

## ***Laginketa-erroreen kalkuluari buruzko txostena***

Biztanleria jardueraren arabera sailkatzeko inkesta  
(BJA)



## AURKIBIDEA

<b>1. Sarrera.....</b>	<b>3</b>
<b>2. Taylorren hedapen-metodoa .....</b>	<b>3</b>
<b>3. Erroreen kalkulua BJA .....</b>	<b>4</b>
3.1 Laginaren diseinua .....	4
3.2 Kalkulurako prozedura .....	5
3.3 BJA inkestako erroreak kalkulatzeko estatistikoak eta eremuak. ....	5
3.4 Emaitzak eta interpretazioa .....	7
<b>Bibliografia .....</b>	<b>9</b>

## 1. Sarrera

Laginketa-errorea definitzean zera esan dezakegu dela, aztergai den biztanleriaren ezaugarri bat (parametroa) biztanleria horren atal edo lagin batetik ateratako balioaren bidez (estatistikoa) zenbatestean egiten den zehazgabetasuna.

Errore hori faktore askoren mende dago, besteak beste, biztanleriaren atal hori hartzeko prozeduraren (laginaren diseinua), hartzen den unitate kopuruaren (laginaren tamaina), zenbatetsi nahi den ezaugarria nolakoa den eta abarren mende. Honako hau litzateke laginketa-errorearen kasuetan askotan erabiltzen diren esamoldeetako bat:

$$\text{Error de muestreo} = \sqrt{\text{Var}(\hat{\theta})} \quad (1)$$

Adierazpen horretan,  $\hat{\theta}$  estatistikoak du interesa (batezbestekoa, guztizkoa, proportzioa...). Estatistiko horrek balio desberdinak hartuko ditu, ateratako laginaren arabera. Estatistikoak laginketan duen aldakortasunaren mende egongo da laginketa-errorea.

Errore horren adierazpena desberdina izango da laginketarako erabili den teknikaren arabera; izan ere, zenbat eta konplexuagoa izan laginaren diseinua, hainbat eta zailagoa izango da kalkulua. Gainera, jasokarien eta azken pisuen kalkuluan aldaketak egin behar dira, informazioa biltzean izaten diren gorabeheren, biztanleriaren ezaugarri jakin batzuetarako egokitzapenaren (estratifikazio osteko fasea) eta inkesta burutzen den artean dauden beste faktore batzuen ondorioz

Literaturak laginketa-erroreak kalkulatzeko ohiko metodoen aldean zenbait ordezkari aukera iradoki izan ditu. Teknika heuristiko horien bidez laginketa-errorearen zenbatespen ona egin daiteke, laginaren diseinuaren azken pisuak eta ezaugarriak aintzat hartuta [3], [5].

Jarraian, metodo horiek aurkeztuko ditugu, eta azalduko dugu nola aplikatu diren, zehatz-mehatz, biztanleria jardueraren arabera sailkatzeko 2015eko inkestan.

## 2. Taylorren hedapen-metodoa

Taylorren hedapen-metodoaren bidez [4], [6] laginketa-akatsen zenbatespenak kalkula daitezke, guztizkoetarako, batezbestekoetarako eta proportzioetarako, estratifikazio, kluster eta probabilitate ezberdinak dituzten laginetan, EUSTATen estatistika eragiketa askotan gertatzen den moduan. Metodoak zenbateslearen hurbilketa linealak ateratzen ditu, eta bariantza kalkulatu du haren laginketa-errorearen zenbatesle gisa erabiliz.

Biztanleriaren batezbestekorako zenbatetsi den bariantza kalkulatzeko, adierazpen hau erabiltzen da:

$$\hat{V}(\hat{Y}) = \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_{h..})^2 \quad (2)$$

Non,

$$e_{hi} = \frac{\sum_{j=1}^{m_{hi}} w_{hij} (y_{hij} - \hat{Y})}{w_{...}}$$

$$\bar{e}_{h..} = \frac{\sum_{j=1}^{n_h} e_{hi}}{n_h}$$

y

$$w_{...} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$$

#### Idazkera:

$h = 1, 2, \dots, H$ -k geruza adierazten du, eta  $H$  geruza ditu guztira.

$i = 1, 2, \dots, n_h$ -k adierazten du zenbat kluster dauden  $h$  geruzan, jakinda  $n_h$  kluster daudela orotara.

$j = 1, 2, \dots, m_{hi}$ -k adierazten du  $h$  geruzako  $i$  klusterraren barruko unitate-zenbakia, eta guztira  $m_{hi}$  unitate daude

$n = \sum_{h=1}^H \sum_{i=1}^{n_h} m_{hi}$  lagineko behaketa-kopuru osoa da.

$w_{hij}$  da  $j$  behaketaren jasokaria  $h$  geruzako  $i$  **KLUSTERREAN**

$y_{hij} = (y_{hij}(1), y_{hij}(2), \dots, y_{hij}(P))$  dira  $Y$  aldagaien behatutako balioak,  $j$  behaketan,  $i$  klusterrean  $h$  geruzan. (zenbaki eta kategoria aldagaiak).

SAS [5] estatistika-multzoko PROC SURVEYMEANS prozedurak laginketa-erroreen zenbatespenerako metodoa aplikatzen du, eta eragiketa honetan laginketa-erroreak kalkulatzeko erabiliko dugu.

### 3. Erroreen kalkulua BJA

#### 3.1 Laginaren diseinua

Eragiketak probabilitatezko lagin bat du oinarri; lurralde historikoen arabera estratifikatuta dago, eta Eustaten Etxebizitzen Direktorioan jasotako unitateetatik aukeratuta. Etxebizitza nagusi guztiak daude zerrendan sartuta; hau da, familia baten ohiko bizileku guztiak [2]. Lagina etxebizitza-panel jarraitu batez dago osatuta, eta 1/8 errotazioarekin eguneratzen da hiruhileko bakoitzean. Etxebizitza bakoitzak bi urtez irauten du panelean.

2015ean hainbat aldaketa txertatu dira laginketaren diseinuan [3], eta haietatik honako hauek nabarmenduko ditugu: eskualdeen arabera estratifikazio osagarria, 20 eskualdeak eta hiru hiriburuak bereizita; laginketa orekatuaren aplikazioa, Cube metodoa [1], probabilitatezko laginketari, populazio-unibertsoaren ordezkari hobe lortuta, xede-aldagaiak zenbatesteko garrantzitsuak diren ezaugarri soziodemografikoei eta merkatu-ezaugarri dagokienez; eta laginaren denbora-banaketa uniformeagoa urte osoan zehar. Azkenik, laginketa-neurria 6.344 etxebizitzatara zabaldu da, dagoeneko ez baitira ordezkatzeko erantzuten ez duten etxebizitzak.

Informatzaile baten bitartez, etxebizitzako pertsona guztiei buruzko informazioa jasotzen da eta, hala, etxebizitzek konglomeratu txiki bezala jardungo dute, diseinuari dagokionez. Etxebizitzak dira hautatutako unitateak eta emaitzak norbanakoei buruzkoak dira.

### 3.2 Kalkulurako prozedura.

Hona hemen erroreen kalkulurako aplikatu den SASaren prozeduraren oinarritzko sintaxia [5]:

```
PROC SURVEYMEANS < fitxategiaren_izena > < irteerarako aukerak >;  
  BY aldagaiak ; /*erroreen kalkulua, azpipopulazio independenteen arabera*/  
  CLASS aldagaiak ; /*erroreen kalkulua, aldagai kualitatiboetarako*/  
  CLUSTER aldagaiak ; /*laginketan dagoen klusterra adierazten duen aldagaia,  
konglomeratuen arabera*/  
  DOMAIN aldagaiak ; /*erroreak kalkulatu nahi diren eremua/gurutzaketa mugatzen duten  
aldagaiak*/  
  RATIO aldagaia/aldagaia ; /*laginketa-errorea kalkulatu nahi den ratio-aldagaiak*/  
  STRATA aldagaiak < / option > ; /*estratifikatutako laginketan geruza adierazten duen  
aldagaia*/  
  VAR aldagaiak ; /* laginketaren erroreak kalkulatu nahi diren aldagai kuantitatibo eta  
kualitatiboak*/  
  WEIGHT aldagaia ; /* aurretik kalkulaturako pisu-aldagaia (aukerakoa)*/
```

BJAraiko soilik, honako hauek izango dira sintaxi honen parametro orokorrak:

CLUSTER = etxebizitzaren identifikatzailea.  
STRATA = Eustateko eskualdea, hiriburua  
WEIGHT = pertsonen hiruhilekoko jasokaria /familien hiruhilekoko jasokaria.  
Pertsonen urteko jasokaria /familien urteko jasokaria.  
RATIO = langabezia-tasak, jarduera eta lanbidea.  
VAR = langabezian eta lanean dauden biztanleak, biztanle aktiboak... guztira.  
DOMAIN = gurutzaketak, aldagai sozio-demografikoen eta ekonomikoen arabera. (ikus 3.3 atala)

### 3.3 BJA inkestako erroreak kalkulatzeko estatistikoak eta eremuak.

Hiruhileko laginketa-erroreen taula hauek argitaratuko dira, honako informazio hau bilduta:

- hamasei urteko edo gehiagoko biztanleak Euskal Autonomia Erkidegoan, jardueraren eta aldagai soziodemografikoen arabera; aldakuntza-koefizienteak eta konfiantza-tarteak, (%);
- Euskal Autonomia Erkidegoko populazio-tasa nagusiak, aldagai soziodemografikoen arabera; aldakuntza-koefizienteak eta konfiantza-tarteak (%);

Hurrengo taula hauetan, arestian aipatutakoa laburtzen da, estatistikoaren eta gurutzaketa-aldagaiaren arabera:

Estatistikoa/gurutzaketa aldagaia	Guztizkoa	Lurralde historikoa	Sexua	Adina (3 talde)	Sektore ekonomikoa	Beste ezaugarri batzuk
16 urteko eta gehiagoko biztanleak	X					
Biztanleria aktiboa	X	X	X			
Biztanleria okupatua	X	X	X		X	
Langabeak	X	X	X	X		
Biztanle ez-aktiboak	X					
Familia-etxeak	X					X
Jarduera-tasa	X	X	X			
Okupazio-tasa	X	X	X			
Langabezia-tasa	X	X	X	X		
Ez-aktiboaren tasa	X					

### 3.4 Emaitzak eta interpretazioa.

Laginketa-errorea zenbatestean gainera (2), SASek beste neurri baliagarri batzuk ere ematen ditu, errorea interpretatzen laguntzen dutenak. Hona hemen, besteak beste, interesgarrienak:

- **Aldakuntza-koefizientea.** Zehaztapenak multzo edo populazio desberdinen artean alderatzen lagatzen duen errorearen neurri erlatiboa da. Dimentsiorik ez duen magnitudea da, eta asko erabiltzen da laginketa-errorearen neurri gisa; hauxe du adierazpidea:

$$CV = \frac{\sqrt{\text{Var}(\hat{\theta})}}{\hat{\theta}} \quad (3)$$

- **Konfiantza-tartea** % 95era. Konfiantza-tarte hori estatistikoaren laginketaren banaketan oinarritzen da (proportzioa, batezbestekoa, tasa,...). Limitearen Teorema Nagusiaren bidez, gehien-gehienetan lege Normal<sup>1</sup> bat onar dezakegu estatistikorik ohikoenetarako; beraz, tarte horren eraikuntza honako adierazpide honen ondorio izango da:

$$\left[ \hat{\theta} - 1,96\sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} + 1,96\sqrt{\text{Var}(\hat{\theta})} \right] \quad (4)$$

1,96 balioa % 95eko probabilitatea daukan batezbestekoa 0 eta desbideratze tipikoa 1 dituen banaketa Normal baten pertzentila da. Hala, baieztatu daiteke  $\hat{\theta}$  estatistikorako kalkulaturako tarteak biztanleriaren parametroaren benetako balorea daukala kasuen % 95ean (balizko laginak).

SASek emandako informazioarekin, estatistikoaren zenbatespena, % 95eko konfiantza-tartearen beheko eta goiko mugak eta aldakuntza-koefizientea (ehunekoa) jasoko duten erroreen behin betiko taulak egingo dira. Jarraian erroreak zabaltzeko taularen eredu bat ageri da:

<sup>1</sup> Behar besteko lagin-neurria hartzen da ( $n > 30$ ). Hala egin ezin dezakegunean, konfiantza-tartea Student-en  $t$  banaketaren % 95eko pertzentilarekin kalkulaturiko da,  $n-1$  libertate-gradurekin.

**BIZTANLERIA JARDUERAREN ARABERA SAILKATZEKO INKESTA (BJA)**

**Euskal Autonomia Erkidegoko populazio-tasa nagusiak, aldagai soziodemografikoen arabera; aldakuntza-koefizienteak eta konfiantza-tarteak (%); II/2018**

Iturria: Eustat. Biztanleria jardueraren arabera sailkatzeko inkesta

	Zenbatespena	Beheko muga % 95	Goiko muga % 95	CV
	%	%	%	%
<b>JARDUERA-TASA</b>	<b>57,0</b>	<b>55,9</b>	<b>58,0</b>	<b>1,0</b>
Gizonezkoak	61,8	60,4	63,1	1,1
Emakumezkoak	52,5	51,2	53,8	1,3
Araba	59,5	56,9	62,1	2,2
Bizkaia	56,6	55,0	58,2	1,4
Gipuzkoa	56,4	54,7	58,2	1,6

Informazio hori interpretatzeko beste modu bat konfiantzaren **% 95ari dagokion errore erlatiboa kalkulatzeko da**; 1,96 pertzentila aldakuntza-koefizienteaz bideratuz ateratzen da hori. Errore erlatibo horrek zenbatespenaren balioaren ehuneko puntutan hitz egiteko aukera ematen digu.

Aurreko taularako, Euskal Autonomia Erkidegoko jarduera-tasaren %95erako errore erlatiboa % 2,1 da ( $1,96 \cdot 1,0$ ). Edo, bestela esanda, %95eko konfiantza mailan EAEko Jarduera tasaren benetako balioa emandako zenbatespenaren  $\pm 1,96$ ko tartean dabil. Taulan ageri diren datuak eskuratzeko (edo, hobeto esanda, haietara gerturatzeko, biribilduta baitaude), formula hau baliatu dezakegu:

$$[57,0 \pm (0,0196 \cdot 57,0)] \cong [55,9, 58,0]$$

Garrantzitsua da % 95erako errore erlatiboaren ehuneko jakin bat gainditzen duten zenbatespenak zein diren aipatzea, erabiltzaileak beharrezko zuhurtziaz joka dezan emandako informazio interpretatzeko orduan. Errore erlatiboaren % 20 gainditzen duten estimazioetan legoke zentzuzko atalase bat (bariazio-koefizientea > % 10 gutxi gorabehera), errore hori %30etik gorakoa den eremuak adieraziz bereziki (bariazio-koefizientea > % 15 gutxi gorabehera).

## **Bibliografia**

- [1] Deville, J.-C., Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91:893-912. [https://www.jstor.org/stable/20441151?seq=1#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/20441151?seq=1#page_scan_tab_contents)
- [2] EUSTAT (2015), "*Biztanleria jardueraren arabera sailkatze inkesta. Fitxa metodologikoa.*" [http://www.eustat.eus/document/poblact\\_c.html](http://www.eustat.eus/document/poblact_c.html)
- [3] EUSTAT (2015), "Populazioaren inkestan egindako aldaketa nagusiak, 2015eko BJA-jarduerari dagokionez." [http://www.eustat.eus/document/datos/Cambios\\_PRA\\_2015\\_c.pdf](http://www.eustat.eus/document/datos/Cambios_PRA_2015_c.pdf)
- [4] Fuller, W. A. (1975), "Regression Analysis for Sample Survey," *Sankhyā*, 37, Series C, Pt. 3, 117 - 132.
- [5] Sas Institute Inc. (2017), "SAS/STAT® 14.3 User guide. The SURVEYMEANS Procedure". [http://documentation.sas.com/#/?docsetId=statug&docsetTarget=statug\\_surveymeans\\_toc.htm&docsetVersion=14.3&locale=es](http://documentation.sas.com/#/?docsetId=statug&docsetTarget=statug_surveymeans_toc.htm&docsetVersion=14.3&locale=es)
- [6] Woodruff, R. S. (1971), "A Simple Method for Approximating the Variance of a Complicated Estimate" *Journal of the American Statistical Association*, 66, 411 -414.