
Ezkongabea - Ezkondua == 0     -0.22024 -0.88293  0.44245

> cld(.Pairs) # compact letter display
Dibortziatua    Ezkondua    Ezkongabea    Alarguna
      "a"           "b"           "b"           "b"

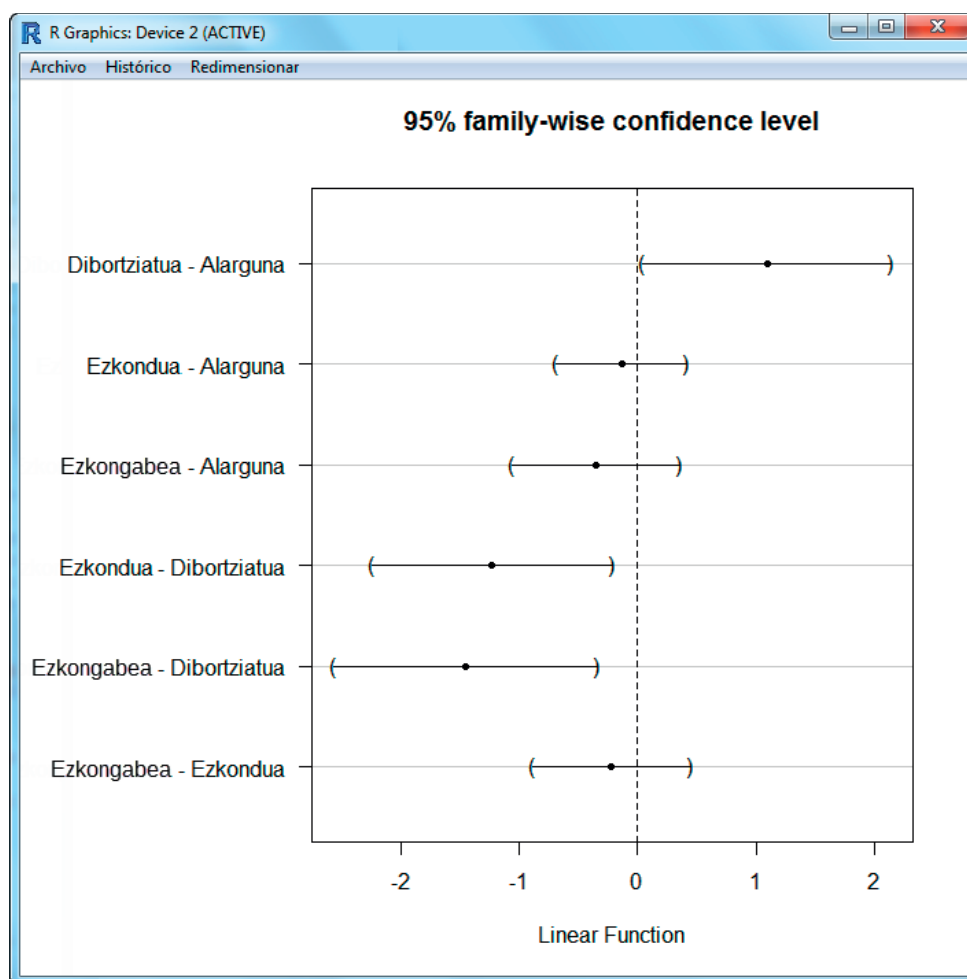
```

104. irudia.

Emaitzen lerro bakoitzean bi taldeen batezbestekoen arteko berdintasuna hipotetizatzen da. Talde arteko desberdintasunak (Estimate zutabeko balioa dagozkion batezbestekoen arteko kenketarekin eskura daiteke) eta % 95eko KTaren mugak (behekoa lwr da eta goikoa upr) zehazten dituzte balioek. Dibortziatua kategoriaren konparaketako desberdintasunetan 0 balioa (hasierako hipotesia) KTko mugetatik kanpo geratzen da. Beste guztietan 0 balioa KTko mugen barnean dago; hortaz, kasu hauetan hasierako hipotesia ez dugu baztertuko. Zehatzago, soilik Dibortziatua taldeak Elikadura aldagaian beste kategoriek baino batezbesteko handiagoa duela baieztatzeko ebidentzia dugu. Beste kasuetan ez daukagu batezbestekoen arteko berdintasuna baztertzeko adina «segurtasun». Laburtuz, % 5eko errore-mailarekin elikadura-maila desberdina duen talde bakarria dibortziatuena dela esango dugu.

Azken lerroetan R Commander-ek desberdintasunen esanguratsutasuna behatzeko laguntza ematen du. a eta b etiketak esleitzen dizkien bi talde homogeneoren presentzia deskribatzen du. b taldeko desberdintasunak (alargunak, ezkonduak eta ezkongabeak) ez dira esanguratsuak. a taldeak dibortziatuak bakarrik jasotzen ditu. Ondorioz, emaitzen arabera Elikadura aldagaian bi talde desberdinu ditzakegu: batean dibortziatuak daude, eta bestean, gainerakoak.

Azkenik, R Commander-ek emaitza hauek grafiko batean irudikatzen ditu. Grafikoa kategorien arteko binakako konparaketa desberdinen KTekin bat dator. 0 balioa jasotzen ez duten KTak dibortziatuen kategoria jasotzen ez duten konparaketak dira.



7. grafikoa. Kategoria pare bakoitzak Elikadura aldagaian duen % 95eko konfiantza-tartea.

10.3.5. Faktore anitzeko ANOVA

Aurreko ataletan batezbestekoak taldeen artean alderatu ditugu sailkapen-irizpide bakarra erabilita (faktore bakarra). Atal honetan, aldi berean, sailkapen irizpide bat baino gehiago (faktoreak) ditugun egoeretaraino zabalduko dugu analisia. Faktore anitzeko ANOVAk faktore bakarrekotik desberdintzen duen ezaugarri osagarri bat dauka; eta hau, faktore bakarreko ANOVA analisien errepikapen hutsarekin ez da lortzen. Horri INTERAKZIOA deitzen diogu. Pentsa Bizitzal aldagaian emakumeek eta gizonek lortutako emaitzak konparatu nahi ditugula. Horretarako Studenten t testa erabil dezakegu (Estadísticos > Medias > Test t para muestras independientes).

```

Ventana de resultados
Ejecutar

> t.test(Bizitzal~Sexua, alternative='two.sided', conf.level=.95, var.equal=FALSE,
+ data=Inkesta)

Welch Two Sample t-test

data: Bizitzal by Sexua
t = -0.0044, df = 139.578, p-value = 0.9965
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.098320  4.080323
sample estimates:
mean in group Emakumea  mean in group Gizona
                20.38925                20.39825

```

105. irudia.

Oro har, gizonen eta emakumeen bizitzarekiko gogobetetze-mailen (Bizitzal) desberdintasuna txikia dela (20.38 eta 20.39) eta desberdintasun hori esanguratsua izatera iristen ez dela ikus daiteke ($p\text{-value} = 0.9965$).

Bizitzal aldagaien Egoerazibila faktoreko kategoria bakoitzean lortutako emaitzak ere aztertu ditzakegu; horretarako Estadísticos > Medias > ANOVA de un factor... sakatuko dugu. Azpiko irudian antzeman daitekeen moduan, dibortziatuek eskuratu dute balio altuena bizitzarekiko gogobetetze-mailan; haiei ezkonduak jarraitu diete; hirugarren lekuan alargunak kokatu dira; eta azkenik, puntuazio baxuena ezkongabeek eskuratu dute.

```

Ventana de resultados
Ejecutar

> AnovaModel.3 <- aov(Bizitzal ~ Egoerazibila, data=Inkesta)

> summary(AnovaModel.3)
              Df Sum Sq Mean Sq F value Pr(>F)
Egoerazibila   3    356    118.5   0.697  0.555
Residuals    146  24831    170.1

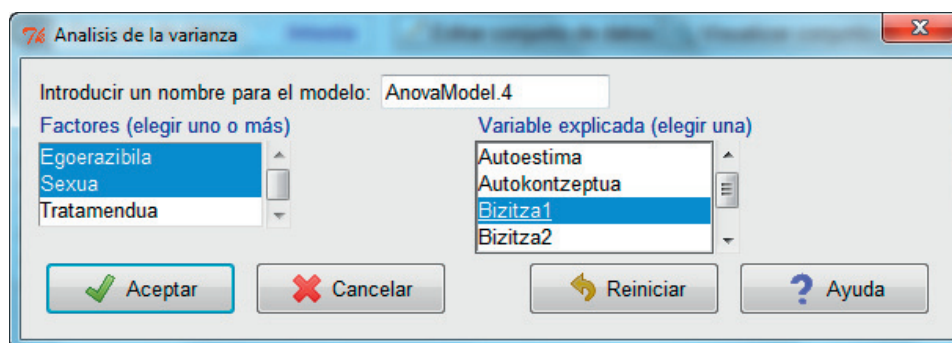
> numSummary(Inkesta$Bizitzal, groups=Inkesta$Egoerazibila,
+ statistics=c("mean", "sd"))
      mean      sd data:n
Alarguna  20.00682 13.31431   44
Dibortziatua 24.07000 10.34387   10
Ezkondua    21.08714 13.27461   70
Ezkongabea   17.76154 12.78468   26

```

106. irudia.

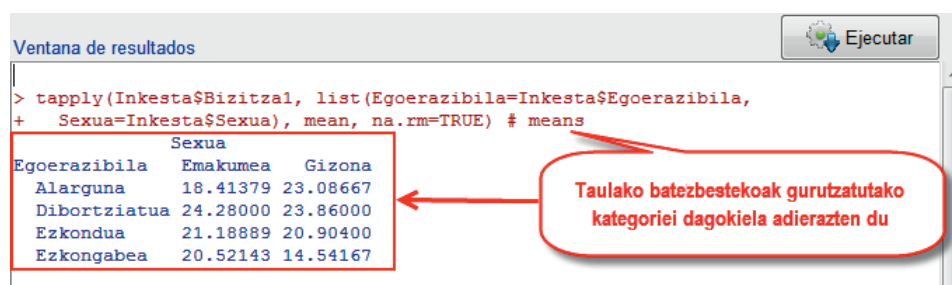
Egoera zibil bakoitzean gizon eta emakumeen batezbestekoen arteko desberdintasunak aztertzeke faktore anitzeko ANOVA erabiliko dugu Estadísticos > Medias > ANOVA de múltiples factores... bidez.

Adibide honetan bi faktoreen (Egoerazibila eta Sexua) bariantza-analisia eskatuko dugu. Ikertuko dugun aldagaia (aldagai azaldua edo menpekoea) Bizitza1 da.



107. irudia.

Emaitzek aurreko analisietatik ondorioak bizkorregi atera ditugula adierazten dute. Hurrengo taulako xehetasun guztiak behatzen baditugu, nahiz eta dibortzietuek oro har gogobetetze-maila altuagoa eduki, bi faktoreko beste kategoria guztietan joerak pairatzen duen aldagetaz jabetuko gara. Gizonen artean ezkongabeek bizitzarekiko gogobetetze-maila baxuena duten bitartean, emakumeen artean alargunak dira hori betetzen dutenak. Lagina Egoerazibila-ko kategorien arabera zatikatzean gizon eta emakumeen batezbestekoak berdinak izateari uzten diote. Azaldutako aldagaiaren portaera diferentzial honi interakzioaren efektua deitzen zaio.



108. irudia.

Batezbestekoen deskribapena osatzeko, R Commander-eko irteerak bi faktoreen kategoriak gurutzatuz atera diren zortzi taldeen desbideratze tipikoak eta laginaren tamainak ere aurkezten ditu.

Ventana de resultados Ejecutar

```
> tapply(Inkesta$Bizitza1, list(Egoerazibila=Inkesta$Egoerazibila,
+   Sexua=Inkesta$Sexua), sd, na.rm=TRUE) # std. deviations
```

	Sexua	
Egoerazibila	Emakumea	Gizona
Alarguna	14.06190	11.557735
Dibortziatua	12.57346	9.085043
Ezkondua	13.97301	12.190792
Ezkongabea	16.20732	6.327064

Desbideratze tipikoak

```
> tapply(Inkesta$Bizitza1, list(Egoerazibila=Inkesta$Egoerazibila,
+   Sexua=Inkesta$Sexua), function(x) sum(!is.na(x))) # counts
```

	Sexua	
Egoerazibila	Emakumea	Gizona
Alarguna	29	15
Dibortziatua	5	5
Ezkondua	45	25
Ezkongabea	14	12

Gurutzatutako kategoria bakoitzeko laginaren tamaina

109. irudia.

Irteera hasieran ageri den atalak `Bizitza1` aldagaiaren bariantzaren analisia aurkezten du. Faktore bakoitzaren eta interakzioaren balioen karratuen batuketaren emaitza (Sum Sq), askatasun-graduak eta laginaren F balioak erakusten dute. Eskuineko zutabean, hasierako hipotesizat populazioaren batezbestekoen arteko berdintasuna eta interakzio eza hartuta, laginean F balio horietako bakoitza topatzeko daukagun probabilitatea aurkituko dugu. Probabilitate horiek hasierako hipotesiarekin bateragarriak dira, ez daukagu hori baztertzeko adina ebidentzia. Irakurleak ikus dezakeen moduan, lagina bi faktoreren arabera zatitzean hainbat talderen parte-hartzaile kopurua dexente jaitsi da (5 gutxiago kasu batzuetan); eta aldi berean, kategoria bakoitzaren barneko desbideratze tipikoa (taldeen barneko distantzia) handia da. Aurkitutako desberdintasunak hasierako hipotesia baztertzeko nahiko ez izateari laguntzen diote aipatutako bi zirkunstantziek.

```
> Anova(AnovaModel.5)
Anova Table (Type II tests)
```

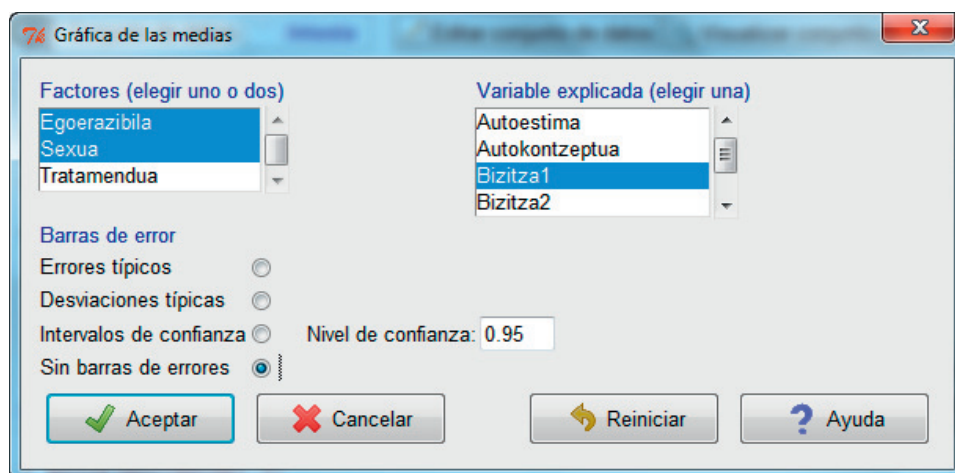
Response: Bizitza1

	Sum Sq	Df	F value	Pr(>F)
Egoerazibila	355.6	3	0.6904	0.5593
Sexua	0.1	1	0.0006	0.9801
Egoerazibila:Sexua	448.6	3	0.8708	0.4579
Residuals	24382.0	142		

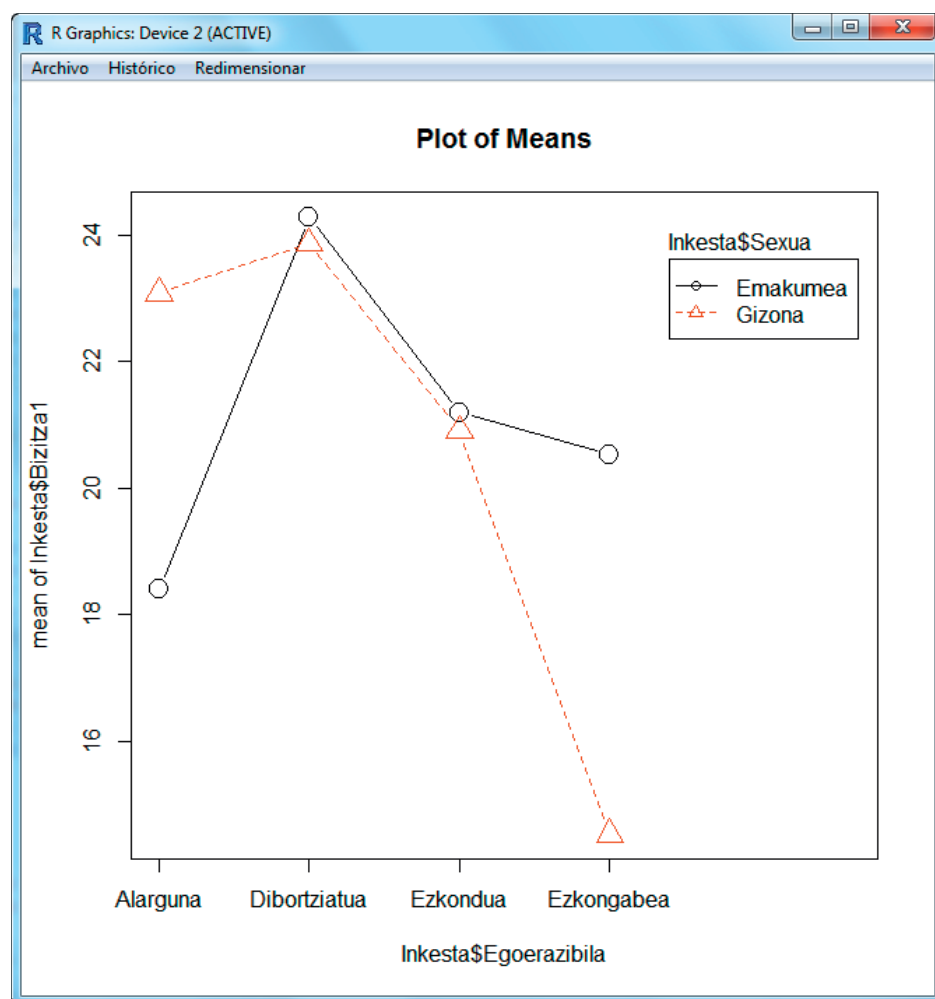
Diagrama: Irudiaren gainean hiru burbuila gorri daude. Lehenburua, 'Estatistikoak eta p: Egoerazibila' deitzen du, eta 'Egoerazibila' erregistroa eta bere 'Pr(>F)' balioa (0.5593) lotzen ditu. Bigarrena, 'Estatistikoak eta p: Sexua' deitzen du, eta 'Sexua' erregistroa eta bere 'Pr(>F)' balioa (0.9801) lotzen ditu. Hirugarrena, 'Estatistikoak eta p: interakzioa' deitzen du, eta 'Egoerazibila:Sexua' erregistroa eta bere 'Pr(>F)' balioa (0.4579) lotzen ditu.

110. irudia.

Nahiz eta grafikoak sortzeko prozesua hurrengo kapituluan aztertuko dugun, aurreko analisia batezbestekoen grafikoarekin osa genezakeela aipatu nahi genuke. Grafiko horrekin aztertutako faktoreen arteko interakzioa ikus baikenezake. Grafikoa lortzeko bidea Gráficas > Gráfica de las medias... izango litzateke.



111. irudia.

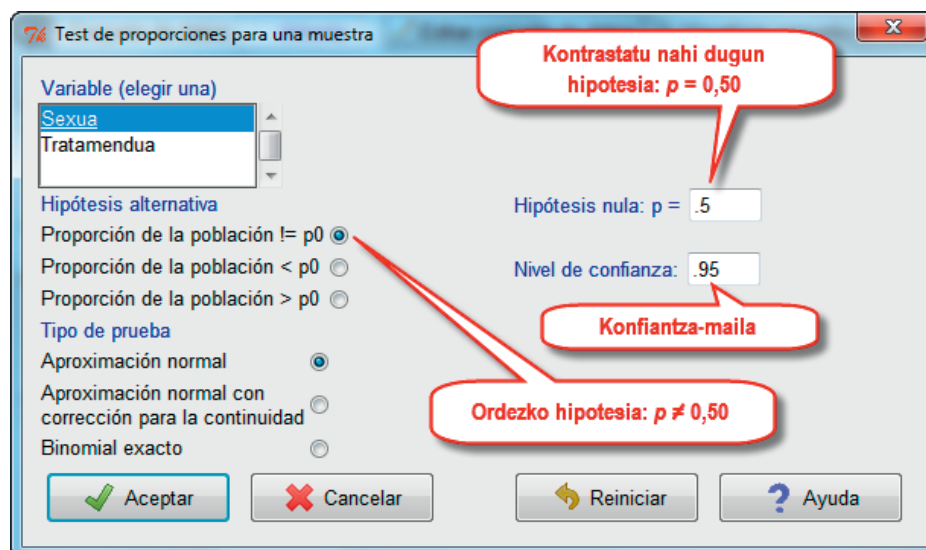


8. grafikoa. Bizitza1 aldagaiaren batezbestekoa Egoerazibila eta Sexua aldagaien gurutzaketa bakoitzean.

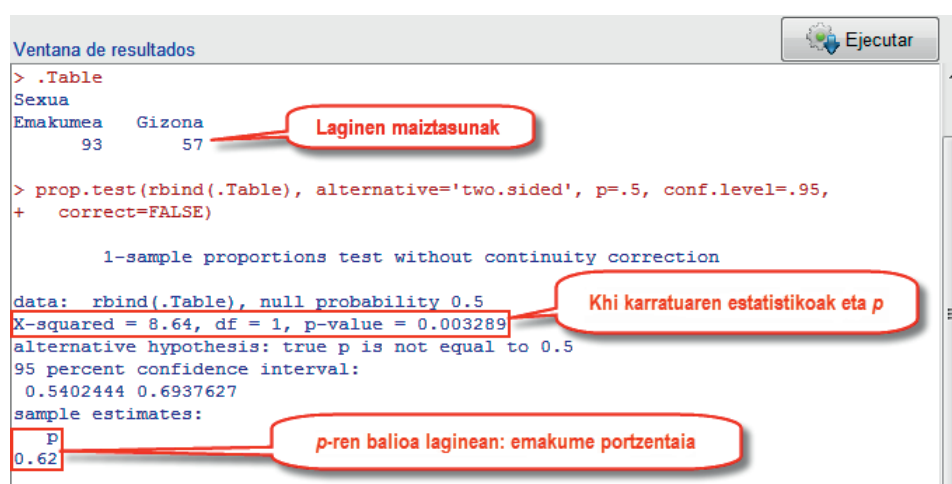
10.4. PROPORZIOAK

Populazio batetik zoriz 93 emakume eta 57 gizon hautatzen baditugu adierazgarria den lagina ($N = 150$) sortuz, esan al dezakegu populazio horretan emakumezkoak gizonezkoak baino gehiago direla? Izan al lieke zoriaren ondorio desberdintasun hori? Txanpon bat 150 aldiz jaurtitzen badugu 93 aurpegi eta 57 gurutze eskuratuz, txanpona gaizki egin dagoela (aurpegia eskuratzeko probalitatea ez da 0,50) pentsatu behar al genuke? Izan al liteke eskuratutako emaitza zoriaren ondorio eta 0,50eko probabilitatearekiko bateragarria? Azpimenu honetan galdera horiek aztertu ahal izango ditugu. Horretarako Estadísticos > Proporciones > Test de proporciones para una muestra... sakatuko dugu.

Lehenengo galderari dagokion analisia eskatu dugu. Bateragarriak al dira zoriz hautatutako 93 emakumek eta 57 gizonen (% 62 emakume) osatutako lagina eta erdiak gizonak eta beste erdiak emakumeak dituen populazioa?



112. irudia.



113. irudia.

R Commander-ek eskainitako irteerak behatutako maiztasunak (93 emakume eta 57 gizon) berdintasunaren hasierako hipotesiaren arabera espero diren maiztasunekin konparatzen ditu eta khi karratu estatistikoak laginean duen balioa (8.64) kalkulatzen du. Laginean balio hori eskuratzeko probabilitatea 0.003289 da. % 5eko errore-mailarekin lanean ari garenez baztertu egingo dugu hasierako hipotesia.

R Commander-ek lagineko estatistikotik abiatuz p proportzioari ($p(\text{Emakumea}) = 0.62$) dagokion konfiantza-tartea ere zenbatesten du. % 95eko konfiantza-mailarekin eskuratutako konfiantza-tartea $[0.5402444, 0.6937327]$ da. Erabakia hartzeko prozesua oso sinplea da. Hipotesiak jasotzen duen proportzioa ($p = 0.50$) konfiantza-tarteko mugen barruan edo kanpoan dagoen ikusi behar dugu. Kasu honetan, khi karratu estatistikoarekin hartu dugun erabakiarekin bat dator, baztertu egingo dugu hasierako hipotesia.

Hurrengo adibidean, aurreko egoera orokortuko dugu. Horretarako, bi tratamendu esperimental dauzkagula eta tratamendu bakoitzean parte hartu duen emakume (edo gizon) portzentajea konparatu nahi dugula pentsatuko dugu. Hasteko, Estadísticos > Tabla de contingencia > Tabla de doble entrada... bidez lortu dugun taula aztertuko dugu.



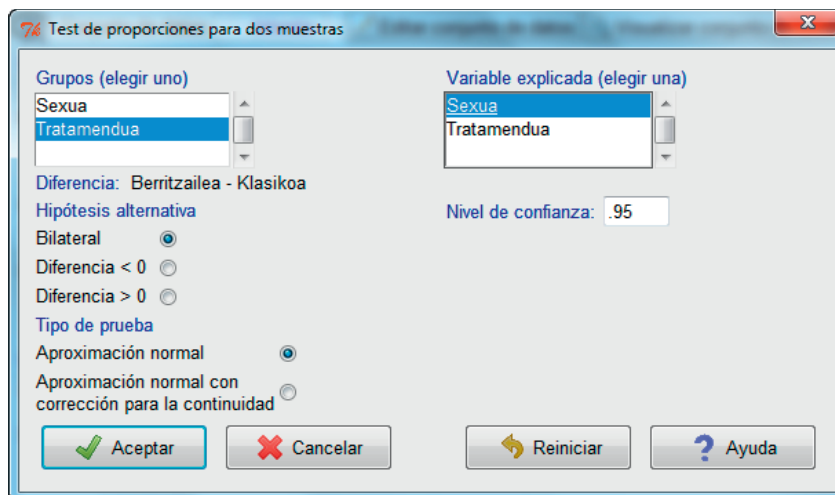
Ventana de resultados

```
> .Table
```

	Tratamendua	
Sexua	Berritzailea	Klasikoa
Emakumea	48	45
Gizona	27	30

114. irudia.

Tratamendu klasikoa esleitu zaion taldean % 60 emakumeak dira ($45/(45+30)$), eta tratamendu berritzailearen taldean emakumeen proportzioa % 64 da ($48/(48+27)$). Bateragarria al da bi taldeen arteko desberdintasuna desberdintasunik ez dagoen errealitate batekin? Analisari ekiteko Estadísticos > Proporciones > Test de proporciones para dos muestras... sakatuko dugu.



7% Test de proporciones para dos muestras

Grupos (elegir uno):

Variable explicada (elegir una):

Diferencia: Berritzailea - Klasikoa

Hipótesis alternativa:
☒ Bilateral
☐ Diferencia < 0
☐ Diferencia > 0

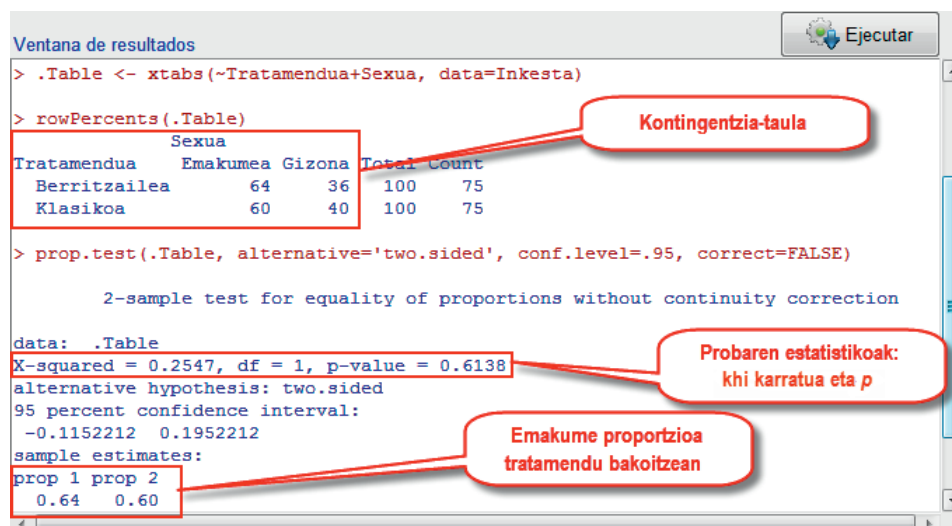
Tipo de prueba:
☒ Aproximación normal
☐ Aproximación normal con corrección para la continuidad

Nivel de confianza: .95

Aceptar Cancelar Reiniciar Ayuda

115. irudia.

Taldeak finkatzen dituen aldagaia Tratamendua da. Sexua aldagaiaren proportziak dira konparatu nahi ditugunak (azaldutako aldagaia²). Aurreko adibidean bezala, R Commander-ek khi karratuaren emaitza eta KTa zenbatesten ditu. Bai batak bai besteak, ondorio berberera bultzatzen gaituzte. Ez daukagu hasierako hipotesia baztertzeko adina froga, diferentzia hori eskuratzeko probabilitatea 0.6138 baita eta % 95eko KTak 0 balioa jasotzen baitu.



116. irudia.

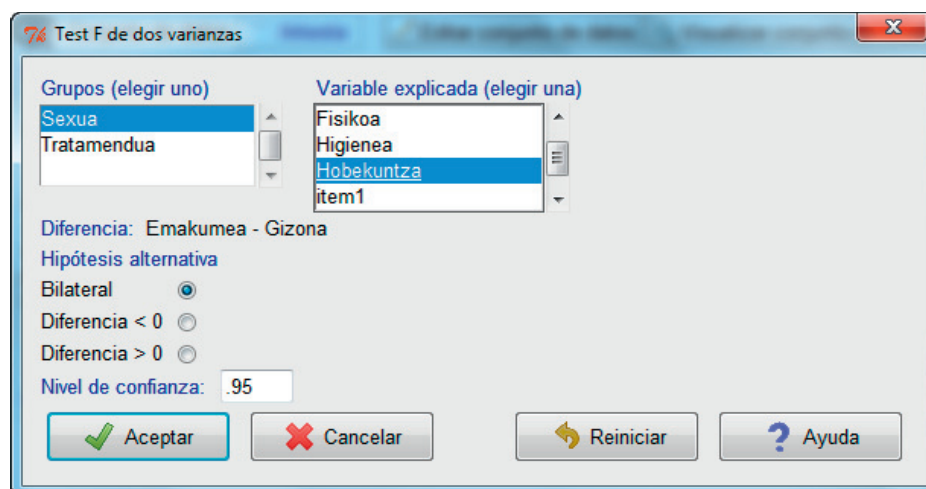
10.5. BARIANTZAK

Atal honetan burutuko dugun analisiak hainbat talderen bariantzak homogeneoak direla hipotetizatuko du. Analisi hau oso garrantzitsua da, oso erabiliak diren hainbat analisis (adibidez: lagin independenteentzako Studenten t testa) bariantzak berdintzat edo desberdintzat hartuko diren erabakitzea eskatzen baitute. Gainera, kasu zehatz batzuetan beharrezkoa izan liteke populazioaren bateko kolektibo desberdinetan aldagai baten aldagarritasuna alderatzea.

R Commander-ek hiru analisi burutzeko aukera ematen digu. Horien artean F testa bi talde ditugunean erabiltzeko egokia da. Beste bi testak, Bartlettrena eta Levenerenena, bi talde baino gehiago ditugunean erabiltzeko diseinatu ziren. Hirurak antzeko moduan gauzatzen eta interpretatzen dira. Hortaz, F testarentzat soilik prestatu dugu adibidea, erabiliena baita. Aukera-sekuentzia honi jarraituko diogu: Estadísticos > Varianzas > Test F para dos varianzas...

2. Proposatutako lan-hipotesiaren ikuspuntutik aldagaien ordena garrantzitsua da. Kasu batean tratamendu bakoitzean dagoen gizon eta emakumeen portzentajeen arteko berdintasuna hipotetizatuko baitugu; eta bestean sexu bakoitzak jaso duen tratamendu klasiko eta berritzaileen portzentajeen arteko berdintasuna.

Hurrengo irudian F testaren bidez gizon eta emakumeen bariantzak Hobekuntza aldagaiaren konparatuko ditugu. Bariantzen arteko proportzioen konfiantza-tartea % 95ean zenbatetsiko da.



117. irudia.

Irteeraren goiko aldean talde bakoitzaren bariantza aurkituko dugu. Emakumeen taldean 369.21 da, eta gizonen taldean 306.92. R Commander-ek bi bariantzen arteko proportzioa kalkulatu du (1.202937), 1 balioa jasotzen duen K^2 aurkezten du (bariantza arteko 1 proportzioak bariantzak berdinak direla esan nahi du), eta bariantzak berdinak balira laginean F balioa ($F = 1.2029$) zoriz eskuratzeko probabilitatea ($p = 0.4578$) adierazten du. Emaitzek bariantzen arteko berdintasuna onartzerantz bultzatzen gaituzte. Hortaz, lagin independenteentzat t testa (ikusi 10.3.2. atala) erabiliz bi taldeen batezbestekoak konparatu nahi baditugu, bariantzak berdinak direla pentsatuko dugu.

```

> tapply(Inkesta$Hobekuntza, Inkesta$Sexua, var, na.rm=TRUE)
Emakumea  Gizona
369.2132  306.9264

> var.test(Hobekuntza ~ Sexua, alternative='two.sided', conf.level=.95,
+ data=Inkesta)

F test to compare two variances

data: Hobekuntza by Sexua
F = 1.2029, num df = 92, denom df = 56, p-value = 0.4578
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7387028 1.9056001
sample estimates:
ratio of variances
      1.202937

```

Kategoria bakoitzaren bariantza

Kontrastearen estatistikoak eta p

Bariantzen arteko proportzioaren konfiantza-tartea

Bariantzen arteko proportzioa

118. irudia.

10.6. TEST EZ-PARAMETRIKOAK

Hipotesien testek populazioarekiko adierazgarria den laginean hasierako hipotesiarekin bat datozen emaitzak zoriz lortzeko probabilitatea kalkulatzeko dute. Proba parametrikotik probabilitatearen kalkulua «ezagunak» diren probabilitate-banaketatik oinarritzen dira; besteak beste, banaketa normala, t , F ... Kasu batzuetan, ordea, bai neurketa-maila tartetakoak ez delako, bai test parametrikoei eskatutako baldintzak betetzen ez direlako, gomendagarria da proba ez-parametrikoei lan egitea. Tejedor eta Etcheberriaren (2006)³ lanean, irakurleak test ez-parametrikoei erabilienean testuinguru teoriko eta praktikoa aurki dezake. Proba ez-parametrikoei:

- Parametroak baino errazago uler daitezke.
- Aurrebaldintza malguagoak dituzte.
- Aurrebaldintzen malgutasunaren ondorioz potentzia gutxiago dute.

Hori dela-eta (oro har) nahiago da proba parametrikoei lan egin. Baina berauei eskatutako baldintzak betetzen ez direnean, beharrezkoa izan liteke proba EZ-parametrikotara jotzea.

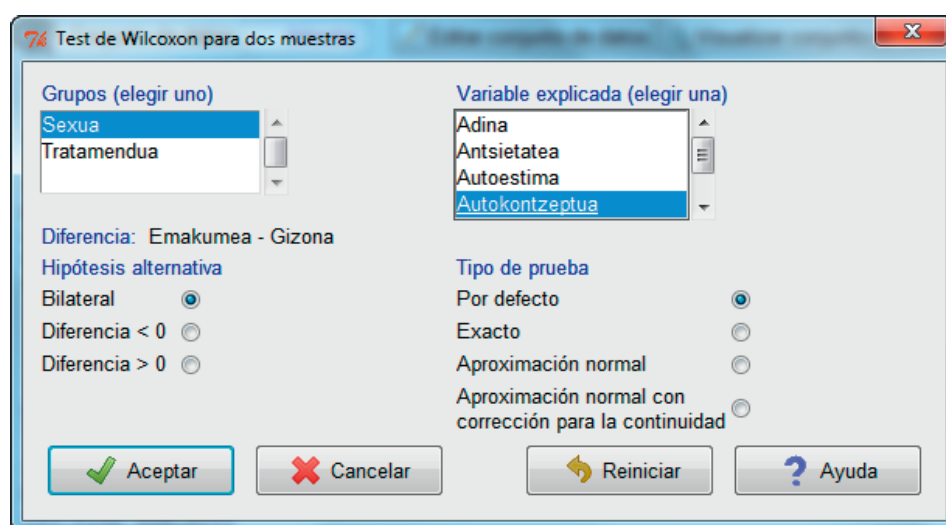
3. Tejedor, F.J. eta Etcheberria, J. (2006). *Análisis Inferencial de Datos en Educación*, La Muralla, Madril.

R Commander-ek Test no paramétricos azpimenuko aukeren bidez lau test ez-parametrikoko erabiltzeko aukera eskaintzen du:

- Bi talde independente – Wilcoxon-en testa.
- Erlazionatutako edo parekatutako bi talde – Wilcoxon-en testa.
- K talde independente – Kruskal Wallisen testa.
- Erlazionatutako edo parekatutako K talde – Friedmanen testa.

10.6.1. Wilcoxon-en testa (bi talde independente)

Proba honek zenbakizko aldagai baten balio zentralak bi taldetan konparatzeko aukera ematen du. Analisia burutzeko jarraituko beharreko pausoak Estadísticos > Test no paramétricos > Test de Wilcoxon para dos muestras... dira.



119. irudia.

Emaitzek aztertutako kategoria bakoitzeko (Emakumea eta Gizona) medianaren balioak adierazten dituzte. Bi kasuetan balioa berdina da, 15. Jarraian R Commander-ek W estatistikoaren balioa ($W = 2759.5$) eta hasierako hipotesiaren baldintzapean balio hori laginean eskuratzeko probabilitatea ($p = 0.6734$) aurkezten du. Emaitza horien aurrean laginak batezbesteko berdinak dituzten populazioetatik datozela baztertzeko ebidentziarik ez daukagula ondoriozta dezakegu.

```

> tapply(Inkesta$Autokontzeptua, Inkesta$Sexua, median, na.rm=TRUE)
Emakumea  Gizona
      15      15

> wilcox.test(Autokontzeptua ~ Sexua, alternative="two.sided", data=Inkesta)

Wilcoxon rank sum test with continuity correction

data:  Autokontzeptua by Sexua
W = 2759.5, p-value = 0.6734
alternative hypothesis: true location shift is not equal to 0

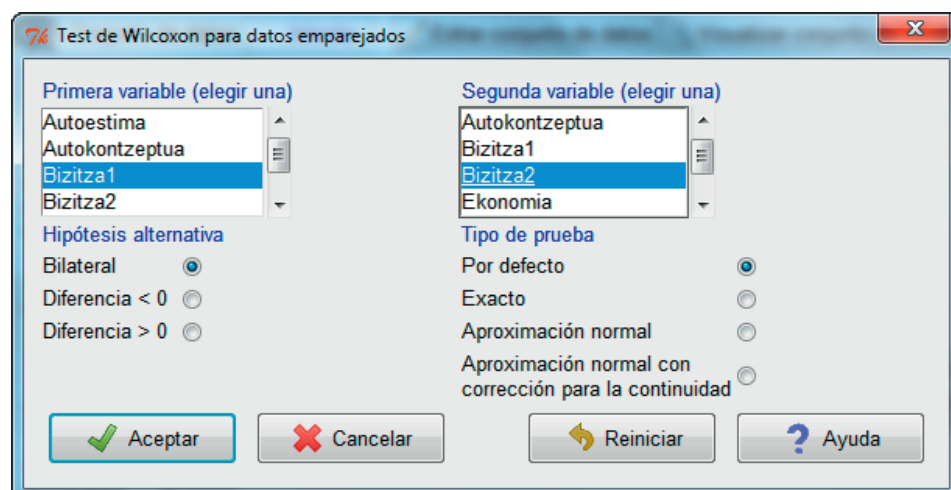
```

120. irudia.

10.6.2. Parekatutako taldeentzako Wilcoxon-en testa

Parekatutako diseinu (*pre-post*) batekin lanean ari bagara eta joera zentraleko neurriak alderatu nahi baditugu, analisi hau egokia da. Estadísticos > Test no paramétricos > Test de Wilcoxon para muestras pareadas... bidez eskatuko diogu R Commander-i.

Adibide honetan, programa edo tratamendu bat aplikatu AURRETIK (Bizitza1) eta ONDOREN (Bizitza2) eskuratutako bizitzarekiko gogobetetze-maila alderatzeko eskatuko dugu.



121. irudia.

Emaitzek bi aldagaien medianen arteko desberdintasuna (-29.7) eta diferentzia positiboen eta negatiboen rangoen gehiketaren balio baxuenari dagokion lagineko estatistikoaren balioa ($V = 204$) aurkezten dituzte. Azkenik, idazketa esponentzialean bi medianak berdinak izatekotan V estatistikoaren balioa aurkitzeko probabilitatea irakur daiteke. Adibide honetan, desberdintasuna esanguratsua dela ondorioztatuko dugu.

The screenshot shows the 'Ventana de resultados' window in R Commander. It displays the output of two R commands. The first command calculates the median difference between two groups, resulting in -29.7 . The second command performs a Wilcoxon signed rank test, showing a test statistic $V = 204$ and a p-value $< 2.2e-16$. Red callouts highlight these key results.

```
> median(Inkesta$Bizitza1 - Inkesta$Bizitza2, na.rm=TRUE) # median difference
[1] -29.7

> wilcox.test(Inkesta$Bizitza1, Inkesta$Bizitza2, paired=TRUE,
+             data.name='Inkesta$Bizitza1 and Inkesta$Bizitza2')

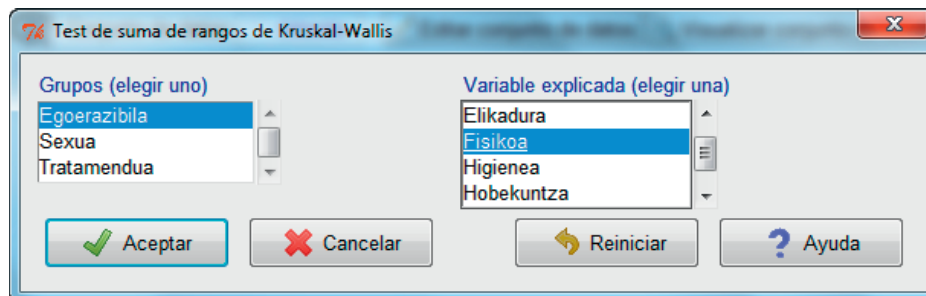
Wilcoxon signed rank test with continuity correction

data: Inkesta$Bizitza1 and Inkesta$Bizitza2
V = 204, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

122. irudia.

10.6.3. Kruskal-Wallisen testa (k talde independente)

Kruskal-Wallisen testa faktore bakarreko bariantza-analisiaren baliokidea da; hasierako hipotesiak jasotzen du lagina banatu den kategorietan neurketa-aldagaiaren mediana berdina dela. Gure kasuan, pertsonen sendotasun fisikoa (Fisikoa) egoera zibilaren arabera eskuratu ditugun kategoria guztietan berdina al den ikusi nahi dugu.



123. irudia.

Neurketa aldagaian taldeak homogeneoak badira eta aldagaiko balio guztiak ordenatzen baditugu txikienari 1 balioa esleituz eta handienari n (parte-hartzaile kopurua), talde bakoitzeko parte-hartzaileen heinen batezbestekoak berdinak edo,

gutxienez, antzekoak izan beharko lirateke. Parte-hartzaileei ordenaren arabera esleitutako balioen ondorioz sortutako taldeen arteko desberdintasunetan oinarrituz, lagineko Kruskal Wallisen balioa eskuratzen da; honek khi karratuaren banaketari jarraitzen dio. Hemendik abiatuz, hasierako hipotesia baieztatzekotan lagineko balio hori lortzeko probabilitatea izango litzatekeena kalkulatzeko da. Emaitzek kategoria guztietako mediana berdina dela adierazten dute (unitate bat handiagoa dibortziatuetan). Kruskal Wallisen H estatistikoaren balioa 4.4256 da, eta hiru askatasun-graduko banaketari jarraitzen dio. Hasierako hipotesia (talde guztietan mediana berbera) baieztatuko balitz, estatistiko hau topatzeko probabilitatea 0.219 izango litzateke. Hortaz, ezin dugu hasierako hipotesia baztertu.

```
> tapply(Inkesta$Fisikoa, Inkesta$Egoerazibila, median, na.rm=TRUE)
```

Alarguna Dibortziatua	Ezkondua	Ezkongabea
6	7	6

```
> kruskal.test(Fisikoa ~ Egoerazibila, data=Inkesta)
```

```

Kruskal-Wallis rank sum test

data: Fisikoa by Egoerazibila
Kruskal-Wallis chi-squared = 4.4256, df = 3, p-value = 0.219

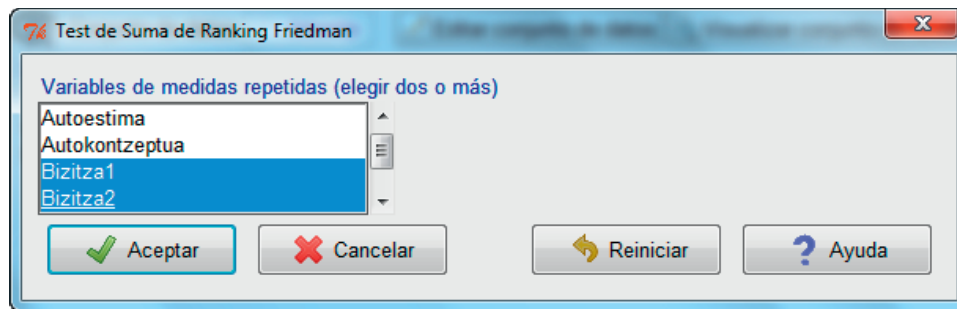
```

124. irudia.

10.6.4. Friedmanen testa (parekatutako k talde)

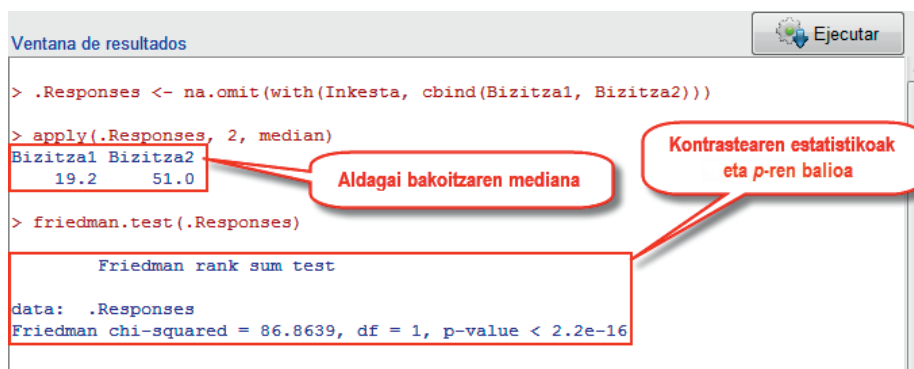
Denboran zehar edo esku-hartze prozesu baten fase zehatz batzuetan ezaugarri baten garapena aztertu nahi bagenu, erabilgarria izan liteke neurketa errepikatuen diseinu batekin lan egitea. Bariantzaren analisiak tarteen eskala batean neurtutako aldagai batean ondoz ondoko emaitzen batezbestekoak konparatzeko aldaera du. Baina neurketa-aldagaiak ez baditu ANOVAk eskatutako aurrebaldintzak betetzen, —esaterako, eskala ordinalean neurtua izan delako—, egokia (eta kasu batzuetan ezinbestekoa) izan liteke proba ez-parametrikoko erabiltzea. Horien artean erabiliena Friedmanen testa da.

Analisia adibide baten bidez erakusteko, ikerketako bi momentuetan bizitzarekiko gogobetetze-mailan eskuratutako emaitzak konparatuko ditugu eta desberdintasunak bi aldagaien medianen berdintasuna jasotzen duen hasierako hipotesiarekin bateragarria ote den ikusiko dugu (normalean, proba hau hiru neurketa errepikaturik edo gehiagori dagozkien aldagaiak konparatzeko erabili ohi da). Estadísticos > Test no paramétricos > Test de Suma de Rangos de Friedman... sakatuko dugu.



125. irudia.

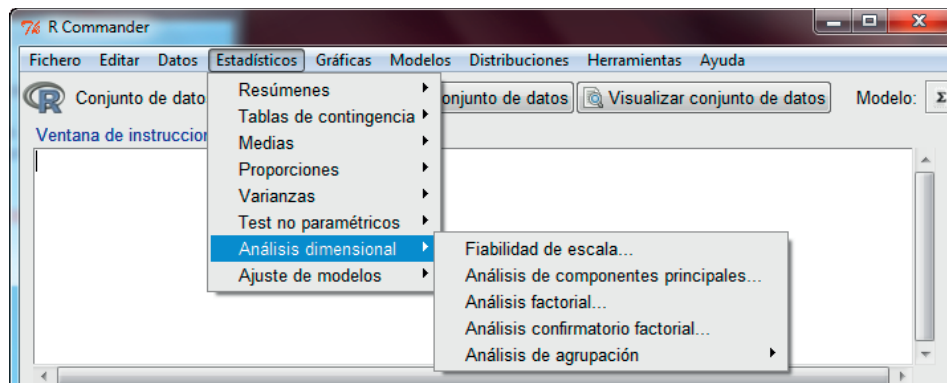
Emaitzen leihoan aldagai bakoitzaren medianaren balioak (19.1 eta 51.0) eta Friedmanen estatistikoaren balioa laginean (86.8639) irakur daitezke. Askatasun-gradu bateko (aldagai kopurua – 1) khi karratuaren banaketari jarraituko lioke estatistikoak. Estatistiko hori laginean eskuratzeko probabilitatea (0,05 baino gutxiago) oso baxua da. Beraz, medianen berdintasuna jasotzen duen hasierako hipotesia baztertuko dugu.



126. irudia.

10.7. DIMENTSIONALITATEA

Portaera-eskalen, nortasunaren inguruko galde-sorten edo zenbakizko gaitasunaren proben ikerketak datuen fidagarritasuna eta horietatik eratortzen diren erabakien baliagarritasuna bermatzera zuzentzen den analisi-multzoa barneratzen du. Fidagarritasuna zehaztasunarekin edo proben barne-kontsistentziarekin erlazionatzen den kontzeptua da, eta baliagarritasuna eskuratutako puntuazioen erabilera abalatzten duten ebidentzien azterketan oinarritzen da (Muñiz, 1996; Martinez Arias, 1995; Elosua, 2003). Análisis Dimensional azpimenuaren bitartez, R Commander-ek fidagarritasunaren eta barne-baliagarritasunaren azterketarako (dimentsionalitatea) teknika erabilienak eskaintzen dizkio ikertzaileari.



127. irudia.

10.7.1. Eskalaren fidagarritasuna

Objektuen propietateei zenbakizko balioak esleitzen dizkiogun bakoitzean neurketa-errorearen presentzia onartzen ari gara. Puntuazioetan duen eragina aztertzea da fidagarritasunaren azterketaren helburua. Puntuazioen zehaztasunaren neurria bailitzan interpretatzen da. Fidagarritasun-koefizientea zenbatesteko eskuragarri dauden prozedura guztien artean Cronbachen alfa da gailendu dena (1951; Elosua eta Zumbo, 2008). Alfa eskala bateko itemen arteko kobariantza-mailaren adierazle bat da; alegia, eskalaren barne-kontsistentziarena.

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_X^2} \right) = \alpha = \frac{n}{n-1} \left(\frac{\sum_{j=1}^n \sum_{\substack{i=1 \\ i \neq j}}^n Cov(i, j)}{\sigma_X^2} \right)$$

Non:

n : testeko item kopurua den.

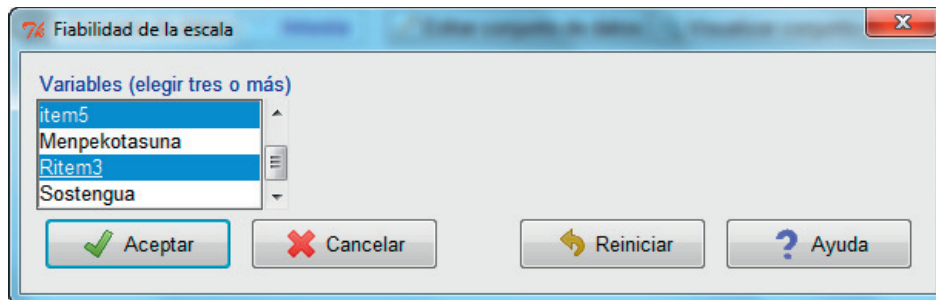
σ_i^2 : i itemaren bariantza den.

σ_X^2 : testaren bariantza den.

$Cov(i, j)$, i eta j arteko kobariantza den.

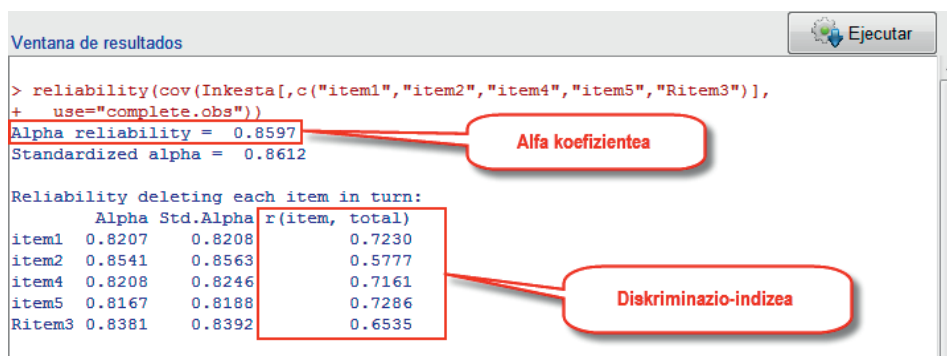
Cronbachen alfa koefizienteak eskala osatzen duten item guztien kontsistentzia-mailaren berri ematen du. Dena den, fidagarritasunaren azterketan oso interesgarria da item bakoitzaren funtzionamenduan sakontzea. Hori dela-eta, R Commander-ek diskriminazio-indize zuzendua (r_X) eskaintzen du. Nahiz eta diskriminazio-indizearen balio teorikoak $[-1, 1]$ tartean kokatzen diren, balio negatiboek itemen analisiaren testuinguruan ez dute zentzurik.

Estadísticos > Análisis Dimensional > Fiabilidad de escala... aukerekin aztertu nahi ditugun itemak hautatu ahal izateko leihoa irekiko da.



128. irudia.

R Commander-ek erakutsitako irteera bi ataletan zatitu daiteke. Lehenengoan Cronbachen alfa koefizientearen balioa irakurri daiteke (Alpha reliability); gure kasuan 0.8597. Jarraian, estandarizatutako aldagaiak (Standardized alfa) erabiliz lortuko genukeen estandarizatutako koefizientearen zenbatespena ageri da. Bi koefiziente horien artean lehenengoa da ohikoena.



129. irudia.

R Commander-ek sortzen duen balioen taulak itemen analisirako garrantzitsua den informazioa barneratzen du. Izan ere, itemek barne-kontsistentzian duten eraginaren berri ematen du eta item bakoitzaren diskriminazio-indizearen balioak jasotzen ditu. Taulako lehen zutabeaz aztertutako itemen balioak daude. Alpha izena duen zutabeak dagokion lerroko itema ezabatuz gero eskalako alfa zein izango litzatekeen adierazten du. Adibidez, lehen balioa (0.8207) lau itemez osaturiko eskalaren alfa koefizientearen balioa da. Std. Alfa izeneko zutabeaz estandarizatutako eskalan zenbatetsitako alfa koefizienteak kokatzen dira. Azken zutabeak lerroko itemaren eta estandarizatutako eskalaren arteko korrelazioaren berri

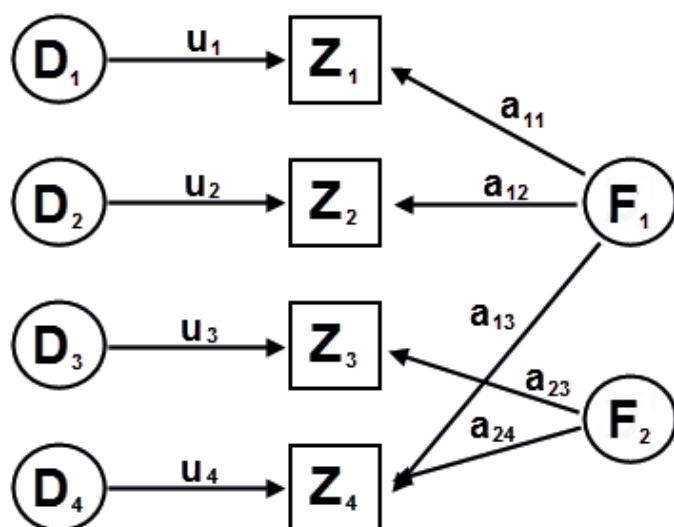
ematen du; alegia, zuzendutako diskriminazio-indizeen balioa ($r(\text{item}, \text{total})$). Antzeman daitekeenez, estandarizatutako eskalaren alfa balioen eta diskriminazio-indizeari dagozkion balioen arteko erlazioa alderantzizkoa da. Item bat ezabatzean alfaren balioa nabarmenki txikitzen bada, item horren diskriminazio-indizea altua da.

10.7.2. Faktore-analisia

Gizarte Zientzientzako ikerketaren helburu diren konstruktutako teorikoak ezin dira zuzenean neurtu; hori dela-eta, beharrezkoa da ageriko indikatzaileak erabiltzea. Konstruktutako teorikoaren eta ageriko indikatzaileen arteko erlazioak definitzen du neurketaren baliagarritasuna (Elosua, 2003). Ageriko eta ezkutuko aldagaien erlazioaren azterketa enpirikoaz arduratzen diren eredu formalen artean faktore-analisia da Gizarte Zientzietan erabiliiena.

Faktore-analisia eskalaren barne-egitura edo dimentsionalitatea aztertzeko teknika multzoa da. Helburutzat du ageriko aldagai multzo (itema, eskala partziala edo eskala) batek (n) jasotzen duen informazioa faktore edo dimentsio gutxiagotan ($m < n$) laburtzea, horien bidez datuetan aurkitutako aldakortasuna azaltzeko.

Aztertuko ditugun datuak aldagai arteko kobariantzen (edo korrelazioen) matrizean daude. Analisi faktorialaren helburuak ebazpenaren partsimonia eta sinpletasuna dira; modu horretan bermatzen baitira ateratako faktoreen esanguratsutasuna eta interpretagarritasuna.



9. grafikoa. Bi faktoreko eredu.

Eredu faktoriala osatzen duten elementuak hurrengoak dira:

- *Behatutako aldagaiak.*
- *Ezkatuko aldagaiak.* Faktore ere deitu izan ohi zaie. Bi motatakoak izan daitezke:
 - *Faktore komunak.* Behatutako aldagai batekin baino gehiagorekin erlazionatzen diren egiturak dira.
 - *Faktore bakarrek.* Behatutako aldagai bakoitzarekin erlazionatzen diren ezkatuko egiturak dira. Faktore bakarren kopurua n da.
- *Pisuak edo saturazioak.* Aldagaiak faktoreetan duten saturazioak dira. Faktoreen arabera bi mota bereiz daitezke:
 - a_{ij} . Faktore komunak pisuak dira.
 - u_i . Faktore bakarren pisuak dira.

Eredu faktorialaren ekuazio orokorra honela idatz daiteke:

$$Z_i = a_{1i}F_1 + a_{2i}F_2 + \dots + a_{mi}F_m + u_iD_i = \sum_{j=1}^m a_{ji}F_j + u_iD_i$$

5. formula.

Analisi faktorialaren ereduak hurrengo ezaugarriak ditu:

- Faktoreak eta behatutako aldagaiak estandarizatuta daude.
- Faktore bakarren eta komunak arteko erlazioa nulua da.
- Faktore espezifikoak estatistikoki independenteak dira.
- Faktore komunak estatistikoki independenteak dira.

Baldintza hauetatik hurrengo ondorioak eratortzen dira:

- Behatutako aldagaien bariantza komuntasunean eta bakartasunean zatitzen da.
- Behatutako bi aldagairen arteko korrelazioaren balioa komunean dituzten faktore bakoitzeko saturazioak biderkatuz eta hortik lortutako balio guztiak batuz lortzen da.
- Behatutako aldagai baten eta faktore baten arteko korrelazioa aldagai horrek faktorean duen pisua da.

10.7.2.a. Erauzketa eta biraketa

Faktoreen erauzketa faktore-analisiko nukleoa da. Asko dira horretarako sortuak izan diren prozedura aljebraiko eta estatistikoak. Guztien aurkezpenak gidaliburu honen mugak gainditzen ditu. Horien guztien artean, R Commander-ek egiantza handieneko zenbatespena erabiltzen du.

Datu multzo bati faktore-analisia aplikatzean sortutako ebazpen faktoriala ez da bakarra. Posibleak diren ebazpen guztien artean ikuspuntu sustantibo batetik interpretagarriena dena aukeratu beharko litzateke; egitura hori Thurstonek (1947) definitutako egitura sinplera gerturatu ohi da. Haren arabera:

- Aldagai bakoitzeko bariantzari ekarpenik egiten ez dion faktore bat dago gutxienez.
- Faktore bakoitzean pisu nulua (edo 0tik gertu dagoen balioa) duen aldagai multzo bat dago.
- Zutabe pare bakoitzeko, batean 0tik gertu dauden balioak eta bestean balio altuak dituzten aldagaiak egongo dira.

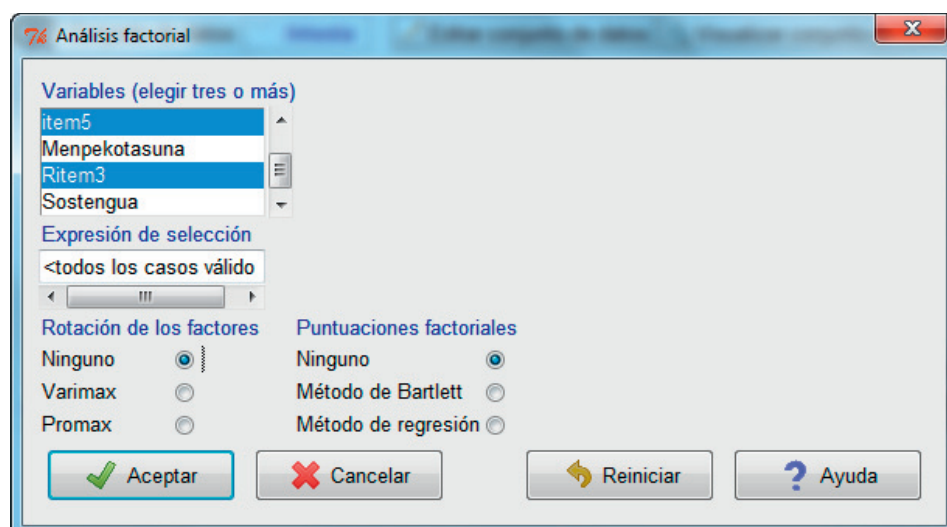
Faktore-biraketak egitura sinplera gerturatzen den faktore-ebazpena bilatzen du. Biraketak ez du eraginik ebazpenaren kalitate formalean. Bi ebazpenen (biratutakoa eta biratu gabekoa) komuntasuna baliokidea da. Biraketak faktore bakoitzarekin lotzen den karga komunari eragiten dien faktoreen pisuak berrantolatzen ditu. Faktore-biraketaren prozedurak bi taldetan banatzen dira: prozedura ortogonalak eta prozedura zeiharrek. Lehenengo taldeko prozedurek (esaterako *varimax*) faktoreen arteko independentzia mantentzen dute, eta bigarren taldekoek (*promax*) korrelazionatuta dauden faktoreak zenbatesten dituzte.

10.7.2.b. Faktoreen puntuazioen zenbatespena

Ikerketa-testuinguru ugaritan, ikertzaileak ateratako faktoreetan parte-hartzai-leek eskuratutako puntuazioak zenbatesteko interesa izaten du. R Commander-ek faktoreetako puntuazioak kalkulatzeko Bartletten metodoa eta erregresioarena eskaintzen ditu. Lehenak, ponderatutako gutxieneko karratuen metodoa proposatu zuen. Metodo horrekin, bakartasunak zorizko terminotzat eta faktoreetako puntuazioak parametrotzat hartuta, behaketak dohitutako pisuen bidez itzultzen dira. Erregresioaren metodoan, faktoreak zorizko aldagaiak bailiran maneiatzen dira.

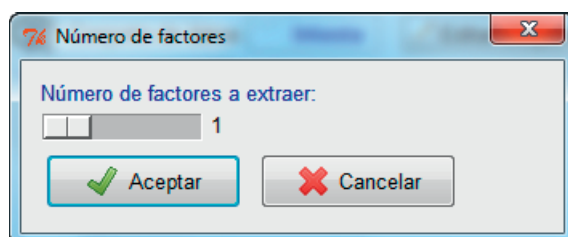
10.7.2.c. Faktore analisia eta R Commander

`Estadísticos > Análisis Dimensional > Análisis Factorial` sakatzean aztertzeraz goazen aldagaiak zeintzuk diren adierazteko aukera emango digun leihoa irekiko da. R Commander-ek faktore-analisia aktibo dagoen datu multzoari aplikatuko dio, edo ikertzailearen interesen arabera, baldintza zehatz bat betetzen duen azpimultzoari.



130. irudia.

Bizitza-kalitateari buruzko eskala bat aztertzen ari garenez, berau osatzen duten 5 itemak aukeratu ditugu. Oraingoz *Expresión de selección* aukera ez dugu markatuko; izan ere, lagin osoan ebaluatu nahi dugu eredua. Helburua datuen dimentsio-bakartasuna baieztatzea denez, faktoreen biraketaren aukerak ez dauka zentzurik eta dagoen bezala (*ninguno*) utziko dugu. Argudio guztiak zehaztu ostean, R Commander-ek atera nahi dugun faktore kopurua zehazteko eskatu digu. Guk 1 balioa finkatuko dugu.



131. irudia.

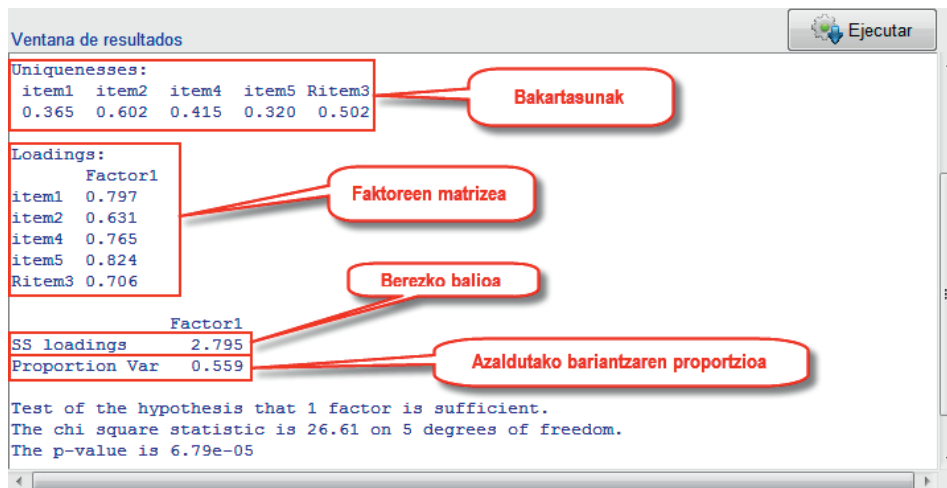
R Commander-ek sortutako emaitzek hiru atal ditu: bakartasuna (*Uniquenesses*), faktoreen pisuak (*Loadings*) eta datuek ereduarekiko erakusten duten egokitzapenari buruzko informazioa.

Bakartasuna aldagai bakoitzaren berezko bariantza da. Antzeman daitekeen moduan, aldagaien bakartasuna handiagoa den heinean, euren komuntasuna edo bariantza komuna txikiagoa da. Izan ere, aldagaien bariantza bakartasunak eta bariantza komunak osatzen dute, eta aldagaiak estandarizatuta daudenean bariantza horien batuketaren emaitza 1 da. Adibidez, lehenengo itemean: $1 = 0.365 + 0.797^2$. Elementuak bariantza totala (1), bakartasuna (0.365) eta bariantza komuna ($h^2 = 0.797^2$) dira. Faktore anitzeko ereduetan bariantza komuna (h^2) berretutako faktoreetako saturazioen batuketa da.

$$h^2 = a_1^2 + a_2^2 + \dots + a_k^2$$

6. formula.

Faktore bakarreko ereduan aldagaiek faktoreetan dituzten saturazioak faktoreen eta aldagaien arteko korrelazioak dira. Erregresio linealaren ikuspuntutik erregresio-koefiziente moduan ezagutuko lirateke. Gure adibidean, saturazio txikiena duena item2 da. Hortaz, aldagai guztien artean hau izango da bakartasun handiena erakutsiko duen itema.



132. irudia.

Erauzitako faktorearen berezko balioa (SS loadings) berretutako saturazioen batuketa da ($2.795 = 0.797^2 + 0.631^2 + 0.765^2 + 0.824^2 + 0.706^2$). Balio hori item kopuruarekin zatitzen badugu ($n = 5$), faktoreak azaltzen duen bariantzaren proportzioa ezagutuko dugu ($2.795/5 = 0.559$). Gure kasuan, erauzitako faktorea datuetan behatutako bariantza osoaren % 55,9 azaltzeko gai da.

Balio hori eskalaren dimentsio-bakartasun maila ebaluatzeko erabili izan da literatura psikometrikoan. % 55,9 Reckasek ezarritako mugaren (% 20) gainetik dago; baita % 40aren gainetik ere. Dimentsionalitatearen ebaluaziorako prozedura guztiek puntu horren azpitik ezartzen dituzte mugak, datu multzo bat dimentsio bakarrekotzat har dadin. Hortaz, aztertutako eskala osatzen duten bost itemen «atzean» bizitzakalitatearekin erlazionatzen den faktore nagusi bat dagoela ondorioztatuko dugu.

Azkenik, kontuan izanda sinegarritasun gehieneko zenbatespen-prozedura erabili dela eta hasierako hipotesizat eredua datuei doitzen zaiela hartuta, R Commander-ek khi karratu estatistikoa eskaintzen du. 5 askatasun-graduak banaketan 83.72 balioak hasierako hipotesiaren aurkako ebidentzia aurkezten du. Baina faktore bakarreko ereduan informazio hori ez da erabakigarria eta faktore-analisi baieztaizalean aurki ditzakegun indikatzaileekin osatu beharko litzateke.

Faktoreetako puntuazioak zenbatesteko aukeretako batetik jo izan bagenu, aktibatutako datu multzoari automatikoki erantsiko litzaiokeen aldagai berri bat sortuko litzateke.

10.7.3. Osagai nagusien analisia

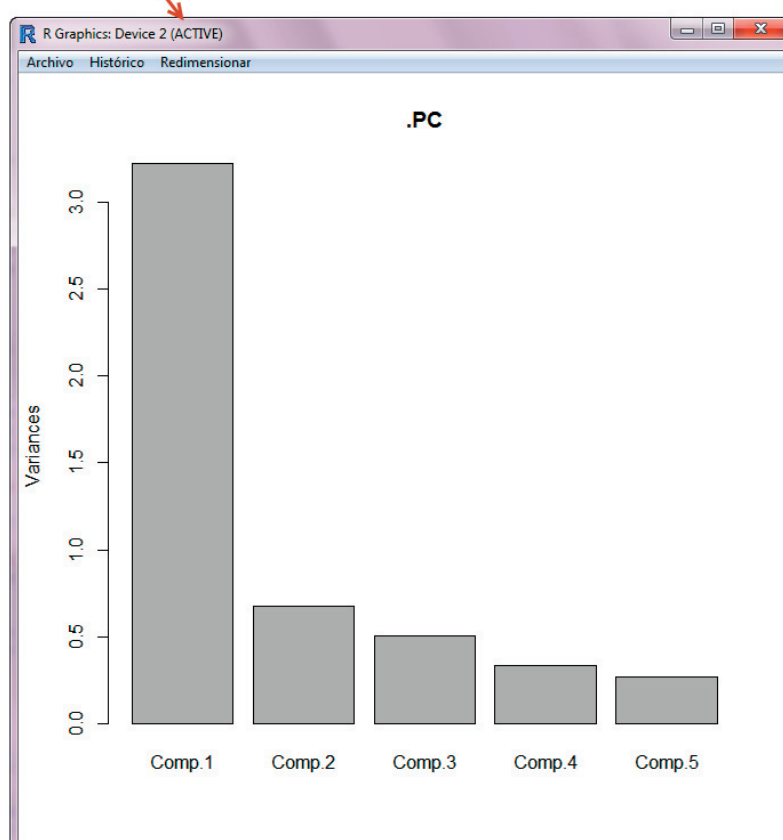
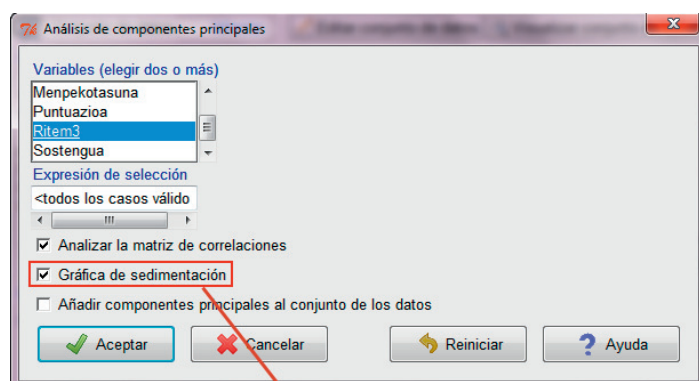
Nahiz eta ikuspuntu formaletik osagai nagusien eredua ez den faktore-analisia, aplikatutako ikerketan erabiltzea arrunta da. SPSS moduko lizentzia pribatiboak dituzten programek osagai nagusien analisia faktore-erazketa metodoa bailitzaan aurkezten dute. Osagai nagusien ereduen helburua bariantza osoa azaltzea da; faktore-analisiarena, berriz, bariantza komuna azaltzea. Helburuen arteko funtsezko desberdintasun horrek ereduen arteko desberdintasunak nabarmentzen ditu.

Egia da, ordea, osagai nagusien eredua faktore-erazketa metodo moduan erabiltzean emaitzak modu berean interpreta daitezkeela batean zein bestean. Bariantza komuna handiagoa den heinean, bi ereduen arteko baliokidetasuna ere handiagoa da.

Osagai nagusien ereduak behatutako n aldagaien konbinazio linealetatik abiatuz, behatutako bariantzaren gehiengoa azaltzen duten Y osagai definitzen ditu. Osagai bat zenbatetsi ostean, haren eragina kendu egiten da eta hondakinezko bariantzarekin bigarren osagaia zenbatesten da.

Osagai nagusien analisisan lehenengo osagaiak bariantza-portzentajea handia azaltzea interesatzen zaigu. Horrela, datuen egitura dimentsio gutxi batzuetara murriztu daiteke. `Estadísticos > Análisis Dimensional > Análisis de componentes principales` aukeren bidez R Commander-ek osagai nagusien analisia burutzeko duen funtzioari hel diezaiokegu. Aukera honekin irekiko zaigun leihoan korrelazio-matrizea aztertzeke, sedimentazio grafikoa sortzeke eta parte-hartzaileek erazitako faktoreetan eskuratutako puntuazioak gehitzeko aukerak topatuko ditugu. Dimentsiozko egitura aztertzean zenbat faktoreko ereduari

entsiko zaion erabakitzeko Cattellen screeplota (sedimentazio-grafikoa) da irudikapenezko prozedura erabilienetako bat.



133. irudia.

10.7.4. Taldekatze-analisia

Cluster moduan ere ezagutzen diren taldekatze-analisiak elementu multzo bat kategoria desberdinetan sailkatzea dute helburu. Kategoria horiei konglomeratu deritze gure testuinguruan. Konglomeratu bati dagozkion elementuak antzekoak dira euren artean (barne-homogeneotasuna) eta desberdinak beste konglomeratuetako elementuen aldean (konglomeratu desberdinen arteko heterogeneotasuna).

Haien parekotasun edo gertukotasunaren arabera, parte-hartzaileak taldekatzen hasi aurretik hainbat galderari erantzun behar zaie. Zer esan nahi du elementuak gertukoak edo antzekoak izateak? Zeintzuk izango dira antzekotasuna aztertzeko erabiliko diren aldagaiak? Nola kalkulatu da distantzia, antzekotasuna edo elementu baten eta talde baten artean dagoen distantzia? Eta taldeen artekoa? Ba al dakigu zenbat talde sortu nahi ditugun? Elementuak hainbat arrazoiengatik taldekatu daitezke: taldeak deskribatzeko, esploratzeko edo baieztatzeak. Baina edozein kasutan, prozedura gauzatzeko erabaki batzuk hartu behar dira.

- Elementuen arteko GERTUTASUNaren definizioa. Nola kalkulatu da elementuen arteko gertutasuna edo distantzia? Galdera horri erantzuteko, gutxienez, beste bi galderari erantzun behar zaie. Gertutasuna neurtzeko orduan, zeintzuk izango dira erabiliko ditugun aldagaiak? Zein izango da bakoitzari emango diogun garrantzia? Nola definituko dugu elementuen ezaugarrien arteko antzekotasuna edo hurruntasuna? Elementuen arteko distantzia taldeen arteko distantziaren definizioarekin osatu behar da. Ohikoak dira distantzia euklidearra, distantzia euklidearraren karratua, Manhattanen distantzia...
- Sailkapen-metodoa (ALGORITMOA). Elementuak taldekatze erabilitako prozesuarekin du zerikusia. Elementu guztiak talde bakar batean sartu arte, ohikoa da pauso bakoitzean elementu pareak edo gertueneko taldeak elkartzea. Taldeak osatzeko jarraituko den metodoa hautatzea oso garrantzitsua da; izan ere, emaitza baldintza dezake. R Commander-ek Warden *lotura sinplearen* metodoarekin, McQuittyen metodoarekin... lan egiteko aukera ematen du
- Konglomeratu kopuruaren zehaztapena. Askotan ez dugu ezagutzen lagina osatzen duten elementuen tipologia eta aldakortasuna. Ezohiko elementuek aurkitzea, prozesuan egon beharko luketeen ala ez erabakitzea eta haien presentziarekin edo presentzia gabe zenbat talde eskuratu nahi ditugun zehaztea beharrezkoa da; izan ere, ezohiko elementuek hurrenez hurreneko faseetako elementuen taldekatzea baldintzatzen dute.

- Eskuratutako taldeen definizioa. Elementuak taldekatu ostean, haien deskribapenari ekingo diogu, konglomeratuak desberdinduko ditugu, eta eskuratutako taldekatzea baliagarria ote den eta proposatutako helburuei erantzuten al dien ikusiko dugu.

Distantzia EUKLIDEARRA gertutasun-neurri erabiliena da. Aldagaiak korrelazionatuta ez daudenean eta Pitagorasen teoremaren aplikazioarekin lortzen denean haren erabilera egokia da. Elementuen arteko distantzia honela definituko dugu:

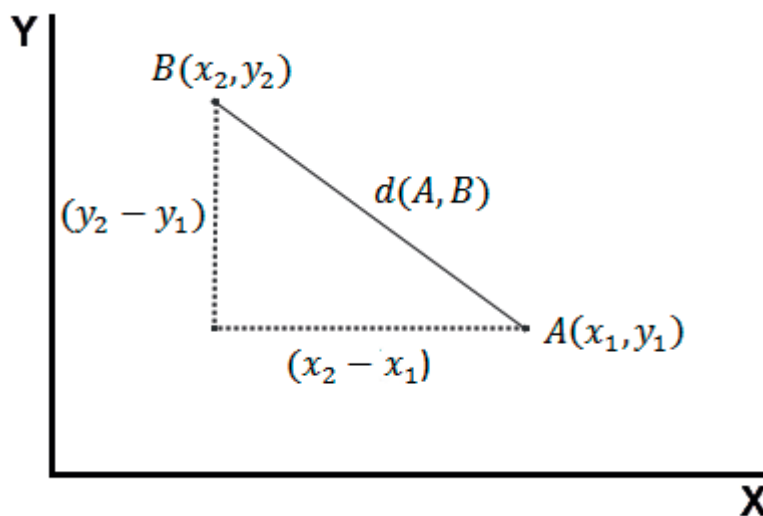
$$d_{A,B} = \sqrt{\sum_{k=1}^n (x_{2,k} - x_{1,k})^2}$$

7. formula.

Soilik bi aldagai hartzen baditugu kontuan, formula eta haren irudikapena hurrengoak dira:

$$d_{A,B} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

8. formula.



10. grafikoa. Distantzia euklidearra.

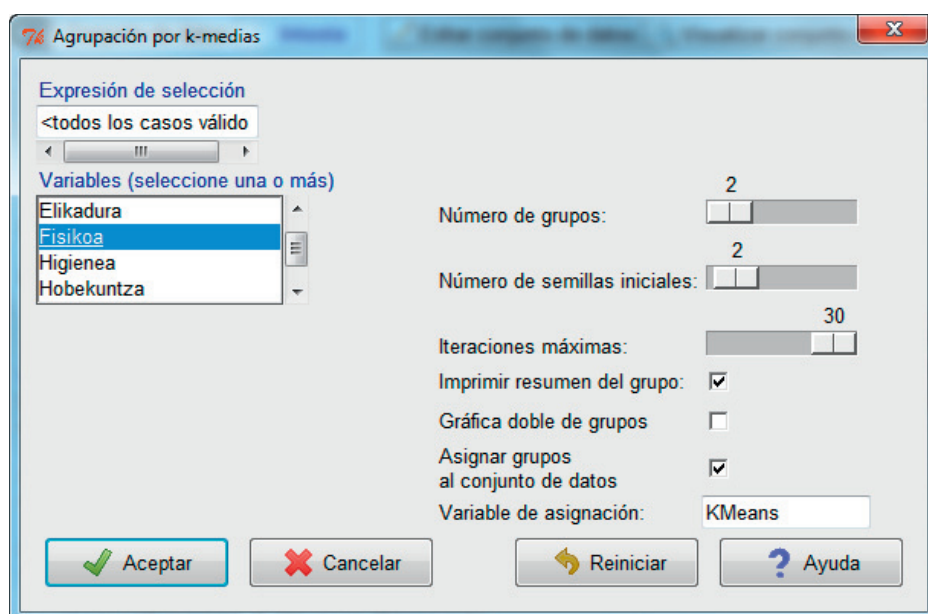
R Commander distantzia euklidearraren karratuarekin eta Manhattanen distantziarekin jarduten du. Honela definitzen da bigarrena:

$$d_{A,B} = \sum_{k=1}^n |x_{2,k} - x_{1,k}|$$

9. formula.

10.7.4.a. k-batezbestekoen bidezko taldekatzea

Analisi honetan lortu beharreko talde kopurua aurrez definituta dago. Taldekatze-prozesua hasierako erdigune (haziak) batzuekin hasten da. Normalean haien hautapena zorizkoa da. Hortik abiatuz, programak lagineko elementu bakoitzari gertuen daukan erdiguneko (haziaren) taldea esleitzen dio. Elementu guztiak talde bati esleitu ostean, erdigune berriak kalkulatzen dira (parte hartzen duten aldagaietan elementuek duten batezbestekoa) eta berriro kalkulatzen dira elementu bakoitzak talde berriekiko dituen distantziak gertuen duen taldeari esleitzeko. Talde egonkorak lortzean edo aurrez ezarritako iterazio kopurua (adibidean 30) gainditzean amaitzen da prozesua.



134. irudia.

Parte-hartzaile bakoitzari esleitu zaion taldea zein den adierazten duen aldagaia datu multzoari Kmeans izenarekin gehitzeko aukera ematen du R Commander-ek.

Eskuratutako taldeak deskribatzeko eta desberdintzeko R Commander-ek bakoitzaren estatistikoak aurkezten ditu. Emaitzen leihoan taldeen tamainak (93 eta 54) eta aldagai bakoitzeko batezbestekoak ikus daitezke. 2 taldeak 1 taldeak baino batezbesteko handiagoa du Adina eta Fisikoa aldagaietan eta txikiagoa Bizitzal aldagaian. Jarraian talde bakoitzaren barnean eta talde artean berretutako distantzien batuketaren emaitza eskaintzen da. Talde barnean 9590.405 eta 6347.269 balioak eskuratu dira eta talde artean 16954.31. Neurri hauek taldeen barneko homogeneousunaren eta talde arteko heterogeneousunaren berri ematen dute.

The screenshot shows the R Commander 'Ventana de resultados' window. The console displays the following R code and output:

```
> .cluster <- KMeans(model.matrix(~1 + Adina + Bizitzal + Fisikoa, Inkesta),
+   centers = 2, iter.max = 30, num.seeds = 2)

> .cluster$size # Cluster Sizes
[1] 93 54

> .cluster$centers # Cluster Centroids
new.x.Adina new.x.Bizitzal new.x.Fisikoa
1 73.98925 12.13226 5.849462
2 72.81481 34.37407 5.407407

> .cluster$withinss # Within Cluster Sum of Squares
[1] 9590.405 6347.269

> .cluster$tot.withinss # Total Within Sum of Squares
[1] 15937.67

> .cluster$betweenss # Between Cluster Sum of Squares
[1] 16954.31
```

Annotations in red boxes point to specific values:

- Analisisiko aldagaiak**: Points to the variable names in the model matrix.
- Taldeen tamaina**: Points to the cluster sizes (93, 54).
- Talde bakoitzak aldagai bakoitzean duen batezbestekoa**: Points to the cluster centroids table.
- Talde bakoitzaren barneko distantzia**: Points to the within-cluster sum of squares.
- Taldeen arteko distantzia**: Points to the between-cluster sum of squares.

135. irudia.

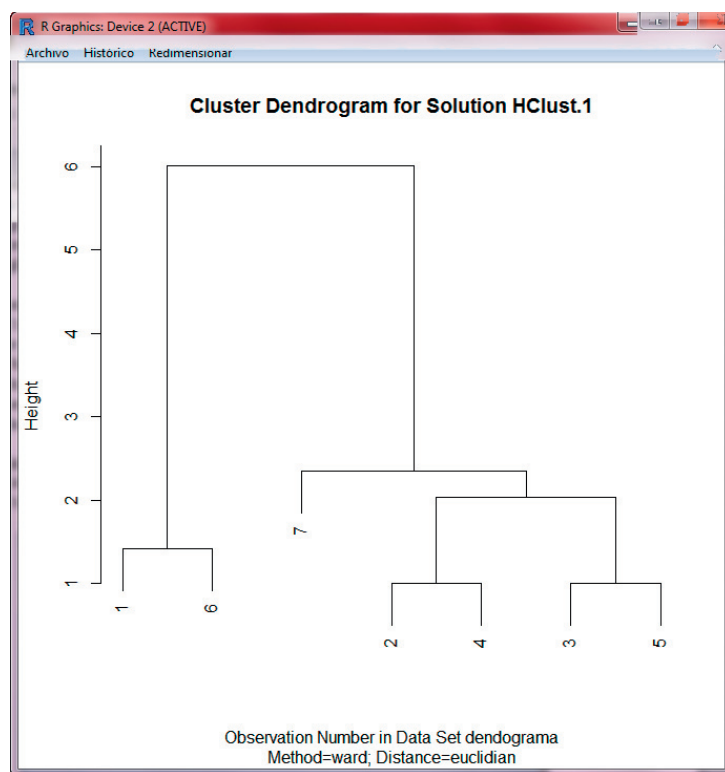
10.7.4.b. Taldekatze hierarkikoa

Taldekatze hierarkikoko prozedurak goranzkoak zein beranzkoak izan daitezke. Lehenengoak elementuak edo gertuen dauden taldeak bata bestearen atzetik elkartzen ditu. Elementu guztiak talde berean elkartzean amaitzen da prozedura. Beheranzko taldekatze hierarkikoa elementu guztiek osatzen duten talde bakar batetik abiatzen da eta hurrenez hurren urrutien dauden elementuak edo taldeak banatzen dira talde berriak sortuz, elementu bakoitzak talde bat osatu arte.

7 parte-hartzailek osatzen duten laginaren dendograma tipiko baten adibidea aurkezten da jarraian:

	1	2	3	4	5	6	7
x aldagaia	1	2	3	2	3	2	4
y aldagaia	9	6	7	5	6	8	5

Dendograman (prozesuaren irudikapena) ikus daitekeen bezala, beheko atalean 2 eta 4, eta 3 eta 5 elementuak elkartzen dira. Bi kasuetan aldagai batean (x) eskuratutako balioa bat dator eta beste aldagai batean (y) eskuratutakoan unitate bakarrean desberdintzen dira. Gainerakoan (1,6) pareak elkartzen da. Gainerakoan (2,4) taldea (3,5) taldearekin elkartzen da. Konglomeratuak (2,4,3,5) 7 taldearekin elkartuz jarraitzen du, eta azkenik bi aldeak elkartzen dira talde bakar batean. Elkartzen diren lerroen altura taldeen distantziarekiko proportzionala da.



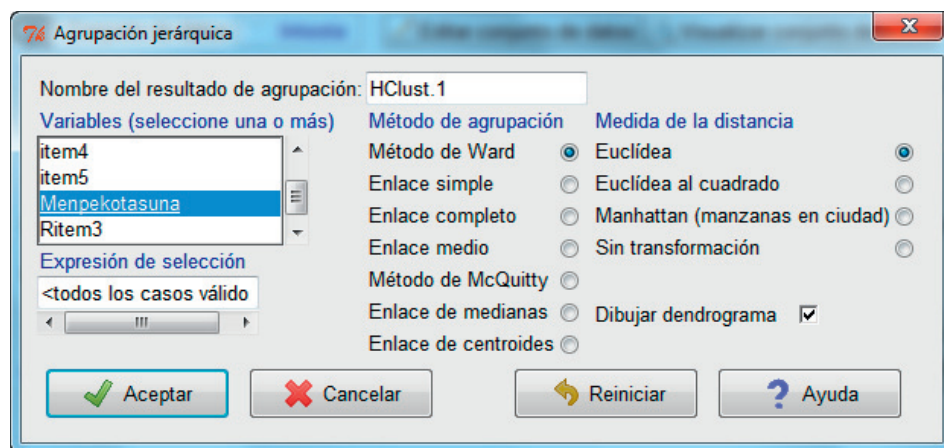
135B. irudia.

Taldekatze hierarkikoen metodoak

Gehien erabiltzen den taldekatze hierarkikoen metodoa R Commander-ek besterik esan ezean erabiltzen duenarekin bat dator, Warden metodoa. Konglomeratuen arteko distantziak kalkulatzeko bariantza-analisan oinarritzen da. Etapa bakoitzean taldekatze bakoitzak dakarren karratuen gehiketa hartzen du kontuan eta balio hori minimizatzen dutenak elkartzen ditu talde berrian. Hortaz, bateratzen diren konglomeratuek informazio-galera txikiena dakarte. Prozedurak elementu gutxik osatzen dituzten konglomeratuak elkartzeko joera du, azkenean antzeko tamainako konglomeratuak eskuratu arte.

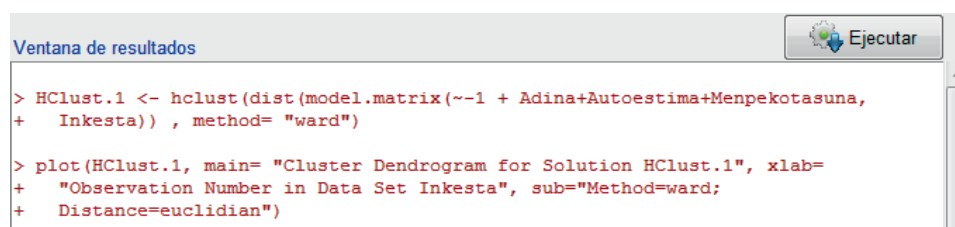
Gainera lotura sinpleari, osoari eta batezbestekoari dagozkien algoritmoak jasotzen ditu R Commander-ek. Lotura sinpleko metodoak konglomeratu bateko edozein puntutatik beste konglomeratuko edozein puntutara egon daitekeen gutxieneko distantzia esleitzen dio bi konglomeratuen arteko distantziari. Lotura osoko metodoak gehieneko distantzia da esleitzen duena. Eta batezbesteko loturak, izenak dioten bezala, bi konglomeratuen elementuen arteko distantzia guztien batezbestekoa hartzen du bi konglomeratuen arteko distantziatzat.

Taldekatze hierarkikoa gauzatzeko hurrengo aukerak baliatuko ditugu: Estadísticos > Análisis dimensional > Análisis de agrupación > Agrupación jerárquica.

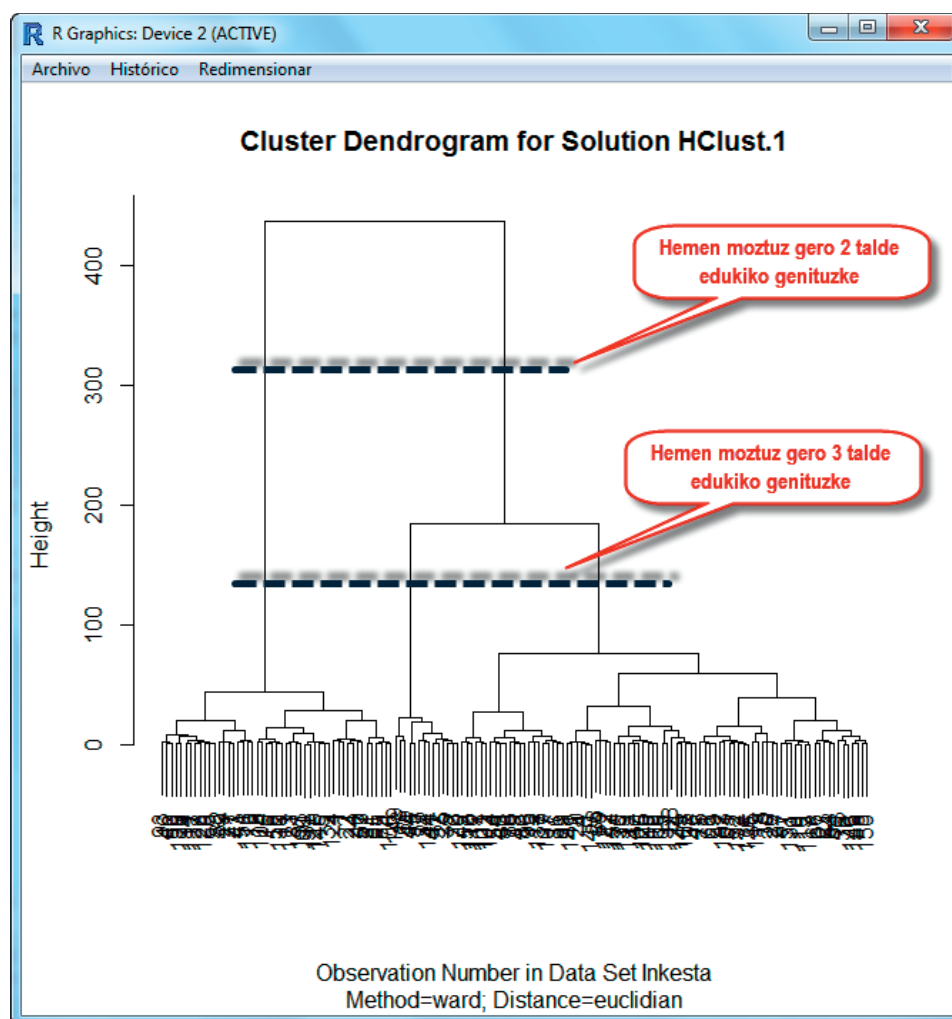


136. irudia.

Analisiaren emaitza beste leiho independente batean agertzen den dendrograma da. Emaitzen leihoan, hurrengo irudian ikus daitekeen moduan, agertzen den gauza bakarra komando bat da.



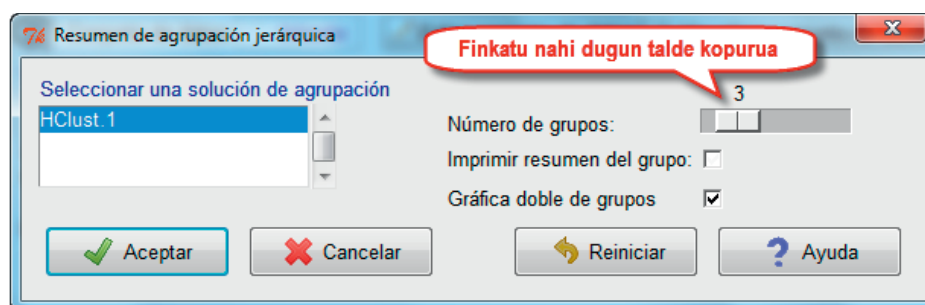
137. irudia.



138. irudia.

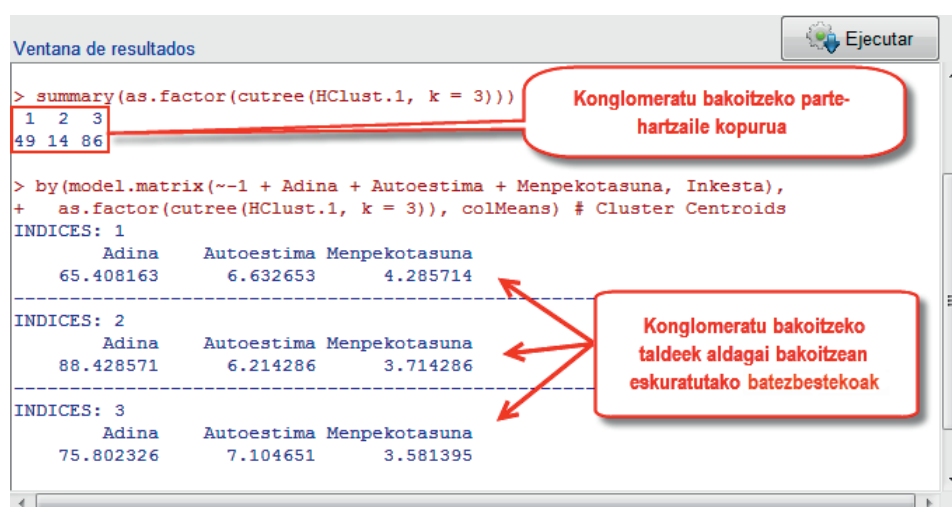
Dendrogramak taldekatze-prozesuaren egitura begiz aztertzeko aukera ematen du. «Adarrak» moztean taldeen homogeneotasuna zein den sumatu dezakegu. Hiru talde lortu nahi baditugu, nahikoa da azpitik hiru adar bertikal uzten dituen altueran moztea dendograma. Adar horietako bakoitzak eskuratutako taldeetako bat ordezkatuko luke. Berdina gertatuko da bi talde lortu nahi baditugu.

Lortu nahi dugun talde kopurua zein den erabaki ostean, parte-hartzaile bakoitzari esleitutako taldea zein den jasotzen duen aldagaia eska diezaiokegu R Commander-i. Horretarako Estadísticos > Análisis dimensional > Análisis de agrupación > Resumir la Agrupación jerárquica... sakatzean irekiko den leihoan lortu nahi dugun talde kopurua definitu dezakegu, baita taldearen laburpena eta taldeen grafiko bikoitza sortzeko aukerak markatu ere.



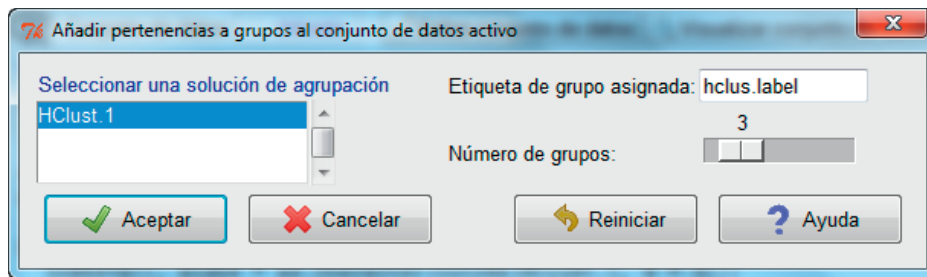
139. irudia.

Azpiko irudian agertzen den taldearen laburpenean lortutako talde bakoitzean konglomeratuak eskuratzeko prozesuan erabilitako aldagai bakoitzaren batezbestekoa aurkezten da.



140. irudia.

Elementu bakoitzari esleitu zaion taldea zein den adierazten duen aldagaia sortzeko (datu multzoaren amaieran agertuko da) Estadísticos > Análisis dimensional > Análisis de agrupación > Agregar la Agrupación jerárquica al conjunto de datos... sakatu behar da.



141. irudia.

Modu horretan, gure datu multzoari 1, 2 eta 3 kategoriak dituen `hclus.label` aldagaia gehituko zaio. Balio horietako bakoitza eskuratutako konglomeratu bati dagokio. Informazio honek talde desberdinak deskribatzen eta desberdintzen lagunduko digu.

10.8. EREDUEN DOIKETA

Ereduak ekuazio moduan adierazten diren eta aldagaien funtzionamendua azaltzen duten prozedurak dira. Gehien ezagutzen eta erabiltzen dena Eredu Lineal Orokorra da. Hainbat aldagai azaltzaileraren (aldagai independenteren) bitartez aldagai kuantitatibo batek (aldagai azaldua, erantzun-aldagaia edo mendeko aldagaia) pairatzen dituen aldaketak azaltzea ahalbidetzen du. Aldagai azaltzaileak kualitatiboak (faktoreak) zein kuantitatiboak izan daitezke. Eredu hauek Orokortutako Eredu Linealera (OEL) zabaltzean, aldagai azaldua kualitatiboa nahiz kuantitatiboa izan daiteke. Hurrengo atalean Eredu Lineal Orokorreko formulazioa azaletik azalduko dugu. R Commander-ek *Ajuste de modelos* azpimenuan jasotzen duen lehen aukera izateaz gain, Gizarte Zientzietan bere aldaera guztietan ere du erabiliena denez, hurrengo atalean erregresio-analisiaren inguruan arituko gara.

10.8.1. Erregresio anizkoitza

Aldagai azaltzaile edo independenteetan ($x_1, \dots, x_k$) gertatzen diren aldaketen bidez aldagai azalduaren edo mendekoaren (y) aldakortasuna azaltzen (edo testuinguruaren arabera, aurreratu) saiatzen da erregresioaren eredua. Aldagai azaltzaile bakarra dagoenean, erregresio bakunaz arituko gara. Aldiz, aldagai azaltzaile bat baino gehiago badaukagu, gure lana erregresio anizkoitzaren esparruan kokatuko da.

Erregresio anizkoitzak $x_1, x_2, \dots, x_k$ izendatuko ditugun aldagai azaltzaileen eta y aldagai azalduaren artean korrelazioa badela onartzen du.

$$y = f(x_1, x_2, \dots, x_k) + \varepsilon$$

10. formula.

f funtzioaren bidez y aldagaiaren baliora gerturatzen saiatzean pairatzen dugun errorea ε karaktereak adierazten du. Errore hori mendeko aldagaiaren eta aldagai independenteen arteko erlazioa erabatekoa ez izatearen ondorioa da.

Hasierako baldintza moduan aldagaien arteko erlazioa linealtzat eta gehigarritzat hartzen da; hortaz, aurreko ekuazioa honela adieraz daiteke:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_k x_k + \varepsilon$$

11. formula.

«Benetako» erregresioaren ekuazioa zenbatestea da helburua. Alegia, aztertu nahi dugun populazioko puntuen lainoari hoberen doitzen zaion planoaren ekuazioa aurkitzea. Horretarako populazioarekiko adierazgarria den lagin batetik eskuratutako datuak behar ditugu. Datuei esker lortzen baitira erregresioaren ekuazioko $\beta_0, \dots, \beta_k$ koefizienteen $b_0, \dots, b_k$ zenbatespenak. Erregresio-planoa puntuen lainoari doituko diogu gure laginean. Erregresio-planoaren ekuazioa hau da:

$$y = b_0 + b_1 x_1 + b_2 x_2 \dots + b_k x_k + e$$

12. formula.

Puntuen lainoari hoberen doitzen zaion planoaren erregresio-plano moduan ezagutuko dugu. Ereduaren aplikazioa egokia izan dadin, hainbat aurrebaldintza bete behar dira:

- **Eredua ondo zehaztea.**

- Aldagai azaltzaileen (independenteak) eta aldagai azalduaren (mendekoa) arteko erlazioak lineala eta gehigarria izan behar du. Baldintza hau onartzen dugunean $x_1, \dots, x_k$ balioen konbinazio bakoitzetik eratortzen den y -ren balioa aldagai azaltzaileen konbinazio lineala dela onartzen dugu.
- EREDUA ondo zehaztuta dago. Erregresio ereduak berau azaltzeko garrantzitsuak diren aldagai guztiak biltzen ditu eta garrantzitsuak ez direnak ereditik at geratzen dira.
- Ez dago MULTIKOLINEALITATERIK. Aldagai azaltzaileek independenteak izan behar dute euren artean. Hots, ereduak ez ditu jaso behar ez aldagai errepikatuak, ezta beste aldagaien konbinazio linealak direnak.

- **Aldagaiak neurketa-errorik ez edukitzea.** Aldagaiak errore sistematiko gabe neurtuta egon behar dute.
- ε_i **neurketa-erroreak** hurrengo baldintzak bete behar ditu:
 - ε_i aldagaiaren banaketa normala da.
 - $(x_1, \dots, x_k)$ balioen konbinazio bakoitzeko

$$E(\varepsilon_i) = 0 \text{ eta } E(\varepsilon_i^2) = \sigma^2$$

13. formula.

- $i \neq j$ guztientzat

$$E(\varepsilon_i, \varepsilon_j) = 0$$

14. formula.

- i guztientzat

$$E(\varepsilon_i, (x_{1i}, \dots, x_{ki})) = 0$$

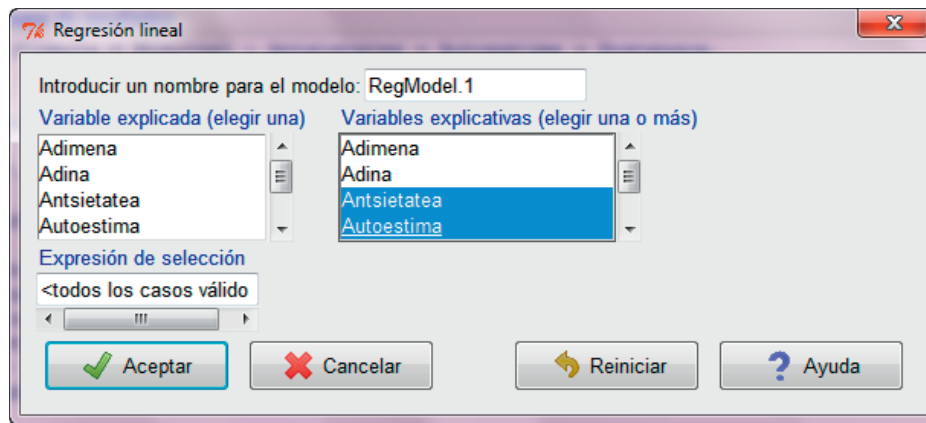
15. formula.

Azken baldintza hori matrizeen bidez adieraz daiteke a) ε_i zorizko errorearen bektoreak n dimentsioko banaketa normalari jarraitzen diola, b) bere batezbestekoa 0 zutabe-matrizea dela, eta c) bariantzen-kobariantzen matrizea σ^2 konstanteaz biderkatutako identitate-matrize diagonal dela zehaztuz.

$$\underline{\varepsilon}_i \rightarrow N_n(\underline{0}, \sigma^2 I_n)$$

16. formula.

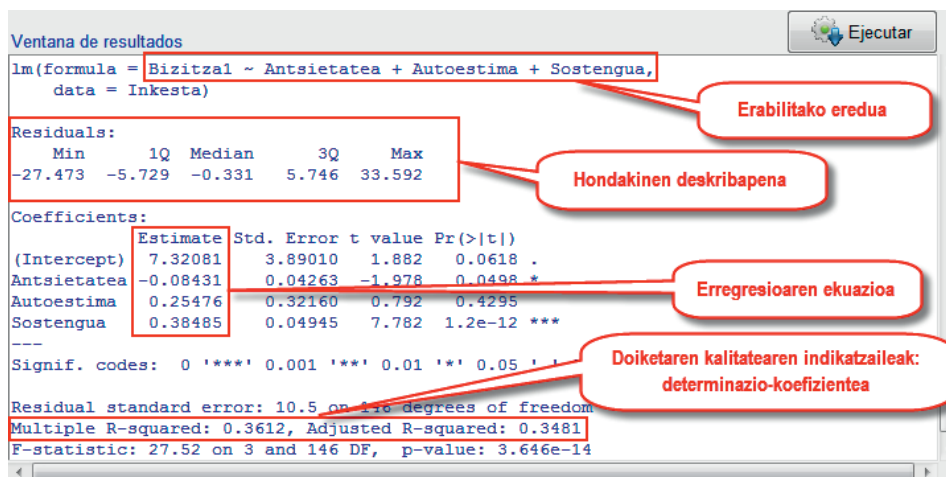
Pertsona batzuek bizitzarekiko gogobetetze-maila (Bizitzal) beste batzuek baino handiagoa zergatik duten azaltzen saiatuko gara. Horretarako antsietate-mailari (Antsietatea), sostegu sozialari (Sostengua) eta autoestimuari (Autoestima) buruzko informazioa daukagu. Analisia burutzeko jorratu beharreko bidea Estadísticos > Ajuste de modelos > Regresión lineal... da.



142. irudia.

Emaitzek erakusten duten lehenengo gauza erregresio-planoaren bidez mendeko aldagaia zenbatesten saiatzean sortzen den hondakina da. Jarraian adiera hertsian ekuaziotzat ulertzen dena aurkezten da. Lehenengo zutabean β_i koefizientearen zenbatespenak agertzen dira dagozkion errore tipikoekin, t balioekin eta esanguratsutasun-mailekin.

Ondoren, R Commander-ek doiketaren kalitateari buruzko informazioa eta determinazio-koefizientea ($\text{Multiple Rsquared} = 0.3612$) erakusten du. Alegia, ereduan sartu diren hiru aldagaiek azaltzen duten Bizitzal aldagaiaren aldakortasunaren portzentajea. Koefizientea doitzean %34.81era jaisten da azaldutako bariantza. Amaitzeko, bariantzaren analisiaren F balioa aurkezten da dagozkion askatasun-graduekin eta esanguratsutasun-mailarekin.



143. irudia.

Erregresio-ekuazioaren zenbatespena hurrengoa da:

$$\text{Bizitza1} = 7,32 - 0.08 \times \text{Antsietatea} + 0.80 \times \text{Sostengua} + 0.25 \times \text{Autoestima}$$

Antsietatea aldagaiaren koefizientea negatiboa da; hortaz, zenbat eta handiago izan antsietate-maila, are eta txikiagoa izango da bizitzarekiko gogobetetze-maila. Beste aldagaiek koefiziente positiboak dituzte. Beraz, zenbat eta sostengu sozial gehiago eta zenbat eta autoestimu-maila handiagoa eduki, orduan eta bizitzarekiko gogobetetze-maila handiagoa edukiko du parte-hartzaileak.

Nahiz eta erregresio-ereduaren emaitzen aurkezpena hemen amaitzen den, R Commander-ek eredu estatistikoen jarduna zaintzen du; eta eredu bat landu ostean, haren azterketarako aukera gehiago eskaintzen dituen menu berezia aktibatzen da.

11. Ereduak – Eredu lineal orokorra

Aurreko ataletan bariantza- eta erregresio-analisiak aurkeztu ditugu. Ikuspuntu estatistiko-matematiko batetik, euren oinarri eta prozedurari dagokienez, biak antzekoak dira; aldagai batek pairatzen dituen aldaketak (aldagaiaren bariantza) azaltzen saiatzen dira kuantitatiboak edo faktoreak (bariantzaren analisisan) izan daitezkeen beste aldagai batzuen aldakortasunaren bidez. Bi analisiak Orokortutako Eredu Linealekoak (ELO) dira.

Ereduaren adierazpen orokorra hau da:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_k x_k + \varepsilon$$

16. formula.

Azaltzeko asmoa daukagun aldagaia y da; $x_1, x_2, \dots, x_n$ zorizko aldagai independenteak dira; $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ egituraren parametroak dira, eta ε zorizko errorea da. x_i aldagaiak behagarriak direnean, ereduari Eredu Lineal deritzo eta bariantzaren analisisa, kobariantzaren analisisa eta erregresio linealaren ereduak bere aldaera guztietan biltzen ditu. Eredu lianealean aldagai kualitatiboak (faktoreak) sar daitezke aldagai azaltzaileen artean. Normalean, ereduaren zenbatespenerako erabiltzen den prozedura gutxiengo karratuena da.

Aurreko eredu bariantza-analisisan doitzeko, aldagai azaltzaileek kualitatiboak (faktoreak) izan behar dute eta beren balioek faktoreko mailekin edo kategoriekin bat etorri behar dute. Hortaz, faktore bat daukagunean, ereduaren adierazpena hurrengoa izango da:

$$y = \mu + \alpha_i + \varepsilon_i$$

17. formula.

μ benetako eraginaren batezbestekoa da, i tratamenduaren edo faktore-mailaren benetako eragina α_i da eta ε_i zorizko errorea da.

Interakzioa duten bi faktoreren kasuan, adierazpena hurrengoa da:

$$y = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij}$$

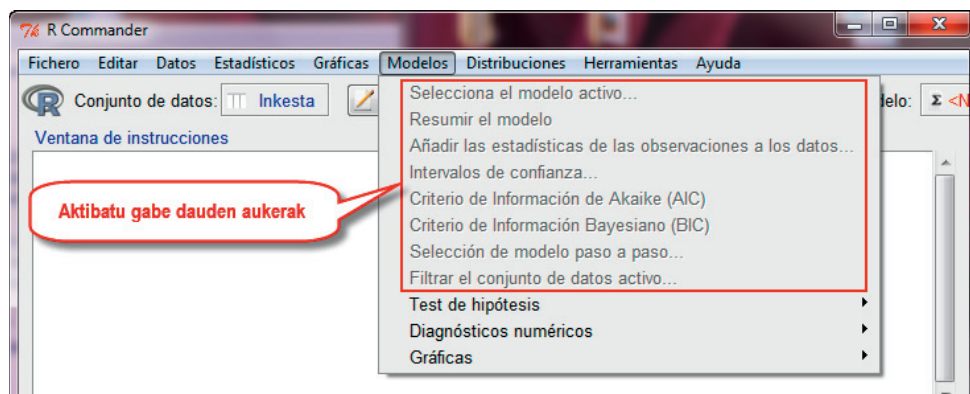
18. formula.

μ benetako eraginaren batezbestekoa da, A faktoreko i tratamenduaren edo faktore-mailaren benetako eragina α_i da, B faktoreko j tratamenduaren edo faktore-mailaren benetako eragina β_j da, A eta B faktoreen arteko interakzioa $(\alpha\beta)_{ij}$ da, eta ε_{ij} zorizko errorea da.

Orokortutako Eredu Linealean (OEL) mendeko aldagai kuantitatibo zein kualitatiboak aurki daitezke. Horretarako, banaketa normalaz gain beharrezkoa da banaketabinomialarekin, Poissonen banaketarekin, banaketahipergeometrikoarekin... jardutea. OELn eredia zenbatesteko prozedura erabat aldatzen da; normalean sinesgarritasun gehieneko zenbatespena erabiltzen da.

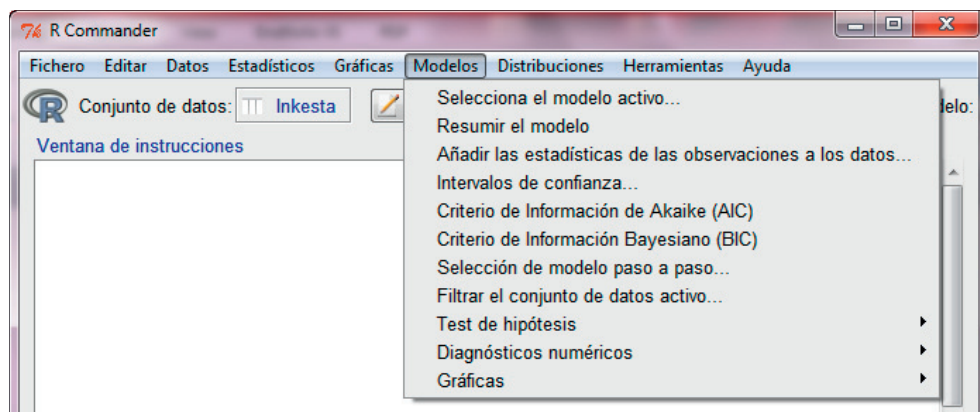
OELren oinarri teorikoen azalpenak eta emaitzen interpretazioak testu honen helburuak gainditzen dituztenez, alde batera utziko ditugu Logit eredu multinomialekin eta erregresio ordinalarekin batera.

R Commander-eko Estadísticos menuan Eredu Lineal Orokorrari dagozkion ereduak defini daitezke. Eredua definitu eta emaitzak lortu ostean, Modelos menuaren bidez sortu diren ereduaren azterketan sakondu daiteke. Menu hori soilik bariantza-analisia edo erregresioa egin ostean aktibatuko da.



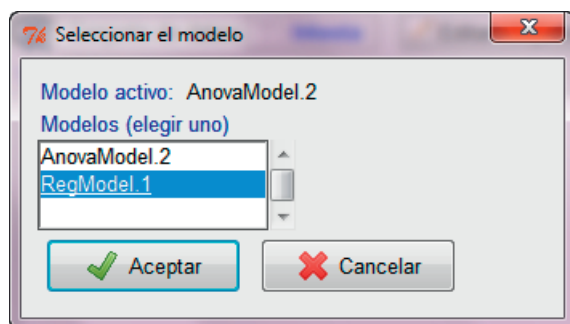
144. irudia.

Menu honetako aukerak azaltzeko erregresio anizkoitz baten analisia burutu dugula imajinatuko dugu. Analisiaren helburua bizitzarekiko gogobetetze-mailaren (Bizitzal), antsietate-mailaren (Antsietatea), sostengu sozialaren (Sostengua) eta autoestimu-mailaren (Autoestima) arabera azaltzea izan da (ikusi aurreko atalean). Bestalde, bi faktoreko bariantzaren analisia gauzatu dugu bizitzarekiko gogobetetze-maila (Bizitzal) tratamenduaren (Tratamendua) eta sexuaren (Sexua) arabera azaltzeko asmoarekin. Analisi hauek gauzatu ostean Modelos menua aktibatuko da eta hurrengo aukeretara jo ahal izango dugu:



145. irudia.

Lehenengo bi aukerekin, Selección del modelo activo eta Resumir el modelo, aurrez zenbatetsitako ereduak laburtu daitezke.

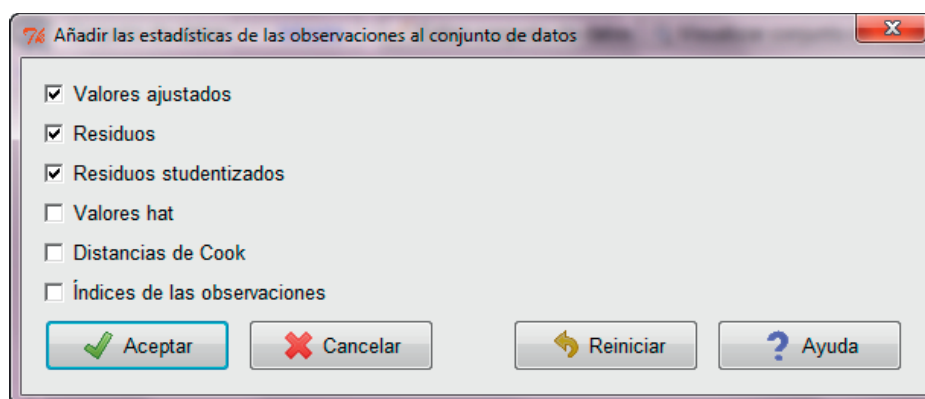


146. irudia.

11.1. EREDUAK. ERREGRESIOAREN ANALISIA

Erregresioaren eredua aktibatu eta gero, R Commander-ek ereduaren laburpena eskuratzeko aukera eskaintzen du. Estadísticos menutik eredua definitzean lortzen diren emaitza berak ematen ditu.

Modelos > Añadir las estadísticas de las observaciones a los datos aukerari esker, aktibatutako datu multzoari hurrengo irudian ageri diren estatistikoak gehitu diezazkiokegu zutabe berrietan.



147. irudia.

	fitted.RegModel.1	residuals.RegModel.1	rstudent.RegModel.1
1	14.741768	17.15823160	1.666025287
2	28.634871	-2.03487071	-0.197716157
3	28.769090	8.83091003	0.853173364
4	27.750046	9.84995353	0.945736623
5	21.024467	-9.62446728	-0.983305
6	34.146098	5.7539020	0.218
7		-12.6190421	0.581
8		-20.87941913	0.141
9		-21.18245421	-2.062229698
10	25.521671	0.66	1.521616956
11	25.521671	0.66	1.521616956
12	23.001723	0.676	1.641561308
13	19.006410	-8.80640966	-0.847244927

Menuko hurrengo aukera sakatuko bagenu, Intervalos de confianza, interesatzen zaigun konfiantza-mailarekin KTetako mugak estimatu ahal izango genituzke erregresioaren ekuazioko parametroentzat. Konfiantza-maila % 95ean finkatuz, KT hauek eskuratu ditugu:

	Estimate	2.5 %	97.5 %
(Intercept)	7.32081110	-0.3673627	1.500898e+01
Ansietatea	-0.08431334	-0.1685691	-5.755634e-05
Autoestima	0.25476087	-0.3808245	8.903463e-01
Sostengua	0.38484899	0.2871094	4.825885e-01

148. irudia.

Erregresioaren ereduko parametroak zenbatesteko oinarritzko prozedura gutxiengo karratuena da. Baina erregresio logistikoko, Poissonen, Coxen... parametroak zenbatesteko sinesgarritasun gehieneko zenbatespen-metodoa erabiltzen da. Gutxiengo karratuen bidez erregresioaren doiketaren kalitatea ebaluatzeko doitu-tako determinazio-koefizientea erabiltzen da. Gehieneko sinesgarritasunaren zenbatespenaren kasuan, ordea, beste neurri mota batzuk erabiltzen dira hautapen-eta diagnostiko-irizpide moduan. Ezagunenak Akaikeren informazioaren irizpidea (AIC) eta Informazioaren Irizpide Bayestarra (BIC) dira. Modelos menuaren bidez eskura daitezke.

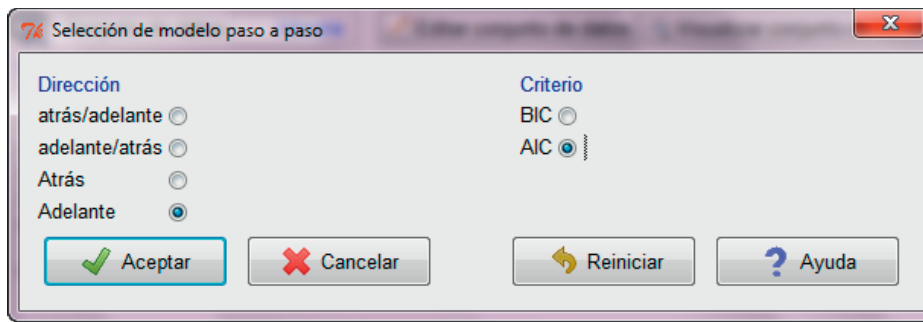
Ereduen diseinuan oso garrantzitsua da aldagaien ordena; izan ere, emaitzak baldintza ditzateke. Gainera, ordenak ikertzen ari garen aldagaiaren funtzionamenduari buruzko informazio garrantzitsua ematen du. Modelos menuko hurrengo aukeran, Selección del modelo paso a paso, aldagaiak ereduan sartzeko lau metodo eskaintzen ditu. Aldagaien ordena baldintzatzen duten metodo hauek ez dira zorizkoak, AICk eta BICk doiketaren irizpideei erantzuten diete.

Dirección: Adelante. Ereduan sartzten den lehen aldagaia eredu hoberen doitzen duena da. Alegia, AICren edo BICren balioa gehien txikitzen duena. Lehenengo aldagaia ereduan sartu ostean, bigarren aldagaia sartuko da irizpide berari jarraituz. Aldagai berriak sartzeko ereduen doiketa hobetzen ez duenean prozesua amaitutzat emango da. AICren eta BICren balioak txikiagotzeak doiketa hobetu egin dela adierazten du.

Dirección: Atrás. Hasiera batean aldagai guztiak sartzten dira ereduan; eta jarraian, eredutik kentzean ereduen doiketa hobetzen duten aldagaiak (baldin eta baleude) kentzen dira gehien hobetzen duenetik hasita. Ez dagoenean eredutik kenduta ereduen doiketa hobetzen duen (AICren edo BICren balioa txikitzen duen) aldagairik, prozesua amaitutzat ematen da.

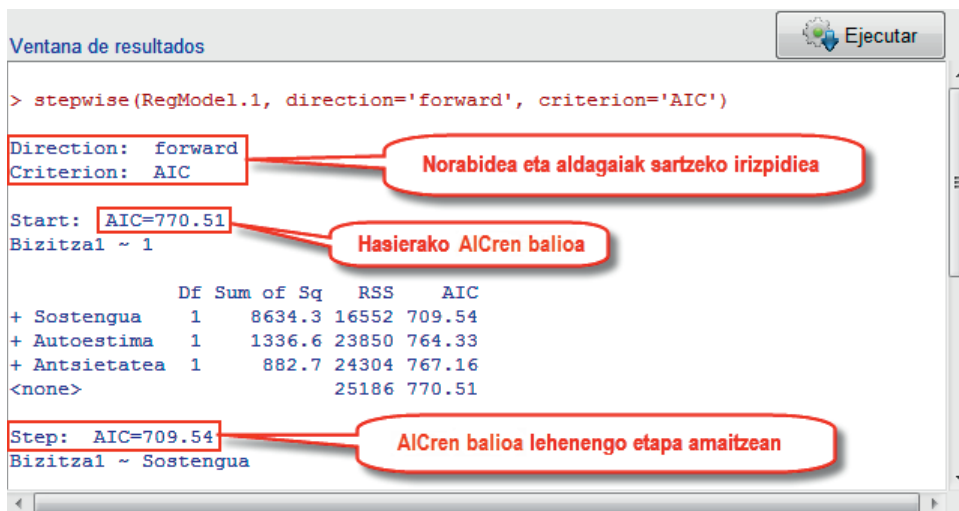
adelante/atrás eta atrás/adelante norabideak aurrekoen konbinazioak dira.

Aldagaiak hautatzeko aurreranzko (Adelante) metodoarekin eta AICren doiketaren irizpidearekin R Commander-en lan egiteko modua azalduko dugu adibide baten bidez. Ereduko parametroen kopuruak AICren balioan eragina duenez, batzuetan aldagai berriak sartzean AICren balioa handitu egiten da eta doiketa okertu. ($AIC = -2(\ln \text{sinesgarritasuna} - \text{parametro kopurua})$).



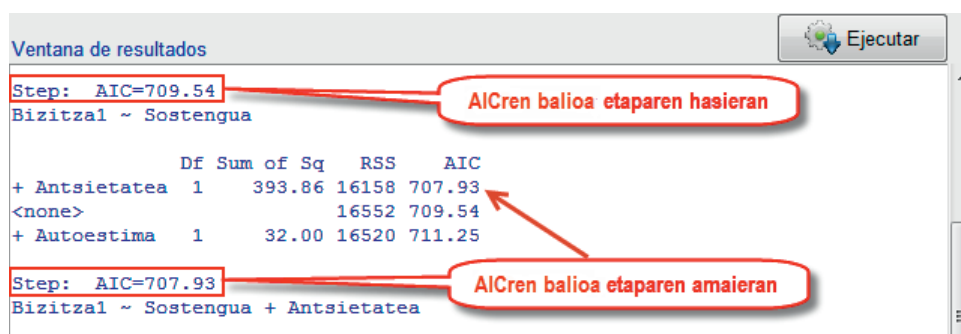
149. irudia.

Prozeduraren emaitzak etapetan aurkezten dira. Lehenengo etapan, R Commander-ek AIC doiketaren indikatzailearen hasierako balioa (770.51) kalkulatzen du, eta erudian adierazitako aldagaietako edozein sartzean lortuko litzatekeen aldaketa erakusten du. Ikus daitekeen moduan indikatzailea gehien txikitzen duen aldagaia Sostengua da; AIC 709.54 balioraino txikitzen du. Hortaz, hau da lehenengo etapan ekuazioari gehituko zaion aldagaia.



150. irudia.

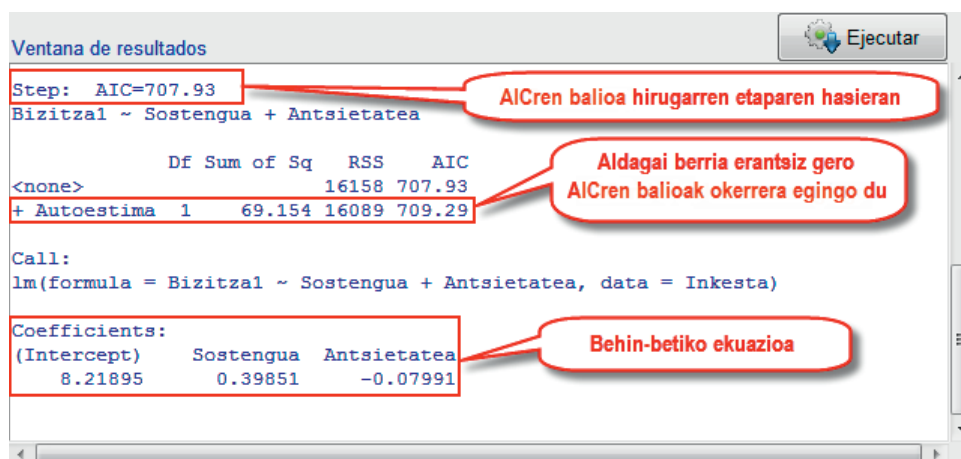
Bigarren etapako emaitzak azpiko irudian daude. AICren 709.54 baliotik abiatzen da. Emaizen leihoan dago aldagaiak sartuz gero indikatzaileak pairatuko lukeen aldaketa. AIC indikatzailea gehien txikitzen duena Antsietatea aldagaia da. Berau izango da ekuazioari bigarren etapan gehituko zaiona.



151. irudia.

Hirugarren etapako emaitzen aurkezpena (eta adibide honetako azkena) hasierako ereduaren datuekin hasten da: ekuazioan Sostengua eta Antsietatea aldagaiak daude eta AICren balioa 707.93 da. Ekuazioan Autoestima aldagaia sartuko bagenu, AICren balioa 709.27 izango litzateke. Hortaz, aldagai hori ekuazioan sartzeak «zarata» ekarriko luke, eta doiketa ez hobetzeaz gain okertuko luke. Egoera honetan programak ez du aldagaia ekuazioan sartuko. Amaierako erregresio-ekuazioaren zenbatespena hurrengoa da:

$$\text{Bizitzal} = 8.21895 + 0.39851 \times \text{Sostengua} - 0.07991 \times \text{Antsietatea}$$

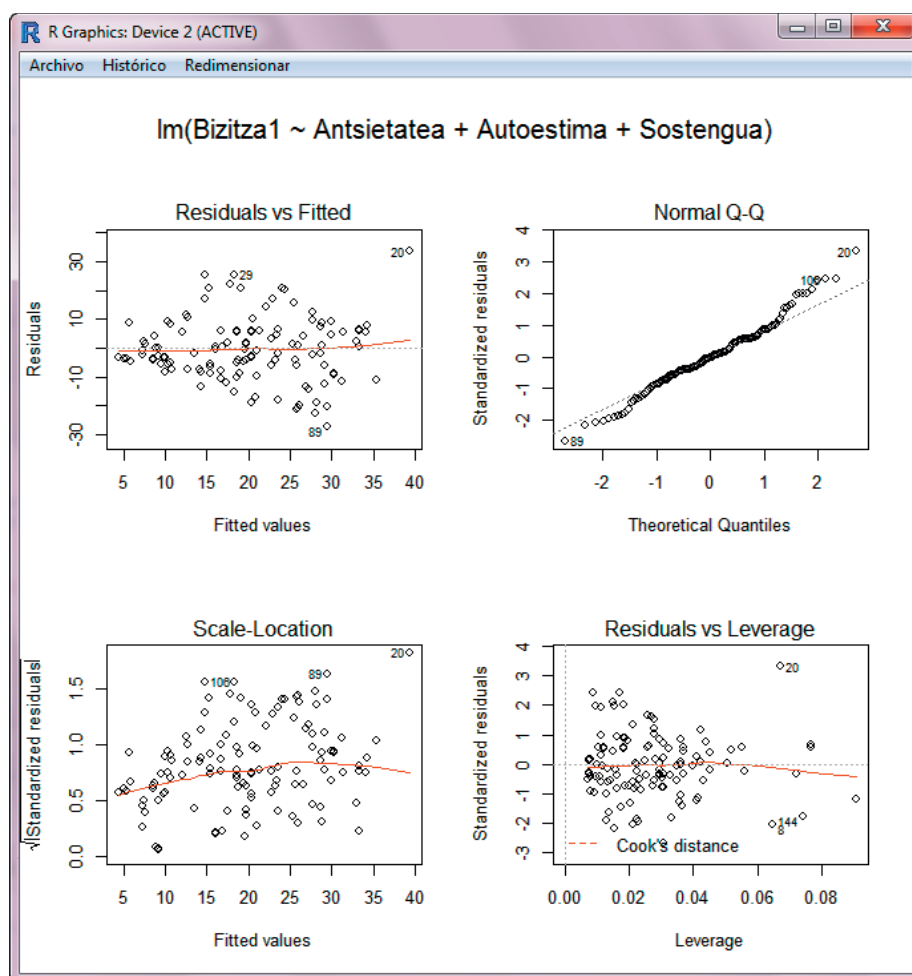


152. irudia.

Gráficas aukerak eredu balioztatzeko eta bere aplikazioaren baldintzak aztertze erremintak eskaintzen ditu. R Commander-ek eskaintzen duen aukera multzotik oinarritzkoenak soilik aztertuko ditugu: Gráficas básicas de diagnóstico. Gure kasuan eskuratutakoak hurrengoak dira:

- Residuals vs Fitted values. Mendeko aldagaia zenbatetsitako balioen arabera erregresio-ekuazioaren bidez gerturatzean eskuratutako hondakinak irudikatzen ditu.
- Normal Q-Q. Kuantil teorikoak estandarizatutako hondakinen bidez behatutakoekin alderatzen ditu.
- Scale-Location. Erregresio-ekuazioaren bidez eskuratutako zenbatespenak eta estandarizatutako hondakinak konparatzen ditu.
- Residuals vs Leverage. Datu atipikoen eta estandarizatutako hondakinen arteko erlazioa irudikatzen du.

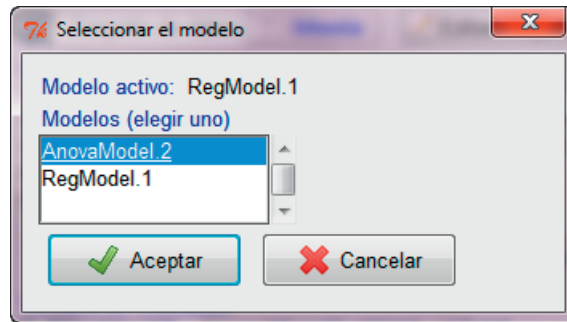
Gainera, grafiko hauekin balio atipikoak identifika (*outliers*) daitezke. Adibidez, 4 parte-hartzaileari beste parte-hartzaile guztientzat finkatutako aldagaien arteko erlazioetik «urruntzen» den hondakina dagokio.



153. irudia.

11.2. EREDUAK. BARIANTZAREN ANALISIA

Sexu (Sexua) eta tratamendu (Tratamendua) bakoitzaren bizitzarekiko gogobetetze-maila (Bizitzal) konparatzeko bi faktoreko bariantzaren analisiari ekin ostean (Estadísticos > Medias > ANOVA de dos factores...) Modelos > Seleccionar el modelo... sakatzean irekitzen den leihoan AnovaModel.2. hautatu dezakegu.



154. irudia.

Lehenengo aukerarekin ereduaren laburpena eskuratu daiteke (Modelos > Resumir el modelo). R Commander-ek aldagai independente moduan bi aldagai dikotomikoak dituen erregresio anizkoitzaren emaitzak aurkezten ditu. Sexua aldagaian 1 balioak parte-hartzailea Gizona dela adierazten du, eta 0 balioak Emakumea dela. Tratamenduei dagokienez, 1 zenbakiak tratamendu Berritzailea ordezkatzten du, eta 0-k tratamendu Klasikoa. Bi aldagaiak gurutzatzean lortutako aldagaian (Sexua x Tratamendua) 1 balioak parte-hartzailea Gizona dela eta tratamendu Berritzailea jaso duela adierazten du; beste hiru kategoriak 0 balioarekin ordezkatu dira. Irakurleak ikus dezakeen moduan beste aukera guztiak erregresioaren ereduari dagozkio.

Ventana de resultados Ejecutar

Call:
lm(formula = Bizitzal ~ Sexua * Tratamendua, data = Inkesta)

Residuals:

Min	1Q	Median	3Q	Max
-20.953	-10.766	-1.098	8.665	50.947

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18.829	1.878	10.028	<2e-16
Sexua[T.Gizona]	3.715	3.129	1.187	0.2371
Tratamendua[T.Klasikoa]	3.224	2.699	1.194	0.2343
Sexua[T.Gizona]:Tratamendua[T.Klasikoa]	-7.302	4.381	-1.667	0.0977

(Intercept) ***
Sexua[T.Gizona]
Tratamendua[T.Klasikoa]
Sexua[T.Gizona]:Tratamendua[T.Klasikoa] .

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

Residual standard error: 13.01 on 146 degrees of freedom
Multiple R-squared: 0.01897, Adjusted R-squared: -0.0011
F-statistic: 0.941 on 3 and 146 DF, p-value: 0.4226

Azaldutako aldagaiak eta faktoreak

Erregresioaren ekuazioa

1 balioarekin kodifikatutako kategoriak

155. irudia.

12. Grafikoak

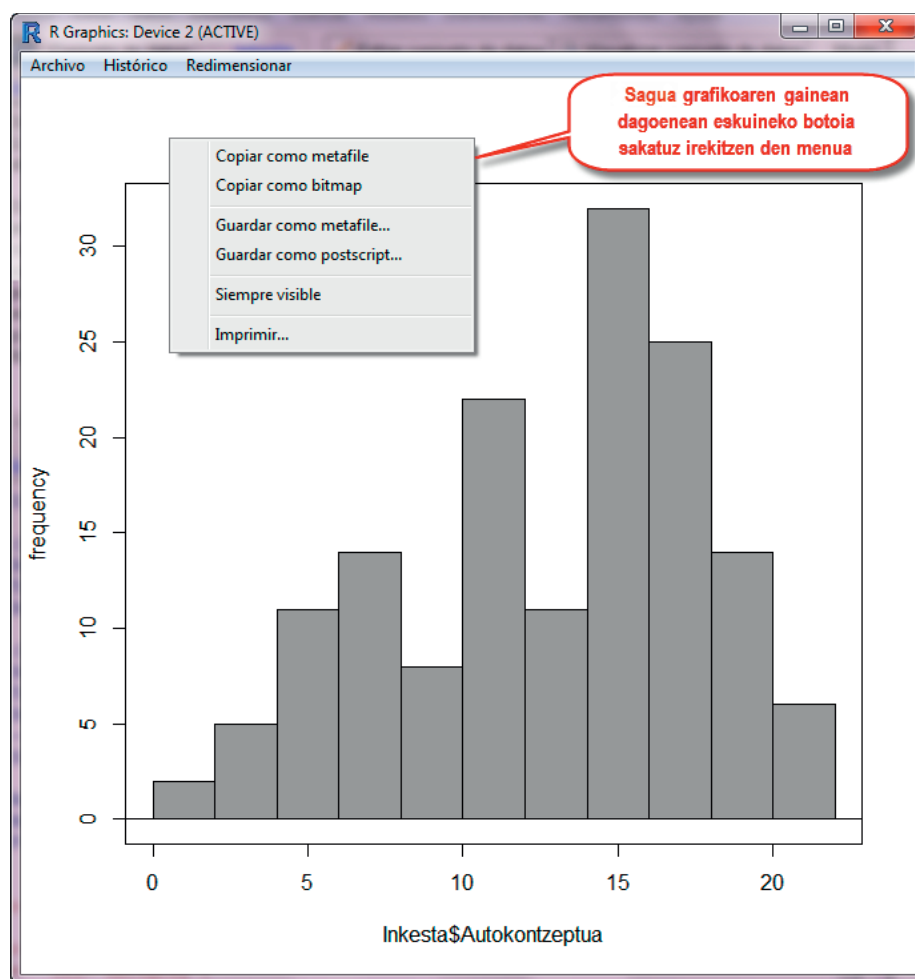
Grafiko mota ugari eskura daiteke modu errazean. Grafikoei dagozkien aukerak erakusteko asmoarekin idatzitako atal honetan, eskuragarriak dauden eredu grafikoak modu laburrean aurkeztuko ditugu.

R Commander-ekin eskuratutako grafikoak R Graphics izeneko gailu (leiho) berri batean aurkezten dira. R Graphics automatikoki aktibatzen da.



156. irudia.

Grafikoa irudikaturik dagoela sagua bere gainean ipintzen badugu eta eskuineko botoia zapaldu, inprimatzeko, kopiaatzeko eta formatu desberdinetan gordetzeko aukerak aurkituko ditugu.



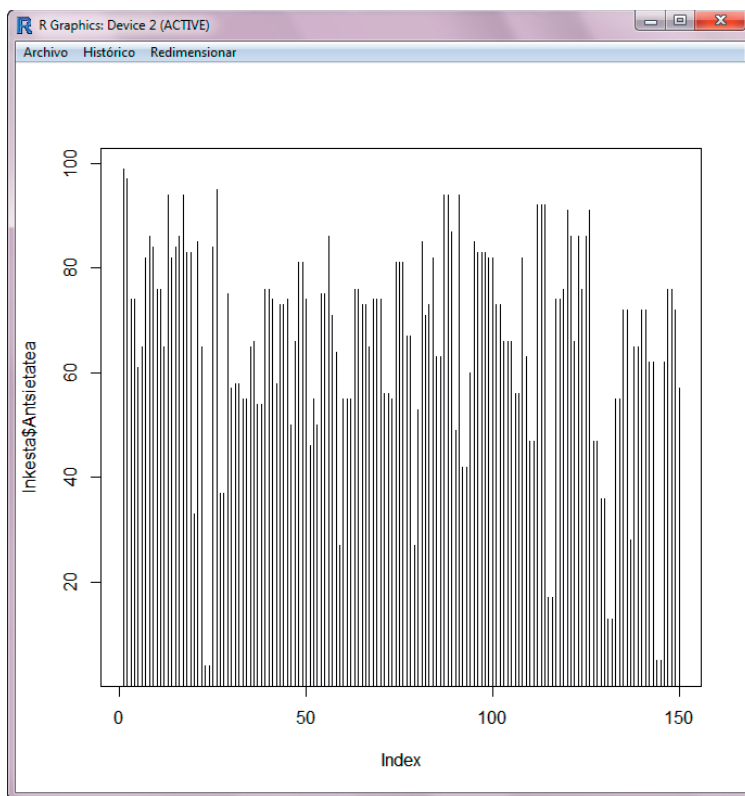
157. irudia.

Sortutako grafikoen formatua oso primarioa da. Nahiz eta grafikoen itxura hobetzeko egin beharreko lana oso sinplea den, gidaliburutxo honetan oinarritzkoenak erakutsiko ditugu soilik. Irakurleak puntu honetan gehiago sakondu nahi izango balu, informazio gehiago bila lezake Elosuaren (2011) lanean.

Grafiko desberdinak eskuratzeko prozesuek antz handia dute. Grafiko desberdinak aurkezteko programak jarraitzen dion ordenari eutsiko diogu.

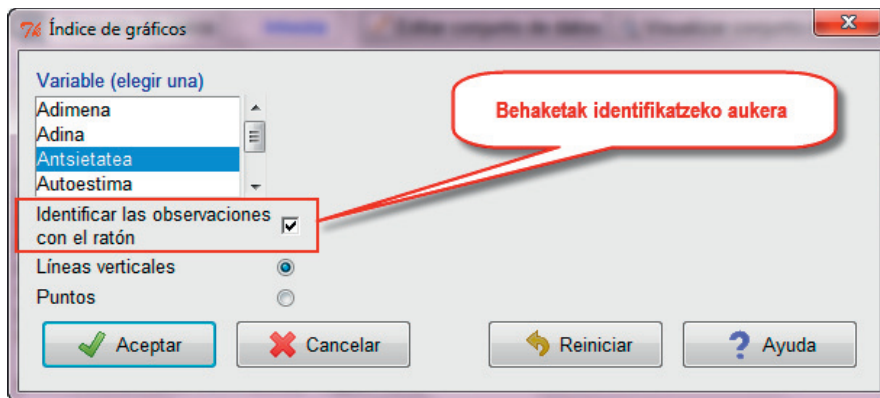
12.1. GRAFIKO SEKUENTZIALA

Grafiko honek datu multzoko ordenari jarraituz, parte-hartzaile bakoitzak hautatu den aldagaian eskuratu duen puntuazioa irudikatzen du. Antsietatea aldagaiaren grafiko sekuentziala eskatzen badugu, hau izango da lortuko dugun emaitza.



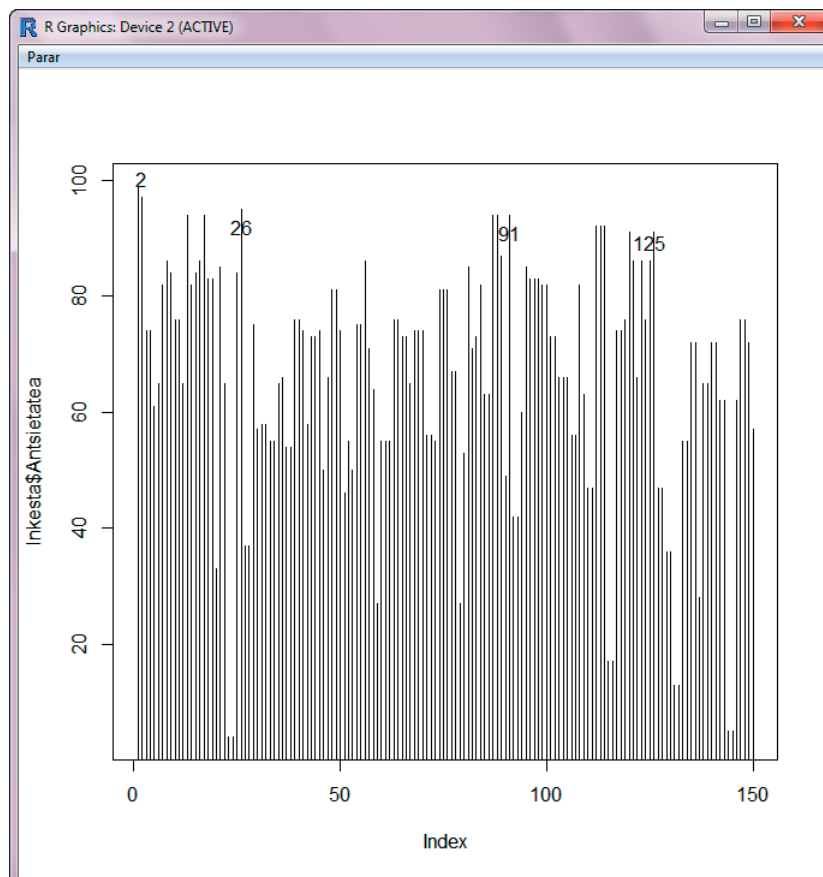
158. irudia.

Programak eskaintzen dituen aukeren artean, ikertzailearentzat interesgarria izan liteke hainbat puntuazio identifikatzeko aukera. Hurrengo leihoan aukera hau markatu dugu puntuen grafiko sekuentzial batentzat (Identificar las observaciones con el ratón). Behatutako balio handienei dagozkien parte-hartzaileak identifikatu ditugu.



159. irudia.

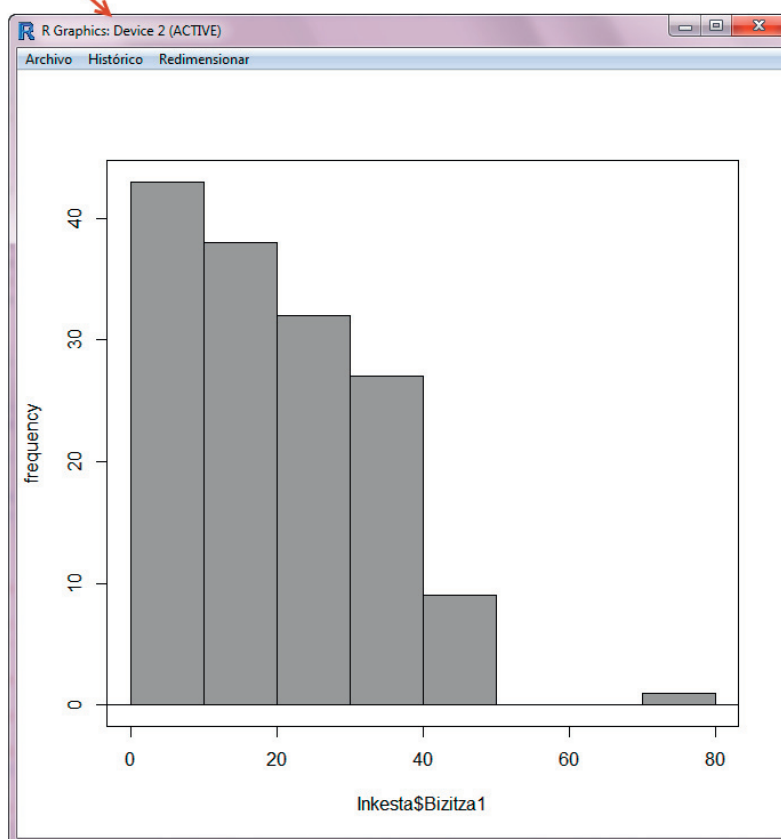
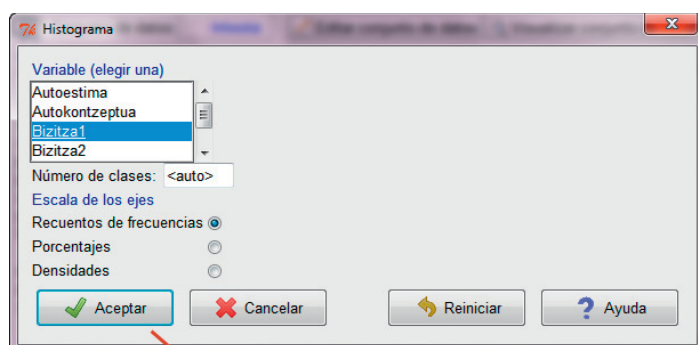
R Commander-ek puntuak irudikatu ostean, hurrengo irudian agertzen den emaitza eskuratzeko saguaren ezkerreko botoiarekin intereseko balioen gainean sakatu dugu. Segidan adierazitako balio bakoitzari dagokion parte-hartzailearen zenbakia eranstean dio programak grafikoari. Klikatzen dugun bakoitzean parte-hartzaileak identifikatzen jarraitu ez dezan, sagua grafikoaren gainean dagoelarik eskuineko botoia sakatu eta `parar` markatu behar da.



160. irudia.

12.2. HISTOGRAMA

Histograma zenbakizko aldagaientzat da egokia. Haren diseinuan balioak zenbat tartetan zatitu nahi diren adieraz daiteke, baita eraikuntzarako irizpideak aukeratu ere: batezbestekoak, portzentajeak edo dentsitatea. Hiru aukerekin lortutako grafikoak baliokideak dira. Lehenengoan, ordenatuen ardatzak maiztasunaren balioari dagokion tartearen zein den adieraziko luke; bigarrenetan portzentajearen balioari dagokiona; eta hirugarrenean lagin guztiarekiko proportziari. Gure kasuan erabilitako aukerak eta emaitza hurrengoak izan dira:



161. irudia.

GRAFIKOEI ERROTULUAK IPINTZEA. Pentsa aurreko grafikoari Autokontzeptua aldagaiaren maiztasunak izenburua ipini nahi diogula eta abzisen eta ordenatuen ardatzei Autokontzeptua eta Maiztasunak deitu nahi diegula hurrenez hurren. Horretarako, azpiko irudian lauki gorri batez inguratuta ageri diren lerroak idatzi beharko genituzke komandoen leihoan:

```
Ventana de instrucciones
Hist(Inkesta$Bizitzal, scale="frequency", breaks="Sturges", col="darkgray")
Hist(Inkesta$Bizitzal, main="Autokontzeptua aldagaiaren maiztasunak",
scale="frequency", xlab="Autokontzeptua", ylab="Maiztasuna"
breaks="Sturges", col="darkgray")
```

162. irudia.

Non:

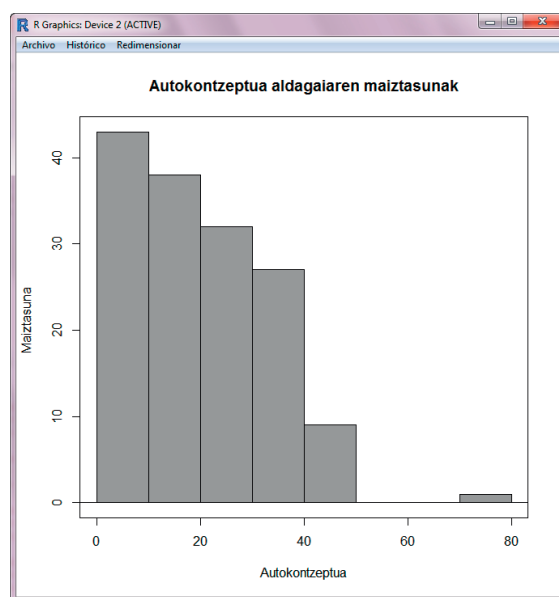
main="Autokontzeptua aldagaiaren maiztasunak": izenburua eransteko komandoa.

xlab="Autokontzeptua": X ardatza izendatzeko komandoa.

ylab="Maiztasuna": Y ardatza izendatzeko komandoa.

Komando osoa aukeratu (kopiatzerakoan bezala) eta sakatu Ejecutar.

KONTUZ! Lehenengo lerroa histograma eskatzerakoan programak besterik esan ezean sortzen duena da. Guk komando hori kopiatu dugu komandoen leihoan eta laukiak inguratutako testua gehitu diogu.



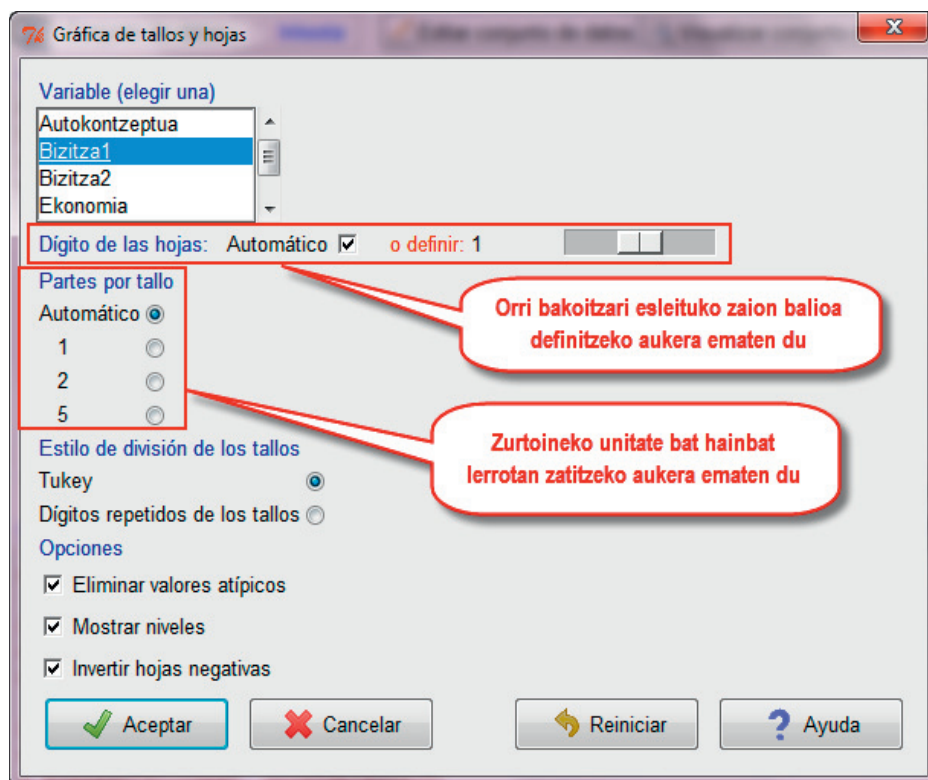
163. irudia.

Komando arrunt hauek programak edozein grafiko eraikitzeke erabiltzen duen komandoekin batera erabili daitezke.

12.3. ADAR ETA HOSTOEN GRAFIKOA

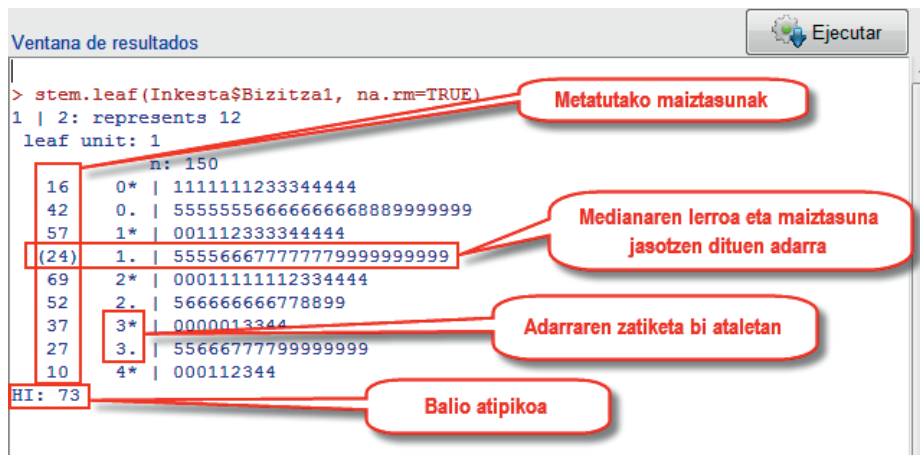
Tukeyri esker estatistika esploratzailea «adin-nagusitasunera» iritsi zen; orduztik, adar eta hostoen grafikoa ospea irabazten aritu da. R Commander-i esker grafiko hau zenbait modutan eta aukera mordo batekin lor daiteke. Bere diseinurako leihoa ireki ostean eta intereseko aldagaia aukeratuta dagoelarik, R Commander-ek hurrengo aukerak eskaintzen ditu:

- Dígito de las hojas. Aldagaia neurtzeko erabilitako eskalaren arabera orri bakoitzari esleitzen zaion balioa aukera daiteke: ehunekoak, hamarrekoak, unitateak... Gizarte Zientzietan erabili ohi dena unitatea da; eta horixe bera da besterik esan ezean R Commander-ek ezartzen duena. Baina balio hori aldatzeko nahikoa da eskuineko aldean ageri den erregeleta mugitzea.
- Partes por tallo. Gure lagineko parte-hartzaile kopurua oso handia balitz, interesekoa izan liteke adarreko unitate bakoitza hainbat lerrotan zatitzea.



164. irudia.

Gaineko irudian markatuta agertzen diren aukerekin, hurrengo izango litzateke irudikapena:



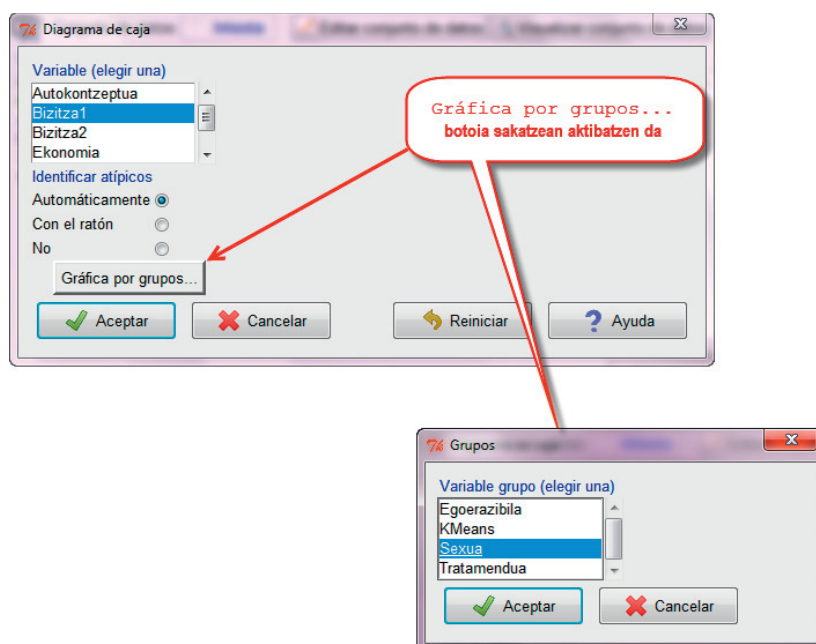
165. irudia.

Aurreko grafikoan laginean Bizitzal aldagaiarentzat eskuratutako balioak ageri dira. Hamarreko bakoitzeko balioak automatikoki banatzeko eskatu zaio R Commander-i. Bi zati sortu dira: bat 5 baino txikiagoak diren balioentzat eta bestea 5 edo handiagoak diren balioentzat. Grafikoak mediana zein lerrotan agertzen den eta honi dagokion parte-hartzaile kopurua (24) ere adierazten du.

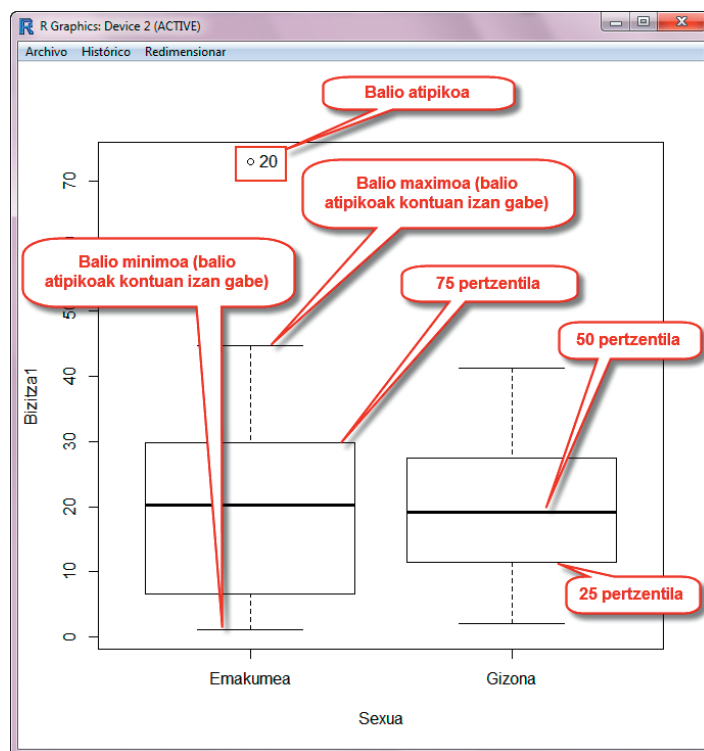
12.4. KAXEN DIAGRAMAK

Kaxen diagramak aldagai baten banaketa erakusten du eta horizontalak diren bost lerroen bidez gutxieneko balioa, lehenengo, bigarren eta hirugarren koartilak eta gehieneko balioa adierazten ditu.

Grafiko berean faktore baten kategoria bakoitzeko kaxaren diagrama bat irudika daiteke. Gure kasuan, gizonen eta emakumeen kaxen diagramak Bizitzal aldagaiaren arabera eskuratu nahi ditugu.



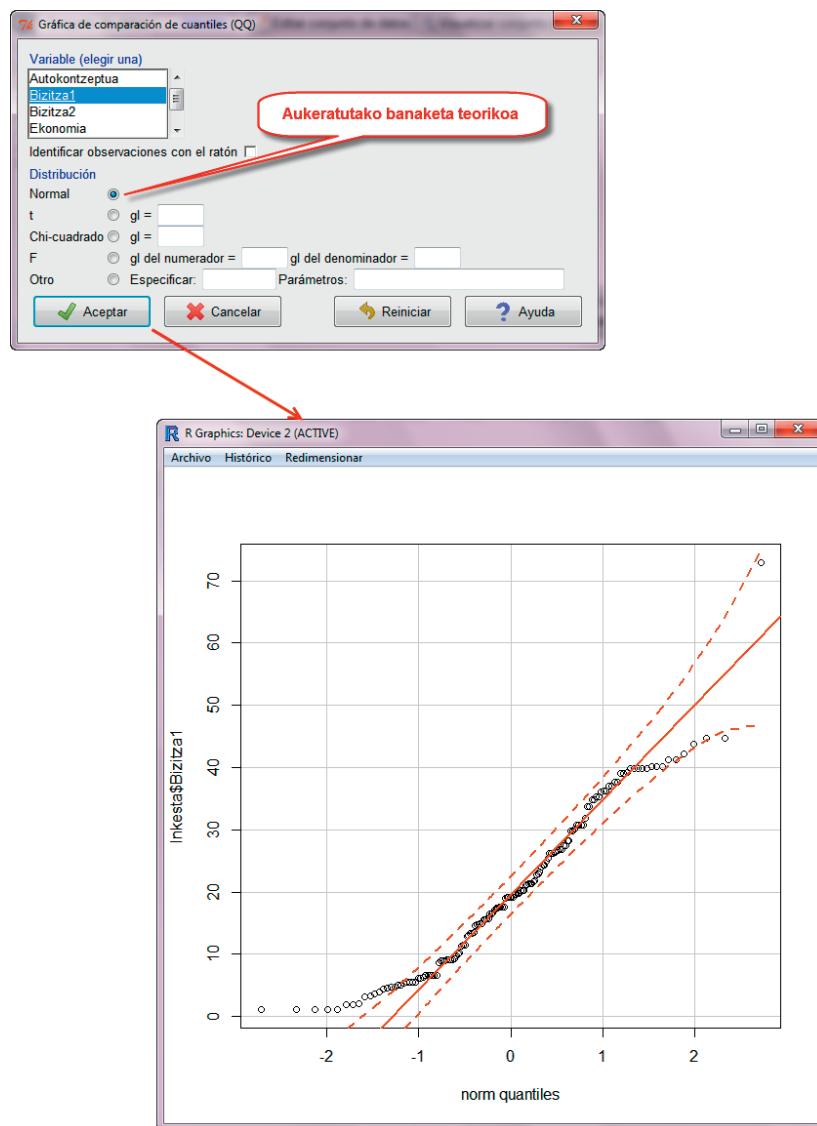
166. irudia.



167. irudia.

12.5. KUANTILAK KONPARATZEKO GRAFIKOA

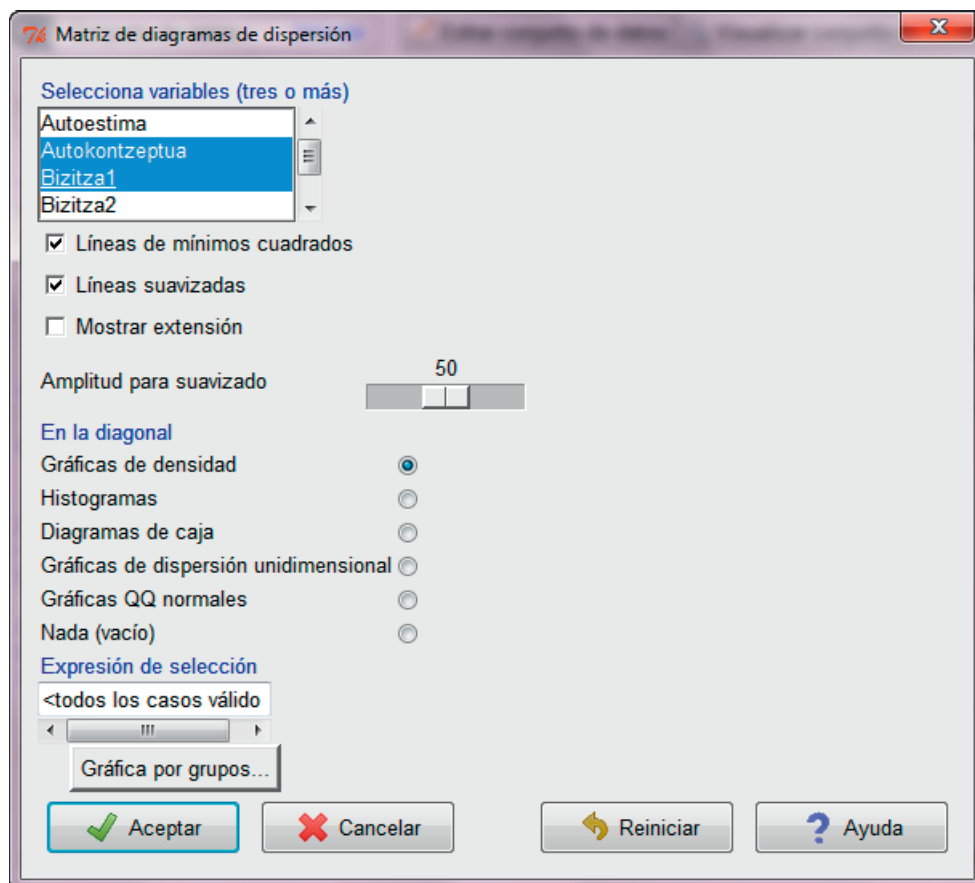
Q-Q grafiko moduan ere ezagutzen diren kuantilak konparatzeko grafiko bidez zenbakizko aldagai baten kuantilak banaketa teoriko baten kuantilekin konpara daitezke. Batzuetan lagingaren banaketak teorikoarekiko duen desbideratzea aztertzeo erabiltzen da; izan ere, lagingeko kuantilen eta teorikoen arteko desberdintasunek aukeratutako banaketarekiko ez dela ongi doitzen adierazten dute. Bizitza1 aldagaiaren kuantilak banaketa normalaren kuantilekin konparatzeko jarraitu beharreko aukera sekuentzia Gráficas > Gráfica de comparación de cuantiles... da.



168. irudia.

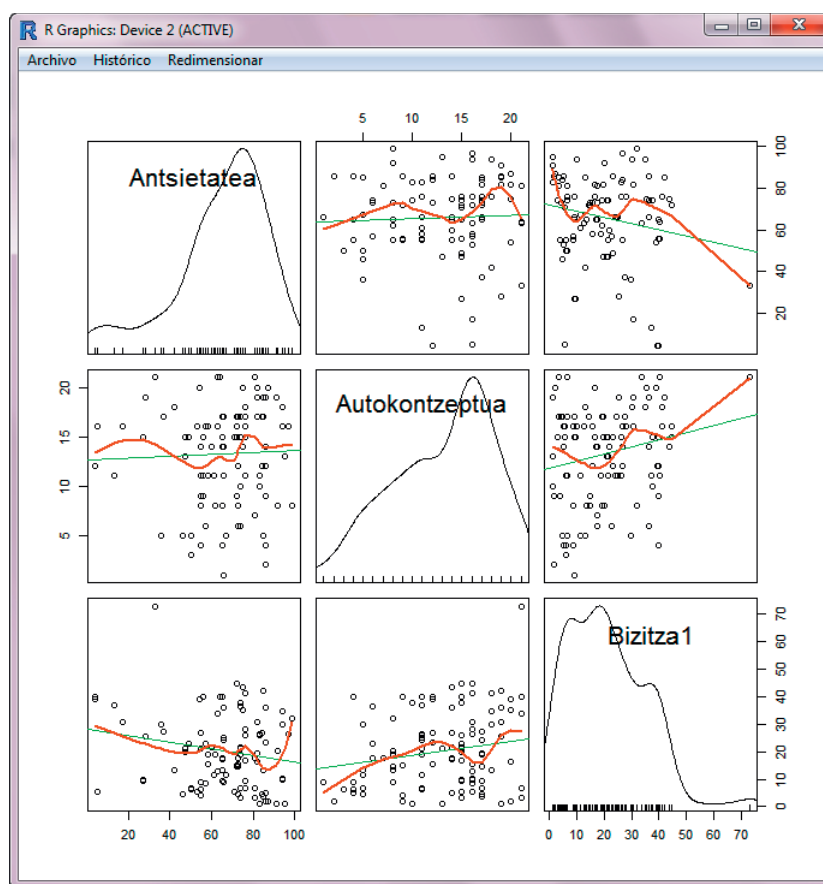
12.6. BANAKETAREN DIAGRAMA. BANAKETEN DIAGRAMEN MATRIZEA

R Commander-ek Diagrama de dispersión eta Matriz de diagramas de dispersión aukerak eskaintzen dizkigu banaketen diagramak eskuratzeko. Diagramas de dispersión sakatuta bi zenbakizko aldagaien informazioa jasotzen duen grafikoa sor daiteke. Matriz de diagramas de dispersión delakoarekin, aldiz, hiru aldagai edo gehiago gurutzatuz lortzen den aldagai bikoteko diagrama bana sor daiteke. Horretarako nahikoa da Gráficas > Matriz de diagramas de dispersión edo Gráficas > Diagrama de dispersión sakatu ondoren irekiko den leihoan zenbakizko aldagaien artean gurutzatu nahi ditugunak aukeratzea.



169. irudia.

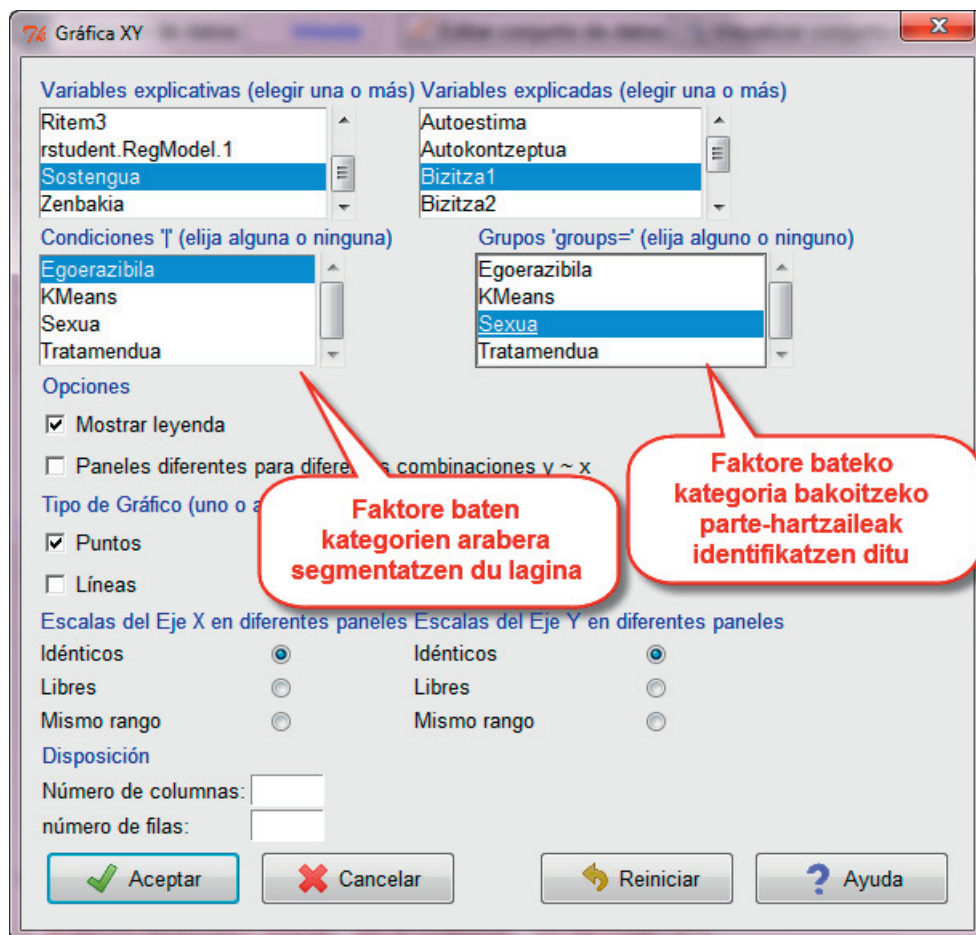
Aurreko leihoan ikus daitekeen moduan, banaketen diagramen matrizeari eusten badiogu, diagonal nagusian dentsitatearen grafikoa, histograma, kaxen diagrama... marrazteko eska diezaiokegu R Commander-i.



170. irudia.

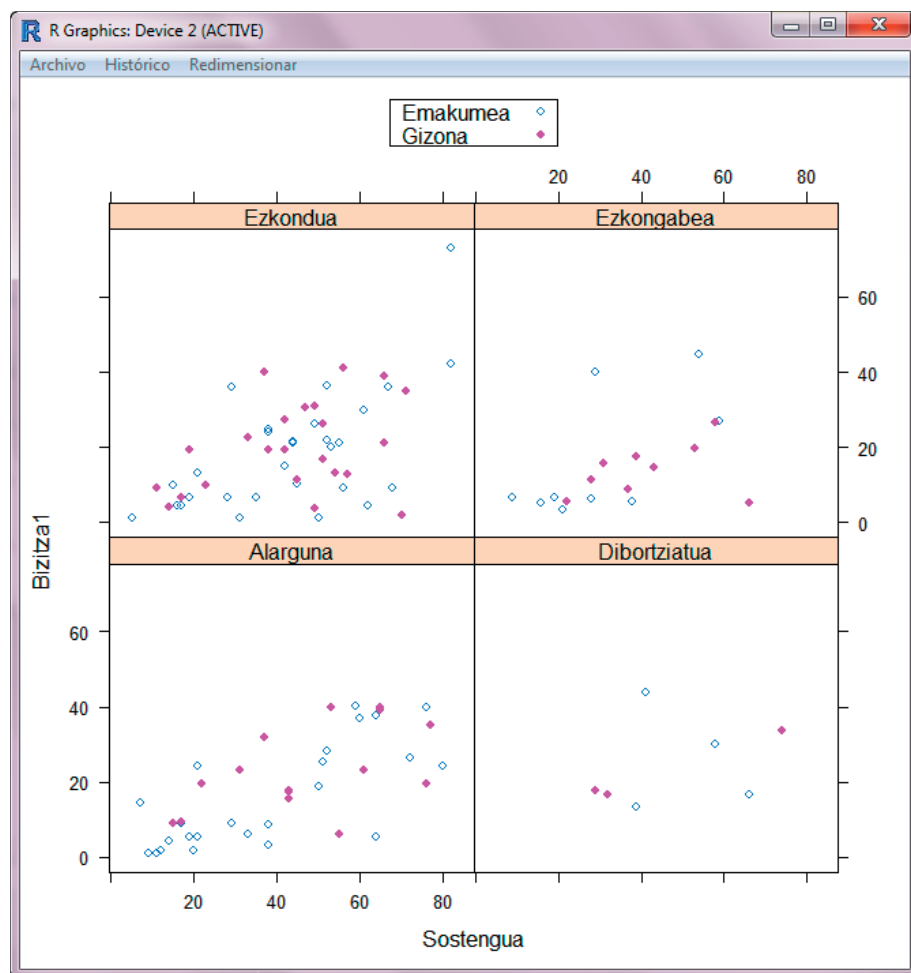
12.7. XY GRAFIKOA

Gráfica XY aukeraren bidez banaketen diagramak (edo banaketen diagramen matrizeak) aukera eta formatu desberdinekin eskura daitezke. Baina aurrez aipatutako banaketen diagramen (eta banaketen diagramen matrizeen) aldean, XY grafikoak faktore baten maila bakoitzeko gurutzatutako aldagaien banaketa bana erakusten du. Gainera, banaketan irudikatutako balioen formatuak beste faktore baten arabera bereiz daitezke. Adibide moduan, egoera zibilaren (Egoerazibila) arabera Sostengua eta Bizitza1 aldagaien banaketa irudikatzeko eta banaketa horietan gizonak eta emakumeak desberdintzeko eskatu diogu R Commander-i. Horretarako, Gráficas > Gráfica XY... sakatu ostean hurrengo irudian ageri diren aukerak markatu ditugu:



171. irudia.

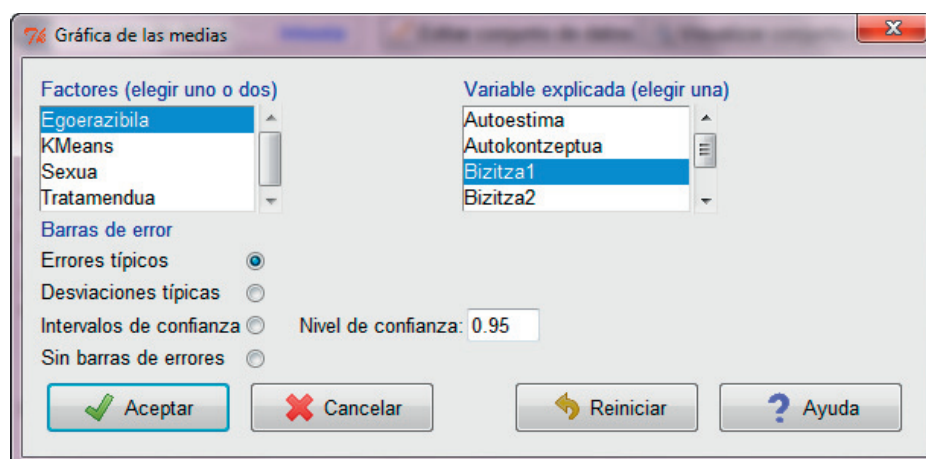
Hau da lortu dugun XY grafikoa:



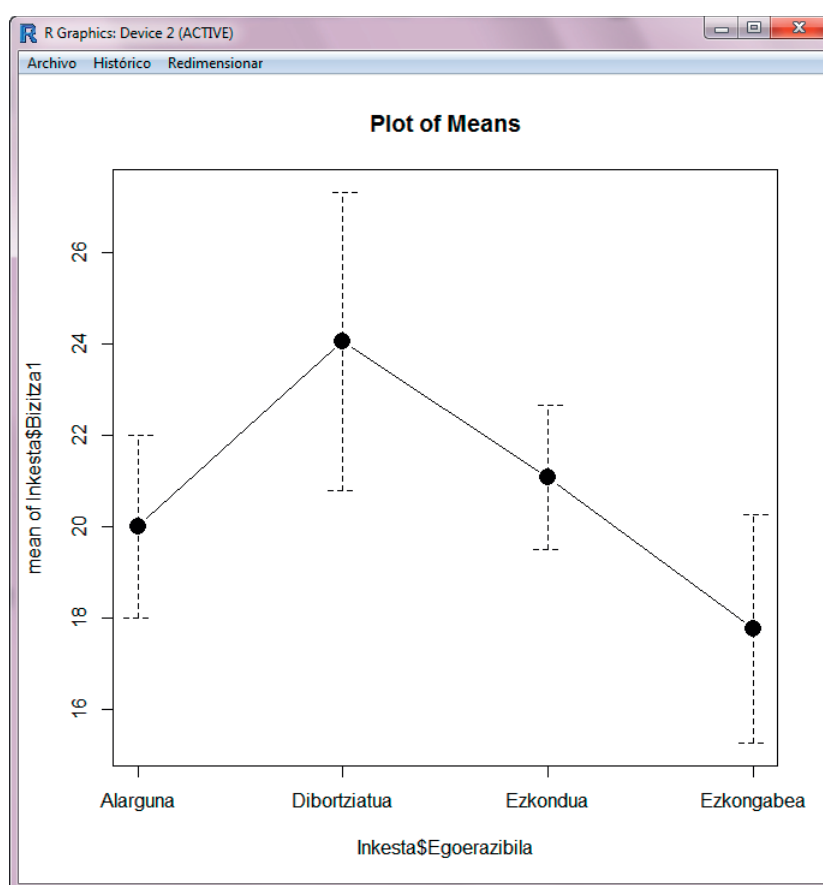
172. irudia.

12.8. BATEZBESTEKOEN GRAFIKOA

Gráfica de las medias delakoarekin faktore baten maila bakoitzak zenbakizko aldagai batean duen batezbestekoa marraztu daiteke. Adibiderako, faktore eta zenbakizko aldagai moduan, hurrenez hurren, Egoerazibila eta Bizitza1 aldagaiak aukeratu ditugu. Batezbestekoen irudikapenak errore, errore tipikoei, desbideratze tipikoei edo maila zehatz bateko (besterik esan ezean % 95) konfiantza-tarteei dagozkien barrekin osatzen dira. Sakatu beharreko aukera sekuentzia Gráficas > Gráfica de las medias... da.



173. irudia.

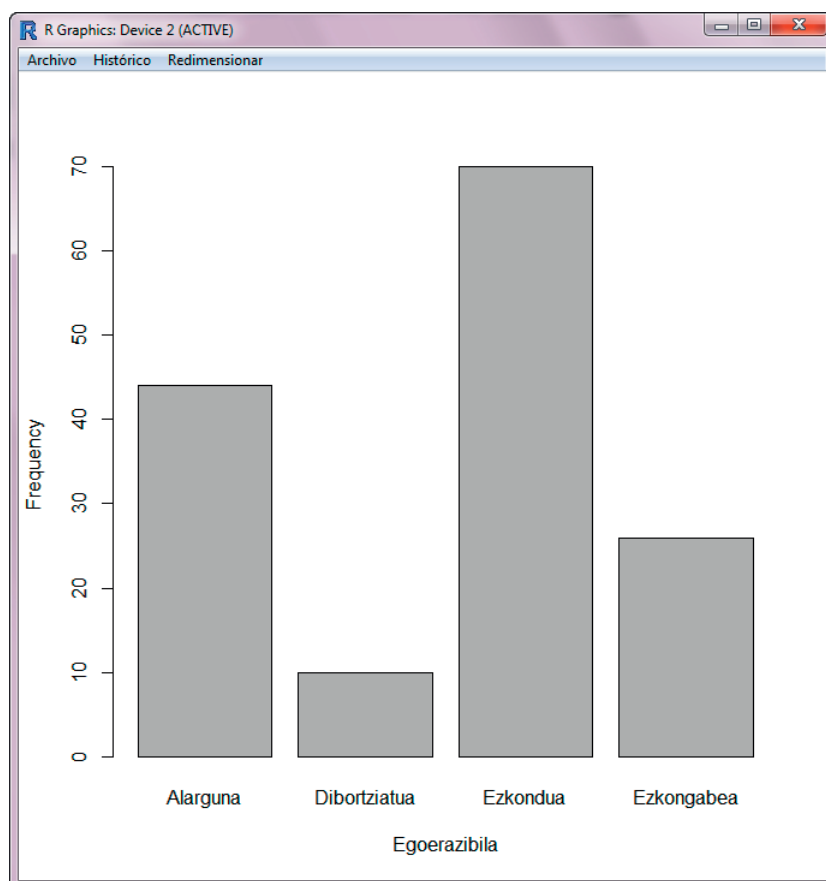


174. irudia.

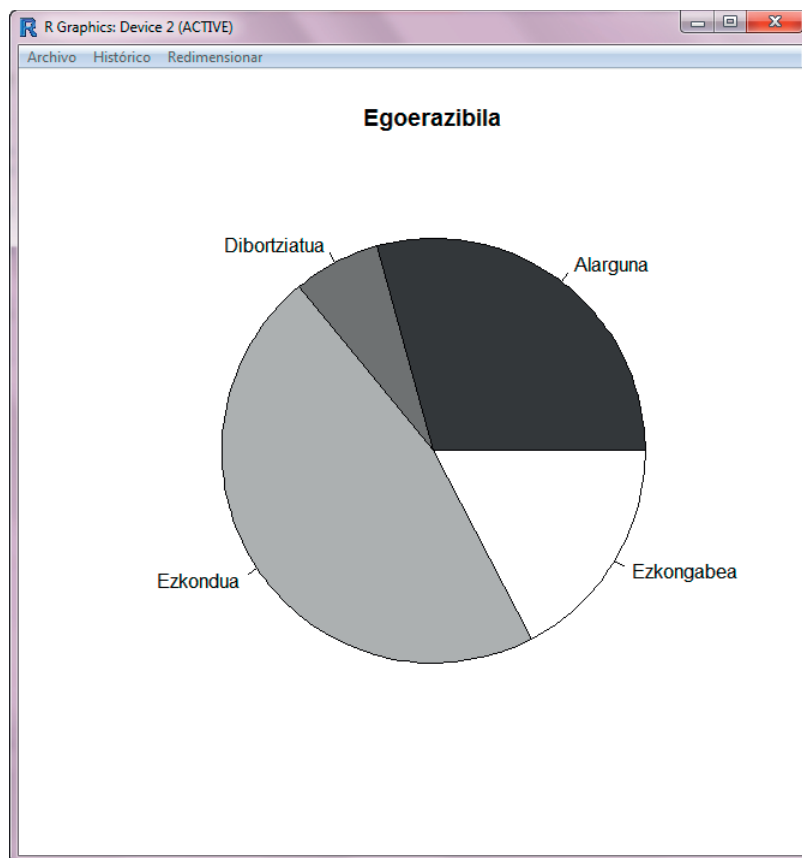
12.9. BARREN ETA SEKTOREEN GRAFIKOA

Gráfica de barras eta Gráfica de sectores direlakoekin faktore bati dagozkion barren edo sektoreen diagramak lor daitezke. Horretarako nahikoa da Gráficas > Gráfica de barras edo Gráficas > Gráfica de sectores... sakatzea eta intereseko faktorea markatzea.

Jarraian Egoerazibila aldagaiaren irudikapenak aurkezten dira:



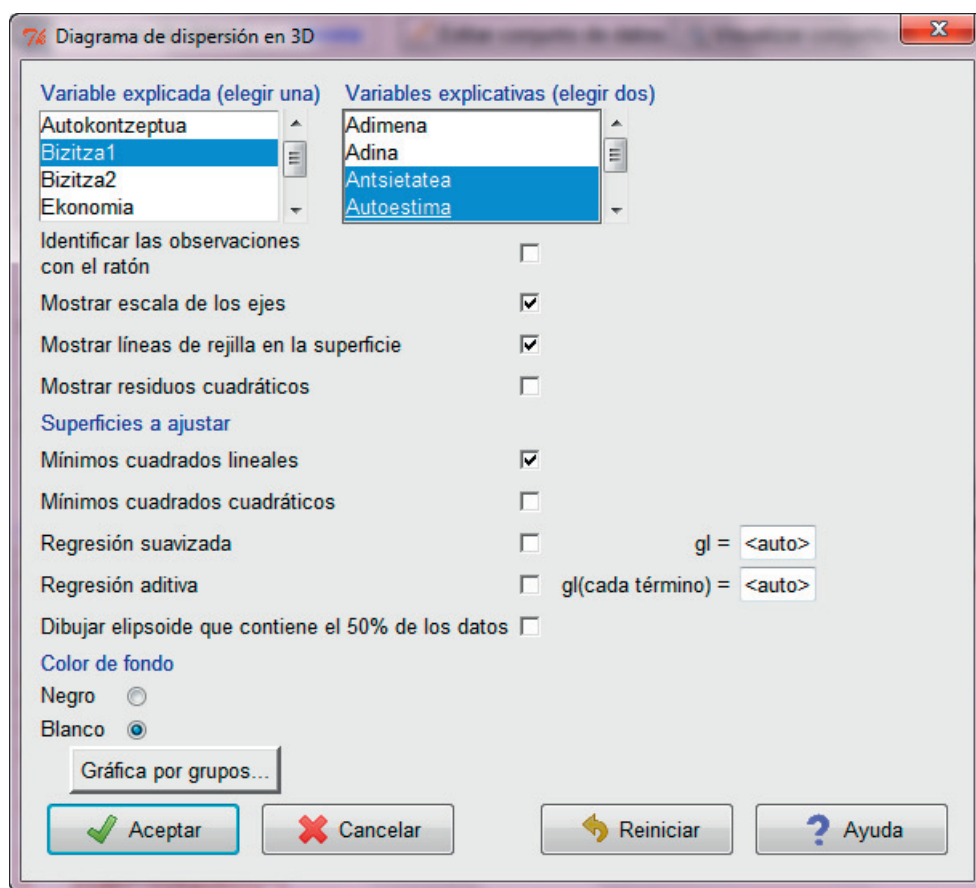
175. irudia.



176. irudia.

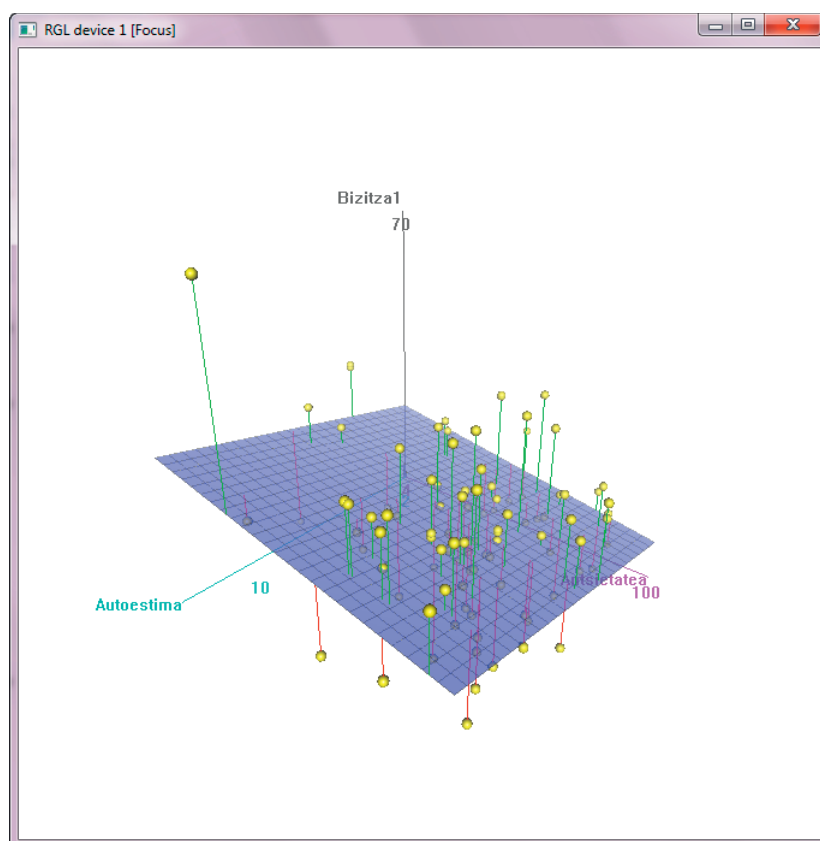
12.10. 3D GRAFIKOAK

Gráficos 3D aukerarekin hiru dimentsioko grafikoak sor daitezke. Besteak beste, gutxiengo karratuen bidez neurtuta mendeko aldagai batek (aldagai azaldua) beste bi aldagai independenteekiko (aldagai azaltzaileak) duen doiketa irudika daiteke. Horregatik, aukeren artean erregresioaren planoak —gutxiengo karratu linealen eta gutxiengo karratu kuadratikoen zentzuan, nahiz leundutako erregresioaren eta erregresio gehigarriaren zentzuan—, irudikatzeko aukerak aurkituko ditugu. 3D grafikoak sortzeko jarraitu beharreko aukera-sekuentzia hurrengoa da: Gráficas > Gráficos 3D > Diagrama de dispersión en 3D...



177. irudia.

Aurreko irudian adierazitako aukerak markatuz gero, hau izango litzateke lortuko genukeen grafikoa:

**178. irudia.**

Grafikoa eskuratu ostean, sagua gainean ipintzen badiogu, ezkerreko botoia sakatuta daukagun bitartean sagua mugituz grafikoa mugi eta bira daiteke. Erreminta honi esker «puntuen lainoaren» azterketan sakon dezakegu, balio atipikoak identifikatu...

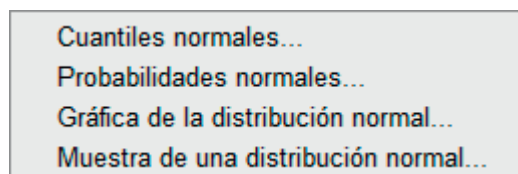
13. Banaketak

Gizarte Zientzietara Aplikatutako Estatistikan erabilienak diren probabilitatearen banaketa teorikoekin lan egitea ahalbidetzen du R Commander-ek. Probabilitatearen banaketa teoriko bakoitzarentzako probabilitateak (taulak ez dira beharrezkoak) eta grafikoak ematen ditu eta zorizko laginak simulatzen ditu (sasizorizkoak zentsu hertsian).

Banaketa teorikoak bi ataletan banatuta aurkituko ditugu: jarraituak eta diskretuak. Aldagai jarraiekin lan egiteko eskainitako aukera guztiak berdinak direnez, banaketa normalari (*Distribución normal*) dagozkionak aurkeztuko ditugu. Modu berean, banaketa diskretuen artean banaketa binomiala aukeratu dugu (*Distribución binomial*).

13.1. BANAKETA NORMALA

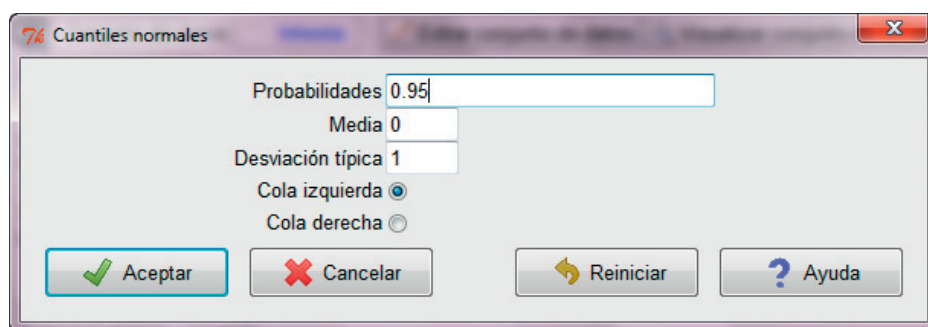
R Commander-ek banaketa normalarekin lan egiteko lau modu biltzen ditu. Aukera horiek hemen aurkituko ditugu: *Distribuciones > Distribuciones continuas > Distribución normal*.



179. irudia.

13.1.1. Kuantil normalak

Behin banaketaren parametroak (batezbestekoa eta desbideratze tipikoa) finkatuta daudela, banaketa normalean zehaztutako probabilitatearen ezkerreko aldean edo eskuineko aldean uzten duen balioa eskuratzen da. *Cuantiles normales...* aukera sakatzean irekiko den leihoan, hurrengo irudian ikus daitekeen moduan, batezbestekoari (μ) 0 balioa eta desbideratze tipikoari (σ) 1 balioa esleitu dizkiegu, eta 0.95eko probabilitatearekin ezkerreko aldean geratzen den aldagaiaren balioa erabili dugu (Cola izquierda).



180. irudia.

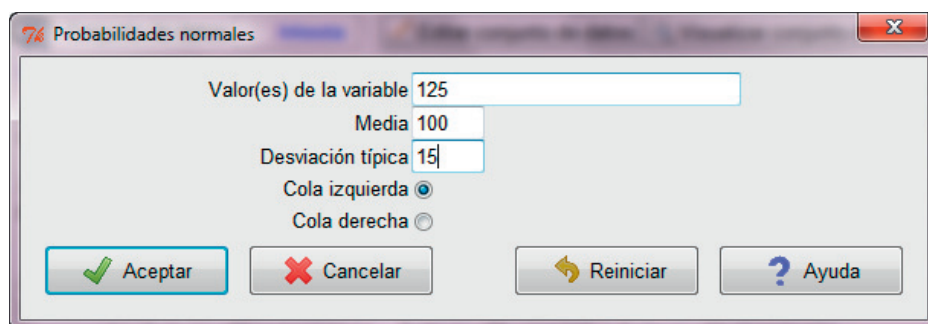
Emaitzen leihoan dagokion balioa zein den ikus daiteke: 1.644854.



181. irudia.

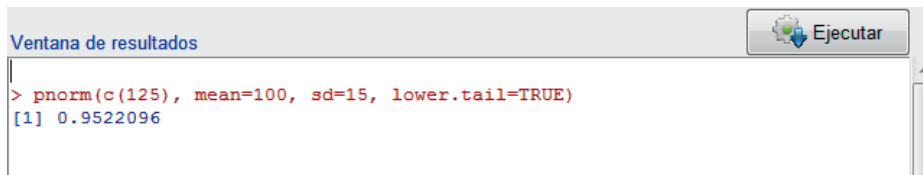
13.1.2. Probabilitateen kalkulua

Behin banaketaren parametroak finkatu ditugula, R Commander-ek emandako balioa baino txikiagoak edo handiagoak diren balioak lortzeko probabilitatea kalkulatu du. Pentsa batezbestekoaren balioetat eta desbideratze tipikotzat 100 eta 15 balioak dituen adimen-maila aldagaiarekin gabiltzala lanean eta adimen-maila aldagaian 125eko puntuazioa lortu duen pertsonari dagokion pertzentila zein den jakin nahi dugula. Horretarako Probabilidades normales... aukera sakatuko dugu eta jarraian ageri diren aukerak markatuko ditugu.



182. irudia.

Emaitzen leihoan lortu den probabilitatea 0.9522096 (P_{95}) dela ikus daiteke.

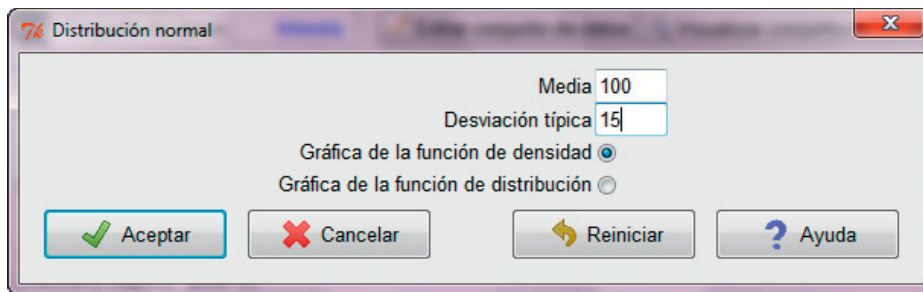


```
> pnorm(c(125), mean=100, sd=15, lower.tail=TRUE)
[1] 0.9522096
```

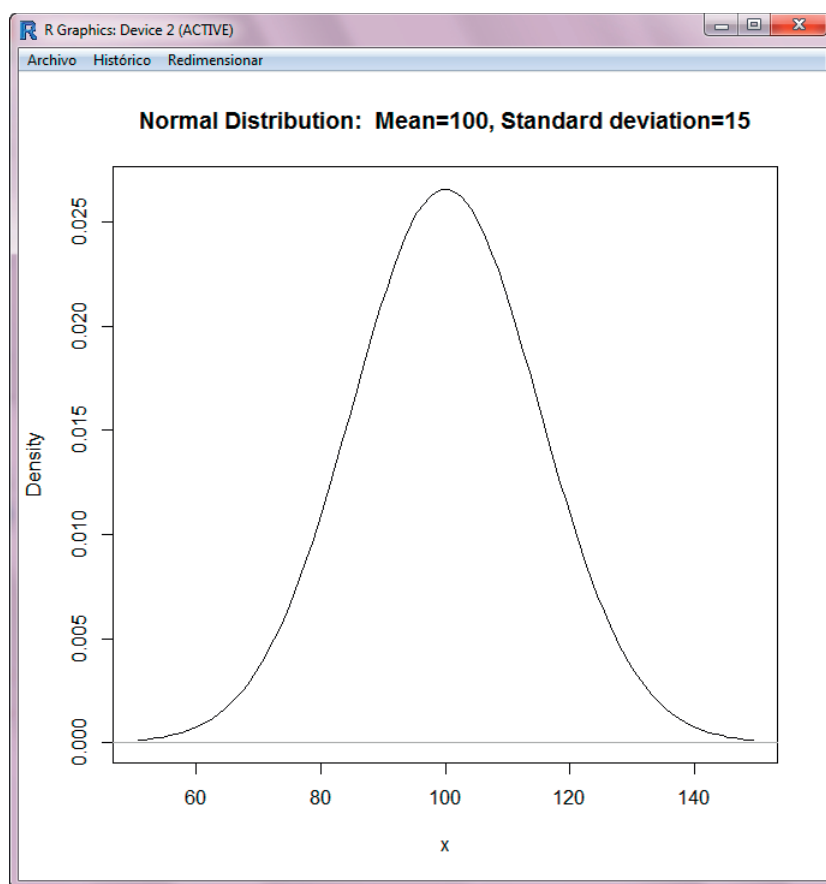
183. irudia.

13.1.3. Banaketa normalaren grafikoa (dentsitatearen funtzioa)

Behin banaketaren parametroak (batezbestekoa eta desbideratze tipikoa) definitu direla, dentsitatearen eta banaketaren grafikoak eskura daitezke Gráfica de la distribución normal... aukerarekin. Adbidean (100,15) parametroen dentsitatearen funtzioa irudikatzeko eskatu diogu R Commander-i.



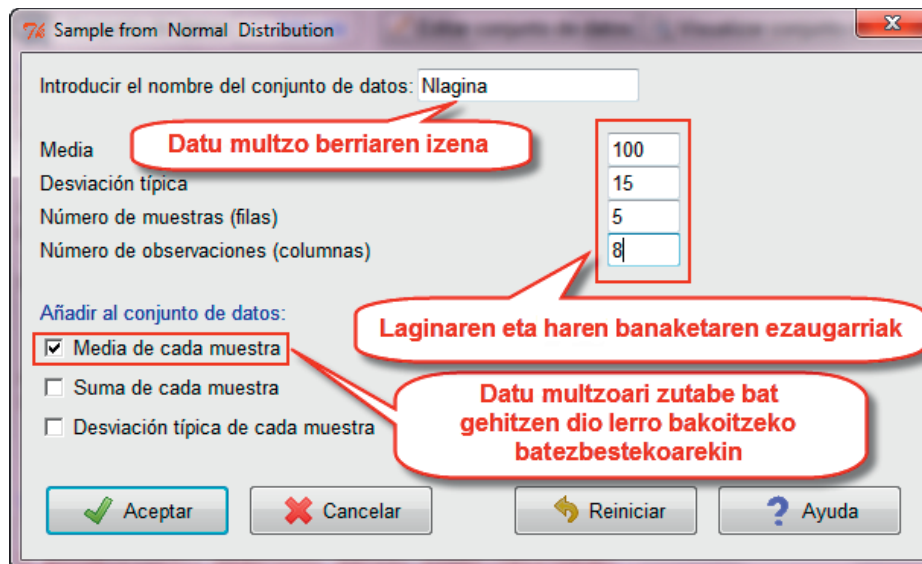
184. irudia.



185. irudia.

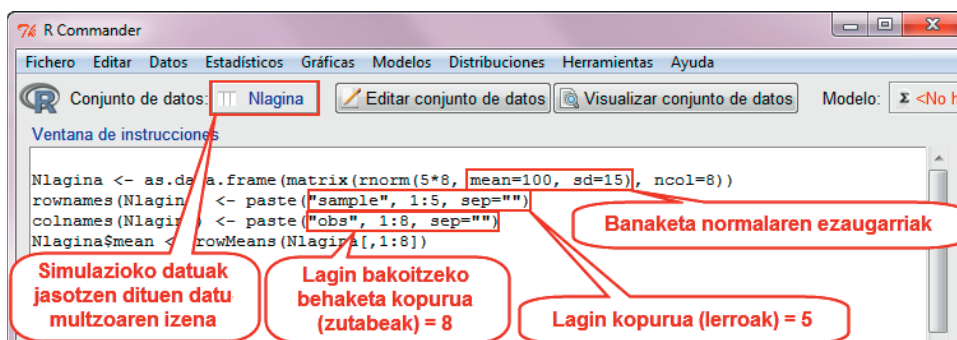
13.1.4. Banaketa baten zorizko lagina

Aukera honen bidez, aurreko kasuetan bezala aurrez parametroak adierazita daudela, banaketa teorikoaren zorizko laginak (sasizorizkoak) lor daitezke. Prozesua oso sinplea da. Pentsa zorizko 5 lagin lortu nahi ditugula, bakoitza behatutako 8 baliorekin. Laginak eskatzeko, lehenik eta behin prozesuaren ondorioz sortuko den datu multzoari izen bat esleitu behar zaio; gure kasuan `Nlagina` izendatuko dugu. R Commander-ek sortuko dituen balioak datu multzo honetan gordeko ditu. Ondoren, banaketaren parametroak zehaztu behar dira; hortaz, batezbestekoari (μ) 100 balioa esleituiko diogu eta desbideratze tipikoari (σ) 15 balioa. Azkenik, sortzen dituen laginen batezbestekoak eska diezazkiokegu.



186. irudia.

Zorizko laginak eskuratu ostean Nlagina aktibatutako datu multzo bihurtuko da. Komandoen leihoan sortutako laginen ezaugarriak agertuko dira.



187. irudia.

Sortutako balioak ikusteko, nahikoa da leiho nagusiko Visualizar conjunto de datos aukeran sakatzea.

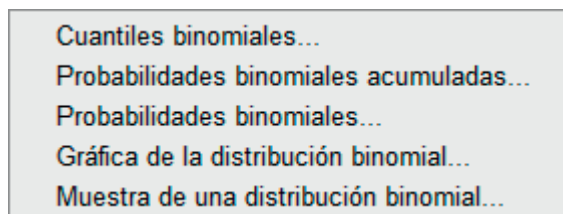
	obs1	obs2	obs3	obs4	obs5	obs6	obs7	obs8	mean
sample1	97.95965	98.58108	109.90606	105.75110	89.46938	75.13407	107.12167	113.28831	99.65141
sample2	91.76045	104.91008	94.09278	99.55168	104.24444	104.69736	90.22478	107.91156	99.67414
sample3	85.97685	108.07222	109.62074	96.90637	77.46629	94.51706	88.58585	96.57087	94.71453
sample4	89.64857	109.28528	97.37850	84.19192	119.87274	100.99834	102.84509	103.10522	101.54071
sample5	98.87698	96.67889	78.89812	107.18824	100.17456	100.17456	100.17456	100.17456	93.03029

Callout: Lagin bakoitzeko batezbestekoak

188. irudia.

13.2. BANAKETA BINOMIALA

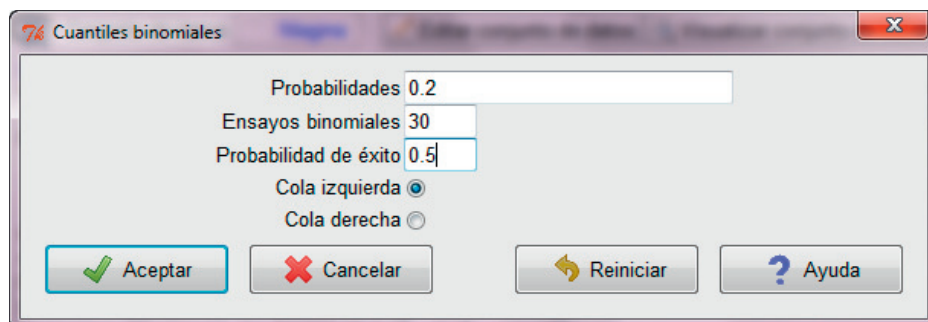
Banaketa binomialentzako aukerak Distribuciones > Distribuciones discretas > Distribución binomial sakatuz aurkituko ditugu. Hurrengoak dira: kuantil binomialak (Cuantiles binomiales...), metatutako probabilitate binomialak (Probabilidades binomiales acumuladas...), probabilitate binomialak (Probabilidades binomiales...), banaketa binomialaren grafikoa (Gráfica de la distribución binomial...) eta banaketa binomial baten lagina (Muestra de una distribución binomial...).



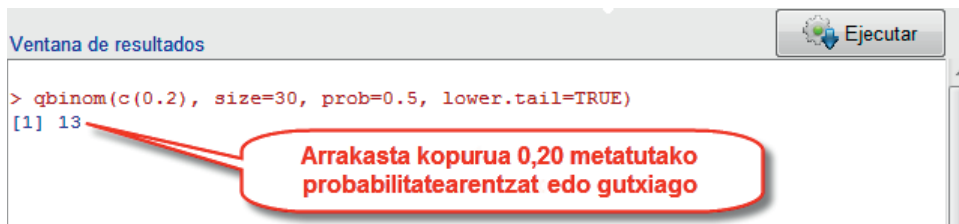
189. irudia.

13.2.1. Kuantil binomialak

Hasteko, aukeratutako banaketaren parametroak zehaztu behar ditugu. Kasu honetan saio kopurua (Ensayos binomiales) eta bakoitzaren arrakasta-probabilitatea (Probabilidad de éxito). Adibidean 30 balioa erabili dugu lehenengoarentzat eta 0.5 bigarrenagoarentzat. Cuantiles binomiales... aukeraren bidez, adierazitako banaketa binomialean metatutako probabilitatearen (Probabilidades) eskuineko edo ezkerreko aldeari (Cola izquierda edo Cola derecha) dagokion arrakasta kopurua kalkulatu du R Commander-ek. Adibidean, metatutako probabilitateari 0.20 balioa esleitu diogu.



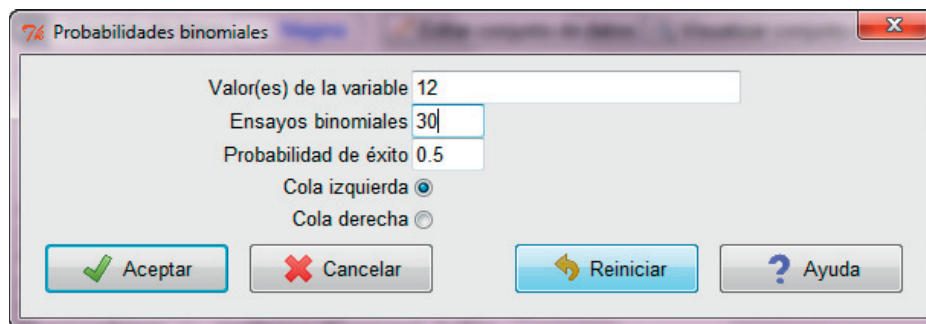
190. irudia.



191. irudia.

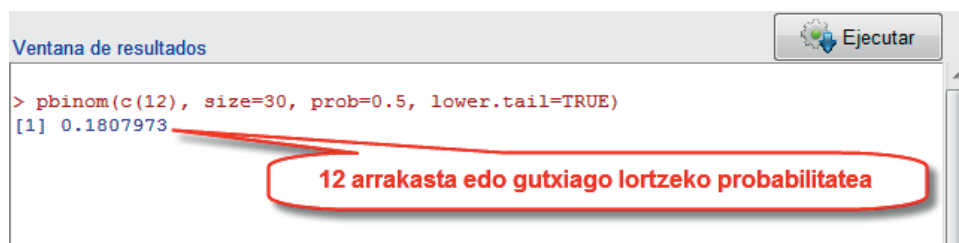
13.2.2. Metatutako probabilitate binomialak

k arrakasta edo gutxiago edo k arrakasta baino gehiago lortzeko probabilitatea kalkula daiteke Probabilidades binomiales acumuladas... aukerarekin. Horretarako, lehenik eta behin banaketaren parametroak aukeratu behar ditugu eta arrakasta kopurua (Valor(es) de la variable) adierazi behar dugu. Hurrengo leihoan 0.5eko arrakasta-probabilitatea eta 30 saio finkatu ditugu. R Commander-i 12 arrakasta edo gutxiago lortzeko probabilitatea kalkulatzeko eskatu diogu.



192. irudia.

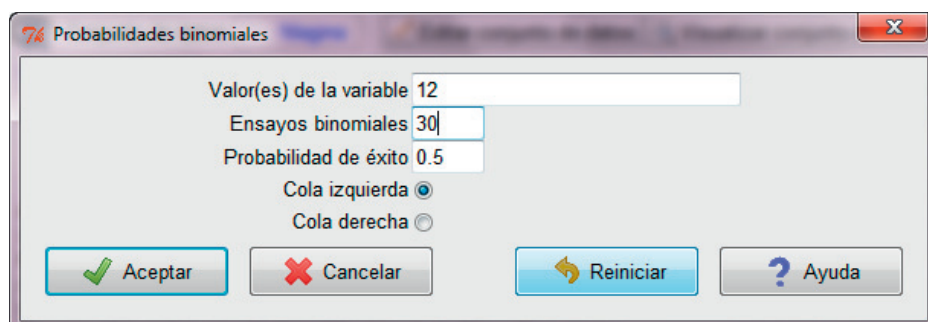
Eraitza 0.1807973 da.



193. irudia.

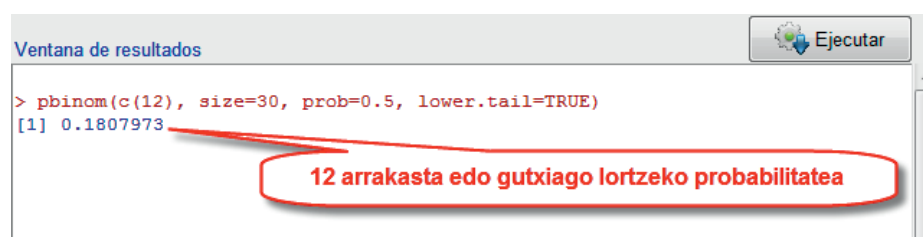
13.2.3. Probabilitate binomialak

Probabilidades binomiales... aukerak probabilitateen taula eskuratzeko aukera ematen du. Saio kopurua eta arrakasta-probabilitatea zehaztuta, arrakasta kopuru bakoitzari dagokion probabilitatea biltzen duen taula aurkezten du. Hurrengo adibidean, 10 saio eta 0.5eko arrakasta probabilitatea finkatu ditugu.



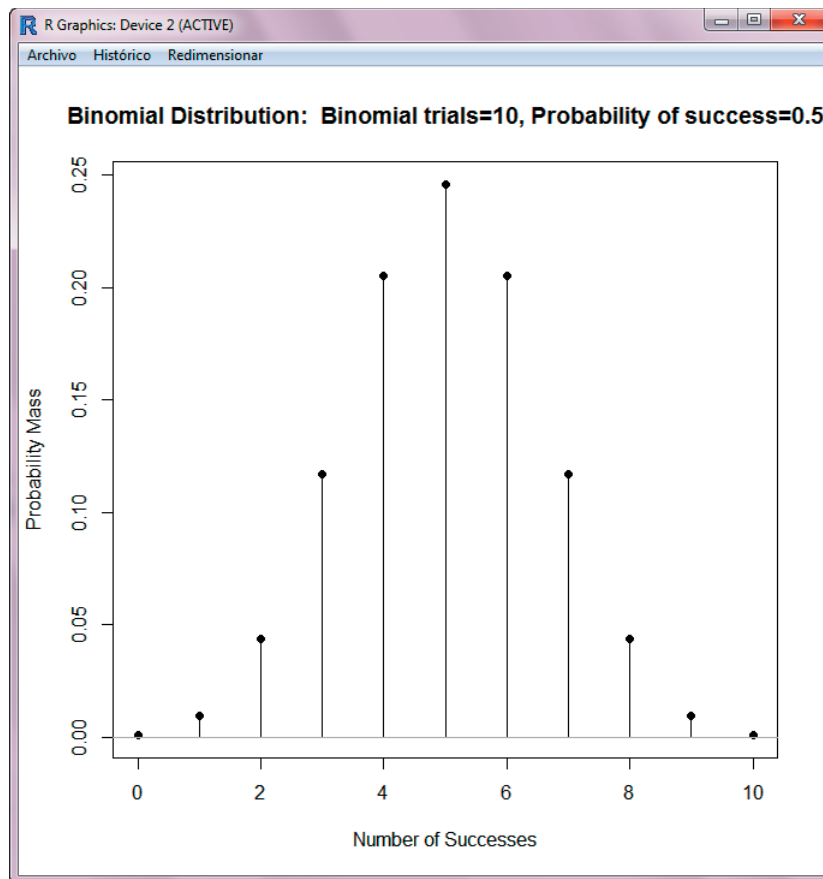
194. irudia.

Emaitzen leihoan probabilitateen taula agertuko da. Adibidez, 3 arrakasta eskuratzeko probabilitatea (0.1171) markatu dugu.



13.2.4. Banaketa binomialaren grafikoa

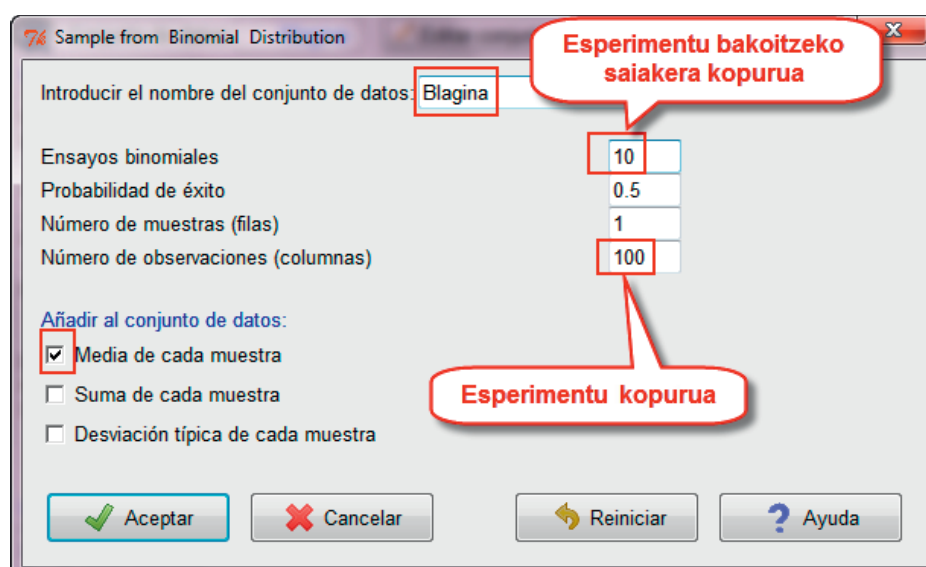
Intereseko saio kopurua eta arrakasta-probabilitatea dituen banaketaren grafikoa lor daiteke Gráfica de la distribución binomial... aukerarekin. Saio kopurua 10 izanik eta arrakasta-probabilitatea 0.50 hurrengo grafikoa lortuko dugu:



195. irudia.

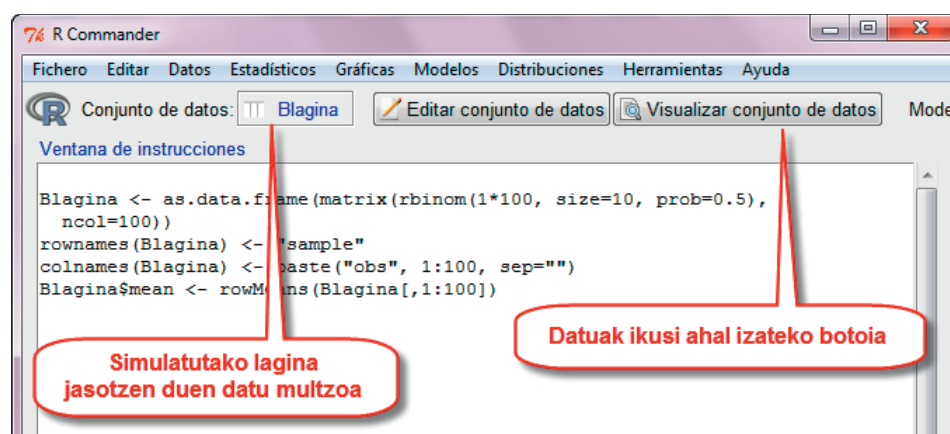
13.2.5. Banaketa binomial baten lagina

R Commander-ekin banaketa binomialaren arabera banatzen diren populazioetako laginak lor daitezke Muestra de una distribución binomial... aukeraren bidez. Laginak lortzeko prozesuaren hasieran R Commander-i eskatutako laginak jasoko dituen datu multzo berriari izen bat esleitu behar zaio; guk Blagina izendatuko dugu. Jarraian simulazioaren ezaugarriak finkatuko ditugu: aldi bakoitzeko saio binomialen kopurua (Ensayos binomiales), arrakasta-probabilitatea (Probabilidad de éxito), lerro kopurua definituko duen lagin kopurua (Número de muestras (filas)) eta zutabe kopurua definituko duen laginaren tamaina (Número de observaciones (columnas)). Adibidean erabilitako balioak hurrengo irudian ageri dira.



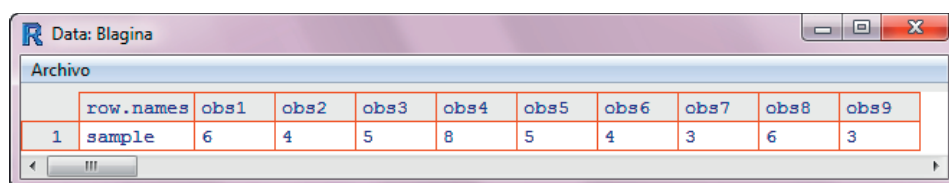
196. irudia.

Simulazioa burutu ostean Blagina datu multzoa eskuratuko dugu. Orain, hau izango da aktibatutako datu multzoa.



197. irudia.

Eskuratutako lagina ikusteko nahikoa da Editar conjunto de datos, edo Visualizar el conjunto de datos sakatzea.



	row.names	obs1	obs2	obs3	obs4	obs5	obs6	obs7	obs8	obs9
1	sample	6	4	5	8	5	4	3	6	3

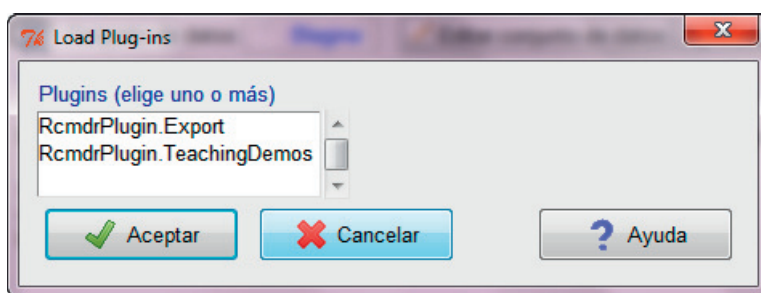
198. irudia.

Ez ahaztu eskuratutako datu multzoarekin berriro lanean aritzeko beharrezkoa dela gordetzea. Horretarako aukera sekuntzia Datos > Conjunto de datos activo > Guardar el conjunto de datos activo... da.

14. TeachingDemos. Estatistika irakasteko erreminta

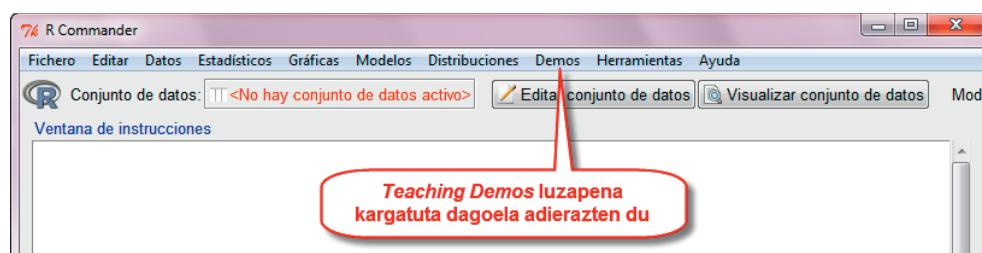
R Commander osatzeko diseinatutako mota desberdineko luzapen (plug-in) ugari dago. Atal honetan estatistikaren irakaskuntzarako bereziki interesgarria iruditzen zaigun luzapen horietako bat aurkeztuko dugu: TeachingDemos. Pakete honek jasotzen dituen aukerek posible egiten dute oinarritzko hainbat kontzeptu estatistiko ikasleei modu erakargarrian azaltzea.

Gidaliburu honen lehen atalburuko 9. atalean luzapenak nola jaisten eta instalatzen diren azaldu dugu. Hortaz, plug-inak jada instalatuta daudela igarritz aurrera jarraituko dugu. Hasteko, TeachingDemos kargatuko dugu. Horretarako sakatu beharreko aukerak Herramientas > Cargar plugin(s) de Rcmdr... dira.



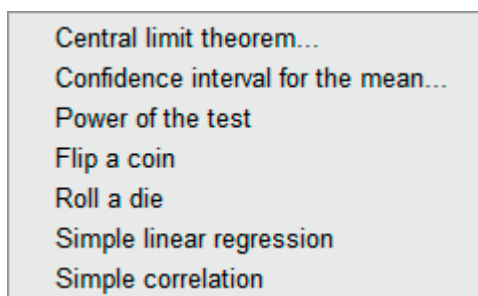
199. irudia.

TeachingDemos luzapena aukeratzean R Commander-ek programa berrabiaraztea eskatuko du. Eskaria onartu aurretik ziurtatu zaitez gordeta dituzula intereseko objektu guztiak, bestela ezabatu egingo baitira. Behin R Commander berrabiarazi dela, menu-barran Demos izeneko menua agertuko da.



200. irudia.

Une honetan prest daude luzapen honen aukera guztiak.



201. irudia.

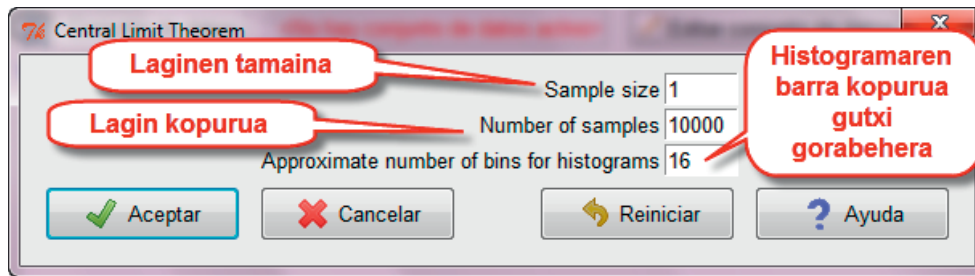
TeachingDemos estatistikaren irakaskuntzan erabiltzeko pentsatuta eta diseinatuta dagoen erreminta da. Funtsezko bost gairi heltzen die: limite zentralaren teoremari, batezbestekoarentzako konfiantza-tarteei, hipotesiaren testaren potentziari, erregresio lineal bakunari eta korrelazio-koefizienteei. Gainera txanpon baten eta dado baten jaurtiketak simula daitezke.

14.1. LIMITE ZENTRALAREN TEOREMA

Limite zentralaren teoremaren arabera aldagai batek jarraitzen duen banaketa edozein dela ere, banaketa horretatik ateratzen ditugun zorizko laginen batezbestekoek banaketa normalari jarraituko diote. Kontzeptu hau ikasteko TeachingDemos-ek eskaintzen duen ariketa banaketa desberdinetatik zorizko laginak erauztean datza, ondoren dagozkien batezbestekoek teorema baieztatzen duten ikusteko.

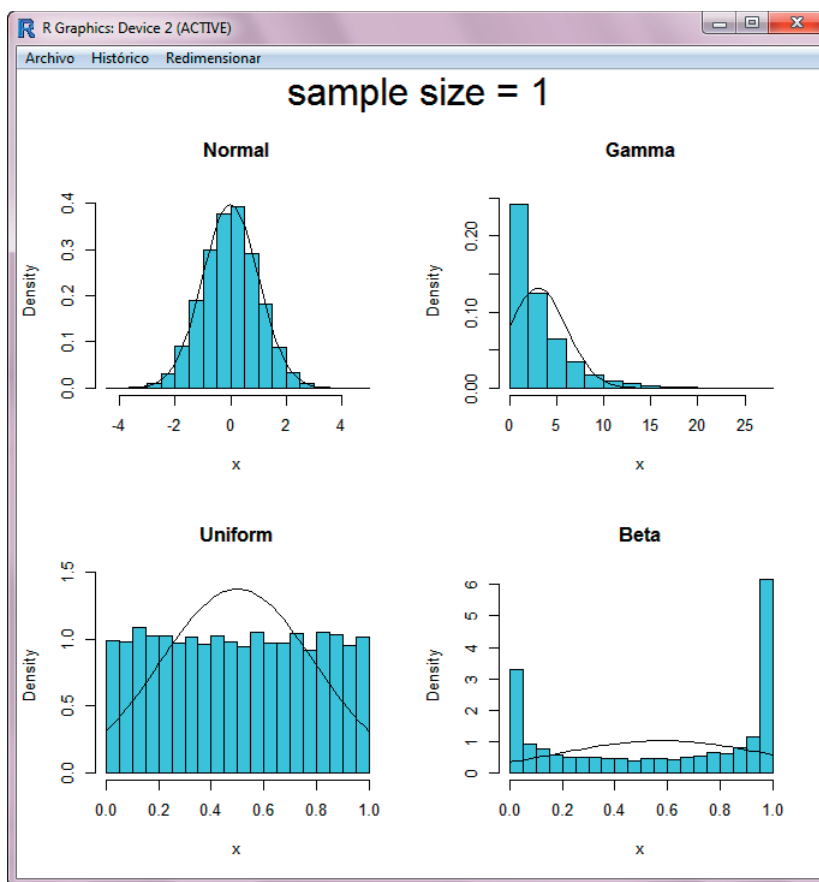
Limite zentralaren teorema baieztatzeko, lehenengo eta behin 1 tamainako lagin kopuru handi bat erauz daiteke banaketa teoriko bakoitzetik. Horrela, banaketa bakoitzak populazioan hartzen duen itxura ikusi ahal izango dugu. Jarraian, tamaina handiagoko laginak erauz daitezke; adibidean, lehen 1 balioa idatzi dugun lekuan 25 idatzi dugu oraingoan. Teoremaren arabera, lagin horien batezbestekoen banaketa banaketa normalaren antzekoa izan beharko litzateke edozein dela ere laginei dagokien populazioen banaketa. Baieztapena modu honetan gauzatzeko `Demos > Central limit theorem` sakatuko dugu.

- Hasteko, 1 tamainako 10000 lagin eskatuko ditugu banaketa mota bakoitzeko.



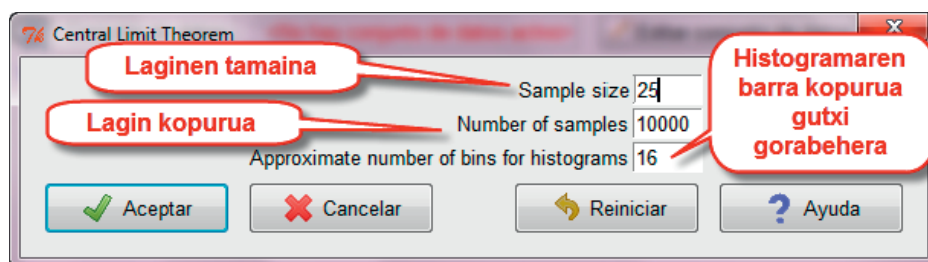
202. irudia.

Hurrengo irudian ageri den moduan, 10000 balioen irudikapen grafikoak banaketa desberdinak irudikatzen dituela ikus dezakegu.



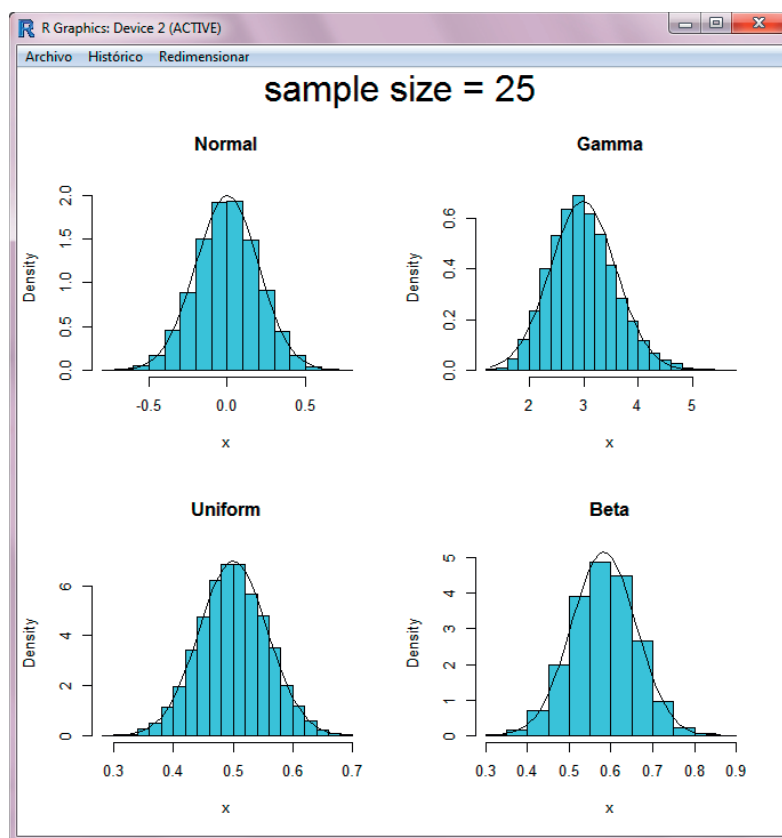
203. irudia.

- Ondoren, 25 tamainako 10000 laging eskatuko ditugu banaketa mota bakoitzeko.



204. irudia.

Hurrengo irudian ikus daitekeen moduan, eta limite zentralaren teorema adierazten duen bezala, banaketa bakoitzetik erauzitako 25 tamainako ($n = 25$) 10000 lagingen batezbestekoen irudikapen grafikoak banaketa normalaren antzekoak dira edozein dela ere jatorrizko banaketa mota.



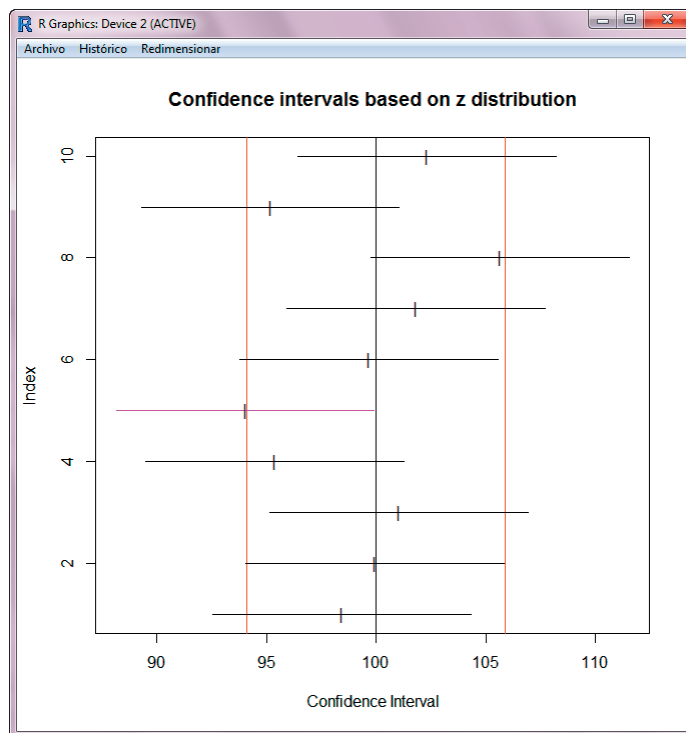
205. irudia.

14.2. POPULAZIOAREN BATEZBESTEKOENTZAKO KONFIANTZA-TARTEAK

Guk aukeratutako batezbestekoa eta desbideratze tipikoa dituen populazio batetik laginak eskuratzen direla simula daiteke Confidence Intervals for the Mean . . . aukerarekin. Interesatzen zaigun lagin kopurua eskuratu ostean, horietako bakoitzaren batezbestekoaren konfiantza-tartea kalkula daiteke aurrez zehaztu dugun konfiantza-mailarekin.

206. irudia.

Hurrengo irudian kalkulaturako 10 konfiantza-tarteak (erauzketa bakoitzeko bat) ikusi daitezke. 8 laginaren konfiantza-tarteak populazioaren batezbestekoa ez duela jasotzen antzeman daiteke.

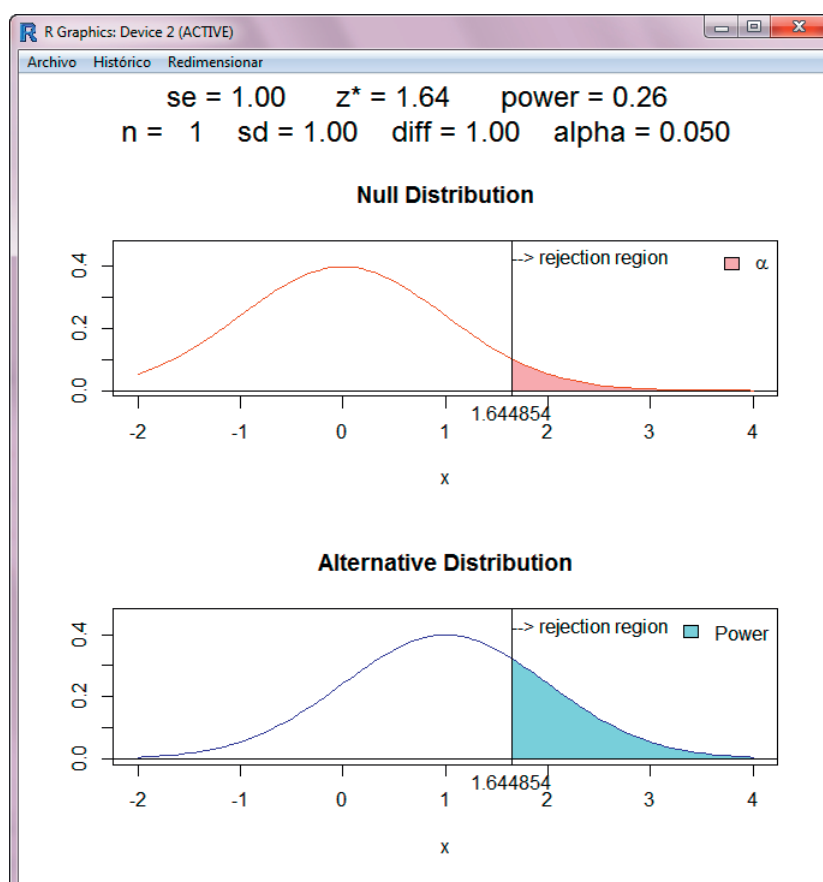


207. irudia.

14.3. TESTAREN POTENZIA

Test estatistikoaren potentzia eta efektuaren tamaina hipotesien testen alderdi oso garrantzitsuak dira. Baina kontzeptu horien asimilazioa ez da batere erraza. Hori dela-eta, TeachingDemos-ek potentzia kontzeptua ulertzen laguntzeko ariketa bat proposatzen du. Horretarako, I motako errorearen (hasierako hipotesia baztertzea benetakoa denean) eta potentziaren kontrastearen (hasierako hipotesia baztertzea faltsua denean) arteko erlazioa simulatzeko aukera ematen du.

Bi kontzeptuekin «jolasteko» hasierako hipotesitik abiatzen da (batezbestekoa = 0) eta ondoko hipotesia ($\text{diff} = \text{intereseko balioa}$), laginaren tamaina, populaziokaren desbideratze tipikoa eta I motako erroreak aldatuz, testaren potentziak pairatzen dituen aldaketak ebalua daitezke. Ariketa hau egiteko `Demos > Power of the test...` sakatuko dugu.

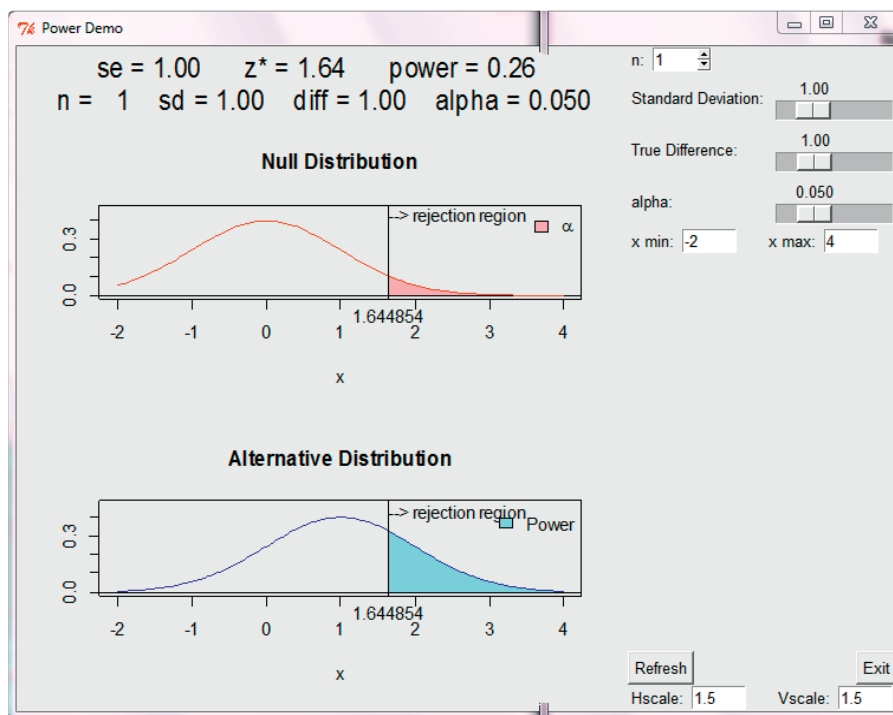


208. irudia.

Aurreko irudiko adibidea estatikoa da, alde bateko test baten egoera adierazten du. % 95eko konfiantza-mailarekin errore-maila (α) % 5ekoa da eta laginaren tamaina 1 da. Hasierako hipotesia baztertzeko espazioa $z^* = 1.644854$ estatistikoaren balioaren gainera aurkitzen da. Kontrastearen balioa kalkulatzeko, ondoko hipotesia (populazioaren batezbestekoa 1 da) benetakoa izatekotan, hasierako hipotesia baztertzeko probabilitatea ($z^* = 1.644854$) kalkulatu behar da. Alegia, batezbestekoaren balioa 1 eta desbideratze tipikoaren balioa 1 dituen populazioan, estatistikoak (aldagaiak) 1.644854 balioa baino handiagoa edukitzeko duen probabilitatea da probaren potentzia. Hortaz, baldin eta populazioaren estatistikoak $N(1,1)$ badira, probaren potentzia 0.26 ($P(x > 1.644854) = 0.26$) da.

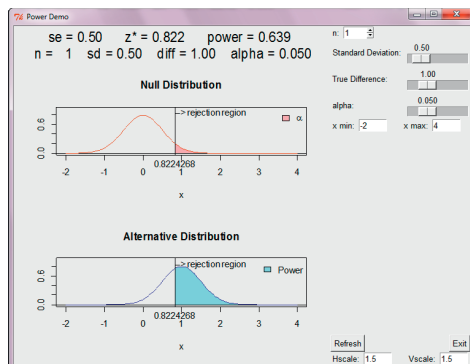
GARRANTZITSUA:

Demosek eskaintzen dizkigun abantailei ahalik eta onura handiena ateratzeko, bere itxura dinamikoa aktibatzea komeni da. Horretarako *tkrplot* paketea instalatu behar da (ikusi 4. atala). Behin instalatuta dagoela, R Commander-i testaren potentzia kalkulatzeko eskatzean, automatikoki leiho interaktibo bat irekiko da. Horrek eskuinean eta goian mugi daitezkeen hiru barra ditu, zeinekin populazioaren desbideratze tipikoaren, ondoko hipotesiaren (*diff*), I motako errore-mailaren, laginaren tamainaren eta koordinatuen ardatzen balioak alda daitezkeen.

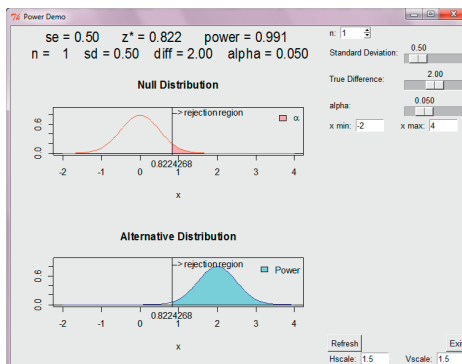


209. irudia.

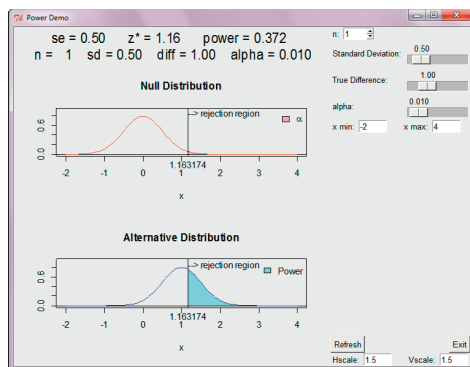
Itxura dinamikoa aktibatu ostean (ikusi aurreko leihoan) balio eta eskala desberdinak aldatu ahal izango dituzu, eta, ondorioz, potentziaren kalkulua egoera desberdinetan simulatu. Hurrengo irudietan desbideratze tipikoa, α errore-maila eta ondoko hipotesiak aldatuz lau adibide diseinatu ditugu.



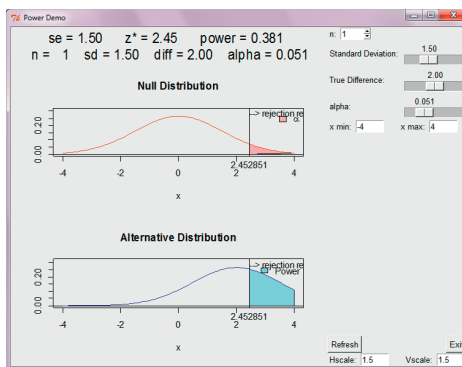
210. irudia.



211. irudia.



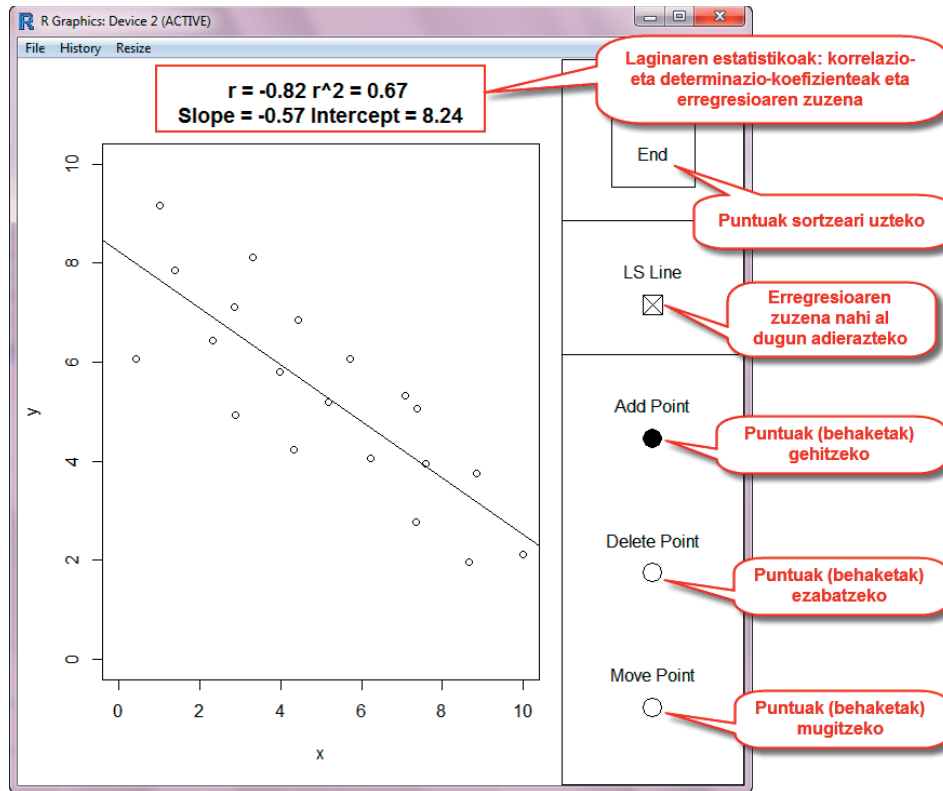
212. irudia.



213. irudia.

14.4. ERREGRESIO LINEAL BAKUNA

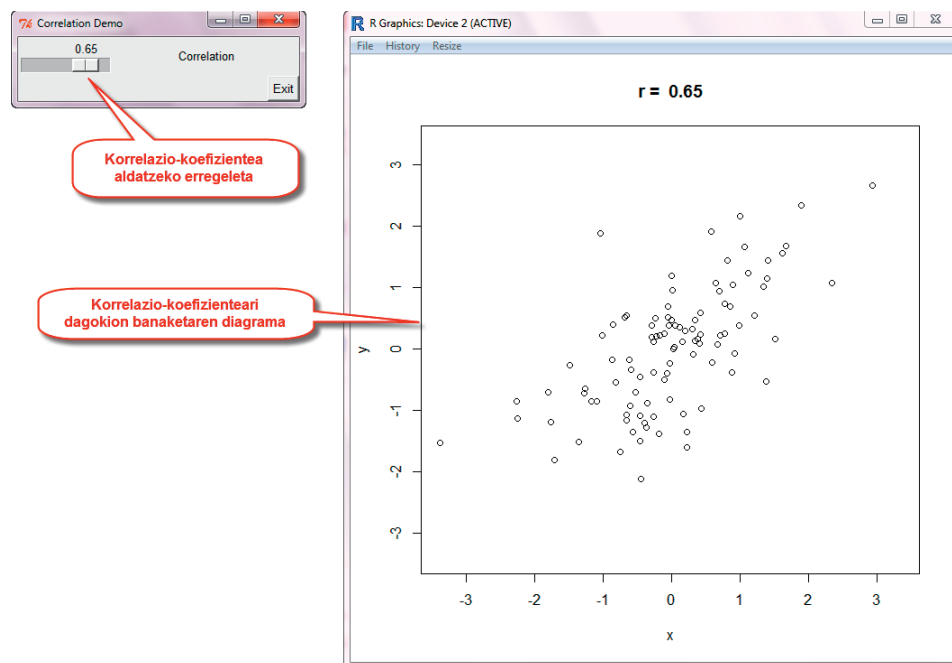
Bi aldagaiko behaketen emaitzak (puntuak) simulatzen ditu eta planoan irudikatu. Puntu berriak eranstean diren heinean TeachingDemos-ek erregresioaren zuzenaren ekuazioa eta korrelazio- eta determinazio-koefizienteen balioak ber kalkulatu, aldatzen eta berrirudikatzen ditu. Horretarako nahikoa da `Demos > Simple linear regression...` sakatzea.



214. irudia.

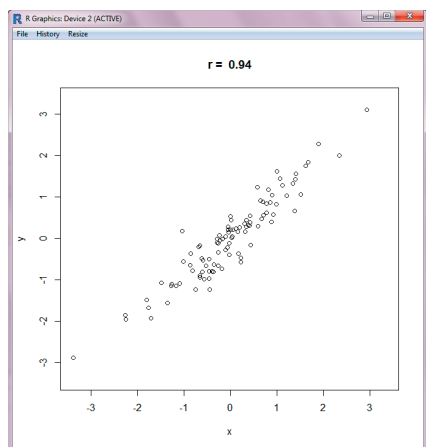
14.5 KORRELAZIO-KOEFIZIENTEAK

«Joko-adibide» ikusgarri baten bidez aurrez finkatutako korrelazio-koefizienteentzako banaketen diagramak sor daitezke Demos > Simple correlation... aukeraren bidez.

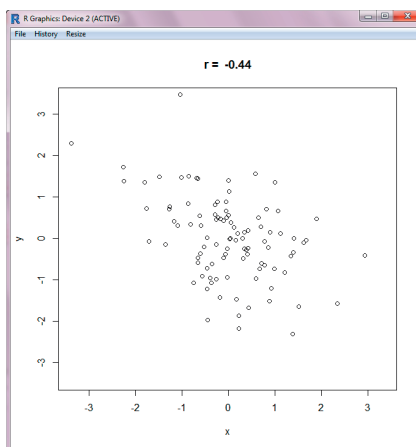


215. irudia.

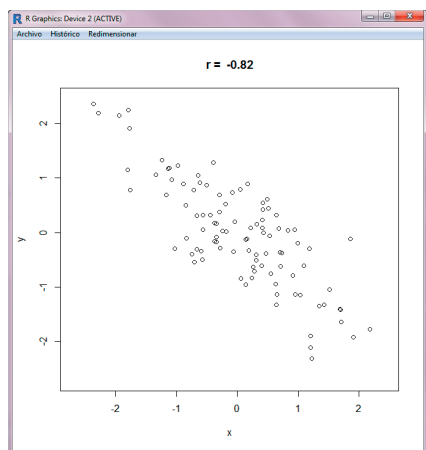
Korrelazio-koefizientearen balio desberdinak finkatuz hurrengo grafikoak lortu ditugu:



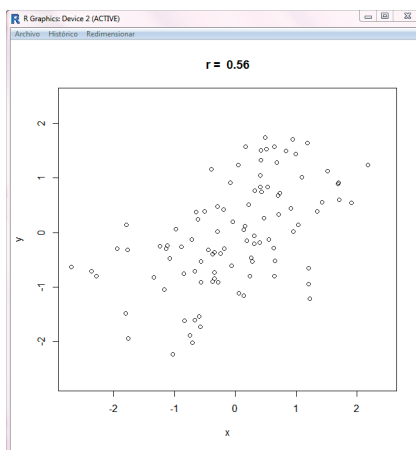
216. irudia.



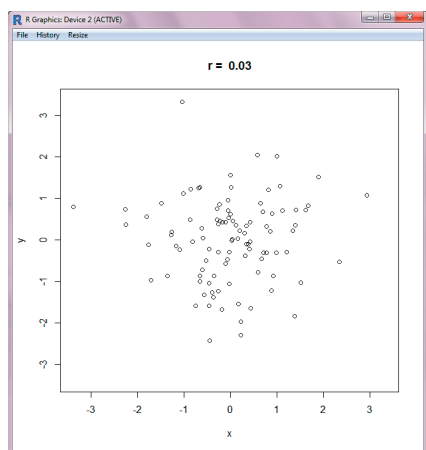
217. irudia.



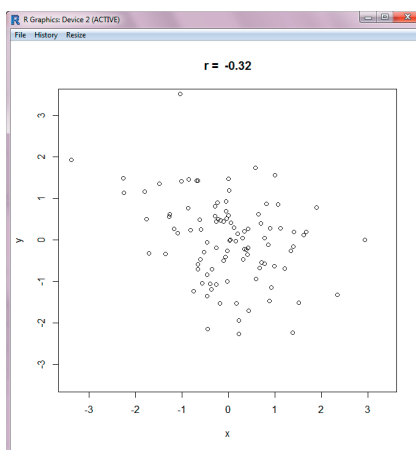
218. irudia.



219. irudia.



220. irudia.



221. irudia.

15. Erreferentzia bibliografikoak

- Becker, R. A.; Chambers, J. M. eta Wilks, A. R. (1988): *The new S language: A programming environment for data analysis and graphics*, CA: Wadsworth, Pacific Grove.
- Chambers, J. M. (1998): *Programming with data: A guide to the S language*, Springer, New York.
- , (2007): *Software for data analysis: Programming with R*, Springer, New York.
- Chambers, J. eta Hastie, T. (1992): *Statistical models in S*, CA: Wadsworth & Brooks/Cole, Pacific Grove.
- Cronbach, L. J. (1951): “Coefficient alpha and the internal structure of tests”, *Psychometrika*, **16(3)**, 297-334.
- Elosua, P. (2003): “Sobre la validez de los tests”, *Psicothema*, **15(2)**, 315-321.
- , (2009): “¿Existe vida más allá del SPSS? Descubre R”, *Psicothema*, **21(4)**, 652-655.
- , (2011): *Introducción al entorno R*, UPV/EHU, Leioa, <http://www.argitalpenak.ehu.es/p291-content/eu/contenidos/libro/se_ccsspdf/eu_ccsspdf/adjuntos/Introduccion%20al%20entorno%20R.pdf>.
- Elosua, P. eta Bully, P. (2011): *Application of new technologies in teaching/learning data analysis using simulation to improve the understanding of statistical concepts*, Comunicación presentada en ICERI, Madril.
- Elosua, P. eta Zumbo, B. (2008): “Coeficientes de fiabilidad para escalas de respuesta categórica ordenada”, *Psicothema*, **20(4)**, 896-901.
- Etxeberria, J. (2000): *Regresión múltiple*, La Muralla, Madril.
- Etxeberria, J. eta Tejedor, F. J. (2005): *Análisis descriptivo de datos en Educación*, La Muralla, Madril.
- Fox, J. (2005): “The R Commander: A Basic-Statistics Graphical User Interface to R”, *Journal of Statistical Software*, **14(9)**, 1-42.
- , (2007): *Getting Started With the R Commander*, <<http://socserv.mcmaster.ca/jfox/Courses/soc3h6/Getting-Started-with-the-Rcmdr.pdf>>.
- , (2009): “Aspects of the Social Organization and Trajectory of the R Project”, *The R Journal*, **1(2)**, 5-13.
- Ihaka, R. eta Gentleman, R. (1996): “R: A Language for Data Analysis and Graphics”, *Journal of Computational and Graphical Statistics*, **5**, 299-314.

- Martínez-Arias, R. (1995): *Psicometría: Teoría de los tests psicológicos y educativos*, Síntesis, Madril.
- Muñiz, J. (1998): *Teoría Clásica de Tests*, Pirámide, Madril.
- R Development Core Team (2011): *R: a language and environment for statistical computing*, R foundation for Statistical Computing, Viena, Austria.
- Tejedor, F. J. eta Etxeberria, J. (2006): *Análisis inferencial de datos en Educación*, La Muralla, Madril.
- Thurstone, L. L. (1947): *Multiple factor analysis*, University of Chicago Press, Chicago.
- Venables, W. N. eta Ripley, B. D. (2000): *S programming*, Springer, New York.

Sailean argitaratu diren beste liburu batzuk

Giza zientziei egokitutako Estatistika

Juan Etxeberria

1989n argitaratua

ISBN: 84-86967-12-0

Bioestatistika

Arantza Urkaregi

1991n argitaratua

ISBN: 84-86967-37-6

Kalkulu diferentziala eta integrala I

Joserra Aizpurua eta Patxi Angulo (koord.)

1992an argitaratua

ISBN: 84-86967-47-3

Kalkulu diferentziala eta integrala II

Joserra Aizpurua eta Patxi Angulo (koord.)

1992an argitaratua

ISBN: 84-86967-48-1

Ekuzio diferentzialak. Aplikazioak eta ariketak

Elena Agirre (euskaratzailea)

1994an argitaratua

ISBN: 84-86967-63-5

Optimizazioa. Programazio lineala

Victoria Fernandez eta Ana Zelaia

1992an argitaratua

ISBN: 84-86967-65-1

Estatistikarako Sarrera

Karmele Fernandez

1996an argitaratua

ISBN: 84-86967-70-8

Estatistikarako Sarrera. Ariketak

Karmele Fernandez, Josu Orbe, Marian Zubia

1997an argitaratua

ISBN: 84-86967-80-5

Estatistika I eta Estatistika II. Ariketak

Karmele Fernandez, Josu Orbe, Marian Zubia
1996an argitaratua
ISBN: 84-86967-79-1

R Hasiberrientzat

Gorka Azkune eta Yosu Yurramendi (itzul.)
2005ean PDFn argitaratua
ISBN: 84-86967-077-7

Estatistikaren oinarriak. Ariketak

Elena Agirre Basurko
2006an argitaratua
ISBN: 84-86967-079-3

Kalkulu diferentziala eta integrala (2. argitalpena)

Patxi Angulo
2009an argitaratua
ISBN: 978-84-8438-236-2

Erregresio lineala, bariantza-analisiak eta hipotesi-testak. Datu-analisirako lanabesak

Xabier Isasi Balanzategi
2010ean argitaratua
ISBN: 978-84-8438-290-4

Estatistikaren Oinarriak. Ariketak (2. argitalpena)

Elena Agirre Basurko
2010ean argitaratua
ISBN: 978-84-8438-288-1

Ekuazio diferentzial arruntak askatzeko zenbakizko metodoak

Elisabete Alberdi Celaya
2012an argitaratua
ISBN: 978-84-8438-419-9