

Laginketa-erroreen kalkuluari buruzko txostena

Informazioaren Gizarteari buruzko Inkesta
(IGI - Administrazio publikoak)



AURKIBIDEA

1. Sarrera.....	3
2. Taylorren hedapen-metodoa	3
3. IGI erroreen kalkulua – Administrazio publikoa.....	4
3.1 Laginaren diseinua	4
3.2 Kalkulurako prozedura	5
3.3 IGIE inkestako erroreak kalkulatzeko estatistikoak eta eremuak.....	5
3.4 Emaitzak eta interpretazioa	7
Bibliografia	9

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 heuristikoren bidez laginketa-errorearen zenbatespen ona egin daiteke, laginaren diseinuaren azken pisuak eta ezaugarriak aintzat hartuta [3], [5].

Hemendik aurrera, aipatutako metodoak sartuko ditugu eta, bereziki, 2005. urtetik aurrerako Enpresetan Informazioaren Gizarteari buruzko Inkestetan aplikatuko.

2. Taylorren hedapen-metodoa [3], [5].

Metodo honen bidez laginketa-errorearen zenbatespenak kalkula daitezke guztizkoetarako, batezbestekoetarako eta ehunekoetarako estratifikazioa, klusterra eta probabilitate ezberdina duten laginketetan; hau da, EUSTATEk egindako estatistika-eragiketa askotan. Metodoak zenbateslearen hurbilketa linealak ateratzen ditu, eta bariantza kalkulatzeko duen haren laginketa-errorearen zenbatesle gisa erabiliz.

Batezbesteko biztanleriarentzat zenbatetsi den bariantza kalkulatzeko, adierazpen hau erabiltzen da:

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

Non,

$$e_{hi} = \left(\sum_{j=1}^{m_{ki}} w_{hij} (y_{hij} - \widehat{Y}) \right) / w_{...}$$

$$\bar{e}_{h..} = \left(\sum_{i=1}^{n_k} e_{hi.} \right) / n_h$$

eta

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

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

$$n = \sum_{h=1}^H \sum_{i=1}^{n_k} m_{hi}$$

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

3. IGI erroreen kalkulua – Administrazio publikoak.

3.1 Lagin-diseinua [1]

Administrazio publikoen IGIa panel-laginketa bat da, eta sektore publikoko establezimenduetatik datozen establezimenduetan orekatuta dago jarduera-establezimenduen kopuruarekin (65 jarduera), baita jarduera-taldeko establezimenduen kopuruarekin ere. Orotara bost talde dira: 1, administrazio publikoa eta defentsa; 2, hezkuntza; 3, osasun-jarduerak; 4, gizarte-zerbitzuak; eta 5, bestelakoak), baita establezimenduen kopurua enplegu-estratuko ere (6 ohiko talde), honako ezaugarri hauekin:

Laginaren esparrua: Sektore publikoko establezimendu guztiak, ujd_cod 3. sektore instituzionalekoa dutenak.

Laginaren neurria. Geruza bakoitzean % 6ko laginketa-errorea, eta % 20 gehiago zabaltzen da, gainlaginketa ziurtatzeko.

Laginketa horretan administrazio publikoak IGIeren iazko panelaren zati izatera igaroko direnez, behin betiko neurria aldatu egingo da aurrekoarekin alderatuta.

Geruzatzea

Lurralde historikoa eta enplegu-geruza (1= 0-9 langile, 2= 10-49 langile, 3= 50-99 langile, eta 4=100 langile eta gehiago).

Deskribatu den diseinu hori ezin hobeto egokitzen da aurreko atalean azaldu den metodo heuristikoaren zehaztapenetara. SASaren prozedurak eskatzen dituen parametroak besterik ez dira adierazi behar, bariantza behar bezala zenbatetsi ahal izateko.

3.2 Kalkulurako prozedura.

Hona hemen erroreen kalkulurako aplikatu den SASaren prozeduraren oinarrizko sintaxia [4]:

```
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)*/
```

IGI – Enpresak inkesta berrirako soilik, honako hauek izango dira sintaxi honen parametro orokorrak:

STRATA = Lurralde historikoa x Jarduera (A31) x Enpleguaren estratua

WEIGHT = Establezimenduen urteko jasokaria.

VAR = Informazioaren teknologien, Interneten eta merkataritza elektronikoaren ekipamendu eta erabileraren aldagaiak.

DOMAIN = Gurutzaketak, aldagai hauen arabera: geografia, jarduera, enplegu-estratua eta enpresaren titulartasuna.

3.3 IGI-Enpresak inkestako erroreak kalkulatzeko estatistikoak eta eremuak.

Hurrengo gurutzaketa eta estatistikoetarako laginketa-erroreak zenbatetsiko dira:

IKTko ekipamenduak administrazio publikoaren establezimenduetan

- Euskal Autonomia Erkidegoaren administrazio publikoetako IKTko ekipamenduak, lurralde historikoaren, jarduera-adarraren eta enplegu-estratuaren arabera (%). Laginketa-erroreak.
- Sare eta truke elektronikoen ekipamenduak Euskal Autonomia Erkidegoaren administrazio publikoetako establezimenduetan, lurralde historikoaren, jarduera-adarraren eta enplegu-estratuaren arabera (%). Laginketa-erroreak.
- Euskal Autonomia Erkidegoaren administrazio publikoetako segurtasun informatikoa, enplegu-estratuaren arabera (%). Laginketa-erroreak.
- Euskal Autonomia Erkidegoaren administrazio publikoetako establezimenduen informazioa kudeatzeko sistemak, enplegu-estratuaren arabera (%). Laginketa-erroreak.
- Euskal Autonomia Erkidegoaren administrazio publikoetako establezimenduetako kode irekiko programa informatikoak, enplegu-estratuaren arabera (%). Laginketa-erroreak.

Administrazio Publikoaren establezimenduen Interneteko erabiltzaileak

- Euskal Autonomia Erkidegoaren administrazio publikoan Interneterako sarbidea duten establezimenduen konexio mota, lurralde historikoaren, jarduera-adarraren eta enplegu-estratuaren arabera (%). Laginketa-erroreak.
- Euskal Autonomia Erkidegoaren administrazio publikoan Interneterako sarbidea duten establezimenduen webguneetako hizkuntzak, lurralde historikoaren, jarduera-adarraren eta enplegu-estratuaren arabera (%). Laginketa-erroreak.
- Euskal Autonomia Erkidegoaren administrazio publikoan Interneterako sarbidea duten establezimenduen webguneetako prestazioak, enplegu-estratuaren arabera (%).

Merkataritza elektronikoa administrazio publikoaren establezimenduetan

- Euskal Autonomia Erkidegoaren administrazio publikoetako establezimenduetako merkataritza elektronikoa, lurralde historikoaren, jarduera-adarraren eta enplegu-estratuaren arabera (%). Laginketa-erroreak.

Administrazio publikoarekin izapideak egiten dituzten establezimenduak eta enpresak

- Euskal Autonomia Erkidegoan Interneterako sarbidea duten administrazio publikoaren establezimenduekin izandako izapide elektronikoak, lurralde historikoaren, jarduera-adarraren eta enplegu-estratuaren arabera (%). Laginketa-erroreak.
- Euskal Autonomia Erkidegoan Interneterako sarbidea duten administrazio publikoaren establezimenduekin izandako izapide elektronikoak, egindako izapidearen, lurralde historikoaren, jarduera-adarraren eta enplegu-estratuaren arabera (%). Laginketa-erroreak.

INFORMAZIOAREN GIZARTEAREN INKESTA (IGI – Administrazio publikoak)

- Euskal Autonomi Erkidegoan Internetarako sarbidea duten administrazio publikoaren establezimenduekin izandako izapide elektronikoak, izapideak egiten diren administrazioaren, enplegu-estratuaren eta izapide motaren arabera (%). Laginketa-erroreak.

3.4 Emaitzak eta interpretazioa.

Laginketa-errorea zenbatestiaz gainera (2), SASek baliagarriak diren eta errorea interpretatzen laguntzen duten beste neurri batzuk ere ematen ditu. 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, laginketa-errorearen neurri gisa oso erabilia; hauxe du adierazpidea:

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

- **Konfiantza-tartea** % 95era. Konfiantza-tarte hori estatistikoaren laginketaren banaketan oinarritzen da (proportzioa, batezbestekoa, tasa,...). Limitearen Teorema Nagusiaren bidez, gehien-gehienetan lege Normal¹ bat onar dezakegu estatistikorik ohikoenetarako; beraz, tarte honen eraikuntza honako adierazpide honen ondorio izango da:

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

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:

Hamar langileko eta hortik gorako establezimenduak, eta Euskal Autonomia Erkidegoko enplegua, informazioaren teknologikoako ekipamenduen eta lurralde historikoaren arabera. Laginketa-erroreak. 2012

	Ordenador personal		Correo electrónico		Internet		Tfno. móvil	Página web
	% s/establ.	% s/empleo	% s/establ.	% s/empleo	% s/establ.	% s/empleo	% s/establ.	% s/establ.
Total	98.6	65.8	97.2	58.0	96.7	55.3	94.5	81.6
L. Inferior 95%	98.0	64.1	96.3	56.2	95.8	53.3	93.3	79.4
L. Superior 95%	99.3	67.5	98.1	59.8	97.7	57.2	95.6	83.7
CV(%)	0.3	1.3	0.4	1.6	0.5	1.8	0.6	1.3
Territorio Histórico								
Araba/Alava	97.6	66.4	95.6	55.6	95.8	51.5	93.9	81.0
L. Inferior 95%	95.2	62.5	92.9	51.9	93.2	47.0	91.6	76.7
L. Superior 95%	99.9	70.4	98.2	59.2	98.4	56.0	96.1	85.4
CV(%)	1.2	3.0	1.4	3.3	1.4	4.5	1.2	2.7
Bizkaia	98.9	67.3	97.3	60.0	96.6	57.7	94.8	81.5
L. Inferior 95%	98.1	64.8	96.0	57.3	95.1	54.9	93.1	78.1
L. Superior 95%	99.7	69.7	98.6	62.7	98.1	60.5	96.5	84.6
CV(%)	0.4	1.9	0.7	2.3	0.8	2.5	0.9	2.1
Gipuzkoa	98.7	63.2	97.8	56.1	97.4	53.4	94.2	82.2
L. Inferior 95%	97.8	60.1	96.7	53.0	96.2	50.3	92.4	78.6
L. Superior 95%	99.6	66.3	98.9	59.2	98.6	56.6	96.0	85.7
CV(%)	0.5	2.5	0.6	2.8	0.6	3.0	1.0	2.2

¹ Lagin-tamaina behar bezain 'handia' izatea asumitzen da (n >30). Hala egin ezin dezakegunean, konfiantza-tartea Student-en t banaketaren % 95eko pertzentilarekin kalkulaturiko da, n-1 libere-tate-gradurekin.

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

Aurreko taularako, Euskal Autonomia Erkidegoan ordenagailua duten establezimenduen ehunekoaren % 95erako errore erlatiboa % 0,588 da ($1,96 \cdot 0,3$). Edo, bestela esanda, % 95eko konfiantza-mailan Euskal Autonomia Erkidegoan ordenagailua duten establezimenduen ehunekoaren benetako balioa emandako zenbatespenaren $\pm\% 0,588$ tartean dabil. Hau da:

$$(98,6 \pm 0,00588 \cdot 98,6) = \% 98,0 \text{ eta } \% 99,3 \text{ artean}$$

Garrantzitsua da % 95erako errore erlatiboaren ehuneko jakin bat gainditzen duten zenbatespenak zein diren aipatzea, erabiltzaileak beharrezko zuhurtziaz joko dezan emandako informazio interpretatzeko orduan. % 95erako 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] EUSTAT (2006), *Informazioaren Gizarteari buruzko Inkesta. IGI – Enpresak. Fitxa metodologikoa*.
http://www.eustat.es/document/esia_c.html
- [2] Ekonomia Jardueren Sailkapen Nazionala (EJSN 1993) – 1. berr.
<http://www.eustat.es/document/datos/CNAE93REV1.xls>
- [3] Fuller, W. A. (1975), "Regression Analysis for Sample Survey," *Sankhyā*, 37, Series C, Pt. 3, 117 - 132.
- [4] Sas Institute Inc. (2004), "SAS/STAT[®] 9.1 Guía de Usuario". Copyright © 2004, Cary, NC, USA. ISBN 1-59047-243-8
- [5] Woodruff, R. S. (1971), "A Simple Method for Approximating the Variance of a Complicated Estimate" *Journal of the American Statistical Association*, 66, 411 -414.