Ekonometria eta GRETL

Marian Zubia Zubiaurre
Susan Orbe Mandaluniz

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

Ekonometria
eta
GRETL

Marian Zubia Zubiaurre
Susan Orbe Mandaluniz

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

CIP. Unibertsitateko Biblioteka
Zubia Zubiaurre, Marian

Ekonometria eta GRETL [Recurso electrónico] / Marian Zubia Zubiaurre, Susan Orbe Mandaluniz. – Datos. – Bilbao : Universidad del País Vasco / Euskal Herriko Unibertsitatea, Argitalpen Zerbitzua = Servicio Editorial, [2019]. – 1 recurso en línea : PDF (130 p.)

Modo de acceso: World Wide Web.

ISBN: 978-84-1319-100-3

1. Econometría. 2. Gretl (Programa de ordenador). I. Orbe Mandaluniz, Susan, coaut.

(0.034)330.43

Euskararen arloko errektoreordetzaren sare argitalpena

UPV/EHUko Euskara Zerbitzuak sustatua eta zuzendua. Euskarazko ikasmaterialgintza sustatzeko deialdiaren bitartez.

ISBN: 978-84-1319-100-3

Aurkibidea

1. gaia

Sarrera

1.1. **Sarrera**

Liburu honen helburua da, errealitate ekonomikoari buruzko informazio estatistikoa interpretatzea. Erabiliko den oinarrizko erreminta eredu ekonometrikoa izango da, ekonomiaren eskema teorikoak eta datuen analisirako teknika estatistikoak bat egiten baititu. Gainera, zenbait ariketa eta praktika burutzeko, Gretl doako softwarea erabiliko da.

1.2. **Ekonometriaren definizioak**

Idazlan edota liburuetan, ekonometriaren definizio asko topa daitezke, baina guztiek kontzeptu edota ideia orokor berberak dituzte amankomunean: Teoria Ekonomikoa, Estatistika, Matematika, aldagai ekonomikoak, erlazioa kuantifikatzea, etab.

— *Ekonomia* eta *metria* hitzetatik dator, eta teknika matematiko eta estatistikoak aplikatzen dituen Zientzia Ekonomikoaren atal bat da, ereduen bitartez, Teoria Ekonomikoak baieztatzeko eta arazo ekonomikoak konpontzeko xedea duena (Espainiako Errege Akademia).

— Ekonometria ez da estatistika ekonomikoa. Ez da Teoria Ekonomiko Orokorra deritzonarekin identifikatzen, nahiz eta azken teoria horren zati handienak ikuspegi kuantitatibo nabaria izan. Gainera, ez da soilik matematikaren aplikazioa ekonomian. Esperientziaren arabera, hiru ikuspuntu horiek —Estatistika, Teoria Ekonomikoa eta Matematika— beharrezko baldintzak dira gaur egungo bizitza ekonomikoko erlazio kuantitatiboak ulertzeko, baina ez dira nahikoak. Hiru ikuspuntuen baterakuntza da boteretsua, eta baterakuntza hori da ekonometria. (Frisch (1933)[1]).

— Gaur egungo fenomeno ekonomikoen analisi kuantitatiboa da, teoria eta behaketen garapenean oinarritzen da, eta inferentziako metodoekin erlazionatzen da (Samuelson, P.A., Koppmans, T.C. eta Stone, J. (1954)[2]).

[1] Frisch, R. (1933). *Econometrica,* 1. lib., 1. zb.. The Econometric Society.

[2] Samuelson, P.A., Koppmans, T.C. eta Stone, J. (1954). Report of the evaluative committee for Econometrics. *Econometrica* 22.

— Ekonometria da, erlazio ekonomikoak neurtzeko metodo estatistikoak erabiltzen dituen arte eta zientzia (Chow (1983)[3]).
— Ekonometria da, datu ekonomikoetan metodo estatistikoak aplikatzea (Maddala (1985)[4]).
— Ekonometriak aldagai ekonomiko baten ezaugarriak edo propietateak analizatzeko lagungarriak diren egiturak ikertzen ditu, beste aldagai ekonomiko batzuk kausa bezala erabiliz (Novales, (1993)[5]).
— Ekonometria da, aldagai ekonomikoen arteko erlazioak neurtu eta kuantifikatzeko estatistika erabiltzen duen ekonomiaren adar bat. Ekonomiaren teoria, matematika, estatistika eta konputu-metodoak erabiltzen dituen diziplina anizkoitzeko materia da (Ramanathan (2002)[6]).

1.3. Ekonometriaren erabileraren adibideak

Ikerketa ekonometriko baten helburua da, fenomeno ekonomikoak hobeto ulertzea, aldagai ekonomikoen arteko erlazioak aztertzea, eta, ondorioz, aldagai horien etorkizuna aurresatea. Oinarrizko erreminta, noski, eredu ekonometrikoa da, aldagai ekonomikoen arteko erlazioak ulertzeko lagungarria baita; baina, ezinbestekoak izango dira Teoria Ekonomikoa, Estatistika eta Matematika ere. Gainera, gaur egun, datu-baseak eskuragarri daude, eta, bestalde, ordenagailu eta software berriekin, ikerketak egitea errazagoa bihurtu da.

Honako hauek dira ekonometriaren erabilgarritasunaren zenbait adibide:

— Gas Natural Fenosa enpresako zuzendariak gasaren eskaeran zein faktorek eragiten duten ikertzen du.
— Edozein saltokik bere salmentetan eta mozkinetan publizitate-maila desberdinek duten eragina kuantifikatzen du, edota publizitate mota desberdinen eragina salmentetan.
— Gobernuak diru- eta fiskal-politikek langabezian, inflazioan, esportazio eta inportazioan, interes tipoan, eta abarretan duten eragina aztertzen du.
— Enpresa bateko sindikatuek, langileen soldatetan sexuarekiko diskriminaziorik dagoen ala ez aztertzen dute.
— Etxebizitza baten salerosketaren kasuan, etxe-merkatuaren egoera nolakoa den interesatzen da, eta baita etxebizitzaren prezioan eragiten duten faktoreak ere.
— Herri bateko udalari populazioaren garapena aurreikustea interesatzen zaio, gizarte-zerbitzuen beharra eta ondorioztatzen den finantzaketa planifikatzeko.
— Pertsona batek mailegu bat kontratatu nahi badu, komeni zaio interes-tipoen garapena zein den jakitea.
— Disko-etxe bati komeni zaio oraindik merkatura atera ez den disko baten salmenten zenbatespen bat egitea.
— Bideko segurtasun-neurrien eraginkortasuna jakiteko (segurtasun-gerrikoa jantzi beharra, adibidez), trafiko-istripuetako hilkortasun-tasa murrizten duen ala ez aztertzen da.
— Seguru-etxe batek bezeroei eskaini behar dien seguru desberdinen azterketa egiten du, prezioan eragiten duten baldintzak garbi izanik.

---

3 Chow, G.C. (1983). *Econometrics*. Ed. McGraw-Hill. New York.
4 Maddala, G. (1985). *Econometría*. Ed. Mc.Graw-Hill. Mexiko.
5 Novales, A. (1993). *Econometría*. Ed. Mc. Graw-Hill. Mexiko.
6 Ramanathan, R. (2002). *Introductory Econometrics with Applications*. Harcourt College Publishers, 5. argit.

— Familia batek bere seme-alabak ikastetxe publiko edo pribatu batera eramateko, faktoreak aztertzen ditu: matrikula, errenta, seme-alaba kopurua, etab.
— Herrialde-hauteskundeen eta Europako hauteskundeen emaitzetan diferentziarik dagoen aztertzen da.
— Kantari baten kontzertu bat antolatzeko eta irabaziak izateko, kantaldi horretan eragingo duten faktoreak aztertzen dira: kantariaren soldata, sarrera-txartelen prezioa, eguraldia, bestelako ekitaldiak, etab.

Adibidea:

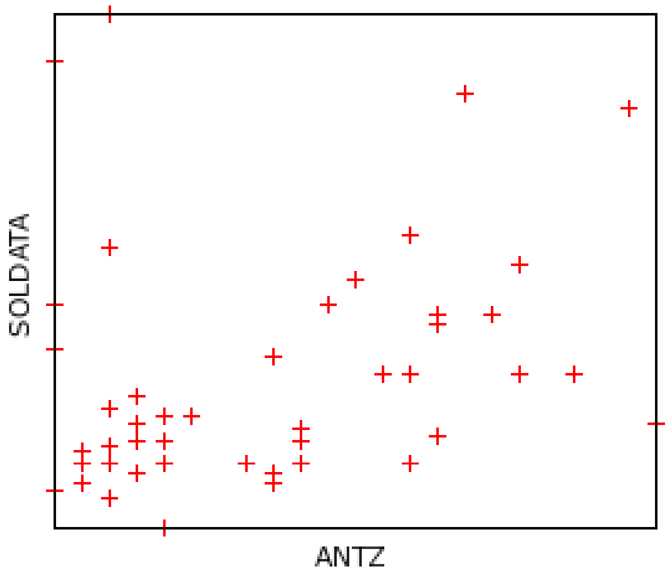
Demagun enpresa bateko langileen soldata aztertu nahi dela, eta honako aldagai hauen informazioa bildu dela[7]:

— **SOLDATA$_i$:** langilearen hileroko soldata eurotan.
— **HEZ$_i$:** langilearen hezkuntza-urteak.
— **ANTZ$_i$:** langilearen antzinatasuna enpresan (urteak).
— **ADINA$_i$:** langilearen adina urtetan.
— **GENEROA.** Bi kategoria dituenez (gizona eta emakumea), bi fikzio-aldagai definitu dira. Alde batetik, GIZ$_i$ = 1, *i* langilea gizona bada eta 0 bestelako kasuan, eta, bestetik, EMAK$_i$ = 1, *i* langilea emakumea bada eta 0 bestelako kasuan.
— **LANPOSTUA.** Enpresa honetan, hiru lanpostu desberdin daudela suposatuz (bulegoa edo ofizina, mantenu-lanak eta lantegia), bakoitzari dagokion fikzio-aldagaia definitu da. Horrela, OFIZINA$_i$ = 1, *i* langileak bulegoan lan egiten badu eta 0 bestelako kasuetan; MANT$_i$ = 1, *i* langileak mantenu-lanak egiten baditu eta 0 bestelako kasuetan; eta LANT$_i$ = 1, *i* langileak lantegi edo tailerrean lan egiten badu eta 0 bestelako kasuetan.

Aldagai horien datuak honako taula honetan ageri dira (lehen 5 langileenak), eta jarraian dagoen grafikoan langile bakoitzaren soldata marraztu da, langilearen antzinatasunaren funtzioan. Informazio hori kontuan izanik, edozein langileren ezaugarriak jakin daitezke. Adibidez, hirugarren langilearen hileroko soldata 1.715 euro da, 6 urteko hezkuntza du, 4 urte daramatza enpresan, 45 urte ditu, gizona da, arraza zurikoa, eta mantenu-lana du.

	Soldata	Hezkuntza-urteak	Antzinatasuna	Adina	Generoa	Lan mota
1	1.345	6	2	38	Emak	Bulegoa
2	2.435	4	18	52	Gizon	Lantegia
3	1.715	6	4	45	Gizon	Mantenu
4	1.461	6	4	58	Gizon	Bulegoa
5	1.639	9	3	30	Gizon	Lantegia

[7] Ramanathan data 7-2.gdt datu-fitxategitik modaltua.



Datuen adierazpen grafikoan oinarrituz, garbi ikusten da soldataren eta antzinatasunaren arteko erlazioa zuzena dela; hau da, antzinatasuna handitzen den heinean, soldata ere handitu egiten da. Bi aldagai horien arteko erlazio hori eredu baliagarri batean gauza daiteke, eredu ekonometriko batean hain zuzen ere, eta, teknika ekonometrikoekin, antzinatasunak soldatan duen eragina neur daiteke. Baita gainontzeko aldagairena ere, eta, horrela, langilearen soldatan zein aldagaik eragiten duten froga daiteke, erlazio horiek kuantifikatuz eta balioetsiz. Bestalde, grafiko honetan datuen sakabanatzea ez dela konstantea esan daiteke; hau da, antzinatasun gutxiko langileen soldatak oso antzekoak dira, baina antzinatasun handikoenak ez. Hori, ekonometrian, heterozedastizitate bezala ezagutzen da, eta kontuan hartu beharrekoa izango da aldagaien arteko erlazioa. Aurrerago aztertuko da arazo hori nola ekidin, eta baita irakasgaia garatzen den heinean sortuko direnak ere.

1.4. **Eredu ekonometrikoa**

Ikerketa ekonometriko batek, honako etapa hauek ditu:

— *Problemaren zehaztapena*. Aztertu nahi diren aldagaiak garbi izan behar dira, eta beren arteko erlazioari buruzko jakintza izan behar da, ikerketaren amaieran emaitzak balioetsi ahal izateko. Horretarako, ekonomiaren teoria lagungarri izaten da, batez ere arazoa fokuratzeko. Hortaz, lehen pausoa hasierako teoria edo eredu ekonomikoa planteatzea da ($Y = f(X)$).
— *Ikerketarako beharrezkoak diren datuen bilketa*. Azken finean, ikerketaren emaitza guztiak bildutako datuen kalitatearen araberakoak dira. Hala ere, analisirako aipagarriak diren datuak lortzea ez da beti erraza izaten, eta, noski, arazoak sor daitezke: datu bat falta, aldagai baten definizio-aldaketa, datu nahikorik ez izatea, edota aldagai baten informaziorik ez izatea.
— *Ereduaren zehaztapena eta zenbatespena*. Teorian eta lehen etapan planteatutako galderen loturatik, eredu ekonometriko bat zehaztatu behar da. Eredu ekonomikotik ekonometrikora pasatzeko, funtzio sortzailearen, $f(.)$, forma matematikoa zehaztatu behar da; hau da, aldagaien arteko erlazioaren funtzioa aukeratu behar da, aldagai endogenoaren (Y) izaera estokastikoa kontuan hartuz.

Hurrengo pausoa eredu ekonometrikoko parametro ezezagunak zenbatestea da. Ereduko elementu guztien hipotesiak kontuan hartuz, ereduko parametroen zenbatesle desberdinak proposatzen dira, eta, bildutako datuak erabiliz, parametro horien balioespenak lortzen dira, hau da, aldagaien arteko erlazioa kuantifikatzen da. Aipatu bezala, balioespen-metodo asko daude. Bata edo beste hautatzea, aukeratutako eredu ekonometrikoaren propietateen baitan dago.

— *Ereduaren analisia*. Aukeratutako eredua aldagaien portaera edo erlazioa jasotzeko egokia den aztertzea da. Adibidez, langilearen antzinatasunak bere soldatan eragiten duela jasotzeko egokia den, bi aldagai horien arteko erlazioa zuzena den, etab. Horretarako, kontrasteetan oinarritzen da, eta, lortutako emaitzak kontuan hartuz, eredua baieztatzen da, edo baliteke eredua aldatu beharra izatea.

— *Ereduaren aplikazioa*. Eredu zuzena lortu ondoren, interesgarriak diren galderak erantzuteko eta iragarpenak egiteko erabiltzen da.

1.5. **Eredu ekonometrikoen sailkapena**

Eredu ekonometrikoen sailkapena irizpide desberdinen arabera egin daiteke:

1. Eredu estatikoak eta eredu dinamikoak:

— Eredu estatikoak: aldagai guztiak une berean neurtuta daude.

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

— Eredu dinamikoak: aldagaiak une desberdinetan neurtuta daude.

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t-1} + u_t.$$

2. Ekuazio bakarreko ereduak eta ekuazio anizkoitzeko ereduak:

— Ekuazio bakarreko ereduak. Aldagaien arteko erlazio bakarra dago.

$$K_t = \beta_1 + \beta_2 R_t + u_t$$

— Ekuazio anizkoitzeko ereduak. Aldagaien arteko erlazio bat baino gehiago dago.

$$\begin{gathered} K_t = \beta_1 + \beta_2(1-\tau)Y_t + \beta_3 r_t + u_t \\ I_t = \alpha_1 + \alpha_2(Y_{t-1} - Y_{t-2}) + \alpha_3 r_{t-1} + \varepsilon_t \\ Y_t = K_t + I_t + G_t \end{gathered}$$

Non K kontsumoa baita; I, inbertsioa; Y, Barne Produktu Gordina; G, gastu publikoa; r, interes tipoa eta τ zerga tipoa.

1.6. Erregresio Lineal Orokorraren ereduaren (ELOE) elementuak

Eredu ekonometriko orokorra, K aldagai azaltzaile dituena, honako hau da:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + u_i$$

— Koefiziente edo parametroek, β_k, aldagaien arteko erlazioa kuantifikatzen dute. Koefiziente horiek ezezagunak direnez, zenbatetsi egin behar dira, eta balioetsitako koefizientea $\widehat{\beta_k}$ bezala jarriko da.

— Aldagaiak:

- *Aldagai azaldua,* endogenoa edo mendekoa: azaldu nahi den aldagaia da (Y).
- *Aldagai azaltzaileak,* exogenoak, independenteak edo erregresoreak ($X_2, X_3, \ldots, X_K$) dira interesatzen den aldagaia azaltzeko erabiltzen diren aldagaiak.

— Eredu ekonometrikoan ageri den perturbazio aleatorioak (u_i) aldagai azalduaren gain lagin behaketa desberdinek izan ditzaketen eragin guztiak biltzen ditu, eta baita ereduan barneratu ez diren aldagaien eraginak ere. Aipatu perturbazioa ez dela behagarria; izaera aleatorioa du, eta ereduan barneratu ez diren aldagaien eraginak, agente ekonomikoen jokabide aleatorioak, neurketa-erroreak, eta abar jasotzen ditu.

1.7. Datuak

Datu ekonomikoak ez dira kontrolatutako esperimentuetatik lortzen. Kontrolatutako esperimentu batean, ikertzaileak ikerketaren baldintza guztiak kontrolatzen ditu. Ongarrien eragina ikertzeko, adibidez, kantitate desberdinak aplika daitezke lursailetan, hezetasuna edo/eta landare bakoitzak jasotzen duen argi kopurua ere kontrolatuz. Gainera, esperimentu hau nahi adina errepika daiteke, baldintza berdinak mantenduz edo batzuk aldatuz. Horrela lortutako datuak, hau da, kontrolatutako esperimentuetatik lortutako datuak, *datu esperimental* bezala ezagutzen dira. Gizartean emandako prozedura baten ondorioz lortutako datuak, berriz, ez dira kontrolatuak (*datu ez-esperimentalak*).

Mota desberdineko datu ekonomikoak daude, eta gauzatuko den analisian eragina izango du. Lehen sailkapen batek datu *kuantitatiboak* eta *kualitatiboak* bereizten ditu. Lehenengoek zenbakizko balioak hartzen dituzte; adibidez, langileen soldata, enpresan duen antzinatasuna eta adina. Bigarrengoak, ordea, kategoria edo atributuetan agertzen dira; adibidez, generoa, lan mota edota arraza.

Bigarren sailkapen bat *denbora-segidako datuak* eta *sekzio gurutzatuko datuak* dira. Lehenak, une jarraituetan jasotako behaketak dira (normalean, erregularrak); urteak, hiruhilekoak edo hilabeteak. Adibidez, herri bateko Barne Produktu Gordina (BPG) 1980 urtetik 2016 urtera arte, hileroko langabezia-tasa, IBEX35aren eguneroko balioa. Bigarrenak, ordea, indibiduo desberdinek denbora momentu batean hartzen dituzten behaketak dira; adibidez, Europar Elkarteko herri bakoitzaren 2016. urteko populazioa, enpresa bateko langileen soldata, familia desberdinek egindako gastua ikuskizunetan. Posible da denbora-segidako eta sekzio gurutzatuko datuen konbinazio bat izatea ere, eta horiek paneleko datu edo luzetarako datu bezala ezagutzen dira: 2015-2016 eta 2016-2017 ikasturteetan estatistikako irakasgaian ikasleek lortutako kalifikazioa, Nazio Produktu Gordina 2000 urtetik 2016 urtera arte Europako Elkarteko herrietan.

Hirugarren sailkapen bat bateratzearen mailaren funtzioan ezartzen da. Horrela, *datu mikroekonomikoak* edo *mikrodatuak* daude. Horiek, indibiduo, familia edo enpresa bezalako agente ekonomikoen portaera jasotzen dute. Bestalde, *datu makroekonomikoak* edo *makrodatuak* daude. Banakako agenteen baterakuntzaren ondorioz sortutako herri, eskualde edo nazioei buruzko datuak dira horiek.

Datuen jatorriak

Askotan, datuak bilatu eta biltzea ez da erraza izaten, batzuetan lan handia izaten baita egoerari egokitzen zaizkion datuak lortzea eta maneiatzea. Hala ere, azken urteetan datuen bilketa asko hobetu da, batez ere erakunde askok beren datu-baseetan sartzen uzten dutelako. Honako datu hauek dira datuak publikatzen dituzten zenbait erakunde:

— Euskal Estatistika Erakundea (EUSTAT): http://www.eustat.es
— Espainiako Bankua: http://www.bde.es
— Nazio Estatistika Erakundea (INE): http://www.ine.es. Inebase edo Banco tempus. Eskuragarri daude, adibidez, Biztanleria Aktiboaren Inkestako emaitzak, Nazio Kontabilitatekoak eta Hileroko Aldizkarikoak.
— EUROSTAT: http://europa.eu.int/comm/eurostat
— Ekonomia Lankidetza eta Garapenerako Antolakundea (ELGA): http://www.oecd.org. Main Economic Indicators (hilekoa) edo Nazioarteko Merkataritza argitalpeneneko zenbait segida eskuragarri dira.
— Nazioarteko Diru Funtsa (NDF): http://www.imf.org

1.8. **Software ekonometrikoa**

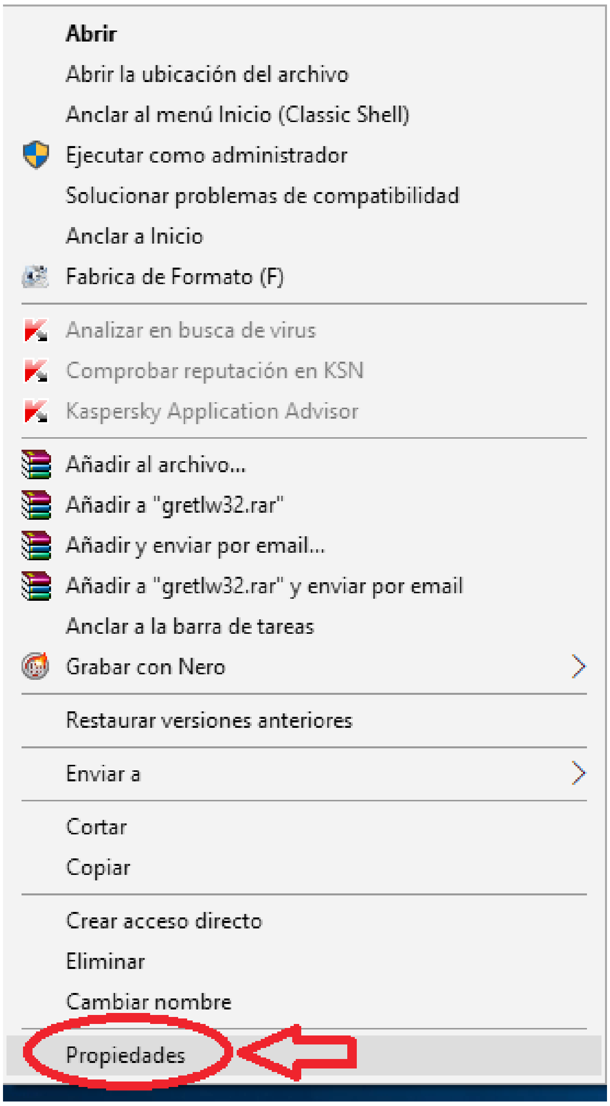
Ordenagailuen garapenak datu kopuru handi bat gordetzeko aukera eskaintzen du, eta, aldi berean, bere maneiua errazten. Gaur egun, analisi ekonometriko bat burutzeko software asko daude, eta, jarraibide erraz batzuk erabiliz, eragiketa konplexuak egin daitezke. Datuak paperean bakarrik badira eskuragarriak, datuak sartzeko, prestatzeko eta eragiketa errazak egiteko kalkulu-orriak (adibidez, EXCEL) erabil daitezke. Hala ere, orokorrean, komenigarri izaten da programa ekonometriko zehatzak erabiltzea. Ekonometriako ikastaroetan ohikoenak honako hauek dira:

— **Gretl.** Allin Cottrellek (Wake Forest Unibertsitatea) garatua. Software askea da, eta erabiltzeko erraza. Gaztelaniako bertsioa honako web-orritik jaits daiteke: http:\\gretl.source.forge.net\gretl_espanol.html, baina euskarazko bertsiora bihur daiteke. Orri horretan, Ekonometriako zenbait testu barneratzen dituen datu-base zabal bat dago erabilgarri; besteak beste, Ramanathan (2002), Wooldridge (2003) eta Greene (2003).
— **EViews.** Quantitative Micro Softwareak garatua. Analisi ekonometrikoko teknika asko biltzen ditu. Programaren informazioa jasotzeko web-orria: http:\\www.eviews.com
— **SHAZAM.** British of Columbia Unibertsitateak (Kanada) garatua. Eredu ekonometriko asko zenbatesteko teknikak biltzen ditu. Informazio gehiago honako web-orri honetan lor daiteke: http:\\shazam.econ.ubc.can
— **RATS.** Regression Analysis of Time Series. http:\\ www.estima.com
— **R.** Dohako softwarea da, kalkulu estatistikoak eta grafikoak egiteko aproposa.

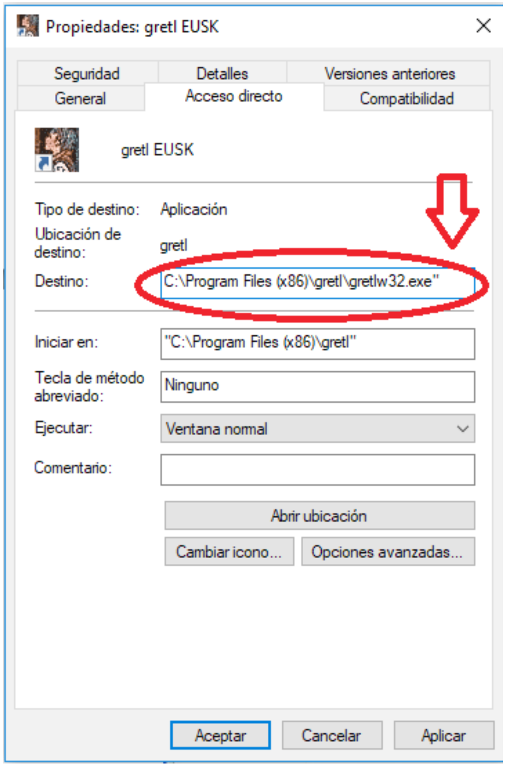
1.9. Gretl sarrera

Ekonometria garatzeko programa informatiko asko izan arren, irakasgai honetan Gretl programa erabiliko da, besteak beste, doakoa delako eta euskarazko bertsioa duelako. Atal honetan, lehenengo kontaktua egingo da Gretl programarekin. Gretl jaisteko web gunea http://gretl.sourceforge.net/win32 da. Bertan dituzue Gretl software-aren azken bertsioak hizkuntza desberdinetan. Bestela, nahi izanez gero, klikatu eGelan duzuen **gret-1.7.0.exe** exekutagarria. Behin programa instalatzen denean, Gretl ikonoa agertuko da ordenagailuko mahai gainean.

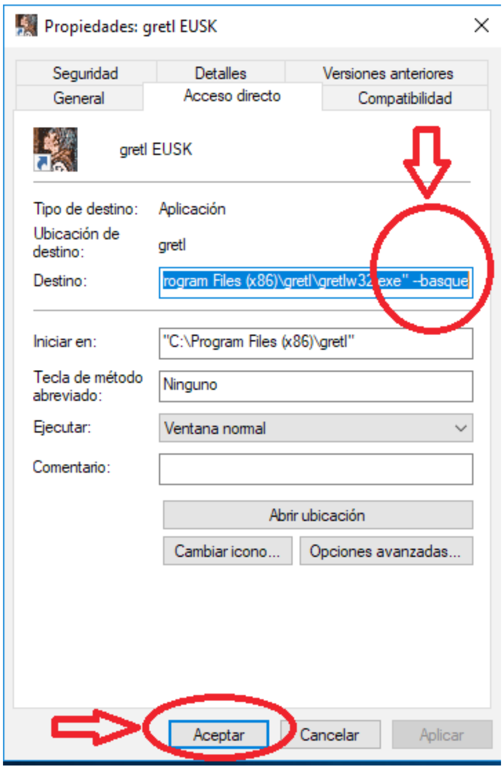
Euskarazko bertsioa lortzeko: Gretl ikonoaren gainean jarri, eta saguaren eskuineko botoiarekin irteten den menuan *Propiedades* aukeratu:



Hemen, *Acceso directo* atalean, *Destino* aldatu behar da:

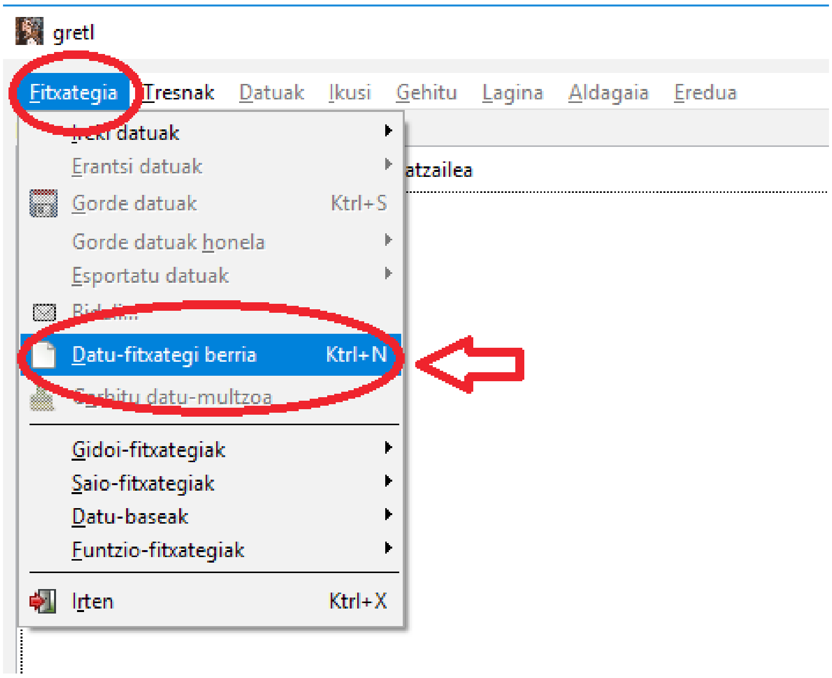


Agertzen den helbidean, azkeneko kakoaren atzetik hutsune bat utzi behar da, eta, **--basque** idatzi ondoren, *Aceptar* klikatu.



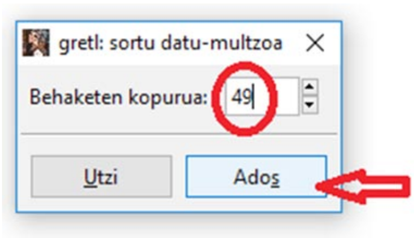
Oraindik ez denez fitxategirik kargatu, menu nagusiko aukera batzuk gris argian agertzen dira, ez baitaude erabilgarri.

Datu-fitxategiak sortu egin daitezke, Gretl softwareak dituenak erabil daitezke, edota beste formatu batean daudenak erabili. Lehenengo kasuan, hau da, fitxategia sortu nahi bada, honako aukera honetara joan behar da: *Fitxategia Datu fitxategi berria*

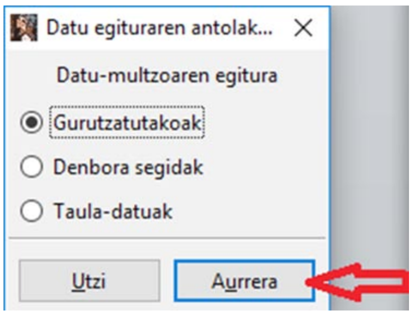


Eta jarraian, programak eskatzen duen informazioa osatu behar da:

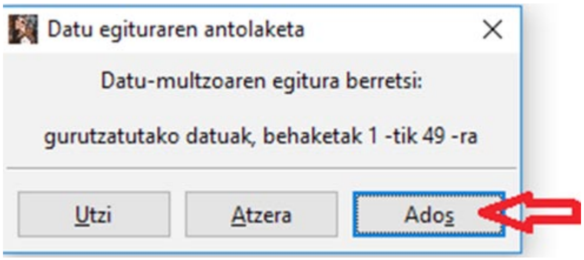
Behaketa kopurua, 49 langile direnez,



Nolako datuak diren jarriko da. Adibidean, *Gurutzatutako datuak* dira.

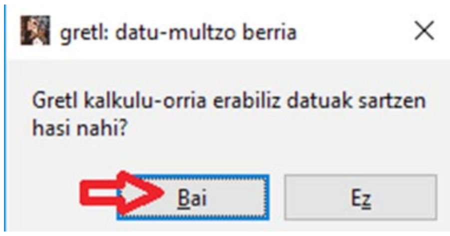


Aurreko pausoa zuzen eman bada, datuen egitura baieztatuko da *Ados* klikatuz.

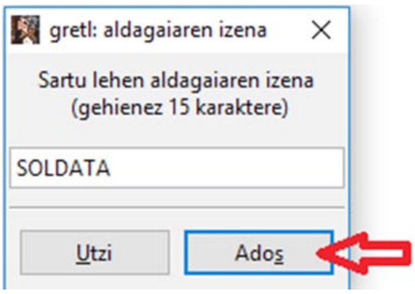


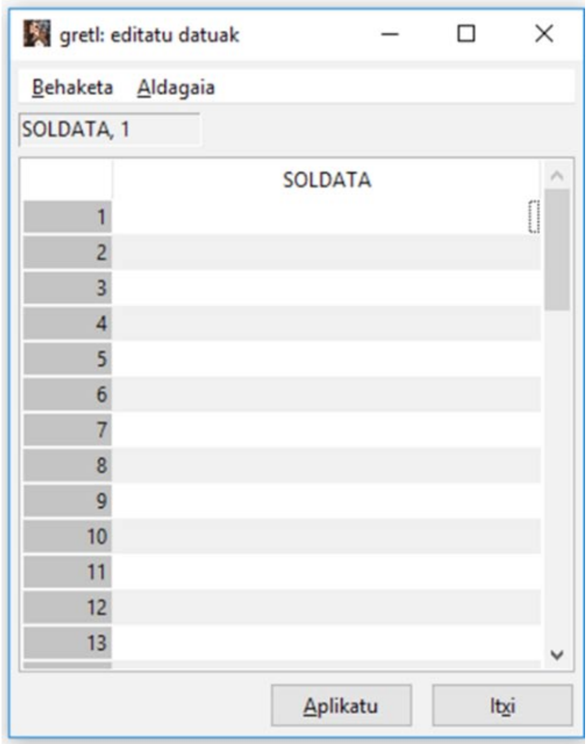
Nolako datuak diren azaltzen den leihatila berreskuratu nahi bada, *Atzera* klikatu behar da, baina aukera honek ez du behaketa kopuruan izan den akatsa zuzentzen uzten.

Azken leihatilan *Bai* klikatuko da datuak sartu ahal izateko.

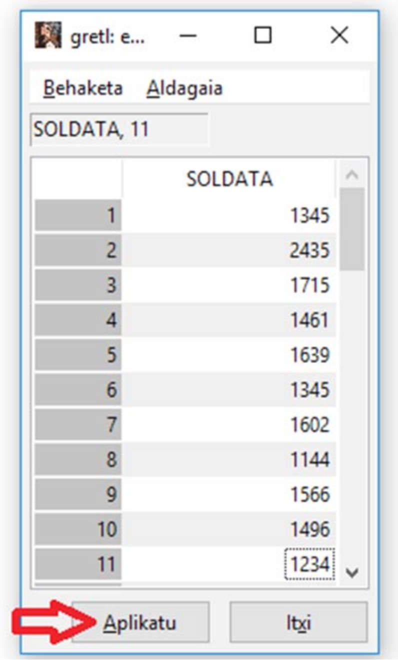


Ondorengo leihatilan lehenengo aldagaiari ezarriko zaion izena jarriko da; adibidez, SOLDATA. Jarraian, *Ados* klikatu eta kalkulu-orri bat irekiko da, eta pantailan honako hau azalduko:

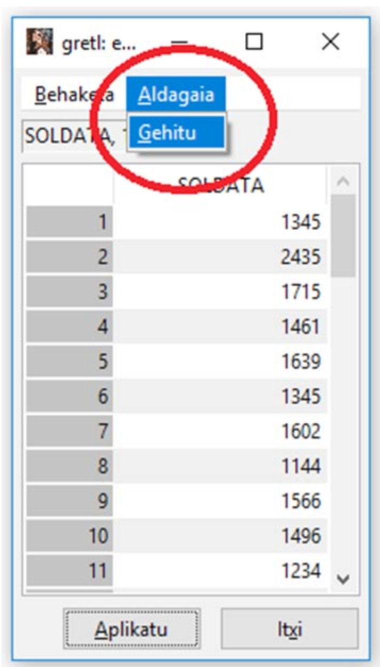


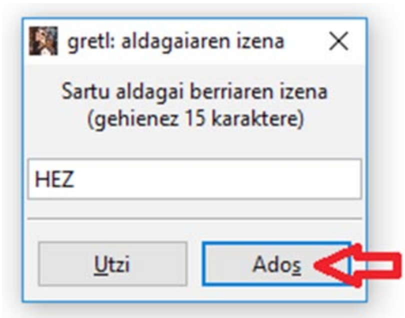


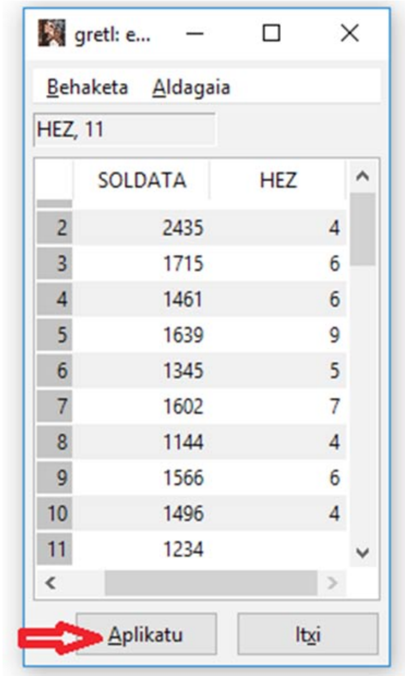
SOLDATA aldagaiaren datuak sartzeko, saguarekin dagokion gelaxkara joan (adibidez, lehenengora) eta, 1345 zifra idatzi ondoren, *intro* zapalduko da. Datu bat sartzerakoan akats bat egiten bada, adibidez bigarren behaketa (2435) ez bada sartu, ondorengo errenkadan jarri (adibidean 1715), eta, *Behaketak Erantsi behaketak* klikatuz, errenkada txuri berri bat irekiko da aurrekoaren gainean. Gogoratu lan-saioko aldaketak gordetzeko *Aplikatu* klikatu behar dela.



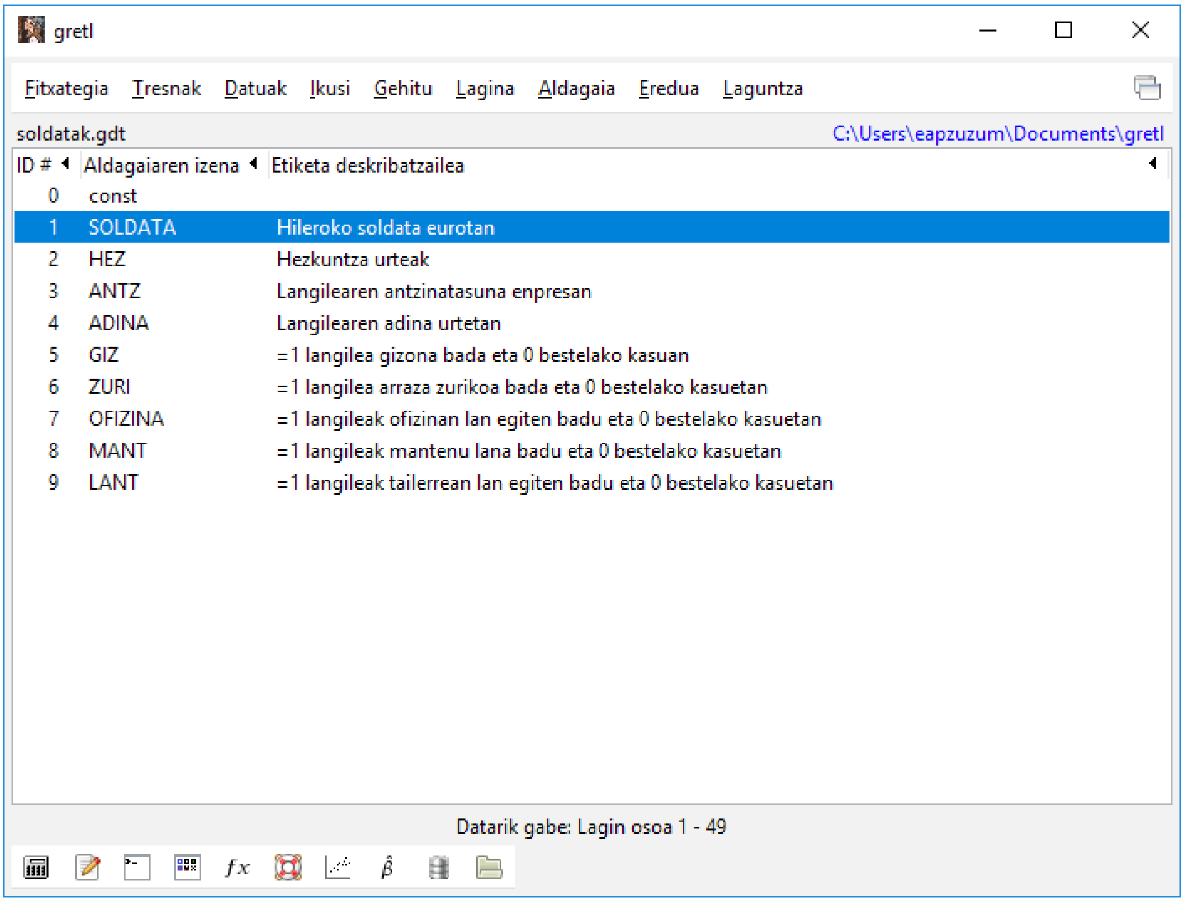
Aldagai gehiago erants daitezke; kalkulu-orriko menuko *Aldagaia Gehitu* aukerarekin egin daiteke. Adibidez, HEZ izeneko aldagai bat sortuko da, eta, datu guztiak sartu ondoren, *Aplikatu*ra joan eta kalkulu-orria itxiko da.







Azken aldaketak ez baldin badira gorde, kalkulu-orria ixterakoan koadro berri bat azalduko da aldaketen berrespena eskatuz. Osatutako segidak horrela azalduko dira pantailan:

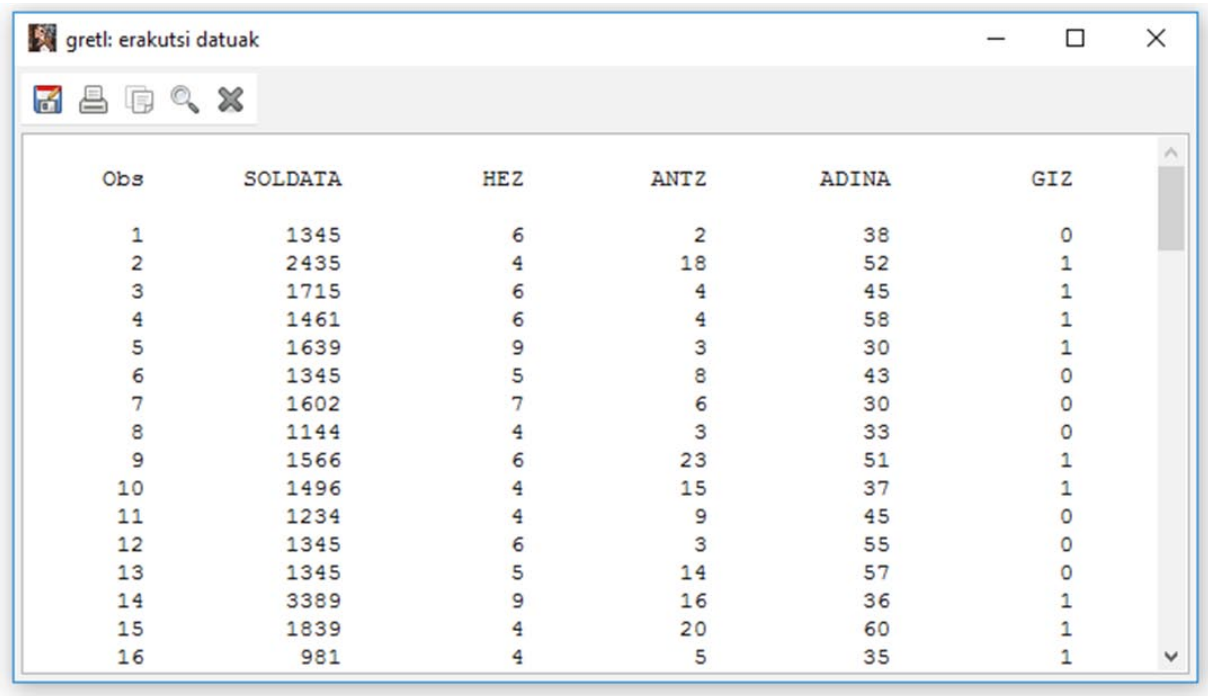


Adibide honetan, aldagai guztien datuak sartu ondoren, aldagai bakoitzaren ezaugarriak zein diren ere jarri da, aldagai bakoitzean saguaren eskuineko botoiari sakatu eta *Ezaugarriak editatu* aukerarekin. Komenigarria izaten da artxibo edo fitxategi batean jadanik sartutako datuak gordetzea, menu nagusiko *Fitxategia* aukeran, *Gorde datuak* erabiliz. Hurrengo leihatilan direktorioa erantsiko diogu, eta datuen fitxategiaren izena; adibidez, *SOLDATAK*. Datuak *gdt* luzapenarekin gordeko ditu; hurrengo saio batean fitxategia irekitzeko, fitxategian bi aldiz klikatu besterik ez da egin behar.

Askotan, datuak EXCEL bezalako beste kalkulu-orri batean egoten dira gordeta. Adibidez, soldatak.xls artxiboan. Orduan:

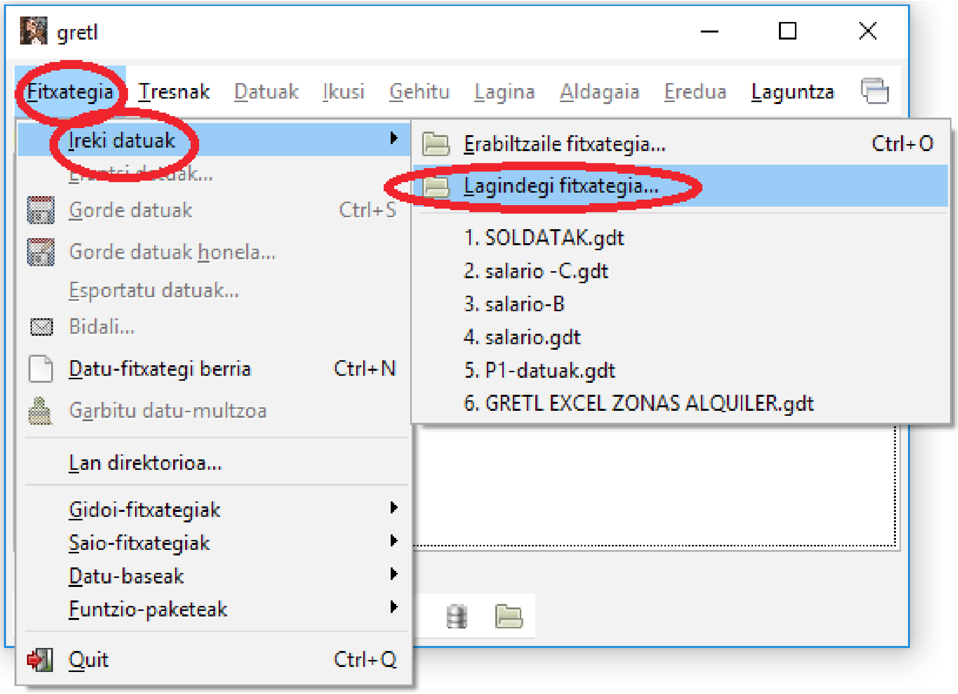
— Egiaztatu Gretl-eko kalkulu-orria geldi dagoela.
— Menu nagusiko *Fitxategia* → *Ireki datuak* → *Inportatu* → *EXCEL* erabiliz.
— Eman EXCEL artxiboaren izena eta kokapena.
— Datuak inportatu behar diren hasierako gelaxka zehaztu, eta *Ados* klikatu.

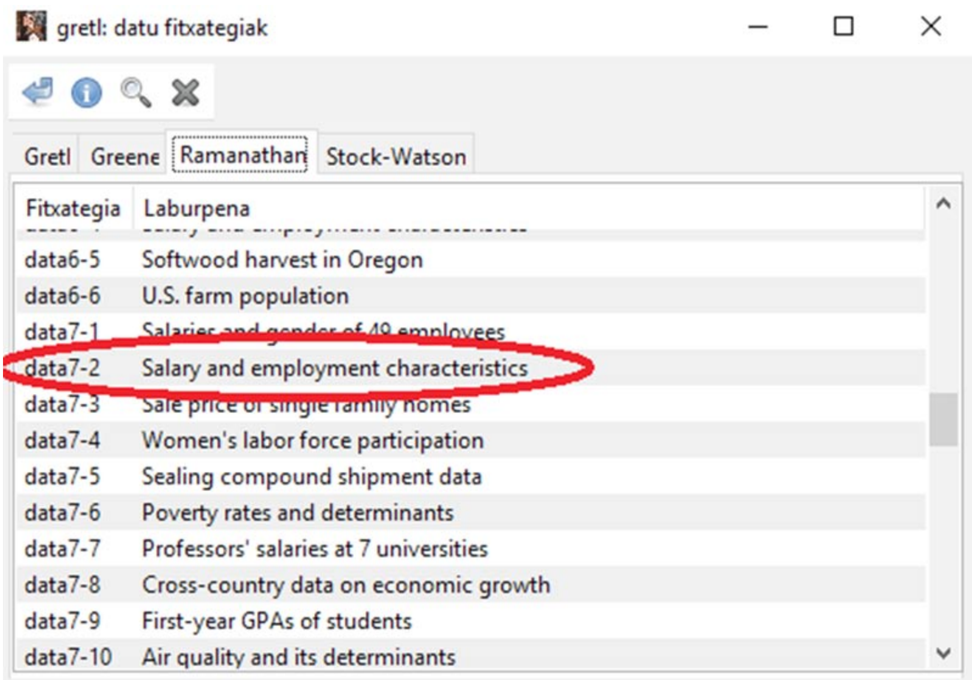
Datuen inportazioa ongi gauzatu den frogatzeko, *Datuak* aukeran *Aukeratu guztiak* atalera joan, eta *Datuak* → *Erakutsi balioak* aukerarekin froga daiteke. Adibidean, honako leihatila hau agertuko da:



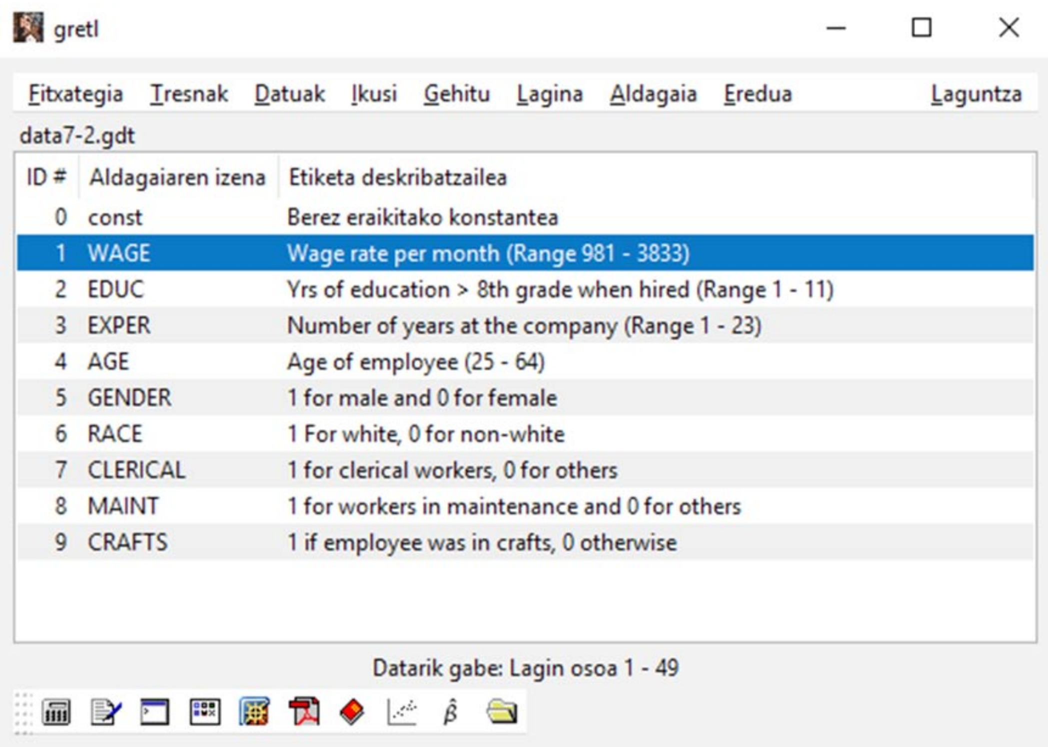
Obs	SOLDATA	HEZ	ANTZ	ADINA	GIZ
1	1345	6	2	38	0
2	2435	4	18	52	1
3	1715	6	4	45	1
4	1461	6	4	58	1
5	1639	9	3	30	1
6	1345	5	8	43	0
7	1602	7	6	30	0
8	1144	4	3	33	0
9	1566	6	23	51	1
10	1496	4	15	37	1
11	1234	4	9	45	0
12	1345	6	3	55	0
13	1345	5	14	57	0
14	3389	9	16	36	1
15	1839	4	20	60	1
16	981	4	5	35	1

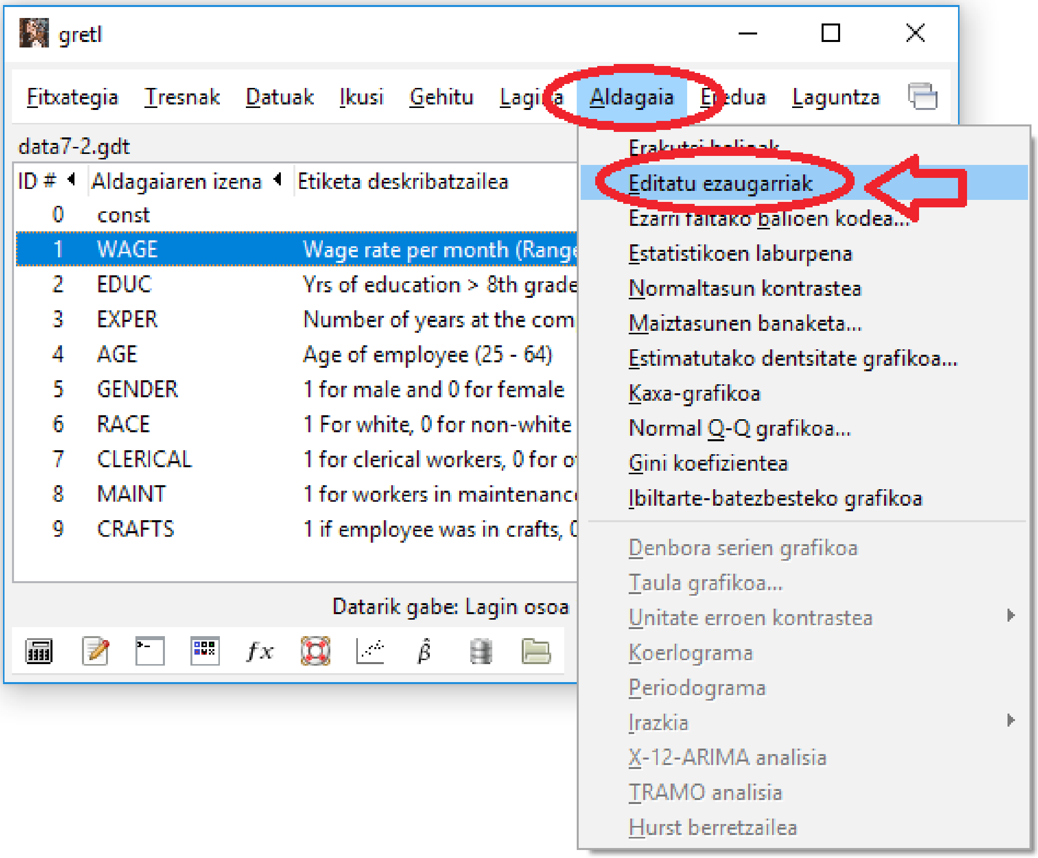
Azkenik, aipatu bezala, Gretl softwarea instalatzean, jada baditu zenbait datu-fitxategi. Hainbat liburutan adibide gisa erabiltzen direnak dira, baina Gretl-eko web-orritik beste datu-base asko jaisteko aukera ere badago. Adibideko datu-fitxategia Ramanathan-eko data 7-2 da. Hortaz, fitxategi hau irekitzeko, *Fitxategia* klikatu, ondoren *Lagindegi fitxategia*, jarraian *Ramanathan*, eta aukeratu *data7-2*, bi aldiz klikatuz:

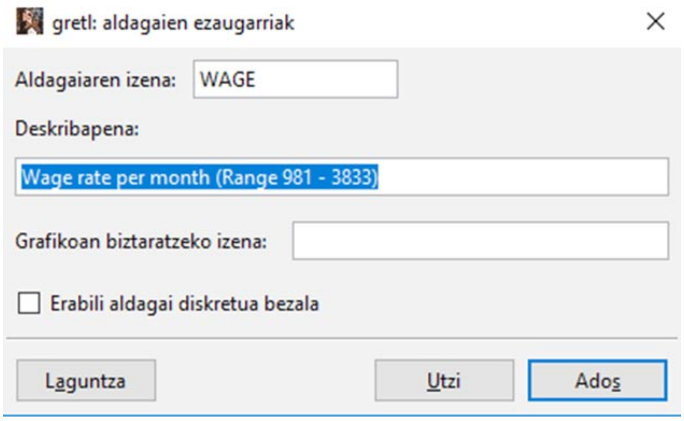


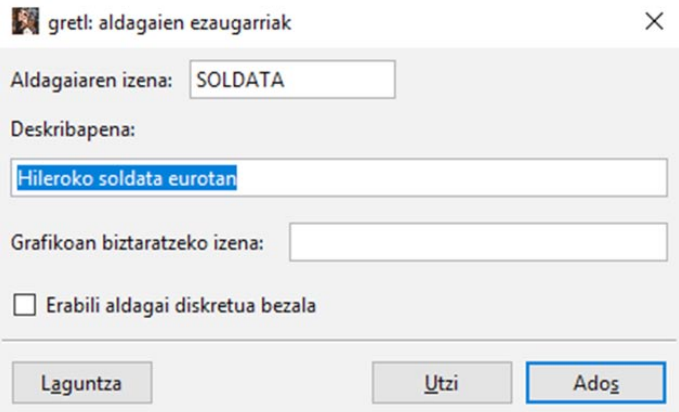


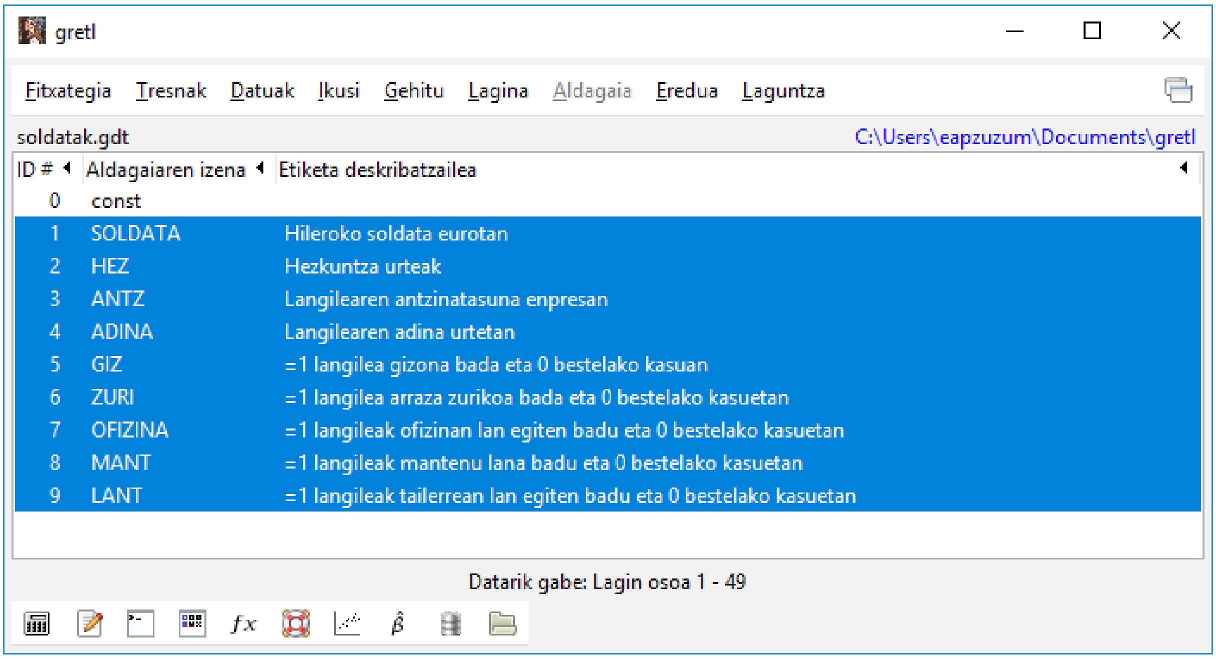
Aldagaien izenak eta ezaugarriak ingelesez ageri direnez, aldagai bakoitzaren ezaugarriak aldatuko dira lehen azaldutako pausoak jarraituz.


gretl
Fitxategia Tresnak Datuak Ikusi Gehitu Lagina Aldagaia Eredua Laguntza
data7-2.gdt
ID # Aldagaiaren izena Etiketa deskribatzailea
0 const Berez eraikitako konstantea
1 WAGE Wage rate per month (Range 981 - 3833)
2 EDUC Yrs of education > 8th grade when hired (Range 1 - 11)
3 EXPER Number of years at the company (Range 1 - 23)
4 AGE Age of employee (25 - 64)
5 GENDER 1 for male and 0 for female
6 RACE 1 For white, 0 for non-white
7 CLERICAL 1 for clerical workers, 0 for others
8 MAINT 1 for workers in maintenance and 0 for others
9 CRAFTS 1 if employee was in crafts, 0 otherwise
Datarik gabe: Lagin osoa 1 - 49


gretl
Fitxategia Tresnak Datuak Ikusi Gehitu Lagina Aldagaia Eredua Laguntza
data7-2.gdt
ID # Aldagaiaren izena Etiketa deskribatzailea
0 const
1 WAGE Wage rate per month (Range
2 EDUC Yrs of education > 8th grade
3 EXPER Number of years at the com
4 AGE Age of employee (25 - 64)
5 GENDER 1 for male and 0 for female
6 RACE 1 For white, 0 for non-white
7 CLERICAL 1 for clerical workers, 0 for o
8 MAINT 1 for workers in maintenanc
9 CRAFTS 1 if employee was in crafts, 0
Datarik gabe: Lagin osoa
Editatu ezaugarriak
Ezarri faltako balioen kodea...
Estatistikoen laburpena
Normaltasun kontrastea
Maiztasunen banaketa...
Estimatutako dentsitate grafikoa...
Kaxa-grafikoa
Normal Q-Q grafikoa...
Gini koefizientea
Ibiltarte-batezbesteko grafikoa
Denbora serien grafikoa
Taula grafikoa...
Unitate erroen kontrastea
Koerlograma
Periodograma
Iragazkia
X-12-ARIMA analisia
TRAMO analisia
Hurst berretzailea





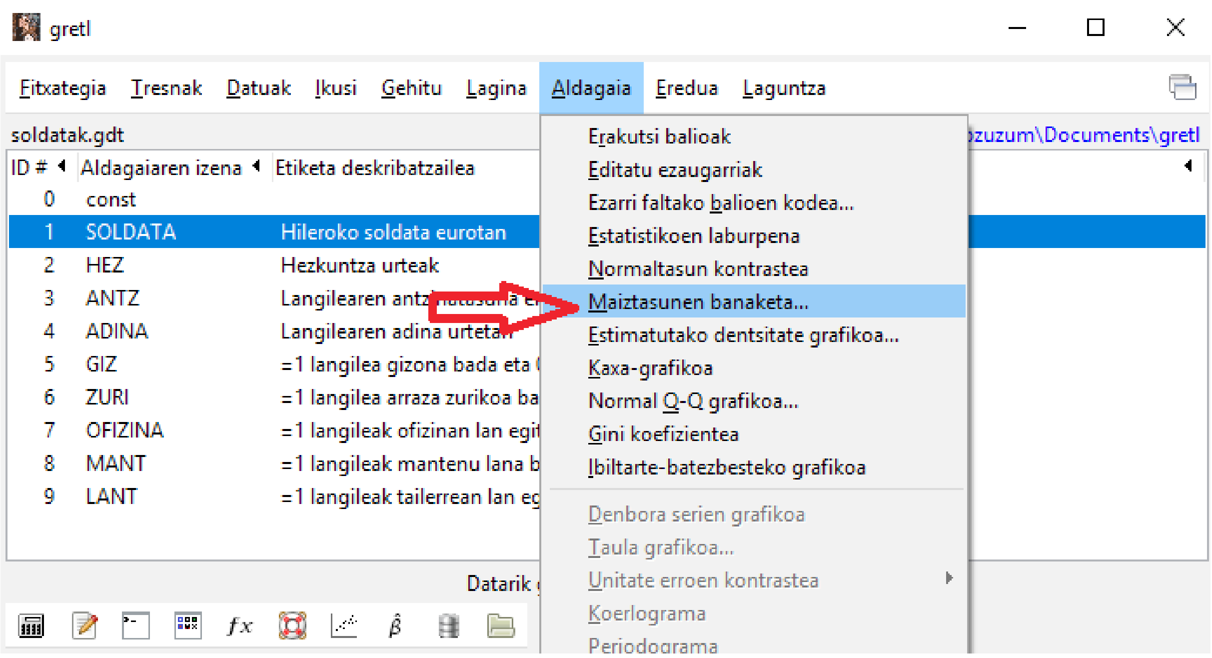


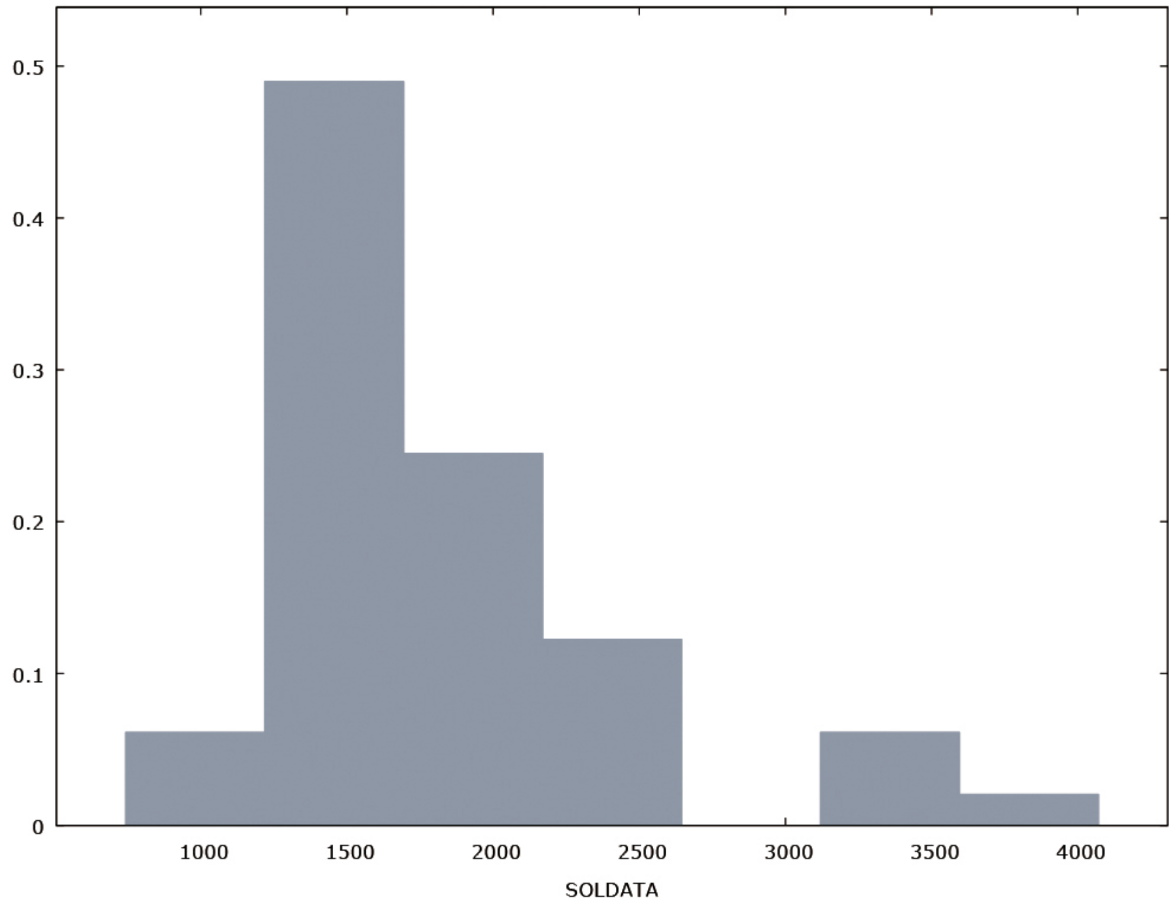
Gogoratu, fitxategia osatu denean eta gustuko ezaugarriak jarri direnean, komenigarria izaten dela datu-fitxategia gordetzea hurrengo batean erabili ahal izateko.

Jarraian, Gretl programa erabiltzen hasteko, estatistikan ikasitako aldagaien analisi deskribatzailea egingo da. Lehen etapa batean, datuen ikuspegi orokor bat egin behar da, eta analisi deskribatzaile batekin datu multzoaren ezaugarri nagusiak laburbiltzen dira, ikerketarako ezaugarri eta informazio nabariena ateraz.

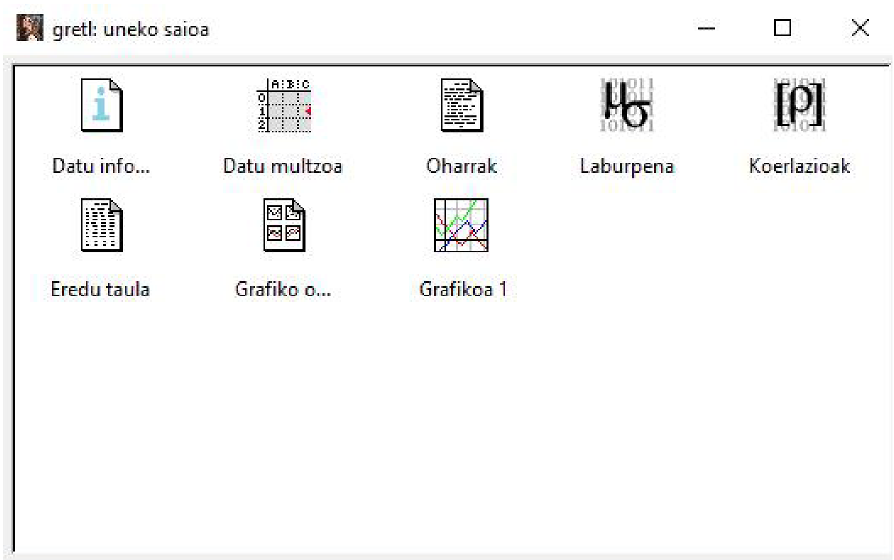
Lehendabizi, aldagai bakoitzaren informazioa laburbilduko da, eta, ondoren, aldagaien arteko erlazioen irudiak jasoko dira. Aldagai baten analisi deskribatzailerako elementu nagusiak menu laguntzailean agertzen dira, aldagaiaren gainean jarri eta saguaren eskuineko tekla klikatuz edo menu nagusiko *Aldagaia* aukera klikatuz.

Aldagai ekonomiko baten sekzio gurutzatuko datuak laburbiltzeko gehien erabiltzen den grafikoa histograma da, menu laguntzailean, *Aldagaia* aukeran, *Maiztasun grafikoa* aukeran. Barra-diagrama bat da, ardatz horizontalean aldagaien balioak tarteka banatuak agertzen direlarik. Tarteak zabalera berdinekoak badira, tarte bakoitzaren gainean tarte horretako behaketa kopuruko altuerako barra bat marrazten da. «Soldata» aldagaiaren gainean saguaren eskuineko tekla eta *Maiztasun banaketa* klikatuz, honako histograma hau lortzen da:

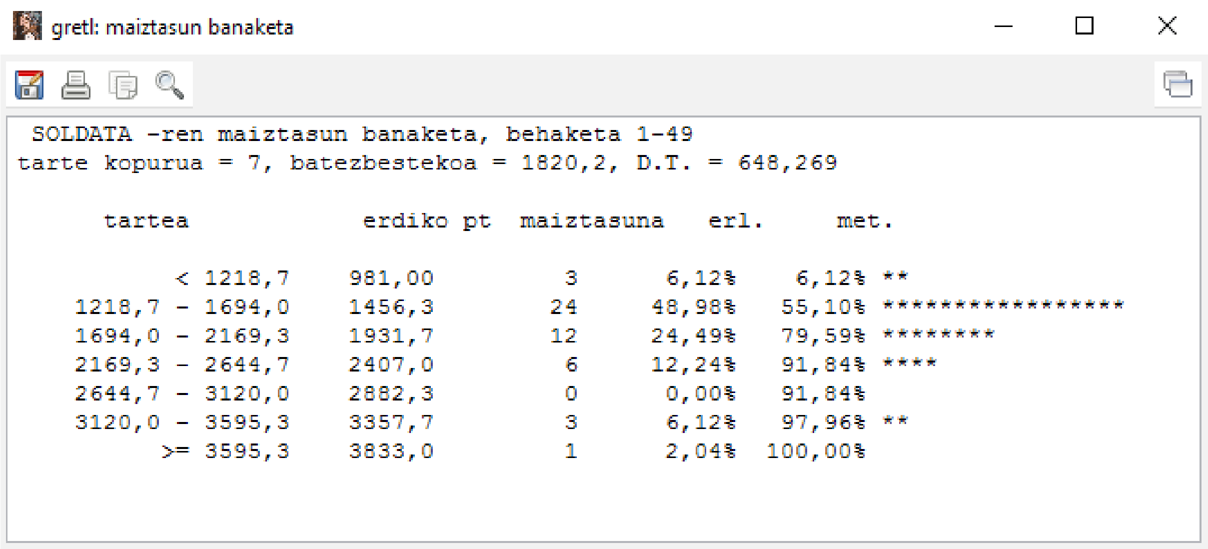




Edozein grafikotan klikatuz, menu laguntzaile batek irtengo du, eta aldaketak egin daitezke edo formatu desberdinetan gordetzeko (postcript, pdf, etab.) aukera dago. Saioan zehar, *Gorde ikono bezala saioan* aukerarekin grafikoa gorde daiteke, eta, horrela, edozein momentutan berreskuratu daiteke behean eskuinaldean dagoen laugarren sinboloan (saioaren ikono-ikuspegia) eta gero Grafikoa 1-en klikatuz.

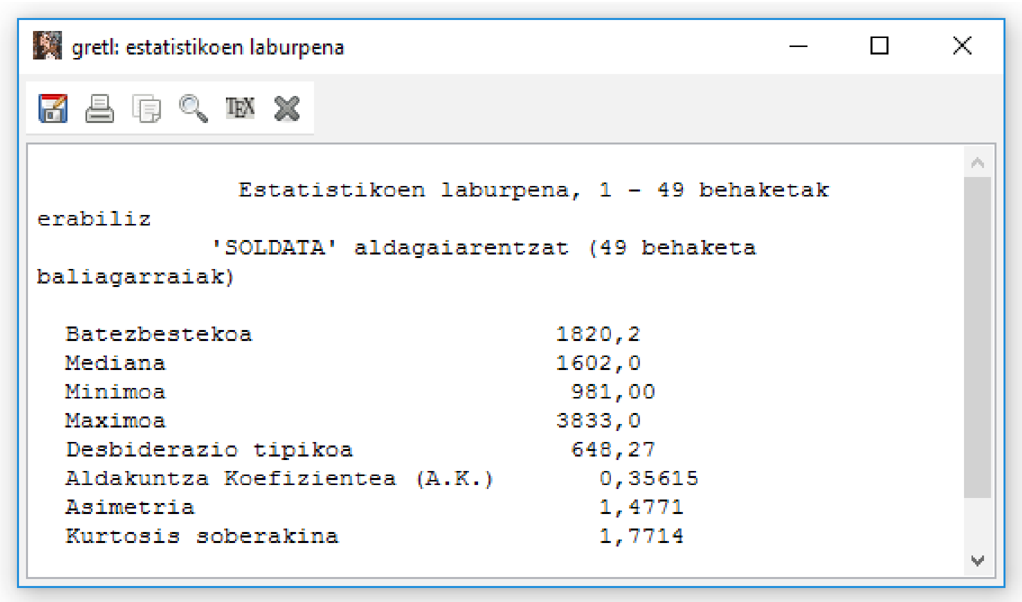


Histograman adierazitako maiztasun-banaketak ikusteko, dagokion aldagaia markatu eta, *Aldagaia* aukeran, *Maiztasun-banaketa* aukerara zuzendu behar da. Honako hau da adibideko soldata aldagaiaren maiztasun-banaketari dagokiona:


gretl: maiztasun banaketa

```
 SOLDATA -ren maiztasun banaketa, behaketa 1-49
tarte kopurua = 7, batezbestekoa = 1820,2, D.T. = 648,269

      tartea            erdiko pt  maiztasuna   erl.     met.

           < 1218,7    981,00         3      6,12%    6,12% **
    1218,7 - 1694,0    1456,3        24     48,98%   55,10% *****************
    1694,0 - 2169,3    1931,7        12     24,49%   79,59% ********
    2169,3 - 2644,7    2407,0         6     12,24%   91,84% ****
    2644,7 - 3120,0    2882,3         0      0,00%   91,84%
    3120,0 - 3595,3    3357,7         3      6,12%   97,96% **
          >= 3595,3    3833,0         1      2,04%  100,00%
```

Aldagai baten deskribapena, berriz, menu laguntzaileko *Ikusi* → *Estatistiko deskribatzaileak* aukeran agertzen da, edo menu nagusiko *Aldagaia* → *Estatistikoen laburpena* aukeran; bertan, lagineko batez bestekoa, desbideratze tipikoa, kurtosi-soberakina, aldakuntza-koefizientea, maximoa, minimoa eta asimetria- koefizientea agertzen dira. «Soldata» aldagaiari dagokiona, adibidez, honako hau da:


gretl: estatistikoen laburpena

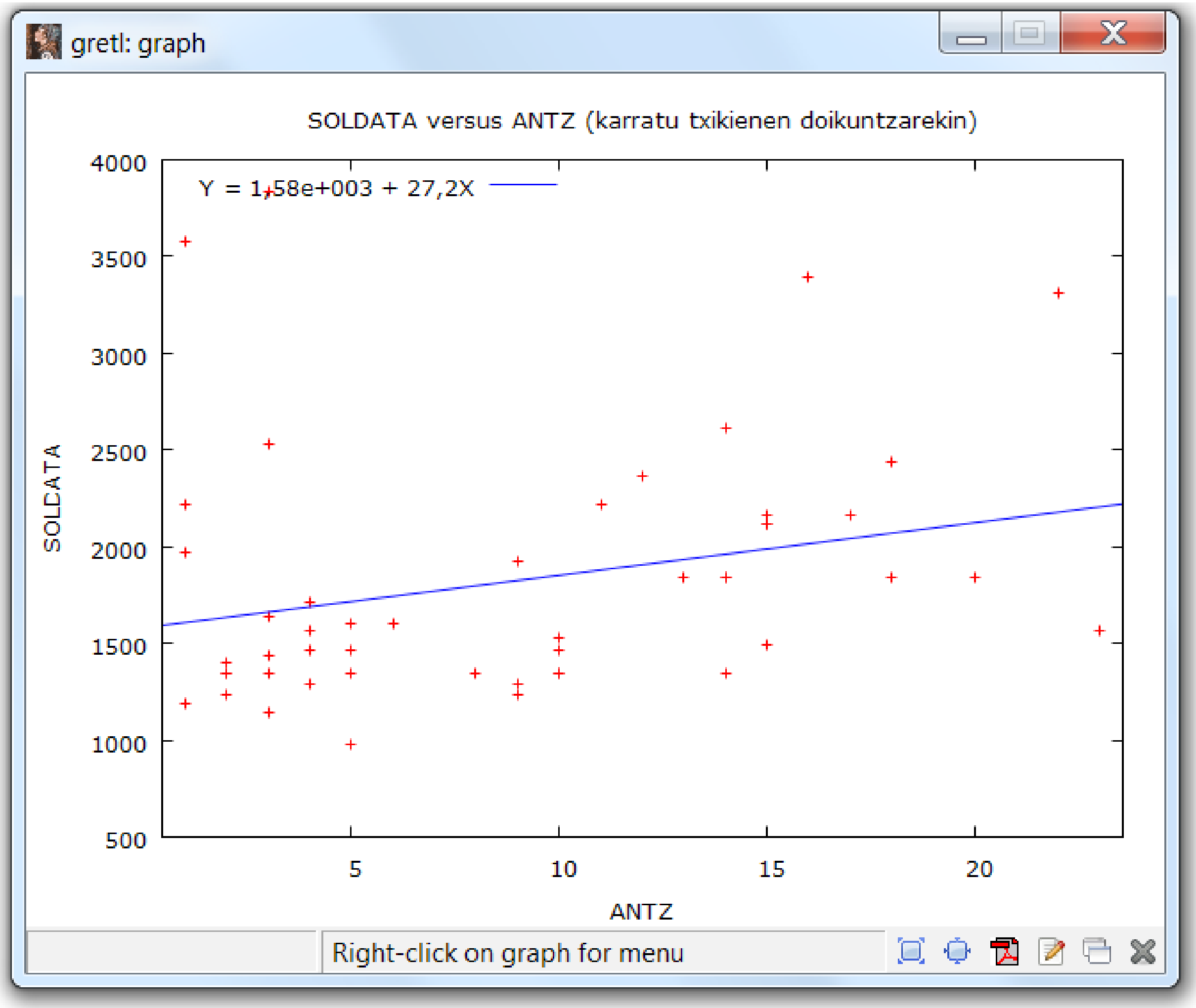
```
              Estatistikoen laburpena, 1 - 49 behaketak
erabiliz
            'SOLDATA' aldagaiarentzat (49 behaketa
baliagarraiak)

  Batezbestekoa                     1820,2
  Mediana                           1602,0
  Minimoa                            981,00
  Maximoa                           3833,0
  Desbiderazio tipikoa               648,27
  Aldakuntza Koefizientea (A.K.)       0,35615
  Asimetria                            1,4771
  Kurtosis soberakina                  1,7714
```

Bestalde, aldagaiak kuantitatiboak direnean, beren arteko erlazioa aztertzea komenigarria izaten da, grafikoak, kobariantzak eta korrelazioak aztertuz adibidez. Gretl softwareak aukera desberdinak eskaintzen dizkigu.

Hau da, bi aldagaien arteko erlazioaren ideia bat sakabanatze-diagramak eman diezaguke, planoan (X_i; Y_i); $i = 1, \ldots, N$ puntu guztiak adieraziz: aldagai bat (X) ardatz horizontalean adierazten da, eta beste aldagaia (Y), ordea, jatorri-ardatzean. Adibidez, Gretl erabiliz zazpigarren orriko diagrama lortzeko (soldata *vs* antzinatasuna), honako pauso hauek jarraitu behar dira:

— *Ikusi* → *Grafikoak* → *X-Y grafikoa* → *Definitu grafikoa*
X-ardatzaren aldagaia → *Aukeratu* → *Antzinatasuna*
Y-ardatzaren aldagaia → *Gehitu* → *Soldata*

Beste aukera bat *soldata* eta *antzinatasuna* aldagaiak aukeratzea da, saguaren ezkerreko tekla eta *control* batera klikatuz eta menu laguntzaileko *Ikusi* → *X-Y grafikoa* aukeratuz.



2. gaia

Erregresio Lineal Bakunaren Eredua. Zehaztapena

2.1. **Sarrera**

Gai honetan Erregresio Lineal Bakunaren Eredua (ELBE) aztertuko da, hau da, aldagai azaldua, aldagai azaltzaile bakarra (kuantitatiboa edo kualitatiboa), bi koefiziente eta perturbazioa dituen eredua.

$$Y_i = \alpha + \beta X_i + u_i \qquad i = 1, 2, \ldots, N$$

— Y_i aldagai endogenoa edo azaldua da.
— Xi, aldagai azaltzailea edo exogenoa da.
— Ekuazioaren lehen zatia ($\alpha + \beta X_i$) Populazioko Erregresio Funtzioa (PEF) da, eta, eredu bakuna izanik, Populazioko Erregresio Zuzena da. Populazioarentzat *X* eta *Y* aldagaien arteko batez besteko erlazioa jasotzen du, eta, beraz, *X* aldagaiaren balioa jakinez gero, zuzen hori erabiliz, *Y* aldagaiaren batez besteko balioa iragarri daiteke.
— α eta β koefizienteak edo parametroak dira (ezezagunak).
— u_i perturbazio aleatorioa da (ez behagarria), eta *Y* aldagaian eragiten duten beste faktoreen eragina jasotzen du.

Lagineko banako bakoitzarentzat (*i* bakoitzarentzat) eredua berridatziz, honako ekuazio hauek lortzen dira:

$$\left\{\begin{array}{ll} Y_1 = \alpha + \beta X_1 + u_1 & i = 1 \\ Y_2 = \alpha + \beta X_2 + u_2 & i = 2 \\ Y_3 = \alpha + \beta X_3 + u_3 & i = 3 \\ & \\ Y_N = \alpha + \beta X_N + u_N & i = N \end{array}\right.$$

Eredua matrizeak erabiliz idatz daiteke:

$$Y_{(Nx1)} = X_{(Nx2)}\,\beta_{(2x1)} + U_{(Nx1)}$$

$$Y_{(Nx1)} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_i \\ . \\ . \\ Y_N \end{pmatrix} \qquad X_{(Nx2)} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ 1 & X_i \\ \vdots & \vdots \\ 1 & X_N \end{pmatrix}$$

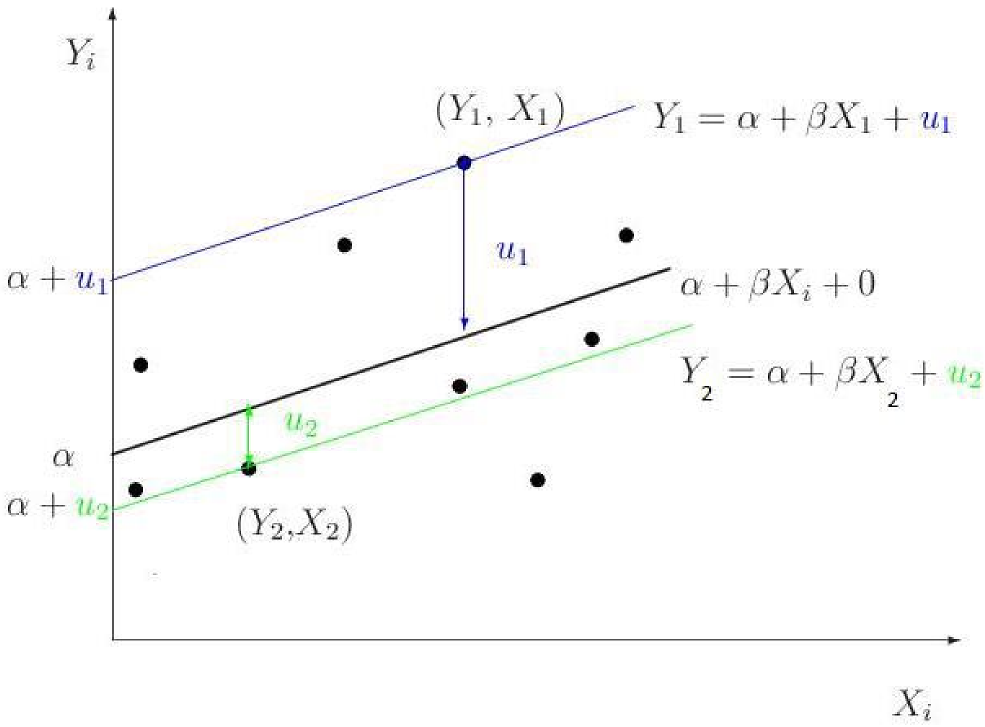
$$\beta_{(2x1)} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \qquad U_{(Nx1)} = \begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_N \end{pmatrix}$$

Aldagai azaltzailea kuantitatiboa denean

Ereduko aldagai azaltzailea kuantitatiboa denean, α termino independentea edo jatorria da, eta, β, berriz, malda da; honako hau da koefiziente hauen interpretazioa:

— α: Y-ren batez besteko balioa da, X zero denean (Oharra: eredu ekonometriko batzuetan jatorriak ez du interpretazio ekonomikorik izaten).
— β: Y-ren batez bestekoaren gehikuntza da, X unitate bat handitzen denean.

Ekuazioetan oinarrituz, (Y_i, X_i) puntu bakoitzetik β maldako zuzen bat pasatzen dela ikusi daiteke. Eta malda berekoak direnez, N zuzen guztiak paraleloak dira. Jatorriari dagokionez, zuzen bakoitzaren jatorria lortzeko α «batez besteko jatorri» ari zuzen bakoitzari dagokion perturbazioa (u_i) gehitu behar zaio; hau da, zuzen bakoitzaren jatorria $\alpha + u_i$ balioa da. Ondorengo grafikoan, N zuzen hauetatik bi marraztu dira, lehenengoko biak hain zuzen, lehena urdinez eta bigarrena berdez. Beltzez dagoen lerroa, berriz, «erdiko zuzena» da, perturbazioa zero denean lortzen den zuzena.



Lehen behaketari (Y_1, X_1) dagokion zuzena (lerro urdina) erdikoarekiko paraleloa da, malda berekoa baita, eta berekiko duen distantzia u_1 balioak neurtzen du. Bigarren behaketari (Y_2, X_2) dagokion lerroak ezaugarri berberak ditu, erdiko zuzenarekiko paraleloa da, eta u_2 distantziara agertzen da. Era berean marraztu daitezke zuzen guztiak.

Erregresio Lineal Bakunaren Eredu baten zenbatespenak ez du N zuzen guztien zenbatespena helburutzat, «erdiko zuzenaren» zenbatespena baizik, hau da, aldagai endogenoaren batez besteko portaeraren zenbatespena. Eredu bakun bat zenbatesterakoan datuei hoberen doitzen zaien malda eta batez «batez besteko jatorria» besteko jatorria zenbatetsi nahi dira. Ekonometrian oinarrituz, laginetik eratorritako (Y_1, X_1), (Y_2, X_2), …, (Y_N, X_N) behaketak ematen duten informazioa kontuan izanik, interesatzen den aldagaiaren (Y) batez besteko portaera ($\alpha + \beta X_i$) zein den kalkulatu nahi da.

Demagun *SOLDATA*ren adibidea. Langileen soldata aztertu nahi izanez gero enpresan duen antzinatasunaren funtzioan, honako hau da zehaztatuko den eredu bakuna:

$$SOLDATA_i = \alpha + \beta\, ANTZ_i + u_i \qquad i = 1, 2, ..., 49$$

Kasu honetan, α koefizienteak, langilearen batez besteko soldata (eurotan) jasotzen du enpresan sartu berria denean, hau da, *ANTZ* zero denean. Bestalde, β koefizienteak, berriz, langilearen antzinatasuna urtebete handitzen denean izango duen batez besteko soldataren gehikuntza jasotzen du; hau da, enpresan urte gehigarri bakoitzeko bere batez besteko soldatak izango duen gehikuntza (euro).

Matrizeak erabiliz, honako hau da eredu bakuna:

$$Y_{(49x1)} = X_{(49x2)}\ \beta_{(2x1)} + U_{(49x1)}$$

Non:

$$Y_{(49x1)} = \begin{pmatrix} SOLDATA_1 \\ SOLDATA_2 \\ SOLDATA_3 \\ SOLDATA_4 \\ \vdots \\ SOLDATA_{46} \\ SOLDATA_{47} \\ SOLDATA_{48} \\ SOLDATA_{49} \end{pmatrix} = \begin{pmatrix} 1345 \\ 2435 \\ 1715 \\ 1461 \\ \vdots \\ 2115 \\ 1839 \\ 1288 \\ 1288 \end{pmatrix}; X_{(49x2)} = \begin{pmatrix} 1 & ANTZ_1 \\ 1 & ANTZ_2 \\ 1 & ANTZ_3 \\ 1 & ANTZ_4 \\ 1 & ANTZ_5 \\ \vdots & \vdots \\ 1 & ANTZ_{46} \\ 1 & ANTZ_{47} \\ 1 & ANTZ_{48} \\ 1 & ANTZ_{49} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 18 \\ 1 & 4 \\ 1 & 4 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & 15 \\ 1 & 13 \\ 1 & 9 \\ 1 & 4 \end{pmatrix}$$

$$\beta_{(2x1)} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}; \qquad U_{(49x1)} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{47} \\ u_{48} \\ u_{49} \end{pmatrix}$$

Aldagai azaltzailea kualitatiboa denean

Aldagai azaltzailea kualitatiboa denean, dituen kategorien arabera, hainbat fikzio-aldagai definitu behar dira. Demagun enpresa bateko soldatak aztertu nahi direla, langileen generoaren arabera; adibidez, sexuarekiko soldataren diskriminaziorik dagoen ala ez frogatzeko. Hortaz, eredu ekonometrikoaren aldagai endogenoa soldata izango da (*SOLDATA*, eurotan), eta aldagai azaltzailea, berriz, sexua edo generoa. Azken horrek bi kategoria ditu (emakumea eta gizona), eta bakoitzarentzat fikziozko aldagaiak definituko dira. Adibidez, $EMAK_i$ fikzio-aldagaiak 1 balioa hartzen du langilea emakumea denean eta 0 gizona denean. Halaber, GIZ_i fikzio-aldagaiak 1 balioa hartzen du langilea gizona denean eta 0 emakumea denean. Zehaztatu daitezkeen eredu ekonometrikoak (baliokideak) honako hauek dira:

— $SOLDATA_i = \alpha_1 + \beta_1 EMAK_i + u_i \qquad i = 1, 2, ..., 49$

Kasu honetan, α_1 gizonezko baten batez besteko soldata da (eurotan), eta $\alpha_1 + \beta_1$, berriz, emakume den langile baten batez besteko soldata (eurotan). Beraz, β_1, langile emakume baten batez besteko soldataren diferentzia da (eurotan) gizon batek izango duenarekiko.

— $SOLDATA_i = \alpha_2 + \beta_2 GIZ_i + u_i \qquad i = 1, 2, ..., 49$

Kasu honetan, α_2 emakume baten batez besteko soldata da (eurotan), eta $\alpha_2 + \beta_2$, berriz, gizon langile baten batez besteko soldata (eurotan). Beraz, β_2, gizonezko langile baten batez besteko soldataren diferentzia da (eurotan) emakume batek izango duenarekiko.

Adibide honetan, sexuarekiko diskriminaziorik dagoen ala ez aztertzeko, batez besteko soldaten artean alderik dagoen ala ez aztertu beharko da. Hau da, lehen zehaztapenean, β_1 zero den ala ez aztertzea da, eta bigarrenean, aldiz, β_2 zero den ala ez. Aztertu nahi dena gizonezkoek emakumezkoekin alderatuz batez besteko soldata handiago dutela baldin bada, ordea, kontrastatu beharko den hipotesia $\beta_1 < 0$ izango da lehenengo zehaztapenean, eta $\beta_2 > 0$ bigarren zehaztapenean.

(OHARRA. Eredu bat zenbatetsita izanez gero, baliokidetasunak erabiliz bestea ere zenbatetsi ahal izango da; hau da, $\beta_1 + \alpha_1 = \alpha_2$ eta $\beta_2 + \alpha_2 = \alpha_1$ direla kontuan izanik.)

2.2. **Oinarrizko hipotesiak**

Eredua zehaztatu ondoren, zenbatespenaren testuingurua finkatu behar da. Horretarako, oinarrizko hipotesi batzuk betetzen direla pentsatuko da. Zenbatesten, doikuntzaren egokitasuna neurtzen, beharrezkoak diren kontrasteak egiten eta iragarpenak lortzen ikasten denean, oinarrizko hipotesi batzuk erlaxatuko dira. Baina, bitartean, hipotesi guztiak betetzen direla pentsatuko da.

Oinarrizko hipotesi hauek ereduko elementu desberdinei buruzkoak dira:

— Forma funtzionalari dagokionez: eredua koefizienteekiko lineala da. Kurtsoan zehar zenbatetsiko diren ereduak koefizienteekiko linealak edo linealizagarriak dira.
— Koefizienteei dagokienez: koefizienteak konstante mantentzen dira laginean zehar.
— Aldagai endogenoari dagokionez: aldagai endogenoa kuantitatiboa da. Oinarrizko ikasturte honetan zehar aldagai azaldua kuantitatiboa izango da, zeren aldagai dependente kualitatiboa daukan eredu bat zenbatesteko, ikasturte honetan ikasiko diren zenbatespen-metodoek ez baitute balio.
— Aldagai azaltzaileari dagokionez:

1. *X* aldagai azaltzailearen lagin-bariantza (Sx^2) ezin da zero izan, eta, gainera, behaketa kopuruak zenbatetsi behar diren koefiziente kopurua baino handiagoa izan behar du; hau da, eredu bakunean, $N > K = 2$ izan behar da. Hipotesi hau koefizienteak identifikatzeko beharrezkoa da. Hasteko, zenbatetsi behar diren koefizienteen kopurua behaketen kopurua baino handiago izanez gero, orduan zenbatespena aurrera eramateko ez daukagu informazio nahikorik. Bestalde, aldagai azaltzailearen lagin-bariantza zero izango balitz, orduan erregresio zuzenaren malda identifikatzea ezinezkoa izango litzateke.
2. *X* datu-matrizea hein osokoa da zutabeetan, eta, beraz, eredua zenbatesgarria izango da ($h(X) = 2$).

— Ereduari dagokionez, eredua ondo zehaztatuta dago. Orokorrean azalduz, ereduak ezin du aldagai esanguratsurik barneratu gabe utzi, eta ezta kontrakorik ere, aldagai ez-nabariren bat kontuan hartu. Eredu bakunean, hipotesi hau inposatzerakoan, *Y* aldagai dependentea azaltzeko behar den aldagai exogeno bakarra *X* aldagaia dela eskatzen ari gara.
— Perturbazioari dagokionez:

1. Perturbazioen populazio-batez bestekoa edo itxaropena zero da; hau da, zeron zentratutako aldagai aleatorioa da ($E_X(u_i) = 0$, $i = 1, 2, \dots, N$). Zenbatetsi ezina den errorearen batez bestekoa zero izateak ereduaren alde sistematikoa edo analizatu nahi den batez besteko portaera $E_X(Y_i) = \alpha + \beta X_i$ izatea dakar.

2. Perturbazioen populazio-bariantza konstantea da (homozedastizitatea). Aldagai aleatorioaren edo perturbazioaren aldakortasuna laginean zehar konstante mantentzen dela pentsatuko da. Hau da, $Bar(u_i) = E_X(u_i - E(u_i))^2 = E_X(u_i)^2 = \sigma_u^2$ konstantea da $i = 1, 2, \ldots, N$.
   Horrela, aldagai azaltzailearen balioak emanik, aldagai azalduak har ditzakeen balio posibleen tartearen zabalera berdina da, eta balio bakoitzak irteteko duen probabilitatea X aldagaiak hartzen duen balioarekiko independentea da.
   Kontrako kasuan, perturbazio heterozedastikoak dira, perturbazioen bariantzak laginean zehar aldatzen direlarik.
   Homozedastizitateak eta heterozedastizitateak duten interpretazioa ulertzeko, har dezagun soldataren adibidea. Langilearen antzinatasuna urte gutxikoa bada, langileen soldatek har ditzaketen balio posibleak oso antzekoak izateko probabilitatea handia da, eta, beraz, bariantza txikia dago. Baina langileen antzinatasuna handitzen doan heinean, soldatek har ditzaketen balio posibleen tartea handiagoa da, eta, beraz, bariantza handiagoa du. Horrela, bada, antzinatasun-urte gutxiko langileen soldatak nahiko antzekoak izan ohi dira, baina enpresan igarotako urteak handitzerakoan soldata posibleen tartea ere handitu egiten da, eta, ondorioz, antzinatasun handiko langileen soldatak oso desberdinak izan daitezke. Kasu honetan, heterozedastizitate kasuan egongo ginateke.
3. Perturbazioen artean ez dago autokorrelaziorik. Perturbazio desberdinen arteko korrelazioa zero da ($Kor(u_i, u_j) = r_{ui,uj} = 0;\ i \neq j$). Hortaz, beren arteko kobariantza ($Kob((u_i, u_j) = 0;\ i \neq j$) ere zero izango da.

$$Kob(u_i, u_j) = E_X[(u_i - E_X(u_i))(u_j - E_X(u_j))] = E_X(u_i\, u_j) = 0 \qquad i \neq j$$

4. Perturbazioek Banaketa Normala jarraitzen dute. Azken hipotesi hau, aurrerago ikusiko den bezala, ez da beharrezkoa eredua zenbatesteko, ezta zenbateslearen propietateak lortzeko ere. Inferentzia (kontrasteak) egiteko edota konfiantza-tarteak lortzeko izango da beharrezkoa.

   Perturbazioen propietate hauek batera idatz daitezke:

$$u_i \sim NIB(0, \sigma^2) \qquad i = 1, 2, \ldots, N$$

   Edota matrizeak erabiliz:

$$U_{(Nx1)} \sim N(0_{(Nx1)}, \sigma^2\, I_{(NxN)})$$

   Non:

$$E_X(U)_{(Nx1)} = \begin{pmatrix} E_X(u_1) \\ E_X(u_2) \\ \vdots \\ E_X(u_N) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0_N$$

$$Bar(U)_{NxN} = E_X\big(U - E_X(U)\big)\big(U - E_X(U)\big)' = E_X(U\,U') \quad =$$

$$= \begin{pmatrix} E_X(u_1)^2 & E_X(u_1\,u_2) & E_X(u_1\,u_3) & \dots & E_X(u_1\,u_N) \\ E_X(u_2\,u_1) & E_X(u_2)^2 & E_X(u_2\,u_3) & \dots & E_X(u_2\,u_N) \\ E_X(u_3\,u_1) & E_X(u_3\,u_2) & E_X(u_3)^2 & \dots & E_X(u_3\,u_N) \\ \vdots & \vdots & \vdots & & \vdots \\ E_X(u_N\,u_1) & E_X(u_N\,u_2) & E_X(u_N\,u_3) & \dots & E_X(u_N)^2 \end{pmatrix} =$$

$$= \begin{pmatrix} \sigma_u^2 & 0 & 0 & & \dots & 0 \\ 0 & \sigma_u^2 & 0 & & & 0 \\ 0 & & \sigma_u^2 & & & \vdots \\ \vdots & & & \ddots & & \\ & & & & \sigma_u^2 & 0 \\ 0 & 0 & \dots 0 & & 0 & \sigma_u^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & & \dots & 0 \\ 0 & 1 & 0 & & & 0 \\ 0 & & 1 & & & \vdots \\ \vdots & & & \ddots & & \\ & & & & 1 & 0 \\ 0 & 0 & \dots 0 & & 0 & 1 \end{pmatrix} = \quad \sigma^2\, I_N$$

3. gaia

Erregresio Lineal Bakunaren Eredua. Zenbatespena

3.1. Karratu Txikien Arrunten (KTA) zenbateslea

Demagun honako eredu ekonometriko bakun hau, non oinarrizko hipotesi guztiak betetzen baitira.

$$Y_i = \alpha + \beta X_i + u_i \qquad i = 1, 2, ..., N$$

Aurreko gaietan aipatu bezala, ereduko koefizienteak ezezagunak direnez, aldagai endogenoaren (Y) batez besteko portaera ere ezezaguna da; hau da, ereduko alde sistematikoa edo Populazioko Erregresio Funtzioa (PEF) ezezaguna da.

$$\text{PEF} \longrightarrow \text{E}_X(\text{Y}_i) = \alpha + \beta X_i \qquad i = 1, 2, \ldots, N$$

Hortaz, ereduko perturbazioa aldagai azalduaren benetako balioaren eta Populazioko Erregresio Funtzioaren arteko diferentzia da, eta ereduko alde sistematikoak azaldu ezin duena jasotzen du ($u_i = Y_i - \alpha - \beta X_i \;\; i = 1, 2, ..., N$).

Ekonometriaren helburua da eredu ekonometrikoa ahalik eta hobekien zenbatestea; beraz, ahalik eta hondar ($\hat{u}_i$ txikienak lortzea du xede, hau da, aldagai azalduaren benetako balioen eta zenbatetsitako balioen arteko diferentziak ahalik eta txikienak izatea. Hortaz, Karratu Txikien Arrunten (KTA) bitartez eredua zenbatesterakoan minimizatuko den helburu-funtzioa Hondar Karratuen Batura (HKB) izango da.

Eredu ekonometriko bakun batean bi koefiziente izanik ($K = 2$):

$$min_{\hat{\alpha}\,\hat{\beta}} \sum_{i=1}^{N} \hat{u}_i^2 \equiv min_{\hat{\alpha}\,\hat{\beta}} \sum_{i=1}^{N} (Y_i - \hat{\alpha} - \hat{\beta}\, X_i)^2$$

Hemendik, lehen ordenako baldintzak deribatuz, bi Ekuazio Normal lortzen dira:

$$\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\alpha}} = 0 \rightarrow 2\sum_{i=1}^{N}\big(Y_i - \hat{\alpha} - \hat{\beta}\, X_i\big)(-1) = 0 \;\; \rightarrow \sum_{i=1}^{N} Y_i \; = \; N\,\hat{\alpha} \; + \hat{\beta}\,\sum_{i=1}^{N} X_i$$

$$\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\beta}} = 0 \rightarrow 2\sum_{i=1}^{N}\big(Y_i - \hat{\alpha} - \hat{\beta}\, X_i\big)(-X_i) = 0 \;\; \rightarrow \sum_{i=1}^{N} Y_i\, X_i = \hat{\alpha}\; \sum_{i=1}^{N} X_i + \hat{\beta}\; \sum_{i=1}^{N} X_i^2$$

Ekuazio Normal hauek askatuz, ereduko bi koefizienteen KTA zenbatesleak lortzen dira:

$$\hat{\alpha} \; = \; \bar{Y} - \hat{\beta}\; \bar{X}$$

$$\hat{\beta} \; = \; \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N} X_i\, Y_i - N\,\bar{X}\,\bar{Y}}{\sum_{i=1}^{N} X_i^2 - N\,\bar{X}^2} = \frac{S_{XY}}{S_X^2}$$

Dagokion Lagineko Erregresio Funtzioa honako hau da:

$$\text{LEF} \;\longrightarrow\; \widehat{E_X(Y_i)} = \hat{Y}_i = \hat{\alpha} + \hat{\beta}\, X_i \qquad i = 1, 2, \dots N$$

Soldataren adibidean:

$$\sum_{i=1}^{49} SOLDATA_i = 89180, \;\; \sum_{i=1}^{49} ANTZ_i = 433,$$

$$\sum_{i=1}^{49} SOLDATA_i^2 = 182515200, \;\; \sum_{i=1}^{49} ANTZ_i^2 = 5706$$

$$\sum_{i=1}^{49} SOLDATA_i\, ANTZ_i = 839174$$

Eta, beraz, dagozkion Ekuazio Normalak honako hauek dira:

$$89190 = \; 49\,\hat{\alpha} + \hat{\beta}\;\; 433$$

$$839174 = \; \hat{\alpha}\, 433 + \hat{\beta}\, 5706$$

Askatuz, $\hat{\alpha}$ = 1580,29 eta $\hat{\beta}$ = 27,15 direnez, honako hau da dagokion Lagineko Erregresio Funtzioa, hau da, zenbatetsitako eredua :

$$\widehat{SOLDATA_i} \; = 1580{,}29 + 27{,}15\, ANTZ_i \qquad i = 1{,}2, \dots, 49$$

Edozein eredu ekonometrikotan, matrizeak erabiliz ere lor daitezke Ekuazio Normalak, eta, beraz, baita eredu bakun batean ere. Aipatu den bezala, Karratu Txikien Arrunten (KTA) helburu-funtzioa da Hondar Karratuen Batura (*HKB*) minimo egitea. *HKB* honela lortzen da:

$$HKB = \sum_{i=1}^{N} \hat{u}_i^2 \; = \; \hat{U}'\hat{U}$$

Non $\hat{U} = Y - \hat{Y} = Y - X\,\hat{\beta}$ baita, hau da, $N \times 1$ zutabe bektorea. Hortaz, zenbatespen-irizpidea honela jarriko da:

$$min_{\hat{\beta}}\ HKB \ \equiv min_{\hat{\beta}}\ \hat{U}'\hat{U} \ \equiv min_{\hat{\beta}}\left(Y - X\,\hat{\beta}\right)'(Y - X\,\hat{\beta})$$

$$\frac{\partial\,\hat{U}'\hat{U}}{\partial\,\hat{\beta}} = 0 \ \rightarrow\ -2\,X'\left(Y - X\,\hat{\beta}\right) = 0\ \rightarrow X'Y = X'X\,\hat{\beta}\ \rightarrow \hat{\beta} = \ (X'X)^{-1}\,X'Y$$

Non:

$$\hat{\beta} = \begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}\ ,\quad X'X = \begin{pmatrix} N & \sum_{i=1}^{N} X_i \\ \sum_{i=1}^{N} X_i & \sum_{i=1}^{N} X_i^2 \end{pmatrix}\ ,\quad X'Y = \begin{pmatrix} \sum_{i=1}^{N} Y_i \\ \sum_{i=1}^{N} X_i\,Y_i \end{pmatrix}$$

Soldaten adibidera bueltatuz:

$$X'X = \begin{pmatrix} N & \sum_{i=1}^{N} ANTZ_i \\ \sum_{i=1}^{N} ANTZ_i & \sum_{i=1}^{N} ANTZ_i^2 \end{pmatrix} = \begin{pmatrix} 49 & 433 \\ 433 & 5706 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 49 & 433 \\ 433 & 5706 \end{pmatrix}^{-1} = \frac{1}{92105}\begin{pmatrix} 5706 & -433 \\ -433 & 49 \end{pmatrix} = \begin{pmatrix} 0{,}0619 & -0{,}0047 \\ -0{,}0047 & 0{,}000532 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} \sum_{i=1}^{N} SOLDATA_i \\ \sum_{i=1}^{N} ANTZ_i\,SOLDATA_i \end{pmatrix} = \begin{pmatrix} 89190 \\ 839174 \end{pmatrix}$$

$$\hat{\beta} = (X'X)^{-1}\,X'Y = \begin{pmatrix} 1580{,}29 \\ 27{,}15 \end{pmatrix}\ \rightarrow\ \widehat{SOLDATA}_i = 1580{,}29 + 27{,}15\,ANTZ_i$$

$$i = 1, \dots, 49$$

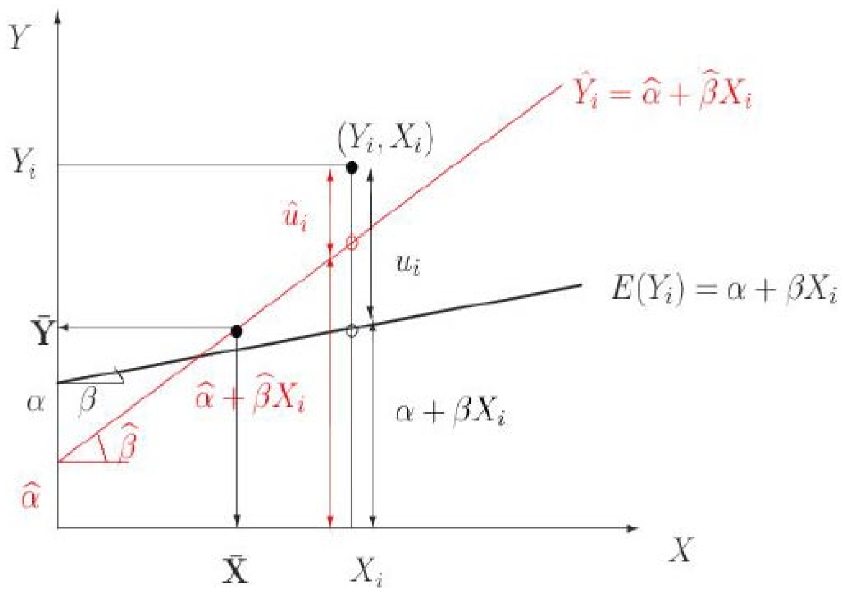
Ereduko koefizienteen zenbatespenen interpretazioa honako hau da:

— $\hat{\alpha} = 1580{,}29$ → Enpresan sartu berria den ($ANTZ_i = 0$) langile baten hileroko batez besteko soldata 1.580,29 eurokoa dela zenbatesten da.

— $\hat{\beta}$ = 27,15 $\rightarrow$ Lan-urte gehigarri bakoitzeko (*ANTZ* urte bat handitzen denean), hileroko batez besteko soldata 27,15 euro handitzen dela zenbatesten da.

Aurretik adierazi bezala, ereduko koefiziente guztiak zenbatetsiz, Lagineko Erregresio Funtzioa (LEF) lortzen da, eta, horrekin, edozein langileren hileroko batez besteko soldataren zenbatespena eta dagokion **hondarra** ($\hat{u}_i = SOLDATA_i - \widehat{SOLDATA_i}$). Hondarra, zenbatespen-errore bat da, eta eredua zenbatesterakoan egiten diren errore guztiak biltzen ditu. Bi motatako erroreak bereiziko dira. Bata, koefizienteak ezezagunak izanik, horiek zenbatestean egiten den erroretik dator, eta bestea, zenbatetsi ezinezko perturbazioa (u_i) dagoelako ereduan (garrantzitsua da perturbazioak eta hondarrak bereiztea eta ez nahastea).

Orokorrean, $\hat{u}_i = Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i) \quad i = 1, 2, ..., N$



Grafikoan, beltzez, Populazioko Erregresio Funtzioa dago irudikatuta. Bertan, erraz ikusi daiteke, aldagai azalduaren edozein *i* behaketa (Y_i) lortzeko, ereduaren zati sistematikoari, hau da ($\alpha + \beta X_i$) terminoari, perturbazioa (u_i) gehitu egin behar zaiola ($Y_i = \alpha + \beta X_i + u_i$).

Lagineko Erregresio Funtzioa eta zenbatetsitako koefizienteak gorriz adierazita daude. PEF eta LEF arteko diferentzia, koefizienteak zenbatesterakoan egiten diren erroreengatik ematen da; hau da, $\alpha \neq \hat{\alpha}$ eta $\beta \neq \hat{\beta}$ delako.

LEF oinarritzat hartuz, aldagai azalduaren edozein *i* behaketa (Y_i) lortzeko, zenbatetsitako zati sistematikoari ($\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$) dagokion hondarrak hartzen duen balioa ($\hat{u}_i$) gehitu behar zaio, horrela $Y_i = \hat{Y}_i + \hat{u}_i$ lortuz. Azkenik, aipatu LEF batez bestekoen puntutik ($\bar{Y}, \bar{X}$) pasatzen dela.

3.2. Lagineko Erregresio Funtzioaren (LEF) propietateak

Eredu ekonometriko bakun batean, honako bost propietate hauek betetzen dira (frogapenak Erregresio Lineal Orokorraren Ereduan egingo dira).

1. Hondarrak eta aldagai azaltzailea ortogonalak dira: $\sum_{i=1}^{N} X_i \hat{u}_i = 0$.
2. Hondarrak eta zenbatetsitako aldagaia ortogonalak dira: $\sum_{i=1}^{N} \hat{Y}_i \hat{u}_i = 0$.
3. Hondarren batura zero da: $\sum_{i=1}^{N} \hat{u}_i = 0$.
4. *Y*ren lagineko batez bestekoa eta *Y*ren zenbatespenen lagineko batez bestekoa berdinak dira: $\bar{Y} = \bar{\hat{Y}}$.
5. LEF lagineko batez bestekoen puntutik pasatzen da: $\bar{Y} = \hat{\alpha} + \hat{\beta}\,\bar{X}$.

3.3. Doikuntza-neurria: mugatze-koefizientea (R^2)

Eredua zenbatetsita dagoenean, Lagineko Erregresio Funtzioa datuekiko nola doitzen den analizatzea interesatzen da. Hau da, zenbatetsitako funtzioa eta datuak nola doitzen diren.

Aldagai endogenoaren balio guztiak berdinak izango balira, horiek eta aldagai endogenoaren lagineko batez bestekoa berdinak lirateke; horrela, aldagai endogenoaren aldakuntza zero izango da. Orokorrean, aldagai endogenoaren aldakortasuna, behatutako balioen eta aldagai horren lagineko batez bestekoaren arteko distantzia bezala definitzen da, eta kalkulatutako aldakortasunaren karratuen batura (*KTB*: Karratu Totalaren Batura) aldagai azaltzaileekin azaldu nahi den aldakuntza da. Horrela, zenbat eta handiagoa izan azaldutako aldakortasunaren proportzioa, hainbat eta hobe izango da egindako doikuntza.

Mugatze-koefizientea (R^2) erregresioaren doikuntza-neurri bat da. Aldagai azaltzaileak zenbat azaltzen duen aldagai endogenoaren aldakuntza neurtzen du, hau da, aldagai endogenoaren Karratu Totalaren Baturaren (*KTB*) zein ehuneko azaltzen duen.

$$KTB = \sum_{i=1}^{N}(Y_i - \bar{Y})^2 = \sum_{i=1}^{N} Y_i^2 - N\,\bar{Y}^2 = Y'Y - N\,\bar{Y}^2$$

Aurretik azaldu bezala, $Y_i = \hat{Y}_i + \hat{u}_i$ da, eta zenbatetsitako balioak eta hondarrak ortogonalak direnez,

$$\sum_{i=1}^{N} Y_i^2 = \sum_{i=1}^{N}(\hat{Y}_i + \hat{u}_i)^2 = \sum_{i=1}^{N}\hat{Y}_i^2 + \sum_{i=1}^{N}\hat{u}_i^2 + 2\sum_{i=1}^{N}\hat{Y}_i\hat{u}_i = \sum_{i=1}^{N}\hat{Y}_i^2 + \sum_{i=1}^{N}\hat{u}_i^2$$

Bi aldeetan $N\bar{Y}^2$ kenduz eta $\bar{Y} = \bar{\hat{Y}}$ denez,

$$\sum_{i=1}^{N} Y_i^2 - N\,\bar{Y}^2 = \sum_{i=1}^{N}\hat{Y}_i^2 - N\,\bar{Y}^2 + \sum_{i=1}^{N}\hat{u}_i^2$$

$$\sum_{i=1}^{N} Y_i^2 - N\,\bar{Y}^2 = \sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{N}\hat{u}_i^2$$

$$\sum_{i=1}^{N} Y_i^2 - N\,\bar{Y}^2 = \sum_{i=1}^{N}(\hat{Y}_i - \bar{\hat{Y}})^2 + \sum_{i=1}^{N}\hat{u}_i^2 \qquad \rightarrow \qquad KTB = KAB + HKB$$

Erregresioaren mugatze-koefizientea honela definitzen da:

$$R^2 = \frac{KAB}{KTB} = 1 - \frac{HKB}{KTB}$$

Eredu bakunean, gainera, froga daiteke mugatze-koefizientea aldagai endogenoaren eta aldagai azaltzailearen arteko korrelazio-koefizientearen berbidura dela, hau da, $R^2 = (r_{XY})^2$. Bestalde, eredu bakunean, mugatze-koefizientea [0, 1] tartean egongo da, 1etik hurbil dagoenean, doikuntza hobea izanik.

3.4. **Ereduko koefizienteen KTA zenbatesleen lagin txikitako propietateak**

Demagun Erregresio Lineal Bakunaren Eredua non oinarrizko hipotesiak betetzen baitira.

$$Y_i = \alpha + \beta X_i + u_i \qquad i = 1, 2, ..., N$$

Ereduko koefizienteen KTA zenbatesleak honako hauek dira:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\ \bar{X}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N} X_i\, Y_i - N\,\bar{X}\,\bar{Y}}{\sum_{i=1}^{N} X_i^2 - N\,\bar{X}^2} = \frac{S_{XY}}{S_X^2}$$

Lagin txikitan honako propietate hauek betetzen dituzte (frogapenak eredu orokorrean):

- Linealtasuna. KTA zenbateslea perturbazioen (baita aldagai endogenoaren) konbinazio lineal bat dela kontuan harturik, perturbazioekiko lineala da.
- Alboragabetasuna. $E_X(u) = 0$ dela jakinik, alboragabeak dira, hau da, bere itxaropena ereduko benetako koefizienteen berdina da: $E_X(\hat{\beta}) = \beta$.
- Bariantza minimodunak dira.

$$Bar(\hat{\beta}) = E_X\big(\hat{\beta} - E(\hat{\beta})\big)\big(\hat{\beta} - E(\hat{\beta})\big)' = \begin{pmatrix} Bar(\hat{\alpha}) & Kob(\hat{\alpha},\hat{\beta}) \\ Kob(\hat{\alpha},\hat{\beta}) & Bar(\hat{\beta}) \end{pmatrix} =$$

$$= \sigma^2 \begin{pmatrix} \dfrac{\sum_{i=1}^{N} X_i^2}{N\,\sum_{i=1}^{N}(X_i - \bar{X})^2} & \dfrac{-\bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \\ \dfrac{-\bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} & \dfrac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \end{pmatrix} = \sigma^2 \begin{pmatrix} \dfrac{\frac{1}{N}\sum_{i=1}^{N} X_i^2}{S_X^2} & \dfrac{-\bar{X}}{N\,S_X^2} \\ \dfrac{-\bar{X}}{N\,S_X^2} & \dfrac{1}{N\,S_X^2} \end{pmatrix}$$

Gauss-Markov-en Teorema: erregresio linealeko ereduaren oinarrizko hipotesiak kontuan izanik, zenbatesle lineal eta alboragabe guztien artetik, KTA zenbateslea efizientea da, hau da, bariantza txikiena du.

Propietate horiek kontuan izanik eta oinarrizko hipotesiak betetzen direnean, ereduko bi koefizienteen KTA zenbatesleen banaketak honako hauek dira:

$$\hat{\alpha} \sim N\Big(\alpha\ ,\ \sigma^2 \frac{\sum_{i=1}^{N} X_i^2}{N\,\sum_{i=1}^{N}(X_i - \bar{X})^2}\Big)$$

$$\hat{\beta} \sim N\Big(\beta\ ,\ \sigma^2 \frac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\Big)$$

3.5. Perturbazioen bariantzaren zenbateslea

Perturbazioen bariantza ezezaguna izaten da, eta, hemendik aurrera, honako zenbatesle alboragabe hau erabiliko da:

$$\hat{\sigma}^2 = \frac{\hat{U}'\hat{U}}{N-K} = \frac{\hat{U}'\hat{U}}{N-2} = \frac{HKB}{N-2}$$

Frogapena:

$Y = X\,\beta + U$ eredua izanik non oinarrizko hipotesiak betetzen baitira, $\hat{U} \sim N(0, \sigma^2\, M)$ betetzen dela froga daiteke, non $M = I_N - X\,(X'X)^{-1}\,X'$ matrizea, simetrikoa, idenpotentea, heina $N - K = N - 2$ eta X matrizearekiko ortogonala baita.

$$\hat{U} = Y - \hat{Y} = Y - X\,\beta = Y - X\,(X'X)^{-1}X'Y = (I_N - X\,(X'X)^{-1}X')Y =$$

$$= (I_N - X\,(X'X)^{-1}X')(X\beta + U) = X\beta - X\,(X'X)^{-1}X'X\,\beta + (I_N - X\,(X'X)^{-1}X')\,U =$$

$$(I_N - X\,(X'X)^{-1}X')\,U = M\,U$$

— M simetrikoa da: $M = M'$

$$M' = (I_N - X\,(X'X)^{-1}X')' = I'_N - (X\,(X'X)^{-1}X')' =$$
$$= I_N - (X')'\,[(X'X)^{-1}]'X' = I_N - X\,(X'X)^{-1}X' = M$$

— M idenpotentea da: $MM = M$

$$MM = (I_N - X\,(X'X)^{-1}X')(I_N - X\,(X'X)^{-1}X') =$$
$$= I_N - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X'$$
$$= I_N - X(X'X)^{-1}X' = M$$

— $h(M) = tr(M) = N - K$

$$tr(M) = tr[I_N - X(X'X)^{-1}X'] = tr(I_N) - tr[X(X'X)^{-1}X'] =$$
$$= tr(I_N) - tr[(X'X)^{-1}X'X] = tr(I_N) - tr(I_K) = N - K$$

— M matrizea X datu-matrizearekiko ortogonala da: $MX = 0$

$$MX = [I_N - X\,(X'X)^{-1}X'\,]X = X - X\,(X'X)^{-1}X'X = X - X = 0$$

$\hat{U} = M\,U$ izanik,

$$E_X(\hat{U}) = M\,E_X(U) = M\,0 = 0$$

$$Bar(\hat{U}) = E_X(\hat{U}\hat{U}') = E_X(MUU'M') =$$

$$= M\,E_X(UU')M' = M\,\sigma^2 I_N\,M' = \sigma^2 MM' = \sigma^2 MM = \sigma^2 M$$

M matrizearen diagonal nagusiko elementu guztiak ez dira berdinak, eta, ondorioz, hondarrek ez dute bariantza berdina. Bestalde, diagonal nagusitik kanpoko elementu guztiak ez dira zero; hortaz, hondarrak ez dira korrelatu gabeak. Nahiz eta perturbazioak homozedastikoak eta autokorrelatu gabeak izan, hondarrekin EZ DA BERDINA gertatzen.

Horrela,

$$E_X(\hat{\sigma}^2) = E_X\left(\frac{\hat{U}'\hat{U}}{(N-K)}\right) = \frac{E_X(\widehat{U}'\widehat{U})}{(N-K)} = \frac{E_X(U'M'\,M\,U)}{(N-K)} = \frac{E_X(U'M\,U)}{(N-K)} =$$

$$\frac{E_X(tr(U'\,M\,U))}{(N-K)} = \frac{E_X(tr(M\,U\,U')}{(N-K)} = \frac{tr(E_X(M\,U\,U'))}{(N-K)} = \frac{tr(M\,E_X(\,UU'))}{(N-K)} =$$

$$\frac{tr(M\,\sigma^2\,I_N)}{(N-K)} = \frac{\sigma^2 tr(M\,I_N)}{(N-K)} = \frac{\sigma^2 tr(M\,)}{(N-K)} = \frac{\sigma^2(N-K)}{(N-K)} = \sigma^2$$

3.6. Ereduko koefizienteen KTA zenbatesleen bariantza-kobariantza matrizearen zenbateslea

Erregresio Lineal Bakunaren Eredu batean oinarrizko hipotesiak betetzen direnean, aurreko ataletan ikusi denez, honako hau da ereduko koefizienteen KTA zenbateslearen bariantza-kobariantza matrizea:

$$Bar(\hat{\beta}) = \sigma^2 \begin{pmatrix} \dfrac{\sum_{i=1}^N X_i^2}{N\,\sum_{i=1}^N (X_i-\bar{X})^2} & \dfrac{-\bar{X}}{\sum_{i=1}^N (X_i-\bar{X})^2} \\ \dfrac{-\bar{X}}{\sum_{i=1}^N (X_i-\bar{X})^2} & \dfrac{1}{\sum_{i=1}^N (X_i-\bar{X})^2} \end{pmatrix}$$

Eta perturbazioen bariantzaren zenbatesle alboragabea, berriz, $\hat{\sigma}^2 = \frac{\hat{U}'\hat{U}}{N-2}$. Hortaz, ereduko koefizienteen KTA zenbatesleen bariantza-kobariantza matrizearen zenbateslea honako hau izango da:

$$\widehat{Bar}(\hat{\beta}) = \hat{\sigma}^2 \begin{pmatrix} \dfrac{\sum_{i=1}^N X_i^2}{N\,\sum_{i=1}^N (X_i-\bar{X})^2} & \dfrac{-\bar{X}}{\sum_{i=1}^N (X_i-\bar{X})^2} \\ \dfrac{-\bar{X}}{\sum_{i=1}^N (X_i-\bar{X})^2} & \dfrac{1}{\sum_{i=1}^N (X_i-\bar{X})^2} \end{pmatrix}$$

3.7. Ereduko koefizienteen konfiantza-tartea

Ikusi bezala, Erregresio Lineal Bakunaren Eredu batean oinarrizko hipotesiak betetzen direnean:

$$\hat{\alpha} \sim N(\alpha \ , \ \sigma^2 \frac{\sum_{i=1}^{N} X_i^2}{N \sum_{i=1}^{N}(X_i - \bar{X})^2})$$

$$\hat{\beta} \sim N(\beta \ , \ \sigma^2 \frac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2})$$

Bestalde, froga daiteke $\frac{\hat{\sigma}^2 (N-2)}{\sigma^2} \sim \aleph^2_{(N-2)}$ dela, eta, gainera, $\hat{\sigma}^2$ eta $\hat{\beta}$ independenteak direnez, maldaren kasuan adibidez:

$$t = \frac{\hat{\beta} - \beta}{\hat{\sigma}\sqrt{\frac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2}}} = \frac{\hat{\beta} - \beta}{\widehat{des}\,(\hat{\beta})} \sim t_{(N-2)}$$

Hortaz:

$$\Pr\left[- t_{(N-2)\alpha/2} \le \frac{\hat{\beta} - \beta}{\widehat{des}\,(\hat{\beta})} \le \ -t_{(N-2)\alpha/2}\right] = 1 - \alpha$$

$$\Pr\left[\hat{\beta} - t_{(N-2)\alpha/2}\,\widehat{des}\,(\hat{\beta}) \le \beta \le \hat{\beta} + t_{(N-2)\alpha/2}\,\widehat{des}\,(\hat{\beta})\right] = 1 - \alpha$$

$$\mathrm{KT}(\beta)_{1-\alpha} = \left[\hat{\beta} - t_{(N-2)\alpha/2}\,\widehat{des}\,(\hat{\beta})\ ;\ \hat{\beta} + t_{(N-2)\alpha/2}\,\widehat{des}\,(\hat{\beta})\right]$$

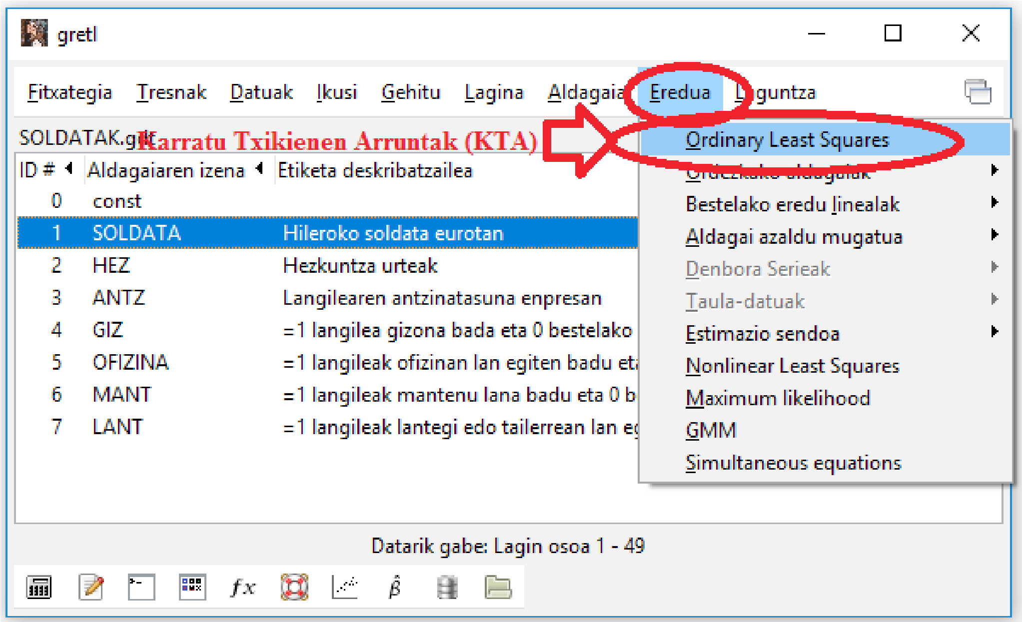
3.8. Zebatespena Gretl-ekin

Gai honetan jorratu dena Gretl-ekin nola gauzatzen den azalduko da jarraian; hau da, eredu bat nola zenbatesten den, koefizienteen zenbatesleen bariantza-kobariantza matrizea nola lortzen den, ereduko koefizienteen konfiantza-tarteak, etab.

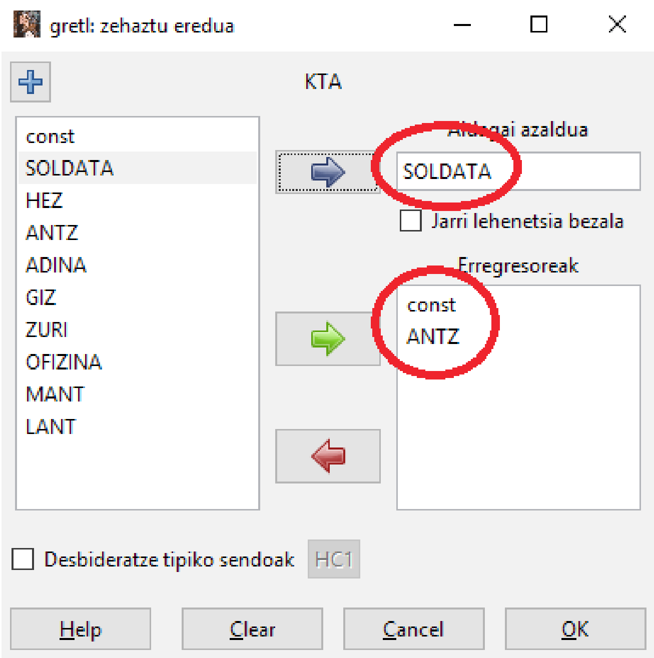
SOLDATAren adibidearekin jarraituz,

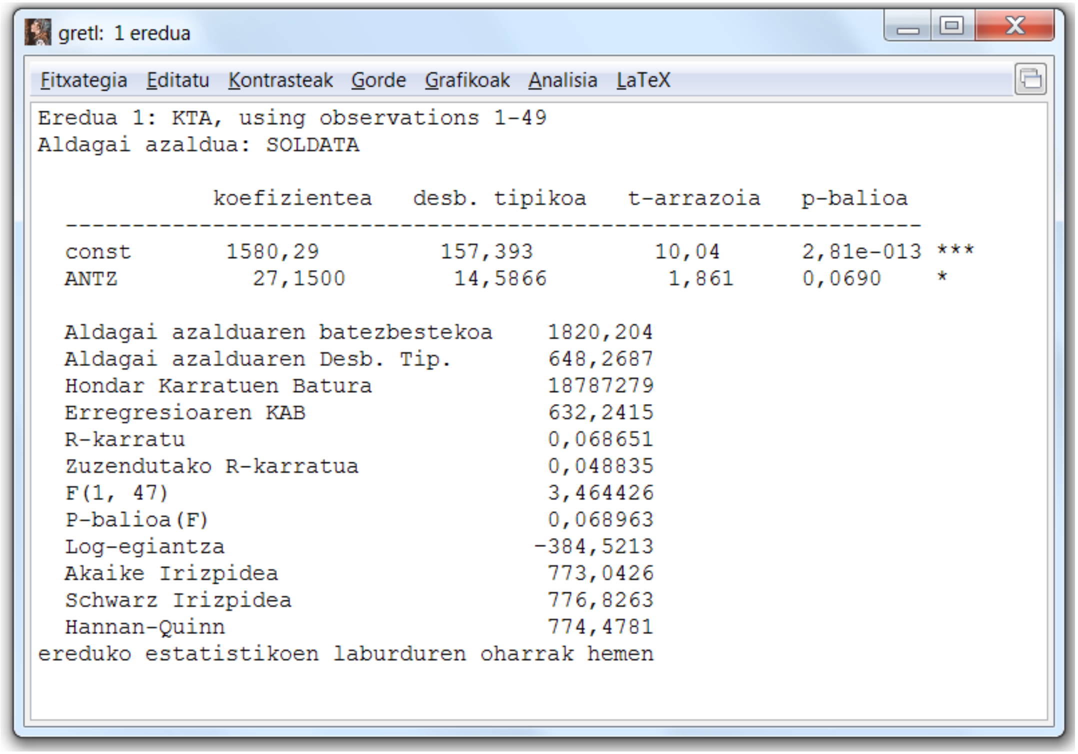
$$SOLDATA_i = \alpha + \beta\, ANTZ_i + u_i \qquad i = 1, 2, ..., 49$$

Eredu hau KTA bitartez zenbatesteko, *Eredua* aukeratu eta *Karratu Txikienen Arrunta* klikatuko da.

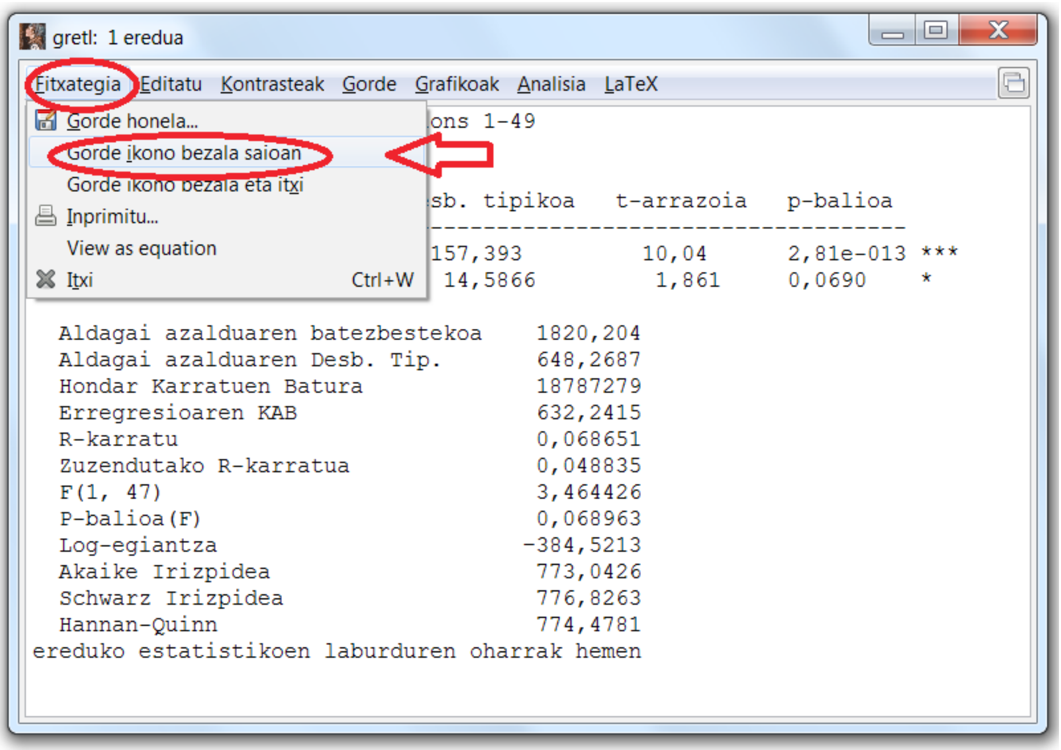


Jarraian, aldagai azaldua eta azaltzaileak aukeratuko dira, eta *Ok* klikatuko zenbatespen-emaitzak lortzeko:

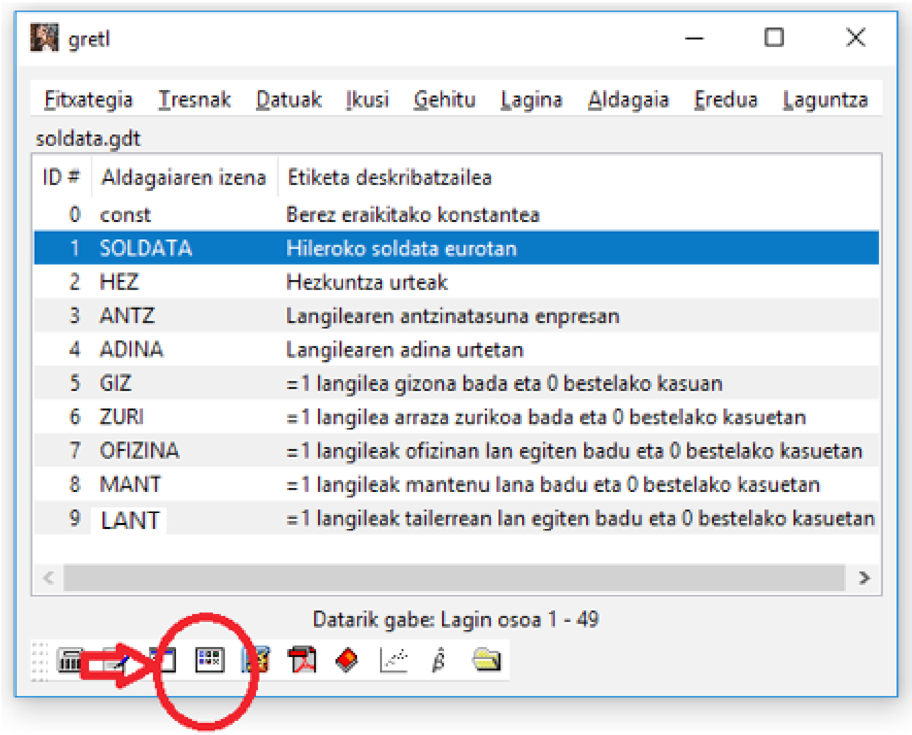


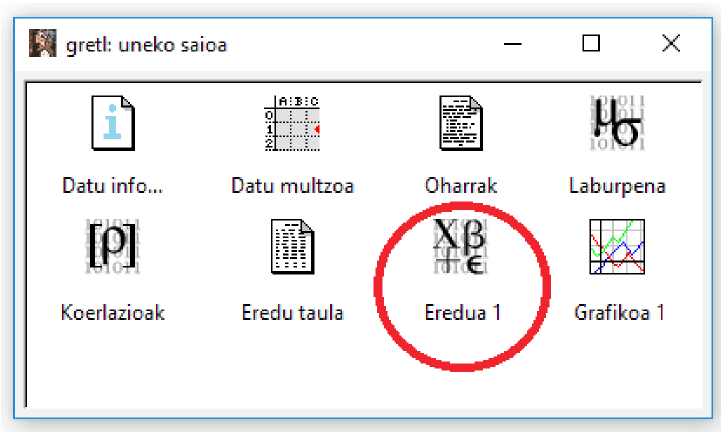


Emaitzak aztertzen hasi baino lehen, nahi izanez gero, eredua ikono bezala gorde daiteke, geroago berreskuratzeko asmoarekin. Gordetzeko *Fitxategia* aukeratu, eta *Gorde ikono bezala saioan* klikatu:



Horrela, eredua ikono bezala gordeta gelditzen da GRETL programaren barneko USER karpetan. Berreskuratzeko, beheko barrako *ikonoen menuan* (ezkerretik hasita laugarren botoia) aukeratu *saioaren ikono ikuspegia*, eta, *Eredua 1* jartzen duen ikonoari klik bikoitza emanik, ereduen zenbatespen-emaitzak berreskuratzen dira.





Emaitzen taulara itzuliz, lehen zutabean ereduan barneratutako aldagai azaltzaileen izenak ageri dira, kasu honetan, langilearen antzinatasuna. Bigarren zutabean, aldiz, ereduko koefizienteen KTA zenbatespenak dauzkagu: $\hat{\alpha} = 1580{,}29$ eta $\hat{\beta} = 27{,}15$. Horrela, kasu honetan, honako hau da dagokion Lagineko Erregresio Funtzioa:

$$\widehat{SOLDATA_i} = 1580{,}29 + 27{,}15\ ANTZ_i \qquad i = 1, 2, \dots, N$$

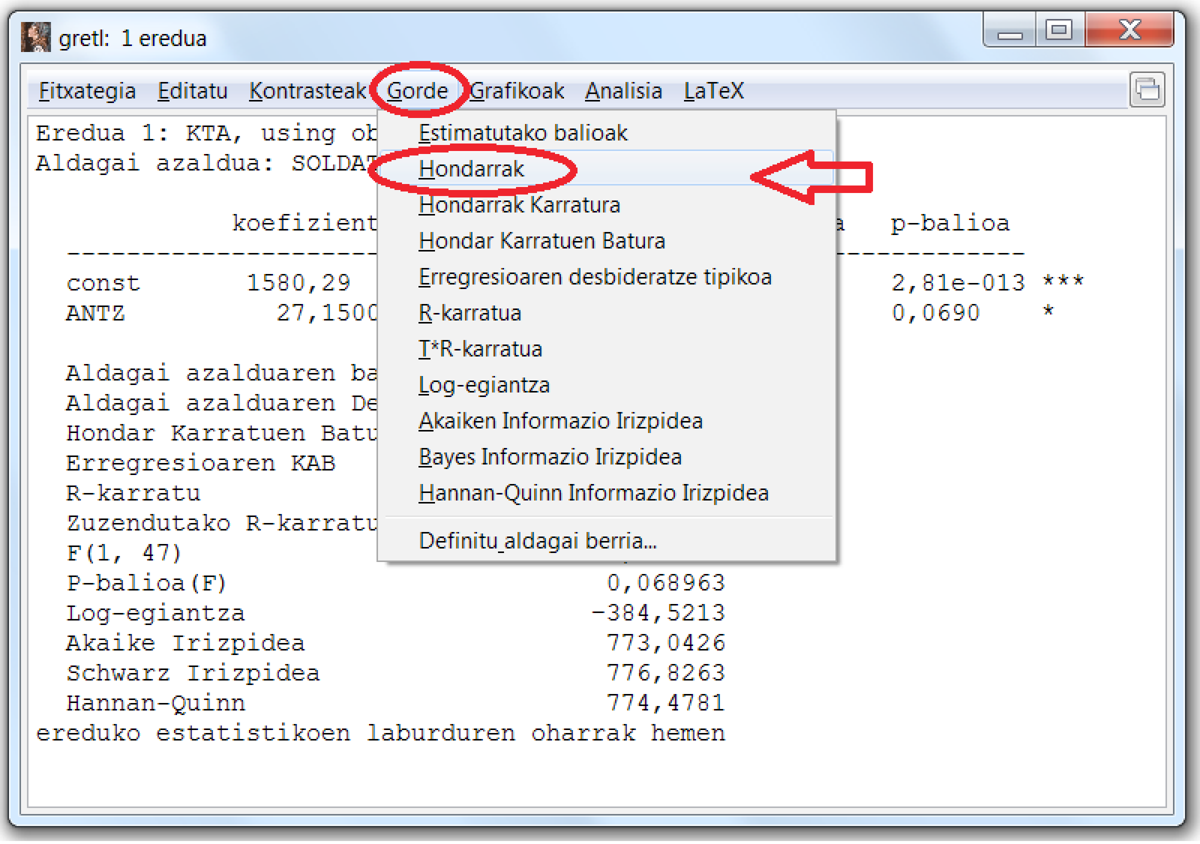
Hirugarren zutabean, berriz, ereduko koefizienteen KTA zenbatesleen desbideratze zenbatetsiak ageri dira; hau da, bariantza zenbatetsien erro karratuak:

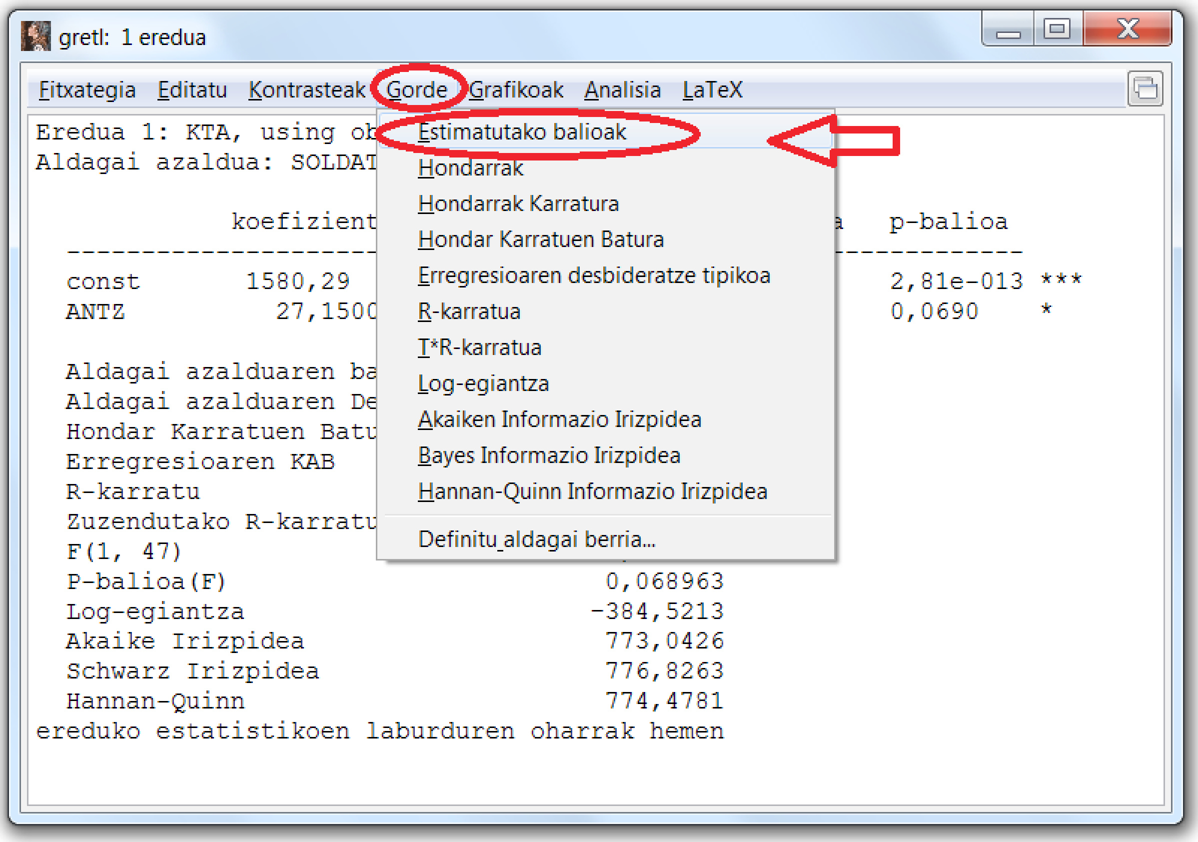
$$\widehat{des}\ (\hat{\alpha}) = \sqrt{\hat{\sigma}^2 \frac{\sum_{i=1}^{N} X_i^2}{N\ \sum_{i=1}^{N}(X_i - \bar{X})^2}} = 157{,}393$$

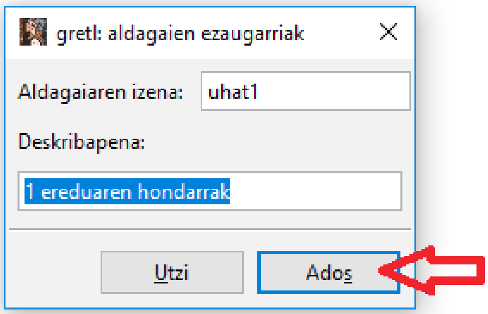
$$\widehat{des}(\hat{\beta}) = \sqrt{\hat{\sigma}^2 \frac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2}} = 14{,}58$$

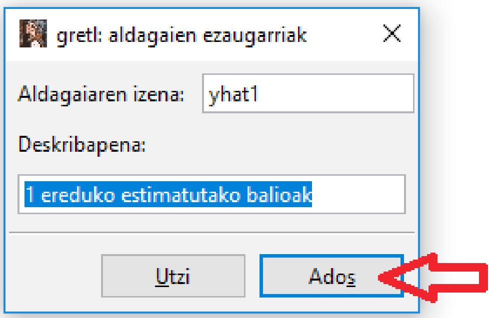
Zenbatespen-emaitzetan ikusten denez, erregresioari dagokion mugatze- koefizientea $R^2 = 0{,}068$ da; hau da, *antzinatasuna* aldagai azaltzaileak *langileen soldata* aldagai azalduaren aldakuntzaren % 6,8 azaltzen du. Aldi berean, askotan komenigarria izaten da erregresio honi dagokion Hondar Karratuen Batura gordetzea; kasu honetan, *HKB* = 18787279.

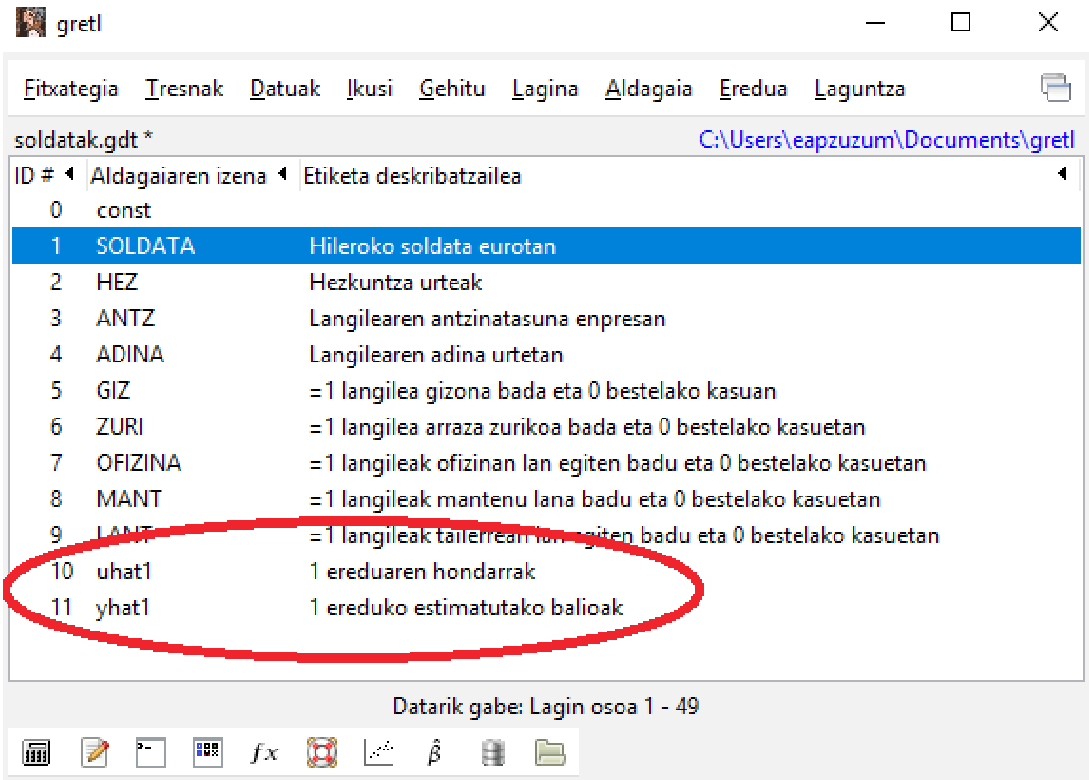
Beti izango da zenbatetsitako balioak edota hondarrak gordetzeko, ikusteko edota grafikoak egiteko aukera, askotan komenigarria izaten baita eskura izatea. Gretl-ek automatikoki *uhat* eta *yhat* izendatzen ditu, eta gainontzeko aldagaien zerrendara gehitzen ditu; gogoratu beti alda ditzakezula izenak eta ezaugarriak:



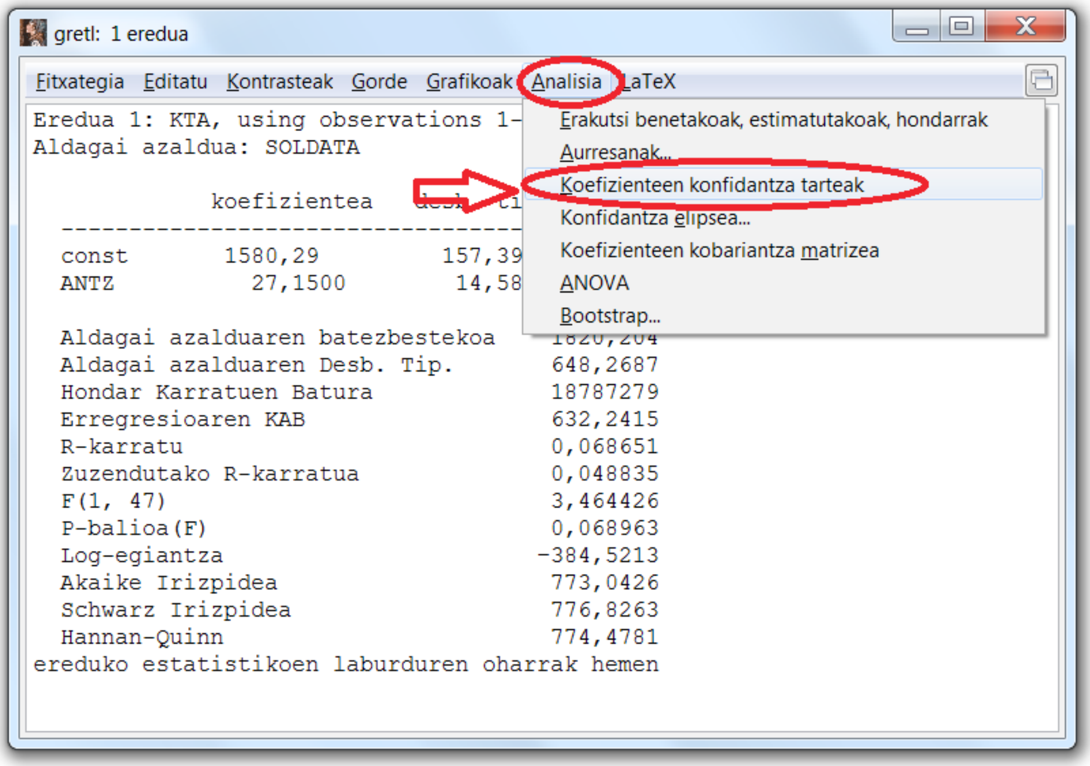

gretl: 1 eredua
Fitxategia Editatu Kontrasteak Gorde Grafikoak Analisia LaTeX
Eredua 1: KTA, using ob
Aldagai azaldua: SOLDAT
koefizient
p-balioa
const 1580,29
2,81e-013 ***
ANTZ 27,1500
0,0690 *
Aldagai azalduaren ba
Aldagai azalduaren De
Hondar Karratuen Batu
Erregresioaren KAB
R-karratu
Zuzendutako R-karratu
F(1, 47)
P-balioa(F) 0,068963
Log-egiantza -384,5213
Akaike Irizpidea 773,0426
Schwarz Irizpidea 776,8263
Hannan-Quinn 774,4781
ereduko estatistikoen laburduren oharrak hemen
Estimatutako balioak
Hondarrak
Hondarrak Karratura
Hondar Karratuen Batura
Erregresioaren desbideratze tipikoa
R-karratua
T*R-karratua
Log-egiantza
Akaiken Informazio Irizpidea
Bayes Informazio Irizpidea
Hannan-Quinn Informazio Irizpidea
Definitu aldagai berria...

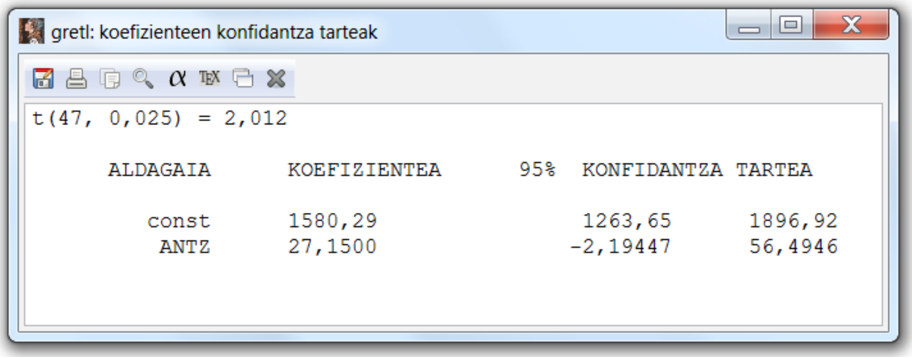

gretl: aldagaien ezaugarriak
Aldagaiaren izena: uhat1
Deskribapena:
1 ereduaren hondarrak
Utzi
Ados


gretl: aldagaien ezaugarriak
Aldagaiaren izena: yhat1
Deskribapena:
1 ereduko estimatutako balioak
Utzi
Ados



Aldi berean, ereduko koefizienteen KTA zenbatesleen tartezko zenbatespena ere lor daiteke zenbatespen-emaitzetako leihatilan, *Analisia* → *Koefizienteen konfiantza tarteak* klikatuz:

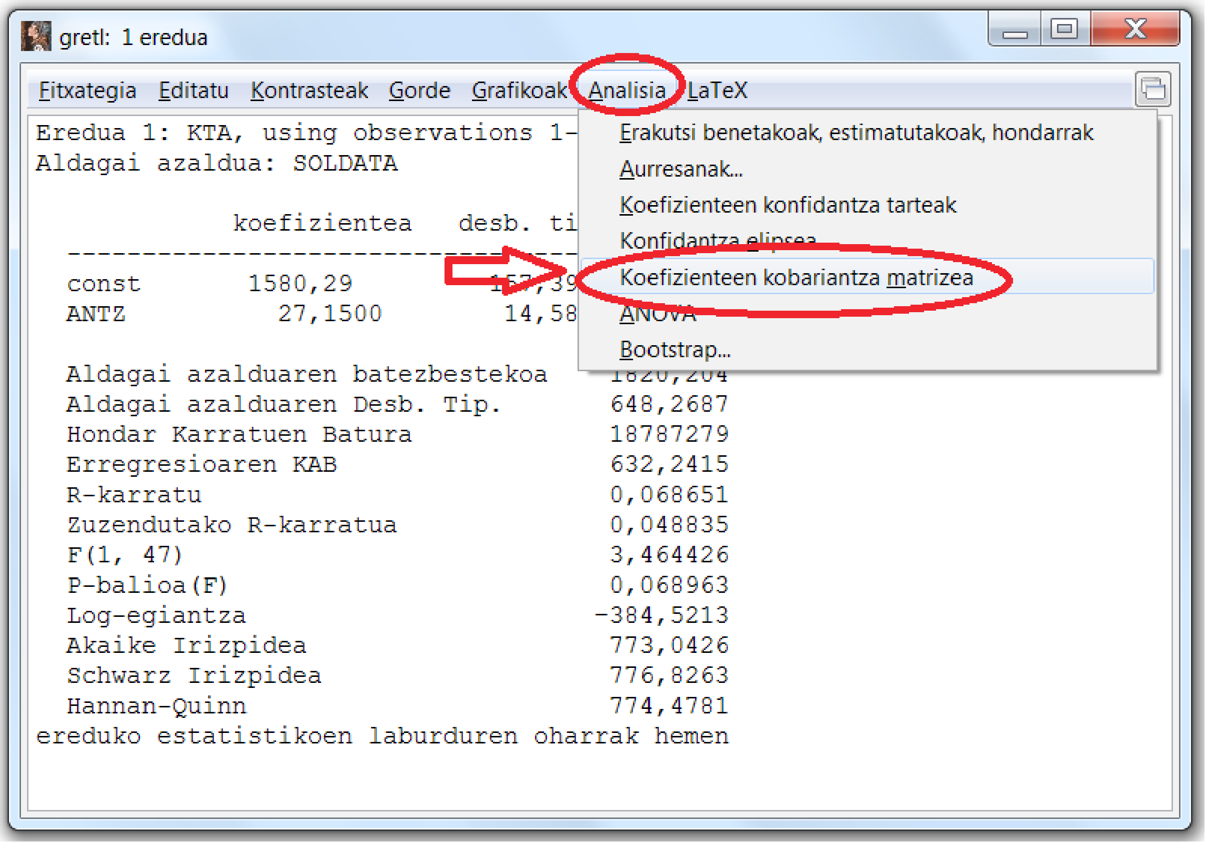


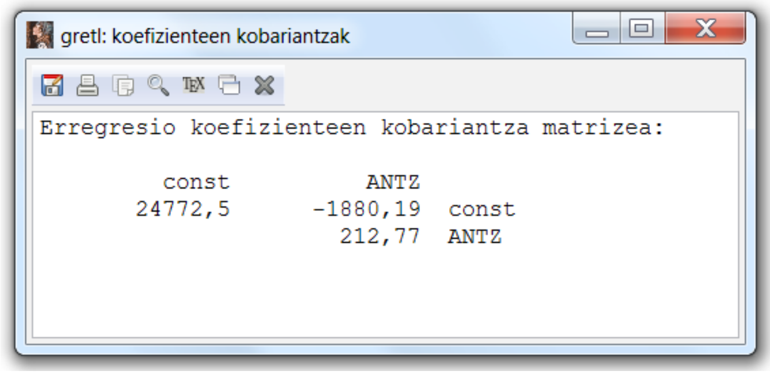


Azkenik, nahiz eta zenbatespen-emaitzako leihatilan ereduko koefizienteen desbideratze zenbatetsiak izan, zenbait kontraste gauzatzeko, komenigarri izaten da ereduko koefizienteen bariantza-kobariantza matrizearen zenbatespena izatea; hau da, honako matrizea honen emaitzak:

$$\widehat{Bar}(\hat{\beta}) = \hat{\sigma}^2 \begin{pmatrix} \dfrac{\sum_{i=1}^{N} X_i^2}{N\ \sum_{i=1}^{N}(X_i - \bar{X})^2} & \dfrac{-\bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \\ \dfrac{-\bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} & \dfrac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \end{pmatrix} = \begin{pmatrix} 24772{,}5 & -1880{,}19 \\ -1880{,}19 & 212{,}77 \end{pmatrix}$$

Analisia → *Koefizienteen kobariantza-matrizea* klikatuz,




gretl: koefizienteen kobariantzak
Erregresio koefizienteen kobariantza matrizea:
const ANTZ
24772,5 -1880,19 const
212,77 ANTZ

4. gaia

Erregresio Lineal Bakunaren Eredua. Murrizketa linealen kontrasteak

4.1. **Sarrera**

Ekonometriaren funtsezko zeregin bat da ekonomia erreal baten ezaguera kuantitatibo bat ekartzea. Ekonomialariek portaera ekonomikoari buruzko teoriak garatu eta ebaluatzen dituzte, eta hipotesien kontrasteak teoria horiek ebaluatzeko prozedurak dira.

4.2. **Maldaren kontrasteak**

Atal honetan, ereduko maldaren zenbait kontraste nola gauzatzen diren azalduko da.

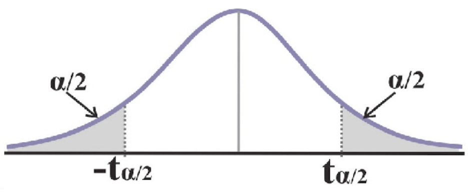
Bi aldetako kontrasteak

Adibidez, ereduko malda, β, «c» konstante bat den ala ez kontrastatzeko, honako estatistiko hau erabiliko da:

$$t = \frac{\hat{\beta} - \beta}{\widehat{des}\,(\hat{\beta})} \sim t_{(N-2)}$$

Eta honako hipotesi huts eta aurkako hau izanik:

$$\begin{cases} H_0: \beta = c \\ Ha: \beta \neq c \end{cases} \qquad t = \frac{\hat{\beta} - c}{\widehat{des}\,(\hat{\beta})} \overset{H_0}{\sim} t_{(N-2)}$$



Hipotesi hutsa egiazkoa bada, t estatistikoak zerotik oso urruneko balioa izatea gertaezina da. Hortaz, H_0 baztertuko da α esangura-mailarekin, baldin eta $|t| > t_{(N-2)\alpha/2}$ bada.

Ereduko koefizienteen konfiantza-tartearekin ere gauza daitezke horrelako kontrasteak; nahikoa izango da «c» balioa konfiantza-tartean dagoen ala ez ikustea. Hau da, «c» balioa tartean baldin badago, hipotesi hutsa ez da baztertzen, eta, alderantziz, tartetik kanpo baldin badago, hipotesi hutsa baztertzen da α esangura-mailarekin.

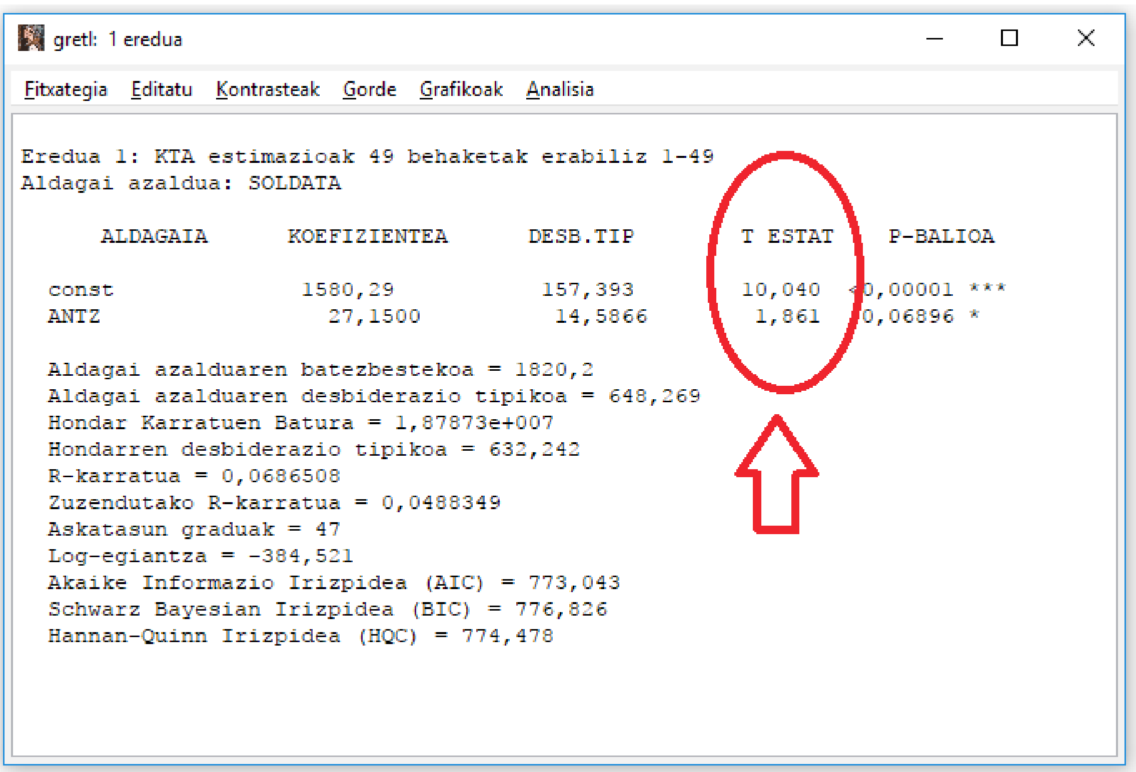
$$\mathrm{KT}(\beta)_{1-\alpha} = \left[\hat{\beta} - t_{(N-2)\alpha/2}\, \widehat{des}\,(\hat{\beta})\,;\; \hat{\beta} + t_{(N-2)\alpha/2}\, \widehat{des}\,(\hat{\beta})\right]$$

Hipotesi hutsean jarri den «c» terminoaren balioa zero denean (H_0: $\beta = 0$), ereduko aldagai azaltzailea nabaria edo esanguratsua den kontrastatzen ari gara; hau da, X_i aldagaiaren **banakako esangura-kontrastea** egiten.

$$\begin{cases} \mathrm{H_0}: \beta = 0 \\ \mathrm{Ha}: \beta \neq 0 \end{cases} \qquad t = \frac{\hat{\beta}}{\widehat{des}\,(\hat{\beta})} \overset{H_0}{\sim} t_{(N-2)}$$

Kasu honetan, eta $|t| > t_{(N-2)\alpha/2}$ bada, hipotesi hutsa baztertzen da α esangura- mailarekin, eta X_i aldagaia nabari dela esango da. Bestela, hau da, eta $|t| < t_{(N-2)\alpha/2}$ bada, hipotesi hutsa ez da baztertzen, eta X_i aldagaia ez-nabaria dela esango da.

t estatistiko honi *t-arrazoia* edo *lagineko t* esaten zaio, eta Gretl-eko zenbatespen-leihatilan laugarren zutabean ageri diren balioak dira. Hau da,



gretl: 1 eredua

Fitxategia Editatu Kontrasteak Gorde Grafikoak Analisia

```
Eredua 1: KTA estimazioak 49 behaketak erabiliz 1-49
Aldagai azaldua: SOLDATA

      ALDAGAIA       KOEFIZIENTEA       DESB.TIP         T ESTAT      P-BALIOA

  const               1580,29           157,393          10,040     <0,00001 ***
  ANTZ                  27,1500          14,5866          1,861      0,06896 *

  Aldagai azalduaren batezbestekoa = 1820,2
  Aldagai azalduaren desbiderazio tipikoa = 648,269
  Hondar Karratuen Batura = 1,87873e+007
  Hondarren desbiderazio tipikoa = 632,242
  R-karratua = 0,0686508
  Zuzendutako R-karratua = 0,0488349
  Askatasun graduak = 47
  Log-egiantza = -384,521
  Akaike Informazio Irizpidea (AIC) = 773,043
  Schwarz Bayesian Irizpidea (BIC) = 776,826
  Hannan-Quinn Irizpidea (HQC) = 774,478
```

Bestalde, hurrengo zutabean ageri diren *p-balioa*k ere, lagungarri dira esangura- kontrasteak egiteko, hipotesi hutsa zein esangura-mailatik aurrera baztertzen den adierazten baitute, hau da, ze esangura-mailatik aurrera den nabaria aldagaia. Adibidean, antzinatasuna % 6,89ko esangura-mailatik aurrera da nabaria, eta, beraz, % 5arekin ez-esanguratsua da; % 10arekin, ordea, esanguratsua.

p-balio horien ondoan posible da izar batzuk azaltzea ala ez. Horiek ere esangura-kontrasteei buruzko informazioa ematen dute:

— Izar bakarra dagoenean (*): % 10eko esangura-mailarekin nabaria da.
— Bi izar daudenean (**): % 5eko esangura-mailarekin nabaria da.
— Hiru izar daudenean (***): % 1eko esangura-mailarekin nabaria da.

OHARRA: *p* balioak eta izarrak, banakako esangura-kontrastea egiteko bakarrik erabil daitezke.

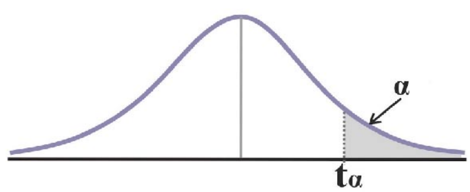
Alde bateko kontrasteak

Adibidez, ereduko β malda «c» konstantea baino handiagoa den kontrastatu daiteke. Bi aldetako kontrastea egiteko erabili den estatistiko berdinarekin gauzatzen da, baina erabaki-irizpidea desberdina da, eskualde kritiko alde batera egongo baita.

Alde bateko kontrasteak egiteko, kontu handia izan behar da hipotesi hutsa eta aurkakoa osatzeko orduan, hipotesi hutsean ageri behar baitu berdintasunak. Kasu honetan, «c» konstantea baino handiagoa dela kontrastatu nahi denez, aurkako hipotesian jarri beharko da, jarraian adierazten den bezala:

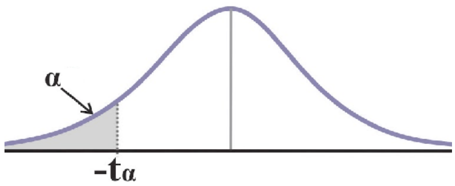
$$\begin{cases} H_0: \beta \le c \\ H_a: \beta > c \end{cases} \qquad t = \frac{\hat{\beta} - c}{\widehat{des}\,(\hat{\beta})} \overset{H_0}{\sim} t_{(N-2)}$$

Kasu honetan, aurkako hipotesian oinarrituko gara eskualde kritikoa osatzeko, eskuinerantz osatuz eta hipotesi hutsa baztertuz $t > t_{(N-2)\alpha}$ denean.



Bestalde, ereduko β malda «c» konstantea baino txikiagoa den kontrastatu nahi izanez gero, aurkako hipotesian jarri beharko da aukera hori. Lehen bezala, estatistiko berdinarekin gauzatzen da kontrastea, baina erabaki-irizpidea desberdina izango da; kasu honetan, eskualde kritikoa alde batera egongo da, baina ezkerrerantz, hipotesi hutsa baztertuz $t < -t_{(N-2)\alpha}$ denean.

$$\begin{cases} H_0: \beta \geq c \\ Ha: \beta < c \end{cases} \qquad t = \frac{\hat{\beta} - c}{\widehat{des}\,(\hat{\beta})} \overset{H_0}{\sim} t_{(N-2)}$$



5. gaia

Erregresio Lineal Orokorraren Eredua. Zehaztapena

5.1. **Sarrera**

Gai honetan Erregresio Lineal Orokorraren Eredua (ELOE) aztertuko da, hau da, aldagai azaltzaile bat baino gehiago dituzten ereduak; kuantitatiboak edota kualitatiboak izan daitezke. Horrela, ereduan konstante bat izateaz gain, beste aldagai azaltzaile bat baino gehiago izango dira; hau da, $K > 2$ izango da:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + u_i \qquad i = 1, 2, \ldots, N$$

non

— Y aldagai endogenoa edo azaldua da.
— X_k, $k = 1, 2, \ldots, K$ aldagai azaltzaileak edo exogenoak dira.
— β_k, $k = 1, 2, \ldots, K$ koefiziente edo parametroak dira (ezezagunak).
— u_i perturbazio aleatorioa da, ez da behagarria, eta izaera aleatorioa du. Ereduan barneratu gabeko aldagaien eraginak, agente ekonomikoen jokabide aleatorioak edota neurketa-erroreak jaso ditzake.

Lagineko banako bakoitzarentzat (i bakoitzarentzat) eredua berridatziz, honako ekuazio hauek lortzen dira:

$$\begin{cases} Y_1 = \beta_1 + \beta_2 X_{21} + \cdots + \beta_K X_{K1} + u_1 & i = 1 \\ Y_2 = \beta_1 + \beta_2 X_{22} + \cdots + \beta_K X_{K2} + u_2 & i = 2 \\ Y_2 = \beta_1 + \beta_2 X_{23} + \cdots + \beta_K X_{K3} + u_3 & i = 3 \\ \\ Y_N = \beta_1 + \beta_2 X_{2N} + \cdots + \beta_K X_{KN} + u_N & i = N \end{cases}$$

Matrizialki idatzita:

$$Y_{(Nx1)} = X_{(NxK)}\, \beta_{(Kx1)} + U_{(Nx1)}$$

$$Y_{(Nx1)} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_i \\ . \\ . \\ Y_N \end{pmatrix} \qquad X_{(NxK)} = \begin{pmatrix} 1 & X_{21} & X_{31} & \cdots & X_{K1} \\ 1 & X_{22} & X_{32} & \cdots & X_{K2} \\ & & & & \\ 1 & X_{2i} & X_{3i} & \cdots & X_{Ki} \\ & & & & \\ 1 & X_{2N} & X_{3N} & \cdots & X_{KN} \end{pmatrix}$$

$$\beta_{(Kx1)} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_K \end{pmatrix} \qquad U_{(Nx1)} = \begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_N \end{pmatrix}$$

5.2. **Oinarrizko hipotesiak**

Eredua zehaztatu ondoren, zenbatespenaren testuingurua finkatu behar da. Horretarako, gogora ditzagun oinarrizko hipotesiak.

— Forma funtzionalari dagokionez: eredua koefizienteekiko lineala da; kurtsoan zehar zenbatetsiko diren ereduak koefizienteekiko linealak edo linealizagarriak dira.

— Koefizienteei dagokienez: koefizienteak konstante mantentzen dira laginean zehar.

— Aldagai endogenoari dagokionez: aldagai endogenoa kuantitatiboa da. Oinarrizko kurtso honetan zehar, aldagai azaldua kuantitatiboa izango da, zeren, aldagai dependente kualitatiboa daukan eredu bat zenbatesteko, irakasgai honetan ikasiko diren zenbatespen-metodoek ez baitute balio.

— Aldagai azaltzaileei dagokienez:

1. X_k aldagai azaltzaileen lagin-bariantzak (Sx_k^2) ezin dira zero izan, eta, gainera, behaketa kopurua zenbatetsi behar den koefiziente kopurua baino handiagoa izan behar da; hau da, eredu orokorrean, $N > K$ izan behar da. Hipotesi hau beharrezkoa da koefizienteak identifikatzeko. Hasteko, zenbatetsi behar den koefiziente kopurua behaketen kopurua baino handiago izanez gero, orduan zenbatespena aurrera eramateko ez daukagu informazio nahikorik. Bestalde, aldagai azaltzaile baten lagineko bariantza zero izango balitz, orduan erregresio funtzioaren maldak identifikatzea ezinezkoa izango litzateke.
2. X datu-matrizea hein osokoa izango da zutabeetan, eta, beraz, eredua zenbatesgarria izango da ($h(X) = K$).

— Ereduari dagokionez: eredua ondo zehaztatuta dago, hau da, ereduak ezin du aldagai nabaririk barneratu gabe utzi, eta ezta kontrakorik ere, aldagai ez-nabariren bat kontuan hartu.

— Perturbazioari dagokionez:

1. Perturbazioen populazio-batez bestekoa edo itxaropena zero da; zeron zentratutako aldagai aleatorioa da, hau da, $E_X(u_i) = 0$, $i = 1, 2, \ldots, N$. Iragarri ezina den errorearen batez bestekoa zero izateak ereduaren alde sistematikoa edo analizatu nahi den batez besteko portaera $E_X(Y_i) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki}$ izatea dakar.
2. Perturbazioen populazio-bariantza konstantea da (homozedastizitatea). Aldagai aleatorioaren edo perturbazioaren aldakortasuna laginean zehar konstante mantentzen dela pentsatuko da. Hau da, $Bar(u_i) = E_X(u_i - E(u_i))^2 = E_X(u_i)^2 = \sigma_u^2$ konstantea da $i = 1, 2, \ldots, N$.
   Horrela, aldagai azaltzaileen balioak emanik, aldagai azalduak har ditzakeen balio posibleen tartearen zabalera berdina da, eta balio bakoitzak irteteko duen probabilitatea independentea da X aldagaiek hartzen dituzten balioekiko.
3. Perturbazioen artean ez dago autokorrelaziorik. Perturbazio desberdinen arteko korrelazioa zero dela pentsatuko da ($Kor(u_i, u_j) = r_{ui,uj} = 0$; $i \neq j$). Hortaz, beren arteko kobariantza ($Kob((u_i, u_j) = 0$; $i \neq j$) ere zero izango da.

$$Kob(u_i, u_j) = E_X[(u_i - E_X(u_i))(u_j - E_X(u_j))] = E_X(u_i\, u_j) = 0 \qquad i \neq j$$

4. Perturbazioek Banaketa Normala jarraitzen dute. Koefizienteak zenbatesteko beharrezkoa ez izan arren, normaltasuna behar da inferentzia egiteko, eta baita koefizienteen konfiantza-tarteak lortzeko ere.
   Perturbazioen propietate hauek batera idatz daitezke:

$$u_i \sim NIB(0, \sigma^2) \qquad i = 1, 2, \ldots, N$$

Edota matrizeak erabiliz:

$$U_{(Nx1)} \sim N(0_{(Nx1)}, \sigma^2\, I_{(NxN)})$$

Non:

$$E_X(U)_{(Nx1)} = \begin{pmatrix} E_X(u_1) \\ E_X(u_2) \\ \vdots \\ E_X(u_N) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0_N$$

Bar(U)= $E_X\big(U - E_X(U)\big)\big(U - E_X(U)\big)' = E_X(U\,U')_{(NxN)} =$

$$= \begin{pmatrix} E_X(u_1)^2 & E_X(u_1\,u_2) & E_X(u_1\,u_3) & \dots & E_X(u_1\,u_N) \\ E_X(u_2\,u_1) & E_X(u_2)^2 & E_X(u_2\,u_3) & \dots & E_X(u_2\,u_N) \\ E_X(u_3\,u_1) & E_X(u_3\,u_2) & E_X(u_3)^2 & \dots & E_X(u_3\,u_N) \\ \vdots & \vdots & \vdots & & \vdots \\ E_X(u_N\,u_1) & E_X(u_N\,u_2) & E_X(u_N\,u_3) & \dots & E_X(u_N)^2 \end{pmatrix} =$$

$$= \begin{pmatrix} \sigma_u^2 & 0 & 0 & & \dots & 0 \\ 0 & \sigma_u^2 & 0 & & & 0 \\ 0 & & \sigma_u^2 & & & \vdots \\ \vdots & & & \ddots & & \\ & & & & \sigma_u^2 & 0 \\ 0 & 0 & \dots 0 & & 0 & \sigma_u^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & & \dots & 0 \\ 0 & 1 & 0 & & & 0 \\ 0 & & 1 & & & \vdots \\ \vdots & & & \ddots & & \\ & & & & 1 & 0 \\ 0 & 0 & \dots 0 & & 0 & 1 \end{pmatrix} = \sigma^2\, I_N$$

5.3. **Koefizienteen interpretazioa**

Erregresio Lineal Orokorraren Ereduan aldagai azaltzaileak asko izan daitezke, eta ereduko koefizienteen interpretazioa eredu bakunean baino «zailagoa» da. Atal honetan, soldataren adibidea landuz, Erregresio Lineal Orokorraren Eredu desberdinak zehaztatuko dira, eta, kasu bakoitzean, koefizienteen interpretazioak aztertuko dira. Eredu guztietan, langilearen soldata azaldu nahi da, langilearen hezkuntza-urteak, antzinatasuna, generoa eta lanpostuaren arabera; hau da, guztietan aldagai azaldu eta azaltzaile berberak ageri dira. Proposatuko diren eredu horiek linealtasunaren oinarrizko hipotesia betetzen dute, guztiak koefizienteekiko linealak baitira; baina zenbait kasutan ez dira aldagaiekiko linealak izango, aldagaien berbidura edota aldagaien arteko iterazioa barneratuko baitugu.

A) *ADIBIDEA → ELOE-1 eredua*

Soldata = *f* (Hezkuntza-urteak, Antzinatasuna, Generoa, Lanpostua)

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + u_i \qquad i = 1, 2, ..., 49$$

***E*(*SOLDATA*)**	**Emakumea**	**Gizona**
BULEGOA	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i$	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4$
MANTENU	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_5$	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4 + \beta_5$
LANTEGIA	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_6$	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4 + \beta_6$

β_1: Hezkuntza-urterik ez duen ($HEZ = 0$) eta enpresako bulegoan sartu berria den ($ANTZ = 0$) emakume baten hileroko batez besteko soldata da (eurotan).

β_2: Hezkuntza-urte gehigarri bakoitzeko, hileroko batez besteko soldataren gehikuntza (eurotan) da, antzinatasuna konstante mantenduz. Gehikuntza hori berdina da langile guztientzat (emakume nahiz gizon, eta edozein lanpostu izanik).

β_3: Antzinatasun-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza (eurotan) da, hezkuntza-urteak konstante mantenduz. Gehikuntza hori berdina da langile guztientzat (emakume nahiz gizon, eta edozein lanpostu izanik).

β_4: Gizon baten hileroko batez besteko soldata (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun, hezkuntza-urte eta lanpostu berdineko— emakume batenarekin alderatuta.

β_5: Mantenu-lanetan ari den langile baten hileroko batez besteko soldata (eurotan) ezaugarri berdinak dituen —hau da, hezkuntza-urte, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen langilearenarekin alderatuta.

β_6: Lantegian edo tailerrean lan egiten duen langile baten hileroko batez besteko soldata (eurotan), ezaugarri berdinak dituen —hau da, hezkuntza-urte, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen langilearenarekin alderatuta.

B) *ADIBIDEA* → *ELOE-2 eredua*

Soldata = f (Hezkuntza-urteak, Antzinatasuna, Generoa, Lanpostua)

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i\, HEZ_i + \beta_6\, LANT_i\, HEZ_i + u_i$$

$$i = 1, 2, ..., 49$$

***E*(*SOLDATA*)**	**Emakumea**	**Gizona**
BULEGOA	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i$	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4$
MANTENU	$\beta_1 + (\beta_2 + \beta_5)\, HEZ_i + \beta_3\, ANTZ_i$	$\beta_1 + (\beta_2 + \beta_5)\, HEZ_i + \beta_3\, ANTZ_i + \beta_4$
LANTEGIA	$\beta_1 + (\beta_2 + \beta_6)\, HEZ_i + \beta_3\, ANTZ_i$	$\beta_1 + (\beta_2 + \beta_6)\, HEZ_i + \beta_3\, ANTZ_i + \beta_4$

β_1: Enpresan sartu berria ($ANTZ = 0$) eta hezkuntza-urterik ez duen ($HEZ = 0$) emakume baten hileroko batez besteko soldata da (eurotan).

β_2: Bulegoan lan egiten duen langile baten hezkuntza-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza da (eurotan), antzinatasuna konstante mantenduz. Gehikuntza hori berdina da langile guztientzat, emakume nahiz gizon.

β_3: Antzinatasun-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza (eurotan) da, hezkuntza-urteak konstante mantenduz. Gehikuntza hori berdina da langile guztientzat (emakume nahiz gizon, eta edozein lanpostu izanik).

β_4: Gizon baten hileroko batez besteko soldata (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun, hezkuntza-urte eta lanpostu berdineko— emakume batenarekin alderatuta.

β_5: Mantenu-lanetan ari den langile baten hezkuntza-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza (eurotan), bulegoan lan egiten duen batenarekin alderatuta, antzinatasuna konstante mantenduz. Diferentzia hori berdina da emakume nahiz gizonezkoentzat.

β_6: Lantegi edo tailerrean ari den langile baten hezkuntza-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza (eurotan), bulegoan lan egiten duen batenarekin alderatuta, antzinatasuna konstante mantenduz. Diferentzia hori berdina da emakume nahiz gizonezkoentzat.

C) *ADIBIDEA → ELOE-3 eredua*

Soldata = *f* (Hezkuntza-urteak, Antzinatasuna, Generoa, Lanpostua)

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + \beta_7\, ANTZ_i\, GIZ_i + u_i$$

$i = 1, 2, ..., 49$

***E*(*SOLDATA*)**	**Emakumea**	**Gizona**
BULEGOA	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i$	$\beta_1 + \beta_2\, HEZ_i + (\beta_3 + \beta_7)\, ANTZ_i + \beta_4$
MANTENU	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_5$	$\beta_1 + \beta_2\, HEZ_i + (\beta_3 + \beta_7)\, ANTZ_i + \beta_4 + \beta_5$
LANTEGIA	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_6$	$\beta_1 + \beta_2\, HEZ_i + (\beta_3 + \beta_7) ANTZ_i + \beta_4 + \beta_6$

β_1: Enpresako bulegoan sartu berria (*ANTZ* = 0) eta hezkuntza-urterik ez duen (*HEZ* = 0) emakume baten hileroko batez besteko soldata da (eurotan).

β_2: Hezkuntza-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza da (eurotan), antzinatasuna konstante mantenduz. Gehikuntza hori berdina da langile guztientzat, emakume nahiz gizon eta edozein lanpostutan.

β_3: Antzinatasun-urte gehigarri bakoitzeko edozein lanpostu duen emakume baten hileroko batez besteko soldataren gehikuntza da (eurotan), hezkuntza-urteak konstante mantenduz.

β_4: Gizon baten hileroko batez besteko soldata da (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun, hezkuntza-urte eta lanpostu berdineko— emakume batenarekin alderatuta.

β_5: Mantenu-lanetan ari den langile baten hileroko batez besteko soldata (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen batenarekin alderatuta.

β_6: Lantegi edo tailerrean ari den langile baten hileroko batez besteko soldata da (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen batenarekin alderatuta.

β_7: Hezkuntza-urteak konstante mantenduta, antzinatasun-urte gehigarri bakoitzeko gizon baten hileroko batez besteko soldataren gehikuntza da (eurotan), emakume batek izango duenarekin alderatuta, edozein lanpostu izanik.

D) *ADIBIDEA* → *ELOE-4 eredua*

Soldata = f (Hezkuntza-urteak, Antzinatasuna, Generoa, Lanpostua)

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + \beta_7\, ANTZ_i^2 + u_i$$

$i = 1, 2, ..., 49$

***E*(*SOLDATA*)**	**Emakumea**	**Gizona**
BULEGOA	$\beta_1 + \beta_2 HEZ_i + \beta_3 ANTZ_i + \beta_7 ANTZ_i^2$	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4 + \beta_7 ANTZ_i^2$
MANTENU	$\beta_1 + \beta_2 HEZ_i + \beta_3 ANTZ_i + \beta_5 + \beta_7 ANTZ_i^2$	$\beta_1 + \beta_2\, HEZ_i + \beta_3 ANTZ_i + \beta_4 + \beta_5 + \beta_7 ANTZ_i^2$
LANTEGIA	$\beta_1 + \beta_2 HEZ_i + \beta_3 ANTZ_i + \beta_6 + \beta_7 ANTZ_i^2$	$\beta_1 + \beta_2 HEZ_i + \beta_3 ANTZ_i + \beta_4 + \beta_6 + \beta_7 ANTZ_i^2$

β_1: Enpresako bulegoan sartu berria ($ANTZ = 0$) eta hezkuntza-urterik ez duen ($HEZ = 0$) emakume baten hileroko batez besteko soldata da (eurotan).

β_2: Hezkuntza-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza da (eurotan), antzinatasuna konstante mantenduz. Gehikuntza hori berdina da langile guztientzat (emakume nahiz gizon, eta edozein lanpostutan).

$\beta_3 + 2\ \beta_7\ ANTZ_i$: Antzinatasun-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza da (eurotan), hezkuntza-urteak konstante mantenduz. Gehikuntza hori berdina da langile guztientzat (emakume nahiz gizonezko, eta edozein lanpostu izanik), baina langilearen antzinatasunaren araberakoa da. Kasu honetan, antzinatasunaren efektu marjinala soldataren gain ez da konstantea, langile bakoitzaren antzinatasunaren araberakoa da.

β_4: Gizon baten hileroko batez besteko soldata da (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun, hezkuntza-urte eta lanpostu berdineko— emakume batenarekin alderatuta.

β_5: Mantenu-lanetan ari den langile baten hileroko batez besteko soldata da (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen batenarekin alderatuta.

β_6: Lantegi edo tailerrean dagoen langile baten hileroko batez besteko soldata da (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen batenarekin alderatuta.

E) *ADIBIDEA → ELOE-5 eredua*

Soldata = f (Hezkuntza-urteak, Antzinatasuna, Generoa, Lanpostua)

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + \beta_7\, ANTZ_i\, MANT_i + \\ + \beta_8\, ANTZ_i\, LANT_i + u_i$$

$$i = 1, 2, ..., 49$$

E(*SOLDATA*)	Emakumea	Gizona
BULEGOA	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i$	$\beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4$
MANTENU	$\beta_1 + \beta_2\, HEZ_i + (\beta_3 + \beta_7)\, ANTZ_i + \beta_5$	$\beta_1 + \beta_2\, HEZ_i + (\beta_3 + \beta_7)\, ANTZ_i + \beta_4 + \beta_5$
LANTEGIA	$\beta_1 + \beta_2\, HEZ_i + (\beta_3 + \beta_8)\, ANTZ_i + \beta_6$	$\beta_1 + \beta_2\, HEZ_i + (\beta_3 + \beta_8)\, ANTZ_i + \beta_4 + \beta_6$

β_1: Enpresako bulegoan sartu berria (*ANTZ* = 0) eta hezkuntza-urterik ez duen (*HEZ* = 0) emakume baten hileroko batez besteko soldata da (eurotan).

β_2: Hezkuntza-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza da (eurotan), antzinatasuna konstante mantenduz. Gehikuntza hori berdina da langile guztientzat (emakume nahiz gizon, eta edozein lanpostutan).

β_3: Bulegoan lan egiten duen langile baten antzinatasun-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza da (eurotan), hezkuntza-urteak konstante mantenduz. Gehikuntza hori berdina da langile guztientzat, emakume nahiz gizonezko.

β_4: Gizon baten hileroko batez besteko soldata da (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun, hezkuntza-urte eta lanpostu berdineko— emakume batenarekin alderatuta.

β_5: Mantenu-lanetan ari den langile baten hileroko batez besteko soldata da (eurotan), ezaugarri berdinak dituen —hau da antzinatasun eta genero berdineko— baina bulegoan lan egiten duen batenarekin alderatuta.

β_6: Lantegi edo tailerrean ari den langile baten hileroko batez besteko soldata da (eurotan), ezaugarri berdinak dituen —hau da hezkuntza urte, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen batenarekin alderatuta.

β_7: Mantenu-lanetan ari den langile baten antzinatasun-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza da (eurotan), bulegoan lan egiten duen batenarekin alderatuta, hezkuntza-urteak konstante mantenduz. Diferentzia hori berdina da langile guztientzat, emakume nahiz gizonezko.

β_8: Lantegi edo tailerrean lan egiten duen langile baten antzinatasun-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza da (eurotan), bulegoan lan egiten duen batenarekin alderatuta, hezkuntza-urteak konstante mantenduz. Diferentzia hori berdina da langile guztientzat, emakume nahiz gizonezko.

6. gaia

Erregresio Lineal Orokorraren Eredua. Zenbatespena

6.1. **Karratu Txikien Arrunten (KTA) zenbateslea**

Demagun honako eredu ekonometriko orokor hau, non oinarrizko hipotesi guztiak betetzen baitira.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + u_i \qquad i = 1, 2, \dots, N$$

Aurreko gaietan aipatu bezala, ereduko koefizienteak ezezagunak direnez, aldagai endogenoaren (*Y*) batez besteko portaera ere ezezaguna da; hau da, ereduko alde sistematikoa edo Populazioko Erregresio Funtzioa (PEF) ezezaguna da.

$$\text{PEF} \longrightarrow E_X(Y_i) = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} \qquad i = 1, 2, \dots, N$$

Hortaz, ereduko perturbazioa da, aldagai azalduaren benetako balioaren eta Populazioko Erregresio Funtzioaren arteko diferentzia eta ereduko alde sistematikoak azaldu ezin duena jasotzen du:

$$u_i = Y_i - E_X(Y_i) = Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_K X_{Ki} \qquad i = 1, 2, \dots, N$$

Ekonometriaren helburua eredu ekonometrikoa ahalik eta hobekien zenbatestea denez, ahalik eta hondar ($\hat{u}_i$) txikienak lortzea du xedetzat, hau da, aldagai azalduaren benetako balioen eta zenbatetsitako balioen arteko diferentziak ahalik eta txikienak izatea. Hortaz, Karratu Txikien Arrunten (KTA) bitartez eredua zenbatesterakoan minimizatuko den helburu-funtzioa Hondar Karratuen Batura (HKB) izango da. Jarraian, eredu bakunean egin bezala, eredu orokor batean Ekuazio Normalak lortuko dira, eta, askatuz, ereduko koefizienteen KTA zenbatesleak.

$$min_{\hat{\beta}} HKB \equiv min_{\hat{\beta}}\ \hat{U}'\hat{U} \equiv min_{\hat{\beta}} \left(Y - X\hat{\beta}\right)'(Y - X\hat{\beta})$$

$$\frac{\partial\, \hat{U}'\hat{U}}{\partial\, \hat{\beta}} = 0 \rightarrow -2X'\left(Y - X\hat{\beta}\right) = 0 \rightarrow X'Y = X'X\hat{\beta} \rightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

Non:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}, \quad X'X = \begin{pmatrix} N & \sum_{i=1}^{N} X_{2i} & \sum_{i=1}^{N} X_{3i} & \dots & \sum_{i=1}^{N} X_{Ki} \\ \sum_{i=1}^{N} X_{2i} & \sum_{i=1}^{N} X_{2i}^2 & \sum_{i=1}^{N} X_{2i}\, X_{3i} & \dots & \sum_{i=1}^{N} X_{2i}\, X_{Ki} \\ \sum_{i=1}^{N} X_{3i} & \sum_{i=1}^{N} X_{2i}\, X_{3i} & \sum_{i=1}^{N} X_{3i}^2 & \dots & \sum_{i=1}^{N} X_{3i}\, X_{Ki} \\ & & & & \\ \sum_{i=1}^{N} X_{Ki} X_{2i} & \sum_{i=1}^{N} X_{Ki} X_{2i} & \sum_{i=1}^{N} X_{Ki}\, X_{3i} & \dots & \sum_{i=1}^{N} X_{Ki}^2 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} \sum_{i=1}^{N} Y_i \\ \sum_{i=1}^{N} X_{2i}\, Y_i \\ \sum_{i=1}^{N} X_3\, Y_i \\ \vdots \\ \vdots \\ \sum_{i=1}^{N} X_{Ki}\, Y_i \end{pmatrix}$$

Soldaten adibideko ELOE-1 eredura bueltatuz, hau da:

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + u_i$$

$$i = 1, 2, \dots, 49$$

ereduan $X'X$ matrizea honako hau da:

$$\begin{pmatrix} 49 & \sum_{i=1}^{49} HEZ_i & \sum_{i=1}^{49} ANTZ_i & \sum_{i=1}^{49} GIZ_i & \sum_{i=1}^{49} MANT_i & \sum_{i=1}^{49} LANT_i \\ \sum_{i=1}^{49} HEZ_i & \sum_{i=1}^{49} HEZ_i^2 & \sum_{i=1}^{49} HEZ_i ANTZ_i & \sum_{i=1}^{49} HEZ_i GIZ_i & \sum_{i=1}^{49} HEZ_i MANT_i & \sum_{i=1}^{49} HEZ_i LANT_i \\ \sum_{i=1}^{49} ANTZ_i & \sum_{i=1}^{49} HEZ_i ANTZ_i & \sum_{i=1}^{49} ANTZ_i^2 & \sum_{i=1}^{49} GIZ_i ANTZ_i & \sum_{i=1}^{49} MANT_i ANTZ_i & \sum_{i=1}^{49} LANT_i ANTZ_i \\ \sum_{i=1}^{49} ANTZ_i & \sum_{i=1}^{49} HEZ_i GIZ_i & \sum_{i=1}^{49} GIZ_i ANTZ_i & \sum_{i=1}^{49} GIZ_i^2 & \sum_{i=1}^{49} GIZ_i MANT_i & \sum_{i=1}^{49} GIZ_i LANT_i \\ \sum_{i=1}^{49} MANT_i & \sum_{i=1}^{49} HEZ_i MANT_i & \sum_{i=1}^{49} MANT_i ANTZ_i & \sum_{i=1}^{49} GIZ_i MANT_i & \sum_{i=1}^{49} MANT_i^2 & \sum_{i=1}^{49} MANT_i LANT_i \\ \sum_{i=1}^{49} LANT_i & \sum_{i=1}^{49} HEZ_i LANT_i & \sum_{i=1}^{49} ANTZ_i LANT_i & \sum_{i=1}^{49} GIZ_i LANT_i & \sum_{i=1}^{49} MANT_i LANT_i & \sum_{i=1}^{49} LANT_i^2 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} \sum_{i=1}^{49} SOLDATA_i \\ \sum_{i=1}^{49} HEZ_i\, SOLDATA_i \\ \sum_{i=1}^{49} ANTZ_i\, SOLDATA_i \\ \sum_{i=1}^{49} GIZ_i\, SOLDATA_i \\ \sum_{i=1}^{49} MANT_i\, SOLDATA_i \\ \sum_{i=1}^{49} LANT_i\, SOLDATA_i \end{pmatrix} \quad \Rightarrow \quad \widehat{\beta} = (X'X)^{-1}\, X'Y = \begin{pmatrix} 747{,}14 \\ 83{,}025 \\ 42{,}15 \\ 802{,}399 \\ -734{,}078 \\ -343{,}814 \end{pmatrix}$$

Hortaz, dagokion Lagineko Erregresio Funtzioa honako hau da:

$$\widehat{SOLDATA}_i = 747{,}14 + 83{,}025\, HEZ_i + 42{,}15\, ANTZ_i + 802{,}399\, GIZ_i$$
$$- 734{,}078\, MANT_i - 343{,}814\, LANT_i \quad i = 1,2, \dots, 49$$

$\hat{\beta}_1 = 747{,}14 \rightarrow$ Hezkuntza-urterik ez duen (*HEZ* = 0) eta enpresako bulegoan sartu berria den (ANTZ=0) emakume baten hileroko batez besteko soldata 747,14 € dela zenbatesten da.

$\hat{\beta}_2 = 83{,}025 \rightarrow$ Hezkuntza-urte gehigarri bakoitzeko hileko batez besteko soldataren gehikuntza (antzinatasuna konstante mantenduz) 83,025 € dela zenbatesten da. Gehikuntza hori berdina da langile guztientzat, emakume nahiz gizon eta edozein lanpostu izanik.

$\hat{\beta}_3 = 42{,}15 \rightarrow$ Antzinatasun-urte gehigarri bakoitzeko hileroko batez besteko soldataren gehikuntza (hezkuntza-urteak konstante mantenduz) 42,15 € dela zenbatesten da. Gehikuntza hori berdina da langile guztientzat, emakume nahiz gizon eta edozein lanpostu izanik.

$\hat{\beta}_4 = 802{,}399 \rightarrow$ Gizon baten hileroko batez besteko soldata (eurotan), ezaugarri berdinak dituen —hau da, antzinatasun, hezkuntza-urte eta lanpostu berdineko— baina emakume batenarekin alderatuta 802,399 € dela zenbatesten da.

$\hat{\beta}_5 = -734{,}078 \rightarrow$ Mantenu-lanetan ari den langile baten hileroko batez besteko soldata, ezaugarri berdinak dituen —hau da, hezkuntza-urte, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen batenarekin alderatuta, 734,078 € gutxiago dela zenbatesten da.

$\hat{\beta}_6 = -343{,}814 \rightarrow$ Lantegian lan egiten duen langile baten hileroko batez besteko soldata, ezaugarri berdinak dituen —hau da, hezkuntza-urte, antzinatasun eta genero berdineko— baina bulegoan lan egiten duen batenarekin alderatuta, 343,814 € gutxiago dela zenbatesten da.

6.2. Lagineko Erregresio Funtzioaren (LEF) propietateak

Erregresio Lineal Orokorraren Ereduan, honako propietate hauek betetzen dira:

1. Hondarrak eta aldagai azaltzaileak ortogonalak dira: $X'\hat{u} = 0 \rightarrow \sum_{i=1}^{N} X_{ki}\,\hat{u}_i = 0$.

$$X'\hat{U} = X'\left(Y - \hat{Y}\right) = X'\left(Y - X\hat{\beta}\right) = X'(Y - X(X'X)^{-1}X'Y)$$
$$= X'Y - X'X\,(X'X)^{-1}X'Y = X'Y - X'Y = 0$$

2. Hondarrak eta zenbatetsitako aldagaia ortogonalak dira: $\hat{Y}'\hat{U} = 0 \rightarrow \sum_{i=1}^{N} \hat{Y}_i\,\hat{u}_i = 0$.

$$\hat{Y}'\hat{U} = \left(X\,\hat{\beta}\right)'\hat{U} = \hat{\beta}'\underbrace{X'\hat{U}}_{=0} = 0$$

3. Hondarren batura zero da: $\sum_{i=1}^{N} \hat{u}_i = 0$.
   Ereduak termino independentea izanez gero, hau da, $X_{1i} = 1$ baldin bada eta $X'\hat{u} = 0$ denez, zuzenean frogatzen da propietatea, X' matrizeko lehenengo errenkada batekoz osatuta baitago.

$$X'\hat{U} = 0 \quad \Leftrightarrow \quad \begin{pmatrix} \sum_{i=1}^{N}\hat{u}_i \\ \sum_{i=1}^{N}X_{2i}\hat{u}_i \\ \sum_{i=1}^{N}X_{3i}\hat{u}_i \\ \vdots \\ \vdots \\ \sum_{i=1}^{N}X_{Ki}\hat{u}_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} \quad \Rightarrow \quad \sum_{i=1}^{N}\hat{u}_i = 0$$

4. *Y*ren lagineko batez bestekoa eta *Y*ren zenbatespenen lagineko batez bestekoa berdinak dira: $\bar{Y} = \bar{\hat{Y}}$.

$$\hat{u}_i = Y_i - \hat{Y}_i \;\Rightarrow\; Y_i = \hat{Y}_i + \hat{u}_i$$

$$\sum_{i=1}^{N} Y_i = \sum_{i=1}^{N}\hat{Y}_i + \underbrace{\sum_{i=1}^{N}\hat{u}_i}_{=0}$$

$$\frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{1}{N}\sum_{i=1}^{N}\hat{Y}_i \quad \Rightarrow \quad \bar{Y} = \bar{\hat{Y}}$$

5. LEF lagineko batez bestekoen puntutik pasatzen da:

$$\sum_{i=1}^{N} \hat{u}_i = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_K X_{Ki}\right) = 0$$

$$\sum_{i=1}^{N} Y_i - N\hat{\beta}_1 - \hat{\beta}_2 \sum_{i=1}^{N} X_{2i} - \ldots - \hat{\beta}_K \sum_{i=1}^{N} X_{Ki} = 0$$

$$\sum_{i=1}^{N} Y_i = N\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{N} X_{2i} + \ldots + \hat{\beta}_K \sum_{i=1}^{N} X_{Ki}$$

$$\frac{1}{N}\sum_{i=1}^{N} Y_i = \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{N}\sum_{i=1}^{N} X_{2i} + \ldots + \hat{\beta}_K \frac{1}{N}\sum_{i=1}^{N} X_{Ki}$$

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \ldots + \hat{\beta}_K \bar{X}_K$$

Oharra: ELOE bateko LEFak lehenengo bi propietateak **beti** betetzen ditu. Azken hiru propietateak, berriz, ereduak termino independentea baldin badu beteko dira.

6.3. Doikuntza-neurria: mugatze-koefizientea (R^2)

Eredua zenbatetsita dagoenean, Lagineko Erregresio Funtzioa datuekiko nola doitzen den analizatzea interesatzen da. Hau da, zenbatetsitako funtzioa eta datuak nola doitzen diren.

Aldagai endogenoaren balio guztiak berdinak izango balira, horiek eta aldagai endogenoaren lagineko batez bestekoa berdinak lirateke, eta horrela, aldagai endogenoaren aldakuntza zero izango da. Orokorrean, aldagai endogenoaren aldakortasuna behatutako balioen eta aldagai horren batez bestekoaren arteko distantzia bezala definitzen da, eta kalkulatutako aldakortasunaren karratuen batura (KTB: Karratu Totalaren Batura) aldagai azaltzaileekin azaldu nahi den aldakuntza da. Horrela, zenbat eta handiagoa izan azaldutako aldakortasunaren proportzioa, hainbat eta hobea izango da egindako doikuntza.

Mugatze-koefizientea (R^2) erregresioaren doikuntza-neurri bat da. Aldagai azaltzaileek aldagai endogenoaren aldakuntza zenbat azaltzen duten neurtzen du; hau da, aldagai endogenoaren Karratu Totalaren Baturaren (KTB) ehunekoa azaltzen duten.

$$KTB = \sum_{i=1}^{N} (Y_i - \bar{Y})^2 = \sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2 = Y'Y - N\bar{Y}^2$$

Aurretik azaldu bezala, $Y_i = \hat{Y}_i + \hat{u}i$ da, eta zenbatetsitako balioak eta hondarrak ortogonalak direnez:

$$\sum_{i=1}^{N} Y_i^2 = \sum_{i=1}^{N} (\hat{Y}_i + \hat{u}_i)^2 = \sum_{i=1}^{N} \hat{Y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2 + 2\sum_{i=1}^{N} \hat{Y}_i \hat{u}_i = \sum_{i=1}^{N} \hat{Y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2$$

Bi aldeetan $N\bar{Y}^2$ kenduz eta $\bar{Y} = \bar{\hat{Y}}$ denez:

$$\sum_{i=1}^{N} Y_i^2 - N\,\bar{Y}^2 = \sum_{i=1}^{N} \hat{Y}_i^2 - N\,\bar{Y}^2 + \sum_{i=1}^{N} \hat{u}_i^2$$

$$\sum_{i=1}^{N} (Y_i - \bar{Y})^2 = \sum_{i=1}^{N} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{N} \hat{u}_i^2$$

$$\sum_{i=1}^{N} (Y_i - \bar{Y})^2 = \sum_{i=1}^{N} (\hat{Y}_i - \bar{\hat{Y}})^2 + \sum_{i=1}^{N} \hat{u}_i^2 \qquad \rightarrow \qquad KTB = KAB + HKB$$

Erregresioaren mugatze-koefizientea honela definitzen da:

$$R^2 = \frac{KAB}{KTB} = 1 - \frac{HKB}{KTB}$$

6.4. **Ereduko koefizienteen KTA zenbatesleen lagin txikitako propietateak**

Demagun Erregresio Lineal Orokorraren Eredua non oinarrizko hipotesiak betetzen baitira.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + u_i \qquad i = 1, 2, \ldots, N$$

Ereduko koefizienteen KTA zenbateslea honako hau da:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Lagin txikitan honako propietate hauek betetzen dituzte:

— Linealtasuna. KTA zenbateslea perturbazioen (eta aldagai endogenoaren) konbinazio lineal bat dela kontuan hartuz, KTA zenbateslea lineala da perturbazioekiko.

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + U) = \beta + (X'X)^{-1}X'U$$

— Alboragabetasuna. $E_X(u) = 0$ dela jakinik, alboragabeak dira; hau da, bere itxaropena ereduko benetako koefizienteen berdina da: $E_X(\hat{\beta}) = \beta$.

$$E_X(\hat{\beta}) = E_X(\beta + (X'X)^{-1}X'U) = E_X(\beta) + (X'X)^{-1}X'E_X(U) = \beta$$

— Bariantza minimoak dituzte:

$$Bar(\hat{\beta}) = E_X\left(\hat{\beta} - E(\hat{\beta})\right)\left(\hat{\beta} - E(\hat{\beta})\right)' = E_X([(X'X)^{-1}X'U][(X'X)^{-1}X'U]')$$

$$= E_X([(X'X)^{-1}X'U][U'X\,(X'X)^{-1}]) = E_X((X'X)^{-1}\,X'\,U\,U'\,X\,(X'X)^{-1}) =$$

$$= (X'X)^{-1}\,X'\,E_X(U\,U')\,X(X'X)^{-1} = (X'X)^{-1}\,X'\,\sigma^2 I\,X\,(X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1}\,X'\,I\,X\,(X'X)^{-1} = \sigma^2\,(X'X)^{-1}$$

Gauss-Markov-en Teorema: erregresio linealeko ereduaren oinarrizko hipotesiak betetzen direnean, zenbatesle lineal eta alboragabe guztien artetik, KTA zenbateslea efizientea da, hau da, bariantza txikiena du.

6.5. Perturbazioen bariantzaren zenbateslea

Eredu bakunean bezala, perturbazioen bariantza ezezaguna izaten denez, honako zenbatesle alboragabe hau erabiliko da:

$$\hat{\sigma}^2 = \frac{\hat{U}'\hat{U}}{N-K} = \frac{HKB}{N-K}$$

6.6. Ereduko koefizienteen KTA zenbatesleen bariantza-kobariantza matrizearen zenbateslea

Erregresio Lineal Orokorraren Ereduan oinarrizko hipotesiak betetzen direnean, honako hau da ereduko koefizienteen KTA zenbateslearen bariantza-kobariantza matrizea:

$$Bar(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Eta perturbazioen bariantzaren zenbatesle alboragabea, berriz, $\hat{\sigma}^2 = \frac{\hat{U}'\hat{U}}{N-K}$. Hortaz, ereduko koefizienteen KTA zenbatesleen bariantza-kobariantza matrizearen zenbateslea honako hau izango da:

$$\widehat{Bar}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$$

Non, diagonalean, ereduko koefizienteen zenbatesleen bariantza zenbatetsiak ageri baitira, eta, diagonaletik kanpora, koefizienteen zenbatesleen kobariantzak. Hau da:

$$\widehat{Bar}(\hat{\beta}) = \begin{pmatrix} \widehat{Bar}(\hat{\beta}_1) & \widehat{Kob}(\hat{\beta}_1,\hat{\beta}_2)\ldots & \widehat{Kob}(\hat{\beta}_1,\hat{\beta}_K) \\ \widehat{Kob}(\hat{\beta}_1,\hat{\beta}_2) & \widehat{Bar}(\hat{\beta}_2)\ \ldots & \widehat{Kob}(\hat{\beta}_2,\hat{\beta}_K) \\ \vdots & \ddots & \vdots \\ \widehat{Kob}(\hat{\beta}_2,\hat{\beta}_K) & & \widehat{Bar}(\hat{\beta}_K) \end{pmatrix}$$

$$\widehat{Bar}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{NS^2_{X_j}\,(1-R^2_j)}$$

Non:

N: laginaren tamaina da.
R_j^2: X_j aldagaia beste aldagai azaltzaileen funtzioan zenbatetsiz lortzen den mugatze-koefizientea da (X_j aldagaiak eskaintzen duen informazioa da, baina beste aldagaiek ere jasotzen dute).
$1 - R_j^2$: X_j aldagaiak eskaintzen duen informazioa da, baina beste aldagaiek ez dute jasotzen.
$S^2_{X_j}$: X_j aldagaiaren lagineko bariantza da.
$\hat{\sigma}^2 = \frac{HKB}{(N-K)}$

6.7. Ereduko koefizienteen konfiantza-tartea

Erregresio Lineal Orokorraren Ereduan oinarrizko hipotesiak betetzen direnean:

$$\hat{\beta} \sim N(\beta_k \quad , \quad \sigma^2 a_{kk})$$

Non a_{kk}, $(X'X)^{-1}$ matrizeko diagonaleko k elementua baita. Hortaz, $\frac{\hat{\sigma}^2 \, (N-K)}{\sigma^2} \sim \aleph^2_{(N-K)}$ dela jakinik, eta, gainera, $\hat{\sigma}^2$ eta $\hat{\beta}_k$ independenteak direnez:

$$t = \frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}\sqrt{a_{kk}}} = \frac{\hat{\beta}_k - \beta_k}{\widehat{des}\,(\hat{\beta}_k)} \sim t_{(N-K)}$$

Beraz:

$$\Pr\left[- t_{(N-K)}{}^{\alpha}/_{2} \leq \frac{\hat{\beta}_k - \beta_k}{\widehat{des}\,(\hat{\beta}_k)} \leq -t_{(N-)}{}^{\alpha}/_{2}\right] = 1 - \alpha$$

$$\Pr\left[\hat{\beta}_k - t_{(N-K)}{}^{\alpha}/_{2}\, \widehat{des}\left(\hat{\beta}_k\right) \leq \beta \leq \hat{\beta}_k + t_{(N-K)}{}^{\alpha}/_{2}\, \widehat{des}\left(\hat{\beta}_k\right)\right] = 1 - \alpha$$

$$\mathrm{KT}(\beta_k)_{1-\alpha} = \left[\hat{\beta}_k - t_{(N-K)}{}^{\alpha}/_{2}\, \widehat{des}\left(\hat{\beta}_k\right);\ \hat{\beta}_k + t_{(N-K)}{}^{\alpha}/_{2}\, \widehat{des}\left(\hat{\beta}_k\right)\right]$$

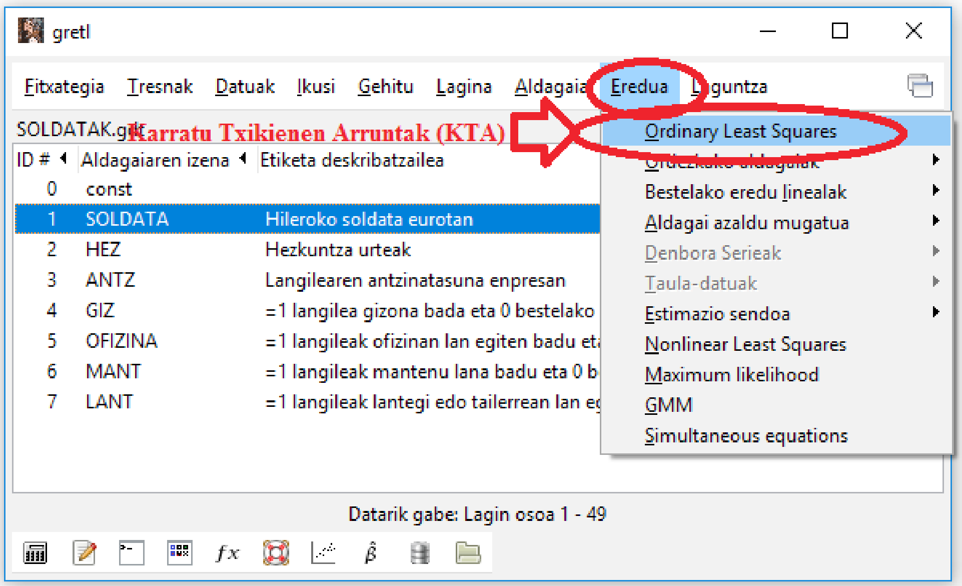
6.8. Zenbatespena Gretl-ekin

Soldaten adibideko ELOE-1 eredura bueltatuz, hau da,

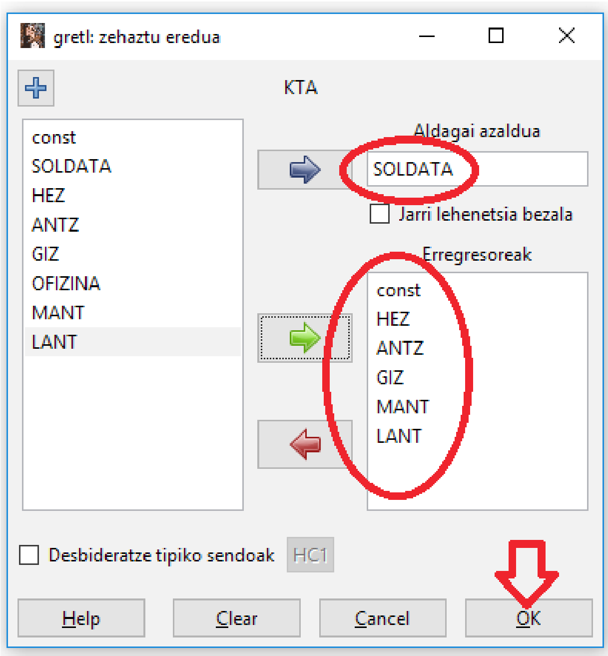
$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + u_i$$

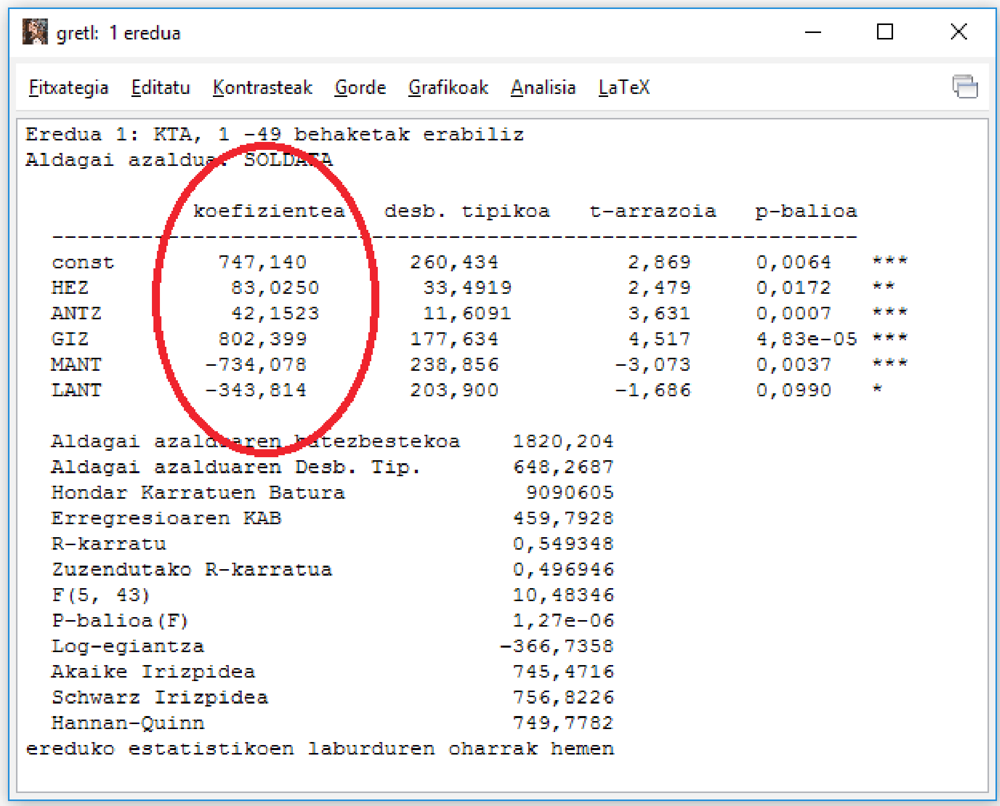
$$i = 1, 2, ..., 49$$

Eredu hau KTA bitartez zenbatesteko *Eredua* aukeratu eta *Karratu Txikien Arrunta* klikatuko da:



Jarraian aldagai azaldua eta azaltzaileak aukeratuko dira, eta *OK* klikatuko zenbatespen-emaitzak lortzeko.






```
Eredua 1: KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: SOLDATA

             koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ---------------------------------------------------------------
  const        747,140       260,434          2,869      0,0064   ***
  HEZ           83,0250       33,4919         2,479      0,0172   **
  ANTZ          42,1523       11,6091         3,631      0,0007   ***
  GIZ          802,399       177,634          4,517      4,83e-05 ***
  MANT        -734,078       238,856         -3,073      0,0037   ***
  LANT        -343,814       203,900         -1,686      0,0990   *

  Aldagai azalduaren batezbestekoa    1820,204
  Aldagai azalduaren Desb. Tip.       648,2687
  Hondar Karratuen Batura              9090605
  Erregresioaren KAB                  459,7928
  R-karratu                           0,549348
  Zuzendutako R-karratua              0,496946
  F(5, 43)                            10,48346
  P-balioa(F)                         1,27e-06
  Log-egiantza                       -366,7358
  Akaike Irizpidea                    745,4716
  Schwarz Irizpidea                   756,8226
  Hannan-Quinn                        749,7782
ereduko estatistikoen laburduren oharrak hemen
```

Dagokion LEFa honakoa da:

$$SO\widehat{LDA}TA_i == 747{,}14 + 83{,}02HEZ_i + 41{,}15\,ANTZ_i + 802{,}39\,GIZ_i - 734{,}07\,MANT_i - 343{,}81\,LANT_i$$

$$i = 1, 2, ..., 49$$

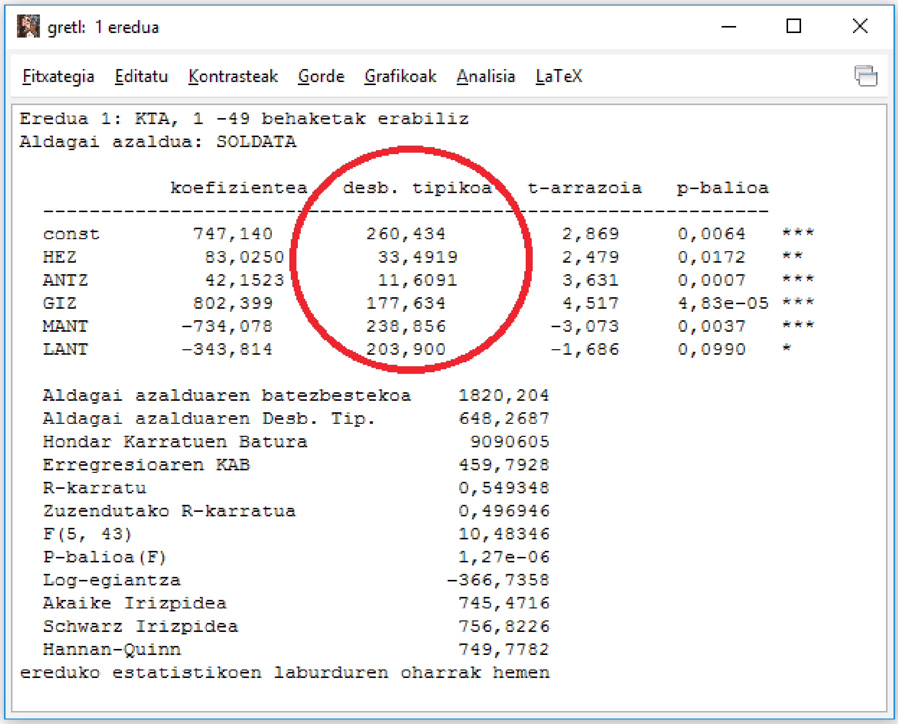
Ikus daitekeenez, zenbatespen honen doikuntza-neurria (mugatze-koefizientea), 0,54 denez, hezkuntza-urteak, antzinatasunak, generoak eta lanpostuak soldataren aldakuntzaren % 54 azaltzen dute.

```
gretl: 1 eredua

Fitxategia  Editatu  Kontrasteak  Gorde  Grafikoak  Analisia  LaTeX

Eredua 1: KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: SOLDATA

             koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ---------------------------------------------------------------
  const         747,140        260,434         2,869     0,0064    ***
  HEZ            83,0250        33,4919        2,479     0,0172    **
  ANTZ           42,1523        11,6091        3,631     0,0007    ***
  GIZ           802,399        177,634         4,517     4,83e-05  ***
  MANT         -734,078        238,856        -3,073     0,0037    ***
  LANT         -343,814        203,900        -1,686     0,0990    *

  Aldagai azalduaren batezbestekoa    1820,204
  Aldagai azalduaren Desb. Tip.       648,2687
  Hondar Karratuen Batura              9090605
  Erregresioaren KAB                  459,7928
  R-karratu                           0,549348
  Zuzendutako R-karratua              0,496946
  F(5, 43)                            10,48346
  P-balioa(F)                         1,27e-06
  Log-egiantza                       -366,7358
  Akaike Irizpidea                    745,4716
  Schwarz Irizpidea                   756,8226
  Hannan-Quinn                        749,7782
ereduko estatistikoen laburduren oharrak hemen
```

Era berean, baita dagokion Hondar Karratuen Batura zenbatekoa den ere:

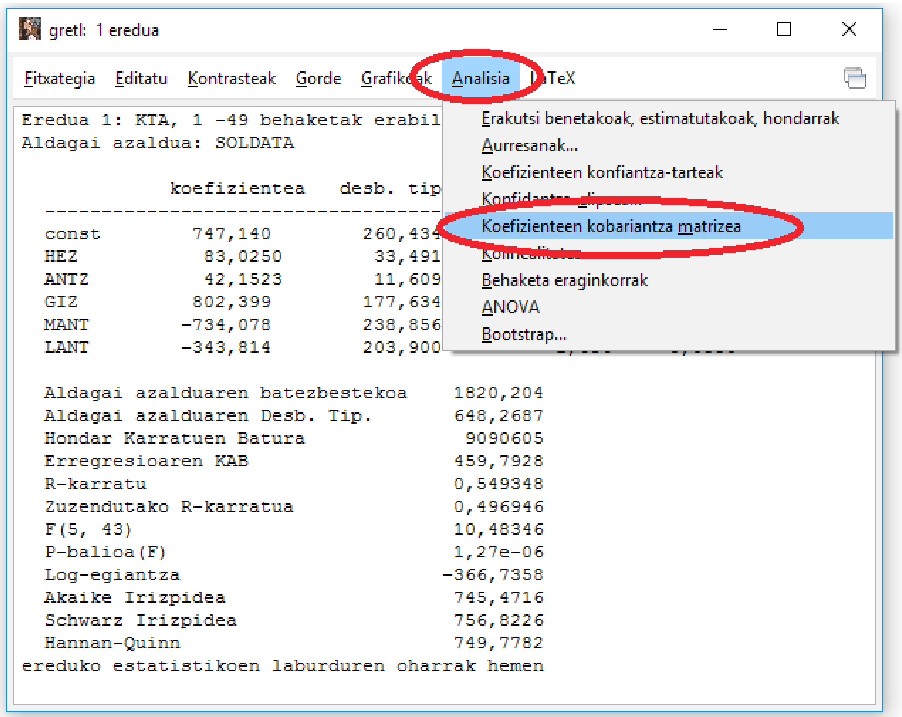
```
gretl: 1 eredua

Fitxategia  Editatu  Kontrasteak  Gorde  Grafikoak  Analisia  LaTeX

Eredua 1: KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: SOLDATA

             koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ---------------------------------------------------------------
  const         747,140        260,434         2,869     0,0064    ***
  HEZ            83,0250        33,4919        2,479     0,0172    **
  ANTZ           42,1523        11,6091        3,631     0,0007    ***
  GIZ           802,399        177,634         4,517     4,83e-05  ***
  MANT         -734,078        238,856        -3,073     0,0037    ***
  LANT         -343,814        203,900        -1,686     0,0990    *

  Aldagai azalduaren batezbestekoa    1820,204
  Aldagai azalduaren Desb. Tip.       648,2687
  Hondar Karratuen Batura              9090605
  Erregresioaren KAB                  459,7928
  R-karratu                           0,549348
  Zuzendutako R-karratua              0,496946
  F(5, 43)                            10,48346
  P-balioa(F)                         1,27e-06
  Log-egiantza                       -366,7358
  Akaike Irizpidea                    745,4716
  Schwarz Irizpidea                   756,8226
  Hannan-Quinn                        749,7782
ereduko estatistikoen laburduren oharrak hemen
```

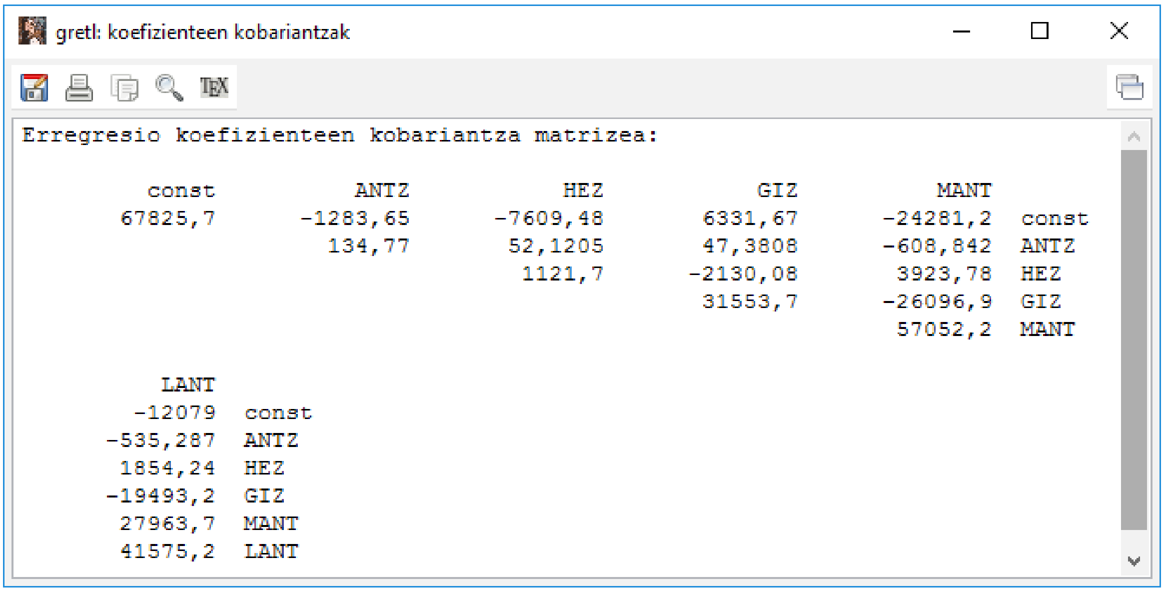
Zenbatespen-emaitzetan, hurrengo zutabean, ereduko koefizienteen KTA zenbatesleen desbideratze tipiko zenbatetsiak daude.



LANT

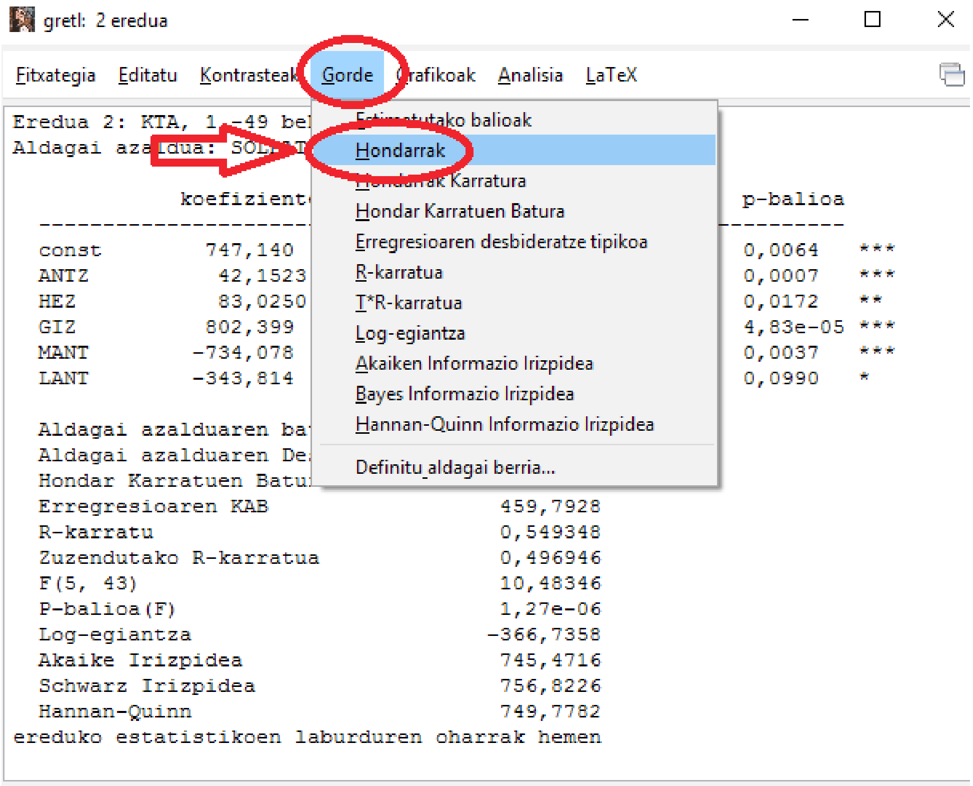
Ereduko koefizienteen KTA zenbatesleen bariantza-kobariantza matrize zenbatetsia lortzeko, KTA zenbatespen-leihatilan *Analisia* klikatuko da, eta, ondoren, *Koefizienteen kobariantza matrizea*.

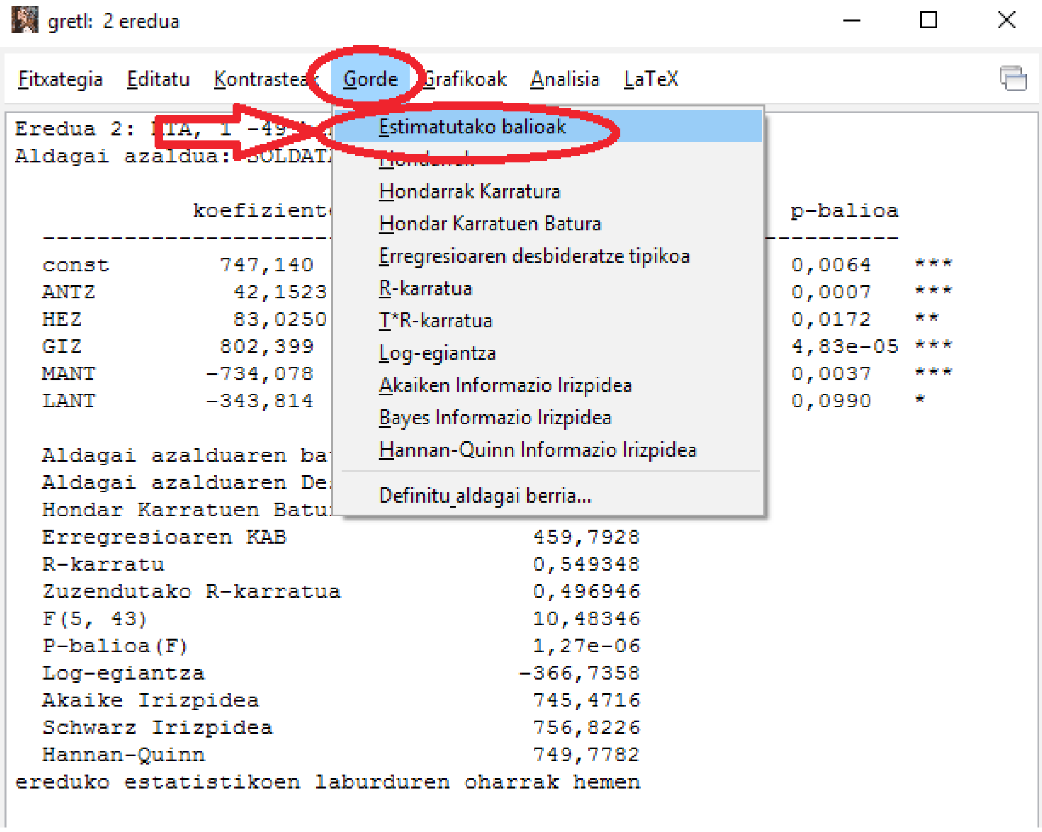


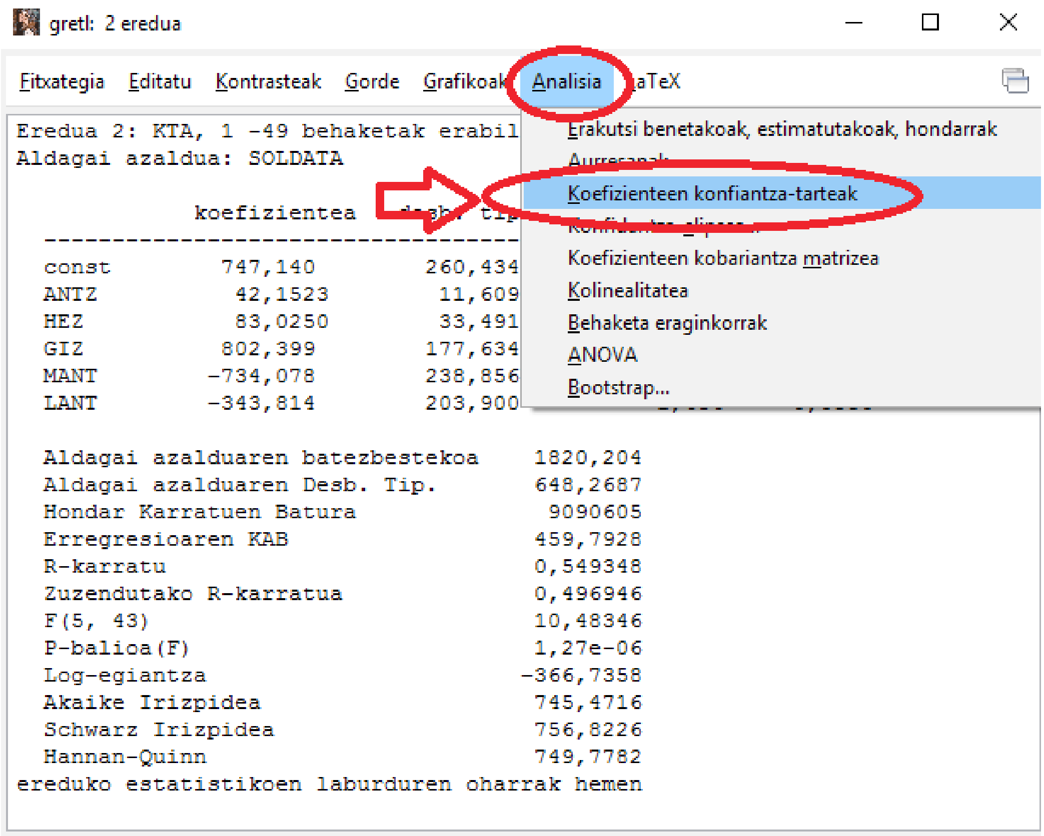


Horrela, bariantza eta kobariantza zenbatetsi guztiak lortuko dira. Bariantza zenbatetsiak diagonaleko balioak dira, eta diagonaletik kanpokoak, berriz, kobariantza zenbatetsiak. Hau da, $\hat{\beta}_3$ren bariantza zenbatetsia diagonaleko hirugarrena da ($\widehat{Bar}(\hat{\beta}_3)$ = 134,77 = 11,6091²), eta $\hat{\beta}_3$-ren eta $\hat{\beta}_4$-ren arteko kobariantza zenbatetsia, berriz, $\widehat{Kob}(\hat{\beta}_3, \hat{\beta}_4) = 47{,}38$.

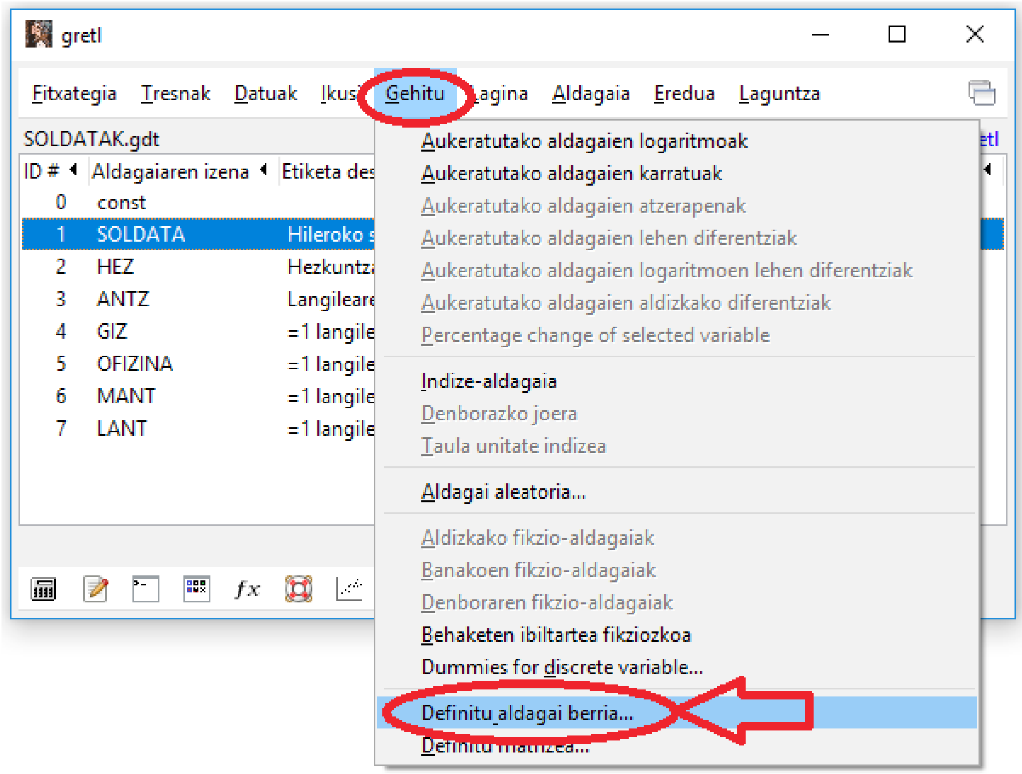
Eredu bakunean bezala, erregresioko hondarrak eta zenbatetsitako balioak gorde daitezke, grafiko desberdinak egin, koefizienteen konfiantza-tarteak lortu, etab.



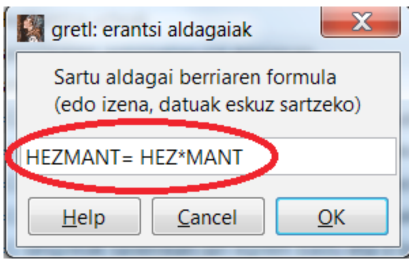

gretl: 2 eredua
Fitxategia Editatu Kontrasteak Gorde Grafikoak Analisia LaTeX
Estimatutako balioak
Hondarrak Karratura
Hondar Karratuen Batura
Erregresioaren desbideratze tipikoa
R-karratua
T*R-karratua
Log-egiantza
Akaiken Informazio Irizpidea
Bayes Informazio Irizpidea
Hannan-Quinn Informazio Irizpidea
Definitu aldagai berria...
Eredua 2:
Aldagai azaldua:
p-balioa
const 747,140 0,0064 ***
ANTZ 42,1523 0,0007 ***
HEZ 83,0250 0,0172 **
GIZ 802,399 4,83e-05 ***
MANT -734,078 0,0037 ***
LANT -343,814 0,0990 *
Erregresioaren KAB 459,7928
R-karratu 0,549348
Zuzendutako R-karratua 0,496946
F(5, 43) 10,48346
P-balioa(F) 1,27e-06
Log-egiantza -366,7358
Akaike Irizpidea 745,4716
Schwarz Irizpidea 756,8226
Hannan-Quinn 749,7782
ereduko estatistikoen laburduren oharrak hemen

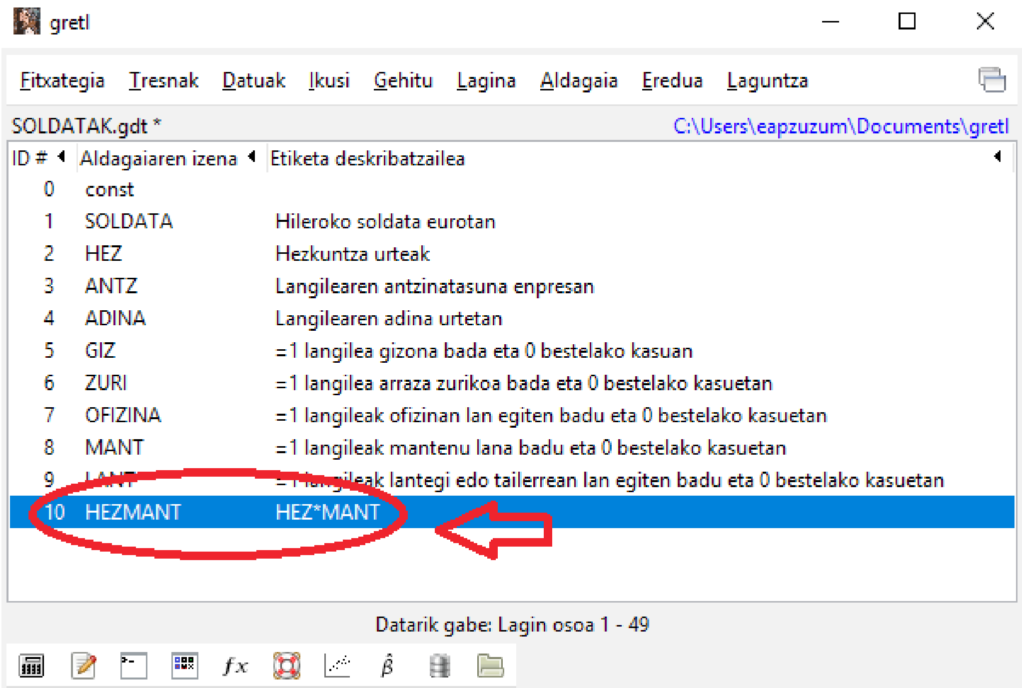

gretl: 2 eredua
Fitxategia Editatu Kontrasteak Gorde Grafikoak Analisia LaTeX
Erakutsi benetakoak, estimatutakoak, hondarrak
Koefizienteen konfiantza-tarteak
Koefizienteen kobariantza matrizea
Kolinealitatea
Behaketa eraginkorrak
ANOVA
Bootstrap...
Eredua 2: KTA, 1 -49 behaketak erabil
Aldagai azaldua: SOLDATA
koefizientea
const 747,140 260,434
ANTZ 42,1523 11,609
HEZ 83,0250 33,491
GIZ 802,399 177,634
MANT -734,078 238,856
LANT -343,814 203,900
Aldagai azalduaren batezbestekoa 1820,204
Aldagai azalduaren Desb. Tip. 648,2687
Hondar Karratuen Batura 9090605
Erregresioaren KAB 459,7928
R-karratu 0,549348
Zuzendutako R-karratua 0,496946
F(5, 43) 10,48346
P-balioa(F) 1,27e-06
Log-egiantza -366,7358
Akaike Irizpidea 745,4716
Schwarz Irizpidea 756,8226
Hannan-Quinn 749,7782
ereduko estatistikoen laburduren oharrak hemen

Era berean, proposatutako beste eredu orokorrak zenbatets daitezke, baina, zenbaitetan, aldagai berriak definitu beharko dira. Adibidez, $HEZ_i\ MANT_i$ iterazioa sortzeko jarraitu beharreko pausoak honako hauek dira:

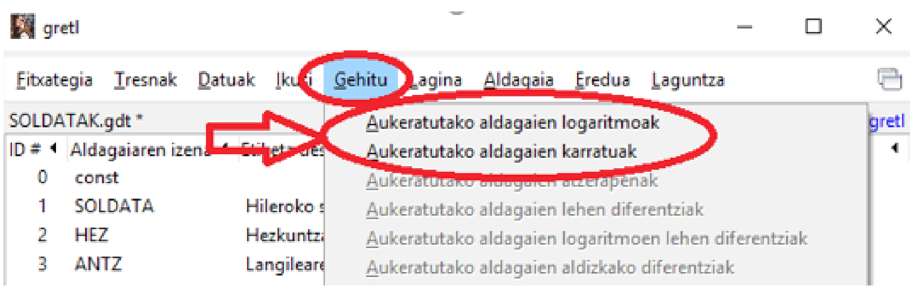


Betiere, aldagai berriaren izena jarri behar da, eta «=» ondoren, dagokion formula, non +, –, * eta / erabiltzen baitira, batuketa, kenketa, biderketa eta zatiketak egiteko. Amaieran, *OK* klikatuko da, aldagaien zerrendan, sortutako aldagai berria izateko.





Aldagaien berbidurak edota logaritmo nepertarrak lortzeko, beste aukera bat dago; hau da, *Gehitu* atalean, zuzenean aukera horiek eskaintzen ditu Gretl-ek.



7. gaia

Erregresio Lineal Orokorraren Eredua. Murrizketa linealen kontrasteak eta iragarpenak

7.1. **Sarrera**

Ekonometriaren funtsezko zeregin bat ekonomia erreal baten ezaguera kuantitatibo bat ekartzea da, eta, lortzeko metodo bat balizko alternatiboen kontrasteak dira. Hau da, ekonomialariek portaera ekonomikoari buruzko teoriak garatzen eta ebaluatzen dituzte, eta teoria horiek ebaluatzeko prozedurak dira hipotesien kontrasteak.

7.2. **Koefizienteen murrizketa linealen kontrasteak**

Eredu bakunean bezala egiten dira, kontuan izanik koefiziente bakoitzaren zenbatespena eta bere desbideratze zenbatetsia.

ELOEren oinarrizko hipotesiak kontuan izanik (perturbazioen normaltasuna barneratuz) eta perturbazioen bariantza ezezaguna izanik, honako estatistiko hau erabiliko da murrizketa bakarreko hipotesiak kontrastatzeko:

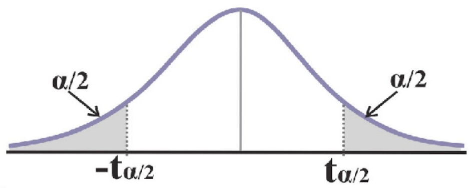
Koefiziente baten bi aldetako kontrasteak

Adibidez, ereduko edozein koefiziente β_k «c» konstante bat den ala ez kontrastatzeko, honako estatistiko hau erabiliko da:

$$t = \frac{\hat{\beta}_k - \beta_k}{\widehat{des}\,(\hat{\beta}_k)} \sim t_{(N-K)}$$

Eta honako hipotesi huts eta aurkako hau izanik:

$$\begin{cases} H_0: \beta_k = c \\ H_a: \beta_k \neq c \end{cases} \qquad t = \frac{\hat{\beta}_k - c}{\widehat{des}\,(\hat{\beta}_k)} \overset{H_0}{\sim} t_{(N-K)}$$



Hipotesi hutsa egiazkoa bada, t estatistikoak zerotik oso urruneko balioa izatea gertaezina da. Hortaz, H_0 baztertuko da α esangura-mailarekin, baldin eta $| t | > t_{(N-K)\alpha/2}$ bada.

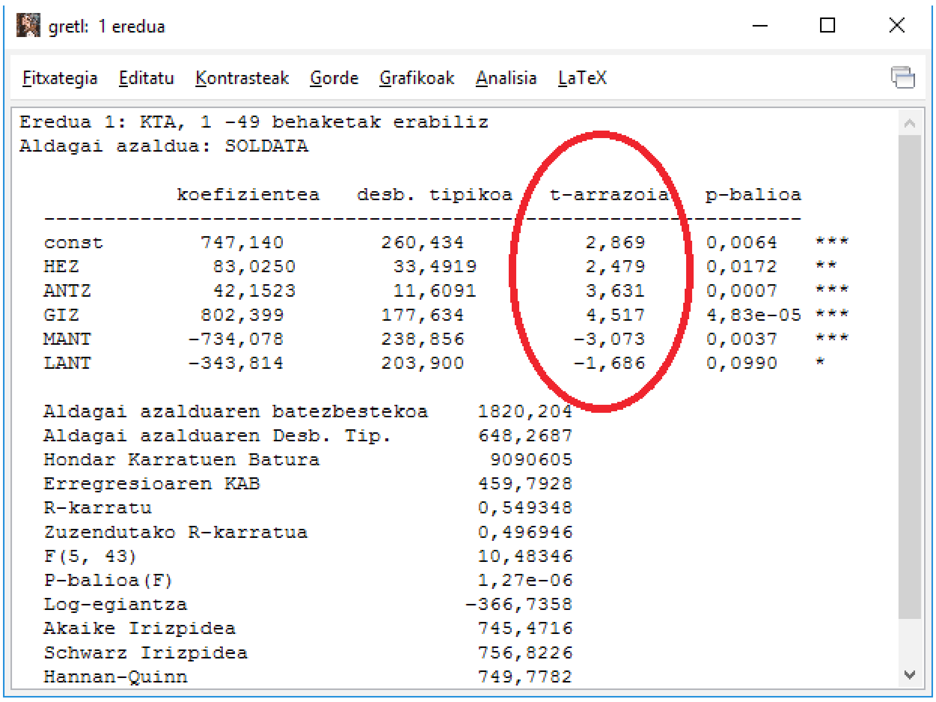
Ereduko koefizienteen konfiantza-tartearekin ere gauza daitezke horrelako kontrasteak, nahikoa izango da c balioa konfiantza-tartean dagoen ala ez ikustea. Hau da, «c» balioa tartean baldin badago, hipotesi hutsa ez da baztertzen, eta, alderantziz, tartetik kanpo baldin badago, hipotesi hutsa baztertzen da α esangura-mailarekin.

$$\mathrm{KT}(\beta_k)_{1-\alpha} = \left[\hat{\beta}_k - t_{(N-K)\alpha/2}\,\widehat{des}\left(\hat{\beta}_k\right);\ \hat{\beta}_k + t_{(N-K)\alpha/2}\,\widehat{des}\left(\hat{\beta}_k\right)\right]$$

Hipotesi hutsean jarri den «c» terminoaren balioa zero denean (H_0: $\beta_k = 0$), koefiziente horri dagokion aldagai azaltzailea nabaria edo esanguratsua den kontrastatzen ari gara, hau da, X_k aldagaiaren **banakako esangura-kontrastea** egiten.

$$\begin{cases} H_0: \beta_k = 0 \\ H_a: \beta_k \neq 0 \end{cases} \qquad t = \frac{\hat{\beta}_k}{\widehat{des}\left(\hat{\beta}_k\right)} \overset{H_0}{\sim} t_{(N-K)}$$

Kasu honetan, eta $| t | > t_{(N-K)\alpha/2}$ bada, hipotesi hutsa baztertzen da α esangura- mailarekin, eta X_k aldagaia nabari da. Bestela, hau da, $| t | < t_{(N-K)\alpha/2}$ bada, hipotesi hutsa ez da baztertzen eta X_k aldagaia ez-nabaria dela esango da. t estatistiko honi *t-arrazoia* edo *lagineko t* esaten zaio, eta Gretl-eko zenbatespen-leihatilan laugarren zutabean ageri diren balioak dira. Hau da:


gretl: 1 eredua

Fitxategia Editatu Kontrasteak Gorde Grafikoak Analisia LaTeX

```
Eredua 1: KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: SOLDATA

             koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ---------------------------------------------------------------
  const        747,140        260,434          2,869     0,0064    ***
  HEZ           83,0250        33,4919         2,479     0,0172    **
  ANTZ          42,1523        11,6091         3,631     0,0007    ***
  GIZ          802,399        177,634          4,517     4,83e-05  ***
  MANT        -734,078        238,856         -3,073     0,0037    ***
  LANT        -343,814        203,900         -1,686     0,0990    *

  Aldagai azalduaren batezbestekoa    1820,204
  Aldagai azalduaren Desb. Tip.       648,2687
  Hondar Karratuen Batura              9090605
  Erregresioaren KAB                  459,7928
  R-karratu                           0,549348
  Zuzendutako R-karratua              0,496946
  F(5, 43)                            10,48346
  P-balioa(F)                         1,27e-06
  Log-egiantza                       -366,7358
  Akaike Irizpidea                    745,4716
  Schwarz Irizpidea                   756,8226
  Hannan-Quinn                        749,7782
```

Bestalde, gogoratu hurrengo zutabean ageri diren p-balioa ere esangura-kontrasteak egiteko lagungarriak direla, hipotesi hutsa zein esangura-mailatik aurrera baztertzen den adierazten baitute; hau da, aldagaia ze esangura-mailatik aurrera den nabaria.

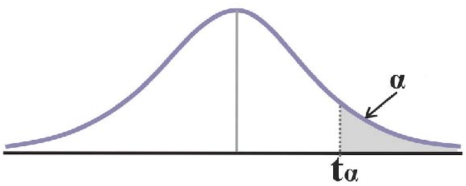
Koefiziente baten alde bateko kontrasteak

Adibidez, ereduko β_k koefizientea «c» konstantea baino handiagoa den kontrasta daiteke. Bi aldetako kontrastea egiteko erabili den estatistiko berdinarekin gauzatuko da, baina erabaki-irizpidea desberdina izango da, eskualde kritikoa alde batera egongo baita.

Alde bateko kontrasteak egiteko, kontu handia izan behar da hipotesi hutsa eta aurkakoa osatzeko orduan, hipotesi hutsean ageri behar baitu berdintasunak. Kasu honetan, «c» konstantea baino handiagoa dela kontrastatu nahi denez, aurkako hipotesian jarri beharko da, jarraian adierazten den bezala:

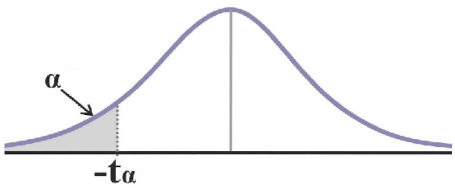
$$\begin{cases} H_0: \beta_k \leq c \\ H_a: \beta_k > c \end{cases} \qquad t = \frac{\hat{\beta}_k - c}{\widehat{des}\,(\hat{\beta}_k)} \overset{H_0}{\sim} t_{(N-K)}$$

Kasu honetan, aurkako hipotesian oinarrituko gara eskualde kritikoa osatzeko, eskuinerantz osatuz eta hipotesi hutsa baztertuz $t > t_{(N-K)\alpha}$ denean.



Bestalde, ereduko β_k malda «c» konstantea baino txikiagoa den kontrastatu nahi izanez gero, aurkako hipotesian jarri beharko da aukera hori. Lehen bezala, estatistiko berdinarekin gauzatuko da kontrastea, baina erabaki-irizpidea desberdina izango da; kasu honetan, eskualde kritikoa alde batera egongo da, baina ezkerrerantz, hipotesi hutsa baztertuz $t < -t_{(N-K)\alpha}$ denean.

$$\begin{cases} H_0: \beta_k \geq c \\ H_a: \beta_k < c \end{cases} \qquad t = \frac{\hat{\beta}_k - c}{\widehat{des}\,(\hat{\beta}_k)} \overset{H_0}{\sim} t_{(N-K)}$$



*Koefizienteen konbinazio baten kontrasteak (*q *= 1)*

Koefiziente bakar batekin egiten diren bezala egin daitezke, bai bi aldetakoak eta baita alde batekoak ere. Jarraian, zenbait adibiderekin azalduko da prozedura.

Demagun soldataren adibideko eredu orokorra:

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + u_i$$

Lehenik, mantenu-lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, lantegi edo tailerrean lan egiten duenak izango duen diferentziaren bikoitza dela kontrastatu nahi izanez gero, hau da, $\beta_5 = 2\,\beta_6$ kontrastatu nahi izanez gero, $w = \beta_5 - 2\,\beta_6$ izanik, $w = 0$ kontrastatzearen baliokidea da. Hortaz,

$$\begin{cases} H_0: w = 0 \\ H_a: w \neq 0 \end{cases} \qquad t = \frac{\hat{w}}{\widehat{des}\,(\hat{w})} \overset{H_0}{\sim} t_{(N-K)}$$

Non:

$$\hat{w} = \hat{\beta}_5 - 2\,\hat{\beta}_6 = \text{-734,078- 2 (-343,814) = -46,45}$$

$$\widehat{Bar}(\hat{w}) = \widehat{Bar}(\hat{\beta}_5) + (2)^2\widehat{Bar}(\hat{\beta}_6) - 2\,(1)(2)\widehat{Kob}(\hat{\beta}_5, \hat{\beta}_6) =$$

$$= 57.052{,}2 + (2)^2\, 41.575{,}2 - 2\,(1)(2)\, 27.963{,}7 = 111.498{,}2$$

Eta kontrastearen prozedura honako hau da: $|\, t \,| > t_{(N-K)\alpha/2}$ bada, hipotesi hutsa baztertzen da α esangura-mailarekin, eta, beraz, mantenu-lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, tailerrean lan egiten duenak izango duen diferentziaren bikoitza dela ez da baieztatuko. Bestela gertatzen bada, hau da, $|\, t \,| < t_{(N-K)\alpha/2}$ bada (adibidean bezala $t = 0{,}139 < t_{(43)0{,}025}$ $\alpha = \%\ 5$), α esangura-mailarekin hipotesi hutsa ez da baztertzen, eta mantenu-lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, tailerrean lan egiten duenak izango duen diferentziaren bikoitza dela esango da.

Kontraste hori gauzatzeko, w-ren konfiantza-tartea ere erabil daiteke. Hau da, $\beta_k$ren konfiantza-tartearen antzera:

$$\text{KT(w)}_{1-\alpha} = \left[\hat{w} - t_{(N-K)\alpha/2}\,\widehat{des}\,(\hat{w})\,;\ \hat{w} + t_{(N-K)\alpha/2}\,\widehat{des}\,(\hat{w})\right]$$

Kasu honetan, $w = 0$ tartearen barnean baldin badago, hipotesi hutsa ez da baztertzen, eta mantenu-lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, tailerrean lan egiten duenak izango duen diferentziaren bikoitza dela baieztatuko da.

Era berean, mantenu-lanetan ari den langile baten batez besteko ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, lantegi edo tailerrean lan egiten duenak izango duen diferentziaren bikoitza baino gehiago dela kontrastatu nahi izanez gero, hau da, $\beta_5 > 2\,\beta_6$ kontrastatu nahi izanez gero, $w = \beta_5 - 2\,\beta_6$ izanik, $w > 0$ kontrastatzearen baliokidea da. Hortaz:

$$\begin{cases} H_0: w \leq 0 \\ H_a: w > 0 \end{cases} \qquad t = \frac{\hat{w}}{\widehat{des}\,(\hat{w})} \overset{H_0}{\sim} t_{(N-K)}$$

Non:

$$\hat{w} = \hat{\beta}_5 - 2\,\hat{\beta}_6 = -734{,}078 - 2\,(-343{,}814) = -46{,}45$$

$$\widehat{Bar}(\hat{w}) = \widehat{Bar}(\hat{\beta}_5) + (2)^2\widehat{Bar}(\hat{\beta}_6) - 2\,(1)(2)\widehat{Kob}(\hat{\beta}_5, \hat{\beta}_6) =$$

$$= 57052{,}2 + (2)^2\, 41.575{,}2 - 2\,(1)(2)\; 27963{,}7 = 111498{,}2$$

Eta kontrastearen prozedura honako hau da: $t > t_{(N-K)\alpha}$ bada, hipotesi hutsa baztertzen da α esangura-mailarekin, eta, beraz, mantenu-lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, tailerrean lan egiten duenak izango duen diferentziaren bikoitza baino gehiago dela baieztatuko da. Bestela gertatzen bada, hau da, $t < t_{(N-K)\alpha}$ bada (adibidean $t = 0{,}139 < t_{(43)0{,}05} = 1{,}68$ $\alpha = \%\ 5$), α esangura-mailarekin hipotesi hutsa ez da baztertzen, eta mantenu-lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, lantegi edo tailerrean lan egiten duenak izango duen diferentziaren bikoitza baino gehiago ez dela esango da.

*Aldagai azaltzaileen baterako esangura-kontrastea (*q = K – *1)*

Kasu honetan, ereduko aldagai azaltzaile guztiak denak batera nabariak diren ala ez kontrastatzen da. Horretarako, estatistiko berezi bat erabil daiteke. Demagun Erregresio Lineal Orokorraren Eredua:

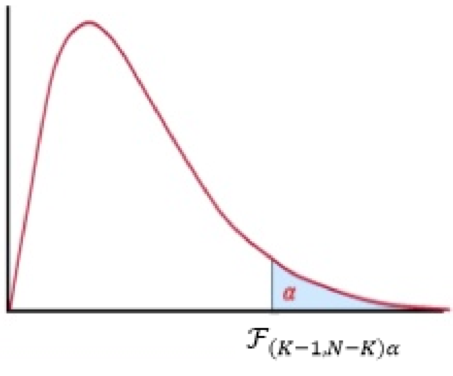
$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + u_i \qquad i = 1, 2, \ldots, N$$

$$\begin{cases} H_0: \beta_2 = \beta_3 = \cdots = \beta_K = 0 \\ H_a: \beta_2 \neq 0\ edo/eta\ \beta_3 \neq 0 \ldots edo/eta\ \beta_K \neq 0 \end{cases}$$

Hipotesi hutsa kontuan izanik:

$$F = \frac{R^2 / (K-1)}{(1-R^2) / (N-K)} \sim \mathcal{F}_{(K-1,\ N-K)}$$

Hortaz, α esangura-mailarekin hipotesi hutsa baztertzen da, eta aldagai azaltzaileak batera nabariak direla ondorioztatuko da $F > \mathcal{F}_{(K-1, N-K)\,|\,\alpha}$ bada, eta batera ez-nabariak $F < \mathcal{F}_{(K-1, N-K)\,|\,\alpha}$ bada.



Adibidean,

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + u_i$$

$$\begin{cases} H_0: \beta_2 = \beta_3 = \cdots = \beta_6 = 0 \\ H_a: \beta_2 \neq 0\; edo/eta\; \beta_3 \neq 0 \ldots edo/eta\; \beta_6 \neq 0 \end{cases}$$

$$F = \frac{R^2/(K-1)}{(1-R^2)/(N-K)} \sim \mathcal{F}_{(K-1,\; N-K)}$$

Non, R^2 = 0,549, $(K - 1) = q = 5$ eta $(N - K) = 43$ izanik, $F = 10{,}4835$ denez eta $\mathcal{F}_{(5,43)\,|\,0{,}05} = 2{,}43$ denez, hau da $F = 10{,}4835 > \mathcal{F}_{(5,43)\,|\,0{,}05} = 2{,}43$ denez, % 5eko esangura-mailarekin hipotesi hutsa baztertzen da, eta hezkuntza-urteak, antzinatasuna, generoa eta lan-postua batera nabariak direla ondorioztatzen da.

*Koefizienteen murrizketa linealen kontraste orokorra (*q*)*

Atal honetan azalduko den kontraste-prozedura edozein murrizketa kontrastatzeko baliagarria da, eta, beraz, aurretik azaldutako kontrasteak ere gauza daitezke estatistiko honekin: banakako esangura-kontrastea, baterako esangura-kontrastea, koefizienteen arteko erlazio bat edo batzuk, etab.

Demagun Erregresio Lineal Orokorraren Eredua:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + u_i \qquad i = 1, 2, \ldots, N$$

Edozein murrizketa $R\,\beta = r$ adierazpenarekin idatz daiteke. Eredu orokorrean, $\beta' = [\beta_1\; \beta_2\; \beta_3\; \ldots\; \beta_K]$ izanik, R eta r matrizeak ezagunak dira eta murrizketen funtzioan osatuko dira. Adibidez:

— $2\,\beta_2 - 6\,\beta_3 = 8 \quad \rightarrow \quad R = [0\; 2\; -6\; 0\; 0\; \ldots\; 0]$ eta $r = [8]$

— $\begin{cases} 2\,\beta_2 + 3\,\beta_4 = 6 \\ \beta_1 - 2\,\beta_3 = 3 \end{cases} \rightarrow \quad R = \begin{bmatrix} 0 & 2 & 0 & 3 & \ldots 0 \\ 1 & 0 & -2 & 0 & \ldots\; 0 \end{bmatrix}\; eta \quad r = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$

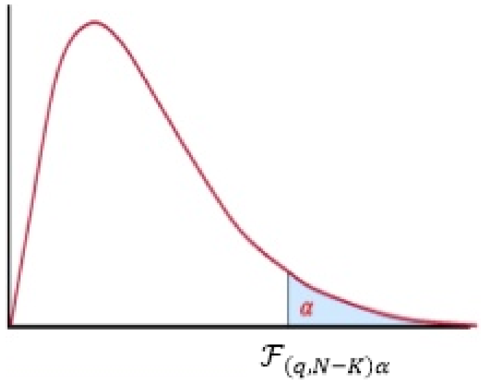
Hortaz:

$$\begin{cases} H_0: \mathrm{R}\,\beta = r \\ H_a: R\,\beta \neq r \end{cases}$$

Froga daiteke, hipotesi hutsa kontuan izanik:

$$\mathrm{F} = \frac{\left(R\,\hat{\beta} - r\right)'(R\,(X'X)^{-1}R)^{-1}(R\,\hat{\beta} - r)/q}{HKB/(N-K)} \sim \mathcal{F}_{(q,\; N-K)}$$

Eta α esangura-mailarekin hipotesi hutsa baztertzen da eta koefizienteen murrizketak ez direla betetzen ondorioztatzen da $F > \mathcal{F}_{(q,\, N-K) \mid \alpha}$ bada, eta murrizketak egiazkoak direla ondorioztatuko da, $F < \mathcal{F}_{(q,\, N-K) \mid \alpha}$ bada.



Hipotesi hutsa kontuan izanik, estatistiko honen baliokidea den beste estatistiko bat, Hondar Karratuen Baturetan oinarritzen dena, honako hau da:

$$F = \frac{(HKB_M - HKB_{EM})/q}{HKB_{EM}/(N-K)} \sim \mathcal{F}_{(q\,,\; N-K)}$$

Non HKB_{EM} ez-murriztutako ereduko Hondar Karratuen Batura baita, HKB_M murriztutako ereduko Hondar Karratuen Batura (hau da, murrizketak kontuan hartzen dituen eredua), q murrizketa kopurua, eta, azkenik, $(N-K)$ ez-murriztutako ereduko askatasun-graduak.

Demagun hurrengo erregresio-eredua eta $2\,\beta_2 - 6\,\beta_3 = 8$ murrizketa. Murrizketa hori kontuan hartzen duen murriztutako eredua lortzeko, honako pauso hauek jarraituko dira:

$$Y_i \;=\; \beta_1 + \beta_2\, X_{2i} + \beta_3 X_{3i} + u_i \qquad i = 1, 2, \ldots, N$$

$2\,\beta_2 - 6\,\beta_3 = 8 \rightarrow \beta_2 = 3\,\beta_3 + 4$ denez,

$$Y_i \;=\; \beta_1 + (3\beta_3 + 4)X_{2i} + \beta_3 X_{3i} + u_i^*$$

$$Y_i \;=\; \beta_1 + 4\, X_{2i} + \beta_3(3X_{2i} + X_{3i}) + u_i^*$$

$4\,X_{2i}$ ezaguna denez, berdintzaren ezker aldera pasatuko da aldagai endogeno berri bat osatuz. Lortuko den murriztutako eredua, kasu honetan eredu bakuna, honako hau da:

$$Y_i - 4\, X_{2i} = \beta_1 + \beta_3(3X_{2i} + X_{3i}) + u_i^*$$

Murriztutako ereduko aldagai endogeno berria $Y_i^* = Y_i - 4\,X_{2i}$ da, eta eredu bakuneko aldagai azaltzailea, berriz, $X_i^* = 3X_{2i} + X_{3i}$. Eredu hori KTA bitartez zenbatetsiko da:

$$Y^*_{(Nx1)} = X^*_{(Nx2)}\,\beta_{(2x1)} + U^*_{(Nx1)}$$

$$Y^* = \begin{pmatrix} Y_1^* \\ Y_2^* \\ \cdot \\ \cdot \\ Y_i^* \\ \cdot \\ \cdot \\ Y_N^* \end{pmatrix} \qquad X^* = \begin{pmatrix} 1 & 3\,X_{21} + X_{31} \\ 1 & 3\,X_{22} + X_{32} \\ \vdots & \vdots \\ 1 & 3\,X_{2i} + X_{3i} \\ \vdots & \vdots \\ 1 & 3\,X_{2N} + X_{3N} \end{pmatrix} \qquad \beta^*_{(2x1)} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Hortaz, eredua KTA bitartez zenbatetsiz ($\hat{\beta}^* = (X^{*\prime}\ X^*)^{-1}\ X^{*\prime}\, Y^*$), dagokion Hondar Karratuen Batura lortuko da, hau da, $HKB_M = Y^{*\prime}Y^* - \hat{\beta}^{*\prime}\, X^{*\prime}\, Y^*$.

7.3. **Koefizienteen murrizketa linealen kontrasteak Gretl-ekin**

Soldataren adibidea oinarritzat hartuz, Gretl-ekin zenbait kontraste nola gauzatzen diren azalduko da atal honetan.

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + u_i$$

Eredua KTA bitartez zenbatetsiko da:

gretl: 1 eredua

Fitxategia Editatu Kontrasteak Gorde Grafikoak Analisia LaTeX

```
Eredua 1: KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: SOLDATA

             koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ---------------------------------------------------------------
  const        747,140        260,434          2,869     0,0064   ***
  HEZ           83,0250        33,4919         2,479     0,0172   **
  ANTZ          42,1523        11,6091         3,631     0,0007   ***
  GIZ          802,399        177,634          4,517     4,83e-05 ***
  MANT        -734,078        238,856         -3,073     0,0037   ***
  LANT        -343,814        203,900         -1,686     0,0990   *

  Aldagai azalduaren batezbestekoa    1820,204
  Aldagai azalduaren Desb. Tip.       648,2687
  Hondar Karratuen Batura              9090605
  Erregresioaren KAB                  459,7928
  R-karratu                           0,549348
  Zuzendutako R-karratua              0,496946
  F(5, 43)                            10,48346
  P-balioa(F)                         1,27e-06
  Log-egiantza                       -366,7358
  Akaike Irizpidea                    745,4716
  Schwarz Irizpidea                   756,8226
  Hannan-Quinn                        749,7782
ereduko estatistikoen laburduren oharrak hemen
```

Zenbatespen-emaitzetan ikusten den bezala, banakako esangura-kontrasteak eginez gero, % 5eko esangura-mailarekin hezkuntza-urteak, antzinatasuna, eta generoa nabariak dira soldata azaltzeko. Bestalde, % 5eko esangura-mailarekin, mantenu-lana duen langile baten batez besteko soldataren diferentzia ezaugarri berdineko baina bulegoan lan egiten duen langile batenarekiko esanguratsua dela ondorioztatzen da, eta, azkenik, % 5eko esangura-mailarekin, tailerrean lan egiten duen langile baten batez besteko soldataren diferentzia ezaugarri berdineko baina bulegoan lan egiten duen batenarekiko ez dela esanguratsua esango da; bai, ordea, % 10eko esangura-mailarekin.

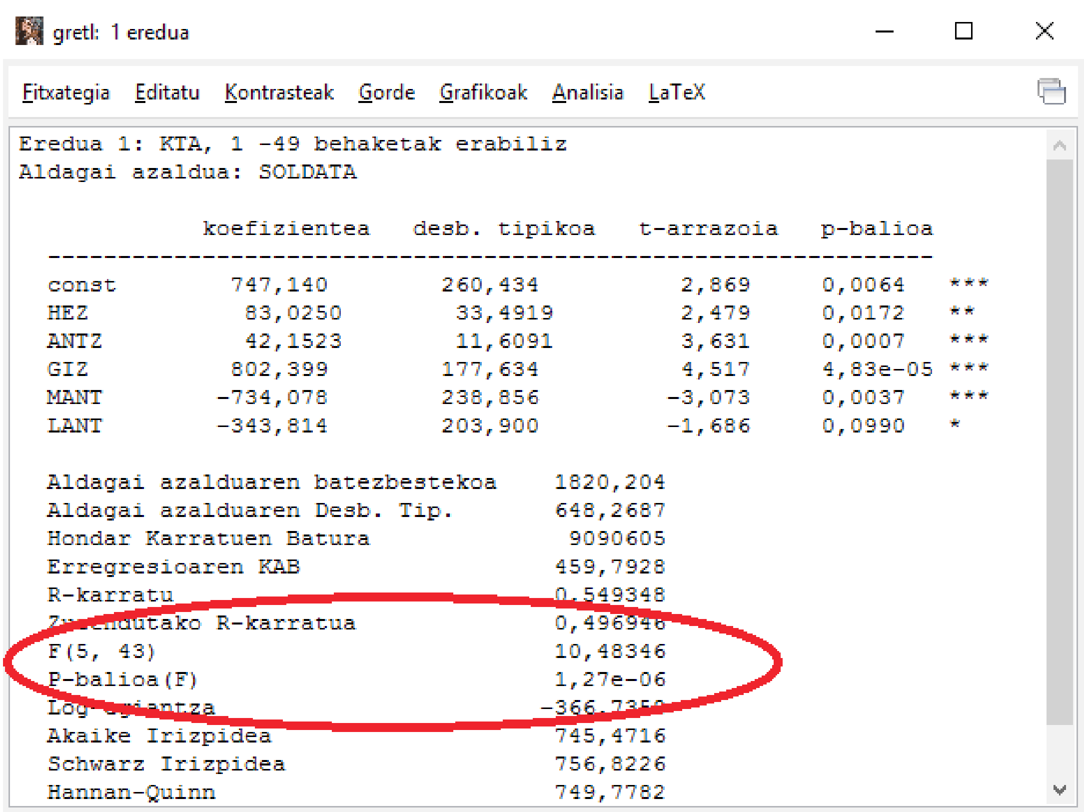
Bestalde, aldagai guztien baterako esangura kontrastatuz gero, hau da,

$$\begin{cases} H_0: \beta_2 = \beta_3 = \cdots = \beta_6 = 0 \\ H_a: \beta_2 \neq 0 \; edo/eta \; \beta_3 \neq 0 \dots edo/eta \; \beta_6 \neq 0 \end{cases}$$

kontrastatu nahi izanez gero, aukera desberdinak daude. Alde batetik, hipotesi hutsa kontuan izanik, honako estatistiko hau erabil daiteke, eta zenbatespen-emaitzetan mugatze-koefizientea azaltzen denez ($R^2 = 0{,}549$), $K - 1 = 5$ izanik eta $N - K = 43$, F-ren balioa eskuz kalkula daiteke.

$$F = \frac{R^2 / (K-1)}{(1-R^2) / (N-K)} \sim \mathcal{F}_{(K-1,\ N-K)}$$

Edo, zenbatespen-emaitzetan fijatuz, zuzenean F-ren balioa ageri dela ohartu. Hau da:



```
gretl: 1 eredua
Fitxategia  Editatu  Kontrasteak  Gorde  Grafikoak  Analisia  LaTeX

Eredua 1: KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: SOLDATA

             koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ---------------------------------------------------------------
  const        747,140        260,434          2,869     0,0064   ***
  HEZ           83,0250        33,4919         2,479     0,0172   **
  ANTZ          42,1523        11,6091         3,631     0,0007   ***
  GIZ          802,399        177,634          4,517     4,83e-05 ***
  MANT        -734,078        238,856         -3,073     0,0037   ***
  LANT        -343,814        203,900         -1,686     0,0990   *

  Aldagai azalduaren batezbestekoa    1820,204
  Aldagai azalduaren Desb. Tip.       648,2687
  Hondar Karratuen Batura              9090605
  Erregresioaren KAB                  459,7928
  R-karratu                           0,549348
  Zuzendutako R-karratua              0,496946
  F(5, 43)                            10,48346
  P-balioa(F)                         1,27e-06
  Log-egiantza                       -366,7359
  Akaike Irizpidea                    745,4716
  Schwarz Irizpidea                   756,8226
  Hannan-Quinn                        749,7782
```

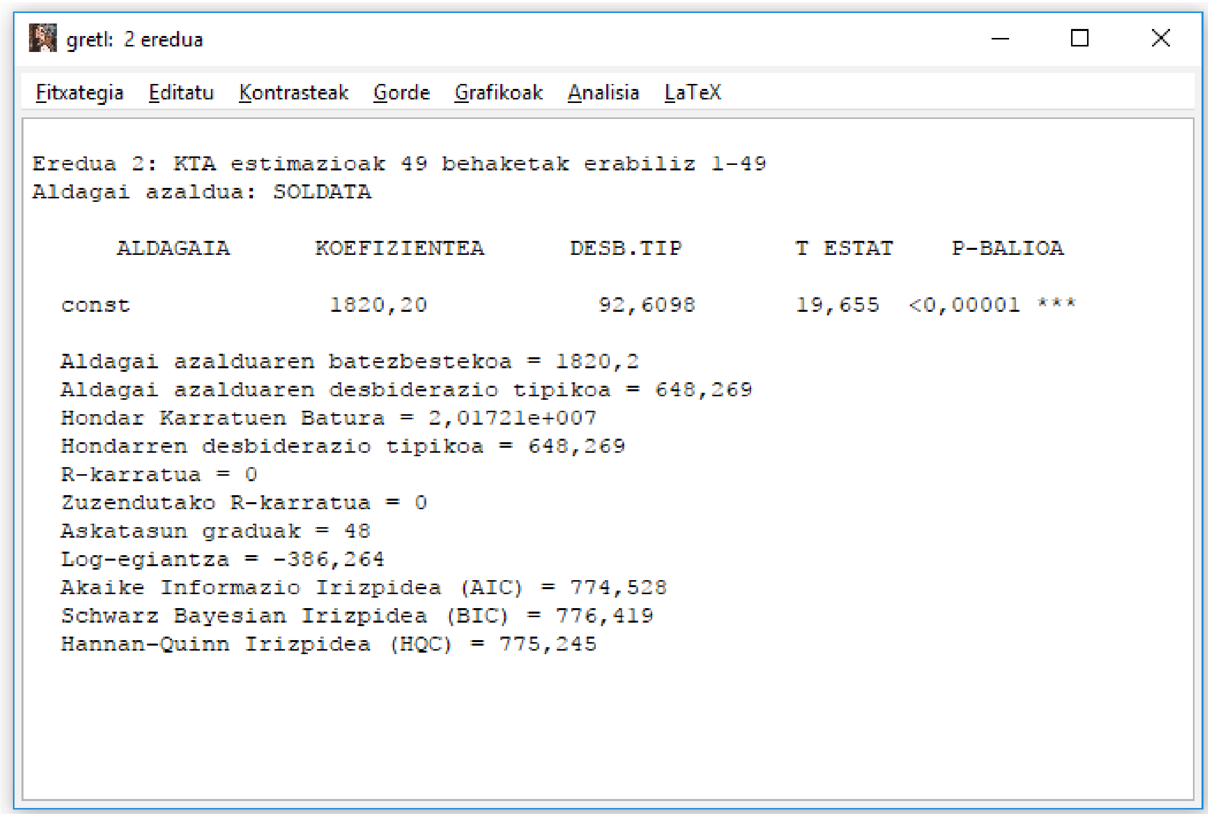
Hortaz, $F = 10{,}48 > \mathcal{F}_{(5,43)\,|\,0{,}05} = 2{,}43$ denez, % 5eko esangura-mailarekin aldagai azaltzaileak batera nabariak dira.

Beste aukera bat estatistiko orokorrarekin egitea da, hau da, hipotesi hutsa kontuan izanik:

$$F = \frac{(HKB_M - HKB_{EM})/q}{HKB_{EM}/(N-K)} \sim \mathcal{F}_{(q\,,\ N-K)}$$

Non ez-murriztutako eredua atal honetan adierazitako soldataren eredua baita, eta, beraz, $HKB_{EZ} = 9094696$ da, $q = 5$ eta $N - K = 43$. Murriztutako eredua, berriz, aldagai azaltzailerik gabeko eredua da:

$$SOLDATA_i = \beta_1 + u_i^*$$




```
Eredua 2: KTA estimazioak 49 behaketak erabiliz 1-49
Aldagai azaldua: SOLDATA

      ALDAGAIA       KOEFIZIENTEA       DESB.TIP         T ESTAT    P-BALIOA

  const               1820,20            92,6098       19,655   <0,00001 ***

  Aldagai azalduaren batezbestekoa = 1820,2
  Aldagai azalduaren desbiderazio tipikoa = 648,269
  Hondar Karratuen Batura = 2,01721e+007
  Hondarren desbiderazio tipikoa = 648,269
  R-karratua = 0
  Zuzendutako R-karratua = 0
  Askatasun graduak = 48
  Log-egiantza = -386,264
  Akaike Informazio Irizpidea (AIC) = 774,528
  Schwarz Bayesian Irizpidea (BIC) = 776,419
  Hannan-Quinn Irizpidea (HQC) = 775,245
```

Ohartu $HKB_M = 2{,}0172\ 10^7$ dela, hau da, *KTB*ren berdina. Aldagai azaltzailerik ez izatean, erregresio honi dagokion mugatze-koefizientea zero da. Logikoa denez, aukera honekin lortuko den emaitza aurreko aukerarekin lortutakoaren berdina da, aldagai azaltzaile guztiak batera nabariak direla, alegia.

Azkenik, edozein murrizketa kontrastatzeko erabil daitekeen aukera jarraituz, hau da, Gretl-ek edozein murrizketa kontrastatzeko aukera erabiliz, ondorio berdinera heltzen da. Edozein murrizketa kontrastatzeko segi beharreko pausoak honako hauek dira:

Lehendabizi, zenbatespen-leihatilan *kontrasteak* → *Murrizketa linealak* klikatu behar da, eta jarraian irtengo den leihatilan, kontrastatu nahi diren murrizketak idatziko dira. Murrizketa bakoitza errenkada batean idatzi behar da, eta, gainera, β_k erabili beharrean, b_k idatzi beharko da. Adibidean, baterako esangura-kontrastea egiteko $\beta_2 = \beta_3 = \cdots = \beta_6 = 0$ bost murrizketak bost errenkadetan idatziko dira, eta ados klikatu.

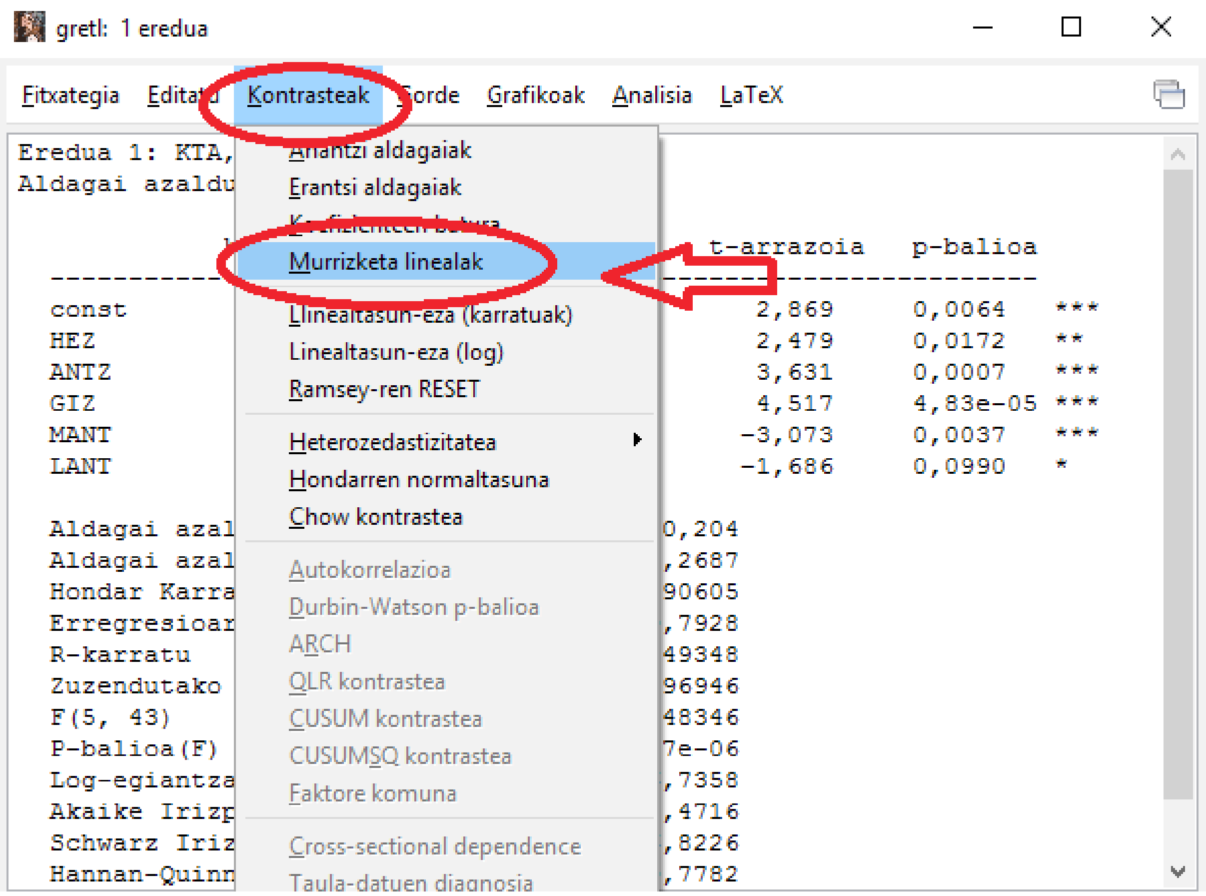
b2 = 0

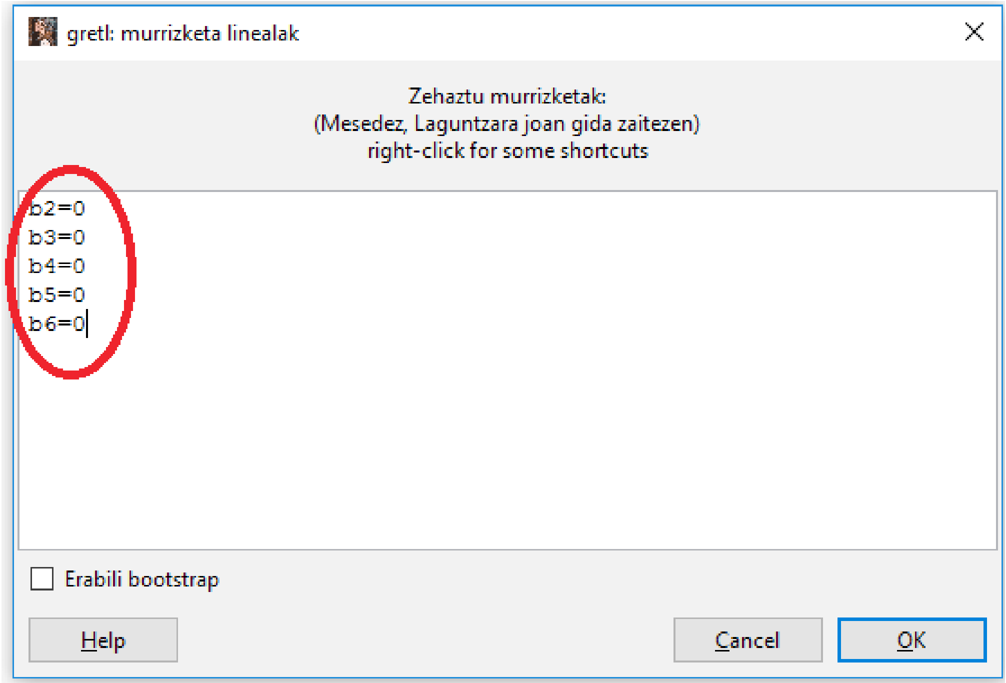
b3 = 0

b4 = 0

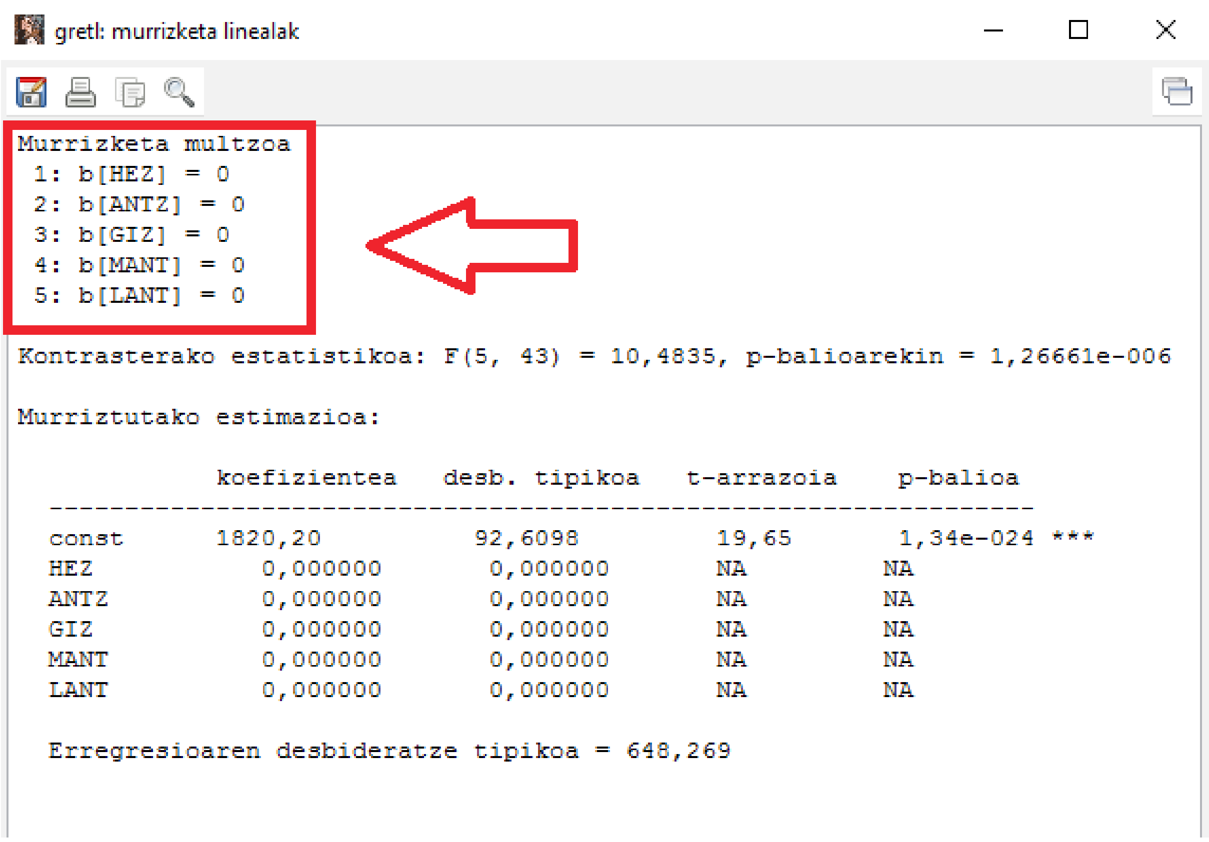
b5 = 0

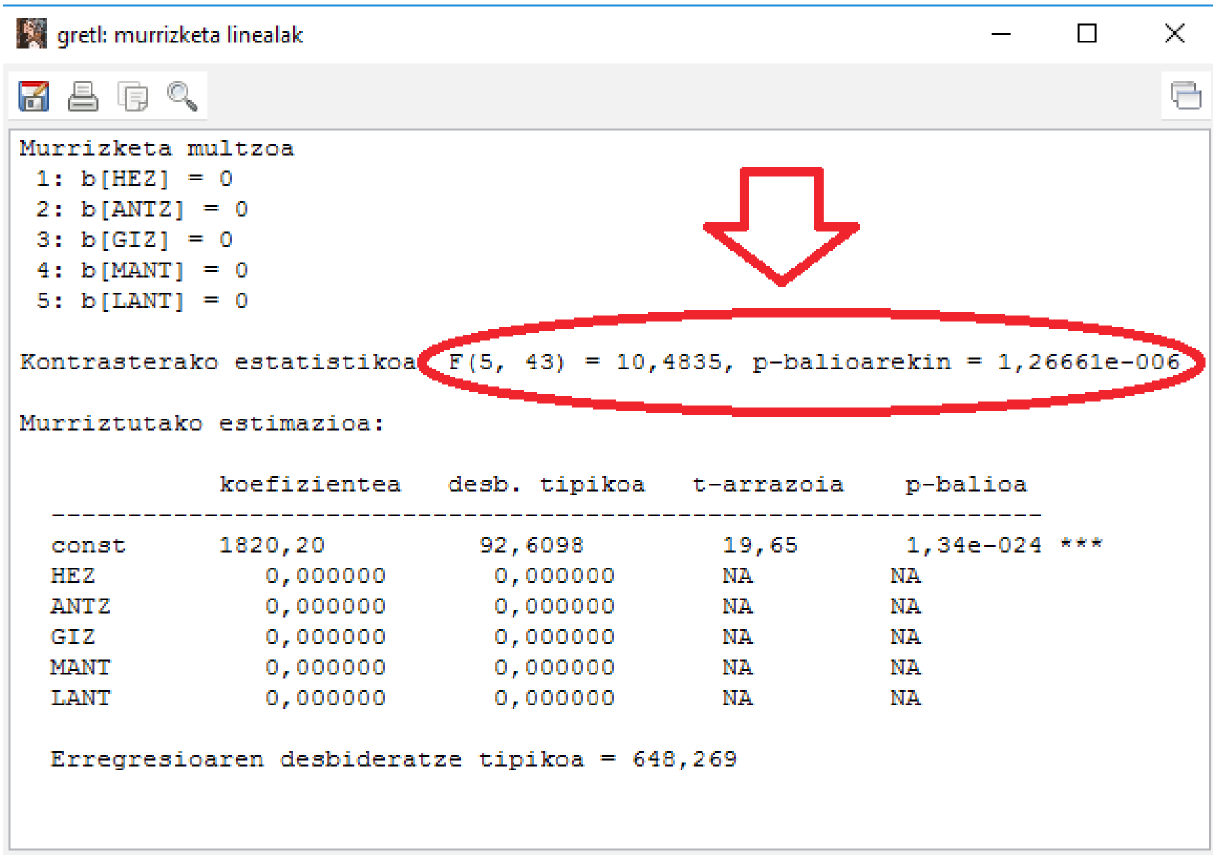
b6 = 0





Irtengo den leihatilan informazio desberdina ageri da, eta, askotan, komenigarria izaten da ezarritako murrizketa ondo idatzi den jakiteko. Bestalde, leihatila honetan zuzenean F estatistikoaren balioa ageri da, eta baita dagokion p-balioa ere. Jakina, aurreko aukerekin lortutako balio berdinak irtengo du.






```
Murrizketa multzoa
 1: b[HEZ] = 0
 2: b[ANTZ] = 0
 3: b[GIZ] = 0
 4: b[MANT] = 0
 5: b[LANT] = 0

Kontrasterako estatistikoa: F(5, 43) = 10,4835, p-balioarekin = 1,26661e-006

Murriztutako estimazioa:

             koefizientea   desb. tipikoa   t-arrazoia    p-balioa
  ---------------------------------------------------------------
  const       1820,20         92,6098         19,65       1,34e-024 ***
  HEZ            0,000000      0,000000       NA          NA
  ANTZ           0,000000      0,000000       NA          NA
  GIZ            0,000000      0,000000       NA          NA
  MANT           0,000000      0,000000       NA          NA
  LANT           0,000000      0,000000       NA          NA

  Erregresioaren desbideratze tipikoa = 648,269
```

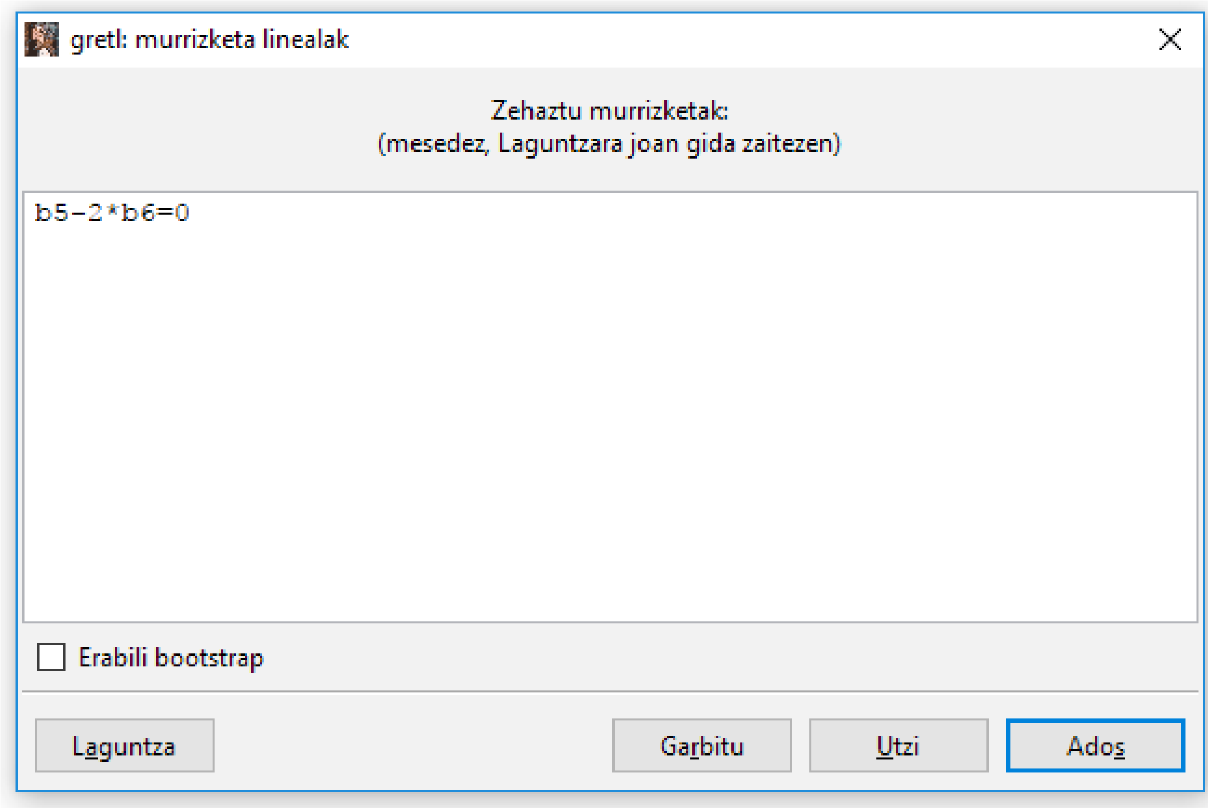
HKBtan oinarritzen den estatistikoarekin, edozein murrizketa kopuru kontrasta daiteke. Demagun gai honetan azaldu den murrizketa kontrastatu nahi dela; mantenu- lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, lantegi edo tailerrean lan egiten duenak izango duen diferentziaren bikoitza dela, alegia. Aurretik azaldu bezala, $\beta_5 = 2\,\beta_6$ kontrastatu nahi da, eta aukera bat da, azaldutako $w = \beta_5 - 2\,\beta_6 = 0$rekin egitea.

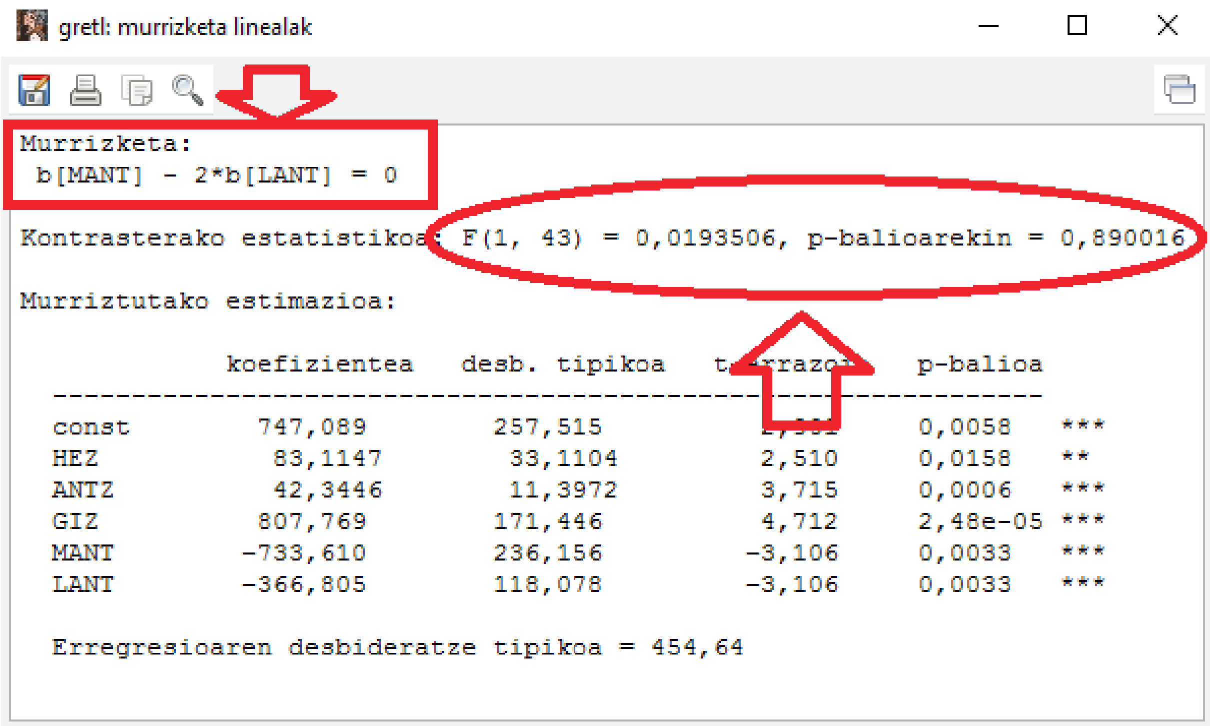
Beste aukera bat, ordea, estatistiko orokorrarekin egitea da, hau da:

$$\begin{cases} H_0\colon \beta_5 = 2\beta_6 \\ H_a\colon \beta_5 \neq 2\beta_6 \end{cases}$$

$$F = \frac{(HKB_M - HKB_{EM})/q}{HKB_{EM}/(N-K)} \sim \mathcal{F}_{(q,\ N-K)}$$

KTA zenbatespen-emaitzetako leihatilan murrizketa ezarriz gero, hau da, b5-2*b6=0, Gretl-ek zuzenean F estatistiko honen balioa ematen du, eta baita dagokion p-balioa ere, kontrastea erraz gauzatuz.


gretl: murrizketa linealak
Zehaztu murrizketak:
(mesedez, Laguntzara joan gida zaitezen)
b5-2*b6=0
Erabili bootstrap
Laguntza
Garbitu
Utzi
Ados


gretl: murrizketa linealak
Murrizketa:
b[MANT] - 2*b[LANT] = 0
Kontrasterako estatistikoa: F(1, 43) = 0,0193506, p-balioarekin = 0,890016
Murriztutako estimazioa:
koefizientea desb. tipikoa p-balioa
const 747,089 257,515 0,0058 ***
HEZ 83,1147 33,1104 2,510 0,0158 **
ANTZ 42,3446 11,3972 3,715 0,0006 ***
GIZ 807,769 171,446 4,712 2,48e-05 ***
MANT -733,610 236,156 -3,106 0,0033 ***
LANT -366,805 118,078 -3,106 0,0033 ***
Erregresioaren desbideratze tipikoa = 454,64

$F = 0{,}0193 < \mathscr{F}_{(1,43)\,|\,0{,}05} = 4{,}06$ denez, % 5eko esangura-mailarekin hipotesi hutsa ez da baztertzen, eta, beraz, mantenu-lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, tailerrean lan egiten duenak izango duen diferentziaren bikoitza dela ondorioztatzen da.

Ohartu horrela lortzen den F-ren balioa, 0,0193506, murrizketa hori w-rekin egiten genuen t-estatistikoaren berbidura dela, hau da, $(0{,}139)^2$.

$\beta_5 = 2\,\beta_6$ murrizketa kontrastatzeko osatu behar den F estatistikoaren elementu guztiak lortu nahi izanez gero, ez-murriztutako ereduko HKB_{EM} eta askatasun-graduak beharko dira, eta baita murriztutako ereduko HKB_M ere.

Ez-murriztutako eredua:

$$SOLDATA_i = \beta_1 + \beta_2\,HEZ_i + \beta_3\,ANTZ_i + \beta_4\,GIZ_i + \beta_5\,MANT_i + \beta_6\,LANT_i + u_i$$

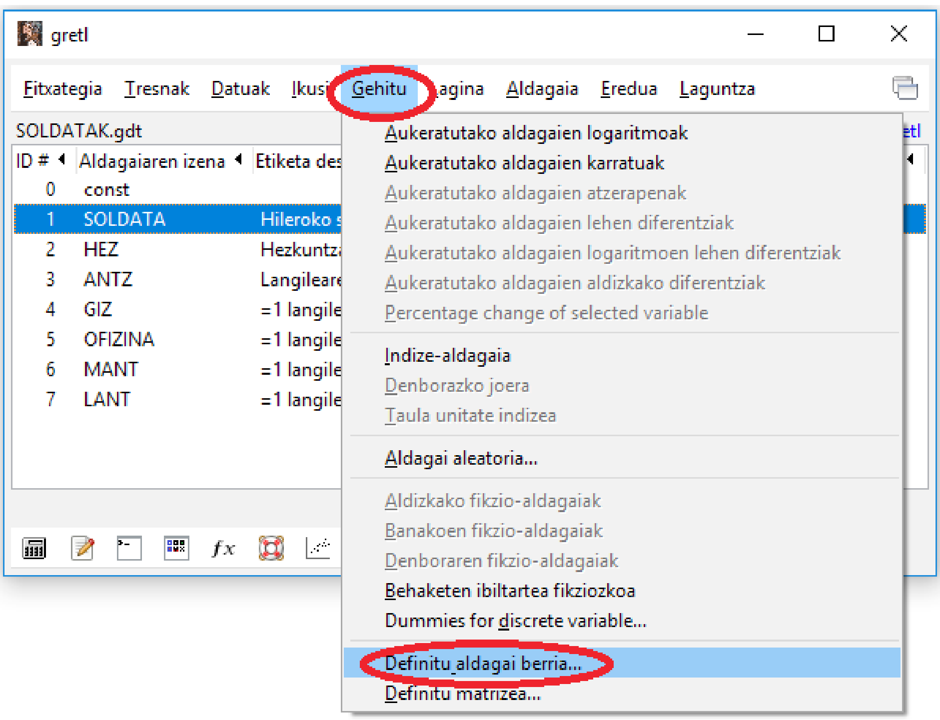
$$HKB_{EM} = 9{,}0906 \quad 10^6 \qquad\qquad N - K = 43$$

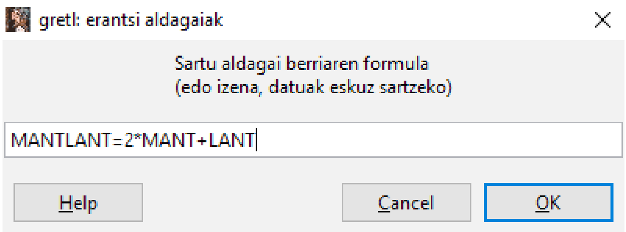
Murriztutako eredua lortzeko, murrizketa ereduan barneratuko da, eta elementuak ordenatu, ezagutzen den guztia berdintzaren ezkerraldera pasatuz (kasu honetan ez).

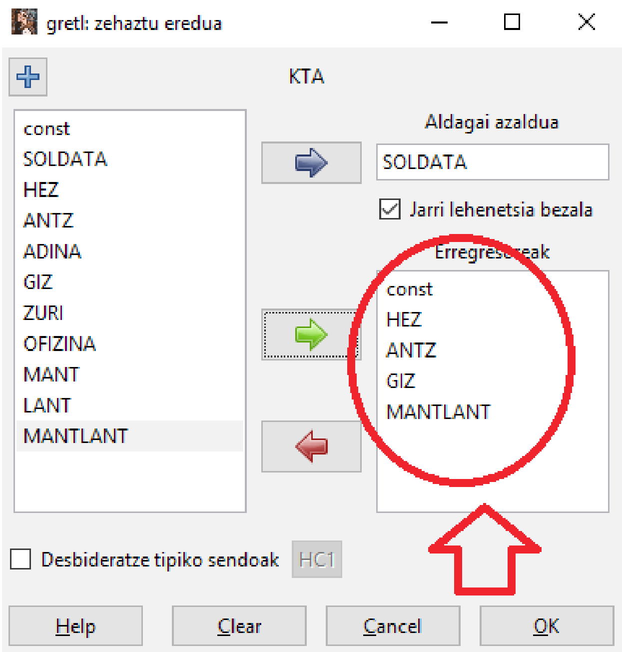
$$SOLDATA_i = \beta_1 + \beta_2\,HEZ_i + \beta_3\,ANTZ_i + \beta_4\,GIZ_i + 2\,\beta_6\,MANT_i + \beta_6\,LANT_i + u_i^*$$

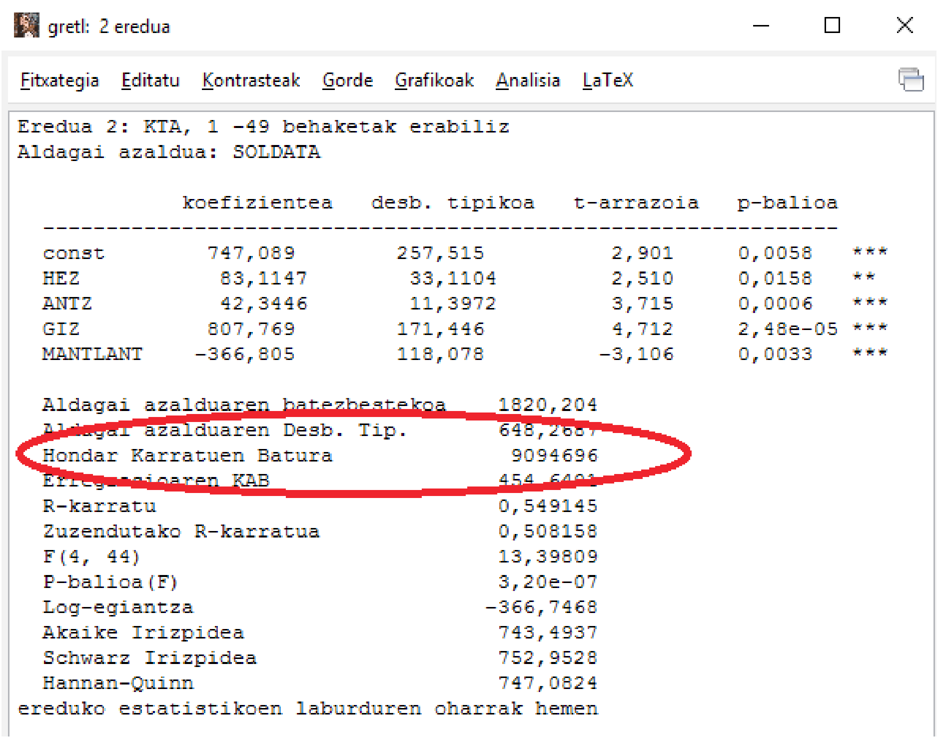
$$SOLDATA_i = \beta_1 + \beta_2\,HEZ_i + \beta_3\,ANTZ_i + \beta_4\,GIZ_i + \beta_6\,(2\,MANT_i + LANT_i) + u_i^*$$

Gretl-ekin murriztutako eredu hori zenbatetsi behar denez, aldagai bat sortu behar da, $(2\,MANT_i + LANT_i)$ alegia. Horretarako, *Gehitu → Definitu aldagai berria* klikatuko da, eta, ondoren, sortu nahi den aldagaiaren formula jarriko da, betiere, aldagai berriaren izena jarri behar dela kontuan izanik (adibidean, MANTLANT izena jarriko da).










```
Eredua 2: KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: SOLDATA

              koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ---------------------------------------------------------------
  const         747,089        257,515         2,901      0,0058    ***
  HEZ            83,1147        33,1104        2,510      0,0158    **
  ANTZ           42,3446        11,3972        3,715      0,0006    ***
  GIZ           807,769        171,446         4,712      2,48e-05  ***
  MANTLANT     -366,805        118,078        -3,106      0,0033    ***

  Aldagai azalduaren batezbestekoa    1820,204
  Aldagai azalduaren Desb. Tip.       648,268
  Hondar Karratuen Batura              9094696
  Erregresioaren KAB                  454,640
  R-karratu                           0,549145
  Zuzendutako R-karratua              0,508158
  F(4, 44)                            13,39809
  P-balioa(F)                         3,20e-07
  Log-egiantza                       -366,7468
  Akaike Irizpidea                    743,4937
  Schwarz Irizpidea                   752,9528
  Hannan-Quinn                        747,0824
ereduko estatistikoen laburduren oharrak hemen
```

$$F = \frac{(HKB_M - HKB_{EM})/_q}{HKB_{EM}/_{(N-K)}} = \frac{(9094696-9090605)/_1}{9090605/_{43}} = 0{,}0193 < \mathcal{F}_{(1,\ 43)|\,0{,}05} = 4{,}06 \text{ denez,}$$

% 5eko esangura-mailarekin hipotesi hutsa ez da baztertzen, eta, beraz, mantenu-lanetan ari den langile baten batez besteko soldata ezaugarri berdinak dituen baina bulegoan lan egiten duen langilearenarekin alderatuta, lantegi edo tailerrean lan egiten duenak izango duen diferentziaren bikoitza dela ondorioztatzen da.

7.4. **Puntuzko eta tartezko iragarpenak**

Ekonometriaren helburu nagusia ereduaren zenbatespen on bat lortzea dela pentsatzen bada ere, sarritan, iragarpen zehatzak lortzea ere oso garrantzitsua da. Eredua zuzenki zehaztu ondoren, parametroak zenbatetsiko dira, eta, kontrasteak eginez, ereduari oniritzia emango zaio ala ez. Emaitza ezezkoa bada, eredua berriro zehaztu beharko da, eta pauso guztiak berregin. Eredua onartuz gero, iragarpenak egiteko erabil daiteke, edota «zein izango litzateke aldagai azalduaren balioa aldagai azaltzaileek balio konkretu bat hartzen badute» motako galderei erantzun ahal izango diegu.

Demagun Erregresio Lineal Orokorraren Eredua:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + u_i \qquad i = 1, 2, \ldots, N$$

Eta dagokion Lagineko Erregresio Funtzioa:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_K X_{Ki} \qquad i = 1,2, \ldots, N$$

Orduan, aldagai azaltzaileen behaketa berri bat izanik —hau da $X_{2p}, X_{3p}, \ldots, X_{Kp}$ behaketak adibidez, non $p \notin (1, 2, \ldots, N)$— honako bektore hau osatuko da:

$$X_p' = [1 \;\; X_{2p} \;\; X_{3p} \ldots X_{Kp}]$$

Lagineko Erregresio Funtzioa erabil daiteke aldagai azalduak izango lukeen balioa aurresateko (puntuzko iragarpena):

$$\hat{Y}_p = \hat{\beta}_1 + \hat{\beta}_2 X_{2p} + \cdots + \hat{\beta}_K X_{Kp}$$

Hala ere, aldagai azalduaren p momentuko edo banakoaren benetako balioa jakitean, errore bat egin dela antzemango da (iragarpenaren errorea: $e_p = Y_p - \hat{Y}_p$), aldagai azaltzaileen balioetan errorea dagoelako, koefizienteen zenbatesleak erabili direlako, Y_p u_p perturbazioaren menpekoa delako (hau da, behaketa horri dagokion perturbazioaren mende), etab. Horregatik, komenigarria izaten da tartezko iragarpen bat egitea, nolabait iragarpenaren zehaztasun neurri bat kontuan hartzen baitu.

Tartezko iragarpena lortzeko, iragarpenaren errorearen banaketan oinarrituko gara, u_p eta $\hat{\beta}$ aldagai aleatorio normalak badira, aurresandako edo iragarpenaren errorea ere normala baita:

$$e_p \sim N(0, \sigma^2(1 + X_p' \, (X'X)^{-1} \, X_p))$$

Orokorrean, σ^2 ezezaguna izaten denez, bere zenbatesle alboragabea erabiliz $\left(\hat{\sigma}^2 = \frac{HKB}{(N-K)}\right)$, honako estatistiko eta banaketa hauek lortuko genituzke:

$$\frac{e_p}{\hat{\sigma}\sqrt{1 + X_p'(X'X)^{-1}X_p}} = \frac{e_p}{\hat{\sigma}_{e_p}} \sim t_{(N-K)}$$

Eta hemendik, puntuzko iragarpenaren inguruan, aldagai azalduak p momentuan hartuko lukeen balioaren iragarpen-tarte bat lortuko da 1-α konfiantza-mailarekin.

$$\Pr\left[-t_{(N-K)\alpha/2} \leq \frac{e_p}{\hat{\sigma}_{e_p}} \leq t_{(N-K)\alpha/2}\right] = 1-\alpha$$

$$\Pr\left[-t_{(N-K)\alpha/2} \leq \frac{Y_p - \hat{Y}_p}{\hat{\sigma}_{e_p}} \leq t_{(N-K)\alpha/2}\right] = 1-\alpha$$

$$\Pr\left[\hat{Y}_p - t_{(N-K)\alpha/2}\,\hat{\sigma}_{e_p} \leq Y_p \leq \hat{Y}_p + t_{(N-K)\alpha/2}\,\hat{\sigma}_{e_p}\right] = 1-\alpha$$

$$\mathrm{KT}(Y_p)_{1-\alpha}\left[\hat{Y}_p - t_{(N-K)\alpha/2}\,\hat{\sigma}_{e_p} \;\; ; \;\; \hat{Y}_p + t_{(N-K)\alpha/2}\,\hat{\sigma}_{e_p}\right]$$

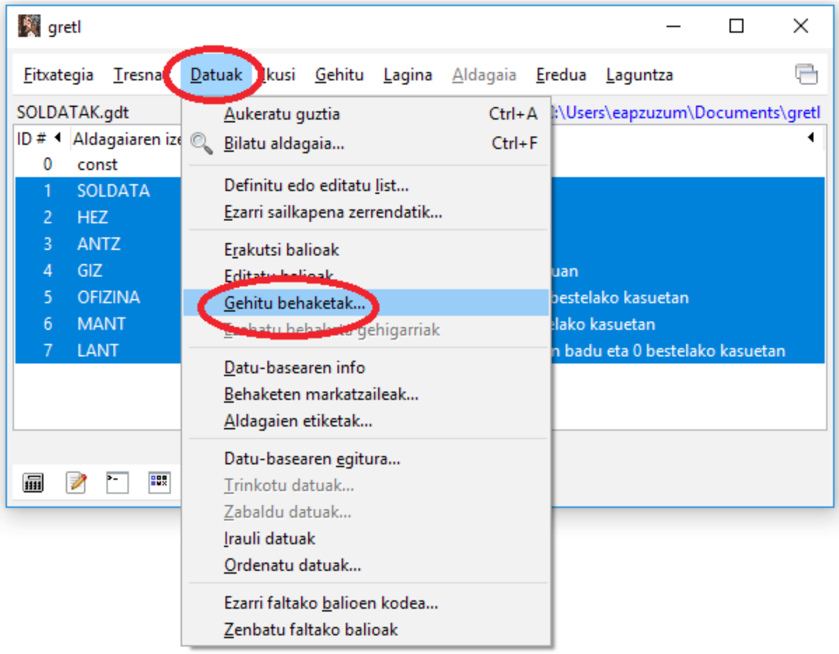
7.5. **Puntuzko eta tartezko iragarpenak Gretl-ekin**

Soldataren adibidea oinarritzat hartuz, Gretl-ekin iragarpenak egiteko jarraitu beharreko pausoak zein diren aztertuko da.

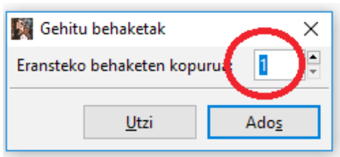
$$SOLDATA_i = \beta_1 + \beta_2\,HEZ_i + \beta_3\,ANTZ_i + \beta_4\,GIZ_i + \beta_5\,MANT_i + \beta_6\,LANT_i + u_i$$

Demagun langile berri baten informazioa lortzen dela; hau da, emakumea da, 15 hezkuntza-urte ditu, eta, enpresako bulegoan sartu berria denez, antzinatasuna 0 du. Langile honi 1.800 € ordainduko omen diote. Lagineko informazioarekin, langile honek izan beharko duen soldataren iragarpena lor daiteke eta, egiaztatu egoki ordainduko zaion ala ez.

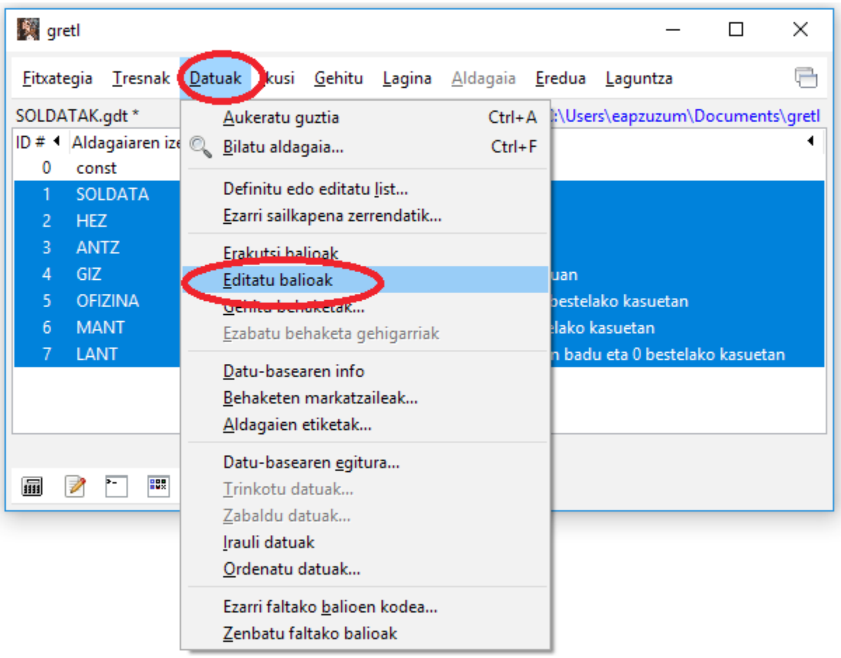
Lehen pausoa datu berria datu-basean barneratzea da. Horretarako, ereduan dauden aldagaiak aukeratuko dira, eta *Datuak* → *Gehitu behaketa* klikatuko da, balioak barneratzen hasteko.

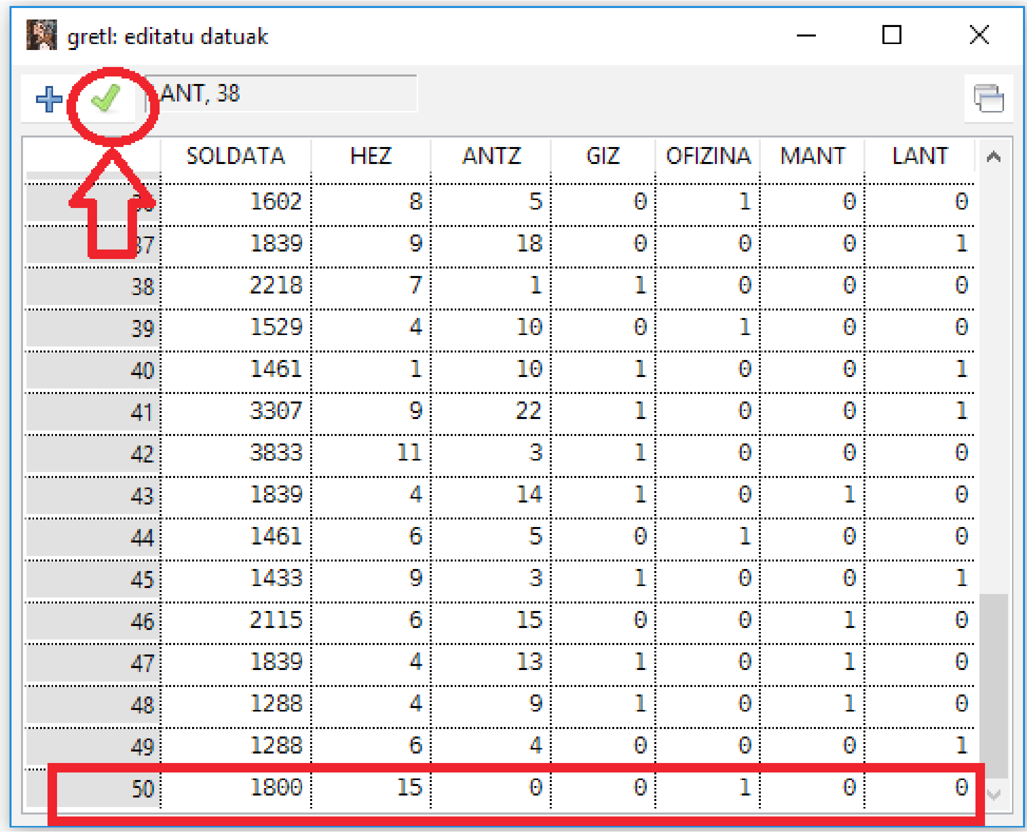


Hori egiteko, barneratu diren behaketa kopurua (adibidean, behaketa bat) zein den adierazi behar da.

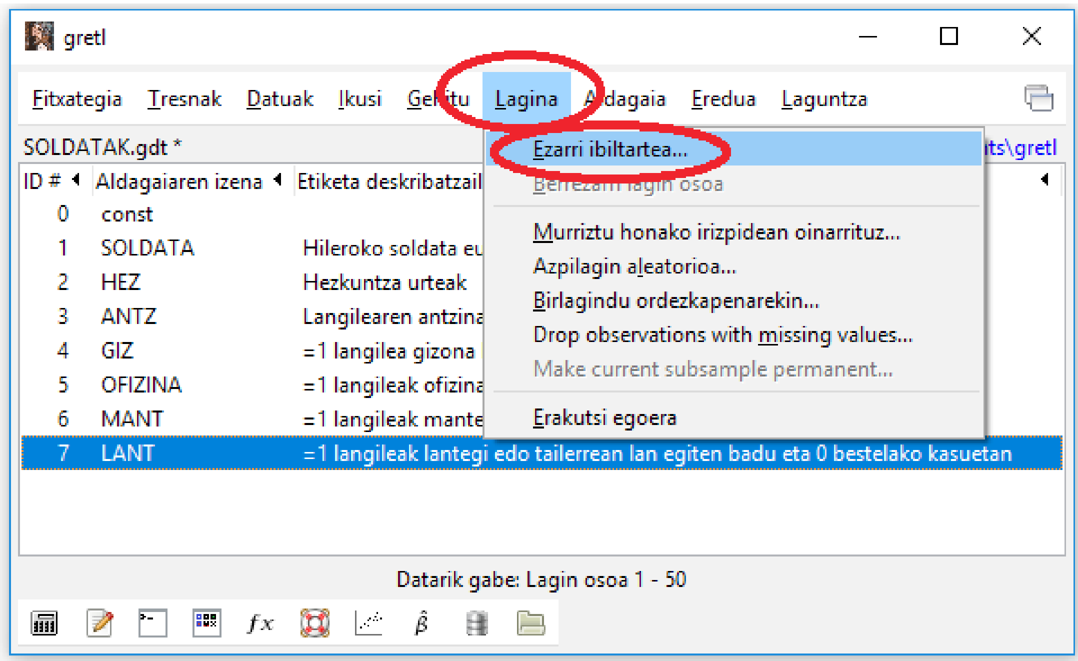


Eta datuak sartu ahal izateko, berriro aldagaiak aukeratu eta *Editatu balioak* klikatuko da.

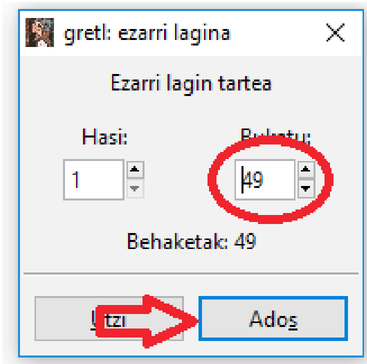




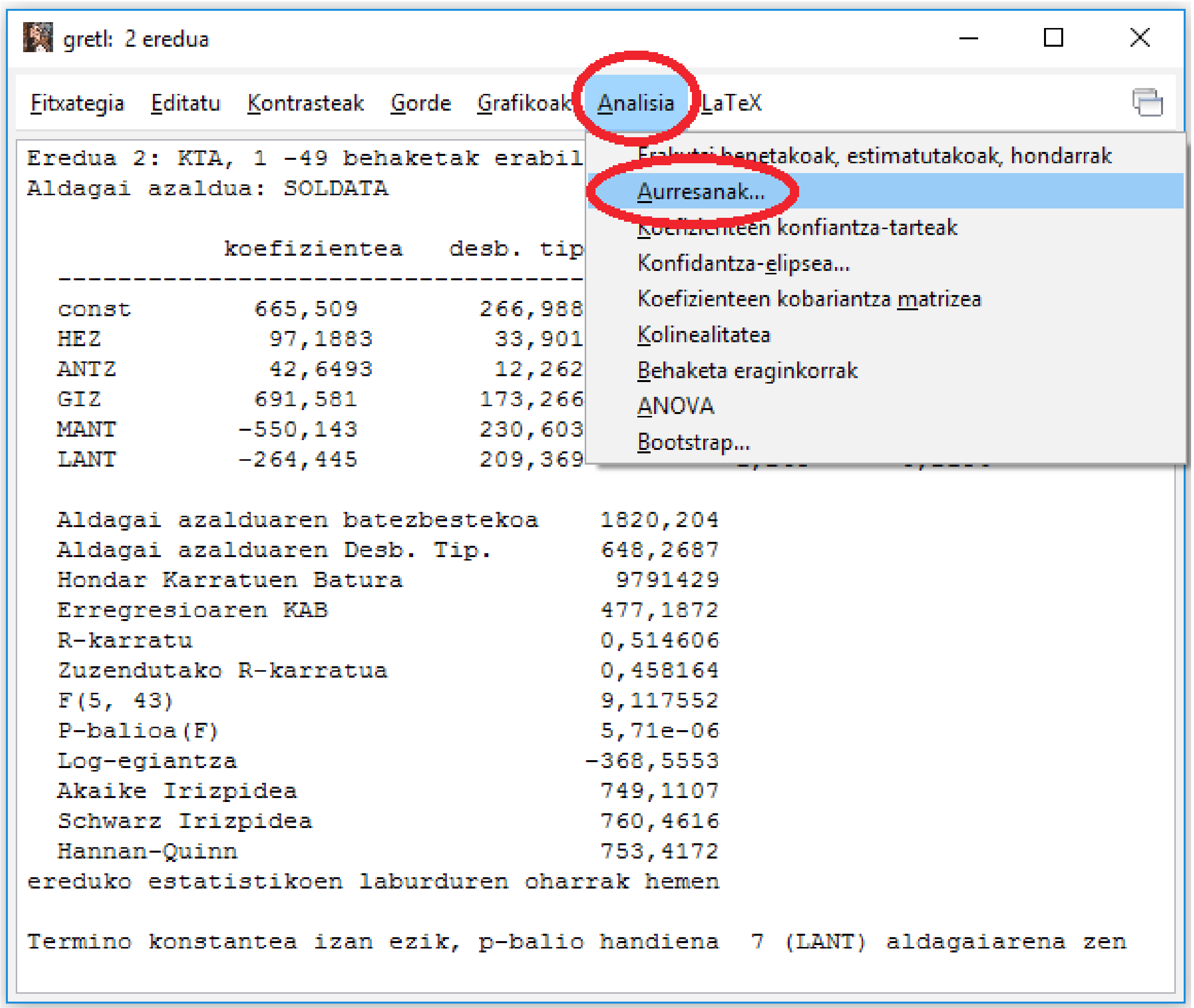
Ondoren, kontu handia izan behar da eredua zenbatesterakoan behaketa berria ez baita kontuan hartu behar (gogoratu, lagina 49 behaketez osaturik dago). Horregatik, *Lagina* → *Ezarri ibiltartea* aukeran



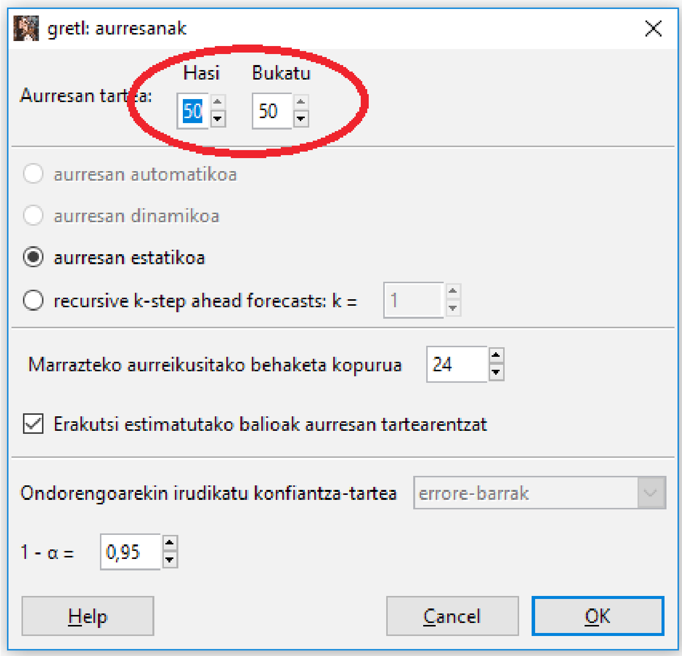
Lagina nola osatzen den adieraziko da (adibidean, 49 behaketa), eta *Ados* klikatuko da:



Lagineko 49 behaketekin eredua KTA bitartez zenbatetsiko da, eta emaitzen leihatilan *Analisia* → *Aurresanak* klikatuko.

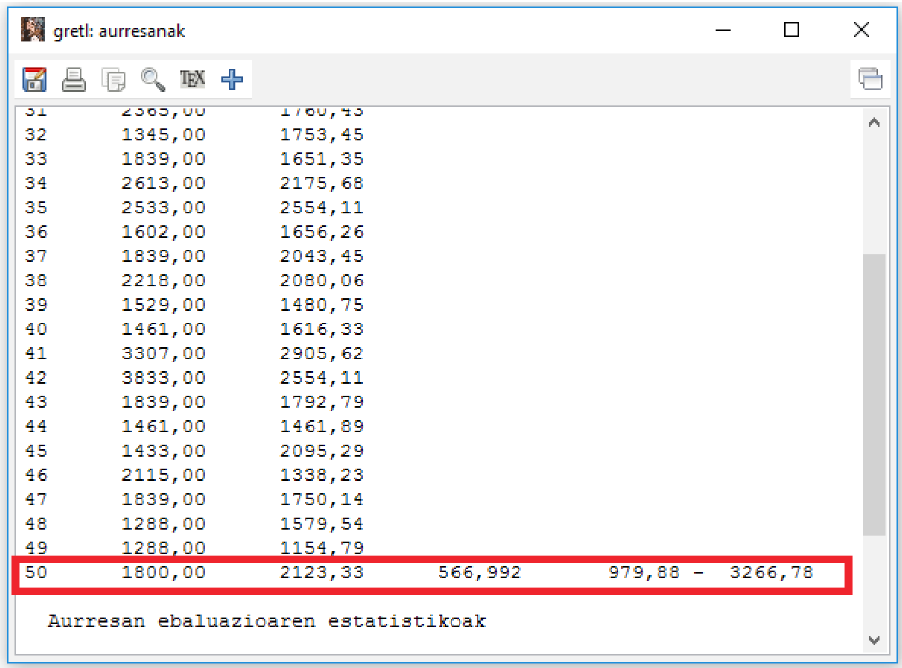


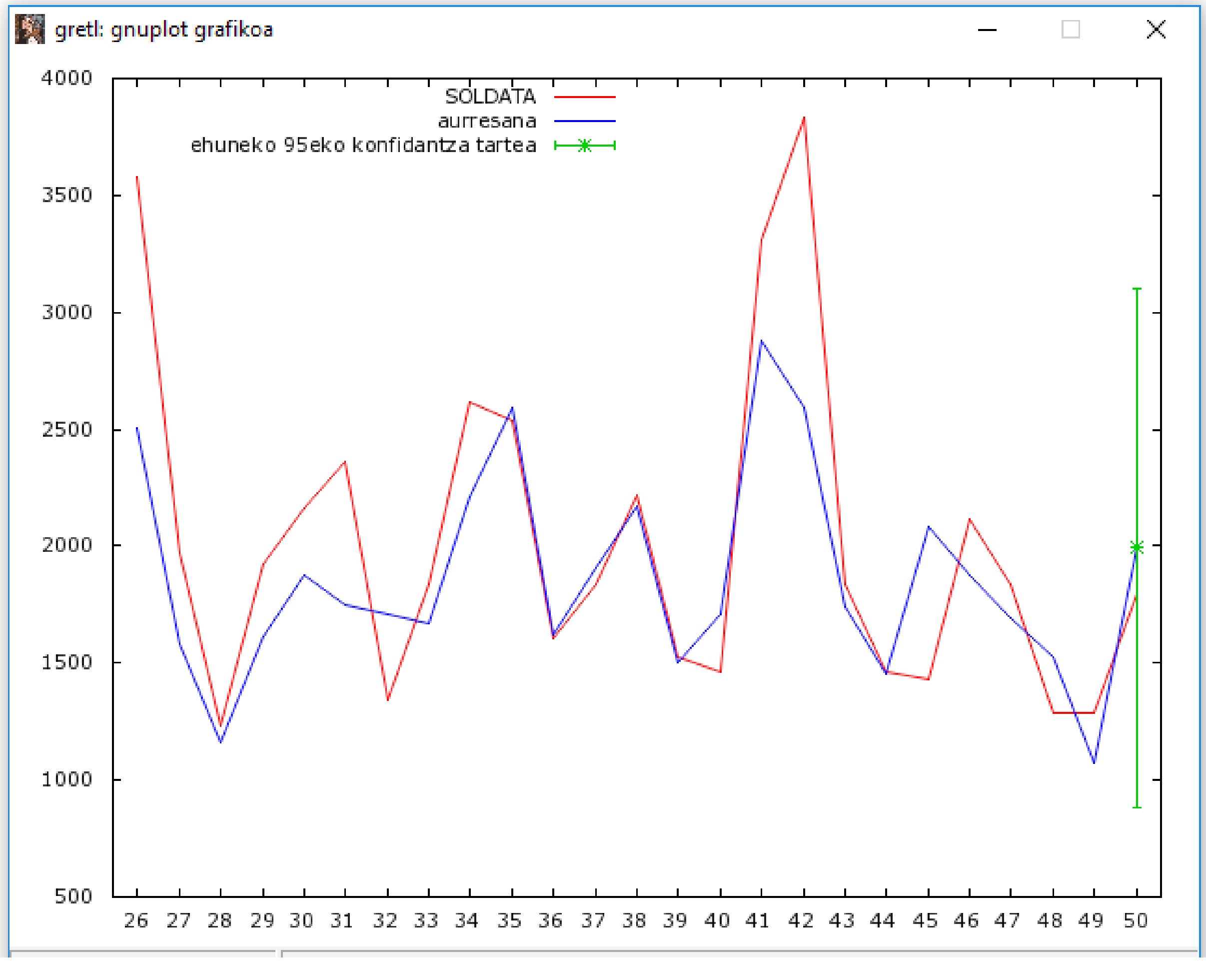
Leihatila berrian 50 behaketa idatziko da, ondoren *Ok* klikatuz.



Lortuko den leihatilan, 50. errenkadan, emakume honen soldataren tartezko iragarpena azalduko da. Adibidean, langile berri horren soldata 881,81 € – 3.103,22 € bitartekoa dela iragarri da, eta, beraz, 1.800 €ko soldata arrazoizkoa da.

Aurreko emaitzarekin batera, irudi bat agertzen da. Bertan, orain arte lortutakoa laburbiltzen da; halaber, interesatzen zaigun aldagaiaren balioaren garapena (gorriz), iragarpenak (urdinez) eta % 95eko konfiantza-mailako tartea (berdez). Askotan, lagungarria izaten da emaitzak horrela ikustea, zenbakiz osatutako taula batean baino hobeto nabari direlako aztergai den aldagaiaren ezaugarri nagusiak.




gretl: gnuplot grafikoa
SOLDATA
aurresana
ehuneko 95eko konfidantza tartea
4000
3500
3000
2500
2000
1500
1000
500
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50

8. gaia

Kolinealitate anizkoitza

8.1. **Sarrera**

Kolinealitate anizkoitza aldagai azaltzaileen arteko lagin-korrelazioa da. Ereduaren zehaztapena emanik, erregresoreetako bat besteen konbinazio lineal zehatz bezala adieraz badaiteke, orduan, kolinealitate anizkoitz zehatza dagoela esango da. Bestalde, posible da, erregresore bat besteen konbinazio zehatza izan gabe, beren arteko erlazioa oso sakona izatea; hau da, kolinealitate anizkoitza altua izatea. Ondorioz, arazoa laginekoa da; hau da, ereduaren zehaztapena zuzena da, baina zenbatesterakoan laginean gertatzen den korrelazioaren arazoa antzemango da.

Jarraian, aipaturiko bi kasuak azalduko dira, eta baita sortzen dituzten arazoak ere.

A) *Kolinealitate anizkoitz zehatza*

Problemaren kasu limitea da, eta aldagai azaltzaileen artean konbinazio lineal zehatza dagoela esan nahi du. Hau da, aldagai baten balioak izanik, beste aldagaiaren balio zehatzak lortzen dira. Orduan, X datu-matrizea osatuz gero, zutabe bat beste zutabe batetik erator daiteke; X datu-matrizea ez da zutabeetan hein osokoa, hau da, $h(X) < K$ izango da, eta, beraz, eredua ezin izango da era bakar batean zenbatetsi.

B) *Kolinealitate anizkoitz altua*

Kasu honetan, aldagai azaltzaileen artean ez dago konbinazio lineal zehatzik, eta, beraz, X datu-matrizea zutabeetan hein osokoa da ($h(X) = K$) eta eredua zenbatesgarria izango da. Baina aldagai azaltzaileen arteko korrelazioa altua bada, bere ondorioak izango ditu zenbatespen-emaitzetan eta baita kontrasteetan ere.

Alde batetik, aldagai azaltzaileen korrelazioa zenbat eta handiagoa izan, ereduko koefizienteen zenbatesleen bariantzak ere orduan eta handiagoak izango dira, hau da, zehaztasun gutxiagoko zenbatesleak lortuko dira. Eta, bestetik, bariantza horiek handiagoak direnean, banakako esangura-kontrasteak egitean, adibidez, erabiliko den t estatistikoaren balioa gero eta txikiagoa izango denez, hipotesi hutsa ez baztertzearen probabilitatea handiago da, eta aldagaia ez esanguratsua izateko aukera handiagotu egiten da. Konfiantza-tarteetan ere eragina izango du, noski, bariantza handiagoak tarte zabalagoa izatea ekarriko baitu.

Hala ere, oinarrizko hipotesiak betetzen direnez eta Gauss-Marcov-en teorema aplikagarria denez, lortzen diren koefizienteen zenbatesleak alboragabeak eta bariantza minimokoak izaten jarraituko dute. Hau da, aldagai azaltzaileen arteko erlazioa sakona izan arren eta, hortaz, koefizienteen zenbatesleen bariantzak handiak izan arren, ezin daiteke bariantza txikiagoko beste zenbatesle lineal eta alboragabe bat aurkitu.

Arazoa nola antzeman

Zenbatetsi nahi den ereduko aldagai azaltzaileen arteko gradu altuko kolinealitatearen susmoa dago, baldin eta honako egoera hauek azaltzen badira:

— Datuetan izandako aldaketa txikiek zenbatespenetan aldaketa handiak sorrarazten dituzte.
— Zenbatetsitako koefizienteek desbideratze tipiko handiak izango dituzte, eta, ondorioz, banakako esangura-kontrasteak egitean, aldagaiak ez-nabariak izango dira. Baina, bestalde, mugatze-koefizientea oso altua izaten da, eta, baterako esangura-kontrastea egitean, batera nabariak direla irtengo da.
— Aldagaien arteko korrelazio-koefizientea oso altua izaten da, eta aldagai azaltzaile bakoitzari beste aldagaien arabera erregresio laguntzaile bat zenbatetsiz gero, dagokion mugatze -koefizientea oso handia izaten da.
— Oharra: aldagai desberdinen arteko erlazio linealak aztertzen dira, eta ereduan aldagai bat eta bere berbidura barneratuz gero, ez da kolinealitaterik izango.

Soluzio posibleak

Arazoa konpontzeko proposatzen diren soluzioak asko dira, baina orokorrean ez dira asebetegarriak.

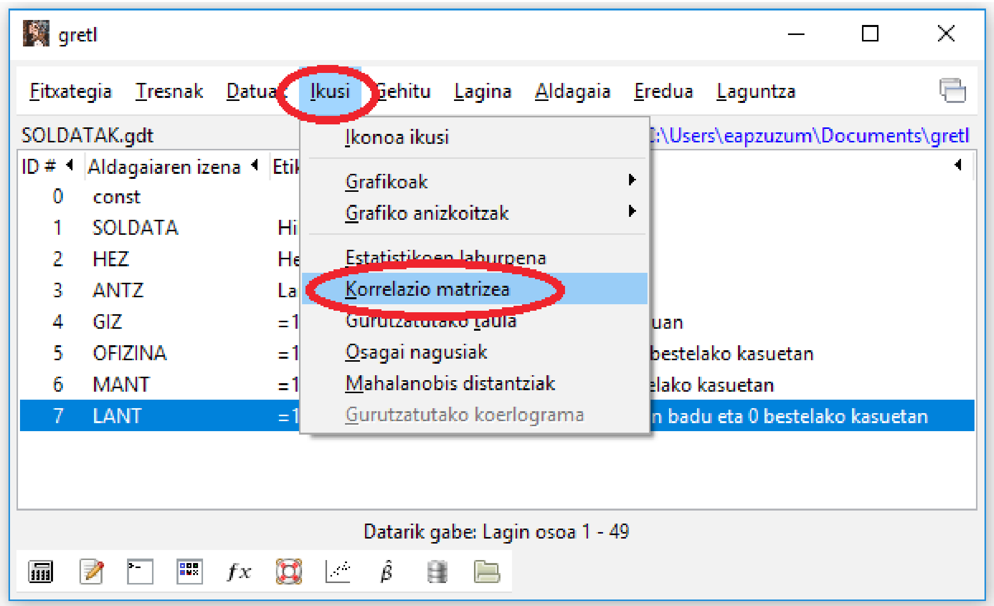
— Datu gehiago lortu edota lagin berri bat lortu, ea $X'X$ matrizearen egitura aldatzen den; hau da, ahalik eta erlazio gutxien duten datu berriak lortu.
— A prioriko murrizketak barneratu. Ereduko koefiziente batek edo batzuek har ditzaketen balioen a prioriko informazioa izanez gero, informazio hori erabil daiteke kolinealitate anizkoitzaren arazoa saihesteko.
— Kolinealitate anizkoitza sorrarazten duten aldagaien ikerketa sakon bat egin eta zenbatespen-prozeduratik kanporatu. Hala ere, ez da soluzio egokia, zehaztapen oker baten arazoa sor baitezake, hau da, aldagai nabarien omisioa.

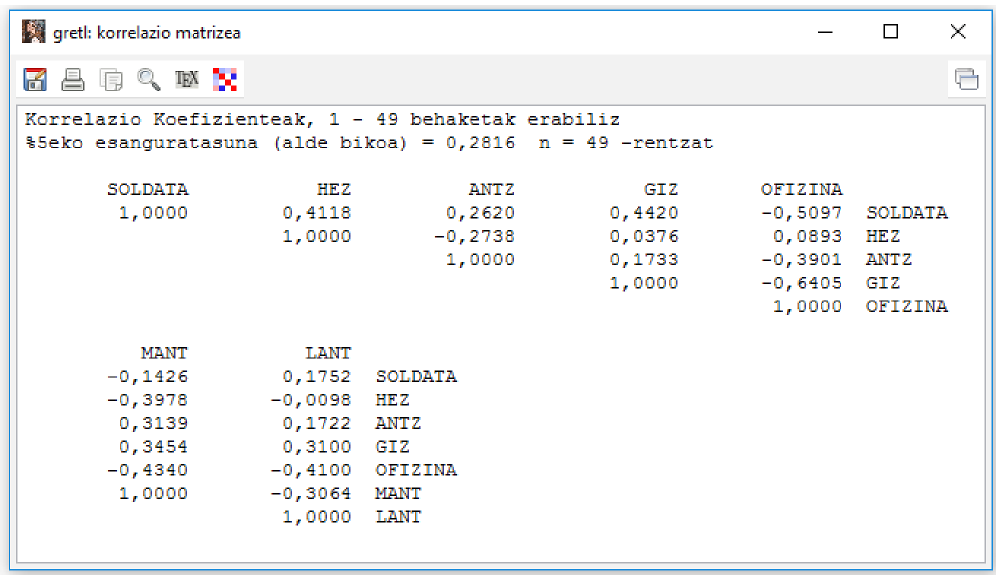
8.2. **Kolinealitate anizkoitza Gretl-ekin**

Soldataren adibidea oinarritzat hartuz, aldagaien artean kolinealitaterik dagoen jakiteko jarraitu beharreko pausoak zein diren aztertuko da Gretl-ekin.

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + u_i$$

Lehen pausoa datu berriak datu-basean barneratzea da; *Ikusi* → *Korrelazio matrizea* klikatuko da, aldagaien arteko korrelazioa nolakoa den aztertzeko, hau da, ea 1 edo –1etik hurbileko korrelaziorik badagoen ikusteko. Leihatila berrian, aldagai guztiak aukeratuko dira, edo aztertu nahi diren aldagai konkretuak.

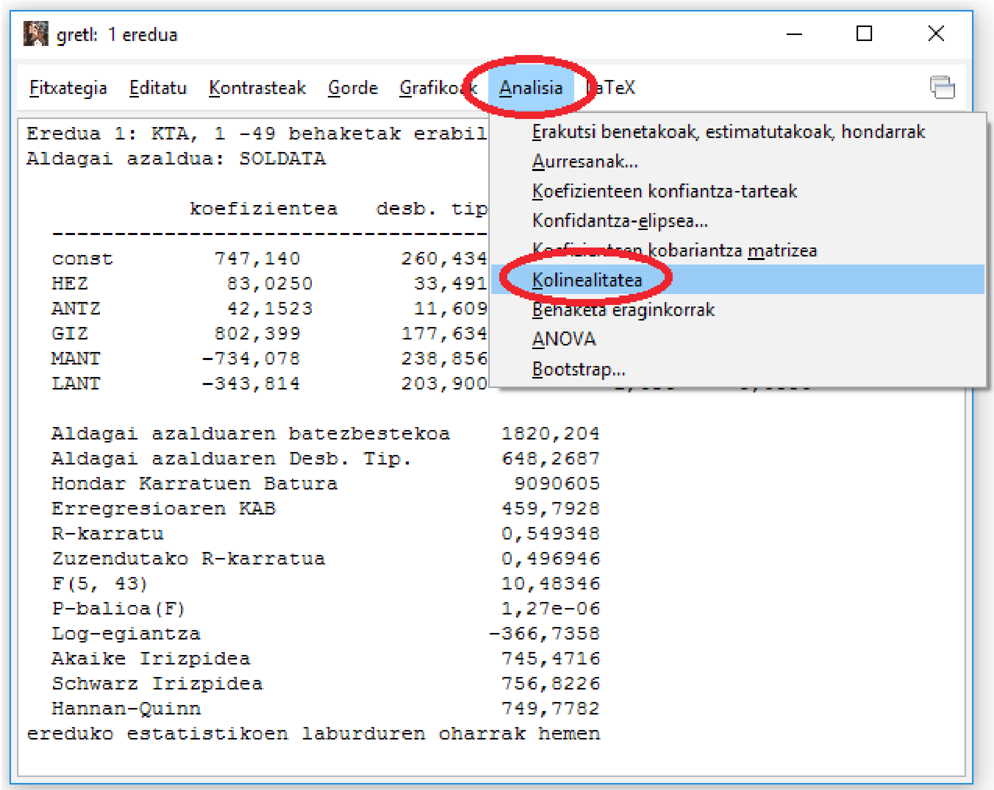


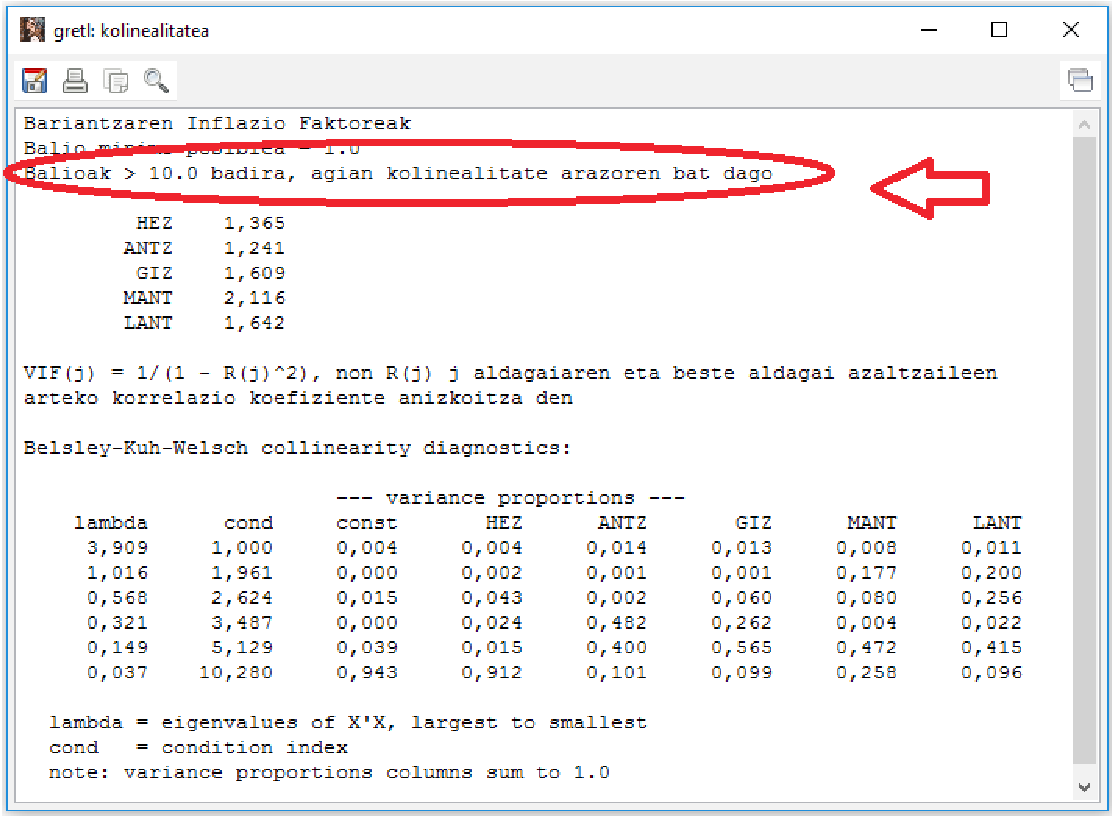


Baina badago beste aukera bat: Bariantzaren Inflazio Faktorea (*VIF*). Hartarako, lehendabizi eredua zenbatetsiko da eta zenbatespen-emaitzako leihatilan a*nalisia* klikatuz, *kolinealitatea* kontrastatzeko aukera dago. Bertan, aldagaien *VIF*ak irtengo dira, eta 10 baino balio handiagoek kolinealitatea dagoela adieraziko dute.

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

Non, R_j^2, j aldagaiaren beste aldagaiekiko erregresioaren mugatze-koefizientea baita.





9. gaia

Karratu Txikien Murriztuen zenbateslea (KTM)

9.1. **Sarrera**

Demagun $Y = X\beta + U$ ereduan : H_0: $R\beta = r$ murrizketaren kontrastea gauzatu dela eta hipotesi hutsa ez baztertzearen ondorioa lortu dela. Interesgarria litzateke eredua berriro zenbatestea, baztertu ez diren murrizketa horiek kontuan hartuta. Matematikoki, honako hau da optimizazio-ariketa:

$$\begin{cases} min_{\hat{\beta}}\ \hat{U}'\hat{U} \\ k.h.\ R\hat{\beta} = r \end{cases}$$

Lortuko den emaitza, hau da, Karratu Txikien Murriztuen zenbateslea (KTM) honako hau da:

$$\hat{\beta}_{KTM} = \hat{\beta}_M = \hat{\beta}_{KTA} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\,(r - R\,\hat{\beta}_{KTA})$$

9.2. $\hat{\boldsymbol{\beta}}_{KTM}$ **zenbateslearen propietateak**

1. Perturbazioarekiko lineala da.
2. Orokorrean alboratua da, baina murrizketa egia denean, hau da, $R\beta = r$ denean, alboragabea da.

$$E_X(\hat{\beta}_{KTM}) = E_X[\hat{\beta}_{KTA} + \underbrace{(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}}_{A}\ (r - R\,\hat{\beta}_{KTA})] =$$

$$= E_X(\hat{\beta}_{KTA}) + A\,(r - R\,E_X(\hat{\beta}_{KTA})) = \beta + A(r - R\beta) = \beta \qquad (R\beta = r\ bada)$$

3. $\hat{\beta}_{KTM}$ zenbatesleak $\hat{\beta}_{KTA}$ zenbatesleak baino bariantza txikiagoak ditu beti, murrizketa egia edo gezurra izan.

$$Bar\left(\hat{\beta}_{KTM}\right) = \sigma^2[(X'X)^{-1} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}] =$$

$$= \sigma^2(X'X)^{-1} - \sigma^2 (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1} =$$

$$= \mathrm{Bar}(\hat{\beta}_{KTA}) - \sigma^2 (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1} < \mathrm{Bar}(\hat{\beta}_{KTA})$$

9.3. $\hat{\boldsymbol{\beta}}_{KTM}$ lortzeko beste aukera bat

Karratu Txikien Murriztuak lortzeko beste aukera bat murriztutako ereduarekin lan egitean datza, hau da, hasierako ereduan murrizketa barneratuz eta lortzen den eredu berria (murriztutako eredua) Karratu Txikien Arrunten bitartez zenbatetsiz.

Adibide batekin azalduko da:

Demagun $Y_1 = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ $i = 1, 2, \ldots, N$ Erregresio Linealaren Eredua eta $\beta_2 + \beta_3 = 1$ murrizketa. Murriztutako eredua lortzeko, murrizketa barneratuko da ereduan:

$$Y_i = \beta_1 + \beta_2 X_{2i} + (1 - \beta_2)X_{3i} + u_i^* = \beta_1 + \beta_2 X_{2i} + X_{3i} + \beta_2 X_{3i} + u_i^*$$

$$\underbrace{Y_i - X_{3i}}_{Y_i^*} = \beta_1 + \beta_2 (X_{2i} + X_{3i}) + u_i^*$$

Jarraian, murriztutako ereduari dagozkion X^* eta Y^* osatuz, ereduko bi koefizienteak ($\hat{\beta}_1$ eta $\hat{\beta}_2$) zenbatetsiko dira:

$$\hat{\beta}^* = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^{*\prime}X^*)^{-1} X^{*\prime}Y^*$$

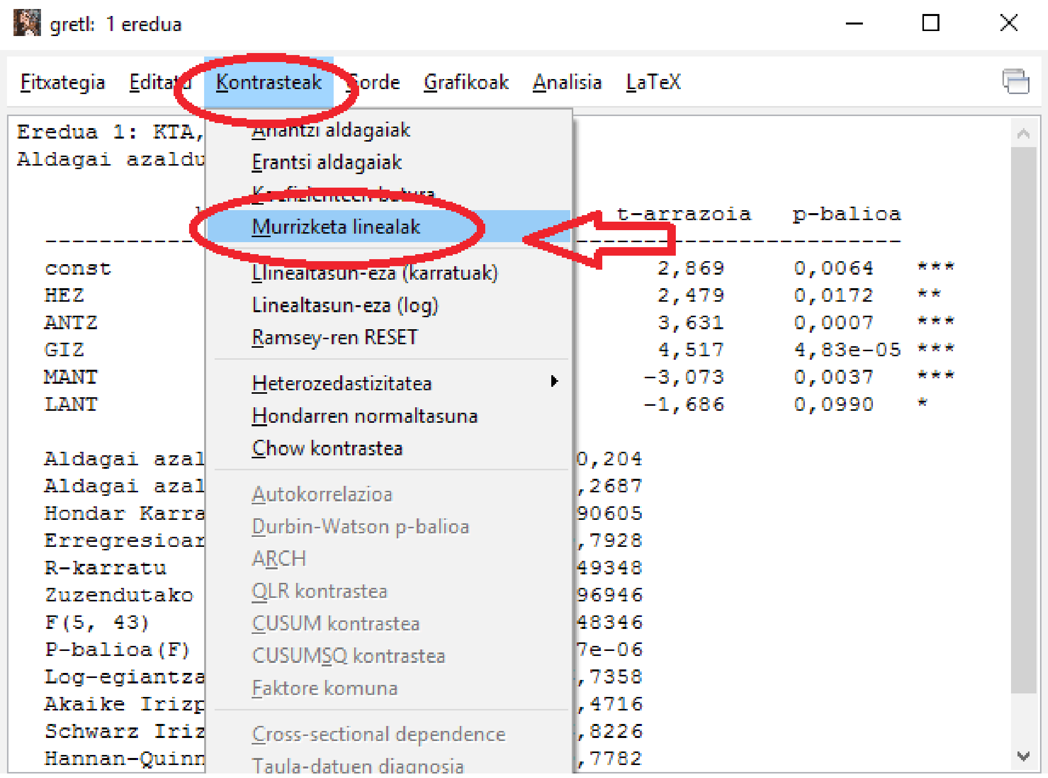
Eta falta den $\hat{\beta}_3$ zenbateslea, murrizketatik lortuko da; hau da, $\beta_2 + \beta_3 = 1$ murrizketatik, horrela, $\hat{\beta}_{KTM}$ bektorea lortuz.

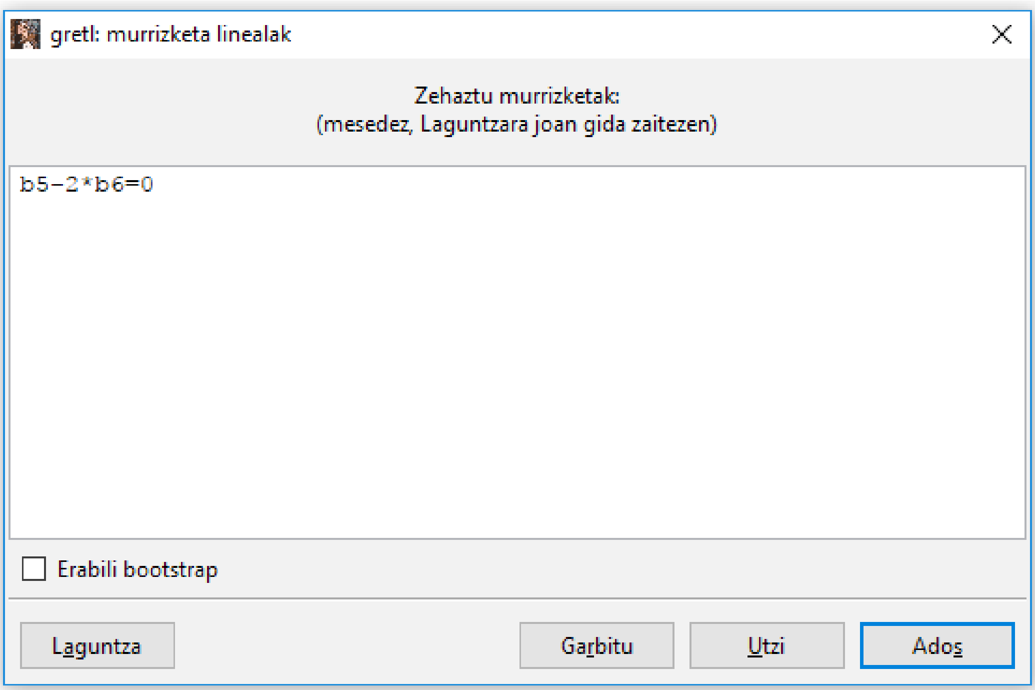
9.4. Karratu Txikien Murriztuen zenbateslea Gretl-ekin

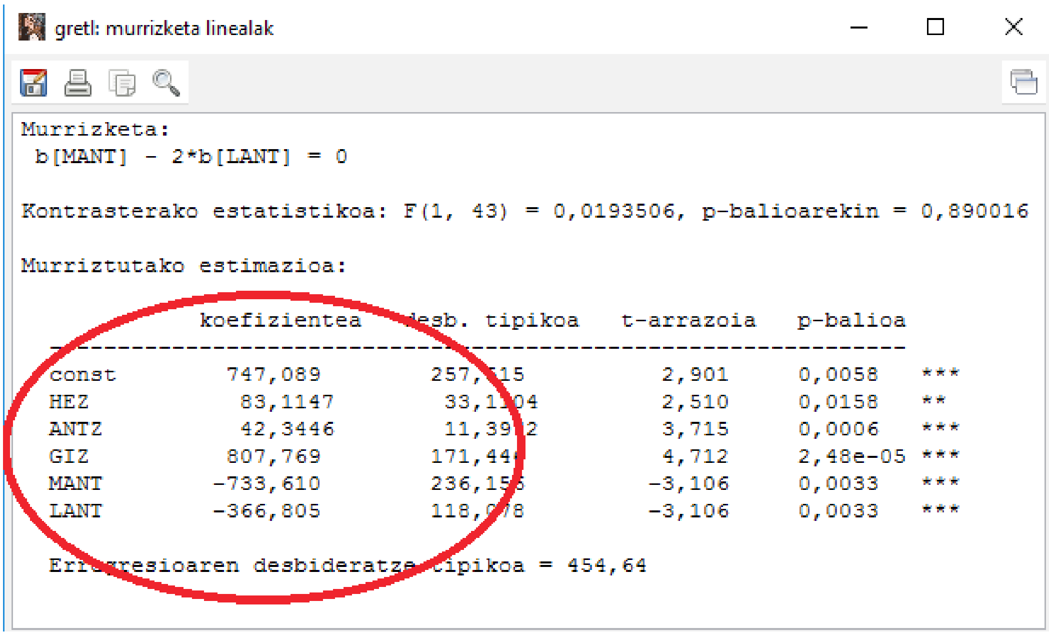
Soldataren adibidea oinarritzat hartuz, Gretl-ekin aldagaien artean *KTM* zenbateslea lortzeko jarraitu beharreko pausoak zein diren aztertuko da, aurreko atal batean azaldutako murrizketarekin; $\beta_5 + 2\,\beta_6$ murrizketarekin, alegia.

$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_5\, MANT_i + \beta_6\, LANT_i + u_i$$

Eredua *KTA* bitartez zenbatestean eta murrizketa kontrastatzean, zuzenean *KTM* zenbateslea lor daiteke, *KTA* zenbatespen-emaitzetako leihatilan murrizketa ezarriz gero, hau da, b5-2*b6=0:


gretl: 1 eredua
Fitxategia Editatu Kontrasteak Gorde Grafikoak Analisia LaTeX
Eredua 1: KTA,
Aldagai azaldu
Arantzi aldagaiak
Erantsi aldagaiak
Murrizketa linealak
Linealtasun-eza (karratuak)
Linealtasun-eza (log)
Ramsey-ren RESET
Heterozedastizitatea
Hondarren normaltasuna
Chow kontrastea
Autokorrelazioa
Durbin-Watson p-balioa
ARCH
QLR kontrastea
CUSUM kontrastea
CUSUMSQ kontrastea
Faktore komuna
Cross-sectional dependence
Taula-datuen diagnosia
t-arrazoia p-balioa
const 2,869 0,0064 ***
HEZ 2,479 0,0172 **
ANTZ 3,631 0,0007 ***
GIZ 4,517 4,83e-05 ***
MANT −3,073 0,0037 ***
LANT −1,686 0,0990 *
Aldagai azal 0,204
Aldagai azal ,2687
Hondar Karra 90605
Erregresioar ,7928
R-karratu 49348
Zuzendutako 96946
F(5, 43) 48346
P-balioa(F) 7e-06
Log-egiantza ,7358
Akaike Irizp ,4716
Schwarz Iriz ,8226
Hannan-Quinn ,7782


gretl: murrizketa linealak
Zehaztu murrizketak:
(mesedez, Laguntzara joan gida zaitezen)
b5-2*b6=0
Erabili bootstrap
Laguntza
Garbitu
Utzi
Ados



```
gretl: murrizketa linealak

Murrizketa:
 b[MANT] - 2*b[LANT] = 0

Kontrasterako estatistikoa: F(1, 43) = 0,0193506, p-balioarekin = 0,890016

Murriztutako estimazioa:

              koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ---------------------------------------------------------------
  const         747,089        257,515          2,901     0,0058   ***
  HEZ            83,1147        33,1104         2,510     0,0158   **
  ANTZ           42,3446        11,3982         3,715     0,0006   ***
  GIZ           807,769        171,446          4,712     2,48e-05 ***
  MANT         -733,610        236,156         -3,106     0,0033   ***
  LANT         -366,805        118,078         -3,106     0,0033   ***

  Errorgresioaren desbideratze tipikoa = 454,64
```

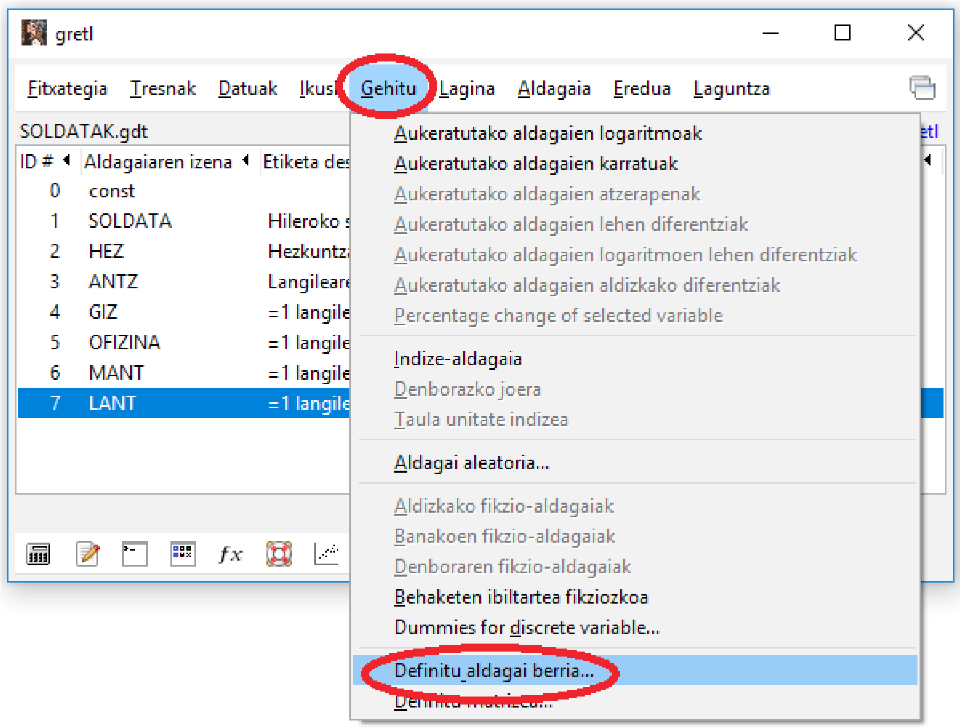
Eta, zuzenean, $\hat{\beta}_{KTM} = \begin{pmatrix} 747{,}089 \\ 83{,}114 \\ 42{,}344 \\ 807{,}769 \\ -733{,}610 \\ -366{,}805 \end{pmatrix}$ lortzen da.

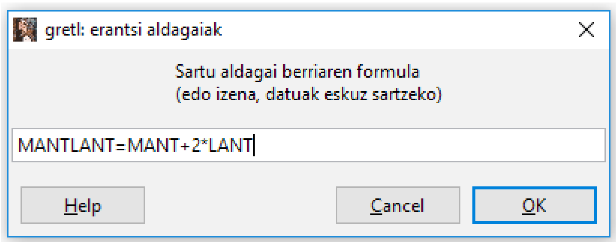
Aurretik aipatu bezala, murriztutako eredua lortzeko murrizketa ereduan barneratuko da, eta elementuak ordenatu, ezagutzen den guztia berdintzaren ezkerraldera pasatuz (kasu honetan ez).

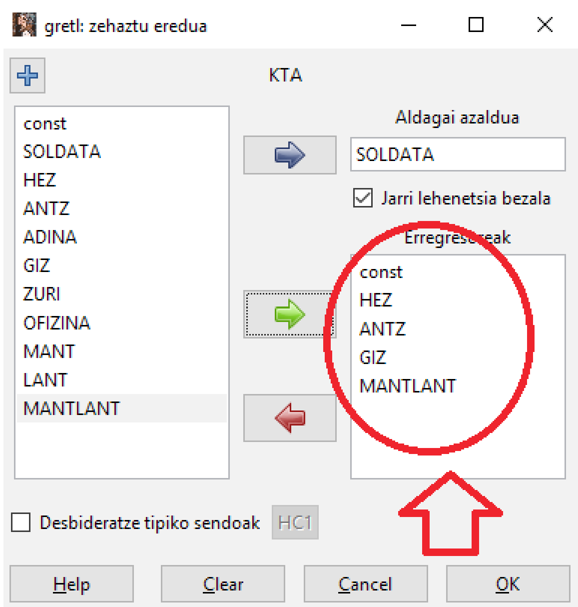
$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + 2\,\beta_6\, MANT_i + \beta_6\, LANT_i + u_i^*$$

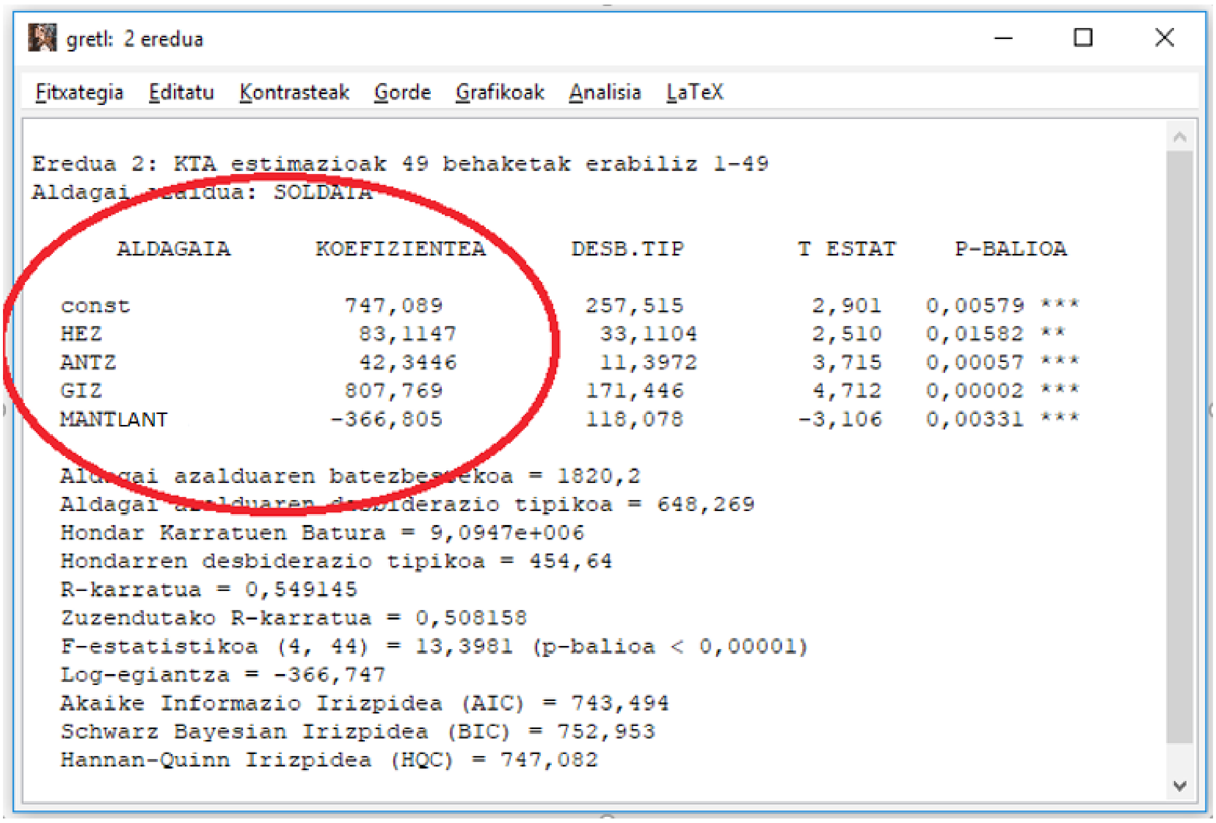
$$SOLDATA_i = \beta_1 + \beta_2\, HEZ_i + \beta_3\, ANTZ_i + \beta_4\, GIZ_i + \beta_6\,(2\, MANT_i + LANT_i) + u_i^*$$

Gretl-ekin murriztutako eredu hau zenbatetsi behar denez, aldagai berri bat sortu behar da, ($2\, MANT_i + LANT_i$) alegia. Horretarako, *Gehitu* → *Definitu aldagai berria* klikatuko da, eta, ondoren, sortu nahi den aldagaiaren formula jarriko da, betiere, aldagai berriaren izena jarri behar dela kontuan izanik (adibidean, MANTLANT izena jarriko da).


gretl
Fitxategia
Tresnak
Datuak
Ikus
Gehitu
Lagina
Aldagaia
Eredua
Laguntza
SOLDATAK.gdt
ID #
Aldagaiaren izena
Etiketa des
0 const
1 SOLDATA Hileroko s
2 HEZ Hezkuntz
3 ANTZ Langileare
4 GIZ =1 langile
5 OFIZINA =1 langile
6 MANT =1 langile
7 LANT =1 langile
Aukeratutako aldagaien logaritmoak
Aukeratutako aldagaien karratuak
Aukeratutako aldagaien atzerapenak
Aukeratutako aldagaien lehen diferentziak
Aukeratutako aldagaien logaritmoen lehen diferentziak
Aukeratutako aldagaien aldizkako diferentziak
Percentage change of selected variable
Indize-aldagaia
Denborazko joera
Taula unitate indizea
Aldagai aleatoria...
Aldizkako fikzio-aldagaiak
Banakoen fikzio-aldagaiak
Denboraren fikzio-aldagaiak
Behaketen ibiltartea fikziozkoa
Dummies for discrete variable...
Definitu aldagai berria...


gretl: erantsi aldagaiak
Sartu aldagai berriaren formula
(edo izena, datuak eskuz sartzeko)
MANTLANT=MANT+2*LANT
Help
Cancel
OK


gretl: zehaztu eredua
KTA
const
SOLDATA
HEZ
ANTZ
ADINA
GIZ
ZURI
OFIZINA
MANT
LANT
MANTLANT
Aldagai azaldua
SOLDATA
Jarri lehenetsia bezala
Erregresoreak
const
HEZ
ANTZ
GIZ
MANTLANT
Desbideratze tipiko sendoak
HC1
Help
Clear
Cancel
OK




```
Eredua 2: KTA estimazioak 49 behaketak erabiliz 1-49
Aldagai azaldua: SOLDATA

      ALDAGAIA        KOEFIZIENTEA       DESB.TIP         T ESTAT     P-BALIOA

  const                  747,089        257,515          2,901    0,00579 ***
  HEZ                     83,1147        33,1104         2,510    0,01582 **
  ANTZ                    42,3446        11,3972         3,715    0,00057 ***
  GIZ                    807,769        171,446          4,712    0,00002 ***
  MANTLANT              -366,805        118,078         -3,106    0,00331 ***

  Aldagai azalduaren batezbestekoa = 1820,2
  Aldagai azalduaren desbiderazio tipikoa = 648,269
  Hondar Karratuen Batura = 9,0947e+006
  Hondarren desbiderazio tipikoa = 454,64
  R-karratua = 0,549145
  Zuzendutako R-karratua = 0,508158
  F-estatistikoa (4, 44) = 13,3981 (p-balioa < 0,00001)
  Log-egiantza = -366,747
  Akaike Informazio Irizpidea (AIC) = 743,494
  Schwarz Bayesian Irizpidea (BIC) = 752,953
  Hannan-Quinn Irizpidea (HQC) = 747,082
```

$$\hat{\beta}^* = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_6 \end{pmatrix} = \begin{pmatrix} 747{,}089 \\ 83{,}114 \\ 42{,}344 \\ 807{,}769 \\ -366{,}805 \end{pmatrix} \rightarrow \hat{\beta}_5 = 2\,\hat{\beta}_6 = -733{,}610 \rightarrow \hat{\beta}_{KTM} = \begin{pmatrix} 747{,}089 \\ 83{,}114 \\ 42{,}344 \\ 807{,}769 \\ -733{,}610 \\ -366{,}805 \end{pmatrix}.$$

$$HKB_M = 9{,}0947 \quad 10^6$$

10. gaia

Zehaztapen-errorea

10.1. **Sarrera**

Ereduan zehaztapen-errorea dagoela esango da, ereduan barneratutako aldagai azaltzailetzat barneratutako aldagai multzoa zuzena ez denean, bai aldagaiak falta direlako (aldagai nabarien omisioa) bai behar direnak baino gehiago daudelako (aldagai ez-nabarien barneratzea).

10.2. **Aldagai nabarien omisioa**

Demagun Erregresio Lineal Orokorraren Eredu bat non oinarrizko hipotesiak betetzen baitira. Ereduko X datu-matrizea $N \times x$ ordenakoa da, baina pentsa dezagun zehaztapen-errore bat egin dela eta azken $s = K - r$ aldagai omititu direla. Barneratu diren aldagaiak A matrizean jasoko dira, eta omititu direnak, berriz, B matrizean. Hortaz, A matrizea $N \times r$ ordenakoa izango da, eta B matrizea $N \times s = N \times (K - r)$.

$$X = [A_{Nxr} \mid B_{Nxs}] = [1 \ X_2 \ \dots \ X_r \ | X_{r+1} \ X_{r+2} \dots \ X_K]$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \\ \beta_{r+1} \\ \beta_{r+2} \\ \vdots \\ \beta_K \end{pmatrix} = \begin{pmatrix} \beta_A \\ \beta_B \end{pmatrix} \rightarrow \beta_A = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{pmatrix} \qquad \beta_B = \begin{pmatrix} \beta_{r+1} \\ \beta_{r+2} \\ \vdots \\ \beta_K \end{pmatrix}$$

Zenbatesten den eredua honako hau litzateke: $Y = A\,\beta_A + V$. Bertako r koefizienteak zenbatetsiko dira:

$$\hat{\beta}_A = (A'A)^{-1}A'Y = (A'A)^{-1}A'(X\beta + U) =$$

$$= (A'A)^{-1}A'[A \mid B]\begin{pmatrix}\beta_A \\ \beta_B\end{pmatrix} + (A'A)^{-1}A'U =$$

$$= (A'A)^{-1}A'A\,\beta_A + (A'A)^{-1}A'B\,\beta_B + (A'A)^{-1}A'U =$$

$$= \beta_A + (A'A)^{-1}A'B\,\beta_B + (A'A)^{-1}A'U$$

Jarraian, egiaztatu egingo da alboragabea edo alboratua den:

$$E(\hat{\beta}_A) = E[\beta_A + (A'A)^{-1}A'B\,\beta_B + (A'A)^{-1}A'U] =$$

$$= \beta_A + (A'A)^{-1}A'B\,\beta_B + (A'A)^{-1}A'E(U) = \beta_A + (A'A)^{-1}A'B\,\beta_B \neq \beta_A$$

Orokorrean, beraz, aldagai nabariak omitituz gero, zenbatesten den ereduko koefizienteen zenbatesleak alboratuak dira, non dagokion alboratzea $E(\beta_A) - \beta_A = (A'A)^{-1}A'B\,\beta_B$ baita. Aldagai nabariak omitituz gero eta barneratutako aldagaiak eta omititutakoak ortogonalak badira, hau da, $A'B = 0$ bada, orduan alboragabea izango da.

Bestalde, perturbazioaren bariantzaren zenbatesleari dagokionez, hau da, $\sigma^2 = \frac{\hat{V}'\hat{V}}{N-r}$, froga daiteke zenbatesle alboratua dela.

$$\hat{V}'\hat{V} = (Y - \hat{Y})'(Y - \hat{Y}) = (Y - A\hat{\beta}_A)'(Y - A\hat{\beta}_A) =$$

$$= (Y - A\,(A'A)^{-1}A'Y)'(Y - A\,(A'A)^{-1}A'Y) =$$

$$= \big((I - A\,(A'A)^{-1}A')Y\big)'\big((I - A\,(A'A)^{-1}A')Y\big) =$$

$$= (M_AY)'(M_AY) = Y'M_AY = ([A \mid B]\begin{pmatrix}\beta_A \\ \beta_B\end{pmatrix} + U)'M_A\,([A \mid B]\begin{pmatrix}\beta_A \\ \beta_B\end{pmatrix} + U) =$$

$$= \beta_A'A'\,M_AA\,\beta_A + \beta_A'A'M_AB\,\beta_B + \beta_A'A'M_AU + \beta_B'B'M_AA\,\beta_A + \beta_B'B'M_AB\,\beta_B$$
$$+ \beta_B'B'M_AU + U'M_AA\,\beta_A + U'M_AB\,\beta_B + U'M_AU =$$

$$\underbrace{=}_{M_AA = A'M_A = 0} \beta_B'B'M_AB\,\beta_B + \beta_B'B'M_AU + U'M_AB\,\beta_B + U'M_AU$$

$$E(\hat{V}'\hat{V}) = E(\beta_B' B' M_A B\,\beta_B + \beta_B' B' M_A U + U' M_A B\,\beta_B + U' M_A U) =$$

$$= \beta_B' B' M_A B\,\beta_B + \beta_B' B' M_A E(U) + E(U') M_A B\,\beta_B + E(U' M_A U) =$$

$$= \beta_B' B' M_A B\,\beta_B + E(U' M_A U) = \beta_B' B' M_A B\,\beta_B + E[tr(U' M_A U)] =$$

$$= \beta_B' B' M_A B\,\beta_B + E[tr(M_A U U')] = \beta_B' B' M_A B\,\beta_B + tr[E(M_A U U')] =$$

$$= \beta_B' B' M_A B\,\beta_B + tr[\,M_A E(UU')] = \beta_B' B' M_A B\,\beta_B + tr\left[\,M_A\,\sigma^2 I\right] =$$

$$= \beta_B' B' M_A B\,\beta_B + \sigma^2\ \ tr[\,M_A\,] = \beta_B' B' M_A B\,\beta_B + \sigma^2\ \ (N - r)$$

$$E(\hat{\sigma}^2) = E\left(\frac{\hat{V}'\hat{V}}{N - r}\right) = \frac{\beta_B' B' M_A B\,\beta_B}{N - r} + \sigma^2 \neq \sigma^2$$

Gainera, $A'B = 0$ denean, hau da, barneratutako aldagaiak eta omititutakoak ortogonalak direnean, nahiz eta $\hat{\beta}_A$ alboragabea izan, perturbazioaren bariantzaren zenbatesleak alboratua izaten jarraituko du, eta, gainera, alboratzea handiagoa izango da.

Alboratzea:

$$E(\hat{\sigma}^2) - \sigma^2 = \frac{\beta_B' B' M_A B\,\beta_B}{N - r} = \frac{1}{(N - r)}\,\beta_B' B'[I - A(A'A)^{-1}A']B\,\beta_B =$$

$$\frac{1}{(N - r)}\left[\beta_B' B' B\,\beta_B - \beta_B' B' A(A'A)^{-1} A' B\,\beta_B\right]$$

Ondorioz, aldagai nabariak omititzen direnean, ereduan barneratu diren aldagaien koefizienteen zenbatesleak alboratuak dira, omititu direnak eta barneratu direnak ortogonalak ez direnean. Baina, perturbazioaren bariantzaren zenbateslea beti alboratua izango da, eta, hortaz, lortuko genituzkeen konfiantza-tarteak eta inferentzia ez dira baliagarriak.

Aldagai nabarien omisioa beste ikuspuntu batetik

Aldagai nabariak omititzen badira, dagozkien koefizienteak 0 direla, murrizketak barneratzen dituen murriztutako eredua zenbatesten arituko ginateke, non murrizketak gezurrak baitira. Hortaz, $\hat{\beta}_{KTM}$ lortuko da.

$$\hat{\beta}_{KTM} = \hat{\beta}_M = \hat{\beta}_{KTA} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\,\hat{\beta}_{KTA})$$

Eta aurreko gaian frogatu bezala, alboratua da murrizketa gezurra denean, hau da, $R\beta \neq r$ denean.

10.3. **Aldagai ez-nabarien barneratzea**

Demagun Erregresio Lineal Orokorraren Eredu bat non oinarrizko hipotesiak betetzen baitira: $Y = X\beta + U$. Ereduko X datu-matrizea $N \times K$ ordenakoa da; baina demagun zehaztapen-errore bat egin dela eta beste s aldagai barneratzen direla. Barneratu diren aldagai ez-nabariak A matrizean jasoko dira. Hortaz, A matrizea $N \times s$ ordenakoa izango da.

Zenbatesten den eredua $Y = Z\,\beta_Z + W$ izango da, non $Z = [X_{N\times K} \mid A_{N\times s}]$ baita, hau da $N \times (K + s)$ ordenakoa. Aldi berean:

$$\beta_Z = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \\ \beta_{K+1} \\ \beta_{K+2} \\ \vdots \\ \beta_{K+s} \end{pmatrix} = \begin{pmatrix} \beta \\ \beta_A \end{pmatrix} \to \beta \quad = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} \qquad \beta_A = \begin{pmatrix} \beta_{K+1} \\ \beta_{K+2} \\ \vdots \\ \beta_{K+s} \end{pmatrix}$$

$Y = Z\,\beta_Z + W$ ereduko $K + s$ koefizienteak zenbatetsiko dira, Hondar Karratuen Batura minimo eginez, Ekuazio Normalen Sistema lortuz, eta hauek askatuz, hau da:

$$min_{\hat{\beta}_Z}\, \widehat{W}'\widehat{W} \;\; \to \;\; Z'Z\,\hat{\beta}_Z = Z'Y$$

$$\begin{pmatrix} X' \\ A' \end{pmatrix} \begin{pmatrix} X & A \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\beta}_A \end{pmatrix} = \begin{pmatrix} X' \\ A' \end{pmatrix} Y$$

$$\begin{pmatrix} X'X & X'A \\ A'X & A'A \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\beta}_A \end{pmatrix} = \begin{pmatrix} X'\,Y \\ A'\,Y \end{pmatrix}$$

Ekuazio Normalak → $\begin{cases} X'X\hat{\beta} + X'A\,\hat{\beta}_A = X'Y & (a) \\ A'X\,\hat{\beta} + A'A\,\hat{\beta}_A = A'Y & (b) \end{cases}$

(a)tik → $\hat{\beta} = (X'X)^{-1}\left[X'Y - X'A\,\hat{\beta}_A\right] \qquad (c)$

(c), (b)n sartuz gero → $A'X\,(X'X)^{-1}\left[X'Y - X'A\,\hat{\beta}_A\right] + A'A\,\hat{\beta}_A = A'Y$

$$A'X\,(X'X)^{-1}X'Y - A'X\,(X'X)^{-1}X'A\,\hat{\beta}_A + A'A\,\hat{\beta}_A = A'Y$$

$$A'A\,\hat{\beta}_A - A'X\,(X'X)^{-1}X'A\,\hat{\beta}_A = A'Y - A'X\,(X'X)^{-1}X'Y$$

$$A'[I - X\,(X'X)^{-1}X']A\,\hat{\beta}_A = A'[I - X\,(X'X)^{-1}X']Y$$

$$A'M\,A\,\hat{\beta}_A = A'M\,Y \quad \to \hat{\beta}_A = (A'M\,A)^{-1}A'M\,Y$$

Non M matrizea simetrikoa baita, idenpotentea, heina edotata traza $N - K$ eta $X'M = 0$. Hortaz:

$$\hat{\beta}_A = (A'M\,A)^{-1}A'M\,Y = (A'M\,A)^{-1}A'M\,(X\,\beta + U) =$$

$$= (A'M\,A)^{-1}A'M\,X\,\beta + (A'M\,A)^{-1}A'MU = (A'M\,A)^{-1}A'M\,U \qquad (d)$$

Beraz, aldagai ez-nabariak barneratzean, beren koefizienteak 0 izatea espero da, hau da, $E(\hat{\beta}_A) = 0$.

(c) eta (d)tik honako hau lortzen da:

$$E(\hat{\beta}) = E\left[(X'X)^{-1}X'Y - (X'X)^{-1}X'A\,\hat{\beta}_A\right] =$$

$$= E\left[(X'X)^{-1}X'\,(X\,\beta + U) - (X'X)^{-1}X'A\,\hat{\beta}_A\,\right] =$$

$$= \beta + (X'X)^{-1}X'\underbrace{E(U)}_{0} - (X'X)^{-1}X'A\,\underbrace{E(\hat{\beta}_A)}_{0} = \beta$$

Hau da, aldagai ez-nabariak barneratzean eta *KTA* bitartez zenbatestean koefizienteen zenbatesle alboragabeak lortzen dira.

Perturbazioen bariantzaren zenbateslea ere alboragabea dela froga daiteke, eta, ondorioz, konfiantza-tarteak eta inferentzia baliagarriak dira.

$$\hat{\sigma}^2 = \frac{\widehat{W}'\widehat{W}}{N - K - s}$$

$$\widehat{W} = Y - \hat{Y} = Y - Z\,\hat{\beta}_Z = Y - Z\,(Z'Z)^{-1}Z'Y =$$

$$= [I - Z\,(Z'Z)^{-1}Z']Y = M_Z Y = M_Z(X\beta + U) = M_Z U$$

Non M_Z simetrikoa, idenpotentea, heina edota traza $N - K - s$ eta $M_Z\,Z = 0$, hau da, $M_Z\,[X\,A] = 0$, eta, ondorioz, $M_Z X = 0$ eta $M_Z A = 0$.

$$E\left(\widehat{W}'\widehat{W}\right) = E\left[U'M_Z{}'\,M_Z\,U\right] = E[U'M_Z U] = E[tr(U'M_Z U)] = E[tr(M_Z UU')] =$$

$$= tr[E(M_Z UU')] = tr[M_Z E(U\,U')] = tr\,[M_Z \sigma^2 I] = \sigma^2 tr[M_Z] = \sigma^2(N - K - s)$$

Hortaz,

$$E(\hat{\sigma}^2) = E\left[\frac{\widehat{W}'\widehat{W}}{N - K - s}\right] = \frac{\sigma^2(N - K - s)}{(N - K - s)} = \sigma^2$$

Hala ere, aldagai ez-nabariak barneratzean aldagai nabarien koefizienteen zenbatesleen zehaztasuna murriztu egiten da. Hau da, beren bariantzak handiagoak izango dira (salbuespena aldagai nabariak eta ez-nabariak ortogonalak badira ($X'A = 0$)).

$$Bar(\hat{\beta}_Z) = E\left[\left(\hat{\beta}_Z - E(\hat{\beta}_Z)\right) - \left(\hat{\beta}_Z - E(\hat{\beta}_Z)\right)'\right] = \sigma^2\,(Z'Z)^{-1} =$$

$$= \sigma^2 \begin{bmatrix} X'X & X'A \\ A'X & A'A \end{bmatrix}^{-1} = \begin{bmatrix} Bar(\hat{\beta}) & Kob(\hat{\beta},\hat{\beta}_A) \\ Kob(\hat{\beta},\hat{\beta}_A) & Bar(\hat{\beta}_A) \end{bmatrix}$$

$$Bar(\hat{\beta}) = \sigma^2\,[X'X - X'A\,(A'A)^{-1}A'X]^{-1} > \sigma^2(X'X)^{-1}$$

11. gaia

Heterozedastizitatea

11.1. **Sarrera**

Gai honetan Erregresio Lineal Orokorraren Ereduko oinarrizko hipotesi bat erlaxatuko da, konkretuki perturbazioen bariantza konstantea. Beraz, $Bar(u_i) = \sigma_i^2 \neq \sigma^2$ izango da, eta gainontzeko oinarrizko hipotesiak beteko dira.

Ondorioz, perturbazioa orain:

— $E(u_i) = 0 \quad \forall i = 1, 2, ..., N$
— $Bar(u_i) = \sigma_i^2 \neq \sigma^2$
— $Kob(u_i u_j) = 0 \quad \forall i \neq j$
— $u_i \sim N(0, \sigma_i^2)$

Perturbazioaren bariantza-kobariantza matrizeak diagonala izaten jarraituko du, baina jada ez du konstante bat izango diagonalean.

$$E(UU') = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2^2 & 0 & & 0 & 0 \\ 0 & 0 & \ddots & \ddots & & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & 0 & & \sigma_{N-1}^2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma_N^2 \end{bmatrix} = \Sigma \neq \sigma^2 \ I$$

Hortaz, $Y = X\beta + U$ eredua zenbatetsiko da, non $U \sim N(0, \Sigma)$.

11.2. **Heterozedastizitatearen egoera posibleak**

Hona hemen heterozedastizitatea ager daitekeen zenbait adibide:

1. Kasua. Sekzio gurutzatuko datuekin lan egitea. Orokorrean, horrelako datuekin lan egitean, perturbazioak heterozedastikoak izan ohi dira, banako bakoitzaren portaera oso des-

berdina izan baitaiteke, eta horrek perturbazioaren bariantza aldakorra izatea ekartzen baitu.

2. Kasua. Taldekatutako datuekin lan egitea. Izan bedi Erregresio Lineal Bakunaren Eredu bat $Y_i = \alpha + \beta X_i + u_i \;\; i = 1, 2, ..., N$ non oinarrizko hipotesiak betetzen baitira, eta, beraz, baita perturbazioenak ere, hau da, $U \sim N(0, \sigma^2 I)$. Demagun N behaketak ez daudela eskuragarri, eta bai, ordea, G1, G2,…, Gm multzotan taldekatutako m datu. Talde bakoitzean N_j behaketa izanik, $N = \Sigma_{j=1}^{m} N_j$. Zenbatets daitekeen eredua honako hau da:

   $$Y_j = \alpha + \beta X_j + \upsilon_j \qquad\qquad j = 1, 2, ..., m$$

   Non:

   — $E(\upsilon_j) = 0 \qquad \forall j = 1, 2, ..., m$
   — $Bar(\upsilon_j) = N_j\, \sigma^2 \neq \sigma^2 \;\; \rightarrow$ Heterozedastizitatea
   — $Kob(\upsilon_j \upsilon_l) = 0 \quad \forall j \neq l$
   — $\upsilon_j \sim N(0, N_j\, \sigma^2)$

3. Kasua. Akats-ikasketa. Denborarekin jendeak ikasten duen bezala, izango diren akatsak murriztuz doaz, eta, beraz, baita beren aldakortasuna ere. Hau da, ekoizpen-kate batean, adibidez, hasieran ekoitzitako unitateek kostu handiagoa edota denbora gehiago beharko dute geroago egiten direnek baino, praktikatzen den heinean, murriztu egiten baita denbora, batez besteko kostua edota akatsekin produzitutako unitate kopurua.
4. Kasua. Eragin handiko behaketak. Askotan, datu multzoaren jarraibide orokorretik kanpo dauden behaketek *(outlier)* edota erregresio-emaitzetan eragin sakona duten behaketek heterozedastizitatea sorraraz dezakete. Diseinu-arazorik ez badago edota datu-bilketa ondo egin bada, behaketak horiek ez dira datu multzotik kendu behar, nahiz eta emaitzak aldatu.
5. Kasua. Aldagai ez-nabari bat barneratzeak, aldagai nabari baten omisioak edota aukeratutako ereduaren forma funtzional desegokiak heterozedastizitatea sorraraz dezakete.

11.3. **Heterozedastizitatearen hautematea**

Heterozedastizitatea hautemateko aukera desberdinak daude. Alde batetik, grafikoetan oinarritu behar da, eta, ondoren, kontraste formal bat gauzatu. Kontraste formal asko daude, baina, gai honetan, kasu bakarra aztertuko da: White-en kontrastea, alegia.

Metodo grafikoak

Grafikoen bitartez, datuen sakabanatzea konstantea den edo ez azter daiteke. Grafiko desberdinak marraztu daitezke, baina erabilgarrienak honako hauek dira:

— KTA hondarrak behaketekiko edota aldagai azaltzaileekiko.
— KTA hondarrak ber bi, behaketekiko edota aldagai azaltzaileekiko.
— Azaldutako aldagaia edota zenbatetsitakoa behaketekiko edota aldagai azaltzaileekiko.

Kontraste formala: white-en kontrastea

$$\begin{cases} H_0:\ \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_N^2 \ \rightarrow\ Homozedastizitatea \\ \\ H_a:\ \sigma_1^2 \neq \sigma_2^2 \neq \ldots \neq \sigma_N^2 \ \rightarrow\ Heterozedastizitatea \end{cases}$$

Prozedura:

1. Eredua zenbatesten da KTA bitartez, eta hondarrak ($\hat{u}_i$) lortzen dira.
2. Erregresio laguntzaile bat zehaztatzen da, non aldagai azaldua KTA hondarrak ber bi baitira, eta honako elementu hauen arabera adierazten: termino independentea, ereduko aldagai azaltzaileak, aldagai azaltzaileak ber bi, aldagai azaltzaileen arteko biderkaketa gurutzatuak eta errore bat. Erregresio laguntzaile hori KTA bitartez zenbatesten da, eta dagokion mugatze-koefizientea (R_w^2) lortzen da.
3. Kontrastearen estatistikoa honako hau da:

$$W = N\,R_W^2 \xrightarrow{H_0 \quad b} \chi^2_{(p-1)}$$

   non p, erregresio laguntzaileko koefiziente kopurua baita.
4. Erabakitze-irizpidea: hipotesi hutsa baztertzen da α esangura-mailarekin baldin eta $W > \chi^2_{(p-1)\alpha}$ bada, eta heterozedastizitatea dagoela ondorioztatzen da.

Kontraste horrek ez du perturbazioen normaltasunik behar, eta, gainera, beste kontraste asko baino orokorragoa da, ez baita heterozedastizitatea sorrarazten duen aldagaia zehaztu behar. Kontraste honen helburua da, azken finean, ereduko aldagai azaltzaileak, beren karratuak eta beren biderkaketa gurutzatuak hondarren karratuak azaltzeko baliagarriak diren aztertzea. Hau da, aldagai azaltzaileen garapenak eta beren bariantza eta kobariantzen eboluzioak esanguratsuak diren hondarren lagineko bariantza zehazterakoan. Erregresio laguntzaileko mugatze-koefizienteak adieraziko digu aukeratutako aldagaiak hondarren sakabanatzea azaltzeko baliagarriak diren ala ez. Bariantza konstanteko perturbazioak izanez gero, mugatze-koefizientea oso txikia izango da, eta aukeratutako aldagai horien aldakuntzek hondar karratuen aldakuntza gutxi azalduko dute.

11.4. Ereduko koefizienteen KTA zenbateslearen eta perturbazioen bariantzaren zenbateslearen propietateak heterozedastizitatepean

Izan bedi Erregresio Lineal Orokorraren Eredua $Y = X\beta + U$, non perturbazioak heterozedastikoak baitira; hau da, $E(UU') = \Sigma \neq \sigma^2 I$.

— Hondar Karratuen Batura minimo eginez, ereduko koefizienteen KTA zenbateslea lortzen da ($\hat{\beta} = (X'X)^{-1} X'Y$), eta dagozkion propietateak honako hauek dira:

1. Perturbazioekiko lineala da, perturbazioen konbinazio lineala baita.

$$\hat{\beta} = \beta + (X'X)^{-1} X'U$$

2. Alboragabea da.

$$E_X(\hat{\beta}) = E_X\,[\beta + (X'X)^{-1}X'U] = \beta + (X'X)^{-1}X'E_X(U) = \beta$$

3. Zenbatesle lineal eta alboragabeen artean ez da bariantza minimoduna.

$$Bar(\hat{\beta}) = E_X\left[\left(\hat{\beta} - E_X(\hat{\beta})\right)\left(\hat{\beta} - E_X(\hat{\beta})\right)'\right] = E_X[\,(X'X)^{-1}\,X'U\,U'\,X\,(X'X)^{-1}] =$$
$$= (X'X)^{-1}X'\,E_X(UU')X\,(X'X)^{-1} = (X'X)^{-1}X'\textstyle\sum X\,(X'X)^{-1} \neq \sigma^2\,(X'X)^{-1}$$

4. Perturbazioen Banaketa Normalpean,

$$\hat{\beta} \sim N\,(\beta\,, (X'X)^{-1}X'\textstyle\sum X\,(X'X)^{-1})$$

— Era berean, froga daiteke perturbazioen bariantzaren ohiko zenbateslea ($\hat{\sigma}^2 = \frac{HKB}{(N-K)}$) alboratua dela.

$$\hat{\sigma}^2 = \frac{HKB}{(N-K)} = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{(N-K)} = \frac{Y'M'M\,Y}{(N-K)} = \frac{Y'M\,Y}{(N-K)} = \frac{U'M\,U}{(N-K)}$$

Non $M = (I - X\,(X'X)^{-1}\,X')$ simetrikoa, idenpotentea, $tr(M) = N - K$ eta $M\,X = 0$ baita. Hortaz:

$$E_X(\hat{\sigma}^2) = \frac{E_X(U'M\,U)}{(N-K)} = \frac{E_X(tr(U'M\,U))}{(N-K)} = \frac{E_X(tr(MUU'))}{(N-K)} = \frac{tr(M\,E_X(UU'))}{(N-K)}$$
$$= \frac{tr(M\textstyle\sum)}{(N-K)} \neq \sigma^2$$

— Ondorioz, kontrasteak egitean edota koefizienteen konfiantza-tarteak lortzean, ezin da $\hat{\sigma}^2$ erabili, ez baitira baliogarriak izango. Ohiko estatistikoekin gauzatutako inferentzia ez da baliagarria izango, $\hat{\sigma}^2$ eta $Bar(\hat{\beta}) = \sigma^2(X'X)^{-1}$ matrizean oinarritzen baitira.

11.5. **Ereduaren zenbatespena eta inferentzia heterozedastizitatepean**

Frogatu den bezala, heterozedastizitatepean ereduko koefizienteen KTA zenbateslea lineala da perturbazioekiko, alboragabea, baina jada ez da bariantza minimoduna. Gainera, perturbazioaren bariantzaren ohiko zenbateslea alboratua denez, ezin dira kontrasteak gauzatu ohiko estatistikoekin. Hala ere, posible da inferentzia egitea KTA zenbateslearekin baina, bere bariantza-kobariantza matrizea era sendoan zenbatetsi beharra dago. White-ek proposatutako $\widehat{Bar(\hat{\beta})}_{WHITE} = (X'X)^{-1}X'S\,X\,(X'X)^{-1}$ erabiliz, non S matrizea hondar karratuz osatutako matrize diagonala baita, hau da:

$$S = \begin{bmatrix} \hat{u}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{u}_N^2 \end{bmatrix}$$

Kasu honetan, koefizienteen q murrizketa lineal kontrastatzeko, honako estatistiko hau erabiliko da:

$$\begin{cases} H_0: R\beta = r \\ H_a: R\beta \neq r \end{cases} \qquad F = \left(R\,\hat{\beta} - r\right)' \left[R\,(X'X)^{-1} X'S\,X\,(X'X)^{-1}\right]^{-1} \left(R\,\hat{\beta} - r\right) \xrightarrow{H_0\ b} \chi_q$$

Eta erabakitze-araua hipotesi hutsa baztertzea izango da, baldin eta $F > \chi_{q|\alpha}$ bada.

Era berean, koefizienteen murrizketa bakarraren kasuan, hau da, $q = 1$ denean:

$$t = \frac{R\,\hat{\beta} - r}{\sqrt{R\,\widehat{Bar(\hat{\beta})}_{WHITE}\,R'}} \xrightarrow{H_0\ \ b} N(0\,,1)$$

Eta erabakitze-araua hipotesi hutsa baztertzea izango da α esangura-mailarekin, baldin eta $|t| > Z_{\alpha/2}$ bada, hau da, $N(0, 1)$ banaketako $\alpha/2$ kuantila baino handiagoa bada.

Adibidez, banakako esangura-kontraste bat egiteko:

$$\begin{cases} H_0:\ \beta_k = 0 \\ H_a: \beta_k \neq 0 \end{cases} \qquad t = \frac{\hat{\beta}_k - 0}{\widehat{des}\,(\hat{\beta}_k)_{WHITE}} \xrightarrow{H_0\ \ b} N(0,1)$$

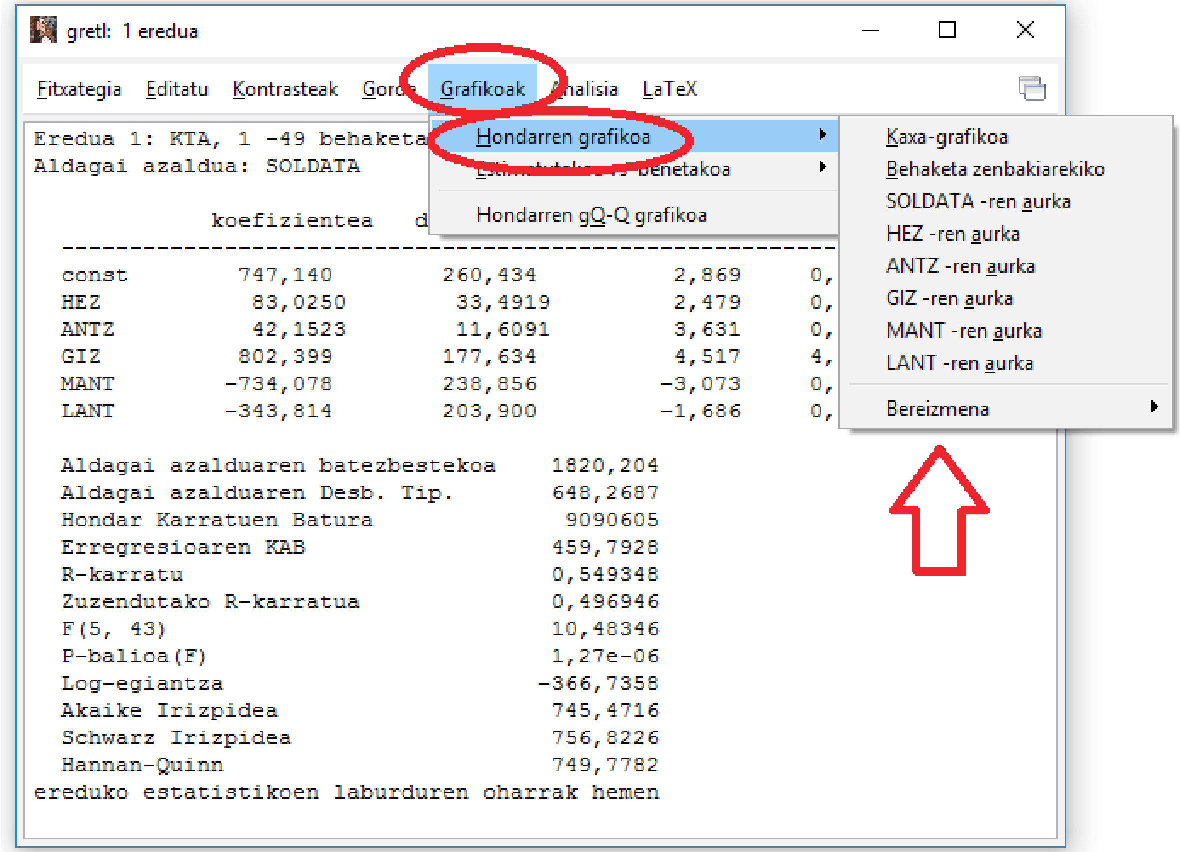
Eta X_k aldagaia nabaria izango da α esangura-mailarekin, baldin eta $|t| > Z_{\alpha/2}$ bada, hau da, $N(0, 1)$ banaketako $\alpha/2$ kuantila baino handiagoa bada.

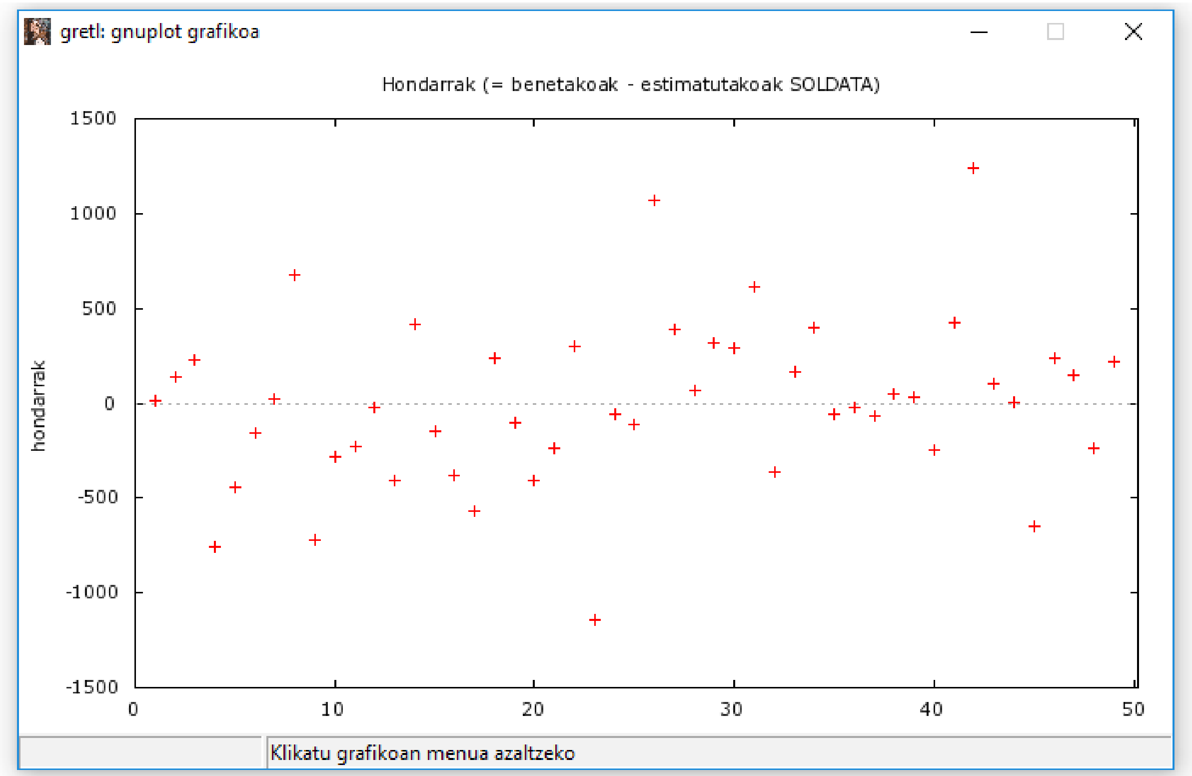
11.6. **Heterozedastizitatea Gretl-ekin**

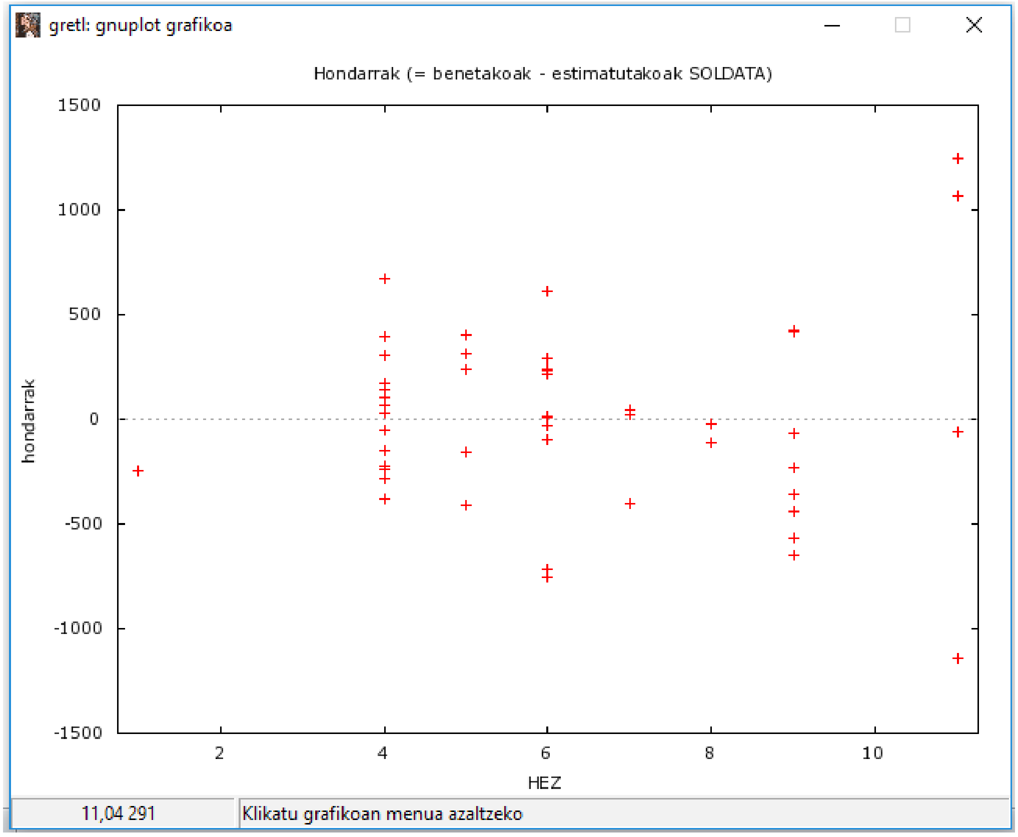
Heterozedastizitatea aztertzeko Gretl bitartez, aurreko gaietan bezala, soldataren adibidea erabiliko da.

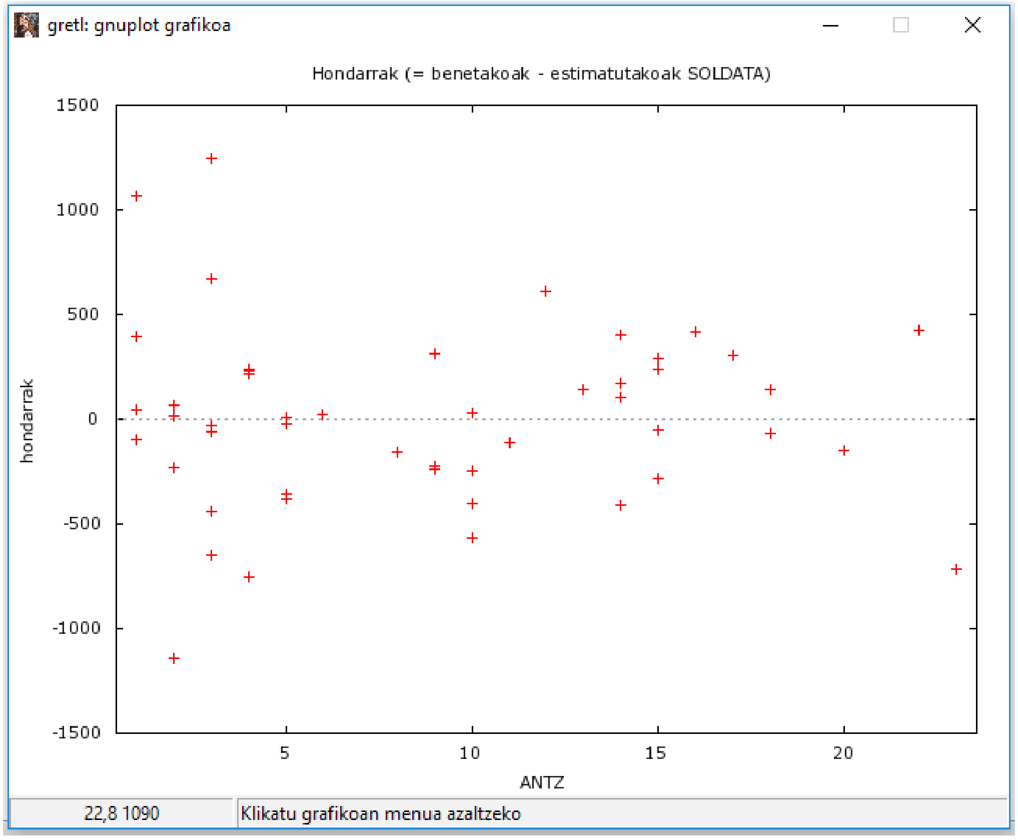
$$SOLDATA_i = \beta_1 + \beta_2 HEZ_i + \beta_3 ANTZ_i + \beta_4 GIZ_i + \beta_5 MANT_i + \beta_6 TAILER_i + u_i$$

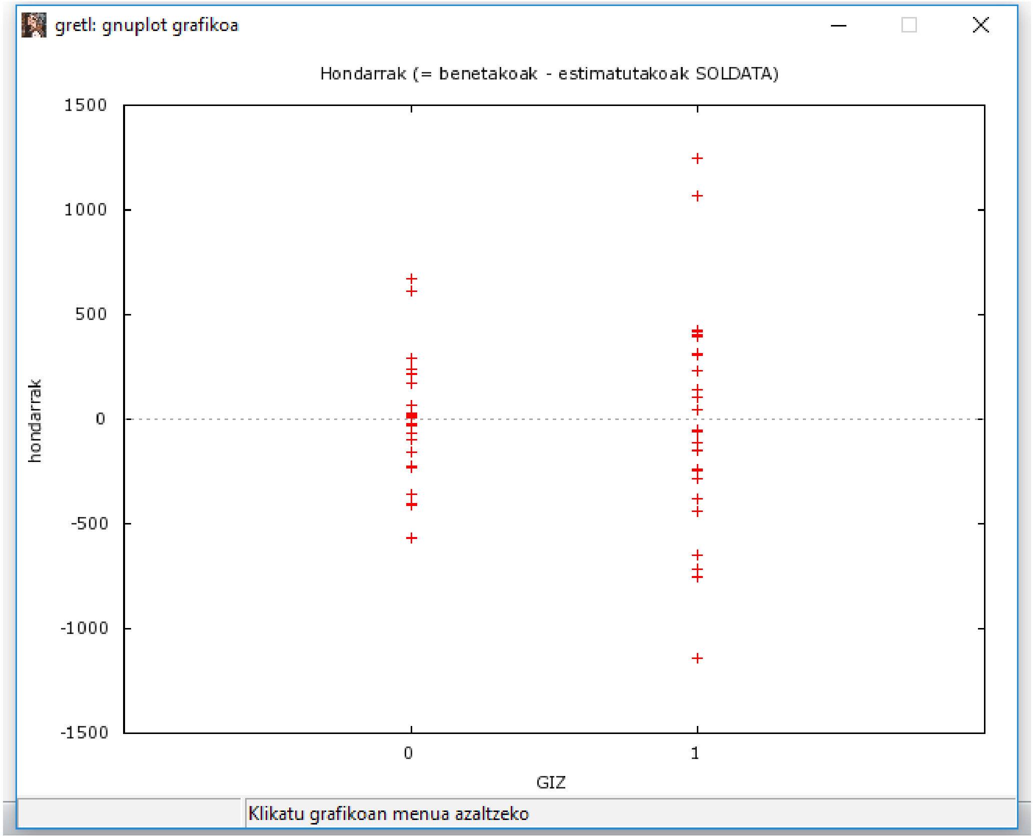
Eredua KTA bitartez zenbatetsi ondoren, hondarren zenbait grafiko lortu dira: hondarrak behaketa-zenbakiarekiko, eta baita ereduan barneratu diren aldagai azaltzaile kuantitatibo eta fikzio-aldagaiekiko ere.

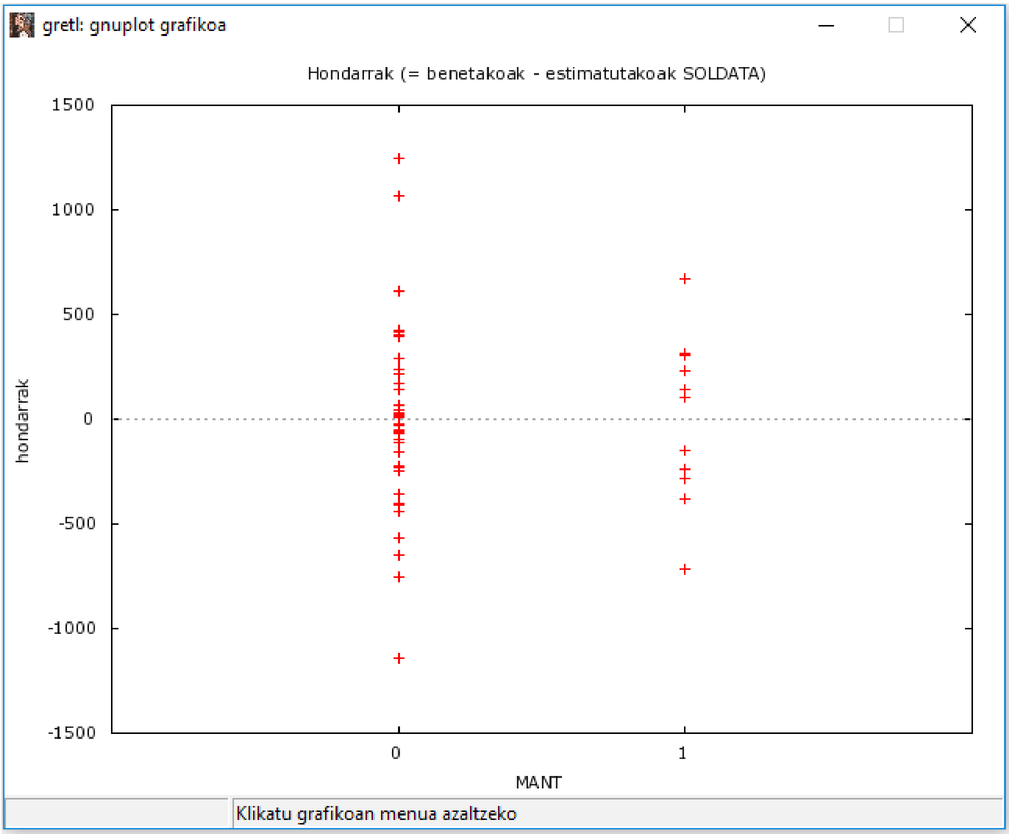

gretl: 1 eredua
Fitxategia Editatu Kontrasteak Gorde Grafikoak Analisia LaTeX
Hondarren grafikoa
Estimatutakoak-benetakoa
Hondarren gQ-Q grafikoa
Kaxa-grafikoa
Behaketa zenbakiarekiko
SOLDATA -ren aurka
HEZ -ren aurka
ANTZ -ren aurka
GIZ -ren aurka
MANT -ren aurka
LANT -ren aurka
Bereizmena
Eredua 1: KTA, 1 -49 behaketa
Aldagai azaldua: SOLDATA
koefizientea
const 747,140 260,434 2,869
HEZ 83,0250 33,4919 2,479
ANTZ 42,1523 11,6091 3,631
GIZ 802,399 177,634 4,517
MANT -734,078 238,856 -3,073
LANT -343,814 203,900 -1,686
Aldagai azalduaren batezbestekoa 1820,204
Aldagai azalduaren Desb. Tip. 648,2687
Hondar Karratuen Batura 9090605
Erregresioaren KAB 459,7928
R-karratu 0,549348
Zuzendutako R-karratua 0,496946
F(5, 43) 10,48346
P-balioa(F) 1,27e-06
Log-egiantza -366,7358
Akaike Irizpidea 745,4716
Schwarz Irizpidea 756,8226
Hannan-Quinn 749,7782
ereduko estatistikoen laburduren oharrak hemen

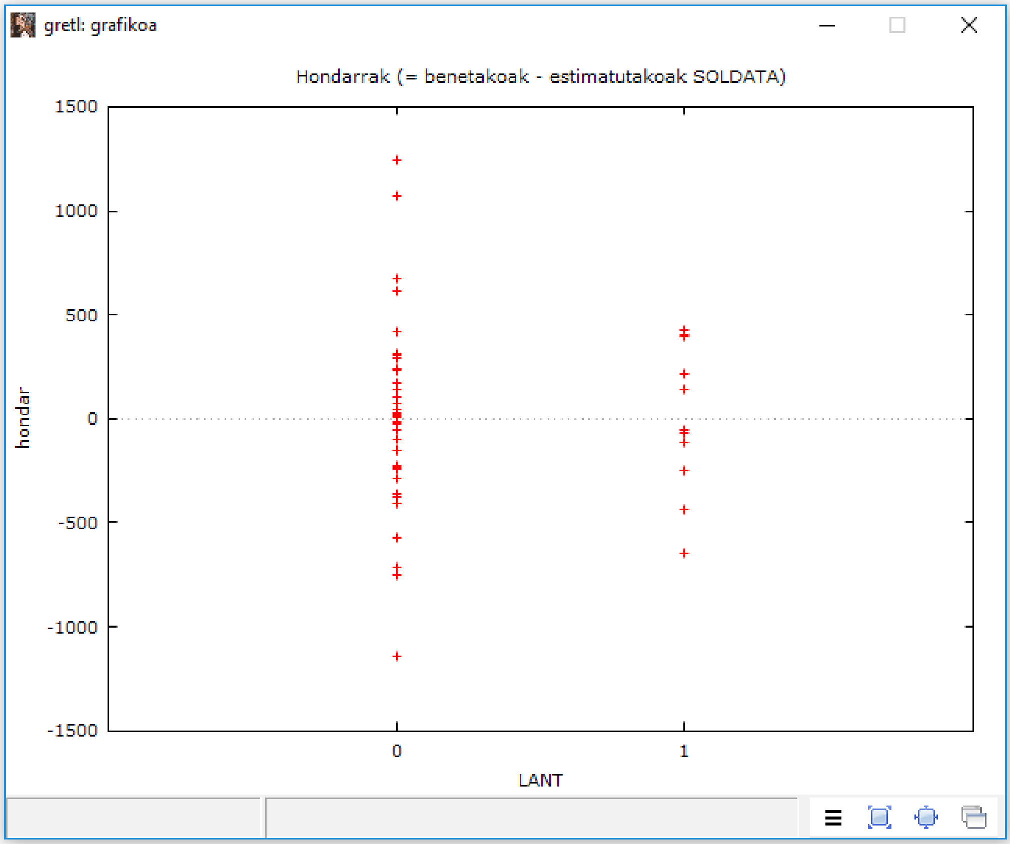

gretl: gnuplot grafikoa
Hondarrak (= benetakoak - estimatutakoak SOLDATA)
1500
1000
500
0
-500
-1000
-1500
hondarrak
0
10
20
30
40
50
Klikatu grafikoan menua azaltzeko


gretl: gnuplot grafikoa
Hondarrak (= benetakoak - estimatutakoak SOLDATA)
1500
1000
500
0
-500
-1000
-1500
hondarrak
2
4
6
8
10
HEZ
11,04 291
Klikatu grafikoan menua azaltzeko

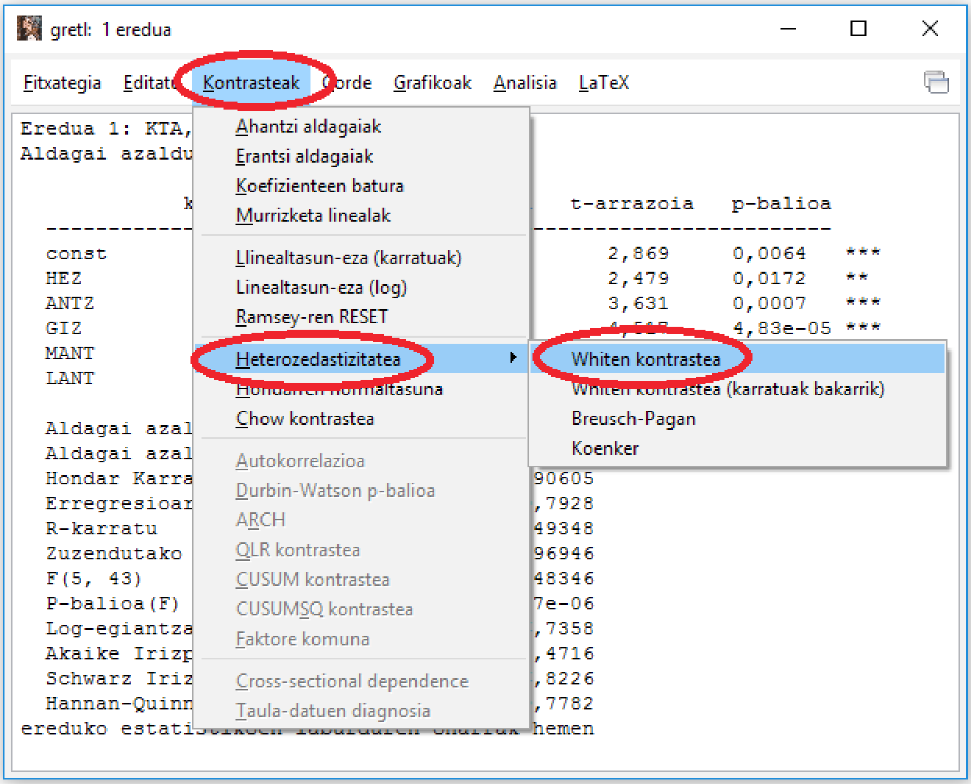

gretl: gnuplot grafikoa
Hondarrak (= benetakoak - estimatutakoak SOLDATA)
1500
1000
500
0
-500
-1000
-1500
hondarrak
5
10
15
20
ANTZ
22,8 1090
Klikatu grafikoan menua azaltzeko

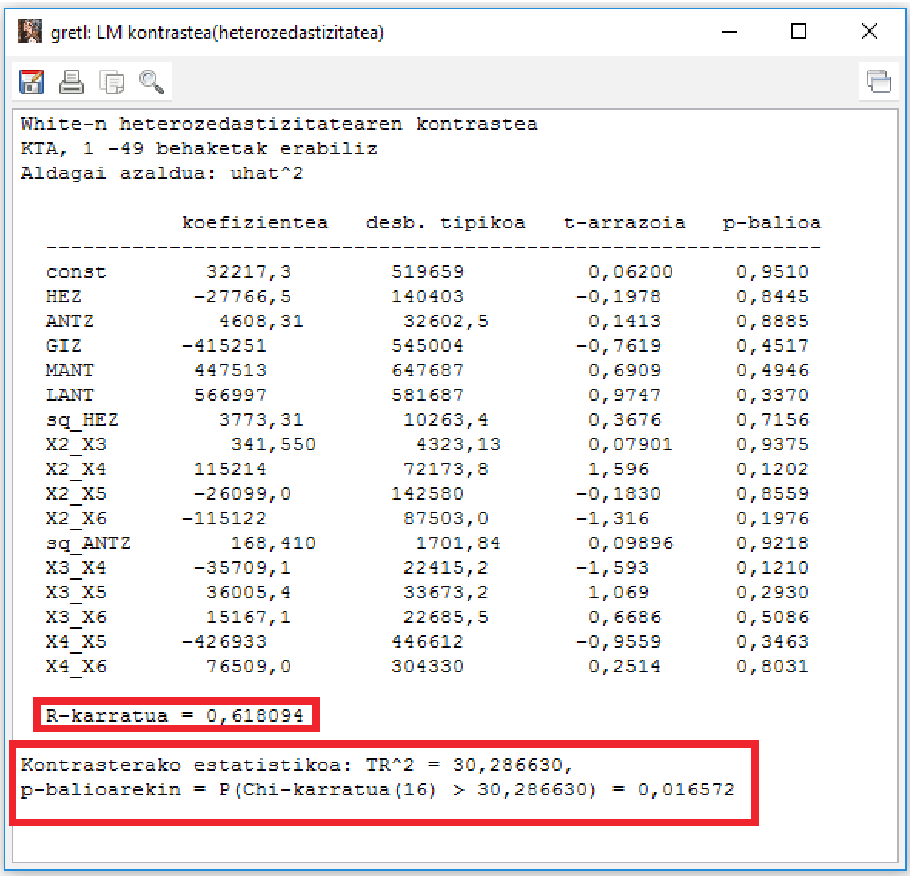

gretl: gnuplot grafikoa
Hondarrak (= benetakoak - estimatutakoak SOLDATA)
1500
1000
500
0
-500
-1000
-1500
hondarrak
0
1
GIZ
Klikatu grafikoan menua azaltzeko


gretl: gnuplot grafikoa
Hondarrak (= benetakoak - estimatutakoak SOLDATA)
1500
1000
500
0
-500
-1000
-1500
hondarrak
0
1
MANT
Klikatu grafikoan menua azaltzeko



Grafiko hauetan zera ikus daiteke: zenbaki-behaketekiko eta aldagai azaltzaile kuantitaboekiko grafikoetan, hondarren sakabanatzea konstantea dela dirudi; fikzio-aldagaiekiko grafikoetan, berriz, 0 balioa edota 1 balioa hartzen dutenean, hondarren sakabanatzea desberdina dela antzeman daiteke. Adibidez, azkenengo grafikoan, (*LANT*ekiko), langileak tailerrean lan egiten ez duenean, hau da *LANT* = 0 denean, sakabanatzea handiagoa da tailerrean lan egiten duenean baino, hau da *LANT* = 1 denean. Hortaz, badirudi heterozedastizitatea egon daitekeela. Egiaztatu ahal izateko, *WHITE*en kontrastea gauzatuko da.

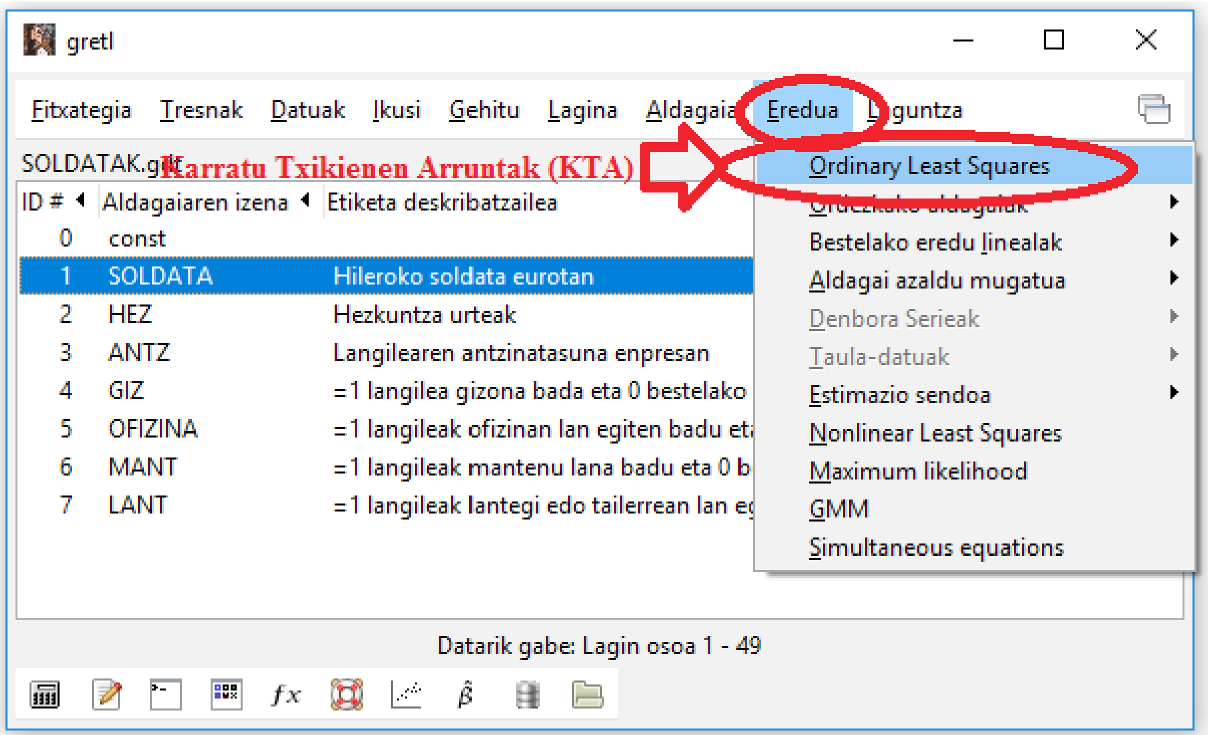


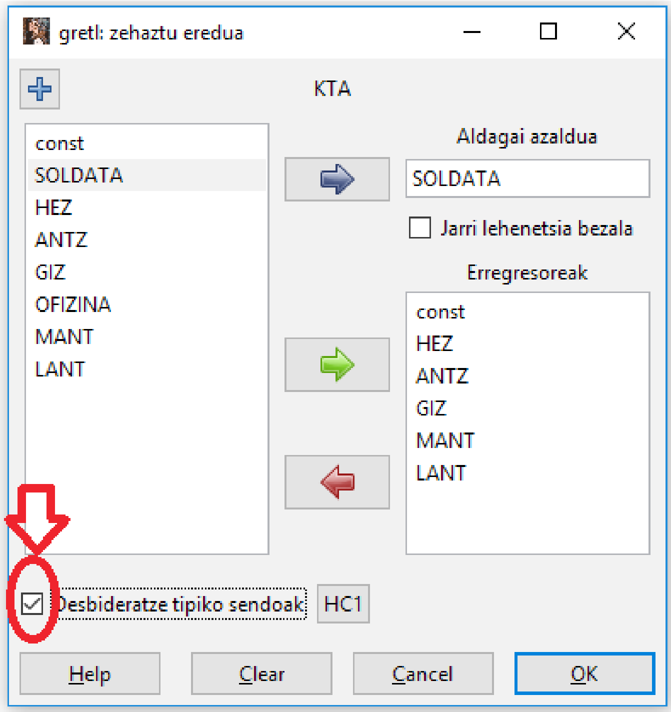


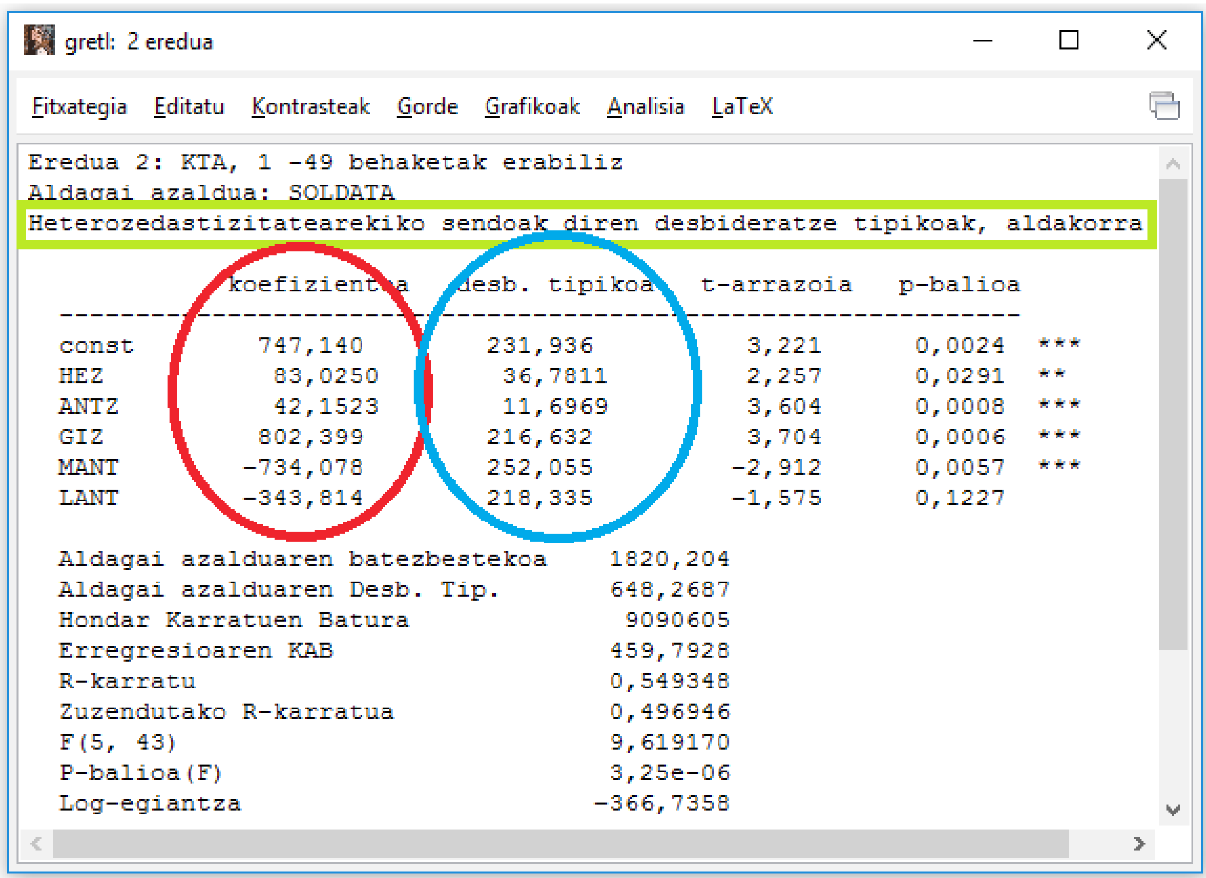

```
White-n heterozedastizitatearen kontrastea
KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: uhat^2

             koefizientea   desb. tipikoa   t-arrazoia   p-balioa
  ----------------------------------------------------------------
  const         32217,3       519659          0,06200     0,9510
  HEZ          -27766,5       140403         -0,1978      0,8445
  ANTZ           4608,31       32602,5        0,1413      0,8885
  GIZ         -415251         545004         -0,7619      0,4517
  MANT         447513         647687          0,6909      0,4946
  LANT         566997         581687          0,9747      0,3370
  sq_HEZ         3773,31       10263,4        0,3676      0,7156
  X2_X3           341,550       4323,13       0,07901     0,9375
  X2_X4        115214          72173,8        1,596       0,1202
  X2_X5        -26099,0       142580         -0,1830      0,8559
  X2_X6       -115122          87503,0       -1,316       0,1976
  sq_ANTZ         168,410       1701,84       0,09896     0,9218
  X3_X4        -35709,1        22415,2       -1,593       0,1210
  X3_X5         36005,4        33673,2        1,069       0,2930
  X3_X6         15167,1        22685,5        0,6686      0,5086
  X4_X5       -426933         446612         -0,9559      0,3463
  X4_X6         76509,0       304330          0,2514      0,8031

  R-karratua = 0,618094

Kontrasterako estatistikoa: TR^2 = 30,286630,
p-balioarekin = P(Chi-karratua(16) > 30,286630) = 0,016572
```

Ikus daiteke erregresio laguntzaileko mugatze-koefizientea $R_w^2 = 0{,}618$ dela; erregresio laguntzailean 17 koefiziente daudenez, p-1=16 izango da, eta *WHITE*en estatistikoaren balioa $WHITE = 30{,}2866$ da. Hortaz, $WHITE > \chi_{16|\alpha=\%5}$ denez, hipotesi hutsa baztertzen da, eta heterozedastizitatea dagoela ondoriotatzen.

Hortaz, eredua *KTA* bitartez zenbatetsiz gero, ereduko koefizienteen *KTA* zenbateslea lineala da perturbazioekiko, alboragabea, baina jada ez da bariantza minimoduna. Gainera, perturbazioaren bariantzaren ohiko zenbateslea alboratua denez, ezin dira kontrasteak gauzatu ohiko estatistikoekin. Hala ere, posible da inferentzia egitea *KTA* zenbateslearekin baina, bere bariantza-kobariantza matrizea era sendoan zenbatetsi beharra dago. White-ek proposatutako $\widehat{Bar(\hat{\beta})}_{WHITE} = (X'X)^{-1}X'S\,X\,(X'X)^{-1}$ erabiliz, non S matrizea hondar karratuz osatutako matrize diagonala baita. Gretl bitartez horrela lortuko genituzke bariantza eta kobariantza zenbatetsi sendoak.




gretl: zehaztu eredua
KTA
const
SOLDATA
HEZ
ANTZ
GIZ
OFIZINA
MANT
LANT
Aldagai azaldua
SOLDATA
Jarri lehenetsia bezala
Erregresoreak
const
HEZ
ANTZ
GIZ
MANT
LANT
Desbideratze tipiko sendoak
HC1
Help
Clear
Cancel
OK


gretl: 2 eredua
Fitxategia
Editatu
Kontrasteak
Gorde
Grafikoak
Analisia
LaTeX
Eredua 2: KTA, 1 -49 behaketak erabiliz
Aldagai azaldua: SOLDATA
Heterozedastizitatearekiko sendoak diren desbideratze tipikoak, aldakorra
koefizientea desb. tipikoa t-arrazoia p-balioa
const 747,140 231,936 3,221 0,0024 ***
HEZ 83,0250 36,7811 2,257 0,0291 **
ANTZ 42,1523 11,6969 3,604 0,0008 ***
GIZ 802,399 216,632 3,704 0,0006 ***
MANT −734,078 252,055 −2,912 0,0057 ***
LANT −343,814 218,335 −1,575 0,1227
Aldagai azalduaren batezbestekoa 1820,204
Aldagai azalduaren Desb. Tip. 648,2687
Hondar Karratuen Batura 9090605
Erregresioaren KAB 459,7928
R-karratu 0,549348
Zuzendutako R-karratua 0,496946
F(5, 43) 9,619170
P-balioa(F) 3,25e-06
Log-egiantza −366,7358

Eredua KTA bitartez zenbatetsi denez, koefizienteen zenbatespenak aurreko gaietan lortutakoen berdinak dira, eta baita mugatze-koefizientea ere; baina desbideratze tipiko zenbatetsiak eta dagozkien t estatistikoak desberdinak dira, oraingo honetan era sendoan zenbatetsi baitira. Emaitza horiekin kontrasteak gauza daitezke, baina asintotikoak izango dira, eta, beraz, lagin handia izan behar da emaitzak fidagarriak izateko.

UNIBERTSITATEKO ESKULIBURUAK
MANUALES UNIVERSITARIOS

INFORMAZIOA ETA ESKARIAK • INFORMACIÓN Y PEDIDOS

UPV/EHUko Argitalpen Zerbitzua • Servicio Editorial de la UPV/EHU
argitaletxea@ehu.eus • editorial@ehu.eus
1397 Posta Kutxatila - 48080 Bilbo • Apartado 1397 - 48080 Bilbao
Tfn.: 94 601 2227 • www.ehu.eus/argitalpenak

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea