

**KUBOAREN METODOA:
LAGINKETA OREKATUAREN APLIKAZIOAK EAE-KO
ESTATISTIKA-ERAKUNDEAN**

Aritz Adin Urtasun



EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA

Donostia-San Sebastián, 1
01001 VITORIA-GASTEIZ
Tel.: 945 01 75 00
Faxa: 945 01 75 01
E-mail: eustat@eustat.es
www.eustat.es

Aurkezpena

Eustat-ek, ikusten baitu egunetik egunera handiagoa dela kalitatezko estatistika gero eta bereiziagoen eskaria, XXIII. Nazioarteko Estatistika Mintegia antolatu zuen, 2010ean, izenburu honekin: "Laginketa orekatu eta eraginkorra: Kuboaren Metodoa".

Eustaten xedea da egungo diseinuak berriz definitzea, kostu bera edo antzekoa eginda eremu edo esparru bereiziagoetarako kalitatezko zenbatesleak ematen dituzten laginak lortu ahal izateko. Helburu berarekin egin zen estatistika- nahiz matematika-metodologiaren alorrean (laginen optimizazioaren gaian, zehazkiago) bi urtez prestatzeko eta ikertzeko beka baterako deialdia.

Ikerketa horren emaitzak zenbait estatistika-jardueratan aplikatu dira 2010-2012ko Estatistiketarako Euskal Planaren barruan: Lehen Hezkuntzako eta Derrigorrezko Bigarren Hezkuntzako ikastetxeetako ikasleen *bullyingari* buruzko ikerketa; Informazioaren Gizarteari buruzko Inkesta. Familiak; berrikuntza teknologikoari buruzko inkesta; pobrezia eta gizarte-desberdintasunei buruzko inkesta; eta EAEko landa-eremuko emakumeei buruzko ikerketa.

Argitalpen honen xedea da bekari esker egindako ikerketa zabaltzea, hala nola material erabilgarria ematea laginketa eraginkor eta orekatuen ezagutzan eta erabilpenean interesa duten erabiltzaile guztiei.

Agiri honek ondo bereizitako atal bi ditu. Lehendabizikoan, laginketaren teoriari dagozkion kontzeptuak eta definizioak daude; bai eta probabilitate-laginketa bakun eta konplexueterako planak ere. Bigarrenengoan, Kuboaren Metodoaren deskribapena jaso da, eta haren aplikazioa Euskal Estatistika Erakundearen zenbait inkesta-motatan.

Gasteiz, 2012ko Abendua .

Javier Forcada Sainz

EUSTATEko zuzendari nagusia

Aurkibidea

AURKEZPENA	1
AURKIBIDEA.....	2
1. SARRERA	4
2. LAGINKETA-TEORIAREN SARRERA	5
DEFINIZIOAK ETA OINARRIZKO NOTAZIOA	5
LAGIN-PROPORTZIOAK	6
HORVITZ-THOMPSONEN ZENBATESLEA	6
3. PROBABILITATE-LAGINKETARAKO PLANAK	7
AUSAZKO LAGINKETA BAKUNA	7
LAGINKETA GERUZATUA	8
CLUSTER EDO KONGLOMERATUEN BIDEZKO LAGINKETA	10
AURKEZTUTAKO METODOEN LABURPENA.....	11
4. LAGINKETA-PLAN KONPLEXUAK	13
LAGINKETA BIKOITZA (EDO BI ETAPAKOA)	13
PROBABILITATE BERAK DITUZTEN LEHEN MAILAKO UNITATEEN AUKERAKETA.....	14
BI ETAPAKO PLAN AUTOHAZTATUA	15
5. KUBOAREN METODOA: LAGINKETA OREKATUA.....	16
KUBO BIDEZKO IRUDIKAPENA.....	16
LAGIN OREKATUAK.....	17
METODOAREN DESKRIBAPENA.....	18
6. SAS MAKROAK LAGIN OREKATUAK AUKERATZEKO	19
EXE_CUBE MAKROA	19
ECHANT_STRAT MAKROA.....	20
DISJUNCTIVE MAKRO LAGUNGARRIA	21
CREAR_ESTRATO MAKRO LAGUNGARRIA	21
MAKROEN ERABILERAREN ADIBIDEA.....	22
7. EUSTATEN KUBOAREN METODOAZ OREKATUTAKO LAGINAK	26
DBHko IKASTETXEEN LAGINA, 'BULLYING'-A AZTERTZEKO EUSKAL AUTONOMIA ERKIDEGOAN	26
INFORMAZIOAREN GIZARTEARI BURUZKO INKESTARAKO LAGINA (IGI-ENPRESAK).....	30
KAPITAL SOZIALARI BURUZKO INKESTARAKO LAGINA.....	33
BERRIKUNTZA TEKNOLOGIKOARI BURUZKO INKESTARAKO LAGINA (BTI).....	38

POBREZIARI ETA GIZARTE DESBERDINTASUNEI BURUZKO INKESTARAKO LAGINA (PGDI)	42
EAEKO LANDA-EREMUKO EMAKUMEAK IKERTZEKO LAGINA.....	47
EUSKADI ETA DROGAK INKESTARAKO LAGINA	52
8. ONDORIOAK.....	56
OREKA ETA GERUZATZEA	56
OREKA-ALDAGAIEN AUKERAKETA.....	56
OREKA ETA KALIBRAZIOA.....	57
Emitzen analisia.....	57
1. 2012ko Euskadi eta Drogak inkestaren kalibrazioa.....	57
2. 2012ko Kapital Sozialari buruzko Inkesta.....	58
LAGINKETA OREKATUAREN INTERESA.....	60
9. BIBLIOGRAFIA	61

1. Sarrera

Koaderno tekniko honetan bildutakoa Euskal Estatistika Erakundeak, 2010. urtean, laginketen optimizaziorako estatistika- eta matematika-metodologietan prestatzeko eta ikertzeko emandako bekari esker egindako lanaren emaitza da.

Dokumentu hau hurrengo kapitulu hauetan banatuta dago:

Lehenengo kapituluan sarrera bat egin da eta koaderno tekniko hau egitean izandako helburuak aipatu dira.

Bigarren kapituluan, laginketa-teoriaren sarrera azaltzen da, honako hauekin batera: laginketa-diseinuaren oinarritzko notazioa eta definizioak, lagin-proportzioak eta Horvitz-Thompsonen zenbateslearen nahiz haren bariantzaren definizioa.

Hurrengo bi kapituluetan, kontzeptu hauek lantzen dira: probabilitate-laginketarako planak eta laginketa konplexuetarako planak; gainera, estatistika ofizialean erabiltzen diren metodo gehienak aurkezten dira.

Bosgarren kapituluan, laginketa orekatuaren kontzeptua jorratzen da, eta lagin orekatuak aukeratzeko Kuboaren Metodoa aurkezten da.

Lagin orekatuak hautatzeko aukera ematen duten SAS makroak zehaztea da seigarren kapituluaren xedea.

Zazpigarrenean, Kuboaren Metodoarekin Eustaten orekatutako laginak aurkezten dira.

Azkenik, orekarekin, geruzatzearekin eta kalibratzearekin zerikusia duten zenbait ondorio azaltzen dira.

Zinez eskertu nahi dut Metodologia, Berrikuntza eta I+G Ataleko kide guztiek emandako laguntza eta, oro har, Eustateko langile guztien adeitasuna.

GAKO-HITZAK: Lagin-diseinuak, sartze-probabilitateak, Horvitz-Thompsonen zenbateslea, lagin orekatuak, kuboaren metodoa, oreka-aldagaiak, geruzatzea, kalibratzea.

2. Laginketa-teoriaren sarrera

Laginketa orekatuak aukeratzeko kuboaren metodoa aurkeztu eta metodoaren baliagarritasuna azaldu baino lehen, laginketa-teoriaren aurkezpen orokorra egin behar dugu.

Definizioak eta oinarritzko notazioa

N tamainako $U = \{1, \dots, N\}$ populazio finitu bat aztertzea da xedea.

y interes-aldagaia definitzen dugu. y_k balioak hartzen ditu, $k \in U$; eta hauek dira haien guztizkoa eta batez bestekoa:

$$Y = \sum_{k \in U} y_k \quad \text{eta} \quad \bar{Y} = \frac{1}{N} \sum_{k \in U} y_k$$

s lagina $s \subset U$ populazioaren azpimultzo bat da.

$p(s)$ lagin-diseinua edo laginketa-plana litezkeen lagin guztiekiko probabilitate-banaketa da; bertan, $\sum_{s \subset U} p(s) = 1$.

S ausazko laginak s balioa hartzen du $\Pr(S = s) = p(s)$ probabilitatearekin.

Sartze-probabilitatea k unitatea S ausazko laginean aukeratua izateko probabilitatea delako definizioa ematen dugu:

$$\pi_k = E(I_k) = \Pr(k \in S) = \sum_{k \in S} p(s) \quad \text{non} \quad I_k = \begin{cases} 1, & k \in S \quad \text{bada} \\ 0, & k \notin S \quad \text{bada} \end{cases}$$

Era berean, hau da 2. ordenako sartze-probabilitatearen definizioa:

$$\pi_{kl} = E(I_k I_l) = \Pr(k \text{ y } l \in S) = \sum_{k, l \in S} p(s)$$

Lagin-diseinua tamaina finkokoa bada, orduan $\sum_{k \in U} \pi_k = n$.

Lagin-proporzioak

Demagun U populazioa kontuan hartuta zehaztutako interes-aldagaia aldagai kualitatiboa dela. Kasu horretan, interes-aldagaiak informazioa ematen digu populazioko unitateen ezaugarriren bati buruz, edo klase jakin batekoak izateaz edo ez izateaz.

Demagun gure interes-aldagaiak populazioko unitateak C eta C' klaseetan sailkatzen dituela.

Populazioko unitate bakoitzerako, honela definitzen dugu y_k ezaugarria:

$$y_k = \begin{cases} 1, & k \in C \text{ bada} \\ 0, & k \notin C \text{ bada} \end{cases} \quad \forall k \in U$$

Hauek dira populazioko elementuen guztizkoa (klasearen guztizkoa) eta C klasekoak diren populazioko elementuen proportzioa (klasearen proportzioa):

$$Y = \sum_{k \in U} y_k = A \quad \text{eta} \quad \bar{Y} = \frac{1}{N} \sum_{k \in U} y_k = \frac{A}{N} = P$$

Kontuan har dezakegu A eta P zenbatesteko problema populazioaren guztizkoa eta batez bestekoa zenbatetsiko bagenu bezala dela, y_k bakoitzak 0 edo 1 balioa hartuz.

S^2 kuasibariantza hauen arabera idazten badugu: P eta $Q = 1-P$

$$S^2 = \frac{\sum_{k \in U} (y_k - \bar{Y})^2}{N-1} = \frac{\sum_{k \in U} y_k^2 - N\bar{Y}^2}{N-1} = \frac{1}{N-1} (NP - NP^2) = \frac{N}{N-1} PQ$$

Haren zenbatesle alboragabea hau izanik:

$$s^2 = \frac{n}{n-1} pq \quad \text{non} \quad p = \frac{\sum_{k \in S} y_k}{n} = \frac{a}{n} \text{ baita}$$

Horvitz-Thompsonen zenbateslea

y interes-aldagaiaren populazioaren guztizkoaren nahiz batez bestekoaren Horvitz-Thompsonen zenbateslea honela definitzen da:

$$\hat{Y}_\pi = \sum_{k \in S} \frac{y_k}{\pi_k} \quad \text{eta} \quad \hat{\bar{Y}}_\pi = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k}$$

Horvitz-Thompsonen zenbateslea alboragabea da $\pi_k > 0$, $k \in U$ izanez gero.

Tamaina finkoko diseinuetarako, bariantza honela zenbatets daiteke:

$$\hat{Var}(\hat{Y}_\pi) = -\frac{1}{2} \sum_{k \in S} \sum_{l \in S, l \neq k} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2 \frac{(\pi_{kl} - \pi_k \pi_l)}{\pi_{kl}}.$$

3. Probabilitate-laginketarako planak

Metodo bat horrelako laginketetako bat da baldin eta litezkeen laginetako edozein ateratzeko edo aukeratzeko probabilitatea kalkula badaiteke metodo horretarako.

Aurrerago azalduko den bezala, kuboaren metodoa diseinuak lagin orekatu bat aukeratzeko zehaztutako sartzeko-probabilitateetatik abiatzen da; hau da, metodoak, egia esan, probabilitate-laginketetarako metodoak optimizatzen ditu.

Ondoren, probabilitate-laginketarako hiru aukera nagusiak definituko dira.

Ausazko laginketa bakuna

Ausazko laginketa bakuna (a.l.b.) deritzon laginketa-metodoan n tamainako lagin bat aukeratzeko N tamainako populazio batetik, halako eran non tamaina bereko lagin guztiek aukeratuak izateko probabilitate bera duten.

n tamaina finkoko a.l.b. baterako lagin-diseinua hau da:

$$p(s) = \begin{cases} \binom{N}{n}^{-1} & \text{card}(s) = n \quad \text{bada} \\ 0 & \text{hala izan ezean} \end{cases}$$

Hortaz, k unitatearen sartzeko-probabilitatea hau da:

$$\pi_k = \sum_{s \in S} p(s) = \sum_{s \in S} \binom{N}{n}^{-1} = \binom{N-1}{n-1} \binom{N}{n}^{-1} = \frac{n}{N}, \quad \forall k \in U$$

Hau da, U -ko banako guztiek aukeratuak izateko probabilitate bera dute.

Ausazko laginketa bakun batean, hau da populazioaren batez bestekorako H-Tren zenbateslea:

$$\hat{\bar{Y}}_{\pi} = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k} = \frac{1}{N} \sum_{k \in S} y_k \frac{N}{n} = \frac{1}{n} \sum_{k \in S} y_k$$

$\hat{\bar{Y}}_{\pi}$ -ren bariantzaren zenbatesle alboragabea hau da:

$$\widehat{Var}(\hat{\bar{Y}}_{\pi}) = (1-f) \frac{s_y^2}{n}$$

Kasu horretan:

$$s_y^2 = \frac{1}{n-1} \sum_{k \in S} (y_k - \hat{\bar{Y}}_{\pi})^2$$

eta

$$f = \frac{n}{N} \text{ laginketa-zati moduan definitzen da.}$$

Laginketa geruzatua

Demagun U populazioa U_h , $h = 1, \dots, H$ azpipopulazioetan edo geruzetan banatuta dagoela, eta geruzek baldintza hauek betetzen dituztela:

- (i) $\bigcup_{h=1}^H U_h = U$
- (ii) $U_h \cap U_i = \emptyset, \quad h \neq i$
- (iii) U_h -ren tamaina N_h bada, orduan $\sum_{h=1}^H N_h = N$

Lagin-diseinu bat geruzatua da baldin eta geruza bakoitzean n_h tamaina finkoko ausazko lagin bakun bat aukeratzen bada, $\sum_{h=1}^H n_h = n$ laginaren tamaina izanik.

Laginketa-teknika hau aztertu beharreko populazioa oso heterogeneoa denean erabiltzen da, eta populazioa zenbait geruza homogeneotan zati dezake. Horrenbestez, zenbatesle zehatzagoak lor ditzakegu geruza bakoitzean, eta haiek bateratu egin ditzakegu populazio osoaren zenbatesle zehatzagoa lortzeko.

Geruza bakoitzean unitateak ausazko laginketa bakun baten bidez aukeratzen direnez, k unitatearen sartze-probabilitatea hau da:

$$\pi_k = \frac{n_h}{N_h}, \quad \forall k \in U.$$

Laginketa geruzatu baterako batez bestekoaren Horvitz-Thompsonen zenbateslea:

$$\hat{\bar{Y}}_{st} = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k} = \frac{1}{N} \sum_{h=1}^H \frac{N_h}{n_h} \sum_{k \in S_h} y_k = \frac{1}{N} \sum_{h=1}^H N_h \hat{\bar{Y}}_h$$

Zenbateslearen bariantza honela zenbatets daiteke alborapenik gabe:

$$\hat{Var}(\hat{\bar{Y}}_{st}) = \frac{1}{N^2} \sum_{h=1}^H N_h^2 (1 - f_h) \frac{s_{yh}^2}{n_h}$$

non $s_{yh}^2 = \frac{1}{n_h - 1} \sum_{k=1}^{n_h} (y_k - \hat{\bar{Y}}_h)^2$ h geruzaren lagin-kuasibariantza baita.

Esleipenak laginketa geruzatuetan

Laginaren tamaina geruzetan banatzeko zenbait irizpide daude. Erabilienak aurkeztuko ditugu.

1. Esleipen proportzionala

Geruza bakoitzari esleitzen zaizkion laginketa-unitateen kopurua haren tamainarekiko proportzionala da.

Hortaz, esango dugu plan geruzatu batek esleipen proportzionala duela hau gertatuz gero:

$$\frac{n_h}{N_h} = \frac{n}{N}, \quad h = 1, \dots, H \text{ izanik}$$

$n_h = \frac{nN_h}{N}$ osoa dela emanez gero, populazioaren batez bestekoaren zenbateslea hau da:

$$\hat{\bar{Y}}_{prop} = \frac{1}{N} \sum_{h=1}^H N_h \hat{\bar{Y}}_h = \frac{1}{n} \sum_{k \in S} y_k$$

Modu berean egin dakizkieke esleipen proportzionalak errodurari, kuboari edo 1 baino txikiagoa den edozein berreturari.

2. Bariantza txikieneko esleipena

Bariantza txikieneko esleipena edo Neyman-en esleipena da n_h -ren balioak zehaztea, lagin finkoaren tamaina n izanik, zenbatesleen bariantza txikiena izan dadin.

Lagrange-ren biderkatzaileak erabiliz, n_h -ren beharrezko balioak lortzen dira

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^H N_h S_h} \quad h = 1, \dots, H \text{ izanik}$$

3. Lagin-tamaina txikieneko esleipena

Kasu honetan, V bariantza finkoa izanik, n^* lagin-tamaina txikiena ematen duen esleipena lortzea da xedea.

Berriz ere, Lagrangeren biderkatzaileei esker, hau dugu:

$$n^* = \frac{\left(\sum_{h=1}^H N_h S_{yh} \right)^2}{V + \sum_{h=1}^H N_h S_{yh}^2}$$

Cluster edo konglomeratuen bidezko laginketa

Demagun U populazioa konglomeratu izeneko M azpimultzotan dagoela zatituta, U_i , $i = 1, \dots, M$ azpimultzoetan, eta horiek ezaugarri hauek dituztela:

- (i) $\bigcup_{i=1}^M U_i = U$
- (ii) $U_i \cap U_j = \emptyset, i \neq j$
- (iii) $\sum_{i=1}^M N_i = N$ non U_i konglomeratuko elementuen kopurua baita N_i .

Lagin-diseinu bat konglomeratuzkoa izateko, nahitaezkoa da s_I denotatuko dugun m tamainako konglomeratu-lagin bat aukeratzea, eta $p_I(s_I)$ planean behatuak izatea hautatutako konglomeratuetako unitate guztiak.

Ausazko lagin osoa $S = \bigcup_{i \in S_I} U_i$ -ren bidez adierazten da, eta $n = \sum_{i \in S_I} N_i$ da haren tamaina. Laginaren tamaina ausazkoa da jeneralean.

Laginketa-teknika hau populazioa ustez populazioaren aldagarritasun guztia jasotzen duten taldetan berez zatituta dagoenean erabiltzen da; hau da, konglomeratu bakoitzak zehatz-mehatz adierazten du ikertu beharreko populazio-ezaugarria (lagin-informazioaren bilketa sinplifikatuz).

Probabilitate berak dituzten konglomeratuen aukeraketa

Eman dezagun konglomeratu guztiek aukeratuak izateko probabilitate bera dutela; orduan, laginketa-plana konglomeratuak m tamainako ausazko laginketa bakun baten arabera aukeratzean datza.

Kasu honetan, $\pi_{Ii} = \frac{m}{M}$ da konglomeratu bat aukeratzeko probabilitatea, eta adierazpen sinplifikatu hau lortzen da batez bestekoaren Horvitz-Thompsonen zenbateslearentzat:

$$\hat{\bar{Y}}_{\pi} = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k} = \frac{1}{N} \sum_{i \in S_I} \frac{N_i \bar{Y}_i}{\pi_{Ii}} = \frac{M}{Nm} \sum_{i \in S_I} N_i \bar{Y}_i$$

non $\bar{Y}_i = \frac{1}{N_i} \sum_{k \in U_i} y_k$ batez bestekoa baita U_i , $i = 1, \dots, M$ konglomeratuarentzat.

Zenbateslearen bariantza honela zenbatets daiteke alborapenik gabe:

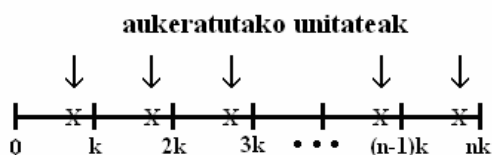
$$\hat{Var}(\hat{\bar{Y}}_{\pi}) = \frac{M-m}{N^2 m} \frac{M}{m-1} \sum_{i \in S_I} \left(Y_i - \frac{\hat{\bar{Y}}_{\pi}}{M} \right)^2$$

Laginketa sistematikoa probabilitate berekin

Demagun U populazioko N unitateei $1, \dots, N$ zenbakiak ipintzen zaizkiela ordenaren baten arabera (ausazkoa edo ordena-irizpideren bat erabiliz).

Laginean aukeratu beharreko unitateen kopurua n bada, $k = N/n$ laginketa-tarte moduan definitzen dugu.

Ausazko $r \in \{1, \dots, k\}$ zenbaki bat aukeratzen dugu hasierako unitate moduan. r -tik aurrera, $l = 1, 2, \dots, n-1$ -rako lk distantziara dauden unitateak aukeratzen dira laginketan.



Laginketa sistematikoz esan daiteke konglomeratu bidezko laginketa dela, eta bertan litezkeen k guztietatik cluster bakar bat hautatzea dela kontua.

Litezkeen k lagin sistematikoen konposizioa

<u>1</u>	<u>2</u>	<u>...</u>	<u>i</u>	<u>...</u>	<u>k</u>
y_1	y_2		y_i		y_k
y_{k+1}	y_{k+2}		y_{k+i}		y_{2k}
...	...		...		...
$y_{(n-1)k+1}$	$y_{(n-1)k+2}$		$y_{(n-k)k+3}$		y_{nk}

Aurkeztutako metodoen laburpena

$\hat{\theta}$ zenbateslearen aldakuntza-koefizientea desbiderapen tipikoaren eta haren θ benetako balioaren arteko zatidura da: $CV(\hat{\theta}) = \frac{\sqrt{Var(\hat{\theta})}}{\theta}$.

Hortaz, $\hat{\theta}$ -ren aldakuntza-koefizientearen zenbateslea $cv(\hat{\theta}) = \frac{\sqrt{\hat{Var}(\hat{\theta})}}{\hat{\theta}}$ da.

Ondoren, zenbateslearen, bariantzaren eta aldakuntza-koefizientearen formulak erakusten dira taula batean, aurkeztutako metodoetako populazioaren batez bestekoari eta proportzioei dagokienez.

		Ausazko laginketa bakuna	Laginketa geruzatua	Konglomeratuen bidezko laginketa
Populazioaren batez bestekoa $\bar{Y}$	Zenbateslea $\hat{\bar{Y}}$	$\hat{\bar{Y}}_{\pi} = \frac{1}{n} \sum_{k \in S} y_k$	$\hat{\bar{Y}}_{st} = \frac{1}{N} \sum_{h=1}^H N_h \hat{\bar{Y}}_h$	$\hat{\bar{Y}}_{\pi} = \frac{M}{Nm} \sum_{i \in S_I} N_i \bar{Y}_i$
	Bariantza $\hat{Var}(\hat{\bar{Y}})$	$Var(\hat{\bar{Y}}_{\pi}) = (1-f) \frac{S_y^2}{n}$	$\hat{Var}(\hat{\bar{Y}}_{st}) = \frac{1}{N^2} \sum_{h=1}^H N_h^2 (1-f_h) \frac{S_{yh}^2}{n_h}$	$\hat{Var}(\hat{\bar{Y}}_{\pi}) = \frac{M-m}{N^2 m} \frac{M}{m-1} \sum_{i \in S_I} \left(Y_i - \frac{\hat{\bar{Y}}_{\pi}}{M} \right)^2$
	Aldakuntza koefizientea $cv(\hat{\bar{Y}})$	$cv(\hat{\bar{Y}}_{\pi}) = \sqrt{\frac{(1-f)}{n}} cv(\bar{Y})$	$cv(\hat{\bar{Y}}_{st}) = \frac{\sqrt{\sum_{h=1}^H N_h^2 (1-f_h) \frac{S_{yh}^2}{n_h}}}{\sum_{h=1}^H N_h \hat{\bar{Y}}_h}$	$cv(\hat{\bar{Y}}_{\pi}) = \frac{\sqrt{\left(1 - \frac{m}{M}\right) \frac{m}{m-1} \sum_{i \in S_I} \left(Y_i - \frac{\hat{\bar{Y}}_{\pi}}{M} \right)^2}}{\sum_{i \in S_I} N_i \bar{Y}_i}$
Proporzioak P	Zenbateslea $\hat{P}$	$\hat{P} = \frac{1}{n} \sum_{k \in S} y_k$	$\hat{P}_{st} = \frac{1}{N} \sum_{h=1}^H N_h p_h$	$\hat{P} = \frac{\sum_{i \in S_I} a_i}{\sum_{i \in S_I} N_i}$; non $a_i = p_i N_i$
	Bariantza $\hat{Var}(\hat{P})$	$\hat{Var}(\hat{P}) = (1-f) \frac{p(1-p)}{n-1}$	$\hat{Var}(\hat{P}_{st}) = \frac{1}{N^2} \sum_{h=1}^H N_h^2 (1-f_h) \frac{p_h q_h}{n_h - 1}$	$\hat{Var}(p) = \frac{M-m}{M} \frac{m}{m-1} \frac{\sum_{i \in S_I} a_i^2 - 2p \sum_{i \in S_I} a_i N_i + p^2 \sum_{i \in S_I} N_i^2}{\sum_{i \in S_I} N_i}$
	Aldakuntza koefizientea $cv(\hat{P})$	$cv(\hat{P}) = \sqrt{(1-f) \frac{p(1-p)}{p(n-1)}}$	$cv(\hat{P}_{st}) = \frac{\sqrt{\sum_{h=1}^H N_h^2 (1-f_h) \frac{p_h(1-p_h)}{n_h - 1}}}{\sum_{h=1}^H N_h p_h}$	$cv(\hat{P})^2 = \frac{M-m}{M} \frac{m}{m-1} \frac{\sum_{i \in S_I} a_i^2 - 2p \sum_{i \in S_I} a_i N_i + p^2 \sum_{i \in S_I} N_i^2}{\left(\sum_{i \in S_I} a_i \right)^2}$

4. Laginketa-plan konplexuak

Aurkeztutako metodoak probabilitate-laginketetarako hiru mota nagusiak izan arren, EUSTAT-ek edo estatistika-organoek egindako inkesten diseinuak zehaztu nahi direnean, diseinu horiek zertxobait konplexuagoak izaten dira.

Laginketa bikoitza (edo bi etapakoa)

Demagun $U = \{1, \dots, k, \dots, N\}$ populazioa M azpipopulazioz osatuta dagoela, U_i , $i = 1, \dots, M$ azpipopulazioez, eta haiei lehen mailako unitate deritzela.

Era berean, lehen mailako U_i unitate bakoitza bigarren mailako N_i unitatez osatuta dago, non $\sum_{i=1}^M N_i = N$ baita.

Oro hor, laginketa bikoitza honela zehazten da:

- Lehen mailako unitatez osatutako m tamainako S_I lagin bat aukeratzen da.
- Lehen mailako unitate bat lehen etapan aukeratzen bada, bigarren mailako unitatez osatutako n_i tamainako S_i lagin bat hautatzen da.
- Plan bikoitzek propietate hauek bete behar dituzte: inbariantza eta independentzia.

Ausazko lagin osoa $S = \bigcup_{i \in S_I} S_i$ -ren bidez adierazten da, eta $n = \sum_{i \in S_I} n_i$ da haren tamaina.

Definizio hauek egin ditzakegu:

- $\pi_{I,i}$: lehen mailako U_i unitatea aukeratzeko probabilitatea.
- $\pi_{k|i}$: U_i aukeratua izan denez, k unitatea aukeratzeko probabilitatea.

Hortaz, k unitatearen sartze-probabilitatea hau da:

$$\pi_k = \pi_{I,i} \pi_{k|i}, \quad k \in U_i$$

Laginketa bikoitz batean, batez bestekoaren H-Tren zenbateslea hau da:

$$\hat{\bar{Y}}_\pi = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k} = \frac{1}{N} \sum_{i \in S_I} \sum_{k \in S_i} \frac{y_k}{\pi_{I,i} \pi_{k|i}} = \frac{1}{N} \sum_{i \in S_I} \frac{N_i \hat{\bar{Y}}_i}{\pi_{I,i}}$$

non $\hat{\bar{Y}}_i = \frac{1}{N_i} \sum_{k \in S_i} \frac{y_k}{\pi_{k|i}}$ lehen mailako U_i unitatearen batez bestekoaren H-Tren zenbateslea baita.

Gainera, bi etapako plan batean $\hat{Var}(\hat{Y}_\pi) = \hat{Var}_{UP} + \hat{Var}_{US}$,

non $\hat{Var}_{UP}$ lehen mailako unitateei dagokien bariantzaren zatia baita, eta $\hat{Var}_{US}$, berriz, bigarren mailako unitateei dagokiena.

Hortaz, laginketa bikoitz batean, batera erabil ditzakegu probabilitate-laginketa egiteko plan nagusiak (ausazko laginketa bakuna, geruzatua eta konglomeratuzkoa) lehen mailako unitateen eta bigarren mailako unitateen hautaketan.

Probabilitate berak dituzten lehen mailako unitateen aukeraketa

Demagun ausazko laginketa bakuna erabiltzen dela laginketaren bi etapetan.

Orduan, lehen zehaztutako probabilitateek forma hau hartzen dute:

$$\pi_{I,i} = \frac{m}{M}, \quad i = 1, \dots, M$$

$$\pi_{k|i} = \frac{n_i}{N_i}, \quad i = 1, \dots, M, \quad k \in S_i$$

Kasu honetan, k unitatearen sartzeko probabilitatea hau da:

$$\pi_k = \frac{mn_i}{MN_i}, \quad k \in U_i$$

Laginketa bikoitzetarako H-Tren zenbateslearen formula aldatzen badugu,

$$\hat{\bar{Y}}_\pi = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k} = \frac{M}{Nm} \sum_{i \in S_I} \sum_{k \in S_i} \frac{N_i y_k}{n_i}$$

Eta bariantzarako haren zenbateslea sinplifikatu egiten da

$$\hat{Var}(\hat{\bar{Y}}_\pi) = \frac{M-m}{N^2 m} M s_I^2 + \frac{M}{N^2 m} \sum_{k \in S_i} N_i \frac{N_i - n_i}{n_i} s_i^2$$

non

$$s_I^2 = \frac{1}{m-1} \sum_{i \in S_I} \left(\hat{Y}_i - \frac{\hat{\bar{Y}}_\pi}{M} \right)^2 \quad \text{baita, eta} \quad s_i^2 = \frac{1}{n_i-1} \sum_{k \in S_i} \left(y_k - \frac{\hat{\bar{Y}}_\pi}{N_i} \right)^2$$

Bi etapako plan autohaztatua

Demagun lehen mailako unitateak, lehen etapan, tamainarekiko proportzionala den sartze-probabilitatearekin aukeratzen direla; hau da,

$$\pi_{I,i} = \frac{N_i}{N} m$$

Bigarren etapan, bigarren mailako unitateak aukeratzen dira $n_i = n_0$ tamaina finkoko ausazko laginketa baten arabera (lehen mailako unitate bakoitzean); hau da,

$$\pi_{k|i} = \frac{n_0}{N_i}$$

Hortaz, k unitatearen sartze-probabilitateak berak dira U populazioko unitate guztientzat:

$$\pi_k = \pi_{I,i} \pi_{k|i} = \frac{N_i}{N} m \frac{n_0}{N_i} = \frac{mn_0}{N}$$

5. Kuboaren metodoa: laginketa orekatua

Kuboaren metodoak (Deville and Tillé, 2004) laginketa orekatuak aukeratzeko modua ematen du, sartze-probabilitateak berak izanik zein ez, probabilitate-laginketetarako metodoak optimizatuz.

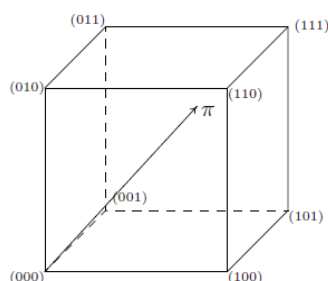
Intuitiboki, metodo honek aukera ematen du jatorriko populazioan zenbait oreka-aldagairi dagokienez (aldagai kualitatiboak) zeuden proportzioak laginean mantentzeko, betiere diseinuaren sartze-probabilitateak kontuan hartuta. Oreka-aldagai horiek korrelazio handia izan behar dute interes-aldagarekin.

Kubo bidezko irudikapena

Demagun N tamainako $U = \{1, \dots, N\}$ populazio finitu bat dugula, eta helburua interes-aldagai batzuen guztizkoa (edo batez bestekoa) zenbatestea dela.

Kuboaren metodoaren funtzionamendua ulertu ahal izateko, demagun lagin bat, egiazki, $\mathbf{s} = (s_1 \dots s_k \dots s_N)^t$ bektorea dela, eta bertan 1 balioa hartzen duela s_k -k, k unitatea laginean badago, edo 0 balioa laginean ez badago.

Geometrikoki, $\mathbf{s}$ bektore bakoitza N -kubo baten erpin bat da.



Litezkeen laginak $N=3$ tamainako populazio batean

Hortaz, $p(\cdot)$ lagin-diseinua $S = \{0,1\}^N$ multzoari legozkiokeen lagin guztien probabilitate-banaketa litzateke, k unitatearen sartze-probabilitatea $\pi_k = \Pr(S_k = 1)$ izanik.

Lagin orekatuak

Demagun $k \in U$ populazioko unitate guztietarako balio jakinak dituzten zenbait aldagai lagungarri ditugula.

Aldagai lagungarri horiek geruzatze-aldagai moduan (kualitatiboak) edo oreka-aldagai moduan (kualitatiboak edo kuantitatiboak) erabil litezke.

		Interes aldagaiak		Geruzatze aldagaiak			Oreka aldagai kuantitatiboak			Oreka aldagai kualitatiboak (indikatzailak)		
		Y_1 Y_2		E_1 ... E_h			X_1 ... X_p			$\tilde{X}_1$... $\tilde{X}_q$		
N tamainako populazioa	1	y_{11}	y_{21}	e_{11}	...	e_{h1}	x_{11}	...	x_{p1}	$\tilde{x}_{11}$	...	$\tilde{x}_{q1}$
	2	y_{12}	y_{22}	e_{12}	...	e_{h2}	x_{12}	...	x_{p2}	$\tilde{x}_{12}$	...	$\tilde{x}_{q2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$		...			...	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$		...			...	
	N	y_{1N}	y_{2N}	e_{1N}	...	e_{hN}	x_{1N}	...	x_{pN}	$\tilde{x}_{1N}$	...	$\tilde{x}_{qN}$
							$\sum_{k=1}^N x_{1k} \dots \sum_{k=1}^N x_{pk}$			$\sum_{k=1}^N \tilde{x}_{1k} \dots \sum_{k=1}^N \tilde{x}_{qk}$		

Lagina guztizko hauetan orekatzen da

Hortaz, esaten da **s lagin bat** $x_1, x_2, \dots, x_p$ **aldagaien gainean orekatua** dela oreka-ekuazioak betetzen badira:

$$\hat{\mathbf{X}}_{\pi} = \mathbf{X} \Leftrightarrow \sum_{k \in S} \frac{x_{kj}}{\pi_k} = \sum_{k \in U} x_{kj} \quad \forall s \in S \text{ con } p(s) > 0$$

$$j = 1, \dots, p$$

Hau da, $x_1, x_2, \dots, x_p$ aldagaien Horvitz-Thompsonen zenbatesleak laginean aldagai horiei populazioan dagozkien guztizkoen berdinak dira.

π sartze-probabilitateen bektorea beti lagin-diseinuak berak aurrez erabakita egongo da.

Oreka-murritzeta horietatik ondorioztatzen diren ekuazioek $N - p$ neurriko (Q) azpiespazio bat mugatzen dute R^N -n. Hortaz, azken batean, N-kuboan Q azpiespazioaren barruan gelditzen den erpin bat (lagin bat) aukeratzea da kontua.

Ezinezkoa denez zehaztasun osoz orekatutako laginak hautatzea, kuboaren metodoak gutxi gorabehera orekatutako laginak aukeratzeko dituen metodo bat ezartzen du.

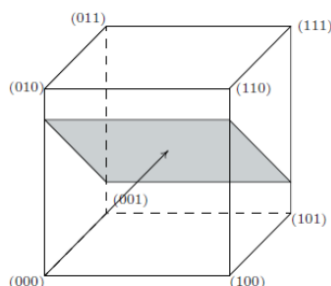
Metodoaren deskribapena

Devillek eta Tillék (2004) proposatutako kuboaren metodoak bi fase ditu:

1. Hegaldi-fasea

Zatiketa-metodoaren orokortzea da (Ikus "*Laginketaren teoria*").

Ausazko bide bat da, π sartze-probabilitatearen bektorean hasten dena eta kuboaren eta oreka-ekuazioek (Q) zehaztutako azpiespazioaren arteko ebakiduran jarraitzen duena.



2. Lurreratze-fasea

Hegaldi-fasearen amaieran lagin (erpin) bat ez bada aukeratu, lurreratze fasea aplikatu beharko da.

Hiru aukera daude fase honetarako:

- Oreka-aldagaiak progresiboki ezabatzea eta hegaldi-fasea berriz aplikatzea (beharrezkoa da aldagaiak garrantzi txikiena dutenetatik handiena dutenetaranzko ordenan ezabatzea).
- Programazio lineala erabiltzea, gutxi gorabehera orekatutako lagin onena kalkulatzeko (orekako aldea minimizatuz).
- Hegaldi-fasean lortzen den probabilitate-bektoretik hurbilen dagoen erpina aukeratzea, oraindik 0 edo 1 ez diren sartze-probabilitateak biribilduz.

Devillek eta Tillék hegaldi-fasearen askoz abiarazte azkarragoa programatu zuten (Ikus "*Fast SAS Macros for balancing simples user's guide*"), hark gauzatze-denboraren zatirik handiena hartzen baitu, eta abantaila hauek lortu zituzten:

- o Ez dago murrizketarik populazioaren tamainan.
- o Gauzatze-denborak lotura lineala du populazioaren tamainarekin.

6. SAS makroak lagin orekatuak aukeratzeko

Ondoren, lagin orekatuak aukeratzeko ahalbidetzen diguten SAS makroak aurkeztuko ditugu.

Bi makro nagusiak (*exe_cube* eta *echant_estrat*) Guillaume Chauvet-ek eta Yves Tillék garatu dituzte; *disjunctive* eta *crear_estrato* makro lagungarriak, berriz, Eustaten landu dira, aurrekoen erabilera azkartzeko.

Eustaten kuboaren metodoa ezartzen duten SAS makroekin lan egitea aukeratu den arren, lagin orekatuak R-n aukeratzeko dituzten funtzioak ere erabilgarri daude (ikus *sampling* paketea: <http://cran.r-project.org/web/packages/sampling/index.html>)

***exe_cube* makroa**

exe_cube izeneko SAS makroak bide ematen du lagin orekatuak aukeratzeko kuboaren metodoa erabiliz (Fast Cube Method).

Sarrerako datuak

SAS taula bat da, lagina aukeratzeko erabiliko den populazioaren unitate guztiak dauzkana.

Hauek izan behar ditu gutxienez:

- Identifikatzeko aldagaia.
- Sartze-probabilitateak dauzkan aldagaia.
- Lagina orekatzeko nahi diren aldagaiak.

Taulak ezin du baliorik eskas izan aipatutako aldagaietan.

Makroaren sintaxia

Hau beharrezko argumentuen deskribapen laburra da:

- BASE = SAS liburutegiaren izena; sarrerako datuak dituen taula dauka.
- DATA = Sarrerako datuak dituen SAS taularen izena.
- ID = Populazioko unitateak identifikatzeko aldagaia.
- PI = Sartze-probabilitateak dauzkan aldagaia.

- CONTR = Lagina orekatzeko nahi diren aldagaiak.
- ATTER = Lurreratze-faserako egindako aukera.

1. Oreka-aldagaiak progresiboki ezabatzen dira.
2. Litezkeen lagin guztiak kontuan hartzen dira gainerako unitateetarako (0 edo 1 ez diren balioak), orekarako alde txikiena ematen dutenak aukeratuz.
3. Bigarren aukerako prozedura bera, baina soil-soilik sartze-probabilitateen baturaren tamaina bereko laginak kontuan hartuz (lagin-tamaina finkoa).
4. Sartze-probabilitateak biribildu egiten dira gainerako unitateetarako, laginaren tamaina aurrez erabakita izanik.

Hirugarren edo laugarren aukerak erabiltzeko, sartze-probabilitateen aldagaia *contr.* parametroan sartu behar da.

- COMPEQ = Berdin 1, laginaren osagarria ere orekatzeko.
- SORT = Irteerako datuak dauzkan SAS taularen izena; *base* parametroan zehaztutako liburutegian gordeko da. Populazioko unitate guztiak dauzka, bai eta *ech* aldagaia ere. 1 balioa du, unitatea aukeratua izan bada; 0 balioa, bestela.

***echant_strat* makroa**

echant_strat deritzon SAS makroari esker, lagin geruzatuak aukera daitezke kuboaren metodoarekin (Fast Cube Method), orokorki orekatuak populazio osoan eta gutxi gorabehera orekatuak geruza bakoitzean.

Hauek dira makroak lagin orekatu bat aukeratzeko egiten dituen urratsak:

1. Hegaldi-fase independentea geruzetako bakoitzean.
2. Baterako hegaldi-fasea geruzetan aukeratuak izan ez diren gainerako unitate guztiekin.
3. Lurreratze-fasea oraindik aukeratu gabeko unitateekin.

Sarrerako datuak

Beharrezkoa da populazioko unitateak izango dituen SAS taula bat egotea geruzatutako laginerako zehaztutako geruza bakoitzerako.

Taula bakoitzak, gutxienez, *exe_cube* makrorako zehaztu ditugun aldagai berak izan behar ditu.

Makroaren sintaxia

Hau beharrezko argumentuen deskribapen laburra da:

- DATA = Geruza bakoitzeko sarrerako datuak dituzten SAS taulen izena.
- ID = Populazioko unitateak identifikatzeko aldagaia.
- PI = Sartze-probabilitateak dauzkan aldagaia.
- CONTR = Lagina orekatzeko nahi diren aldagaiak.
- SORT = Irteerako datuak dituen SAS taularen izena.

***disjunctive* makro lagungarria**

disjunctive izeneko SAS makroak aukera ematen du interes-aldagai bat edo gehiago aldagai disjuntutan banatzeko zenbait kategoriaren arabera. Makroak, gainera, kategoria horien izenak sartzeko aukera ematen du.

Deskripzioa

Demagun N tamainako populazio batean, interes-aldagaia Y izanik, eta aldagai kualitatiboa X ($1, 2, \dots, L$ balioak hartzen ditu), *disjunctive* makroak $Y^1, Y^2, \dots, Y^L$ aldagai disjuntuak itzultzen dizkigula, non

$$y_i^l = \begin{cases} y_i & x_i = l \quad \text{bada} \quad i = 1, \dots, N \\ 0 & x_i \neq l \quad \text{bada} \quad i = 1, \dots, N \end{cases} \quad \text{izanik}$$

Makroaren sintaxia

Hau beharrezko argumentuen deskribapen laburra da:

- DATA = Populazioaren datuak dauzkan SAS taularen izena.
 - VAR = Interes-aldagaia(k).
 - CATEG = Aldagai disjuntuak sortzeko kategoriak dauzkan aldagai kualitatiboa.
 - NOMBRES_CATEG (aukerakoa) = *categ* aldagaiak kategoriaren izenak.
- Besterik adierazi ezean: *categ1, categ2, ..., categL*.

Emaitzak eta irteerak

disjunctive makroak *var* interes-aldagaitik abiatuta sortutako aldagai disjuntuak gehitzen dizkio sarrerako taulari.

Aldagai berri horien izenak *var* aldagaiaren izena eta *nombres_categ* aldagaiak zehaztutako izenak lotuta lortzen dira (tartean “_” ikurra dutela).

Izen horiek *contr_categ* makro-aldagai lokalean gordetzen dira.

***crear_estrato* makro lagungarria**

crear_estrato izeneko SAS makroak aukera ematen du SAS taula bat zenbait taulatan banatzeko geruzatze-aldagai baten arabera.

Makroaren sintaxia

Hau beharrezko argumentuen deskribapen laburra da:

- DATA = Populazioaren datuak dauzkan SAS taularen izena.
- ID = Identifikatzeko aldagaiak.
- VAR_ESTRAT = Geruzatzeko erabili nahi den aldagaiak.

Emaizak eta irteerak

crear_estrato makroak SAS taula bat itzultzen du *var_estrat* aldagaiaren balio bakoitzerako.

Irteerako taulen izenak, besterik adierazi ezean, honelakoak dira: *estrato_{var_estrat}_j* non $\{var_estrat\}_j$ balioa *var_estrat*. aldagaiaren *j*-garren balioa baita.

Izen horiek *datos_estrat*. makro-aldagai lokalean gordetzen dira.

Makroen erabileraren adibidea

Demagun establezimenduen lagin geruzatu bat aukeratu nahi dugula, lagina lurralde historiko bakoitzeko enplegatu-kopurua baliatuz orekatuta.

Gure hasierako SAS taulak, populazioari buruzko datuekin batera, itxura hau izango luke:

datuak

<i>id</i>	<i>estratua</i>	<i>pik</i>	<i>enplegua</i>	<i>LH</i>
1	A	π_1	e_1	48
2	A	π_2	e_2	20
3	B	π_3	e_3	20
4	B	π_4	e_4	01
5	B	π_5	e_5	48
6	C	π_6	e_6	01
7	C	π_7	e_7	20

non

01 = Araba baita, 20 = Gipuzkoa eta 48 = Bizkaia;

π_k hau baita: *k* establezimenduaren sartze-probabilitatea;

e_k : enplegatu-kopurua *k* establezimenduan.

- Lehendabizi, *disjunctive* makroa aplikatuko dugu lurralde historiko bakoitzean enplegatutako kopuruari dagozkion oreka-aldagai disjuntuak kalkulatzeko.

```
% global contr_categ;  
% disjunctive(  
    DATA = datuak,  
    VAR = enplegua,  
    CATEG = LH,  
    NOMBRES_CATEG = Araba Gipuzkoa Bizkaia  
);
```

datuak

id	estratua	pik	emplegua	LH	emplegua _ Araba	emplegua _ Gipuzkoa	emplegua _ Bizkaia
1	A	π_1	e_1	48	0	0	e_1
2	A	π_2	e_2	20	0	e_2	0
3	B	π_3	e_3	20	0	e_3	0
4	B	π_4	e_4	01	e_4	0	0
5	B	π_5	e_5	48	0	0	e_5
6	C	π_6	e_6	01	e_6	0	0
7	C	π_7	e_7	20	0	e_7	0

Esan dugun bezala, lurralde historiko bakoitzean enplegatutako kopuruaz lagin orekatu bat aukeratzea da xedea; hau da, guztizkoez:

$$\sum_{k \in N} \text{emplegua_Araba}_k, \sum_{k \in N} \text{emplegua_Gipuzkoa}_k \text{ eta } \sum_{k \in N} \text{emplegua_Bizkaia}_k$$

Kasu honetan, *contr_categ* makroak balioak gordetzen ditu:

&contr_categ. = emplegua_Araba emplegua_Gipuzkoa emplegua_Bizkaia.

- Ondoren, *crear_estrato* makroa aplikatuko genuke dataset bat lortzeko geruzetako bakoitzari dagozkion datuekin.

```
% global datos_estrat;                                datuak
% crear_estrato(
    DATA = datuak,
    ID = id,
    VAR_ESTRAT = estratua
);
```

A_estratua

id	estratua	pik	emplegua	LH	emplegua _ Araba	emplegua _ Gipuzkoa	emplegua _ Bizkaia
1	A	π_1	e_1	48	0	0	e_1
2	A	π_2	e_2	20	0	e_2	0

B_estratua

id	estratua	pik	emplegua	LH	emplegua _ Araba	emplegua _ Gipuzkoa	emplegua _ Bizkaia
3	B	π_3	e_3	20	0	e_3	0
4	B	π_4	e_4	01	e_4	0	0
5	B	π_5	e_5	48	0	0	e_5

C_estratua

id	estratua	pik	emplegua	LH	emplegua _ Araba	emplegua _ Gipuzkoa	emplegua _ Bizkaia
6	C	π_6	e_6	01	e_6	0	0
7	C	π_7	e_7	20	0	e_7	0

Kasu honetan, *datos_estrat* makro-aldagaiak balioak gordetzen ditu:

&datos_estrat. = A_estratua B_estratua C_estratua

- Azkenik, *echant_strat* makrora joko dugu, geruzatutako laginetarako lagin orekatua aukera dezan kuboaren metodoaren bidez.

```
% echant_strat(
  DATA = &datos_estrat.,
  ID = id,
  PI = pik,
  CONTR = pik &contr_categ.,
  SORT = lagina
);
```

Makroaren irteerak honelako itxura izango luke:

lagina

<i>id</i>	<i>ech</i>
1	ech_1
2	ech_2
3	ech_3
4	ech_4
5	ech_5
6	ech_6
7	ech_7

$$\text{non } ech_k = \begin{cases} 1 & \text{baita, } k \text{ unitatea aukeratua izan bada} \\ 0 & \text{baita, bestela gertatuz gero} \end{cases} \quad k \in \{1, \dots, 7\} \text{ guztietarako}$$

* Oharra:

Batzuetan, xedea izan daiteke laginketa-unitateei berei erreferentzia egiten dieten guztizkoen gainean orekatzea lagina.

Adibidez, aurreko kasuan xedea izan liteke lagina lurralde historiko bakoitzeko establezimendu-kopurua kontuan hartuta orekatzea.

Kasu horretan, unitate guztietarako 1 balioa hartzen duen aldagai bat sortu behar dugu, eta hori % *disjunctive* makroan sartuko dugu, nahi ditugun oreka-aldagaiak sortzeko.

datuak

<i>id</i>	<i>estratua</i>	<i>pik</i>	<i>emplegua</i>	<i>LH</i>	<i>BAT</i>
1	A	π_1	e_1	48	1
2	A	π_2	e_2	20	1
3	B	π_3	e_3	20	1
4	B	π_4	e_4	01	1
5	B	π_5	e_5	48	1
6	C	π_6	e_6	01	1
7	C	π_7	e_7	20	1

```

% global contr_categ;
% disjunctive(
    DATA = datuak,
    VAR = BAT,
    CATEG = TH,
    NOMBRES_CATEG = Araba Gipuzkoa Bizkaia
);

```

datuak

<i>id</i>	<i>estratua</i>	<i>pik</i>	<i>emplegua</i>	<i>LH</i>	<i>BAT</i>	<i>BAT _ Araba</i>	<i>BAT _ Gipuzkoa</i>	<i>BAT _ Bizkaia</i>
1	A	π_1	e_1	48	1	0	0	1
2	A	π_2	e_1	20	1	0	1	0
3	B	π_3	e_1	20	1	0	1	0
4	B	π_4	e_1	01	1	1	0	0
5	B	π_5	e_1	48	1	0	0	1
6	C	π_6	e_1	01	1	1	0	0
7	C	π_7	e_1	20	1	0	1	0

7. EUSTATen kuboaren metodoaz orekatutako laginak

Ondoren, Eustat-en kuboaren metodoaz orekatu diren lagin-diseinuetako batzuk aurkeztuko dira.

Kasu bakoitzean, diseinu metodologikoa deskribatuko da: fitxa teknikoa, geruzatze-aldagaiak, esleipenak eta sartze-probabilitateak, eta lagina orekatzeko baliatutako aldagaiak. Lortutako emaitzetako batzuk ere aurkeztuko dira.

DBHko ikastetxeen lagina, 'bullying'-a aztertzeko Euskal Autonomia Erkidegoan

Hezkuntza, Unibertsitate eta Ikerketa Sailak, Ebaluatzeko eta Ikertzeko Erakundearen bidez, inkesta bat egin zien DBHko ikasleei, Euskal Autonomia Erkidegoko ikastetxeetako eskola-tratu txarrei buruz.

Horretarako, konglomeratu-lagin bat atera behar zen (ikastetxeena), aukeratutako ikastetxe bakoitzeko gehienez ere 40 ikasle ebaluatzeko.

Fitxa teknikoa

- **Eremua**

DBHko 1., 2., 3. eta 4. mailetan gutxienez talde bat duten EAEko Bigarren Hezkuntzako ikastetxeak.

- **Laginaren diseinua**

Konglomeratu desberdinen lagina, eta azpilaginketa bigarren fasean.

1. etapa

Lagin-unitateak

EAEko Bigarren Hezkuntzako ikastetxeak

Geruzatzea

Ikastetxeak aukeratzeko, laginketa geruzatu bat egin zen lurralde historiko eta sare (publiko nahiz pribatu) bakoitzeko.

Esleipena

Geruza bakoitzeko ikastetxe-kopuruaren proportzioan.

Zozketa

Probabilitate-laginketa ikastetxe bakoitzeko ikasle-kopuruaren tamainaren proportzioan.

2. etapa

Lagin-unitateak

EAEko Bigarren Hezkuntzako ikasleak

Geruzatzea

40 neska-ikasle (1. mailako 10, 2. mailako 10, 3. mailako 10 eta 4. mailako 10) aukeratutako ikastetxe bakoitzeko, betiere ahal izanez gero. Ez dago gutxieneko ikasle-kopururik ikastetxe bakoitzeko.

Zozketa

Ausazko laginketa bakuna.

Azken lagina geruzaka autohaztatua da (lurraldea eta sarea).

- **Laginaren tamaina**

Konglomeratu-laginketa baterako lagin hoberenaren tamaina formula honen bidez kalkulatu zen:

$$n_{ikastetxe} = n_a \frac{[(1 + \delta)(\bar{M} - 1)]}{\bar{M}}$$

non n_a ausazko bakun baterako laginaren tamaina baita, eta gainerakoa konglomeratu-laginketako diseinu-efektua delakoa.

Era berean:

$\bar{M}$ = Ikastetxe bakoitzeko batez besteko ikasle-kopurua

δ = Ikastetxe barruko korrelazioa.

$$n_a = \frac{N z_{\alpha/2}^2 S^2}{N e^2 + z_{\alpha/2}^2 S^2} = \frac{N}{\left[1 + (N - 1) \frac{e^2}{z_{\alpha/2}^2 p q} \right]}$$

N = Ikasle-kopuru osoa (oinarrizko unitateak).

e = Gehieneko errore onargarria.

$z_{\alpha/2}^2$ = Balio kritikoa α esangura-mailarako.

- **Oreka-aldagaiak**

Lagina aldagai hauek oinarri hartuta orekatu zen:

- Mailako bakoitzeko ikasle-kopurua eta maila bakoitzeko talde-kopurua.
Horrela, ikastetxe eta talde bakoitzeko ikasleen batez bestekoaren zenbatespenek Hezkuntza Estatistikak emandako datuen antz handiena dute.
- Tamaina-mota bakoitzeko ikastetxe-kopurua.

Ikastetxearen tamaina 5 taldetan kodetzea, klase barruko inertzia minimizatuz ikasle-tamainaren arabera: [0-143], [144-243], [244-361], [362-506] eta [507-708].

Emaitzak

Ondoren, kuboaren metodoaren bidez oreka-aldagaietarako lortutako emaitzak azaltzen dira.

Taula bakoitzean, populazio-banaketa eta laginaren jasotzaileak oinarri hartuta lortutakoa alderatzen dira. Ehunekoak zutabeka ageri dira.

Maila bakoitzeko ikasle-kopuruaren banaketa

	Populazioa	Lagina (jasoa)
DBH 1	19.664 (% 27,21)	19.617 (% 27,14)
DBH 2	18.633 (% 25,78)	18.649 (% 25,80)
DBH 3	17.669 (% 24,45)	17.764 (% 24,58)
DBH 4	16.306 (% 22,56)	16.243 (% 22,47)
GUZTIRA	72.272	72.272

Maila bakoitzeko talde-kopuruaren banaketa

	Populazioa	Lagina (jasoa)
DBH 1	870 (% 25,02)	869 (% 24,04)
DBH 2	852 (% 24,50)	849 (% 24,47)
DBH 3	896 (% 25,77)	896 (% 25,82)
DBH 4	859 (% 24,71)	856 (% 24,67)
GUZTIRA	3.477	3.470

Ikastetxe-kopuruaren banaketa tamaina-motaren arabera

	Populazioa	Lagina (jasoa)
1. tamaina	100 (% 30,12)	95 (% 28,79)
2. tamaina	128 (% 38,55)	129 (% 39,09)
3. tamaina	61 (% 18,37)	63 (% 19,09)
4. tamaina	31 (% 9,34)	31 (% 9,39)
5. tamaina	12 (% 3,61)	12 (% 3,64)
GUZTIRA	332	330

Kontuan hartuta lagina zer aldagai erabiliz orekatu den, ikastetxeko eta taldeko ikasleen batez bestekoaren oso zenbatesle onak ere lortu dira maila bakoitzerako.

2011/2012 IKASTURTEA	Ikasleen batez bestekoa ikastetxeko		Ikasleen batez bestekoa taldeko	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
DBH 1	59.23	59.44	22.60	22.57
DBH 2	56.21	56.51	21.90	21.97
DBH 3	53.22	53.83	19.72	19.83
DBH 4	49.11	49.22	18.98	18.98
GUZTIRA	217.69	219.00	20.79	20.33

Informazioaren Gizarteari buruzko Inkestarako lagina (IGI-Enpresak)

IGIren helburu orokorra (EUSTATen bidez gauzatua) da informazioaren eta komunikazioaren teknologia berriek (IKT) Euskal Autonomia Erkidegoko enpresetan izan duten sartzeko-mailari buruzko aldian behingo informazioa ematea arduradun politikoei, eragile ekonomiko eta sozialei, unibertsitateari, ikertzaile pribatuei eta herritarrei oro har.

IGI-Enpresen laginaren ezaugarria da urtero taula batean jasotzea inkestaren aurreko aldietan erantzun duten enpresa titullarrak. Zenbait gorabeheraren ondorioz (bajak, ordezkapenak, erantzunik eza...), laginaren hasierako banaketak balioa galtzen du; horregatik, lagina lagin-banaketa berri baten arabera eguneratzea erabaki zen, populazioak geruzetan daukan banaketa berria jasotzeko, hasierako diseinua errespetatuz.

2012an, taularen ia % 15 berritzeko erabakia hartu zen. Gainera, kuboaren metodoa sartu da lagin orekatuak aukeratzeko eta EAEko eskualdeetan banaketa orekatua lortzeko.

Fitxa teknikoa

- **Eremua**

EAEko esparruan diharduten edozein jarduera-sektoretako establezimenduak dira, lehen sektorea eta etxe-zerbitzua salbuetsita.

- **Laginaren diseinua**

Etapak bakarreko lagin geruzatua da.

Lagin-unitateak

Aipatu eremuko establezimendu guztiak.

Geruzatzea

Lagin geruzatu bat egiten da aldagai hauek gurutzatuz:

- Lurralde historikoa
1 = Araba; 2 = Bizkaia; 3 = Gipuzkoa
- Enplegu-geruza
1 = 0-5 enplegatu; 2 = 6-9 enplegatu; 3 = 10-19 enplegatu;
4 = 20-49 enplegatu; 5 = 50-99 enplegatu; 6 = 100 enplegatu eta gehiago.
- Jarduera-sektorea (CNAE09; 2 digitu).

Esleipena

Beren burua ordezkatzeko duten elementuak: 100 enplegatu eta gehiago dituzten establezimenduak (6. enplegu-geruza).

Gainerako establezimenduei dagokienez, bi esleipen desberdin egiten dira:

1. Hasierako diseinuan aurrez finkatutako $n=7.000$ ko lagin-tamaina oinarri hartuta, lurralde bakoitzeko establezimendu-kopuruaren erroaren proportzioan dagoen eta geruza (lurralde, jarduera eta enplegu) bakoitzeko establezimendu-kopuruaren zuzenki proportzionala den banaketa bat egiten da.

Geruza bakoitzean, laginaren tamaina formula honen bidez kalkulatzen da:

$$n_{TH_i Jar_j Enp_k} = n_{TH_i} \frac{estab_{Jar_j Enp_k}}{\sum_{j \in Jar} \sum_{k=1}^5 estab_{Jart_j Enp_k}}$$

$$\text{non } n_{TH_i} = (7000 - \text{zentsukoak}) \frac{\sqrt{estab_{TH_i}}}{\sum_{i=1}^3 \sqrt{estab_{TH_i}}} \quad i = 1, 2, 3$$

Azkenik, establezimenduak gehitzen dira, multzokatutako enplegu-geruzetan 5 establezimenduko gutxieneko tamaina lortu arte (10 enplegatu baino gutxiago eta 10 enplegatu baino gehiago).

2. Lagin-errorearen arabera banaketa gehienez ere % 10ekoa jarduera-sektore bakoitzean (zentsu-geruzak kontuan hartu gabe).

Jarduera-sektore bakoitzean, laginaren tamaina formula honen bidez kalkulatzen da:

$$n_h = \frac{N_h z_{\alpha/2}^2 S_h^2}{N_h e^2 + z_{\alpha/2}^2 S_h^2} = \frac{N_h}{\left[1 + (N_h - 1) \frac{e^2}{z_{\alpha/2}^2 pq} \right]}$$

non N_h = h geruzako establezimenduen kopurua baita.
 e = Gehieneko errore onargarria.
 $z_{\alpha/2}^2$ = Balio kritikoa α esangura-mailarako.

Bi esleipenak egin ondoren, falta diren unitateak banatu egiten dira, zentsukoak ez diren unitateetarako beharrezkoa den lagin-tamaina lortu arte. Banaketa hori geruzaren tamainarekiko proportzioan egiten da lehen esleipenari dagokionez gutxiegi ordezkatura dauden sektoreetan.

Azkenik, jarduera-sektorekako esleipen horiek erroarekiko modu proportzionalan banatzen dira multzokatutako lurralde eta enplegu bakoitzean.

Zozketa

Ausazko laginketa bakun bat egiten da geruza bakoitzean, eta lehentasuna ematen zaie eremuan alta moduan zehaztuta dauden establezimenduei.

• Oreka-aldagaiak

Eskualde-mailan zenbatespen hobeak lortzeko, lagina orekatu egin da eskualde bakoitzeko establezimendu-kopurua kontuan hartuta (20 eskualde).

- **Ordezkoak**

3.500 establezimendu inguruko ordezko-poltsa bat behar da lagina osatzeko. Geruzako ordezkoen kopurua lagin teorikoarekiko proportzionala da enplegu- eta lurralde-geruza bakoitzean.

Lagin nagusian bezala, ordezkoen lagina kuboaren metodoaren bidez orekatuko da eskualde bakoitzeko establezimendu-kopurua baliatuta.

Emaitzak

Ondoren, eskualdeko establezimendu-kopurua orekatzean kuboaren metodoaren bidez lortutako emaitzak azaltzen dira.

Establezimendu-kopuruaren banaketa eskualdeka

	Populazioa	Lagina (jasoa)
Arabako Haranak	405 (% 0.22)	523 (% 0.29)
Arabako Lautada	18.903 (% 10.49)	19.063 (% 10.58)
Arabako Mendialdea	248 (% 0.14)	257 (% 0.14)
Arabako Errioxa	1.311 (% 0.73)	1.135 (% 0.63)
Gorbeia Inguruak	780 (% 0.43)	749 (% 0.42)
Kantauri Arabarra	2.180 (% 1.21)	2.099 (% 1.16)
Arratia - Nerbioi	1.787 (% 0.99)	1.399 (% 0.78)
Bilbo Handia	73.572 (% 40.82)	72.517 (% 40.24)
Durangaldea	7.517 (% 4.17)	7.795 (% 4.33)
Enkarterria	2.356 (% 1.31)	2.364 (% 1.31)
Gernika – Bermeo	3.425 (% 1.90)	3.364 (% 1.87)
Markina – Ondarroa	1.828 (% 1.01)	2.446 (% 1.36)
Plentzia – Mungia	4.008 (% 2.22)	4.609 (% 2.56)
Bidasoa Beherea	7.169 (% 3.98)	8.343 (% 4.63)

Debabarrena	4.191 (% 2.33)	4.989 (% 2.77)
Debagoiena	4.197 (% 2.33)	4.742 (% 2.63)
Donostialdea	31.422 (% 17.44)	28.724 (% 15.94)
Goierri	4.929 (% 2.73)	5.192 (% 2.88)
Tolosaldea	4.029 (% 2.24)	4.105 (% 2.28)
Urola Kosta	5.966 (% 3.31)	5.809 (% 3.22)
GUZTIRA	180.223	180.223

Ehunekoak zutabeka ageri dira

Kapital Sozialari buruzko Inkestarako lagina

Kapital soziala baliabide moduan ulertzen da, arlo ekonomiko eta sozialetan modu aktiboan parte hartzeko aukera ematen duten sare pertsonal zabalak edukitzean lortzen dena. Parte-hartze hori, hain zuzen, konfiantzazko giroan egiten da, eta norberaren garapenean, garapen sozialean eta gizarte baten garapen ekonomikoan lagun dezake.

Zehazki, Eustatek egindako Kapital Sozialaren gaineko Inkestan, kapital soziala harremanei eta gizartearen parte-hartzeari lotutako dimentsio multzoa da eta honakoez osatuta dago: senide eta lagunaren sare sozialak, konfiantza pertsonengan eta erakundeengan, gizarte parte hartzea eta lankidetzatza, informazioa eta komunikazioa, kohesio soziala eta gizarteratzea, eta zoriona eta osasuna.

2012an, kapital sozialari buruzko inkestarako lagina kuboaren metodoa erabiliz aukeratzea erabaki da. Eta hala, sexuka eta adinka orekatua den lagin bat lortu dugu lurralde historiko bakoitzean, eskualde-mailan zenbatespen hobeak lortzen laguntzeaz gain.

Fitxa teknikoa

- Eremua**

Kapital sozialari buruzko inkestaren laginaren eremua Euskal Autonomia Erkidegoko etxebizitza eta establezimendu kolektiboetan bizi den 15 urteko eta 15 urtetik gorako populazioa da.

- Laginaren diseinua**

Etapa bakarreko lagin geruzatua da.

Lagin-unitateak

EAEko etxebizitza eta establezimendu kolektiboetan bizi den 15 urteko eta 15 urtetik gorako populazioa.

Laginaren tamaina

n = 7.000 banako aukeratzen dira.

Geruzatzea

Lagin geruzatu bat egiten da aldagai hauek gurutzatuz:

- Lurralde historikoa
01 = Araba; 20 = Gipuzkoa; 48 = Bizkaia
- Udalerriaren tamaina
Hiriburuak, Ertainak (20.000-100.000) eta Txikiak (20.000 eta hortik behera)
- Naziotasuna
0 = Nazionalak; 1 = Atzerritarak

Esleipena

Irizpide bat ezarri da geruzatze-maila bakoitzerako.

1. Lurralde bakoitzeko banakoen kopuruaren erro karratuarekiko banaketa proportzionala.
2. Udalerri-tamaina bakoitzeko banakoen kopuruarekiko banaketa proportzionala.
3. Naziotasun bakoitzeko banakoen kopuruaren 2/3 berreturarekiko banaketa proportzionala.

Esleipen egokiena aukeratzeko hirugarren mailan, egindako aurreko inkestako erantzun gabekoen tasak kontuan hartu dira (ECS 2007). Inkestako informazioa jasotzeko metodoak berak direnez gero, pentsa dezakegu oraingo inkestako erantzun-tasak antzekoak izango direla.

Hortaz, hiriburuei eta atzerritar populazioari dagokienez zenbatespenak eman ahal izateko beharrezkoa den gutxieneko lagin-tamaina (400 unitate inguru) lortzeko aukera ematen duen esleipena bilatu da, erantzun-tasa horiek kontuan hartuta.

Geruza bakoitzean, laginaren tamaina formula honen bidez zehaztuta dago:

$$n_{LH_iUDAT_jNAZ_k} = n_{LH_iUDAT_j} \frac{\sqrt[3]{(N_{LH_iUDAT_jNAZ_k})^2}}{\sum_k \sqrt[3]{(N_{LH_iUDAT_jNAZ_k})^2}}$$
$$\text{non } n_{LH_iUDAT_j} = 7000 \frac{\sqrt{N_{LH_i}}}{\sum_i \sqrt{N_{LH_i}}} \frac{N_{LH_iUDAT_j}}{\sum_j N_{LH_iUDAT_j}} \text{ baita,}$$

$i \in \{Araba, Gipuzkoa, Bizkaia\}$
 eta $j \in \{Hiriburuak, Ertainak, Txikiak\}$
 $k \in \{Nazionala, Atzerritarra\}$

Zozketa

Ausazko laginketa bakuna egiten da geruza bakoitzean.

- Oreka-aldagaiak**

Lagina aldagai hauek oinarri hartuta orekatu da:

- Banako-kopurua lurraldeen (Araba, Gipuzkoa eta Bizkaia), sexuen (gizonak eta emakumeak) eta adinen (15-24, 24-34, 35-44, 45-54, 55-64 eta 65 urtetik gora) gurutzetan.
- Banako-kopurua EAEko 20 eskualdeetako bakoitzean.

- Ordezkoak**

Beste 7.000 banakoko ordezkoko-poltsa behar da lagina osatzeko. Ordezko horiek ateratzean, jatorrizko lagineko geruzakako lagin-banaketa bera egin da, eta lagina titularretan erabilitako aldagai berak baliatuz orekatu da.

Emaitzak

Ondoren, kuboaren metodoaren bidez oreka-aldagaietarako lortutako emaitzak azaltzen dira.

Taula bakoitzean, populazio-banaketa eta laginaren jasogailuak oinarri hartuta lortutakoa alderatzen dira. Ehunekoak zutabeka ageri dira

Banaketa lurraldearen, sexuaren eta adinaren arabera

LH = ARABA (01)

	Gizonak		Emakumeak		GUZTIRA	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
15-24 urte	13.818 (% 10,06)	13.729 (% 10,02)	12.831 (% 9,24)	12.762 (% 9,17)	26.649 (% 9,65)	26.491 (% 9,59)
25-34 urte	23.028 (% 16,77)	22.923 (% 16,73)	21.541 (% 15,51)	21.725 (% 15,60)	44.569 (% 16,13)	44.648 (% 16,16)
35-44 urte	28.954 (% 21,08)	28.948 (% 21,13)	26.298 (% 18,93)	26.278 (% 18,87)	55.252 (% 20,0)	55.226 (% 19,99)
45-54 urte	24.889 (% 18,12)	24.895 (% 18,17)	24.891 (% 17,92)	25.039 (% 17,98)	49.780 (% 18,02)	49.934 (% 18,08)
55-64 urte	20.051 (% 14,60)	19.942 (% 14,55)	20.355 (% 14,65)	20.332 (% 14,60)	40.406 (% 14,63)	40.274 (% 14,58)
65 urtetik gora	26.584 (% 19,36)	26.590 (% 19,40)	33.009 (% 23,76)	33.086 (% 23,76)	59.593 (% 21,57)	59.676 (% 21,60)
GUZTIRA	137.324 (% 100)	137.027 (% 100)	138.925 (% 100)	139.222 (% 100)	276.249 (% 100)	276.249 (% 100)

LH = GIPUZKOA (20)

	Gizonak		Emakumeak		GUZTIRA	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
15-24 urte	30.206 (% 10,18)	30.273 (% 10,22)	28.416 (% 9,09)	28.371 (% 9,07)	58.622 (% 9,62)	58.644 (% 9,63)
25-34 urte	45.461 (% 15,32)	45.452 (% 15,34)	43.313 (% 13,86)	43.517 (% 13,91)	88.774 (% 14,57)	88.968 (% 14,60)
35-44 urte	60.481 (% 20,39)	60.491 (% 20,41)	56.318 (% 18,02)	56.361 (% 18,01)	116.799 (% 19,17)	116.852 (% 19,18)
45-54 urte	54.351 (% 18,32)	54.228 (% 18,30)	54.409 (% 17,41)	54.480 (% 17,41)	108.760 (% 17,85)	108.707 (% 17,84)
55-64 urte	45.126 (% 15,21)	44.881 (% 15,14)	46.428 (% 14,85)	46.525 (% 14,87)	91.554 (% 15,03)	91.406 (% 15,0)
65 urtetik gora	61.051 (% 20,58)	61.021 (% 20,59)	83.677 (% 26,77)	83.638 (% 26,73)	144.728 (% 23,76)	144.659 (% 23,74)
GUZTIRA	296.676 (% 100)	296.346 (% 100)	312.561 (% 100)	312.891 (% 100)	609.237 (% 100)	609.237 (% 100)

LH = BIZKAIA (48)

	Gizonak		Emakumeak		GUZTIRA	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
15-24 urte	47.497 (% 9,80)	47.673 (% 9,83)	45.007 (% 8,59)	45.152 (% 8,62)	92.504 (% 9,17)	92.825 (% 9,20)
25-34 urte	76.941 (% 15,87)	76.969 (% 15,88)	73.755 (% 14,07)	73.658 (% 14,06)	150.696 (% 14,94)	150.627 (% 14,93)
35-44 urte	97.104 (% 20,03)	97.136 (% 20,04)	93.542 (% 17,85)	93.318 (% 17,81)	190.646 (% 18,90)	190.454 (% 18,88)
45-54 urte	90.348 (% 18,64)	90.178 (% 18,60)	93.048 (% 17,75)	92.807 (% 17,71)	183.396 (% 18,18)	182.985 (% 18,14)
55-64 urte	72.330 (% 14,92)	72.308 (% 14,91)	77.119 (% 14,71)	77.329 (% 14,76)	149.449 (% 14,81)	149.637 (% 14,83)
65 urtetik gora	100.487 (% 20,73)	100.558 (% 20,74)	141.669 (% 27,03)	141.762 (% 27,05)	242.156 (% 24,0)	242.320 (% 24,02)
GUZTIRA	484.707 (% 100)	484.821 (% 100)	524.140 (% 100)	524.026 (% 100)	1.008.847 (% 100)	1.008.847 (% 100)

Banakoen kopuruaren banaketa eskualdeka

	Populazioa	Lagina (jasoa)
Arabako Haranak	5.107 (% 0,27)	5.051 (% 0,27)
Arabako Lautada	221.595 (% 11,69)	221.680 (% 11,69)
Arabako Mendialdea	2.855 (% 0,15)	2.886 (% 0,15)
Arabako Errioxa	9.852 (% 0,52)	9.835 (% 0,52)
Gorbeia Inguruak	7.296 (% 0,38)	7.292 (% 0,38)
Kantauri Arabarra	30.043 (% 1,58)	30.004 (% 1,58)
Arratia-Nerbioi	20.289 (% 1,07)	20.386 (% 1,08)
Bilbo Handia	768.311 (% 40,53)	767.962 (% 40,51)
Durangaldea	83.470 (% 4,40)	83.513 (% 4,41)
Enkarterria	27.787 (% 1,47)	27.742 (% 1,46)
Gernika-Bermeo	40.183 (% 2,12)	40.331 (% 2,13)
Markina-Ondarroa	23.128 (% 1,22)	23.333 (% 1,23)
Plentzia-Mungia	46.202 (% 2,44)	46.104 (% 2,43)
Bidasoa Beherea	66.403 (% 3,50)	66.418 (% 3,50)
Debarrena	47.748 (% 2,52)	47.664 (% 2,51)
Debagoiena	53.540 (% 2,82)	53.584 (% 2,83)
Donostialdea	282.424 (% 14,90)	282.508 (% 14,90)
Goierrri	57.859 (% 3,05)	57.781 (% 3,05)
Tolosaldea	40.147 (% 2,12)	40.193 (% 2,12)
Urola Kosta	61.490 (% 3,24)	61.462 (% 3,24)
GUZTIRA	1.895.729	1.895.729

Berrikuntza Teknologikoari buruzko Inkestarako lagina (BTI)

EUSTATEk berrikuntza teknologikoari buruz egindako inkestaren helburu nagusia da ekonomiaren sektoreek berrikuntzan egiten duten ahalegina ezagutaraztea eta, bide batez, Euskal Autonomia Erkidegoan (EAE) lortutako maila inguruko beste herrialde batzuetakoarekin erkatzeko aukera emango diguten adierazleak lortzea.

BTI-ren laginaren ezaugarria da urtero taula batean jasotzea inkestaren aurreko aldietan erantzun duten enpresa titularrak. IGI Eren kasuan bezala, laginaren hasierako banaketak balioa galtzen du zenbait gorabeheraren ondorioz (altak, bajak, aldaketak,...); horregatik, lagina lagin-banaketa berri baten arabera eguneratzen da, populazioak geruzetan daukan banaketa berria jasotzeko hasierako diseinua errespetatuz.

2012an, taularen ia % 7 berritzeko erabakia hartu zen. Gainera, kuboaren metodoa sartu da lagin orekatuak aukeratzeko eta EAEko eskualdeetan eta haietako hiriburuetan banaketa orekatua lortzeko.

Fitxa teknikoa

- **Eremua**

EAEko esparruan ari diren edozein jarduera-sektoretako establezimendu guztiek osatzen dute, honako sektore hauek salbuetsita: lehen sektorea, herri-administrazioa, elkartze-jarduerak, etxeetako jarduerak eta lurraldez kanpoko erakundeen jarduerak nahiz antolaketa-jarduerak.

- **Laginaren diseinua**

Etapak bakarreko lagin geruzatua da.

Lagin-unitateak

Aipatu eremuko establezimendu guztiak.

Geruzatzea

Lagin geruzatu bat egiten da aldagai hauek gurutzatuz:

- Lurralde historikoa
 - 1 = Araba; 2 = Bizkaia; 3 = Gipuzkoa
- Enplegu-geruza
 - 1 = 0-9 enplegatu; 2 = 10-49 enplegatu;
 - 3 = 50-249 enplegatu; 4 = 250 enplegatu eta gehiago
- Jarduera-sektorea (CNAE09; 2 digitu).

Esleipena

Beren burua ordezkatzeko duten elementuak: 250 enplegatu eta gehiago dituzten establezimenduak (4. enplegu-geruza), edo 2. eta 3. enplegu-geruzetan 46. jarduerari dagozkion establezimenduak.

Gainerako establezimenduetarako, honako esleipen teoriko hau egiten da:

- 2.400 establezimendu banatzen dira 10 enplegatuko eta gehiagoko geruzetarako, eta 750 establezimendu 10 enplegatu baino gutxiagoko geruzetarako.
- Banaketa lurralde eta enplegu-geruza bakoitzeko establezimendu-kopuruaren erroarekiko modu proportzionalan egiten da, eta, geroago, jarduera-geruza bakoitzeko establezimendu-kopuruaren erroarekiko proportzionala den beste esleipen bat egiten da.

Hau da, geruza bakoitzean laginaren tamaina formula honen bidez zehaztuta dago:

$$n_{LH_i, Enp_j, Act_k} = n_{LH_i, Enp_j} \frac{\sqrt{estab_{LH_i, Enp_j, Act_k}}}{\sum_{k \in Act} \sqrt{estab_{LH_i, Enp_j, Act_k}}} \quad \begin{matrix} i \in \{01, 20, 48\} \\ j \in \{1, 2, 3\} \end{matrix}$$

$$\text{non } n_{LH_i, Enp_j} = \begin{cases} 750 \frac{\sqrt{estab_{LH_i, Enp_j}}}{\sum_{j=1} \sqrt{estab_{LH_i, Enp_j}}} & \text{enplegua} < 10 \text{ bada} \\ 2400 \frac{\sqrt{estab_{LH_i, Enp_j}}}{\sum_{j \in \{2, 3\}} \sqrt{estab_{LH_i, Enp_j}}} & \text{enplegua} > 10 \text{ bada} \end{cases}$$

- Azkenik, establezimenduak gehitzen dira, geruza bakoitzean 5 establezimenduko gutxieneko tamaina lortu arte.

Geruza bakoitzeko beharrezko tamaina teorikoak kalkulatu ondoren, taulak dauzkan unitateak kentzen ditugu, geruza bakoitzean atera beharreko unitate-kopurua lortzeko. Zehazki, 2012. urtean 771 establezimendu atera behar izan dira.

Zozketa

Ausazko laginketa bakun bat egiten da geruza bakoitzean, eta lehentasuna ematen zaie eremuan alta moduan zehaztuta dauden establezimenduei.

- **Oreka-aldagaiak**

Eskualde-mailan zenbatespen hobeak lortzeko, 2. eta 3. enplegu-geruzei (10 enplegatu baino gehiago) dagokien lagina orekatu egin da eskualde bakoitzean (20 eskualde) eta hiriburuetan dagoen establezimendu-kopurua kontuan hartuta.

- **Ordezkoak**

Ordezko-poltsa behar da lagina osatzeko. Horretarako, 5 establezimendu aterako dira osorik ez dauden geruzetan. 2012. urtean, erreserbako 1.950 establezimendu atera dira.

Lagin nagusian bezala, ordezkoen lagina kuboaren metodoaren bidez orekatuko da eskualde bakoitzean eta hiriburuetan dagoen establezimendu-kopurua baliatuta.

Emaitzak

Ondoren, eskualde bakoitzeko eta hiriburuetak o establezimendu-kopurua orekatzean kuboaren metodoaren bidez lortutako emaitzak azaltzen dira.

Establezimendu-kopuruaren banaketa eskualde bakoitzean eta hiriburuetan (10 enplegatu baino gehiago)

	Populazioa	Lagina (jasoa)
Arabako Haranak	50 (% 0.40)	64 (% 0.51)
Arabako Lautada (hiriburua kanpo)	102 (% 0.81)	69 (% 0.54)
Arabako Mendialdea	14 (% 0.11)	19 (% 0.15)
Arabako Errioxa	105 (% 0.83)	93 (% 0.74)
Gorbeia Inguruak	97 (% 0.77)	156 (% 1.23)
Kantauri Arabarra	185 (% 1.47)	234 (% 1.86)
Arratia - Nerbioi	135 (% 1.07)	114 (% 0.91)
Bilbo Handia (hiriburua kanpo)	2.931 (% 23.26)	2.597 (% 20.61)
Durangaldea	648 (% 5.14)	556 (% 4.41)
Enkarterria	111 (% 0.88)	217 (% 1.72)
Gernika – Bermeo	162 (% 1.29)	271 (% 2.15)
Markina – Ondarroa	103 (% 0.82)	192 (% 1.52)
Plentzia – Mungia	200 (% 1.59)	333 (% 2.64)
Bidasoa Beherea	373 (% 2.96)	385 (% 3.06)
Debabarrena	359 (% 2.85)	290 (% 2.30)
Debagoiena	366 (% 2.90)	490 (% 3.88)
Donostialdea (hiriburua kanpo)	910 (% 7.22)	841 (% 6.67)
Goierri	334 (% 2.65)	387 (% 3.07)

Tolosaldea	311 (% 2.47)	419 (% 3.32)
Urola Kosta	390 (% 3.09)	263 (% 2.09)
Gasteiz	1.548 (% 12.28)	1.467 (% 11.64)
Bilbo	1.979 (% 15.70)	1.988 (% 15.78)
Donostia	1.190 (% 9.44)	1.158 (% 9.19)
GUZTIRA	12.603	12.603

Ehunekoak zutabeka ageri dira

• **Oharrak:**

1. Eskualde bakoitzeko establezimendu-kopuruaren jasogailuen kalkulua egiteko, geruzatze ondoko bat egin da; jarduera-geruzak multzokatu egin dira A38 eranste sektorialaren arabera (CNAE09), zabalkundean erabiltzen dena baita.
2. Hiru hiriburuetan, establezimendu-kopuruaren oso zenbatespen onak lortu dira.
3. Gainerako eskualdeei dagokienez, eta gehienak nahiko ondo zenbatetsita egon arren, errore erlatibo handiko eskualdeak aurki ditzakegu; besteak beste, Gorbeia Inguruak, Enkarterria, Gernika-Bermeo, Markina-Ondarroa, Plentzia-Mungia, Tolosaldea eta Urola-Kosta.
4. Zazpi eskualde horietan, kuboaren metodoak ez du lortu emaitza hobeak lortuko dituen lagin-konponbiderik, diseinuak berak ezarritako murrizketen ondorioz.
 - Laginaren tamaina 2.900 establezimendu ingurukoa izan arren, 410 baino ez dira zozkatu, gainerakoak taulatik nahiz errolda-geruzetatik zetozelako.
 - Gainera, lurraldea, jarduera eta enplegua gurutzatuta zehaztutako 401 geruzetatik, 173 geruzatan soilik aukeratzen dira establezimenduak.
 - Azkenik, zozketa benetan egiten den 173 geruzetatik, haietako 21etan alde zaurretik dago zehaztuta aukeratu beharreko establezimendua (altei lehentasuna eman behar izateagatik).

Pobreziari eta Gizarte Desberdintasunei buruzko Inkestarako lagina (PGDI)

Pobreziari eta Gizarte Desberdintasunei buruzko Inkestak (PGDIk) garrantzi handia dauka Justizia, Enplegu eta Gizarte Segurantza Sailarentzat, haren laguntza ekonomikoen ebaluazioari eta programazioari lotuta dagoelako. Horregatik, berebiziko garrantzia dauka inkestatu beharreko kolektibora ahalik eta hurbiltze egokiena egiteko aukera emango duen lagin-diseinua finkatzea.

Orokorrean, PGDIren helburu nagusia da pobrezia-ildoak, Euskadin duten eragina eta gizarte-desberdintasunari lotutako adierazleak ezagutzea, ikertzea eta ebaluatzea.

2012an, PGDIrako lagina kuboaren metodoa erabiliz aukeratzea erabaki da. Eta hala, lagin orekatu bat lortu ahal izan da, sexua, adina eta naziotasuna kontuan hartzen dituen, bai eta lurralde historiko bakoitzeko familia-tamaina ere.

Fitxa teknikoa

- **Eremua**

Pobreziari eta Gizarte Desberdintasunei buruzko Inkestarako laginaren eremua Euskal Autonomia Erkidegoko eta hartako lurralde historikoetako familia-etxebizitza okupatuek osatzen dute.

- **Laginaren diseinua**

Bi etapako lagina da; lehen etapan geruzatuta dago, eta bigarreanean, berriz, laginaren tamaina finkoa da.

Lagin-unitateak

EAEko familia-etxebizitza okupatuak.

Laginaren tamaina

4.000 inkesta-unitate inguru aukeratzen dira, eta 8.000 unitate inguru dira ordezeko (bi ordezeko lagin-unitate bakoitzeko).

Lehen etapa: Atalen lagina

Lehen etapan, EAEko zentsu-atalen zozketa egiten da.

- o **Geruzatzea**

Lehen etapako unitateak aldagai hauek gurutzatuta geruzatzen dira:

- **Eskualdeak eta koadrilak**

01 = Añana; 02 = Aiara; 03 = Kanpezu-Arabako Mendialdea;
04 = Guardia-Arabar Errioxa; 05 = Agurain;
06 = Gasteiz; 07 = Zuia; 08 = Donostialdea;
09 = Tolosaldea-Goierri; 10 = Debagoiena; 11 = Debarrena;
12 = Eskuinaldea; 13 = Bilbo; 14 = Ezkerraldea;
15 = Bizkaiko kostaldea; 16 = Durangaldea

- Tipologiak

Eustaten zentsu-atalen tipologien azterketa bat egiten da, berariazkoa PGDIrako. Horretarako, oinarritzko aldagai hauek kontuan hartzen dira: adina, sexua, nazionalitatea, jarduerarekiko harremana, etxebizitzako bizilagunen kopurua eta familiaren nahiz pertsonaren batez besteko errenta.

Osagai nagusien azterketa egin ondoren, atalak zazpi tipologiatan sailkatzen dira.

- Pertsona gazteen nagusitasuna:

45 urtetik beherako pertsonen presentzia erlatiboa handia izatea ezaugarri duten ataletan lagina gainordezkatzeko, atalak bi multzotan sailkatu dira:

1 = Gazteak nagusi diren atalak.

0 = Gainerakoa.

Bigarren etapan, 24 etxebizitza zozkatuko dira "gazteak" ataletan, eta 16 etxebizitza gainerakoan.

o **Esleipena**

4.000 etxebizitzaren zozketa esleipen hauen arabera egin da:

1. Lurralde historiko bakoitzeko etxebizitza-kopuruaren erro karratuarekiko banaketa proportzionala.
2. Eskualde/koadriletako etxebizitza-kopuruaren erro kubikoarekiko banaketa proportzionala.
3. Tipologia eta atal-mota ("gazte"/"ez-gazte") bakoitzeko etxebizitza-kopuruarekiko banaketa proportzionala.

Eskualdeko 160 etxebizitzako eta, Araban, koadrilako 112 etxebizitzako gutxieneko tamaina eskatzen da.

o **Zozketa**

Atalen zozketa probabilitistikoa eta tamainaren proportzionala izan da (PPT), eta betetako etxebizitza-kopuruan neurtua.

Bigarren etapa: Etxebizitzaren lagina

o **Esleipena**

Laginaren lehen etapan aukeratutako atal bakoitzeko, 16 edo 24 etxebizitza aukeratzen dira kasuan kasuko atal-motaren arabera.

o **Zozketa**

Ausazko zozketa bakuna egiten da lehen etapan aukeratutako atal bakoitzaren barruan.

• **Oreka-aldagaiak**

Lagina aldagai berak erabilita orekatu da lehen etapan bezala bigarreanean ere. Horri esker, azken lagina etxebizitza-eremu osoari dagokionez orekatuta egotea ziurtatzen dugu.

Aldagai orekatuak hauek dira:

- *Familia-tamaina*: 1 bizilaguneko etxebizitza-kopurua, 2 bizilagunekoa, 3-4 bizilagunekoa eta 5 bizilagunekoa edo gehiagokoa lurralde historiko bakoitzeko.
- *Sexua*: Gizon- eta emakume-kopurua lurralde historiko bakoitzeko.
- *Adina*: 34 urte edo gutxiago dituzten pertsonen kopurua, 35-44 urte dituztenena, 45-54 dituztenena, 55-64 dituztenena eta 65 edo gehiago dituztenena lurralde historiko bakoitzeko.
- *Naziotasuna*: Gizabanako nazionalen eta atzerriarren kopurua lurralde historiko bakoitzeko.
- *Gizabanakoen kopurua* eskualde/koadrila bakoitzean.

• Ordezkoak

Lagina osatzeko, ordezko bat eta erreserba bat zozkatzen dira etxebizitza bakoitzerako. Ordezko horiek lehen etapan aukeratutako zentsu-ataletako bakoitzean atera dira, eta lagina etxebizitza titularretan erabilitako aldagai berak baliatuta orekatu da.

Emaitzak

Ondoren, kuboaren metodoaren bidez oreka-aldagaietarako lortutako emaitzak azaltzen dira.

Taula bakoitzean, populazio-banaketa eta laginaren jasogailuak oinarri hartuta lortutakoa alderatzen dira. Ehunekoak zutabeka ageri dira.

Etxebizitzaren banaketa familia-tamainaren eta lurraldearen arabera

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
1 bizilagun	35.528 (% 27,77)	35.440 (% 27,70)	68.232 (% 24,97)	68.553 (% 25,09)	109.535 (% 24,44)	112.675 (% 25,14)
2 bizilagun	37.537 (% 29,34)	38.174 (% 29,84)	78.075 (% 28,57)	78.039 (% 28,56)	130.825 (% 29,18)	130.322 (% 29,07)
3-4 bizilagun	47.391 (% 37,04)	47.735 (% 37,31)	108.714 (% 39,78)	108.381 (% 39,66)	180.827 (% 40,34)	178.194 (% 39,75)
5 bizilagun edo gehiago	7.485 (% 5,85)	6.592 (% 5,15)	18.248 (% 6,68)	18.295 (% 6,69)	27.079 (% 6,04)	27.075 (% 6,04)
GUZTIRA	127.941	127.941	273.269	273.269	448.266	448.266

Banaketa sexuaren eta lurralde historikoaren arabera

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
Gizonak	157.836 (% 49,91)	155.759 (% 49,63)	344.561 (% 49,02)	347.363 (% 49,48)	553.674 (% 48,49)	551.028 (% 48,53)
Emakumeak	158.392 (% 50,09)	158.111 (% 50,37)	358.350 (% 50,98)	354.687 (% 50,52)	588.197 (% 51,51)	584.492 (% 51,47)
GUZTIRA	316.228	313.870	702.911	702.050	1.141.871	1.135.521

Banaketa adinaren eta lurralde historikoaren arabera

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
34 urte baino gutxiago	108.383 (% 34,27)	109.676 (% 34,94)	233.423 (% 33,21)	234.644 (% 33,42)	366.085 (% 32,06)	363.674 (% 32,03)
35 - 44 urte	55.227 (% 17,46)	49.691 (% 15,83)	116.445 (% 16,57)	116.922 (% 16,65)	188.762 (% 16,53)	194.045 (% 17,09)
45 - 54 urte	49.799 (% 15,75)	49.939 (% 15,91)	109.078 (% 15,52)	107.384 (% 15,30)	182.531 (% 15,99)	179.632 (% 15,82)
55 - 64 urte	40.810 (% 12,91)	43.836 (% 13,97)	92.261 (% 13,13)	91.599 (% 13,05)	151.434 (% 13,26)	146.342 (% 12,89)
65 urte edo gehiago	62.009 (% 19,61)	60.729 (% 19,35)	151.704 (% 21,58)	151.501 (% 21,58)	253.059 (% 22,16)	251.828 (% 22,18)
GUZTIRA	316.228	313.870	702.911	702.050	1.141.871	1.135.521

Banaketa naziotasunaren eta lurralde historikoaren arabera

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
Nazionala	286.633 (% 90,64)	289.847 (% 92,35)	658.599 (% 93,70)	659.521 (% 93,94)	1.067.272 (% 93,47)	1.059.925 (% 93,34)
Atzerritarra	29.595 (% 9,36)	24.023 (% 7,65)	44.312 (% 6,30)	42.529 (% 6,06)	74.599 (% 6,53)	75.595 (% 6,66)
GUZTIRA	316.228	313.870	702.911	702.050	1.141.871	1.135.521

Banakoen kopuruaren banaketa eskualdeka/koadrilaka

	Populazioa	Lagina (jasoa)
Añana	8.617 (% 0,40)	8.350 (% 0,39)
Aiara	34.208 (% 1,58)	33.894 (% 1,58)
Kanpezu - Arabako Mendialdea	3.156 (% 0,15)	3.118 (% 0,14)
Guardia - Arabar Errioxa	11.414 (% 0,53)	11.181 (% 0,52)
Agurain	12.255 (% 0,57)	12.384 (% 0,58)
Gasteiz	237.059 (% 10,97)	235.576 (% 10,95)
Zuia	9.519 (% 0,44)	9.368 (% 0,44)
Donostialdea	472.708 (% 21,87)	472.950 (% 21,98)
Tolosaldea - Goierri	114.584 (% 5,30)	113.420 (% 5,27)
Debagoiena	60.919 (% 2,82)	60.945 (% 2,83)
Debabarrena	54.700 (% 2,53)	54.734 (% 2,54)
Eskuinaldea	161.425 (% 7,47)	157.625 (% 7,33)
Bilbo	349.132 (% 16,16)	348.884 (% 16,22)
Ezkerraldea	386.068 (% 17,87)	379.912 (% 17,66)
Bizkaia-Kostaldea	126.504 (% 5,85)	127.321 (% 5,92)
Durangaldea	118.742 (% 5,49)	121.778 (% 5,66)
GUZTIRA	2.161.010	2.151.441

EAEko landa-eremuko emakumeak ikertzeko lagina

Nekazaritza, Arrantza eta Elikadura Sailak "Emakumea EAEko landa-esparruan. Beharrak, eskariak eta gizarte-gabeziak" gaiari buruz 1998tik egiten ari den ikerketa eguneratu nahi du.

2012an, aurreko diseinuetan ez bezala, Sailak landatartzat jotako udalerrietan bizi diren 15 urteko eta urte gehiagoko emakumeen eta gizonen lagin bana aterako da, irizpide hauek erabiliz: tamaina, biztanle-dentsitatea eta nekazaritzako BPG-aren proportzioa. Lagina 250 gizonen eta 250 emakumeen osatu beharko dute EAEko lurralde historiko bakoitzean.

Gainera, lagina kuboaren metodoa erabiliz aukeratzea erabakitzen da, eta lurralde historiko bakoitzean gizonen eta emakumeen lagin orekatua lortzen da hauek kontuan hartuta: adina, naziotasuna, ikasketa-maila eta etxebizitza-mota (nukleoa edo sakabanatua).

Fitxa teknikoa

- **Eremua**

Nekazaritza, Arrantza eta Elikadura Sailak landatartzat jotako 128 udalerrietako familia-etxebizitzetan bizi diren 15 urteko eta urte gehiagoko biztanleek osatzen dute laginaren eremua.

- **Laginaren diseinua**

Xedea landa-udalerrietan tamaina bereko emakume-lagin eta gizon-lagin bana ateratzea denez, bi etapako lagina egitea erabaki da, eta lehen etapan geruzatzeaz baliatzea. Banakoen azken lagina lurralde historiko bakoitzeko autohaztatua izateko moduan kalkulatzeko dira lehen eta bigarren etapetako esleipenak.

Era horretan, landa-udalerriak zozkatu ondoren, gizon- eta emakume-kopuru bera zozkatuko da udalerri bakoitzean.

Laginaren tamaina

250 gizon eta 250 emakume inguru zozkatzen dira EAEko lurralde historiko bakoitzean. Ez da ordezkorik zozkatuko, gainlagin bat egitea aukeratu baita, zenbatetsitako erantzunik gabekoen tasa kontuan hartuta (% 46 lurralde historiko bakoitzean).

Lehen etapa: Udalerrien lagina

Lehen etapan, EAEko 128 landa-udalerrien zozketa geruzatua egiten da.

- o **Lagin-unitateak**

EAEko landa-udalerriak. Zenbait neurritako banako-konglomeratuak dira.

- o **Geruzatzea**

Lehen etapako unitateak honela geruzatzen dira:

- Lurralde historikoa

01 = Araba; 20 = Gipuzkoa; 48 = Bizkaia

- Udalerrien tamaina

Udalerriak ezin hobeto geruzatuta daude tamainaren arabera; hau da, geruzatze horrek ahalik txikiena egiten du geruza bakoitzeko klase barruko inertzia edo barne-bariantza, guztizko bariantza edo inertzia erreferentzia moduan hartuta.

1 = [0-569]; 2 = [570-1154]; 3 = [1155-1884]; 4 = [1885-3400]

o **Esleipena**

Azken xedea lurralde historiko bakoitzean 250 gizon eta 250 emakume zozkatzea da.

Ez da ordezkorik zozkatuko, gainlagin bat egitea aukeratu baita, zenbatetsitako erantzunik gabekoen tasa kontuan hartuta (% 46 lurralde historiko bakoitzean).

Geruza bakoitzean zenbat udalerri zozkatu behar diren kalkulatzeko, prozedura hau erabili da:

1. Lurralde bakoitzean 500 banako geruzen tamainaren (biztanleen) proportzioan banatzea.
2. Lurralde bakoitzean zozkatu beharreko udalerrien kopurua kalkulatzeko, populazioaren laginketa-zatiaren multiplo bat oinarri hartuta.
3. Geruza bakoitzeko udalerri-kopuruarekiko banaketa proportzionala.
4. Udalerrien lagina handitu egiten da, 4 tamainako geruzakoak aukeratzeko.

o **Zozketa**

Banaketa teorikoa lortu ondoren, ausazko laginketa bakuna eginez zozkatzen dira landa-udalerriak.

Bigarren etapa: Gizonen eta emakumeen lagina

Bigarren etapan, inkestatuko diren gizonak eta emakumeak aukeratu behar ditugu.

o **Lagin-unitateak**

Lehen etapan aukeratutako landa-udalerrietako 15 urteko eta urte gehiagoko emakumeak eta gizonak.

o **Esleipena**

Laginaren lehen etapan aukeratutako landa-udalerri bakoitzeko, udalerriak geruzaren barruan duen tamainarekiko proportzioan kalkulatzeko da zozkatu behar diren gizonen eta emakumeen kopurua. Hau da,

$$n_{UDAL_i} = n_h \frac{Pop_{UDAL_i}}{Pop_h}$$

non MUN_i lehen etapan aukeratutako landa-udalerriak baitira, eta h udalerriari dagokion geruza.

o **Zozketa**

Ausazko bi lagin bakun eta independente ateratzen dira udalerri bakoitzeko gizonen eta emakumeen azpipopulazioen barruan.

Lurralde historikoen arabera autohaztatzen da gutxi gorabehera azken lagina.

• Oreka-aldagaiak

Lagina aldagai berak erabilia orekatu da lehen etapan bezala bigarrenean ere. Horri esker, azken lagina banako-eremu osoari dagokionez orekatuta egotea ziurtatzen dugu.

Aldagai orekatuak hauek dira:

- *Sexua*: Gizon- eta emakume-kopurua lurralde historiko bakoitzeko.
- *Adina*: 15-25 urte, 26-39 urte, 40-54 urte, 55-64 urte eta 65 urte edo urte gehiago dituzten gizabanakoen kopurua lurralde historiko bakoitzeko.
- *Naziotasuna*: Gizabanako nazionalen eta atzerritarren kopurua lurralde historiko bakoitzeko.
- *Ikasketak*: Lurralde historiko bakoitzean lehen mailako ikasketak, erdi-mailakoak edo goi-mailakoak dituzten gizabanakoen kopurua.
- *Etxebizitza-mota*: Nukleo-motako etxebizitzetan edo etxebizitza sakabanatuetan bizi diren gizabanakoen kopurua.

Emaizak

Ondoren, kuboaren metodoaren bidez oreka-aldagaietarako lortutako emaitzak azaltzen dira.

Taula bakoitzean, populazio-banaketa eta laginaren jasogailuak oinarri hartuta lortutakoa alderatzen dira. Ehunekoak zutabeka ageri dira.

Banaketa adinaren eta lurralde historikoaren arabera

SEXUA = GIZONAK

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
15 - 25 urte	1.705 (% 9,70)	1.676 (% 9,53)	1.231 (% 10,41)	1.236 (% 10,45)	1.769 (% 8,90)	1.807 (% 9,09)
26 - 39 urte	3.706 (% 21,08)	3.634 (% 20,67)	2.958 (% 25,01)	2.988 (% 25,26)	4.354 (% 21,91)	4.383 (% 22,06)
40 - 54 urte	5.746 (% 32,68)	5.807 (% 33,03)	3.396 (% 28,71)	3.320 (% 28,07)	6.169 (% 31,05)	6.260 (% 31,51)
55 - 64 urte	2.698 (% 15,35)	2.730 (% 15,53)	1.802 (% 15,23)	1.809 (% 15,29)	3.191 (% 16,06)	3.050 (% 15,35)
65 urte edo gehiago	3.727 (% 21,20)	3.734 (% 21,24)	2.442 (% 20,64)	2.476 (% 20,93)	4.386 (% 22,07)	4.369 (% 21,99)
GUZTIRA	17.582	17.852	11.829	11.829	19.869	19.869

SEXUA = EMAKUMEAK

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
15 - 25 urte	1.552 (% 9,91)	1.624 (% 10,37)	1.164 (% 10,73)	1.133 (% 10,45)	1.716 (% 8,99)	1.655 (% 8,67)
26 - 39 urte	3.351 (% 21,39)	3.309 (% 21,12)	2.709 (% 24,98)	2.658 (% 24,51)	3.970 (% 20,81)	4.058 (% 21,27)
40 - 54 urte	4.694 (% 29,96)	4.749 (% 30,31)	2.880 (% 26,56)	2.870 (% 26,47)	5.398 (% 28,29)	5.403 (% 28,32)
55 - 64 urte	2.133 (% 13,61)	2.067 (% 13,19)	1.416 (% 13,06)	1.481 (% 13,66)	2.714 (% 14,23)	2.708 (% 14,19)
65 urte edo gehiago	3.938 (% 25,13)	3.918 (% 25,01)	2.675 (% 24,67)	2.703 (% 24,93)	5.281 (% 27,68)	5.255 (% 27,54)
GUZTIRA	15.668	15.668	10.844	10.844	19.079	19.079

Banaketa naziotasunaren eta lurralde historikoaren arabera

SEXUA = GIZONAK

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
Nazionala	16.410 (% 93,33)	16.403 (% 93,29)	11.182 (% 94,53)	11.218 (% 94,83)	19.037 (% 95,81)	19.000 (% 95,63)
Atzerritarra	1.172 (% 6,67)	1.179 (% 6,71)	647 (% 5,47)	611 (% 5,17)	832 (% 4,19)	869 (% 4,37)
GUZTIRA	17.582	17.852	11.829	11.829	19.869	19.869

SEXUA = EMAKUMEAK

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
Nazionala	14.694 (% 93,78)	14.673 (% 93,65)	10.300 (% 94,98)	10.278 (% 94,78)	18.270 (% 95,76)	18.251 (% 95,66)
Atzerritarra	974 (% 6,22)	995 (% 6,35)	544 (% 5,02)	566 (% 5,22)	809 (% 4,24)	828 (% 4,34)
GUZTIRA	15.668	15.668	10.844	10.844	19.079	19.079

Banaketa ikasketak-mailaren eta lurralde historikoaren arabera

SEXUA = GIZONAK

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
Lehen mailako ikasketak	7.304 (% 41,54)	7.225 (% 41,09)	5.287 (% 44,70)	5.144 (% 43,49)	6.873 (% 34,59)	6.813 (% 34,29)
Erdi-mailako ikasketak	7.616 (% 43,32)	7.630 (% 43,40)	4.957 (% 41,91)	5.123 (% 43,41)	8.798 (% 44,28)	8.915 (% 44,87)
Goi-mailako ikasketak	2.662 (% 15,14)	2.727 (% 15,51)	1.585 (% 13,40)	1.562 (% 13,20)	4.198 (% 21,13)	4.141 (% 20,84)
GUZTIRA	17.582	17.852	11.829	11.829	19.869	19.869

SEXUA = EMAKUMEAK

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
Lehen mailako ikasketak	6.774 (% 43,23)	6.665 (% 42,54)	4.928 (% 45,44)	4.922 (% 45,39)	7.587 (% 39,77)	7.586 (% 39,76)
Erdi-mailako ikasketak	5.459 (% 34,84)	5.557 (% 35,47)	3.451 (% 31,82)	3.441 (% 31,73)	6.148 (% 32,22)	6.160 (% 32,29)
Goi-mailako ikasketak	3.435 (% 21,92)	3.446 (% 21,99)	2.465 (% 22,73)	2.482 (% 22,89)	5.344 (% 28,01)	5.333 (% 27,95)
GUZTIRA	15.668	15.668	10.844	10.844	19.079	19.079

Banaketa etxebizitza-motaren eta lurralde historikoaren arabera

SEXUA = GIZONAK

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
Nukleoa	16.555 (% 94,16)	16.743 (% 95,23)	7.530 (% 63,66)	7.891 (% 66,71)	11.750 (% 59,14)	12.245 (% 61,63)
Sakabanatua	1.027 (% 5,84)	839 (% 4,77)	4.299 (% 36,34)	3.938 (% 33,29)	8.119 (% 40,86)	7.624 (% 38,37)
GUZTIRA	17.582	17.852	11.829	11.829	19.869	19.869

SEXUA = EMAKUMEAK

	Araba		Gipuzkoa		Bizkaia	
	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)	Populazioa	Lagina (jasoa)
Nukleoa	14.781 (% 94,34)	14.977 (% 95,59)	7.223 (% 66,61)	7.687 (% 70,89)	11.555 (% 60,56)	12.072 (% 63,27)
Sakabanatua	887 (% 5,66)	691 (% 4,41)	3.621 (% 33,39)	3.157 (% 29,11)	7.524 (% 39,44)	7.007 (% 36,73)
GUZTIRA	15.668	15.668	10.844	10.844	19.079	19.079

Euskadi eta Drogak Inkestarako lagina

Euskadi eta Drogak bi urtetik behin egiten den inkesta da eta 15 urtetik 74 urtera bitarteko euskal herritarren droga-kontsumoko ohiturak ezagutzea du helburu. Era berean, herritar horiek drogekin eta drogomenpekotasunarekin lotutako gaien inguruan duen iritzia ere ezagutu nahi da.

2012an, inkestarako lagina kuboaren metodoa erabiliz aukeratzea erabaki da. Modu honetan, indibiduen kopuru totala osasun-eskualdeka, udalerri-tamainaka, sexuka eta nazionalitateka orekatua den lagin bat lortu dugu.

Fitxa Teknikoa

- Eremua**

Inkestaren laginaren eremua Euskal Autonomia Erkidegoko eta hartako lurralde historikoetako familia-etxebizitzetan bizi den 15 urtetik 74 urtera bitarteko populazioa da.

- Laginaren diseinua**

Etapak bakarreko lagin geruzatua da.

Lagin-unitateak

Euskal Autonomia Erkidegoko familia-etxebizitzetan bizi diren 15 urtetik 74 urtera bitarteko populazioa (erreferentzia data: 2012ko uztailaren 15a).

Laginaren tamaina

Eragiketa estatistikoaren zehaztapenen arabera, n=2007 banako titular eta horrenbeste ordezkari zein erreserba aukeratzen dira.

Geruzatzea

Lagin geruzatu bat egiten da aldagai hauek gurutzatuz:

- Lurralde Historikoa
01 = Araba; 20 = Gipuzkoa; 48 = Bizkaia
- Adin taldeak:
Hamar urtekako 6 adin talde
(15-24, 25-34, 35-44, 45-54, 55-64 eta 65-74 urte)

Esleipena

Irizpide bat ezarri da geruzatze-maila bakoitzerako:

1. Lurralde bakoitzeko banakoen kopuruaren erro karratuarekiko banaketa proportzionala
2. Lurralde bakoitzerako, adin talde gazteenen tamaina esleipen bikoitza (15-24 urte, 25-34 urte eta 35-44 urte).

Zozketa

Behin banaketa teorikoa kalkulatu eta gero, ausazko laginketa bakuna egiten da geruza bakoitzean.

- **Oreka aldagaiak**

Lagina aldagai hauek hartuta orekatu da:

- 15 urtetik 74 urtera bitarteko banako-kopurua EAEko 11 osasun-eskualdetan: Araba, Mendebalde Gipuzkoa, Ekialde Gipuzkoa, (Biz) Barrualdea, (Biz) Ezkerraldea-Enkarterri, (Biz) Uribe eta (Biz) Bilbo.
- 15 urtetik 74 urtera bitarteko banako-kopurua udalerritan, tamainaren arabera: kapitalak, 50.000-100.000 biztanle, 25.000-50.000 biztanle, 10.000-25.000 biztanle eta 10.000 biztanle arte..
- Banako-kopurua sexuka.
- Banako kopurura nazionalitateka.

- **Ordezkoak**

Banakoko bi ordezeko poltsa behar dira lagina osatzeko: ordezekoak eta erretserbak, 2007 unitatekoak kasu bakoitzean.

Ordezko horiek ateratzean, jatorrizko lagineko geruzakako lagin-banaketa bera egin da, titularretako laginean orekatutako aldagai berdinekiko orekatuz.

Emaitzak

Ondoren, kuboaren metodoaren bidez oreka-aldagaietarako lortutako emaitzak azaltzen dira.

Taula bakoitzean, populazio-banaketa eta laginaren jasogailuak oinarri hartuta lortutakoa alderatzen da. Ehunekoak zutabeka ageri dira.

Indibiduen banaketa osasun-eskualdeka

	Populazioa	Lagina (jasoa)
Araba	219.042 (% 13,28)	218.966 (% 13,28)
Mendebalde Gipuzkoa	218.155 (% 13,23)	218.335 (% 13,24)
Ekialde Gipuzkoa	328.814 (% 19,94)	329.009 (% 19,95)
(Biz) Barrualdea	227.787 (% 13,81)	228.032 (% 13,83)
(Biz) Ezkerraldea-Enkarterria	225.829 (% 13,70)	224.429 (% 13,61)
(Biz) Uribe	166.287 (% 10,08)	166.029 (% 10,07)
(Biz) Bilbo	263.028 (% 15,95)	264.141 (% 16,02)
GUZTIRA	1.648.942	1.648.942

Indibiduen banaketa udalerri-tamainaren arabera

	Populazioa	Lagina (jasoa)
Kapitalak	587.948 (% 35,66)	589.033 (% 35,72)
50.000 - 100.000 bitartean	184.970 (% 11,22)	184.638 (% 11,20)
25.000 - 50.000 bitartean	239.465 (% 14,52)	239.354 (% 14,52)
10.000 - 25.000 bitartean	300.173 (% 18,20)	300.088 (% 18,20)
10.000 arte	336.386 (% 20,40)	335.829 (% 20,37)
GUZTIRA	1.648.942	1.648.942

Indibiduen banaketa sexuaren arabera

	Populazioa	Lagina (jasoa)
Gizonak	823.310 (% 49,93)	823.742 (% 49,96)
Emakumeak	825.632 (% 50,07)	825.200 (% 50,04)
GUZTIRA	1.648.942	1.648.942

Indibiduen banaketa nazionalitatearen arabera

	Populazioa	Lagina (jasoa)
Nazionala	1.519.906 (% 92,17)	1.518.872 (% 92,11)
Atzerritarra	129.036 (% 7,83)	130.070 (% 7,89)
GUZTIRA	1.648.942	1.648.942

8. Ondorioak

Azkenik, ondorio batzuk aipatuko ditugu, laginketa orekatuak egiteak duen baliagarritasunari, oreka-aldagaiak aukeratzeari eta orekak geruzatzearekin eta kalibrazioarekin duen erlazioari buruzkoak.

Oreka eta geruzatzea

Bai geruzatzeko bai oreka lortzeko, populazioaren unitate guztietarako aldagai lagungarrien balioa zein den jakin behar dugu.

Geruzatzearen abantailarik handiena hau da: populazioa azpipopulazio homogeneoagoetan zatitzeko, zenbatesle zehatzagoak lortzeko eta laginketa-bariantza txikitzeko aukera ematen du. Interes-aldagaiarekin korrelazioan dauden aldagaietatik zenbat eta gehiagok parte hartu, orduan eta hobea da geruzatzea.

Hala ere, geruzatze-aldagai gehiegi erabiltzeak geruza txikiegiak sortzea eragin ditzake, eta lagin-tamaina behar adinakoa ez izatea; aipatu gabe ez erantzuteak geruza horietan ekar ditzakeen arazoak, nahiz eta hori geruza-kolapsoaren bidez konpon daitekeen (geruzatze ondokoa).

Oreka-aldagaiak bide ematen dute geruzatze anizkoitzean sar ez daitezkeen aldagaiak oreka-aldagai moduan gehitzeko, geruzatzearen abantaila guztiak mantenduz gehitu ere, bariantzaren murrizketari dagokionez, eta orekaren berezko abantailak gehituz.

Era berean, zenbait geruza edo arearen gurutzaketan zehaztutako esparrutan lan egiteko aukera ematen dute.

Oreka-aldagaiak kuantitatiboak izan daitezke; geruzatze-aldagaiak, ordea, beti kualitatiboak edo kategorikoak izan behar dute.

Oreka-aldagaien aukeraketa

Lagina orekatzeko aukeratutako aldagai lagungarriek oso korrelazio handia izan behar dute interes-aldagaiekin, baina beraien arteko korrelazioak ez du oso handia izan behar.

Lagina aldagai lagungarri kualitatibo ugari erabilita orekatzean, jatorriko populazioarenak ia bezalakoak diren banaketak dituzten guztizko zenbatetsiak lortzen dira (edo zenbatetsitako batez bestekoak).

Kuboaren metodoa oso interesgarria da lehen mailako unitateak aukeratzeko etapa anitzeko lagin batean. Bigarren etapan lagin orekatu bat ere aukeratuz gero, orekatu beharreko aldagaiak aurrez lehen etapan orekatuta egon behar dute.

Oreka eta kalibrazioa

Orekatzeko eta geruzatzeko ez bezala, kalibratzeko nahikoa da jakitea zer balio duten aldagai lagungarriek lagineko elementuei dagokienez, hala nola zein diren populazioan aldagai horien guztizkoak.

Estrategiarik onena oreka eta kalibrazioa batera erabiltzea da (ikus simulazioa *Deville and Tillé*-n, 2004); izan ere, orokorrean, emaitza hobeak lortzen dira lagina orekan erabilitako aldagai lagungarri berak erabiliz kalibratzen badugu.

Kasu honetan, ordea, kalibrazioa orekarakoak ez bezalako aldagaiak baliatuta erabil daiteke: aldagaiak berak direnean, baina une desberdinetan neurtuta.

Emaizten analisia

Jarraian, kuboaren metodoarekin orekatuak izan diren bi lagin (2012ko *Euskadi eta Drogak* eta 2012ko *Kapital Sozialari buruzko Inkesta*) kalibratzerakoan lortutako emaitzak aurkeztuko dira.

Kalibrazioa CALMAR (*calage sur marges*) bidez egin da bi kausetan, banakoen lagin pisuak “bir-haztatuz” kalibrazio aldagai lagungarrien totaletara doitu ahal izateko.

1. 2012ko Euskadi eta Drogak inkestaren kalibrazioa

Lagina ondorengo aldagaiekiko kalibratzea erabaki da 2012ko Euskadi eta Drogak inkestan ($n=2007$ indibiduo):

- Lurralde historikoa eta adinaren arteko gurutzapena (geruzaketa aldagaiak).
- Osasun-eskualdeak, udalerrien tamaina eta sexua (oreka aldagaiak)

$w_{hi} = w_h \forall i$ hasierako pisuetatik abiatuta (geruza bakoitzaren barruan pisu berdinak), zenbatesleak aldagai lagungarrien total marjinalari doitu ahal izateko w_{hi}^* amaierako pisuak lortu dira CALMAR makroaren ranking ratio metodoa erabiliz.

Amaierako pisuen eta hasierako pisuen arteko arrazoa bezala, $f = \frac{w_{hi}^*}{w_{hi}}$ aldagaia definitzen da.

Aldagai honen probabilitate-banaketa aztertuz, kalibrazio aldagaien total marjinalari doitu ahal izateko hasierako pisuak zenbat “deformatu” diren neurtu daiteke.

Hona hemen f aldagaiaren probabilitate-banaketaren laburpen txiki bat:

Media	1
Mediana	0.9987
Moda	0.9978
Desbiderapen estandarra	0.0875
 Aldakuntz-koefizientea	 8.75%
 Minimoa	 0.8365
Maximoa	1.2484

Ikus daitekeen bezala, pisu finalak ez dira hasierako pisuetatik gehiegi aldentzen (%24ko handitze maximoa eta %16ko txikitze maximoa), geruzatzeari lotutako lagin-unitateen pisuak mantenduz neurri handi batean.

2. 2012ko Kapital Sozialari buruzko Inkesta

Lagina ondorengo aldagaiekiko kalibratzea erabaki da 2012ko Kapital Sozialari buruzko Inkestan (n=4000 indibiduo):

- Lurralde historikoa (Araba, Gipuzkoa y Bizkaia)
- Sexua (gizonezkoak eta emakumezkoak)
- Adina (15-24, 25-34, 35-44, 45-54, 55-64 eta 65 urtetik gorakoak)

Beraz, lagina 36 total marjinaletikiko kalibratua izan da

Aurreko adibidean bezala, amaierako pisuen eta hasierako pisuen arteko arrazoia bezala, $f = \frac{w_{hi}^*}{w_{hi}}$ aldagaia definitzen da. Zenbatesleak aldagai lagungarrien total

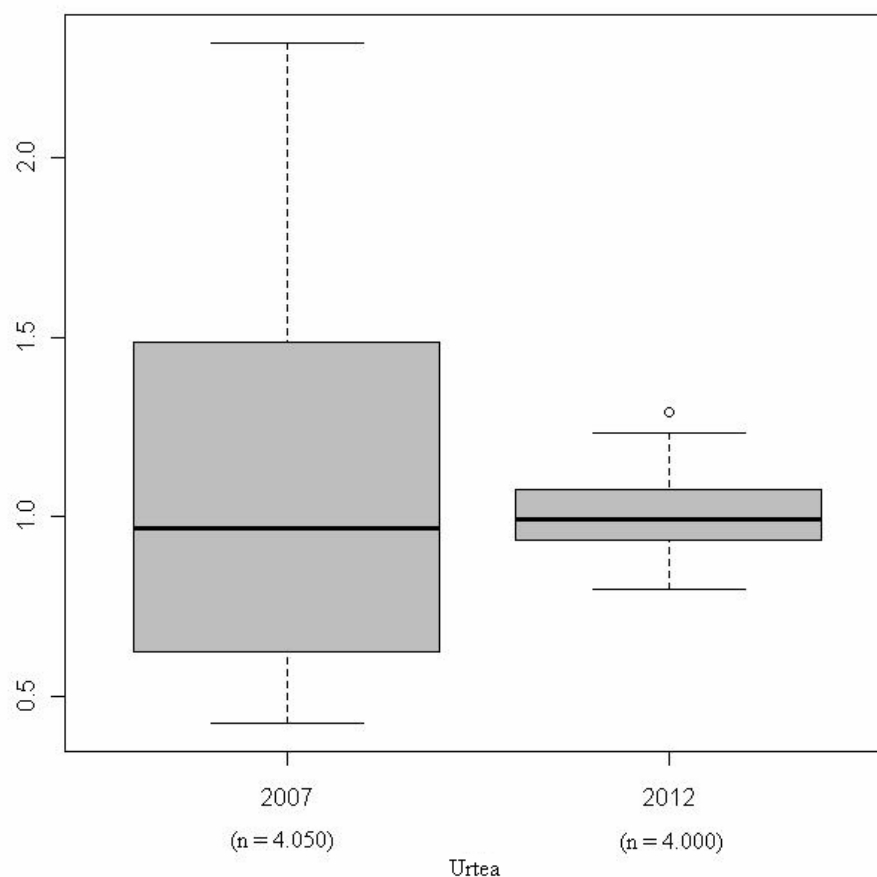
marjinali doitu ahal izateko, w_{hi}^* amaierako pisuak lortu dira CALMAR makroaren ranking ratio metodoa erabiliz.

Kasu honetan, f aldagaiaren probabilitate-banaketa aztertzeaz gain, 2007ko Kapital Sozialari buruzko Inkestarekin konparatuko ditugu emaitzak.

Gogora dezagun, bi inkestek lagin diseinu berdina izan arren, 2012ko ECS inkestaren lagina kuboaren metodoarekin orekatua izan dela. Erabili ditugun oreka aldagaiak, kalibrazioan erabili ditugunak izan dira hain zuzen ere.

Hona hemen 2007 eta 2012 urtetarako lortu ditugun emaitzen konparaketa:

PISUEN ARRAZOLA
(hamaierako pisua / hasierako pisua)



	2007	2012
Media	1.1139	1.0074
Mediana	0.9685	0.9944
Moda	2.0076	1.0287
Desbiderapen estandarra	0.5306	0.1125
Aldakuntz-koefizientea	47.63%	11.17%
Minimoa	0.4223	0.7965
Maximoa	2.3236	1.2915

2012ko ECS-ren lagina kalibrazio aldagaiekiko orekatua izan denez, lortutako emaitzak 2007 urtekoak baino hobeak izan dira, amaierako pisu hurbilagoak lortuz (%29ko handitze maximoa %132arekin alderatuz eta %20ko txikitze maximoa %58arekin alderatuz).

Laginketa orekatuaren interesa

Ereduak lagundutako eta ereduan oinarritutako eremuan, Horvitz-Thompsonen zenbateslearekin orekatutako laginketa-diseinua estrategiarik onena da askotan (ikus Nedyalkova and Tillé, 2009). Egia esan, lagin bat guztiz orekatua denean, aldagai lagungarrien H-Tren zenbatesleen bariantzek zero balioa dute.

Hauek dira laginketa orekatuaren abantailak:

- Probabilitatezko laginketa-diseinuen optimizazioa da, etapa bakarrekoak edo etapa anitzekoak izanda ere, eta haietan diseinuak zehaztutako sartze-probabilitateak ausazko laginak aukeratzeko abiapuntuko gakoa dira.
- H-Tren zenbateslearen zehaztasuna handitzen du; are gehiago, zenbateslearen bariantza soil-soilik interes-aldagaien eta oreka-aldagaien arteko korrelazioaren mende dago (erregresioaren hondarrak).
- Lagin kaltegarrienek, muturrekoenек edo batez bestekotik urrutien daudenek ia ez dute aukeratuak izateko probabilitaterik.
- Laginketa orekatuak laginen tamainak esparru geografikoetan edo barruti partikularretan txikiegiak ez izatea ziurtatzen du.

9. Bibliografia

- ADIN, A.; ARAMENDI, J.; GALBETE, E. AND IZTUETA, A. (2012)
El Método del Cubo: Un Método para seleccionar muestras equilibradas. Euskal Soziologia eta Zientzia Politikoaren Kongresua.
- ARDILLY, P. (1994)
Les Techniques de Sondage. Technip, Paris.
- ARDILLY, P. AND TILLÉ, Y. (2006)
Sampling Methods: Exercises and Solutions. Springer, New York.
- AZORÍN, F. AND SANCHEZ-CRESPO, J. L. (1986)
Métodos y Aplicaciones del Muestreo. Alianza Editorial, Madrid.
- CHAUVET, G. AND TILLÉ, Y. (2005)
Fast SAS Macros for balancing Samples: user's guide. Software Manual, University of Neuchâtel.
- CHAUVET, G. AND TILLÉ, Y. (2007)
Application of fast SAS macros for balanced samples to the selection of addresses. *Case Studies in Business, Industry and Government Statistics*, 1:173-182.
- COCHRAN, W. (1977)
Sampling Techniques. Wiley, New York.
- DEVILLE, J.-C. AND TILLÉ, Y. (2004)
Efficient balanced sampling: the cube method. *Biometrika*, 91:893-912.
- DEVILLE, J.-C. AND TILLÉ, Y. (2005)
Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128:569-591.
- KISH, L. (1965)
Survey Sampling. Wiley, New York.
- NEDYALKOVA, D. AND TILLÉ, Y. (2009)
Optimal sampling and estimation strategies under linear model. *Biometrika*, 95:521-537.

- SÄRNDAL, C.-E.; SWENSSON, B. AND WRETMAN, J. (1992)
Model Assisted Survey Sampling. Springer Verlag, New York.
- TILLÉ, Y. (2000)
Ten years of balanced sampling with the cube method: an appraisal.
Demographic Statistical Methods Division Seminar of the U.S. Census Bureau.
- TILLÉ, Y. (2005)
Teoría de Muestreo. Grupo de Statistique, Université de Neuchâtel, Suisse.
http://www2.unine.ch/files/content/sites/statistics/files/shared/documents/curso_teoría_de_muestreo.pdf
- TILLÉ, Y. AND MATEI, A. (2007)
The R Package Sampling. The Comprehensive R Archive Network, Manual of the Contributed Packages.
<http://cran.r-project.org/web/packages/sampling/sampling.pdf>
- TILLÉ, Y. (2010)
Laginketa orekatu eta eraginkorra: Kuboaren Metodoa. Euskal Estatistika Erakundea, Gasteiz.
http://www.eustat.es/productosServicios/datos/Seminario_52.pdf