

Laginketa-erroreen kalkuluari buruzko txostena

**Ingurumenari eta familiei buruzko inkesta
IFI- 2015**



AURKIBIDEA

1. Sarrera.....	3
2. Taylor-en hedapen-metodoa.....	3
3. Erroreen kalkulua	4
3.1 Laginaren diseinua	4
3.2 Kalkulatzeko prozedura	5
3.3 Inkestako erroreak kalkulatzeko estatistikoak eta eremuak.....	5
3.4 Emaitzak eta interpretazioa	7
Bibliografia.....	9

1. Sarrera.

Laginketa-errorea da biztanleriaren (estatistikoa) zati edo lagin batetik abiatuta lortutako balio baten bidez aztertuko den biztanleriaren (parametroa) ezaugarri bat zenbatesteko egiten den zehazgabetasuna.

Akats hori faktore askoren ondorioz gerta daiteke, besteak beste: biztanleriaren zati hori ateratzeko prozeduragatik (lagin-diseinua), ateratzen den unitate kopuruagatik (laginaren neurria), zenbatetsi beharreko ezaugarriaren izaeragatik, etab. Hona hemen laginketa-errorearen adierazpide bat, oso zabaldua dagoena:

$$\text{ERROR DE MUESTREO} = \sqrt{\text{VAR}(\hat{\theta})}$$

$\hat{\theta}$ intereseko estatistikoa litzateke (batez bestekoa, guztia, proportzioa,...). Estatistiko honek balio desberdinak hartuko ditu ateratako laginaren arabera. Laginketan estatistikoak duen aldakortasunak zehaztuko du laginketako akatsa.

Akats honen adierazpena aldatu egingo da erabilitako laginketa teknikaren arabera eta kalkulatzeko konplexuagoa izango da, laginketaren diseinua ere konplexuagoa den heinean. Gainera, informazioa biltzerakoan gertatzen diren gorabeherak, biztanleriaren ezaugarri batzuetara egokitzeak (estratifikazio ondokoa) eta inkestaren garapenean egondako beste faktore batzuek aldaketak ekartzen dituzte jasokariaren edo amaierako pisuen kalkuluan.

Literaturak laginketa-akatsen kalkuluetarako ohiko metodoentzako alternatibak iradoki ditu. Teknika heuristiko hauek laginketa-akatsen zenbatespen ona eskaintzen dute, amaierako pisuak eta laginketa diseinuko ezaugarriak kontuan izanda [2], [4].

Jarraian metodo hauek eta ezarpen zehatza sartuko ditugu, Ingurumenari eta Familiei buruzko Inkestaren (IFI aurrerantzean) 2008ko datuentzat.

2. Taylor-en hedapen-metodoa.

Metodo honen bidez [4] laginketa-akatsen zenbatespenak kalkula daitezke guztirakoetarako, batezbestekoetarako eta proportzioetarako estratifikazio, kluster eta probabilitate ezberdinak dituzten laginetan, EUSTATen estatistika eragiketa askotan bezala. Metodoak zenbateslearen hurbilketa linealak ateratzen ditu, eta bere bariantza kalkulatu du laginketa-akatsaren zenbatespen bezala erabiliz.

Populazioaren batez bestekorako zenbatetsitako bariantzaren kalkulurako adierazpena hau da:

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

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:

$h = 1, 2, \dots, h$ geruza da eta guztira H geruza daude.

$i = 1, 2, \dots, n_h$, h geruzan dagoen kluster kopurua da, eta guztira n_h kluster daude.

$j = 1, 2, \dots, m_{hi}$, h geruzaren i klusterraren barruko unitate kopurua da, eta guztira m_{hi} unitate daude.

$n = \sum_{h=1}^H \sum_{i=1}^{n_h} m_{hi}$ laginetako behaketen kopuru osoa da.

w_{hij} , j behaketak h geruzaren i klusterrean duen jasokaria da.

$y_{hij} = (y_{hij}(1), y_{hij}(2), \dots, y_{hij}(P))$, h geruzako i klusterraren j behaketan Y aldagaiaren gainean ikusi diren balioak dira (zenbaki eta kategoria aldagaiak).

SAS estatistika paketeko PROC SURVEYMEANS prozedurak [3] laginketa-erroreak zenbatesteko metodo hau inplementatzen du eta gure eragiketako laginketa-erroreak kalkulatzeko erabiliko den tresna izango da.

3. Erroreen kalkulua

3.1 Laginaren diseinua [1].

Laginaren esparrua: Bi etapadun eta zonaren arabera geruzatutako lagina da.

Lagina horrela dago orekatua: okupatzaileen ezaugarriengatik Lurralde Historikoetan eta arlo geografikoetan pertsona-kopuruagatik

Lagina diseinuaren ezaugarri nagusiak hauek dira:

Laginaren tamaina: 7.000

Laginaren esparrua: Eustaten Biztanleria Erregistroa eta Etxebizitzen Direktorioa

Laginaren diseinua: Ausazko Geruza laginketa

Geruzatzearen aldagaiak:

- Lurralde Historikoa
- Eremuak

Laginketaren etapak: Bi etapatan. Aurreneko etapa batean etxebizitzak aukeratzeko dira; bigarrenetan nork erantzun beharko duen norbanakoaren galdetegia.

Ateratzea: Ausazko orekatua

3.2 Kalkulatze prozedura.

Inkesta honetan akatsak kalkulatzeko ezarri den SAS prozedurako oinarritzko sintaxia hau da [3]:

```
PROC SURVEYMEANS < fitxategiaren_izena > < irteerarako aukerak >;  
  BY variables ; /*akatsen kalkulua azpipopulazio independenteen arabera*/  
  CLASS variables ; /*akatsen kalkulua aldagai kualitatiboentzat*/  
  CLUSTER variables ; /*aldagai honek laginketako klusterra konglomeratuka adierazten du eta  
laginketako lehen unitatea (PSU) etapa askoko laginketan*/  
  DOMAIN variables ; /*akatsak kalkulatu diren eremu/gurutzaketa mugatzen duten  
aldagaiak*/  
  RATIO variable/variable ; /*laginketako akatsa kalkulatu nahi zaien ratio aldagaiak*/  
  STRATA aldagaiak < / option > ; /*estratifikatutako laginketan geruza adierazten duen  
aldagaia*/  
  VAR variables ; /* laginketako akatsak kalkulatu nahi zaizkien aldagai kuantitatiboak eta  
kualitatiboak*/  
  WEIGHT variable ; /* aurre-kalkulatutako pisu aldagaia (aukerakoa)*/
```

IFIrako erabilitako sintaxi honen parametro orokorrak ondokoak izango dira:

STRATA = Geruza aldagaia, eremu geografikoak eta ataleko tipologiek eratua.

CLUSTER = Laginketako Lehen Unitatea identifikatzen duen aldagaia. Kasu honetan zentsu esparrua aldagaia izango da.

DOMAIN = sailkapen soziodemografikoko aldagaiak.

VAR = Familiako ingurumenaren aldagai kuantitiboak eta kualitatiboak.

WEIGHT = Etxebizitzaren edo pertsonaren jasotzailea, kalkulatu zenbatespenen arabera.

3.3 IFI inkestako erroreak kalkulatzeko estatistikoak eta eremuak.

Eragiketa honentzat Web orriko estatistika taulen atalean argitaratutako zenbatespen (ehunekoak, batez bestekoak, indizeak, etab.) guztientzako aldakuntza-koefizientearen taulak zabaltzen dira. Akatsen taula hauek argitaratu dira:

Aldakuntza-koefizienteen taula etxebizitzetarako, erreferentziazko pertsonaren ezaugarri soziodemografikoen arabera

- Euskal AEko etxebizitzak, ura aurrezteko ohituren eta gailuen arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, erabilitako energia motaren arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, berogailu motaren arabera (%). Aldakuntza-koefizienteak

- Euskal AEko etxebizitzak, isolatze eta argizatzearen arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, egunez duten tenperatura graduen arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, hondakinen tratamenduaren arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, etxetresna elektriko handien arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, ikus-entzunezko ekipamenduaren arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, etxetresna elektriko txikien arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, zaraten eta ekiditeko neurrien arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, kiratsen eta ekiditeko neurrien arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, erabiltzeko ibilgailu kopuruaren arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, produktu jakin batzuen erabileraren arabera (%). Aldakuntza-koefizienteak
- Euskal AEko etxebizitzak, erosteko faktore garrantzitsuen arabera (%). Aldakuntza-koefizienteak

Aldakuntza-koefizienteen taulak, 16 urte eta gehiagokoentzat, ezaugarri soziodemografikoen arabera

- Euskal AEko 16 urteko eta gehiagoko pertsonak, garraiobidearen arabera (%). Aldakuntza-koefizienteak
- Euskal AEko garraio publikoa erabilitako 16 urteko eta gehiagoko pertsonak (%). Aldakuntza-koefizienteak
- Euskal AEko oinez edo bizikletaz doazen 16 urteko eta gehiagoko pertsonak (%). Aldakuntza-koefizienteak
- Euskal AEko 16 urteko eta gehiagoko pertsonak, ingurumen iritzien arabera (%). Aldakuntza-koefizienteak
- Euskal AEko ingurumen jarduerak egindako 16 urteko eta gehiagoko pertsonak (%). Aldakuntza-koefizienteak
- Euskal AEko ingurumen neurrien aldeko 16 urteko eta gehiagoko pertsonak (%). Aldakuntza-koefizienteak

Aldakuntza-koefizienteen taula ingurumen adierazleentzako, erreferentziarako pertsonaren ezaugarri soziodemografikoen arabera

- Euskal AEn etxebizitzetako ingurumen adierazleak (%). Aldakuntza-koefizienteak
- Euskal AEn etxebizitzak, ingurumen adierazleen mailaren arabera (%). Aldakuntza-koefizienteak

3.4 Emaitzak eta interpretazioa.

Aldakuntza-koefizientea oinarri hartuta, akatsaren beste neurri batzuk kalkulatu daitezke, erabilgarriak izateaz gain, interpretatzen laguntzen baitute. Interesgarrienak hauek dira:

- **Aldakuntza-koefizientea.** Akatsaren neurri erlatibo bat da eta horren bidez talde edo populazio ezberdinen arteko zehaztapenak alderatu daitezke. Laginketa akatsaren batez besteko bezala oso erabiltzen den magnitude adimentsionala da eta honela adierazten da:

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

$\hat{\theta}$ intereseko estatistikoaren balioa da (batez bestekoa, guztizkoa, proportzioa...).

- **Fidagarritasun-tartea % 95.** Fidagarritasun-tartea laginketan estatistikoa banatzean oinarritzen da (proportzioa, batez bestekoa, tasa...). Limitearen teorema zentralaren bidez gehienetan lege Normal bat¹ har dezakegu, ohiko estatistikoetarako, beraz, tarte hori adierazpen honen bidez emango da:

$$(\hat{\theta} - 1,96\sqrt{Var(\hat{\theta})}, \hat{\theta} + 1,96\sqrt{Var(\hat{\theta})})$$

1,96 balioa 0 batez bestekoa duen banaketa Normal baten pertzentila da, ohiko desbiazioa 1 du eta probabilitatea % 95ekoa. Horrekin baieztatu dezakegu $\hat{\theta}$ estatistikorako kalkulaturako tartearik biztanleriaren parametroen egiazko balioa duela kasuen % 95ean (balizko laginak).

- Fidagarritasunaren % 95i dagokion akats erlatiboa: 1,96 pertzentila bider aldakuntza-koefizientea eginda lortzen da. Akats erlatibo horrekin zenbatespenaren balioen ehuneko puntuez hitz egin dezakegu.

Adibidez, Euskal AEn papera eta kartoia birziklatzen duten familiak % 92,3 baldin badira, % 0,5eko aldakuntza-koefizientearekin, % 95i dagokion akats erlatiboa % 0,98 izango da (hau da, 1,96 x 0,5). Edo, bestela esanda, % 95eko konfiantza mailan Euskal AEn papera eta kartoia birziklatzen duten establezimenduen ehunekoaren benetako balioa emandako zenbatespenaren \pm % 0,98ko tartean dabil. Hau da:

$$[92,3 \pm (0,0098 \times 92,3)] = [% 91,39, \% 93,2]$$

Garrantzitsua da akats erlatiboaren ehunekoa % 95ean gaintzen duten zenbatespenak kontuan hartzea, emandako informazioa interpretatzerakoan erabiltzaileak beharrezko neurriak hartzeko. Zentzuzko muga errore erlatiboaren % 20 gaintzen duten zenbatespenetan legoke (AK > % 10 gutxi gorabehera), errore hori % 30etik gorakoa zen laukitxoak bereziki azpimarratuz (AK > % 15 gutxi gorabehera).

¹ Lagin-tamaina behar bezain handia izatea asumitzen da ($n > 30$). Hala ez denean, konfiantza-tartea t-Student banaketaren % 95ari dagokion pertzentilarekin kalkulatu da, $n-1$ libertate-gradurekin.

Bibliografia

[1] EUSTAT. "Ingurumenari eta familiei buruzko inkesta Fitxa metodologikoa."
<http://www.eustat.es/document/EMAF08%5Fc.asp>

[2] Fuller, W. A. (1975), "Regression Analysis for Sample Survey," Sankhyā , 37, Series C, Pt. 3, 117 - 132.

[3] Sas Institute Inc. (2004), "SAS/STAT® 9.1 Guía de Usuario". Copyright © 2004, Cary, NC, USA. ISBN 1-59047-243-8

[4] Woodruff, R. S. (1971), "A Simple Method for Approximating the Variance of a Complicated Estimate" Journal of the American Statistical Association, 66, 411 -414.