A Eranskina: Errepasoak

A.1 Probabilitatearen errepasoa

Zorizko saiakuntza baten ondorioz lortutako datuen interpretazioan oinarritzen dira metodo ekonometrikoak. Aldagai ekonomikoek zati edo osagai sistematiko bat eta aleatorio bat izaten dute, bere behaketa eman aurretik ezin baita zehaztasun osoz hartuko dituzten balioak aurresan. Atal honetan, kurtsoan zehar aplikatuko ditugun probabilitate kontzeptuak gogoratuko ditugu: zorizko aldagaia (*aleatorioa edo estokastikoa*) eta bere propietateak, baita probabilitate banaketa erabilgarrienak ere.

A.1.1 Zorizko aldagaia edo aldagai aleatorioa

Aldagai aleatorio baten (X) balioa ez da ezagutzen bere behaketa eman aurretik eta probabilitatea, emaitzaren ziurgabetasuna adierazteko baliabide bat da. Bi motatako aldagai aleatorioak ezagutzen dira: *diskretuak*, har ditzaken balio posible guztien multzoa finitoa edo zenbagarria denean eta *jarraiak*, egite multzoa infinitoki zatigarria denean eta beraz, ez zenbagarria. Adibidez, etxebizitza baten azalera aldagai jarraia da baina komun kopurua berriz, diskretua. Orokorrean, kurtso honetan aldagai jarraiekin lan egingo dugu.

Baldin eta X aldagai diskretua bada, emaitza posible bakoitzari (x_i) probabilitate bat ezarri diezaiokegu $p(x_i) = Prob(X = x_i)$. *Probabilitate funtzioak* $p(x_i) \geq 0\ \forall x_i$ eta $\sum_i p(x_i) = 1$ bete beharko ditu.

Baldin eta X jarraia bada, puntu zehatz bati dagokion probabilitatea zero da eta X aldagaiaren balioak tarte batean egotearen probabilitateari buruz jardungo dugu, $a \leq X \leq b$.

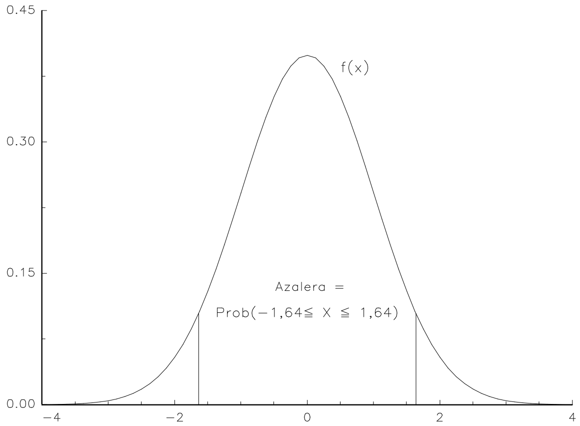
Horrela, X aldagai jarrai baten *dentsitate funtzioak*, $f(x)$, ondorengoa betetzen du:

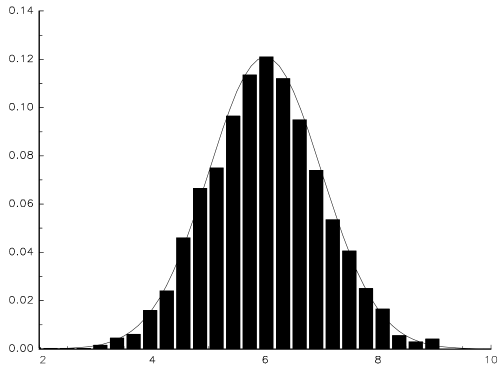
$$Probabilitatea(a \leq X \leq b) = \int_a^b f(x)\, dx$$

Hau da, a eta b bi puntuen artean funtzioaren azpian gelditzen den azalera, aldagaiak hartuko dituen balioak $[a, b]$ tartean egotearen probabilitatea da (ikusi A.1 grafikoko ezkerreko panela). Dentsitate funtzioak $f(x) \geq 0\ \forall x$ eta $\int_{-\infty}^{\infty} f(x)dx = 1$ bete beharko du.

Aldagai aleatorio jarrai baten adibide bat **aldagai Normala** da. Bere dentsitate funtzioak kanpai forma du (ikusi A.1 grafikoko ezkerreko panela). Erdiko balio baten inguruan simetrikoki banatzen diren aldagaiak egituratzeko praktikan asko erabiltzen den funtzioa da, erdiko balio horren inguruko balioetan probabilitate handiena pilatzen delarik eta gutxi urruneko balioetan. A.1 grafikoko eskubiko panelak datuen histograma eta dentsitate funtzioaren ar-

A.1 Irudia: *Normalaren* dentsitate funtzioa eta histograma.





teko erlazioa irudikatzen du. Peña eta Romo (1997) adierazten duten bezala: "*La función de densidad constituye una idealización de los histogramas de frecuencia o un* ***modelo*** *del cual suponemos que proceden las observaciones. El histograma representa frecuencias mediante áreas; análogamente, la función de densidad expresa probabilidades por áreas. Además, conserva las propiedades básicas del histograma: es no negativa y el área total que contiene es uno.* ".

Aldagai aleatorio baten banaketa funtzioa, lekuko neurriak (batezbestekoa, mediana eta moda), sakabanatze neurriak (bariantza, desbideratze tipikoa eta aldakuntz koefizientea) edota forma neurriak (asimetria eta kurtosis koefizienteak) erabiliz laburbildu daiteke. Kontzeptu hauek, datu multzoen ezaugarriak laburbiltzeko erabilitako neurrien antzera definitzen dira. Jarraian, kurtsoan zehar erabiliko ditugun beste elementuak azalduko ditugu.

X aldagai baten **batezbestekoa edo itxarondako balioa** (μ) lekuko neurri bat da. Aldagai horrek har ditzaken balio posible guztien batezbesteko ponderatua da, non ponderazioa balio bakoitzaren probabilitatea den. Aldagaia jarraia bada:

$$\mu \,=\, E(X) \,=\, \int_{-\infty}^{\infty} x\, f(x)dx$$

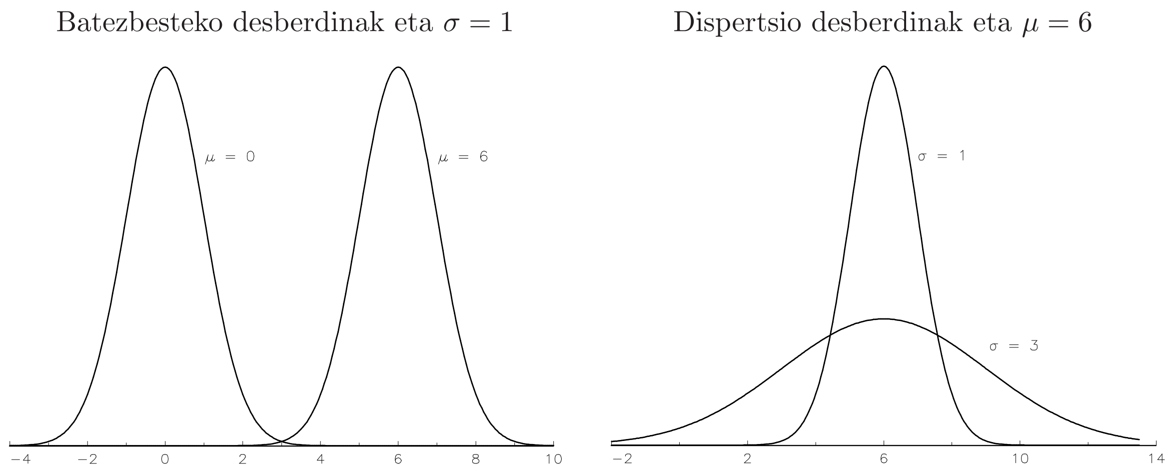
non E, itxaropenaren eragilea den. A.2 grafikoko eskubiko panelan ikusi daitekenez, itxaropen horrek aldagaiaren banaketaren zentrua jasotzen du. Horrela, itxaropena gero eta handiagoa bada, saiakuntzaren gauzatutako balioak handiagoak izatea espero da.

X aldagai aleatorio baten **bariantza**, bigarren ordenako batezbestekoarekiko momentu zentratua da. Hau da:

$$Bar(X) \,=\, \sigma_X^2 \,=\, E[(X-\mu)^2] \geq 0$$

Banaketaren sakabanatze neurri bat da eta bere erro positiboa **desbideratze tipikoa** edo

A.2 Irudia: Banaketa normalaren adibideak



desbideratze estandarra bezala ezagutzen da:

$$desb(X) = \sigma_X = \sqrt{Bar(X)}$$

A.2 grafikoko eskubiko panelak adierazten duen bezala, aldagaiaren bariantza gero eta txikiagoa bada, batezbestekoaren inguruko probabilitatea handiagoa da.

Normal estandar edo sinplearen banaketa. Banaketa Normala bere batezbesteko eta bariantzaren balioengatik zehazten da. Baldin eta Z aldagai aleatoria zero batazbestekodun eta bat barintzaduneko Normala bada, orduan Z aldagai Normal estandarra dela esaten da eta $Z \sim N(0,1)$ bezala adierazten da. Badira banaketa honen taulak, non x emaitz posible bakoitzari, puntu horretaraino metatutako probabilitatea egokitzen zaion, $Prob(X \leq x)$.

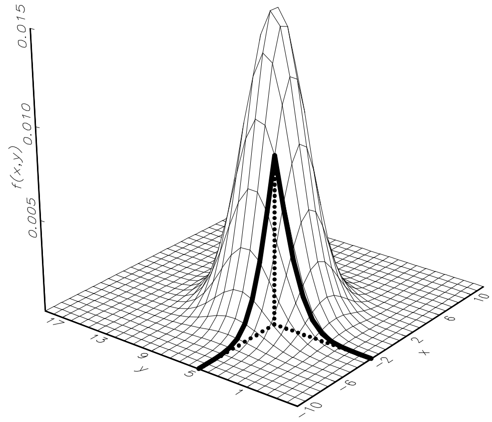
Orokorki, X aldagai Normala bada, orduan μ batezbestekoarekin eta σ^2 bariantzarekin, $X \sim N(\mu, \sigma^2)$ bezala adierazten da. $Z = (X - \mu)/\sigma)$ eraldaketa Normal estandarra izango denez, banaketa Normal honen taularekin $Prob(X \leq x)$ metatutako probabilitatea lortzen da.

A.1.2 Zorizko bi aldagai edo gehiago

Ekonometrian aldagai multzo baten propietateak ikertu nahi dira eta zorizko bi aldagai edo gehiagoei buruzko galderei erantzuteko, baterako dentsitate funtzioa ezagutu behar da. Bi aldagaiak (X eta Y) diskretuak badira, (x_i, y_j) emaitza posible bakoitzari probabilitate bat egokitu diezaiokegu, $p(x_i, y_j)$. Probabilitate multzoa *baterako probabilitate funtzioa da*, non $0 \leq p(x_i, y_j) \leq 1$ eta $\sum_i \sum_j p(x_i, y_j) = 1$ betetzen den.

Aldagaiak jarraiak badira, baterako banaketa *baterako dentsitate funtzioarekin* $(f(x,y))$ lortzen da. Bi aldagaiek banaketa Normala jarraitzen badute, baterako densitate funtzioa ondorengo grafikoan agertzen da:

A.3 Irudia: Aldagai biko banaketa normala.



Funtzio honen azpiko azalera totala probabilitate masa deitzen da eta bere balioa bat izango da, hau da, $\int_x \int_y f(x,y)\,dx\,dy = 1$. Gainera, funtzioak ez duenez balio negatiborik hartzen ($f(x,y) \geq 0$), (a,b) puntuen artean definitutako laukiak zehazten duen azalerak, X aldagaiak a baino balio handiagoak eta Y adagaiak b baino balio txikiagoak hartzeko probabilitatea neurtzen du. Hau da:

$$Probabilitatea(X \leq a,\, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y) dx\, dy$$

Adibide gisa, A.3 grafikoan markatutako azaleraren azpian jasotako bolumena, $X \leq -2$ eta $Y \leq 4.5$ izatearen probabilitatea da.

Hala ere, aldagai bakoitzaren **dentsitate funtzio marginala** ondorengo integralekin kalkulatu daitezke:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \qquad\qquad f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx \tag{A.14}$$

Zorizko bi aldagaien baterako banaketa horrela laburbildu daiteke:

- Aldagai bakoitzaren grabitate zentruak, hau da, μ_X eta μ_Y, (A.14) banaketa marginaletatik lortzen dira.
- Aldagai bakoitzaren sakabanatzea batezbestekoaren inguruan, adibidez X eta Y-ren bariantzak (σ_X^2 eta σ_Y^2), (A.14) banaketa marginaletik eratortzen dira.
- Kobariantzaren bitartez bi aldagaien arteko erlazioa jasotzen da:

  $$Kob(X,Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \tag{A.15}$$

  edota aldagaien koerlazio koefizientearen bitartez

  $Koer(X,Y) = \rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y} \in [-1,1]$. Kobariantzak eta zorizko aldagaien koerlazioak, datuekin egindako antzeko interpretazioa dute. Horrela, $\sigma_{XY} = \rho_{XY} = 0$ bada, X eta Y aldagaiak koerlatugabeak direla esaten da.

Baterako banaketa, batezbesteko bektorean (μ) eta Σ bariantza eta kobariantza matrizean laburtzen da:

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \qquad \Sigma = \begin{pmatrix} Bar(X) & Kob(X,Y) \\ Kob(X,Y) & Bar(Y) \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix} \tag{A.16}$$

Baldintzatutako banaketak. Aldagai multzo bat aztertzean, gertaera bat egin ondoren beste gertaera bat betetzearen posibilitatea ebaluatzea komeni da. Adibidez, zein da ezkondutako eta seme-alabak dituen emakume batek lan merkatuan parte hartzearen probabilitatea? **Baldintzatutako probabilitateak** horrelako galderei erantzuteko aukera ematen du. Aldagaiak diskretuak badira, X aldagaia x_i balioa denean, Y aldagaiaren baldintzatutako banaketa horrela definitzen da:

$$Prob(Y = y_j|X = x_i) = \frac{Prob(Y = y_j, X = x_i)}{Prob(X = x_i)} = \frac{p(y_j, x_i)}{\sum_j p(y_j, x_i)} \quad \text{non } Prob(X = x_i) > 0$$

Aldagaiak jarraiak badira, X aldagaiak x_i balioa hartzen duelarik ($f(x) > 0$ kasuetan) baldintzatutako Y aldagaiaren dentsitate funtzioa horrela definitzen da:

$$f(y|X = x) \;=\; \frac{f(y,x)}{f(x)}$$

Era honetan banaketa berri bat lortzen da, jada ikusitako propietateekin. Banaketa honen momentu interesgarriak, $X = x$ balioari baldintzatutako Y-ren batezbestekoa eta bariantzak dira eta $E(Y|X = x)$ eta $Bar(Y|X = x)$ bezala izendatzen dira.

Independentzia. Bi aldagai aleatorio X eta Y estatatistikoki independenteak dira edo independenteki banatuak daude, baldin eta aldagai batek hartzen duen balioa ezaguturik, ez badu besteak har dezaken balioaren informaziorik ematen. X eta Y aldagaiak independenteak badira, orduan baterako dentsitate funtzioa atalbanatu daiteke jarraian adierazten den bezala:

$$f(x,y) \;=\; f(x)f(y) \qquad -\infty < x,y < \infty$$

Gainera, $f(y|X = x) = f(y)$ betetzen da. Bestalde, X eta Y independenteak badira, orduan $Kob(X,Y) = 0$ izango da. Azkenik, X eta Y aldagaiak batera Normalki banatzen badira eta $Kob(X,Y) = 0$, orduan X eta Y independenteak dira.

Bi aldagai baino gehiago daudenean, aurreko emaitzak n aldagai aleatorioen multzoari orokortu daitezke, hau da, $X_1, X_2, \ldots, X_n$ izanik, bektore batean biltzen dira:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}.$$

Orokorrean, bere baterako banaketaren momentuak **batezbestekoen** bektorea

$$E(\mathbf{X}) = \mu = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

da eta **bariantza eta kobariantza** matrizea Σ_X da:

$$\Sigma_X = \begin{pmatrix} Bar(X_1) & Kob(X_1, X_2) & \dots & Kob(X_1, X_n) \\ Kob(X_1, X_2) & Bar(X_2) & \dots & Kob(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ Kob(X_1, X_n) & Kob(X_2, X_n) & \dots & Bar(X_n) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,n} \\ \sigma_{1,2} & \sigma_2^2 & \dots & \sigma_{2,n} \\ \vdots & \vdots & & \vdots \\ \sigma_{1,n} & \sigma_{2,n} & \dots & \sigma_n^2 \end{pmatrix}$$

non Σ_X matrizea, n ordenako matrize karratua eta definitu ez negatiboa den. Honek, diagonal nagusiko elementuak ez negatiboak direla adierazten du, $\sigma_i^2 \geq 0$, $\forall i$.

Aldagaiak elkarrekiko independenteak badira, orduan elkarrekiko koerlatugabeak dira, hau da, $\sigma_{i,j} = 0, \forall i \neq j$, eta hortaz Σ_X diagonala izango da:

$$\Sigma_X = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

Gainera, $X_1, \dots, X_n$ aldagaiek banaketa berdina jarraitzen badute, batezbesteko eta bariantza berdinekin

$$E(\mathbf{X}) = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} \qquad \Sigma_X = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}$$

orduan aldagai aleatorioak identikoki eta independenteki banatuta egongo dira μ batezbesteko eta σ^2 bariantzarekin: $X_i \sim iib(\mu, \sigma^2), \forall i = 1, \dots n$.

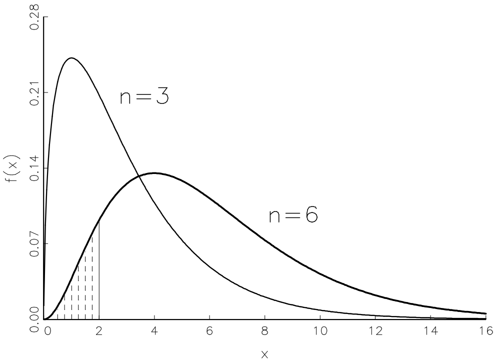
Azkenik, $X_1, \dots, X_n$ zorizko aldagai Normalak badira, $\mathbf{X}$ bektoreak **Banaketa Normal anizkoitza** jarraitzen du eta bere batezbesteko bektorea eta bariantza eta kobariantza matrizeagatik zehazten da, $\mathbf{X} \sim N(\mu, \Sigma_X)$. Aldagaiak independenteak badira, batezbesteko eta bariantza berdinarekin, $X_i \sim NIB(\mu, \sigma^2)$ idatziko dugu.

A.1.3 Zenbait probabilitate banaketa

Kurtsoan zehar, aipaturiko banaketa Normalaz gain, berekin erlazionaturiko beste banaketa batzuk erabiliko ditugu.

Chi-karratu banaketa. $(Z_1, \dots, Z_n)$, zero batezbestekoa eta bat bariantzako banaketa Normaldun zorizko aldagai independenteak badira, hau da $Z_i \sim NIB(0, 1)$, orduan $X =$

A.4 Irudia: Chi karratu banaketaren dentsitate funtzioa



$\sum_{i=1}^{n} Z_i^2$ aldagaia, n askatasun graduko chi-karratu banaketako zorizko aldagaia da eta $X \sim \chi^2(n)$ adierazten da. Bere dentsitate funtzioak ondoko itxura du:

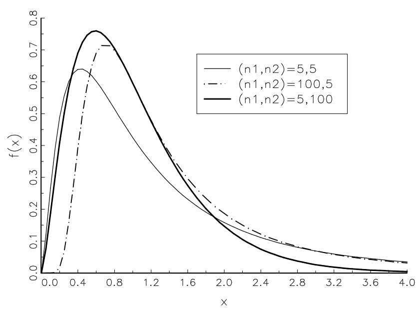
X aldagaiaren balio negatiboentzat $f(x) = 0$ dela ikusten da. Banaketa asimetrikoa da, batezbestekoa n da eta bariantza $2n$. Badira puntu bateraino metatutako probabilitatea ($Prob(X \leq x)$) jasotzen dituen taulak, hau da, n askatasun graduen funtzioan egindako grafikoko marraztutako azalera.

Snedecorren-F banaketa. Izan bitez $Z_1 \sim \chi^2(n_1)$ eta $Z_2 \sim \chi^2(n_2)$ independenteki banatutako aldagaiak, orduan $X = (n_2/n_1)(Z_1/Z_2)$ aldagaiak, n_1 eta n_2 askatasun graduko Snedecorren F banaketa jarraituko du:

$$X = \frac{Z_1/n_1}{Z_2/n_2} \sim \mathcal{F}(n_1, n_2)$$

Hurrengo grafikoak, askatasun gradu desberdinekin dentsitate funtzioak jasotzen ditu:

A.5 Irudia: Snedecorren-F banaketaren dentsitate funtzioa



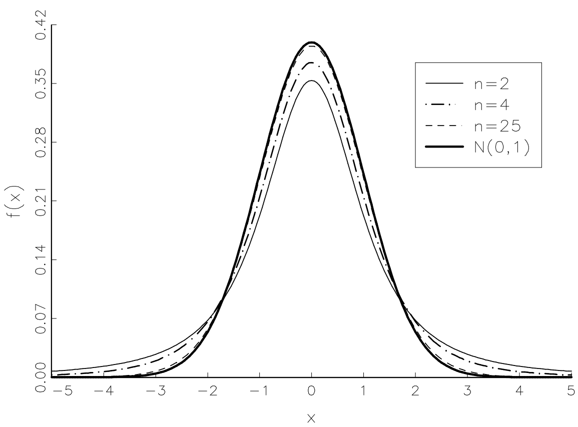
Zuzen errealaren alde positiboan ($x > 0$) metatzen da probabilitatea. Izendatzailearen askatasun graduak handitzen diren heinean, $n_2 \to \infty$, orduan $n1\mathcal{F}(n_1, n_2)$-ren banaketak $\chi^2(n_1)$ banaketara jotzen du.

Student-t banaketa. Izan bitez $Z \sim N(0,1)$ eta $Y \sim \chi^2(n)$ aldagaiak independenteak, orduan $X = Z/\sqrt{Y/n}$ aldagai aleatorioak, n askatasun graduko Student-t banaketa izango du:

$$Y = \frac{X}{\sqrt{Z/n}} \sim t_{(n)}$$

Hurrengo grafikoan, Student-t banaketaren dentsitate funtzioaren zenbait adibide jasotzen dira banaketa normal estandarrakin batera:

A.6 Irudia: Studenten-t banaketaren dentsitate funtzioa



Banaketa guztiak zero balioaren inguruan simetrikoki banatzen dira. Banaketaren batezbestekoa zero da $n > 1$ bada eta $n > 2$ denean, bere bariantza $n/(n-2)$ da. Student-t banaketaren buztanak, Normalarenak baino lodiagoak dira baina askatasun graduak handitzen diren heinean, Normal estandarrera konbergitzen du.

A.2 Inferentzia estatistikoaren errepasoa

Suposa dezagun lizentziatu berrien batezbesteko soldata aztertu nahi dela. Populazioa edo banakoen multzoa zabala denez, lagin bat aukeratzen da, hau da, zoriz aukeratutako lizentziatu berrien azpimultzo bat. Populazioa ikerketaren helburu diren elementuez osatutako multzoa da (herrialde bateko familiak, lizentziatu berriak, herri bateko etxebizitzak, enpresa bateko bezeroak, etab.) Datuak ondo erabiliz, posible da susmoen azterketa sakon bat egitea eta lagin honetako emaitzak ondorioz, populaziora orokortzea.

Inferentzia estatistikoaren helburu bat lagin bat aztertuz populazioaren ezaugarriak ezagutzea da. *Populazioa*, aztertu nahi den elementuetaz osatutako eta ondo definitutako multzoa da.

Lagina berriz, populazioaren azpimultzo adierazgarri bat da.

Populazioa zehaztu ondoren, interesatzen zaizkigun ezaugarriak jasotzen dituzten datuentzat eredu bat zehaztu behar da. Ekonometrian $y_1, y_2, y_3, \ldots, y_N$, Θ parametro ezezagunen menpeko baterako banaketako N aldagai aleatorioen gauzatzeak direla suposatzen da. **Eredu** batek, banaketa eta Θ parametro bektorearen ezaugarri nagusiak jasotzen ditu. Adibidez, hiri bateko etxebizitza baten metro karratuko batezbesteko prezioa ezagutzea interesatzen bazaigu, suposa dezagun jaso ditugun 50 etxebizitzen metro karratuko prezioak $(y_1, y_2, y_3, \ldots, y_{50})$ independenteki eta identikoki banatutako aldagai normalen gauzatzeak direla. Hortaz, datuentzat zehaztutako eredua honako da:

$$Y_i \sim NID(\mu, \sigma^2)$$

Banaketa zehazten duten parametroak, hau da, metro karratuko prezioaren batezbestekoa eta bariantza, $(\Theta = (\mu, \sigma^2))$, ezezagunak dira. Gure ikerketan interesatzen zaiguna batezbestekoa denez, datuak erabiliz, beregaineko informazioa ondorioztatu nahi izango dugu.

Horretarako, estatistikako bi erraminta aplikatuko ditugu: estimazioa eta hipotesien kontrasteak. Estimazioarekin interesatzen zaigun parametroen balio posibleak kalkulatu ahal izango dugu eta hipotesien kontrasteekin berriz, populazioaren hipotesi edo susmo bat ezarri eta gero, datuak aztertuz susmo hori egiazkoa den edo ez ondorioztatu ahal izango dugu.

A.2.1 Estimazioa

Orain arte saiakuntza baten ondorioz datuak nola sortzen diren ikusi dugu. Datu hauek inferentzian erabiliko ditugu populazioko usteak eta erlazioak aztertzeko. Ekonometrian $y_1, y_2, \ldots, y_N$ datuak, $f_\Theta(Y_1, Y_2, \ldots, Y_N)$-rekin izendatzen dugun Θ parametro desberdinen menpeko baterako banaketako N zorizko aldagaien egiteak edo gertaerak direla suposatzen da. Datuentzat zehaztutako *eredu* batek, banaketaren ezaugarri nagusiak jasotzen ditu, Θ parametro ezezagunen bektorearekin batera.

Adibideko 50 etxebizitzen metro karratuko prezioaren datuak $(y_1, \ldots, y_N)$, independenteki banatutako aldagai Normalen egiteak badira, datuentzat zehaztutako eredua

$$Y_i \sim NIB(\mu, \sigma^2)$$

da. Banaketako parametro ezezagunak (batezbestekoa eta bariantza $\Theta = (\mu,\, \sigma^2)$), datuekin estimatu daitezke, lagineko datuen batezbesteko aritmetikoa eta lagin bariantza edo dispertsioa ($\bar{y}$ eta S_y^2) erabiliz adibidez.

Estimazioaren helburua parametro ezezagun multzoaren balioei hurbiltzea da, lagineko behaketen bitartez. Parametro ezezagun bat θ bezala adieraziko dugu eta K parametro ezezaguneko bektorea $\Theta = (\theta_1, \theta_2, \ldots,\, \theta_k)'$ bezala. **Estatistiko** bat, $g(y_1, \ldots, y_N)$ datuen funtzio bat da eta θ-ren **puntuzko estimatzaile bat**. Parametro ezezagunaren hurbilketarentzat estatistiko bat, $\widehat{\theta}$ adierazten delarik. Adibidez, lagineko batezbesteko aritmetikoa, aldagai aleatorioaren batezbestekoaren estimatzailea izan daiteke eta lagineko bariantza berriz, aldagaiaren bariantzaren estimatzailea. Hau da,

$$\hat{\mu} = \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad\qquad \hat{\sigma}^2 = S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2$$

Behaketak izan aurretik osatutako funtzioa da estimatzailea eta datuak funtzio horretan barneratzerakoan lortzen den balioa *estimazioa* da. Horrela, gure adibideko etxebizitzen metro karratuko batezbesteko prezioaren (*Prezioa/m2*) estimazioa ondorengoa izango da:

$$\hat{\mu} = \frac{3,82 + 5,246 + \ldots + 3,434 + 4,20}{50} = 3,91 \text{ mila euro}$$

Hau da, etxebizitza baten batazbesteko prezio estimatua 3910 euro metro karratuko ingurukoa izango da. Hala ere, emaitza honetan nolako konfidantza izan dezakegu? Kantitate berbera berdin baloratuko genuke 5 behaketako lagin batekin lortu izan bagenu? Emaitza noski EZ da, 50 behaketekin lortutako emaitza fidagarriagoa baita 5ekin lortutako baino. Beraz, estimatzaile bat (eta bere estimazioak) zehaztasun edo fidagarritasun neurri batekin batera aztertu behar da.

Estimatzaile bat, Y_i, $i = 1, \ldots, N$ aldagaien menpeko zorizko aldagaia da. Bere probabilitate banaketa lagineko banaketa edo estimatzailearen banaketa enpirikoa bezala ezagutzen da. Aurreko adibidean, $Y_i \sim NIB(\mu, \sigma^2)$ bada, orduan $\hat{\mu} = \bar{y}$ estimatzailea N aldagai Normal independenteen konbinazio lineal bat denez, bere lagineko banaketa hurrengoa da:

$$\hat{\mu} = \bar{y} \sim N(\mu, \sigma^2/N) \tag{A.17}$$

Lagineko batezbestekoa populazioko batezbestekoaren inguruan banatzen da eta laginaren tamaina (N) gero eta handiagoa bada (hau da, bariantza txikiagoa), μ-ren inguruan probabilitate handiagoa bilduko da. Hortaz, μ-ren hurbileko estimazio bat lortzeko probabilitatea handiagoa da 50 behaketekin, 5ekin baino. Kasu honetan, zentzuduna da *zehazpenaren* indikatzaile moduan desbideratze tipikoa ($\sigma/\sqrt{N}$) erabiltzea, desbideratze tipiko txikiagoak zehaztasun handiagoa adierazten duelarik. Hala ere, normalean σ ezezaguna izaten denez, populazioko balioa laginekoarekin ordezkatuko dugu, hau da S_y. Hortaz, $\bar{y}$-ren lagineko banaketaren desbideratze tipikoaren estimazioa honakoa izango da:

$$S_{\bar{y}} = S_y/\sqrt{N}$$

$\bar{y}$-ren *errore tipiko* bezala ezagutzen delarik. etxebizitzen prezioen adibidean lortzen den estimazioaren errore tipikoa $0,993341/\sqrt{50} = 0,14$ da. Bestalde, erraz froga daiteke 5 behaketako lagin batekin $\bar{y}$ eta S_y lortuko bagenitu, errore tipikoa hirukoiztu egingo litzatekela: $S_{\bar{y}} = 0,993341/\sqrt{5} = 0,44$.

Estimatzaileak konparatzeko irizpideak

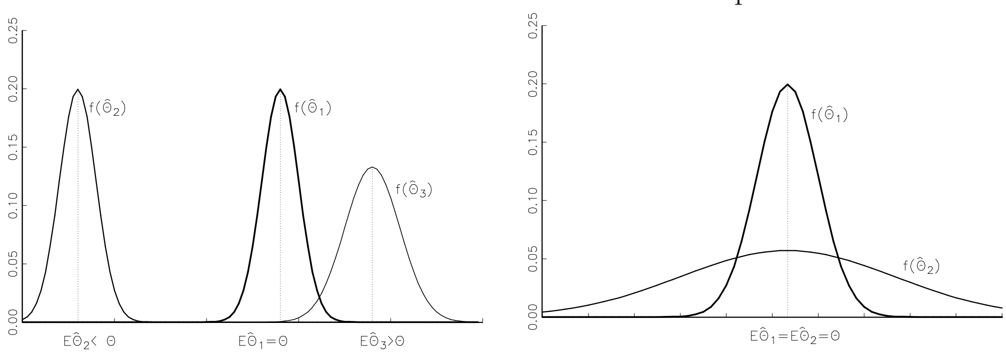
Lehen aipatu bezala, estimatzaileak ugari dira eta noski, batzuk besteak baino hobeagoak. Orokorrean, benetako balioei gehien hurbiltzen diren estimatzaileen bila gabiltza. Estimatzaileek, θ eta $\hat{\theta}$-ren arteko distantzian oinarritzen diren zenbait propietate bete ditzaten espero dugu. Horrela, lagin finituetako propietateak, estimatzaileen ezaugarriak lagineko tamainarekiko independenteki konpara daitezkenak dira. Asintotikoek ordea, portaerak konparatzen dituzte baina estimatzaileak kalkulatzen diren behaketa kopurua *handia* denean. Kurtso honetan, estimatzaileen hautaketa lagin finitu edo txikietako propietateen arabera egingo dugu, hau da, alboragabetasuna, efizientzia eta batezbesteko errore kuadratiko minimoaren arabera.

Alboragabetasuna. Estimatzaile bat alboragabea da baldin eta bere lagineko banaketaren itxaropena, parametroaren benetako balioa baldin bada, hau da:

$$E(\hat{\theta}) = \theta$$

Infinitu lagin lortuz gero eta lagin bakoitzarekin estimazioa kalkulatuz, estimazio guzti hauen batezbestekoa parametroaren benetakoa balioa izango litzateke. Hau da, batezbestekoz, estimazioaren errorea ($\hat{\theta} - \theta$) ezereztatzen da. Kontrako kasuan, estimatzailea alboratua izango da. Horrela, estimatzaile baten alborapena $Alborapena(\hat{\theta}) = (\hat{\theta}) - \theta$ bezala definitzen da. Adibide gisa, A.7 grafikoaren ezkerreko zatiak θ parametroaren hiru estimatzaileren banaketak jasotzen ditu: $\hat{\theta}_1$ estimatzailea alboragabea da; $\hat{\theta}_2$ estimatzaileak alborapen negatiboa du, hau da, batezbestekoak parametroaren balioa azpiestimatzen du eta azkenik, $\hat{\theta}_3$ estimatzailearena positiboa da, hau da, batezbestekoz, estimatzaileak parametroaren balioa gainestimatzen du.

A.7 Irudia: Estimatzaileen alborapena



Efizientzia. Estimatzaile alboragabeetan bakarrik oinarritzen bagara, estimatzaileen artean bat aukeratzeko irizpide bat finkatu behar dugu. Adibidez, A.7 grafikoaren eskubiko aldean bi estimatzaile alboragabeen banaketak adierazten dira. Garbi ikus daitekeenez, $\hat{\theta}_1$ estimatzailea bariantza txikienekoa dela eta hortaz, parametroaren benetako balioaren *urruneko* balioak lortzeko probabilitate txikiagoa du. Horregatik, $\hat{\theta}_1$ estimatzailea $\hat{\theta}_2$ baino efizienteagoa dela esango dugu.

Orokorrean, klase bereko estimatzaileen artean bariantza txikiena duena, *klase hortako efizientea* dela esaten da. Horrela, $\hat{\theta}$ estimatzailea, alboragabeen artean efizientea dela esango dugu, baldin eta ez badago bariantza txikiagodun beste estimatzaile ($\tilde{\theta}$) alboragaberen bat.

$$Bar(\tilde{\theta}) \geq Bar(\hat{\theta}) \qquad \forall\tilde{\theta} \quad alboragabea$$

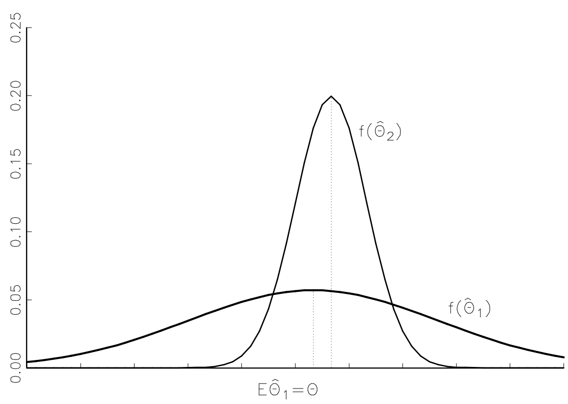
Adibidez, datuen batezbesteko aritmetikoa, aldagai Normal baten populazioko batezbestekoaren (μ) estimatzaile alboragabeen artean, efizientea da. Hau da, $Y_i \sim NIB(\mu, \sigma^2)$ bada, orduan μ-ren estimatzaile alboragabe guztientzat ($\tilde{\mu}$ non $E(\tilde{\mu}) = \mu$) ondorengoa beteko da:

$$Bar(\bar{y}) = \frac{\sigma^2}{N} \leq Bar(\tilde{\mu})$$

Azkenik, K parametrodun multzo bat (Θ) estimatzean, $\widehat{\Theta}$ estimatzaile alboragabe bat beste estimatzaile alboragabe bat ($\tilde{\Theta}$) baino efizienteagoa dela esango dugu baldin eta $[Bar(\tilde{\Theta}) - Bar(\widehat{\Theta})]$ matrizeen diferentzia matrize erdidefinitu positiboa bada. Honek $\widehat{\Theta}$ bektoreko elementu bakoitzaren bariantza, dagokion $\tilde{\Theta}$-ko elementuaren bariantza baino txikiagoa dela esan nahi du.

Batezbesteko errore kuadratikoa Alboragabetasuna desiragarria den propietatea izan arren, ez du esan nahi beti estimatzaile alboragabe bat, alboratu bat baino nahiago izango dugunik. Ondorengo grafikoaren arabera, $\hat{\theta}_1$ estimatzaile alboragabea baztertuko genuke $\hat{\theta}_2$ estimatzaile alboratuarekin alderatuz. Lehenak bariantza *oso handia* duenez, estimazioetan errore handiak egiteko probabilitate handiagoa du, bariantza txikiagoko estimatzailearekin ($\hat{\theta}_2$) konparatuz.

A.8 Irudia: Estimatzaileen banaketen adibideak



Honek, estimatzaileen aukeraketa irizpidetzat estimatzailearen errorearen neurri bat erabiltzea iradokitzen du. Jarraian, estimatzaile baten batezbesteko errore kuadratikoa definituko da:

$$BEK(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Bar(\hat{\theta}) + [alborapena(\hat{\theta})]^2$$

Eta ondoren, estimatzaile multzo batean, batezbesteko errore kuadratiko txikiena duena aukeratzen da.

A.2.2 Hipotesien kontrasteak

Aurretik aipatu bezala, Ekonometriaren helburu bat *hipotesiak kontrastatzea* da. Etxebizitzen prezioen adibideko datuak 3000 euro/m^2 batezbesteko banaketa konkretu batekin bateragarriak diren aztertzea adibidez (7. ariketa). Ikusi dugunez 6. ariketan, datuen batezbesteko aritmetikoa ez da populazioarenaren berdina izango, baina hipotesien kontraste batekin bi batezbesteko horien arteko diferentzia (populaziokoa 3000 euro eta laginekoa 3910 euro) datuen jatorri aleatorioagatik eratorria den edo ez ikusi dezakegu.

Hipotesien kontraste batek hiru etapa ditu (Ramanathan (2002)): (1) Aurkakoak diren bi hipotesi zehaztu; (2) kontrasterako estatistiko bat lortu bere lagineko banaketarekin eta (3) planteaturiko bi hipotesietatik bat aukeratzeko irizpidea ezarri.

Hipotesi bat, aldagai aleatorio bat edo batzuren banaketari buruzko baieztapen bat da. Kontraste batekin, planteatutako hipotesietatik datuei hobekien zein egokitzen den erabaki daiteke. Interesatzen zaigun hipotesia **hipotesi hutsa** (H_0) izango da eta **hipotesi alternatiboa edo aurkako hipotesiarekiko** (H_a) kontrastatzen da. Adibidez, etxebizitzen prezioen adibidean, Y-ren batezbestekoa 3 (mila euro) delaren hipotesi hutsa planteatuko dugu 3-ren desberdina delaren hipotesiaren aurka, hau da:

$$H_0 : \mu = 3 \qquad\qquad H_a : \mu \neq 3$$

Hipotesi hutsean μ-ren balio bakar bat zehazten da eta hipotesi bakuna dela esango dugu. Alternatiboa berriz, osagarrizko hipotesia izango da, gure kasuan ezberdintasuna, ondorioz, alde biko kontrastea da. Kontrastea horrelakoa bada ordea,

$$H_0 : \mu \geq 3 \qquad\qquad H_a : \mu < 3$$

batezbesteko hiru edo hiru baino handiagoa dela kontrastatu nahi dugu, hiru baino txikiagoa delaren aurkako hipotesiaren aurka. Ondorioz, alde bateko kontrastea dela esango dugu.

Azkenean zein hipotesirekin gelditzen garen **kontrasterako estatistikoaren** menpean egongo da, hau da, datuen funtzioaren arabera, konkretuki, datu eta H_0-aren arteko diferentzien funtzioan. Batezbestekoaren bi aldetako kontrastean adibidez, ondorengo diferentzia edo desberdintasuna definitzen da:

$$\frac{\bar{y} - 3}{S_{\bar{y}}}.$$

Diferentzia hau kontrasteko estatistiko gisa erabiliko dugu, ez baita neurri unitateen menpekoa eta $\bar{y}$ batezbesteko aritmetikoan laburbiltzen den datu eta H_0 hipotesian lortutako balioen arteko diferentzia kontuan hartzen den. Gainera, hipotsi hutsa egiazkoa denean aldagai aleatorio honek izango duen banaketa ezagutu beharko dugu. Gure adibidean, $y_1, y_2, \ldots, y_N$ datuak, $Y_i \sim NIB(\mu, \sigma^2)$ aldagaiaren lagina badira, non μ eta σ^2 ezezagunak diren, orduan:

$$\frac{\bar{y} - \mu}{S_{\bar{y}}} \sim t_{(N-1)}$$

eta $\mu = 3$ ordezkatuz, H_0 hipotesi hutsaren menpean estatistikoaren lagineko banaketa izango dugu:

$$t_{est} = \frac{\bar{y} - 3}{S_{\bar{y}}} \overset{H_0}{\sim} t_{(N-1)} \tag{A.18}$$

Estatistiko hau praktikan asko erabiltzen da eta batezbestekoaren t estatistikoa bezala ezagutzen da.

Azkenik, **erabaki irizpidea** zehazteko estatistikoak har ditzaken emaitza multzoa bi eremutan banatzen da: **eskualde kritikoa eta osagarrizkoa**. Eskualde horiek kontuan izanik, hipotesi hutsa baztertuko dugu baldin eta datuekin lortutako estatistikoaren balioa eskualde kritikoko balioa bada. Aurreko kasuan, H_0: $\mu = 3$ hipotesi hutsa dugu H_a: $\mu \neq 3$ hipotesiaren aurka eta H_0 baztertuko dugu batezbestekoaren estimazioa ($\bar{y}$) eta H_0-ren menpean lortutako balioaren arteko distantzia *handia* bada, horrela,

$$|t_{est}| = \left| \frac{\bar{y} - 3}{S_{\bar{y}}} \right| > c \tag{A.19}$$

non c balio kritikoa den eta $|\bar{y} - 3|/S_{\bar{y}} \leq c$ bada, hipotesi hutsa ez dela baztertzen ondorioztatuko dugu. Bestalde c-ren balioa, kontrasteko estatistikoaren banaketaren menpean dago H_0 ematen denean eta baita, jasan nahi izango dugun errorearen menpean. Dena den, kontraste batean ondorengo erroreak egiteko posibilitatea izango da:

- Hipotesi hutsa batertzea egiazkoa denean (*I motako errorea*). Kontraste baten *tamaina edo esanguratasun maila*: I motako errorea egitearen probabilitatea da eta α izendatzen da.
- Hipotesi hutsa ez baztertu egiazkoa ez denean (*II motako errorea*). Kontraste baten *potentzia*: II motako errorea ez egitearen probabilitatea da H_a egiazkoa denean.

Orokorrean, errorerik txikiena egin nahi dugu, baina errealitatean ez da posible izaten goian aipaturiko bi motatako errore horiek ezabatzea, hau da tamaina 0 eta potentzia 1 izatea. Bestalde, I motako errorea murrizteak, II motakoa handitzea dakar. Adibidez, ez genuke I motako errorerik egingo baldin eta irizpide orokortzat hipotesi hutsa ez baztertzea bada; baina kasu horretan, kontrastearen potentzia 0 izango litzateke egiazkoa ez den H_0 ere baztertuko baitugu. Hala ere, I motako erroreari garrantzi gehiago emango diogu, kontrastearen tamaina aukeratuz. Gehien erabiltzen direnak honakoak dira:

$$0,10 \qquad 0,05 \qquad \text{eta} \qquad 0,01$$

Horrela, tamaina aukeratu ondoren, potentzia handien duen kontrastea erabiltzen saiatuko gara.

Adibidea: banaketa Normalaren batezbestekoaren bi aldetako kontrastea. Suposa dezagun $y_1, y_2, \ldots, y_N$ datu multzoa, μ batezbestekoz eta σ^2 bariantzaduneko aldagai aleatorio Normal batetik eratortzen dela, hau da, $Y_i \sim NIB(\mu, \sigma^2)$. Demagun batezbestekoa zero den edo ez kontrastatzea interesatzen zaigula, horretarako H_0: $\mu = 0$ hipotesia planteatuko dugu H_a: $\mu \neq 0$ hipotesiaren aurka.

Kontrasterako estatistikoa osatzeko, μ-ren estimatzaile alboragabe batean oinarrituko gara, adibidez, datuen batezbesteko aritmetikoan ($\bar{y}$). Horrela, zentzuduna izango da H_0 baztertzea, lagineko batezbestekoak muga bat gainditzen badu, hau da:

$$|\bar{y}| > c \tag{A.20}$$

eta ez baztertzea kontrako kasuan, hau da, $-c \leq \bar{y} \leq c$. Muga hori zehazteko, estimatzailearen banaketa erabiliko dugu:

$$\bar{y} \sim N\left(\mu,\, \frac{\sigma^2}{N}\right) \qquad \Rightarrow \qquad \frac{\bar{y}-\mu}{\sigma/\sqrt{N}} \sim N(0,1)$$

Aldagai honek banaketa ezaguna du baina σ^2 bariantza ezezaguna duenez, datuak kontuan izanik, ezin izango genuke estatistikoaren egiterik edo gauzatzerik izan. Baina σ^2 bariantza bere estimatzaile alboragabeagatik ordezka dezakegu, S^2, ondorengoa lortuz:

$$\frac{\bar{y}-\mu}{S/\sqrt{N}} \sim t_{(N-1)}$$

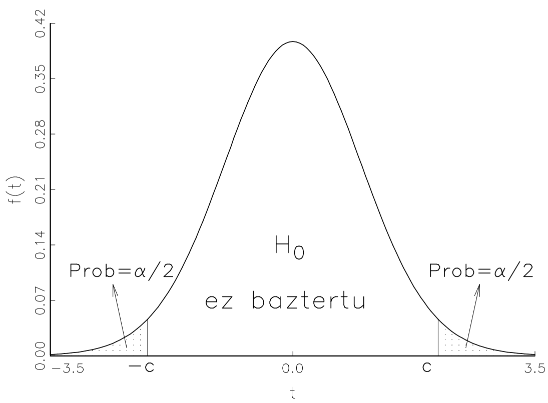
Baldin eta $H_0 : \mu = 0$ egiazkoa bada, aldagai honen banaketa hau da:

$$t_{est} = \frac{\bar{y}}{S/\sqrt{N}} \overset{H_0}{\sim} t_{(N-1)} \tag{A.21}$$

eta baita kontrasterako estatistikoa ere. Estatistiko hau batezbestekoaren t estatistikoa bezala ezagutzen da. Azaldutako (A.20) irizpidea jarraituz, H_0 baztertuko dugu zerotik urruneko balioak hartzen dituenean, hau da $|t_{est}| > c$. Kontrasteko tamainaren arabera aukeratzen da c balioa, eta α esanguratasun maila izendatzen da. Azken finean, esanguratasun maila, egiazkoa den H_0 batertzearen probabilitatea da, hau da:

$$\alpha = Prob(|t_{est}| > c) \qquad t_{est} \sim t_{(N-1)} \text{ denean}$$

A.9 Irudia: Alde biko kontrastearen eskualde kritikoak



c balioa, $t_{(N-1)}$ banaketaren jatorria da eta banaketako bi buztanetan α probabilitatea uzten du. Adibidez, $\alpha = 0,05$ eta $N = 50$ badira, orduan $c = 2,01$. Horrela, H_0 baztertuko dugu %5eko esangura-mailarekin, baldin eta $|t_{est}| > 2,01$ bada edota:

$$\bar{y} > 2,01\,\frac{S}{\sqrt{N}} \qquad \text{edota} \qquad \bar{y} < -2,01\,\frac{S}{\sqrt{N}}$$

A.9 grafikoko ezkerreko aldeak, estatistikoaren banaketa adierazten du H_0: $\mu = 0$ hipotesia egiazkoa denean. Banaketaren buztanetan adierazitako margotutako eremuak eskualde kritikoa osatzen dute, buztan bakoitzean $\alpha/2$ probabilitatea biltzen delarik. Kasu honetan, H_0 baztertuko dugu, datuekin lortutako t-estatistikoaren balioa, $\mu = 0$ batezbestekodun banaketan, *ia ziur* emango ez den balioa baita.

Adibidea: m^2ko prezioaren batezbestekoaren kontrastea Gretlekin.
Demagun metro karratuko prezioa aldagaiak (*pr_m2*) banaketa Normala jaraitzen duela. Kontrasta ezazu H_0: $\mu = 3$ hipotesia H_a: $\mu \neq 3$ hipotesiaren aurka. Jarraitu beharreko pausuak honakoak dira:

1. Ondorengo estatistikoaren lagineko balioa kalkulatu $t = (\bar{y} - 3)/S_{\bar{y}}$, non $\bar{y}$, *pr_m2* aldagaiaren lagineko batezbestekoa den:

$$t = \sqrt{50}(3,9144 - 3)/0,99341 = 6,51$$

Gretleko ondorengo aukera jarraitu dezakegu:
Tresnak →Kontrasteen estatistikoen kalkulagailua
Ondorengo lehiatilan *batezbestekoa* klikatu eta bertan:

– *Erabili datu-multzoko aldagaia* klikatu.
– Hemen *pr_m2* aldagaia aukeratuz, t kalkulatzeko behar diren estatistiko deskribatzaileak agertuko dira:

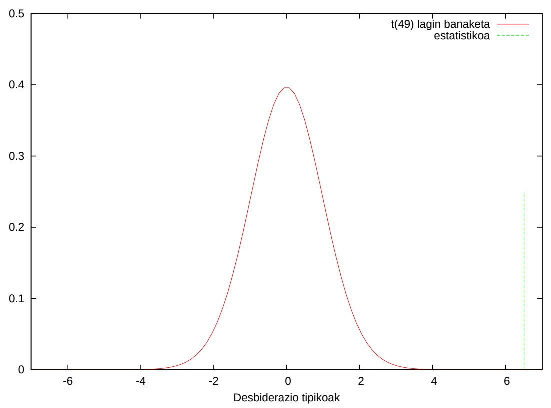
Lagineko batezbestekoa:	3,9144
Desb. tipikoa:	0,99341
Laginaren tamaina:	50

– Kontrastatu nahi den hipotesi hutsa idatzi: *H0: batezbestekoa =* 3.
– Konprobatu *Suposatu desbideratze tipikoa populazioko balioa dela* aukera **ez** dagoela aktibaturik eta klikatu *Ados*.

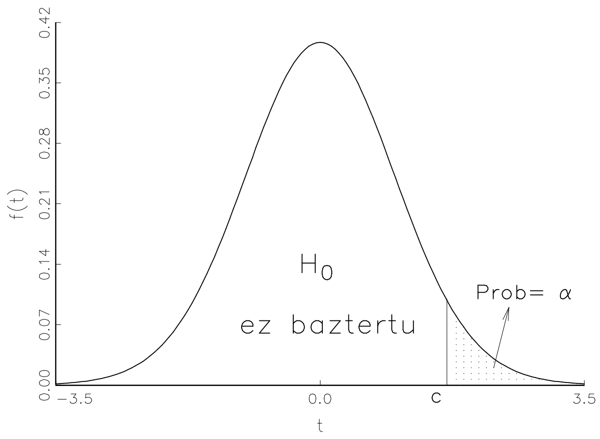
Emaitza honakoa da:

```
Hipotesi hutsa:  populazio batezbestekoa = 3 Lagin tamaina: n =50
Lagin batezbestekoa = 3,91439,  desbideratze tip. = 0,993407
Estatistikoa: t(49) = (3,91439 - 3)/0,140489 = 6,50864 Alde-biko
p-balioa = 3,83e-008 (alde-batekoa = 1,915e-008)
```

Dagokion grafikoa:



A.10 Irudia: Alde bateko kontrastearen eskualde kritikoak



non estatistikoaren balioa konfidantza tartetik kanpo dagoela ikusten den. Bestalde ondorengo grafikoan

estatistikoaren banaketa agertzen da hipotesi hutsaren menpean, kasu honetan $t(49)$, eta baita estatistikoaren lagineko balioa ere (lerro berdea). Ikus dezakegu estatistikoaren lagineko balioa goi buztanean erortzen dela, hau da, H_0 egiazkoa izatea ez da probablea. Beraz, hipotesi hutsa baztertuko dugu. Dena den, eskualde kritikoa kalkulatuko dugu.

2. Eskualde kritikoa edo baztertze eskualdea. Gretl erabiliz c balio kritikoa aurkitu dezakegu, zehazki *Tresnak →Estatistika-taulak* aukeran.

   Lehiatila berrian t aukeratu behar da eta *ag* koadrotxoan askatasun graduak idatzi, gure kasuan 49, eta baita behar dugun esangura. Emaitza bezala 1,299 balioa irtetzen da.

   Balioa interpretatuz, honakoa adierazten du: adibidez $X \sim t(49)$ bada orduan

$$Prob(X > 1,299) = 0,10$$

   Hau da, $[1,299, \infty)$ tartean duen metatutako probabilitatea 0,1 da. Kontrastea bi aldetakoa denez $\alpha =\% \ 5$ esangura-mailarentzat, eskualde kritikoaren eskubiko zatian $\alpha/2$ probabilitatea metatuko da. Beraz, c balio kritikoa lortzeko 49 ag eta $\alpha/2 = 0,025$ barneratzen dugu. Ematen digun balioa $c = 2,010$ da.

3. Erabaki irizpidea aplikatu, hau da, $|6,51| > 2,010$ denez % 5eko esangura-mailarekin, batezbesteko prezioa 3000 euro direla hipotesi hutsa baztertzen da. Azkenik, itxi itzazu lehiatilak.

Adibidea: banaketa Normal baten batezbestekoaren alde bateko kontrastea. Askotan ikerketa ekonometrikoetan alde bateko kontrasteak burutu behar dira. Adibidez, soldatarekiko bereizketarik dagoen edo ez, aztertzeko eta konkretuki emakumena gizonena baino txikiagoa dela kontrastatzeko.

Suposa ezazu, aurreko kasuan batezbestekoa hiru edo txikiagoa dela kontrastatu nahi dela hiru baino handiagoa delaren aurka. Ondorioz, hurrengo hipotesiak planteatuko genituzke:

$$H_0 : \mu \leq 3 \qquad\qquad H_a : \mu > 3.$$

Kontrastea alde batekoa denez, H_0 baztertuko dugu H_a: $\mu - 3 > 0$ hipotesiaren aurka, baldin eta $\bar{y} - 3$ diferentziak balio konkretu bat gainditzen badu, hau da $\bar{y} - 3 > c$. Hipotesi honen kontrasterako estatistikoa ere $\bar{y}$ estimatzailearen banaketatik ondorioztatzen da. Baldin eta $H_0: \mu = 3$ egiazkoa bada, aldagai honen banaketa hurrengoa litzateke:

$$t_{est} = \frac{\bar{y} - 3}{S/\sqrt{N}} \overset{H_0}{\sim} t_{(N-1)} \tag{A.22}$$

Beraz, H_0 baztertuko dugu $\bar{y} - 3$ diferentzia *handia* denean, hau da H_0 baztertuko dugu $t_{est} > c$ denean. Jakina da, c balioa ondorengoagatik zehazten dela: $\alpha = Prob(t_{est} > c); \quad t_{est} \sim t_{(N-1)}$ denean. Hortaz, kontrasteko eskualde kritikoa, A.9 grafikoko eskubiko aldean agertzen den paneleko margotutako eremuak osatuko du. Horrela, H_0 ez dugu baztertuko α esangura-mailarekin, baldin eta t_{est}-ren balioa $t_{(N-1)}$ banaketako c balioa baino txikiagoa bada, honek α probabilitatea metatzen duelarik.

Orokorrean, H_0 *baztertu* edo *ez baztertu* hitzak erabiltzen dira. Kontraste batean H_0 mantendu ohi da, kontrakoaren ebidentzia nahiko ez dagoen bitartean. Datuek H_0 hipotesi hutsa baztertu dezakete baina ezin dute frogatu zuzena edo egiazkoa den eta ondorioz, ez da H_0 *onartu* esaten. Bestalde, H_0 ez baztertzeak, datuekin hipotesiaren faltsukeria ezin dutela frogatu esan nahi du.

Bibliografia

Peña, D. eta J. Romo (1997). *Introducción a la Estadística para las Ciencias Sociales.* Mc-Graw Hill, Madrid.

Ramanathan, R. (2002). *Introductory Econometrics with Applications.* South western, Ohio.