

Laginketa-erroreen kalkuluari buruzko txostena

Kapital sozialari buruzko inkesta (KSI)

AURKIBIDEA

1. Sarrera.....	3
2. Taylor-en hedapen-metodoa.....	3
3. Erroreen kalkulua. KSI	4
3.1. Laginaren diseinua	4
3.2 Kalkulatzeko prozedura	5
3.3 KSIko erroreak kalkulatzeko estatistikoak eta eremuak	5
3.4 Emaitzak eta interpretazioa	6
Bibliografia.....	8

1. Sarrera.

Laginketa-errorearen definizioa honakoa litzateke: aztertutako biztanleriaren (parametro) ezaugarri bat biztanleria horren (estatistikoa) zati edo lagin batetik lortutako balioarekin zenbatzean egiten den zehazgabetasuna.

Akats hori faktore askoren ondorioz eman daiteke, besteak beste: biztanleriaren zati hori (laginaren diseinua) ateratzeko prozedura, ateratzen den unitate kopurua (laginaren neurria), zenbatu beharreko ezaugarriaren izaera, etab. Hona hemen laginketa-errorearen adierazpide bat, oso zabaldua dagoena:

$$\text{LAGINKETA} = \sqrt{V_A(\hat{\theta})}$$

$\hat{\theta}$ intereseko estatistikoa (batez bestekoa, guztizkoa, proportzioa,...) izanda. Estatistiko horrek balio desberdinak hartuko ditu, ateratako laginaren arabera. Estatistiko horrek laginketan izandako aldakortasunak adieraziko du laginketa-errorea.

Errore horren adierazpena aldatu egingo da erabilitako laginketa-teknikaren arabera, eta kalkulatzeko konplexuagoa izango da, laginketako diseinua konplikatuagoa den heinean. Gainera, informazioa biltzerakoan egoten diren gorabeherak, biztanleriaren ezaugarri batzuk zehaztu beharrik (estratifikazio ondorena) eta inkesta baten garapenean egoten diren beste faktoreek aldaketak ekartzen dituzte jasokariak edo amaierako pisuak kalkulatzeko.

Literaturak zenbait alternatiba iradoki ditu, laginketa-erroreak kalkulatzeko ohiko metodoentzako. Teknika heuristiko hauek laginketa-erroreen zenbaketa ona eskaintzen dute amaierako pisuak eta laginketako diseinuaren ezaugarriak oinarri hartuta [2], [4].

Hemendik aurrera aipatu metodoak sartuko ditugu eta, bereziki, Kapital Sozialari buruzko Inkestan (hemendik aurrera KSI) aplikatuko ditugu.

2. Taylor-en hedapen-metodoa.

Metodo honen bidez [4] laginketa-erroreen zenbatespenak kalkulatu daitezke guztirakoetarako, batezbestekoetarako eta proportzioetarako estratifikazioa, klusterra eta probabilitate ezberdinak dituzten laginetan, EUSTATen estatistika-eragiketa askotan bezala. Metodoak zenbateslearen hurbilketa linealak ateratzen ditu, eta bere bariantza kalkulatu du hau laginketa-errorearen zenbatespen bezala erabiliz.

Biztanleriaren batez bestekoarentzat zenbatestutako bariantzaren kalkulurako adierazpen hau erabiltzen da:

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

Kasu honetan:

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

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

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 geruzako i klusterraren barruko unitate kopurua da, eta guztira m_{hi} unitate daude.

$n = \sum_{h=1}^H \sum_{i=1}^{n_h} m_{hi}$ laginketako ohar guztiak dira.

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

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

SAS estatistika-multzoko PROC SURVEYMEANS prozedurak [3], inplementatzen du laginketa-erroreak zenbatzeko metodo hau eta tresna hori erabiliko da eragiketa honetako laginketa-erroreak kalkulatzeko.

3. Erroreen kalkulua – KSI

3.1. Laginaren diseinua [1].

KSIn laginaren diseinuaren ezaugarri nagusiak hauek dira:

Laginketa unitatea: Etxebizitza eta establezimendu kolektiboetan bizi diren 15 urte eta gehiagoko biztanleak.

Laginaren benetako neurria: 7.000

Laginaren esparrua: Biztanleria erregistroa eta Etxebizitzen direktorioa (Eustat).

Laginaren diseinua: Estratifikatutako ausazko laginketa

Estratifikazioaren aldagaiak:

- Lurralde Historikoa
- Udalerriaren tamaina: Hiriburuak (100.000 biztanle baino gehiago), udalerrri ertainak (100.000-20.000 biztanle) eta txikiak (20.000 biztanle eta gutxiago)

Laginketaren etapak: Etapa bakarra, banakoa zuzenean aukeratuta

Ateratzea: Ausazkoa ekilibratua

Laginaren diseinu hau ongi egokitzen da aurreko atalean deskribatutako metodo heuristikoaren berezitasunetara. SAS prozedurak eskatutako parametroak baino ez dira adierazi behar, bariantza zuzen zenbatesteko.

3.2 Kalkulatzeko prozedura.

Inkesta honetako akatsen kalkulurako inplementatutako SAS prozeduraren oinarritzko sintaxia honakoa da [3]:

```
PROC SURVEYMEANS < fitxategiaren izena > < irteerarako aukerak >;  
  BY aldagaiak ; /* erroreen kalkulua mendeko azpipopulazioko*/  
  CLASS aldagaiak ; /*erroreen kalkulua aldagai kualitatiboekiko*/  
  CLUSTER aldagaiak ; /*konglomeratuekiko laginketan klusterra adierazten duen aldagaia*/  
  DOMAIN aldagaiak ; /*akatsak kalkulatzaren domeinu/gurutzaketa mugatzen duten aldagaiak*/  
  RATIO aldagaia/aldagaia ; /*laginketa errorea kalkulatu nahi den ratioen aldagaiak*/  
  STRATA aldagaiak < / option > ; /*estratifikatutako laginketan geruza adierazten duen aldagaia*/  
  VAR aldagaiak ; /* laginetako akatsak kalkulatu nahi diren aldagai kuantitatiboak eta kualitatiboak*/  
  WEIGHT aldagaia ; /* aurre-kalkulatutako pisu aldagia (aukeran)*/
```

KSIrako erabiliko diren sintaxi honen parametro orokorrak ondokoak izango dira:

CLASS = kapital sozilaaren aldagaiak kategorizatuta (kategoriak: altua, ertaina, baxua).
VAR = kapital sozialaren aldagai kuantitatiboak (batez bestekoak)
STRATA = Lurralde Historikoa eta udalerrriaren tamaina.
DOMAIN = Sailkapen soziodemografikoko aldagaiak. (Ikus 3.3 atala)
WEIGHT = Pertsonaren jasokaria.

3.3 KSIko erroreak kalkulatzeko estatistikoak eta eremuak

Web orriko estatistika taulen atalean argitaratutako zenbatespen (ehunekoak eta batez bestekoak) guztientzako aldakuntza-koefizientearen taulak zabaltzen dira. Taula hauek dira:

Aldakuntza-koefizienteen taulak ehunekoentzako eta batez bestekoentzako, ezaugarri soziodemografikoen arabera - KSI

- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: sare zabala eta hurbileko sarea, sareko harreman pertsonalak, ezaugarri soziodemografikoen arabera (%).

- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: lagun sarearen homogeneitatea, emandako eta jasotako laguntza, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: finantza laguntzarako, osasun arazoetarako eta arazo emozionaletarako irisbidea, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: lanbide eta sareetako konfiantza orokorra, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: elkarrekikotasuna eta erakundeekiko konfiantza, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: gizarte eta politika gaiekiko interesa eta hauteskundeetako partaidetza, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: lankidetzat, eragin eta independentzia pertsonalak, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: gizartean, politikan eta elkarteetan duten parte hartzea, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: gizarte eta politika gaietako buruzko informazioa, Interneten eta sare birtualen erabilera, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: komunikabideetarako sarbidea eta informazio-iturri mota, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: gizarteko kohesioa, segurtasuna eta hustelkeria, ezaugarri soziodemografikoen arabera (%).
- Aldakuntza-koefizienteak (%) 15 urteko eta gehiagoko biztanleentzat, honakoak kontuan hartuta: zorientasuna eta osasuna, ezaugarri soziodemografikoen arabera (%).

3.4 Emaizak 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:

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

$\hat{\theta}$ 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:

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

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 tartea 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.

Aurreko taularako, Euskal AEn sare zabalaren batez besteko neurriaren % 95erako errore erlatiboa % 1,53 da (hau da, $1,96 \times 0,78$). Edo, bestela esanda, % 95eko fidagarritasun mailan Euskal AEn sare zabalaren batez besteko neurriaren benetako balioa emandako zenbatespenaren \pm % 1,53ko tartean dabil. Hau da:

$$[22,47 \pm (0,0153 \times 22,47)] = [22,13, 22,81]$$

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

¹ Laginaren neurri behar bezain 'handia' ($n > 30$) hartzen da. Hau horrela ez denean, fidagarritasun-tartea Student-en t banaketaren %95eko pertzentilarekin kalkulatu da, $n-1$ libere-gradurekin.

Bibliografia

[1] EUSTAT. "Kapital sozialari buruzko inkesta Fitxa metodologikoa."

<http://www.eustat.es/document/en%5Fcapital%5Fsocial%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 *Guia 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.