2 Gaia

Erregresio linealeko eredu bakuna

Aurkibidea

2.1 Sarrera

Gai honetan ***erregresio linealeko eredu bakuna*** zehazten, estimatzen eta analizatzen ikasiko dugu. Helburu honetarako behar den teoria adibide batekin batera eskeiniko da, Ramanathan (2002) liburuko data3-1 datu fitxategia erabiliz.

Eredu bakuna bi aldagai linealki erlazionatzen duen eredua da,

$$Y_i = \alpha + \beta X_i + u_i \quad i = 1, \ldots, N \tag{2.1}$$

non,

- Y ***aldagai azaldua***, ***aldagai dependentea edo endogenoa*** den, hau da, azaldu edo analizatu nahi den aldagaia.
- X ***aldagai azaltzailea***, ***aldagai independentea edo exogenoa*** den.
- α eta β, eredu bakuneko jatorria eta malda, ***erregresioaren koefizienteak*** diren. *Estimatu beharreko koefizienteen kopurua* K bezala definitzen badugu, orduan eredu bakunean $K = 2$ koefiziente daude estimatzeko.
- u errorea da, zorizko aldagaia, aldagai aleatorioa edo ***perturbazioa***.
- Erabilitako i azpiindizeak ***behaketa*** adierazten du. Orokorrean, **i** azpiindizea gurutzatutako datuekin lan egiterakoan erabiltzen da eta daukagun laginaren datuak denborazko serieak badira, **t** azpiindizea erabiliko dugu.
- N ***lagin tamaina*** da, ikertzen ari garen aldagaien (Y, X) behaketa kopurua. Denborazko serieak ikertzerakoan, T erabiliko dugu lagin tamaina adierazteko.

Ereduan aldagai aleatorio bat (u) sartzearen arrazoiak desberdinak izan daitezke:

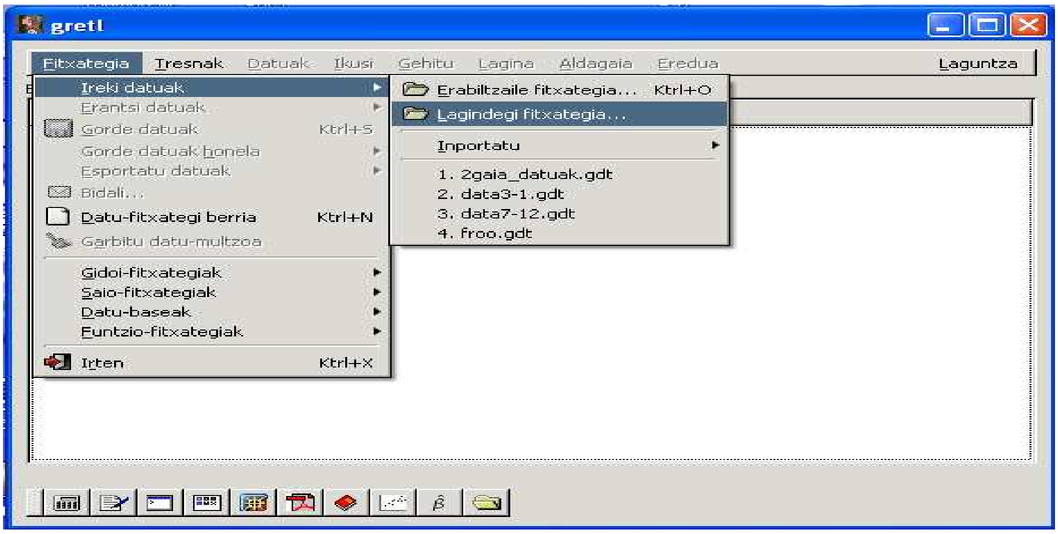
1 Aurresan ezinezko efektuak, bai egoera ekonomiko berezi batetik edo analizatzen ari garen testuinguruaren ezaugarrietatik eratortzen direnak eta banako edo entitate ekonomikoek dituzten lehentasunak edo gustoetatik sortarazten diren zenbagarri-ezinezko efektuak jasotzeko.

2 Interesgarriak zaizkigun aldagaien datuak biltzerakoan sortzen diren neurketa erroreak kontuan hartzeko.

3 Zehazpen erroreak biltzeko, aldagairen bat faltan egoterakoan edo ereduko zati sistematikoan agertzen diren erlazio ez linealak kontuan ez hartzerakoan sortzen direnak.

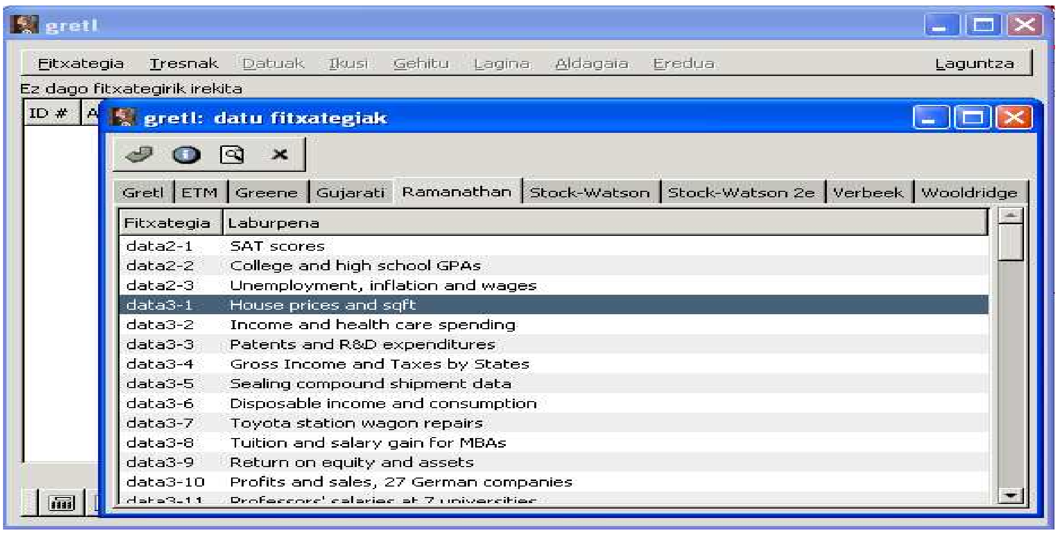
Behin eredu bakunaren zehazpena ikusirik, aurretik esandako datuen analisiarekin hastera goaz. Datu multzo hauek lortzeko: *Fitxategia* $\rightarrow$ *Ireki datuak* $\rightarrow$ *Lagindegi fitxategia...* aukeran.

2.1 Irudia: Datuen bilakaera



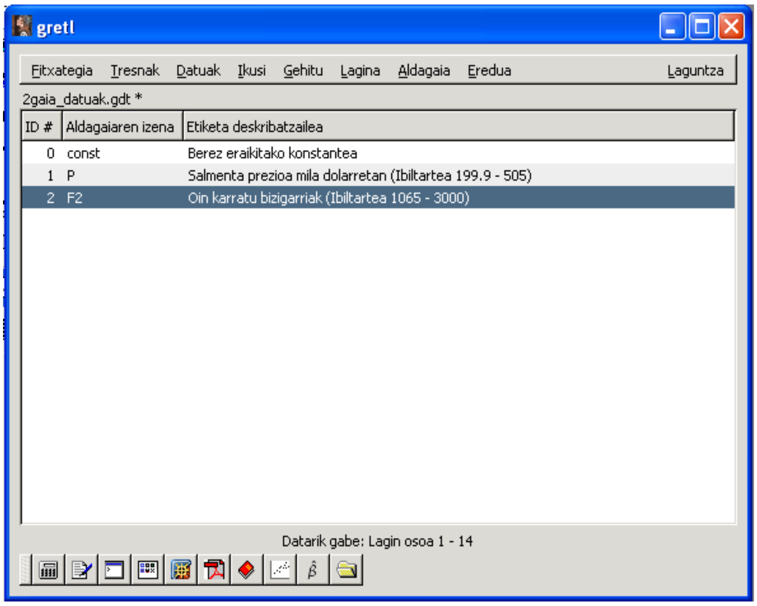
eta Ramanathan karpetaren barnean *data3-1 House prices and sqft* aukeratu:

2.2 Irudia: Datu-fitxategiaren aukerapena



Segidan aldagaien izenak (**"P" eta "F2"**), deskribapena (**"Salmenta prezioa mila dolarretan (Ibiltartea: 199,9-505)" eta "Oin karratu bizigarriak (Ibiltartea: 1065-3000)"**) eta grafikoetan erakutsiko diren izenak (**"P" eta "F2"**) aldatuko ditugu. Horretarako lehen gaian ikusitako pausuak jarraitu behar dira. Lortzen den emaitza honakoa da:

2.3 Irudia: Datuen erakuspena



Ondoren datuak gorde egingo ditugu Gretl formatuan "2gaia-datuak.gdt" izenarekin. Bi aldagaien balioak ondorengo taulan erakusten dira,

2.1 Taula: *"2gaia-datuak.gdt"* fitxategiko datuak

i	P_i	F2	i	P	F2
1	199,9	1065	8	365,0	1870
2	228,0	1254	9	295,0	1935
3	235,0	1300	10	290,0	1948
4	285,0	1577	11	385,0	2254
5	239,0	1600	12	505,0	2600
6	293,0	1750	13	425,0	2800
7	285,0	1800	14	415,0	3000

eta aldagai bien sakabanatzea 2.4 irudian aurkezten da, bertan erlazio lineal positiboa ikusten delarik.

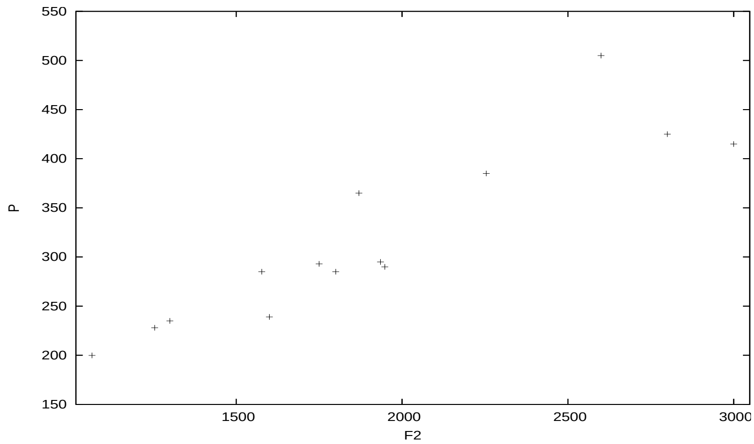
Laburbilduz, adibide honetako eredu bakunaren zehazpena honakoa da:

$$P_i = \alpha + \beta \; F2_i + u_i \qquad i = 1, \ldots, 14 \tag{2.2}$$

analisia aurrera eramateko $N = 14$ behaketaz osaturiko lagina daukagu eta

- P_i "salmenta prezioa" aldagai dependentearen, endogenoaren edo azalduaren i behaketa da.

2.4 Irudia: (Y, X)ren sakabanatze diagrama



- $F2_i$ "oin karratu bizigarri" aldagai independentearen, exogenoaren edo azaltzailearen i behaketa da.

- Estimatu beharreko koefizienteak α eta β dira. 2.4 irudia begiratuz, bai jatorria eta baita malda ere, positiboak direlakoan gaude.

- Testuinguru honetako perturbazioak, u_i, azalera (oin karratutan) berdineko etxebizitzen arteko salmenta prezioen desberdintasunak azaltzen dituen ezaugarriak biltzen ditu: kokalekua, etxearen orientazioa, bistak, argi naturalaren kopurua, etab..

2.2 Oinarrizko hipotesiak

Eredua zehaztu ondoren estimazioaren testuingurua finkatu behar dugu. Horretarako ***oinarrizko hipotesi*** batzuk betetzen direla suposatuko dugu. Behin estimatzen, doikuntzaren egokitasuna neurtzen, beharrezkoak diren kontrasteak egiten eta aurresanak lortzen ikasten dugunean (kurtsoaren amaieran), oinarrizko hipotesi batzuk erlaxatzen hasi gaitezke. Baina bitartean denak betetzen direla suposatuko dugu. Oinarrizko hipotesi hauek erregresioaren elementu desberdinei buruzkoak dira.

Forma funtzionalari dagokionez

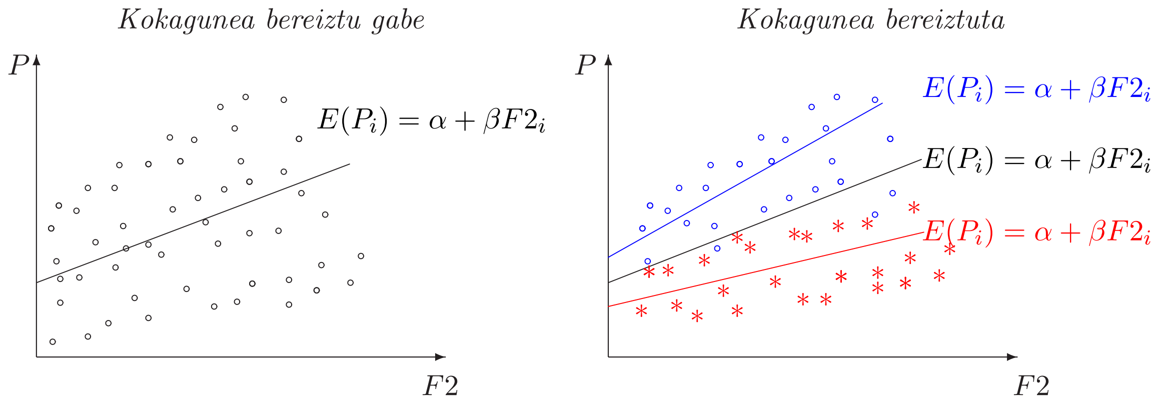
1. *Eredua koefizienteekiko lineala da.* Kurtsoan zehar estimatuko diren ereduak koefizienteekiko linealak dira. Demagun hiri bateko etxebizitzen batezbesteko prezioa (P) azaldu nahi dugula duten azaleraren (F2) menpean (oin karratutan), orduan koefizienteekiko lineala den eredu bat $P_i = \alpha + \beta F2_i + u_i$ litzateke. Bestalde, eredua ez da zertan aldagai azaltzailearekiko lineala izan behar, hau da, $P_i = \alpha + \beta F2_i^2 + u_i$ eredua

estimatzea posible litzateke nahiz eta etxebizitzaren azalerak prezioaren duen eragina lineala ez izan, koadratikoa baizik.

Koefizientei dagokienez

2. *Koefizienteak konstante mantentzen dira laginean zehar.* Hasiera batean, aldagai azaltzaileak duen eragina laginean zehar konstante mantentzen dela suposatuko dugu. Aurreko adibidearekin jarraituz, 2.5 irudiko ezkerreko grafikoko puntu hodeitik pasatzen den "erdiko zuzena" estimatzea interesatuko litzaiguke.

2.5 Irudia: Bilboko etxebizitzen prezioa daukaten azalera bizigarriarekiko



Hala ere, 2.5 irudiko eskubiko grafikoan ikusten den bezala, etxebizitza batzuk hiri zentrukoak baldin badira (urdinez daudenak) eta besteak hiriaren kanpoaldekoak (gorriz daudenak) badira, orduan gerta daiteke "erdiko zuzen bien" (urdina eta gorria) estimazioa interesatzea horrela etxebizitzaren kokapena bereizten baita. Kasu honetan koefizienteak ez dira konstante mantentzen, desberdinak baitira kokagunearen arabera.

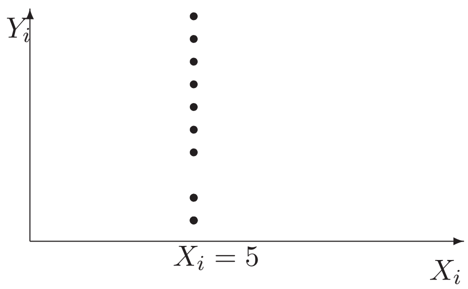
Aldagai azalduari dagokionez

3. *Aldagai azaldua koantitatiboa da.* Aldagai azaldu koalitatiboa daukan eredu bat estimatzeko, kurtso honetan ikasiko ditugun estimazio metodoek ez dute balio.

Aldagai azaltzaileari dagokionez

4. *X aldagai azaltzailearen lagin bariantza (S_X^2) ezin da zero izan eta gainera $N \geq K = 2$* izan behar da. Hipotesi hau koefizienteak (jatorria eta malda) identifikatzeko beharrezkoa da. Hasteko, estimatu behar diren koefizienteen kopurua behaketen kopurua baino handiago izanez gero, orduan estimazioa aurrera eramateko ez daukagu informazio nahikorik. Bestalde, aldagai azaltzailearen lagin bariantza zero izango balitz, (adibidean, $S_{F2}^2 = 0$), hau da, $F2$ termino konstantearen konbinazio lineala izango balitz (adibidez, $F2 = 5 \times$ termino konstantea $= 5 \times 1 = 5$), edo beste era batera esanda, etxebizitzek oin karratu berdina izango balute ($F2_i = 5\ \forall i$, adibidez), orduan malda identifikatzea ezinezkoa izango litzateke. Egitez, 2.6 irudiko grafikoan ikusten den bezala, $P_i = \alpha + \beta 5 + u_i$ moduko eredu batean ezinezkoa da aldagai azalduaren (P) sakabanatzea azaltzea, aldagai azaltzaileak ez duelako inolako aldakortasunik.

2.6 Irudia: Eredua: $Y_i = \alpha + \beta \times 5 + u_i$, non $S_X^2 = 0$ den



5. *Aldagai azaltzailea ez da aleatorioa* eta ondorioz ez dago errorerekin neurtuta. Gure adibidean etxebizitzen azalera oin karratutan zehaztasun osoz (inolako milimetroko errorerik gabe) neurtuta dagoela inplikatzen du.

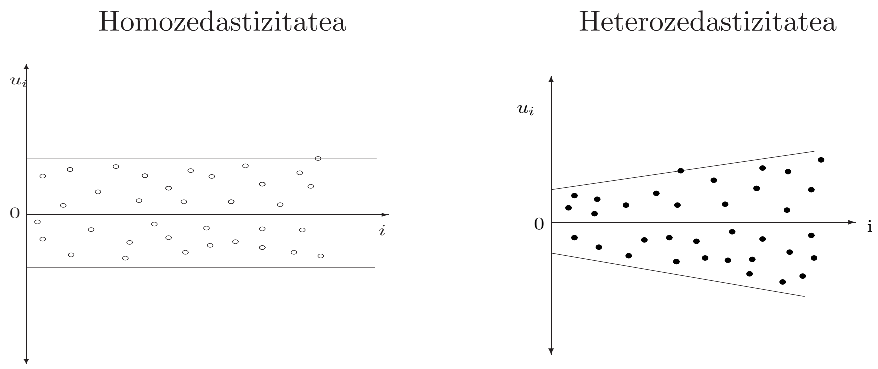
6. *Eredua ondo zehaztuta dago.* Orokorrean azalduz, ereduak ezin du aldagai nabaririk barneratu gabe utzi eta ezta kontrakoa ere, hau da aldagai ez nabariren bat kontuan hartu. Eredu bakunean, hipotesi hau ezartzerakoan, Y aldagai azaldua azaltzeko behar den aldagai azaltzaile bakarra X aldagaia dela eskatzen ari gara.

Perturbazioari dagokionez

7. *Perturbazioen populazio batezbestekoa zero da,* $E(u_i) = 0$. Aurresan edo estimaezina den errorearen batezbestekoa zero izateak ereduaren zati sistematikoa edo analizatu nahi den batezbesteko portaera $E(Y_i) = \alpha + \beta X_i$ izatea dakar (adibidean, $E(P_i) = \alpha + \beta F2_i$).

8. *Perturbazioen populazio bariantza konstantea da.* Aldagai aleatorioaren edo perturbazioaren aldakortasuna laginean zehar konstante mantentzen dela suposatuko dugu 2.7 irudiko ezkerreko grafikoan ikusten den bezala.

2.7 Irudia: Perturbazioen bariantza laginean zehar

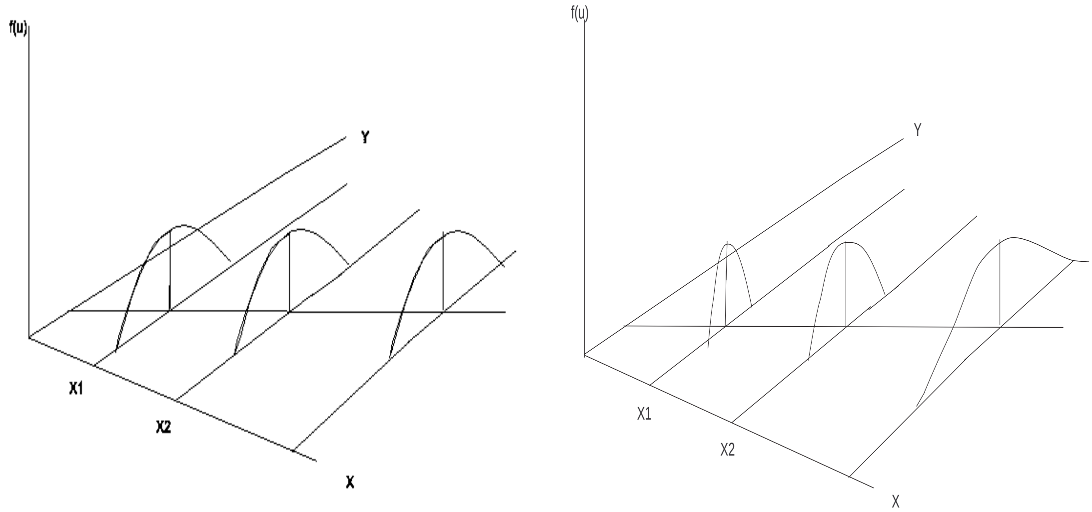


Horrela, 2.8 irudiko ezkerreko grafikoan azaltzen den bezala, aldagai azaltzaileen balioak emanik, aldagai azalduak har ditzakeen balio posibleen tartearen zabalera berdina da (kanpai guztien oinarriaren diametroa berdina da) eta balio bakoitzak irteteko duen probabilitatea X aldagaiak hartzen duen balioarekiko independentea da (kanpai guztien itxurak berdinak dira).

Kontrako kasuan perturbazio heterozedastikoak ditugu, perturbazioen bariantza laginean zehar aldatzen delarik. Horrela, 2.7 irudiko eskubiko grafikoan bariantza gorakorrak dituen perturbazioak marraztu dira.

2.8 Irudia: Perturbazioen bariantza aldagai azaltzailearekiko

Homozedastizitatea Heterozedastizitatea



Aldagai azaltzailearekiko duen interpretazioa ulertzeko, 2.8 irudiaren eskubiko grafikora begiratu behar dugu. Demagun grafikoko aldagaiak gure adibidekoak direla. Etxebizitzak txikiak direnean, ($F2$= oin karratu gutxi) prezioek har ditzaketen balio posibleak oso antzekoak izateko probabilitatea handia da eta beraz bariantza txikia da (kanpaiaren diametroa txikia da eta puntaduna). Baina pisuen azalera handitzen doan heinean, prezioek har ditzaketen balio posibleen tartea handiagoa da eta beraz bariantza handiagoa du (kanpaiaren diametroa zabalagoa eta altuera txikiagokoa).

Horrela bada, azalera txikiko etxebitzen prezioak nahiko antzekoak dira baina azalera handitzerakoan prezio posibleen tartea ere handitu egiten da eta ondorioz prezio oso desberdinetako etxebizitza handiak aurkitu ahal dira.

9. *Perturbazioek ez daukate autokoerlaziorik.* Momentuz, perturbazio desberdinen arteko koerlazioa zero dela suposatuko dugu ($koer(u_i, u_j) = r_{u_i,u_j} = 0;\ i \neq j$). Hortaz, beraien arteko kobariantza zero izango da ere, $kob(u_i, u_j) = 0\ i \neq j$.

10. *Perturbazioek banaketa Normala jarraitzen du.* Azken hipotesi hau, aurrerago ikusten den bezala, ez da estimazioarentzat beharrezkoa, ezta estimatzailearen propietateak

lortzeko ere, inferentzia (kontrasteak) egitea edo konfidantza tarteak ateratzea interesatzen zaigunean, izango da soilik beharrezkoa.

2.3 Karratu Txikienen Arruntetako metodoa

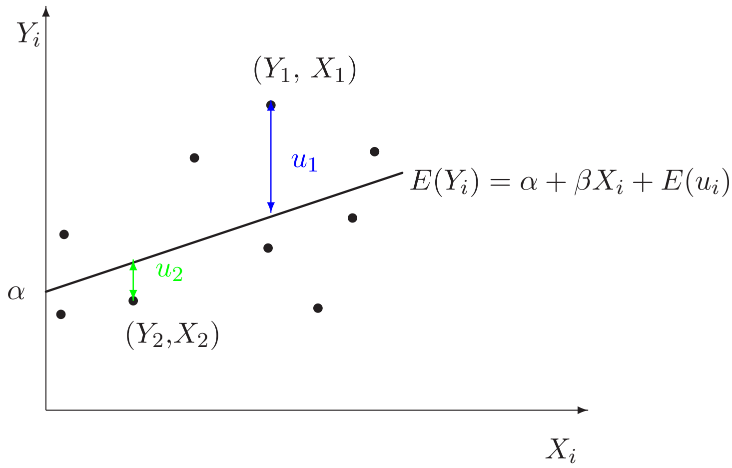
Analizatuko dugun testuinguruaren ezaugarriak (oinarrizko hipotesiak) finkatuz, eredu bakunak zehaztasunez zer adierazten duen ikustera goaz. Hasierako (2.1) ereduak egitura berdineko N ekuazio (berdintzak) biltzen ditu:

$$\begin{aligned}
Y_1 &= \alpha + \beta X_1 + u_1 \\
Y_2 &= \alpha + \beta X_2 + u_2 \\
&\vdots \\
Y_i &= \alpha + \beta X_i + u_i \\
&\vdots \\
Y_N &= \alpha + \beta X_N + u_N
\end{aligned}$$

Ekuazio hauek begiratuz, (Y_i, X_i) punto bakoitzatik β maldako zuzen bat pasatzen dela ikusi daiteke. Eta malda berekoak direnez, N zuzenak paraleloak dira. Jatorriari dagokionez, zuzen bakoitzaren jatorria lortzeko, α "batezbesteko jatorriari" zuzen bakoitzari dagokion perturbazioa (u_i) gehitu behar zaio, hau da, zuzen bakoitzaren jatorria $\alpha + u_i$ balioa da.

Ondorengo 2.9 irudian, N zuzen hauetariko bi marrazten dira, lehenengoko biak hain zuzen. Beltzez dagoen lerroa "erdiko zuzenari" dagokio, perturbazio zero izaterakoan lortzen den zuzena.

2.9 Irudia: Erregresio eredu bakuna



Lehen behaketari (Y_1, X_1) dagokion zuzena (lerro urdina) erdikoarekiko paraleloa da, malda berekoa baita, eta berarekiko duen distantzia u_1 balioak neurtzen du. Bigarren behaketari (Y_2, X_2) dagokion lerroak ezaugarri berberak ditu, erdiko zuzenarekiko paraleloa eta erdikotik

u_2 distantziara aurkitzen da. Era berean, zuzen guztiak marraztu ditzakegu, hau da puntu bakoitzarentzat bat.

Erregresio eredu bakun baten ***estimazioak*** ez du N zuzen guztien estimazioa helburutzat, baizik eta "erdiko zuzenaren" estimazioa. Grafikoki, eredu bakun bat estimatzerakoan datuei hoberen doitzen diren β malda eta α "batezbesteko" jatorria estimatu nahi dira. Teknikoki, ekonometrian oinarrituz, laginetik eratorritako $(Y_1, X_1), (Y_2, X_2), \ldots, (Y_N, X_N)$ behaketek ematen duten informaziotik, interesatzen zaigun aldagaiaren (Y_i) batezbesteko portaera $(\alpha + \beta X_i)$ zein den kalkulatu nahi dugu.

Ondorioz, eredua estimatzera joan baino lehen, kontzeptu berri batzuk definitu beharrean gaude. Estimatzea interesatzen zaigun "erdiko zuzena" ***Populazioaren Erregresio Funtzioa (PEF)*** bezala ezagutzen da eta populazio koefiziente ezezagunen (α eta β) menpekoa da. Ereduko zati sistematikoa (estimagarria den zatia) jasotzen du eta aldagai azalduaren batezbesteko portaera neurtzen du:

$$\begin{aligned} E(Y_i) &= E(\alpha + \beta X_i + u_i) = \\ &= \alpha + \beta X_i + \underbrace{E(u_i)}_{=0} = \alpha + \beta X_i. \end{aligned}$$

Ereduko ***perturbazioa*** (estimaezinezko zatia) aldagai azalduaren benetako balioa eta populazio erregresio funtzioaren arteko diferentzia da:

$$u_i = Y_i - \alpha - \beta X_i.$$

eta ereduko zati sistematikoak azaldu ezin izan duen guztia biltzen du.

Lagin konkretu batetik lorturiko azken emaitza ***Lagin Erregresio Funtzioa (LEF)*** bezala ezagutzen da baina praktikan ***estimatutako ereduari*** buruz hitz egiten da gehienetan. LEF lortzeko erregresio koefizienteak estimatu $(\hat{\alpha}, \hat{\beta})$ behar dira:

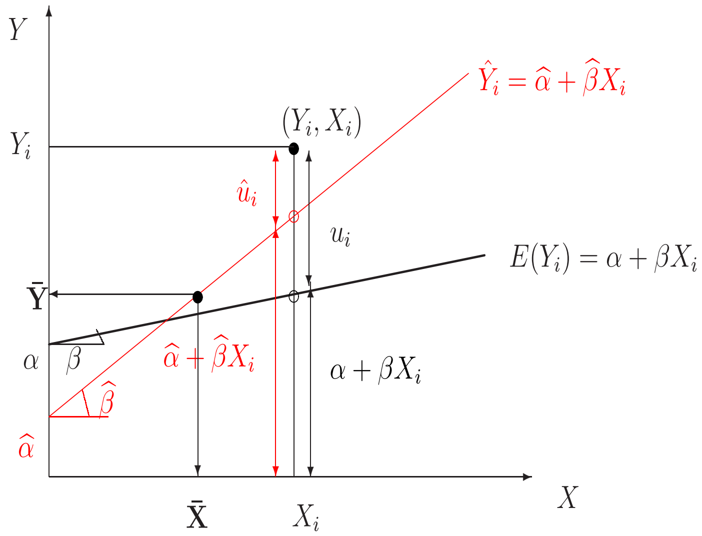
$$\hat{Y}_i = \widehat{E(Y_i)} \quad = \quad \hat{\alpha} + \hat{\beta} X_i.$$

Ereduko ***hondarra*** berriz, azaldu nahi dugun aldagaia eta lagin erregresio funtzioaren arteko diferentzia da:

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\alpha} - \hat{\beta} X_i \\ &= \alpha + \beta X_i + u_i - \hat{\alpha} - \hat{\beta} X_i \\ &= (\alpha - \hat{\alpha}) + (\beta - \hat{\beta}) X_i + u_i \end{aligned} \tag{2.3}$$

Estimazio errore bat da, hau da, eredua estimatzerakoan egiten diren errore guztien batura. Hondarraren barnean bi motako erroreak izango ditugu: elementu estimagarriak (α, β) estimatzerakoan $(\widehat{\alpha}, \widehat{\beta})$ egiten den erroretik eratortzen dena $(\alpha - \widehat{\alpha}, \beta - \widehat{\beta})$ eta estimatu ezinezko elementutik (u_i) eratortzen dena. Ondorioz, benetan garrantzitsua da perturbazio eta hondarrak bereiztea eta ez nahastea. Irudikaturiko 2.10 grafikoan populazio erregresio funtzioa

2.10 Irudia: Populazio eta lagin erregresio funtzioak



(beltzez), bere α jatorria eta β maldarekin (populazio koefizienteak) irudikatuta dago. Bertan, aldagai azalduaren edozein i behaketa (Y_i) lortzeko, ereduaren zati sistematikoari $\alpha+\beta X_i$ (PEF gainean kokaturik) perturbazioa (u_i) gehitu egin behar zaiola ($Y_i = \alpha+\beta X_i+u_i$) erraz ikus daiteke.

Lagin erregresio funtzioa eta estimatutako koefizienteak ($\hat{\alpha}$ eta $\hat{\beta}$) gorriz jarrita daude. PEF eta LEF-ren arteko diferentzia koefizienteak estimatzerakoan egiten den erroreengatik ($\hat{\alpha} \neq \alpha, \hat{\beta} \neq \beta$) ematen da.

LEF oinarritzat hartuz, aldagai azalduaren edozein i behaketa (Y_i) lortzeko, estimatutako zati sistematikoari ($\hat{\alpha} + \hat{\beta}X_i = \widehat{Y}_i$, estimatutako ereduari) dagokion hondarrak hartzen duen balioa ($\widehat{u}_i$) gehitu egin behar zaio, horrela $Y_i = \hat{Y}_i + \hat{u}_i$ lortuz. Azkenik, lagin erregresio funtzioa batezbestekoen puntutik $(\bar{Y}, \bar{X})$ pasatzen dela ere ikus daiteke.

Eredu bakuneko **koefizienteen interpretazioari** dagokionez:

- $\alpha = E(Y_i|X_i = 0)$: aldagai azalduaren batezbesteko balioa edo esperotako balioa da aldagai azaltzaileak hartzen duen balioa zero denean.

- $\beta = \frac{\partial E(Y_i)}{\partial X_i} = \frac{\Delta E(Y_i)}{\Delta X_i}$: aldagai azaltzailea unitate batean handitzerakoan, aldagai azalduaren esperotako gehikuntza edo batezbesteko gehikuntza β unitatekoa da.

Gure adibidera bueltatuz, koefizienteen interpretazioak honakoak dira:

$\alpha = E(P_i|F2_i = 0)$: etxebizitzaren batezbesteko prezioa da (mila dolarretan) etxebizitzak duen azalera zero oin karratu denean. Kasu honetan, koefiziente hau zero izatea esperoko genuke, azalera gabeko etxebizitzari buruz hitz egiteak ez duelako inolako zentzurik. Hala ere, normalean konstantea, eredutik ez da kentzen zeren bestela emaitzen betiko interpretazioa galdu egiten baita. Bestalde adibide honetan etxebizitzak duen hasierako batezbesteko prezioa (tramiteak,...) bezala kontsideratu daiteke eta kasu hau emango balitz, orduan positiboa izatea esperoko genuke.

$\beta = \frac{\partial E(P_i)}{\partial F2_i}$: etxebizitzaren azalera oin karratu batean handitzerakoan espero dugun prezioaren batezbesteko gehikuntza β mila unitatekoa da.

Ondoren ***Karratu Txikienen Arruntetako*** estimatzailea aterako dugu eta horretarako hondar karratuen batura minimizatu behar da:

$$\min_{\hat{\alpha},\hat{\beta}} \sum_{i=1}^{N} \hat{u}_i^2 = \min_{\hat{\alpha},\hat{\beta}} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 = \min_{\hat{\alpha},\hat{\beta}} \sum_{i=1}^{N} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2 \tag{2.4}$$

Minimizazio honetatik α eta β koefizienteen estimatzaileen adierazpenak ($\widehat{\alpha}$ eta $\widehat{\beta}$) lehen ordenako baldintzak askatuz lortzen dira. Hau da, helburu funtzioaren deribatu partzialak berdin zero egiterakoan lortzen diren ekuazioak (**ekuazio normalak**) askatuz:

$$\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\alpha}} = -2 \sum (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0 \Rightarrow \sum \underbrace{(Y_i - \hat{\alpha} - \hat{\beta} X_i)}_{\widehat{u}_i} = 0 \tag{2.5}$$

$$\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\beta}} = -2 \sum (Y_i - \hat{\alpha} - \hat{\beta} X_i) X_i = 0 \Rightarrow \sum \underbrace{(Y_i - \hat{\alpha} - \hat{\beta} X_i) X_i}_{\widehat{u}_i X_i} = 0 \tag{2.6}$$

Ekuazio normal biak askatuz KTAko estimatzaileen adierazpenak lortzen ditugu:

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \tag{2.7}$$

$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{S_{YX}}{S_X^2} \tag{2.8}$$

Ekuazio normalek, estimatzaileen adierazpenak ateratzeko informazioaz gain, informazio gehiago eskaintzen dute. Hasteko, lehen ekuazio normalak (2.5) hondarren batura zero dela adierazten du, hortaz bere batezbestekoa ere zero izango da: $\bar{\hat{u}} = \frac{1}{N} \sum \hat{u}_i = 0$. Bigarren ekuazio normalak (2.6), hondarrak ($\widehat{u}$) aldagai azaltzailearekiko (X) ortogonalak direla ($\sum \widehat{u}_i X_i = 0$) adierazten duenez, beraien arteko lagin kobariantza zero izango da:

$$S_{\hat{u},X} = N^{-1} \sum_{i=1}^{N} (X_i - \bar{X})(\hat{u}_i - \bar{\hat{u}}) = N^{-1} \left(\sum_{i=1}^{N} \widehat{u}_i X_i - N \bar{X} \bar{\hat{u}} \right) = 0 \tag{2.9}$$

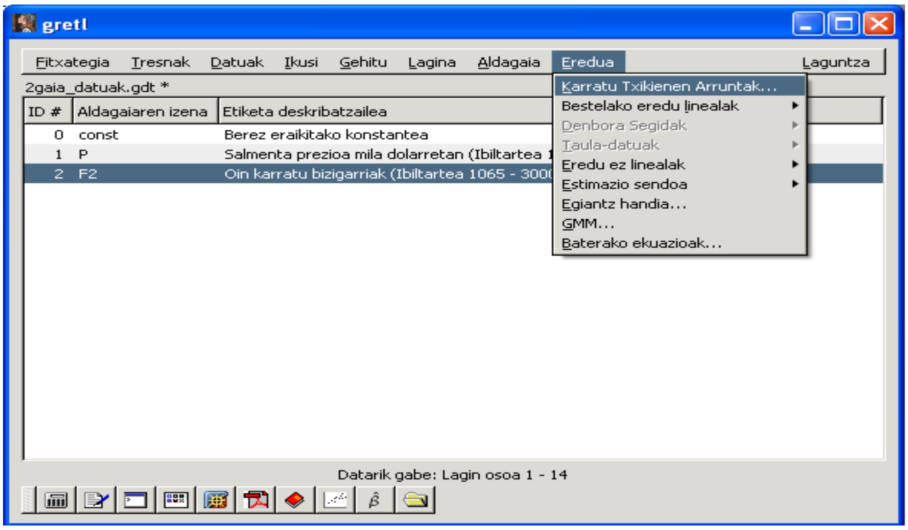
eta ondorioz beraien arteko koerlazio koefizientea ere zero izango da, hau da: $r_{\hat{u},X} = S_{\hat{u},X} / S_{\hat{u}} S_X = 0$. Beraz, eredu bat estimatu ondoren azaldu gabe gelditzen den zatian, hots hondarretan ($\hat{u}$), ez dago X aldagaiak azaldu dezakeenik.

Ondoren adibidera itzuliko gara emandako kontzeptu berriak argitzeko asmoz. Estima dezagun ondorengo eredu bakuna:

$$P_i = \alpha + \beta F2_i + u_i \qquad i = 1, \ldots 14 \tag{2.10}$$

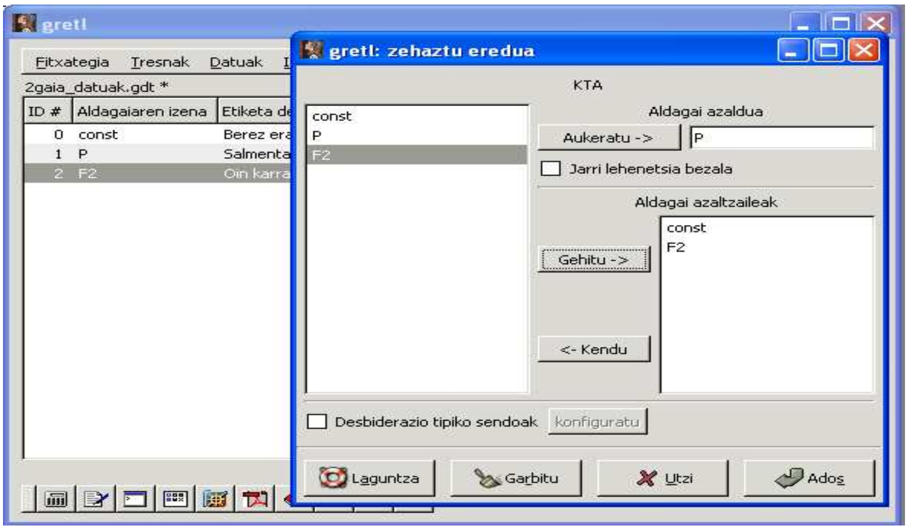
non α eta β koefizienteen interpretazioak jadanik eginda dauzkagun. Eredua karratu txikienen arruntetako metodoaren bitartez estimatzeko, *Eredua* aukeratu eta gero *Karratu Txikienen Arruntak...*

2.11 Irudia: Karratu Txikienen Arruntetako estimatzailea



hemen aldagai azaldu bezala etxebizitzaren prezioa P hartzen dugu eta aldagai azaltzaile bezala etxebizitzaren azalera oin karratutan $F2$:

2.12 Irudia: Ereduaren zehazpena



KTAko estimazioaren emaitzak honako taulan agertzen dira:

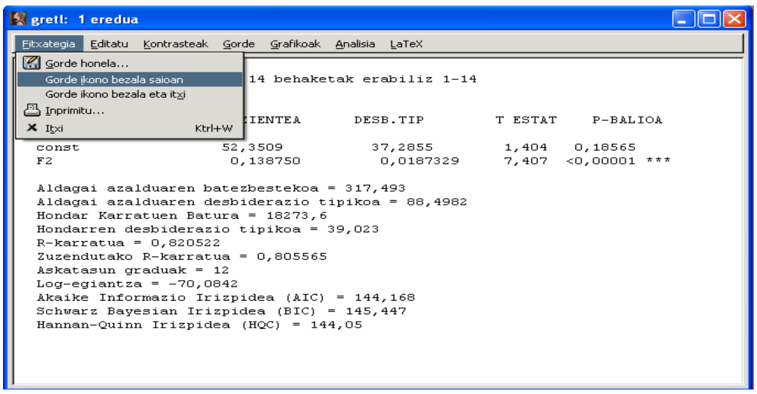
Eredua 1: KTA estimazioak 14 behaketak erabiliz 1–14
Aldagai azaldua: P

Aldagaia	Koefizientea	Desb. Tipikoa	t-estatistikoa	p-balioa
const	52,3509	37,2855	1,4041	0,1857
F2	0,138750	0,0187329	7,4068	0,0000

Aldagai azalduaren batezbestekoa	317,493
Aldagai azalduaren Desb. Tip.	88,4982
Hondar Karratuen Batura	18273,6
Hondarren desbideratze tipikoa ($\hat{\sigma}$)	39,0230
R^2	0,820522
Zuzendutako $\bar{R}^2$	0,805565
Askatasun graduak	12
Log-egiantza	-70,084
Akaike Informazio Irizpidea	144,168
Schwarz Bayesian Irizpidea	145,447
Hannan–Quinn Irizpidea	144,050

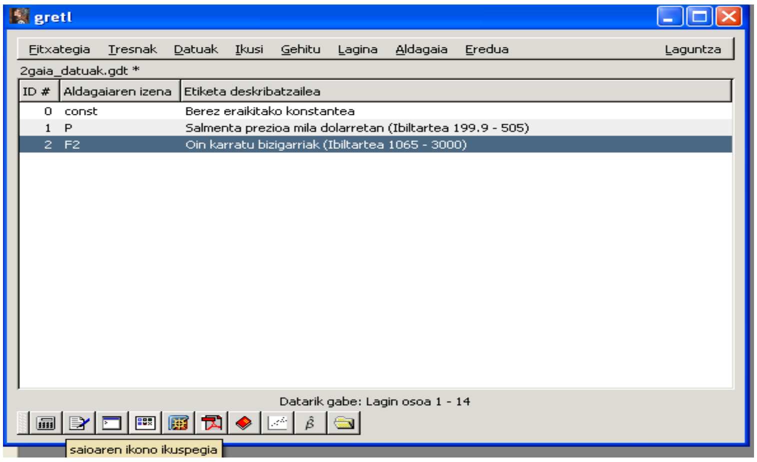
Emaitzak aztertzen hasi baino lehen, eredua ikono bezala gorde egingo dugu gero beranduago berreskuratzeko asmoarekin. Gordetzeko *Fitxategia* aukeratu eta *Gorde ikono bezala saioan* klikatu:

2.13 Irudia: Ikono bezala gordetzen



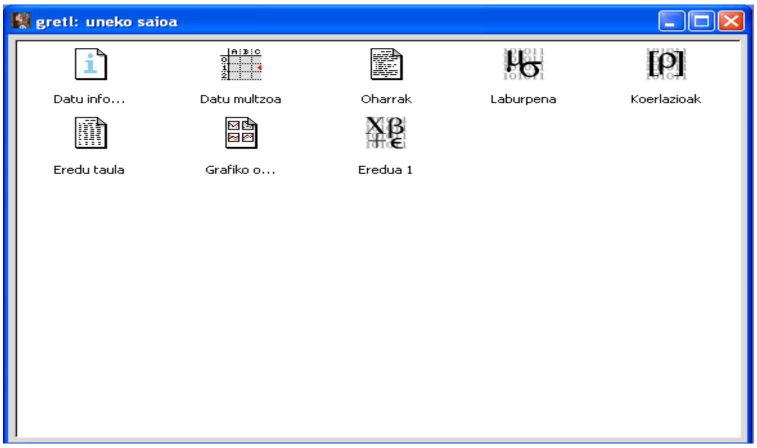
Hau egiterakoan eredua ikono bezala gordeta gelditzen da `GRETL` programaren barneko `USER` karpetan. Berreskuratzeko ezkerreko behealdean agertzen den ikonoen menuan (ezkerretik hasita laugarren laukia) aukeratu *saioaren ikono ikuspegia*

2.14 Irudia: Estimazio emaitzak ikonoetatik berreskuratzen



eta *Eredua 1* jartzen duen ikonoari klik-bikoitza emanik ereduen estimazio emaitzak berreskuratzen dira: Beranduago, beste ereduren bat estimatu eta ikono bezala gordetzen badugu

2.15 Irudia: Ikonoen ikuspegia



Gretlek *Eredua 2* izena jarriko dio.

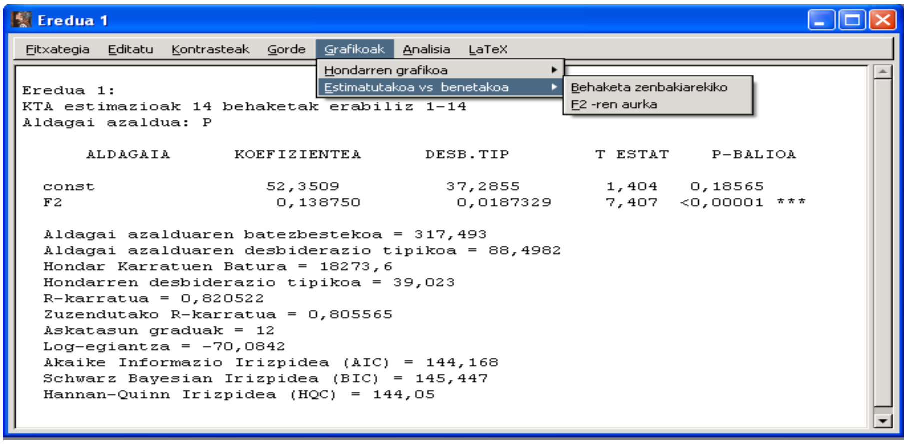
Emaitzen taulara itzuliz lehen zutabean ereduan barneratutako aldagai azaltzaileen izenak agertzen dira: termino konstantea (const) eta etxebizitzen azalera oin karratutan ($F2$). Bigarren zutabean, aldagai azaltzaile bakoitzari dagokion koefizientearen KTAko estimazioak dauzkagu. Honela, lagin erregresio funtzioa lortu dezakegu:

$$\widehat{P}_i = 52,3509 + 0.138750\ F2_i \qquad i = 1, \ldots 14. \tag{2.11}$$

Estimatutako koefizienteak interpretatuz $\widehat{\alpha} = 52,3509$ azalera gabeko etxebizitzen batezbesteko prezio **estimatua** da mila dolarretan edo bestela "hasierako batezbesteko prezioa" estimatua. Etxebizitzaren azalera oin karratu batean handitzerakoan espero dugun prezioaren gehikuntza **estimatua** $\widehat{\beta} = 0,138750$ mila dolarretakoa da edota $0,138750 \times 1000 = 138,750$ dolarretakoa. Hortaz espero genituen zeinuak atera dira.

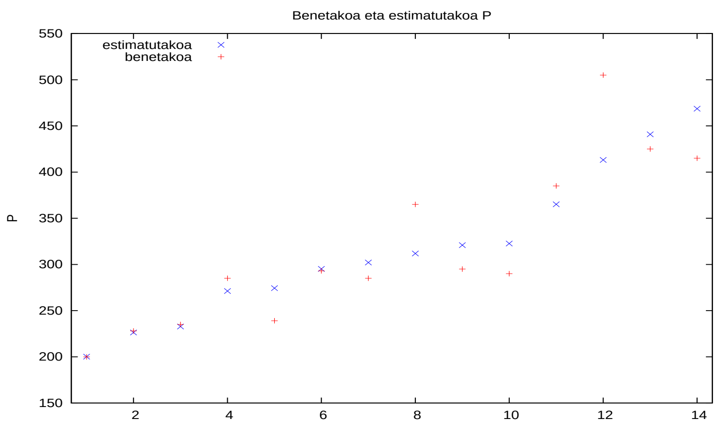
Ondoren interesgarriak diren grafikoak aterako ditugu. Horretarako *Grafikoak* aukeratu eta ondoren $\rightarrow$ *Estimatutakoa vs benetakoa* atalean

2.16 Irudia: Estimatutako vs benetako aldagai azaldua



grafikoa irudikatzeko bi aukera agertuko zaizkigu. Bata *Behaketa zenbakiarekiko* eta bestea *F*2 *-ren aurka*. Lehenengokoa hartzerakoan irteten zaigun grafikoa honakoa da

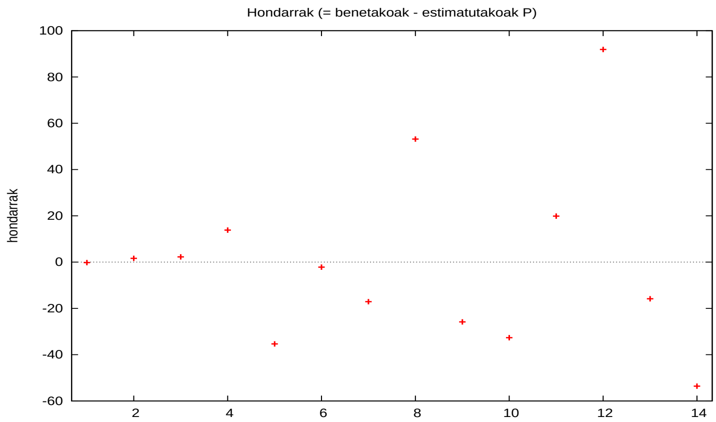
2.17 Irudia: Estimazio emaitzen erakuspena



zeinetan aldagai azalduaren benetako balioak (P_i) eta estimatutakoak ($\widehat{P}_i$) agertzen diren. Grafikoa begiratzerakoan badirudi, urruntzen diren hiru-lau puntutan ezik, nahiko ondo hurbildu garela eta estimazioa nahiko ona dela.

Bestalde, *Grafikoak* atalean *Hondarren grafikoa* → *Behaketa zenbakiarekiko* aukeratzen badugu lortzen den grafikoa honakoa da Grafiko honetan hondarrak zero balioaren inguruan

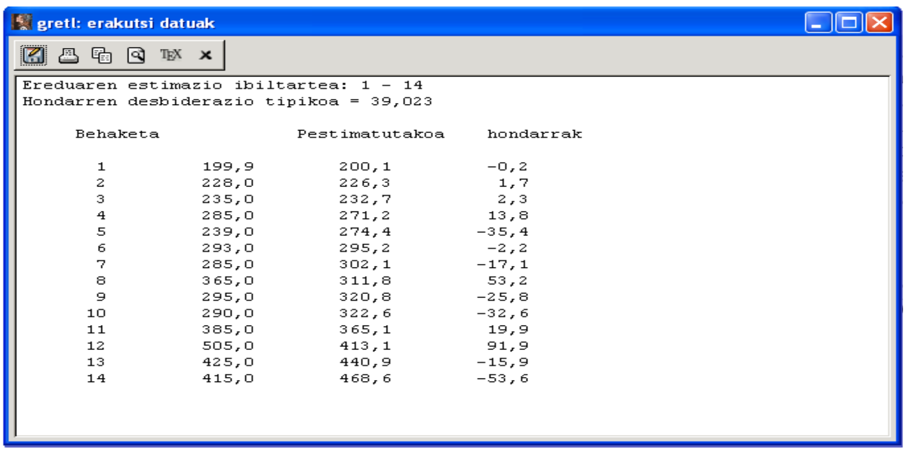
2.18 Irudia: Hondarren grafikoa



banaturik daudela ikus dezakegu eta horrela irten behar da $\bar{\widehat{u}} = 0$ delako. Baina badirudi laginean zehar mugitzen goazen neurrian hondarrak duten sakabanatzea handitzen doala. Emaitza honek arazoren bat dagoela adierazten digu. Posible litzateke homozedastizitatearen oinarrizko hipotesia ez mantentzea gure datuentzat eta hortaz perturbazioak heterozedaztikoak izatea, edo eredua ondo zehaztuta dagoelaren oinarrizko hipotesia ez betetzea, sakabanatze gorakorra duen aldagai nabariren baten omisioa izatea.

Estimatutako balioak eta hondarrak grafikoki ikusteaz gain, hartzen dituzten balioak zeintzuk diren jakiteko, *Analisia* aukeran *Erakutsi benetakoak, estimatutakoak, hondarrak* aukeratuz, ateratzen den emaitza honakoa izango da:

2.19 Irudia: Estimazio emaitzen erakuspena


gretl: erakutsi datuak

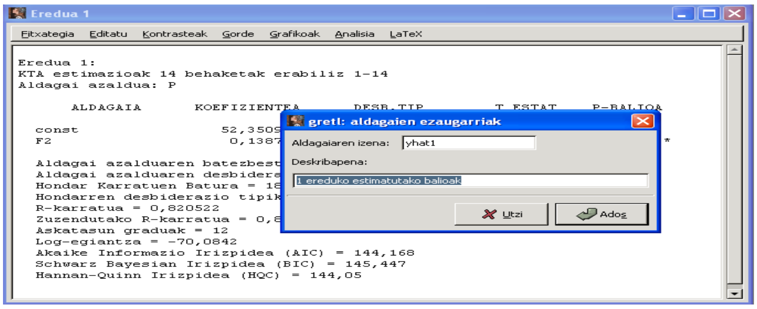
Ereduaren estimazio ibiltartea: 1 - 14
Hondarren desbiderazio tipikoa = 39,023

Behaketa		Pestimatutakoa	hondarrak
1	199,9	200,1	-0,2
2	228,0	226,3	1,7
3	235,0	232,7	2,3
4	285,0	271,2	13,8
5	239,0	274,4	-35,4
6	293,0	295,2	-2,2
7	285,0	302,1	-17,1
8	365,0	311,8	53,2
9	295,0	320,8	-25,8
10	290,0	322,6	-32,6
11	385,0	365,1	19,9
12	505,0	413,1	91,9
13	425,0	440,9	-15,9
14	415,0	468,6	-53,6

Bigarren zutabeko balioak aldagai azalduarenak dira. Hirugarren zutabekoak berriz, estimatutako balioak dira eta hauek ateratzeko $\widehat{P}_i = 52,3509 + 0.138750\ F2_i$ erabili da. Azken zutabeko balioak ateratzeko berriz, $\widehat{u}_i = P_i - \widehat{P}_i$ erabili da, hau da, bigarren eta hirugarren zutabe bien arteko kenketa.

Balio hauek gorde nahi izanez gero *Gorde* barneko aukeren artean *Estimatutako balioak* hartu eta honako leihatila aterako zaigu: non aldagai azaldu estimatuarentzat "yhat1" izena

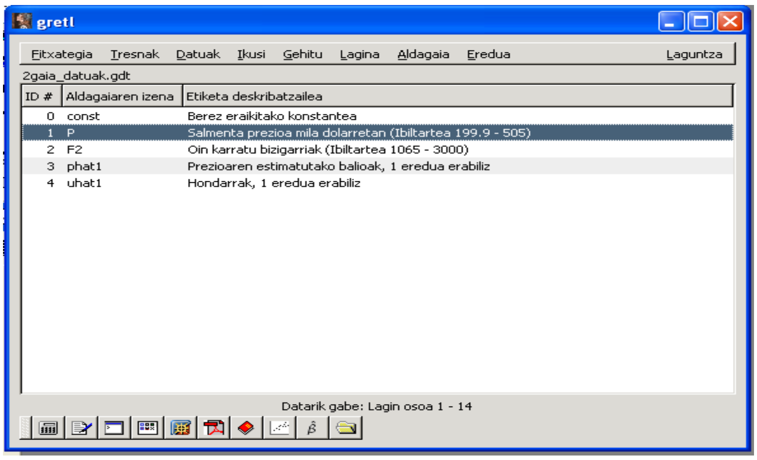
2.20 Irudia: Estimatutako balioak gordetzen



eta deskribapenean "1 ereduko estimatutako balioak" berez agertzen diren. Dena den izen eta deskribapen hauek aldatu ditzakegu. Adibide honetako aldagai azaldua prezioa (P) denez, izen bezala phat1 jarriko dugu eta deskribapenean Prezioaren estimatutako balioak, 1 eredua erabiliz. Horrela "1 eredua" deskribapenean mantentzean, ikono bezala gordeta dauden estimatutako balioak zein eredutik lortu diren jakitea errazagoa da.

Aurreko pausuak hondarrentzat errepikatzerakoan, irtetzen den lehatilan berriz "uhat1" agertzen da hondarren izentzat eta deskribapen bezala "1 ereduaren hondarrak". Hortaz, azken hau aldatuko dugu Hondarrak, 1 eredua erabiliz jarriz. Behin gordeta daudenean, P aldagai azaldua eta $F2$ aldagai azaltzailearekin batera agertuko dira.

2.21 Irudia: Gordetako datuen erakuspena



Ondoren bestelako informazioa aztertuko dugu, dauzkagun aldagai guztien estatistikoak hain zuzen. Horretarako, P, $F2$, phat1 eta uhat1 aldagaiak aukeratu (azpimarratu) ondoren, *Estatistikoen laburpena* klikatzean honako taula irtetzen da:

2.2 Taula: Aldagai guztien estatistiko nagusiak

Estatistikoen laburpena, 1 - 14 behaketak erabiliz

Aldagaia	BATEZ.	MEDIANA	MIN	MAX
P	$317,493$	$291,500$	$199,900$	$505,000$
F2	$1910,93$	$1835,00$	$1065,00$	$3000,00$
phat1	$317,493$	$306,958$	$200,120$	$468,602$
uhat1	$0,000000$	$-1,1919$	$-53,601$	$91,8983$

Aldagaia	D.T.	A.K.	ASIM	KURT. SOB
P	$88,4982$	$0,278741$	$0,653457$	$-0,529833$
F2	$577,757$	$0,302344$	$0,485258$	$-0,672125$
phat1	$80,1640$	$0,252491$	$0,485258$	$-0,672125$
uhat1	$37,4921$	$6,15597e+15$	$1,02687$	$0,817927$

Emaitzak aztertzerakoan:

1. Lehen ekuazio normalatik (2.5) ondorioztatzen den bezala eta lehen zutabean ikus daitekeenez, hondarren, ("uhat1" bezala izendaturik) batezbesteko aritmetikoa (gorriz) zero da: $\bar{\hat{u}} = 0$.

2. Aldagai azalduaren P eta estimatutako aldagai azalduaren $\widehat{P}$ ("phat1") batezbesteko aritmetikoak (laranjaz) berdinak dira: $\bar{P} = \bar{\widehat{P}}$.

3. Estimatutako aldagai azalduaren ($\widehat{P}$) formaren ezaugarriak, asimetria (urdin ilunez) eta kurtosis (urdin argiz), aldagai azaltzailetik ($F2$) eratortzen ditu. Honen arrazoia honakoa da: estimatutako aldagai azalduaren balioak ($\hat{P}_i$), aldagai azaltzailearen balioei ($F2_i$) eskala aldaketa bat ($\times\hat{\beta}$) eta jatorri aldaketa bat ($+\hat{\alpha}$) eginez lortzen dira, eta estimatutako koefizienteen ($\hat{\beta}$ eta $\hat{\alpha}$) balioek ez dute formako estatistikoentzat inolako informaziorik eskaintzen.

Dauzkagun aldagai guztien koerlazio matrizea ateratzeko aldagai guztiak sailkatu ondoren, *Ikusi* → *Koerlazio matrizea* klikatu eta irtetzen den matrizea ondorengoa da:

2.3 Taula: Aldagai guztien koerlazio matrizea

Koerlazio Koefizienteak, 1 - 14 behaketak erabiliz
%5eko esanguratasuna (alde bikoa) = 0,5324 n = 14 -rentzat

P	F2	phat1	uhat1	
1,0000	0,9058	0,9058	0,4236	P
	1,0000	1,0000	−0,0000	F2
		1,0000	−0,0000	phat1
			1,0000	uhat1

Emaitzak aztertuz:

1. Bigarren ekuazio normaletik (2.6) eratorritako emaitzak adierazten duen bezala, hondarren ("uhat") eta aldagai azaltzailearen ($F2$) arteko koerlazioa (urdin ilunez) zero da ($r_{\hat{u},F2} = 0$) zeren hondarrak eta aldagai azaltzailea ortogonalak baitira eta hondarren batezbesteko aritmetikoa zero baita.

2. Aldagai azalduaren (P) koerlazioa estimatutako aldagai azalduarekiko ($\widehat{P}$) eta aldagai azaltzailearekiko ($F2$) (orlegi ilunez) berdina da: $r_{P,\widehat{P}} = r_{P,F2}$ (zeren $\hat{P} = \hat{\beta}F2$ baita).

3. Estimatutako aldagai azalduaren ($\widehat{P}$) eta aldagai azaltzailearen ($F2$) arteko koerlazioa (laranjaz) bat da: $r_{\widehat{P},F2} = 1$ (zeren $\hat{P} = \hat{\beta}F2$ baita).

4. Estimatutako aldagai azalduaren ($\widehat{P}$) eta hondarren ($\widehat{u}$) arteko koerlazioa (gorriz) zero da ($r_{\widehat{u},\widehat{P}} = 0$) zeren $r_{\widehat{u},F2} = 0$ baita.

Teoriara itzuliz, oinarrizko hipotesien menpean lortutako KTAko estimatzaileen propietateak zeintzuk diren jakitea komeni da. Alde batetik, KTAko estimatzaileak **perturbazioekiko linealak** dira, hau da, bai $\widehat{\alpha}$ eta $\widehat{\beta}$ ere, perturbazioen ($u_1, \dots u_N$) konbinazio lineal bat bezala idatzi daitezke Bestaldetik KTAko estimatzaile hauek alboragabeak direnez,

$$E(\hat{\alpha}) = \alpha \qquad\qquad E(\hat{\beta}) = \beta$$

batezbestekoaren inguruan (zentraturik) banatzen dira. Azkenik, estimazioaren zehaztasunari buruz, Gauss-Markoven teoremaren arabera, estimatzaile lineal (perturbazioekiko) eta alboragabe guztien artetik KTAko estimatzaileak **bariantza txikienekoak** dira. Eredu bakunean estimatzaileen populazio edo benetako bariantzen adierazpenak honakoak dira:

$$bar(\hat{\alpha}) = \sigma^2 \left(\frac{\sum_{i=1}^N X_i^2}{N \sum_{i=1}^N (X_i - \bar{X})^2} \right) \tag{2.12}$$

$$bar(\hat{\beta}) = \sigma^2 \left(\frac{1}{\sum_{i=1}^N (X_i - \bar{X})^2} \right). \tag{2.13}$$

Bariantza biak perturbazioen populazioko bariantzaren ($bar(u_i) = \sigma^2$) menpekoak dira eta baita aldagai azaltzailearen (X) dispertsioaren menpekoak (S_X^2). Zenbat eta dispertsio

($\sum_{i=1}^{N}(X_i - \bar{X})^2$) handiagoa izan, aldagai azaltzaileak duen informazioa gero eta handiagoa izaten da. Ondorioz bariantza txikiagoak lortzen dira. Estimatzaileen arteko kobariantzari dagokionez,

$$kob(\hat{\alpha}, \hat{\beta}) = \sigma^2 \left(\frac{-\bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \right) \quad (2.14)$$

ez da zero izango baldin eta aldagai azaltzailearen batezbesteko aritmetikoa desberdin zero bada eta gainera bere zeinua batezbestekoaren aurkakoa izango da.

Populazio momentu hauen estimazioak lortzeko, lehendabiziko pausua perturbazioaren bariantza estimatzea da. Erabiliko dugun estimatzaile alboragabea honako da

$$\widehat{bar}(u_i) = \hat{\sigma}^2 = \frac{1}{N-K}\sum_{i=1}^{N}\hat{u}_i^2$$

non $\sum_{i=1}^{N}\hat{u}_i^2$ **hondar karratuen batura** den eta $N-K$ dauzkagun askatasun graduak. Honela izanik, KTAko estimatzaileen bariantzen eta kobariantzaren estimatzaile alboragabeak hurrengoak dira:

$$\widehat{bar}(\hat{\alpha}) = \frac{\sum_i^N \widehat{u}_i^2}{N-K}\left(\frac{\sum_{i=1}^{N} X_i^2}{N\sum_{i=1}^{N}(X_i - \bar{X})^2} \right) \quad (2.15)$$

$$\widehat{bar}(\hat{\beta}) = \frac{\sum_i^N \widehat{u}_i^2}{N-K}\left(\frac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \right) \quad (2.16)$$

$$\widehat{kob}(\hat{\alpha}, \hat{\beta}) = \frac{\sum_i^N \widehat{u}_i^2}{N-K}\left(\frac{-\bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \right) \quad (2.17)$$

Adibideko emaitzetara itzuliz:

Eredua 1: KTA estimazioak 14 behaketak erabiliz 1–14
Aldagai azaldua: P

Aldagaia	Koefizientea	Desb. Tipikoa	t-estatistikoa	p-balioa
const	52,3509	37,2855	1,4041	0,1857
F2	0,138750	0,0187329	7,4068	0,0000

Aldagai azalduaren batezbestekoa	317,493
Aldagai azalduaren Desb. Tip.	88,4982
Hondar Karratuen Batura	18273,6
Hondarren desbideratze tipikoa ($\hat{\sigma}$)	39,0230
R^2	0,820522
Zuzendutako $\bar{R}^2$	0,805565
Askatasun graduak	12
Log-egiantza	-70,084
Akaike Informazio Irizpidea	144,168
Schwarz Bayesian Irizpidea	145,447
Hannan–Quinn Irizpidea	144,050

2.4 Taula: KTAko estimatzailearen bariantza eta kobariantza matrizea

Erregresio koefizienteen kobariantza matrizea

const	F2	
1390, 21	−0, 670583	const
	3, 50920e − 04	F2

"Desb. Tipikoa" deituriko zutabean, koefizienteen estimatzaileen estimatutako desbideratzeak (aurreko (2.15) eta (2.16) bariantza estimatuen erro positiboa) agertzen dira. Beraien arteko kobariantza estimatua zein den jakiteko bariantza eta kobariantza matrize estimatua kalkulatu behar da. Horretarako: *Analisia* → *Koefizienteen kobariantza matrizea* aukeratu eta ateratzen den emaitza honakoa da:

estimatzen ari garen matrizea ondorengoa dela jakinik

$$\widehat{Bar}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix} = \begin{pmatrix}\widehat{bar}(\widehat{\alpha}) & \widehat{kob}(\widehat{\alpha},\widehat{\beta})\\ \widehat{kob}(\widehat{\alpha},\widehat{\beta}) & \widehat{bar}(\widehat{\beta})\end{pmatrix} = \begin{pmatrix}1390,21 & -0,670583\\ & 0,000350920\end{pmatrix}$$

diagonal nagusian bariantza estimatuak agertzen dira eta diagonal nagusitik kanpo dagoen elementua estimatutako kobariantzari dagokio. Estimatutako momentu hauek kalkulatzerakoan erabili den perturbazioaren bariantza estimatua $\hat{\sigma}^2 = \frac{\sum_i^N \widehat{u}_i^2}{N-K}$ ez da zuzenean ematen, bere erro karratua baizik, hau da, perturbazioaren desbideratze estimatua. Honela, bariantza desbideratzea karratura eginez kalkulatu daiteke $\hat{\sigma}^2 = 39,0230^2 = 1522,8$ edo bestela emandako formula aplikatuz $\hat{\sigma}^2 = 18273,6/12 = 1522,8$ zeren hondar karratuen batura $(18273,6)$ eta askatasun graduak $(N-K = 14-2 = 12)$ emaitzetan agertzen baitira.

Zoritxarrez, nahiz eta estimaturiko bariantzak estimazioaren zehaztasuna neurtu, ateratako emaitzak onak diren edo ez bereizteko gai ez gara. Arrazoia: bariantzak nahi ditugun bezain txikiak egitea posible da aldagai bien (dependentea eta independentea) eskala aldatuz. Ondorioz, estimatutako balioak eta benetakoak hurbil dauden edo ez jakiteko, ontasun neurriren bat behar dugu doikuntzaren egokitasuna aztertzeko.

2.4 Doikuntzaren egokitasuna

Eredu bakunaren doikuntzaren ontasuna neurtzeko erabiliko dugun neurria mugatze koefizientea izango da (R-karratua). Koefiziente honen adierazpena honakoa da:

$$R^2 = r_{X,Y}^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2}. \tag{2.18}$$

Koefiziente honek X aldagai azaltzaile gabeko eredu batetik

$$Y_i = \alpha + u_i$$

X aldagaia barneratzen duen eredu batera pasatzean

$$Y_i = \alpha + \beta X_i + u_i$$

lortzen den irabazia neurtzen du. Hortaz R^2, aldagai azaltzailearen (X) bariantzak, aldagai azalduaren (Y) bariantzaren zenbateko portzentaia azaltzen duen neurtzen du era lineal batean. Adibideko mugatze koefizientea analizatzean $R^2 = 0,820522$, etxebizitzaren azaleraren ($F2$) bariantzarekin salmenta prezioaren bariantzaren $\%82,0522$a azaltzen dela era lineal batean esango genuke.

2.5 Esanguratasun analisia eta konfidantza tarteak

Ereduko X aldagai azaltzailea, aldagai azaldua azaltzeko nabaria den baieztatzeko kontraste bat egin dezakegu. Horretarako, hipotesi hutsean aldagai azaltzaileari laguntzen dion koefizientea berdin zero jarriko genuke eta aurkakoan desberdin zero dela:

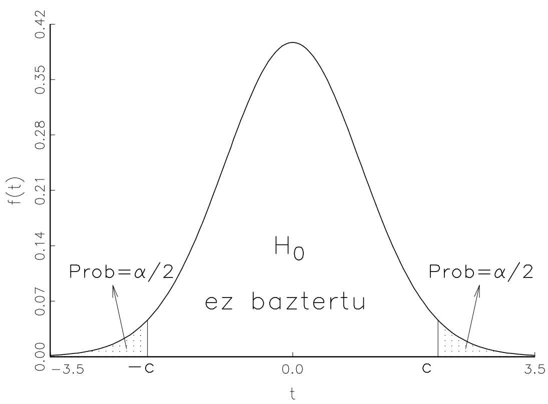
$$\begin{cases} H_0 : \beta = 0 & \text{(X aldagai azaltzailea ez da nabaria)} \\ H_a : \beta \neq 0 & \text{(X aldagai azaltzailea nabaria da)} \end{cases}$$

Kontraste honetarako estatistikoa honakoa litzateke:

$$t_{est} = \frac{\widehat{\beta} - \beta}{\widehat{des}(\widehat{\beta})} \overset{H_0}{\sim} t_{(N-K)}$$

eta erabaki araua hurrengoa izango da:

2.22 Irudia: Erabaki araua



kalkulatu behar den estatistikoaren balioa $[-c = -t_{(N-K)\alpha/2}, c = t_{(N-K)\alpha/2}]$ balioen artean badago, konfidantza tartean aurkitzen denez, ez dugu hipotesi hutsa baztertuko erabilitako esangura-mailarentzat eta konklusio bezala, X aldagai azaltzailea Y aldagai azaldua azaltzeko aldagai nabari bat **ez** dela esango dugu. Bestalde, kalkulatutako estatistikoaren balioa $-c = -t_{(N-K)\alpha/2}$ balioa baino txikiagoa bada, edota $c = t_{(N-K)\alpha/2}$ balioa baino handiagoa, orduan eskualde kritikoan erortzen denez, hipotesi hutsa baztertu egingo dugu erabilitako esangura-mailarekin eta konklusio bezala, X aldagai azaltzailea, aldagai azaldua azaltzeko aldagai nabaria dela baieztatuko dugu.

Adibidera itzuliz, ikus dezagun ea etxebizitzaren azalera aldagai nabaria den etxebizitzen salmenta prezioa zehazteko:

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

Kalkulatu behar den estatistikoaren balioa Gretlek emandako estimazio emaitzetan agertzen denez, egin behar den gauza bakarra tauletako balioarekin ($t_{(14-2)0,05/2} = 2,179$) konparatzea da. Kasu honetan $7,4068 > 2,179 = t_{(14-2)0,05/2}$ ematen denez, estatistikoaren balioa eskualde kritikoan erortzen da eta ondorioz, hipotesi hutsa baztertu egingo dugu %5eko esangura-mailarekin. Konklusio bezala, etxebizitzaren azalera etxebizitzen salmenta prezioa azaltzeko aldagai azaltzaile nabari bat dela esango genuke.

Behin $\beta \neq 0$ dela eta beraz $F2$ aldagai nabari bat dela jakinik, koefizienteak har dezakeen balioa jakitea interesagarria izan daiteke. Adibidez:

• Etxebizitzaren azalera oin karratu batean handitzerakoan, posible litzateke etxebizitzaren batezbesteko salmenta prezioa 100 dolarretan gehitzea? Erantzuteko ondorengo kontrastea burutuko dugu:

$$\begin{cases} H_0 : \beta = 0,1 \\ H_a : \beta \neq 0,1 \end{cases} \qquad t_{est} = \frac{\widehat{\beta} - \beta}{\widehat{des}(\widehat{\beta})} \overset{H_0}{\sim} t_{(N-K)}$$

Kontuan izan behar da lehen erabilitako estatistikoaren balioak (estimazio taulan agertzen dena) banakako esanguratasuna kontrastatzeko soilik balio duela. Beraz, beste kontraste hau egin nahi izanez gero, estatistikoaren balioa berriro kalkulatu behar da $t_{est} = \frac{0,138750-0,1}{0,0187329} = 2,068$, eta tauletako balioa izanik ($t_{(14-2)0,05/2} = 2,179$), lortzen den emaitza $-t_{(14-2)0,05/2} = -2,179 < 2,068 < 2,179 = t_{(14-2)0,05/2}$ da. Estatistikoaren balioa konfidantza tartean erortzen denez hipotesi hutsa ez da baztertzen %5eko esangura-mailarekin eta ondorioz etxebizitza baten azalera oin karratu batean handitzerakoan, bere batezbesteko prezioan 100 dolarreko gehikuntza dakarrela esan dezakegu.

• Eta azaleraren oin karratu bateko gehikuntza berdinarekin, posiblea litzateke etxebizitzaren batezbesteko salmenta prezioa 150 dolarretan gehitzea? Ikus dezagun:

$$\begin{cases} H_0 : \beta = 0,15 \\ H_a : \beta \neq 0,15 \end{cases} \qquad t_{est} = \frac{\widehat{\beta} - \beta}{\widehat{des}(\widehat{\beta})} \overset{H_0}{\sim} t_{(N-K)}.$$

Estatistikoaren balioa kasu honetan $t_{est} = \frac{0,138750-0,15}{0,0187329} = -0,6005$ konfidantza tartean erortzen denez ($-2,179 < -0,6005 < 2,179$) hipotesi hutsa ez da baztertzen %5eko esangura-mailarekin. Orduan, etxebizitza baten azalera oin karratu batean handitzerakoan, bere batezbesteko prezioan 150 dolarreko gehikuntza dakarrela baieztatu dezakegu.

Zer gertatzen ari da? Puntuzko estimazioak β koefizientearen balio posible bat ematen digu, baina errealitatean balio posibleen multzo bat daukagu. Balio posible hauek lortzeko koefizientearen konfidantza tartea atera behar dugu. Goiko kontrasteak aztertuz, erraz ikus daiteke, estatistikoaren balioa $(-2,179; 2,179)$ tartearen barnean erortzea posible egiten duten balio guztiak, koefizientearen balio posible guztiak izango direla.

Teorikoki koefiziente baten konfidantza tartea ateratzeko, Student-t banaketaren taulako $(-t_{(N-K)\alpha/2}\ ;\ t_{(N-K)\alpha/2})$ bi balioen artean egon ahal diren estatistikoaren balio posible guztiak kalkulatuz lortzen da:

$$Pr\left[-t_{(N-K)\alpha/2} < \frac{\widehat{\beta}-\beta}{\widehat{des}(\widehat{\beta})} < t_{(N-K)\alpha/2}\right] = 1-\alpha \tag{2.19}$$

askatuz:

$$Pr\left[\widehat{\beta} - t_{(N-K)\alpha/2}\widehat{des}(\widehat{\beta}) < \beta < \widehat{\beta} + t_{(N-K)\alpha/2}\widehat{des}(\widehat{\beta})\right] = 1-\alpha \tag{2.20}$$

$\%(1-\alpha)$-ri dagokion konfidantza tartea ateratzen da. Tarte hau puntuzko estimazioan zentraturik dago eta zentru honetatik urruntzen den kantitatea jakiteko $t_{(N-K)\alpha/2}$ aldiz estimatzailearen desbideratze tipiko estimatua $(\widehat{des}(\widehat{\beta}))$ kalkulatu behar dugu. Hemendik aurrera, konfidantza tarte bat idazteko erabiliko dugun notazioa hurrengoa izango da:

$$KT(\beta)_{1-\alpha} = \left[\hat{\beta} \pm t_{(T-K)\alpha/2}\ \widehat{des}(\hat{\beta})\right].$$

Bestalde, α koefizienteari dagokion konfidantza tartea era berdinean lortzen da:

$$KT(\alpha)_{1-\alpha} = \left[\hat{\alpha} \pm t_{(T-K)\alpha/2}\ \widehat{des}(\hat{\alpha})\right].$$

Adibidearekin bukatzeko, koefizienteen konfidantza tarteak aterako ditugu, horretarako, *Analisia → Koefizienteen konfidantza tarteak* aukeratu eta ateratzen dena honakoa da:

2.5 Taula: Tartezko estimazioa

```
t(12, .025) = 2,179

     ALDAGAIA      KOEFIZIENTEA       %95eko KONFIDANTZA TARTEA

         const         52,3509        (-28,8872, \ 133,589)
            F2          0,138750      (0,0979349, \ 0,179566)
```

Lortzen den emaitzan, bigarren zutabeko balioak puntuzko estimazioak dira, $\widehat{\alpha} = 52,3509$ eta $\widehat{\beta} = 0,138750$ eta hirugarren zutabean, %95eko konfidantza tartearen behe eta goi kotak agertzen dira, hau da:

$$KT(\alpha)_{0,95} = [-28,887\ ;\ 133,587]$$

$$KT(\beta)_{0,95} = [0,0979349\ ;\ 0,179566]$$

Beraz, etxebizitzaren azaleran oin karratu bateko gehikuntza batek, bere batezbesteko salmenta prezioa 97,9349 eta 179,566 dolar bitarteko gehikuntza ekarriko duela esan dezakegu %95eko konfidantzarekin.

2.6 Laburpena. Emaitzen aurkezpena

Eredu baten estimazio emaitzak laburbildu daitezke, lagineko erregresio funtzioa eta emaitzak ebaluatzeko baliogarriak diren estatistiko multzo bat eskeiniz. Ohituraz, emaitzak bi eratara aukezten dira:

Lehen era:

$$\underset{(\widehat{des})}{\widehat{P}} = \underset{(37{,}285)}{52,3509} + \underset{(0{,}018733)}{0,138750}\, F2$$
$$N = 14 \qquad R^2 = 0,82 \qquad \hat{\sigma} = 39,023$$

non koefiziente estimatu bakoitzaren azpian desbideratze tipiko estimatua agertzen den.

Bigarren era:

$$\underset{(t_{est})}{\widehat{P}} = \underset{(1{,}404)}{52,3509} + \underset{(7{,}407)}{0,138750}\, F2$$
$$\text{Askatasun graduak} = 12 \qquad R^2 = 0,82 \qquad \hat{\sigma} = 39,023$$

non koefiziente estimatu bakoitzaren azpian t-estatistikoa agertzen den.

Bibliografia

Ramanathan, R. (2002), *Introductory Econometrics with Applications*, 5. ed., South-Western, Ohio.