

Laginketa-erroreen kalkulari buruzko txostena

Informazioaren Gizarteari buruzko Inkesta
(IGI- Enpresak)



AURKIBIDEA

1. Sarrera.....	3
2. Taylor-en hedapen-metodoa.....	3
3. IGI-Enpresak inkestako erroreen kalkulua.....	4
3,1 Laginaren diseinua.....	4
3,2 Kalkulurako prozedura.....	5
3,3 IGIE inkestako erroreak kalkulatzeko estatistikoak eta eremuak.....	5
3,4 Emaidzak eta interpretazioa.....	7
Bibliografia.....	9

1. Sarrera

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

Errore hori faktore askoren mende dago, besteak beste, biztanleria atal hori hartzeko prozeduraren (laginaren diseinua), hartzen diren unitate kopuruaren (laginaren tamaina), zenbatetsi nahi den ezaugarria nolakoa den, eta abarren mende. Hona hemen laginketa-errorearen adierazpide bat, oso zabaldua dagoena:

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

KASU HORRETAN, $\hat{\theta}$ INTERESAREN GAIKOA DA (BATEZ BESTEKOA, GUZTIRAKOA, PROPORZIOA...). Estatistiko horrek balio desberdinak hartuko ditu, ateratako laginaren arabera. Estatistikoak laginketan duen aldakortasunaren mende egongo da laginketaren errorea.

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

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

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

2. Taylor-en hedapen-metodoa [3], [5].

Metodo honen bidez laginketa-errorearen zenbatespenak kalkula daitezke guztirakoetarako, batez bestekoetarako eta ehunekoetarako estratifikazioa, klusterra eta probabilitate ezberdina duten laginketetan; hau da, EUSTATEk egindako estatistika-eragiketa askotan. Metodoak zenbateslearen hurbilketa linealak ateratzen ditu, eta bere bariantza kalkulatu du horren laginketa-errorearen zenbatesle bezala erabiliz.

Populazioaren batez besteko bezala zenbatetsi den bariantza kalkulatzeko adierazpena ondokoa da:

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

Kasu horretan:

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

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

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

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

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 behaketa Y aldagaiaren gainean ikusi diren balioak dira (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-Enpresak inkestako erroreen kalkulua.

3.1 Lagin-diseinua [1]

IGI Enpresa inkestan, probabilitatezko lagin jarraitua erabiltzen da. Hau da, Euskal Autonomia Erkidegoko establezimendu ekonomikoen panela da, jarduera-sektore guztietakoak, lehen sektorea eta etxeko lanak izan ezik. Panel honek zentsu-izaera du 100 langile edo gehiagoko establezimenduetan, eta lagin-izaera du 100 langile baino gutxiagokoetan.

Estratifikazioa lurralde historikoaren arabera, establezimenduaren tamainaren arabera (5 mota daude) eta jarduera-adarraren arabera egiten da, A31 sektorizazio normalizatuaren arabera, Jarduera Ekonomikoen Sailkapen Nazionalako jarduera-taldeei dagokiona [2]. Esleipena estratu bakoitzaren tamainaren erro karratuarekin proportzionala da, banaketa honetan behar baino ordezkari txikiagoa duten jarduera-sektoreetan izan ezik; horietan erabiltzen den esleipena tamainarekin proportzionala da. Ia 7.200 establezimendutako lagina osatu da.

Deskribatu den diseinu hori ezin hobeto egokitzen da aurreko idatz-zatian azaldu den metodo heuristikoaren zehaztapenetara. Soilik adierazi beharko dira, beraz, SASaren prozedurak eskatzen dituen parametroak bariantza behar bezala zenbatetsi ahal izateko.

3,2 Kalkulurako prozedura.

Hona hemen errorearen 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 ; /*laginketaren 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 kalkulatuak pisu-aldagaia (aukerazkoa)*/
```

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

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:

IKT ekipamenduak enpresetan eta establezimenduetan

- Euskal AEko establezimenduetako IKT ekipamenduak, lurralde eta jarduera-sektorearen arabera (%). Laginketa hutsegiteak. 2013
- IKT ekipamenduak 10 langile eta gehiagoko establezimenduetan, lurralde eta jarduera-sektorearen arabera (%). Laginketa hutsegiteak. 2013

- Euskal AEko establezimenduetako sareen eta truke elektronikoen ekipamenduak, lurralde historikoaren, jarduera sektorearen eta enplegu-geruzaren arabera (%). Laginketa hutsegiteak. 2013
- Euskal AEko 10 langile eta gehiagoko establezimenduetako sareen eta truke elektronikoen ekipamenduak, lurralde historikoaren eta jarduera sektorearen arabera (%). Laginketa hutsegiteak. 2013
- Euskal AEko establezimenduetako segurtasun informatikoa, enplegu-geruzaren arabera (%) eta jarduera sektorearen arabera (%). Laginketa hutsegiteak. 2013
- Euskal AEko establezimenduetako informazioa kudeatzeko sistemak, enplegu-geruzaren arabera (%) eta jarduera sektorearen arabera (%). Laginketa hutsegiteak. 2013

Internet erabiltzen duten enpresa eta establezimenduak

- Euskal AEko Interneteko sarbidedun establezimenduetako konexio-mota, lurralde historikoaren, jarduera sektorearen (A38) eta enplegu-geruzaren arabera (%). Laginketa hutsegiteak. 2013
- Euskal AEko Interneteko sarbidedun 10 langile eta gehiagoko establezimenduetako konexio-mota, lurralde historikoaren eta jarduera sektorearen arabera (%). Laginketa hutsegiteak. 2013
- Administrazio Publikoarekin eginiko izapide elektronikoak, Euskal AEko Internet duten establezimenduetan, enplegu-geruza eta jarduera sektorearen arabera (%). Laginketa hutsegiteak. 2013
- Euskal AEko web orridun establezimenduetako webaren prestazioak, enplegu-geruza eta jarduera sektorearen arabera (%). Laginketa hutsegiteak. 2013
- Euskal AEko web orridun establezimenduetako webaren hizkuntza erabilgarriak, lurralde historikoaren, jarduera sektorearen eta enplegu-geruzaren arabera (%). Laginketa hutsegiteak. 2013
- Euskal AEko web orridun 10 langile eta gehiagoko establezimenduetako webaren hizkuntza erabilgarriak, lurralde historikoaren eta jarduera sektorearen arabera (%). Laginketa hutsegiteak. 2013

Merkataritza elektronikoa

- Euskal AEko establezimenduetako merkataritza elektronikoa, lurralde historikoaren eta enplegu-geruzaren arabera. Laginketa hutsegiteak. 2012.
- Euskal AEko merkataritza elektronikoko bidezko erosketak eta salmentak, lurralde historikoaren, jarduera sektorearen eta enplegu-geruzaren arabera (milioi euro). Laginketa hutsegiteak. 2012

Lauki honetan, inkesta honetan laginketa-akatsak kalkulatzeko estatistikoaren eta gurutzaketen laburpena ageri da:

Estatistikoa	Inkestak berez dituen aldagaiak			Gurutzaketa aldagaiak			
	IKT ekipamenduak	Interneteko erlazioa	Merkataritza elektronikoa	Lurralde historikoa	Jarduera adarra	Langile kopurua	Enpresaren titulartasuna
Establezimenduen portzentajea	X	X	X	X	X	X	
10 langile edo gehiagoko establezimenduen portzentajea	X	X	X	X	X		X
Enpleguaren portzentajea	X			X	X	X	
10 langile edo gehiagoko establezimenduen enpleguaren portzentajea	X			X	X		X
Establezimenduetan dagoen ekipo kopurua	X			X	X	X	
10 langile edo gehiagoko establezimenduetan dagoen ekipo kopurua	X			X	X		X

3,4 Emaitzak eta interpretazioa.

Laginketako errorea zenbatesteaz gainera (2), SASak 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 hau estatistikoaren laginketaren banaketan oinarritzen da (proportzioa, batez bestekoa, tasa,...). Limitearen Teorema Nagusiaren bidez, gehien-gehienetan lege Normal 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 batez bestekoa 0 eta desbiderapen tipikoa 1 dituen banaketa Normal baten pertzentila da. Horrela baieztatu daiteke θ estatistikorako kalkulaturako tarteak biztanleriaren parametroaren benetako balorea daukala, kasuen %95ean (balizko laginak).

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

IKT1. Aldakuntza koefizienteak (%) eta konfiantza-tarteak %95ean, establezimenduen eta enpleguaren portzentajerako, informazioaren teknologien ekipamenduen arabera eta lurralde historikoaren arabera.

	Ordenagailu pertsonala		Posta elektronikoa		Internet		Sakelako tfnoa
	establ. %	enpleg. %	establ. %	enpleg. %	establ. %	enpleg. %	establ. %
Guztira	68.6	58.9	52.0	47.3	56.4	47.2	72.8
%95etik beherako m.	67.1	57.6	50.5	46.0	54.9	45.8	71.5
%95etik gorako m.	70.1	60.2	53.6	48.6	58.0	48.6	74.1
AK (%)	1.1	1.1	1.5	1.4	1.4	1.5	0.9
Lurralde historikoa							
Araba	71.8	56.9	54.4	43.9	58.2	43.1	74.7
%95etik beherako m.	68.8	53.5	51.1	40.5	54.9	39.6	72.2
%95etik gorako m.	74.9	60.3	57.7	47.2	61.4	46.6	77.2
AK (%)	2.2	3.0	3.1	3.9	2.9	4.2	1.7
Bizkaia	67.3	59.7	51.7	48.6	56.2	48.3	71.7
%95etik beherako m.	65.0	57.8	49.4	46.6	53.9	46.2	69.7
%95etik gorako m.	69.6	61.6	54.1	50.6	58.6	50.3	73.7
AK (%)	1.7	1.6	2.3	2.1	2.2	2.1	1.4
Gipuzkoa	69.3	58.8	51.6	47.0	56.1	47.6	73.7
%95etik beherako m.	66.8	56.7	49.1	45.1	53.6	45.4	71.6
%95etik gorako m.	71.8	60.8	54.1	49.0	58.7	49.8	75.9
AK (%)	1.8	1.8	2.4	2.1	2.3	2.3	1.5

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

Aurreko taularako, Euskal Autonomia Erkidegoan ordenagailua duten establezimenduen ehunekoaren %95erako errore erlatiboa %2,156 da ($1,96 \cdot 1,1$). Edo, bestela esanda, %95eko konfiantza mailan Euskal Autonomia Erkidegoan ordenagailua duten establezimenduen ehunekoaren benetako balioa emandako zenbatespenaren $\pm 2,156$ ko tartean dabil. Hau da,

$$(68,6 \pm 0,02156 \cdot 68,6) = \%67,1 \text{ eta } \%70,1 \text{ artean}$$

Garrantzitsua da %95erako errore erlatiboaren ehuneko jakin bat gainditzen duten zenbatespenak zein diren aipatzea. Zentzuzko muga %95erako errore erlatiboaren %20 gainditzen duten zenbatespenetan legoke (G.K. > %10 gutxi gorabehera), errore hori %30etik gorakoa zen laukitxoak bereziki azpimarratuz (G.K. > %15 gutxi gorabehera).

Bibliografia

- [1] EUSTAT (2006), "Informazioaren Gizarteari buruzko Inkesta. IGI – Enpresak. Metodologi fitxa".
http://www.eustat.es/document/esie_c.html
- [2] Jarduera Ekonomikoen Sailkapen Nazionala (EJSN 1993) – Berr. 1
<http://www.eustat.es/document/datos/CNAE93REV1.xls>
- [3] Fuller, W. A. (1975), "Regression Analysis for Sample Survey," Sankhy, 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.