Z. 9801

LAGINEN EGOKITZAPENA INFORMAZIO OSAGARRIA ERABILIZ

E. Bueno, A. Zarraga eta A. Iztueta

EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA

Duque de Wellington, 2
01010 VITORIA-GASTEIZ
Tel.: 945 18 75 00
Fax: 945 18 75 01
E-mail: eustat@eustat.es
www.eustat.es

LAGINEN EGOKITZAPENA INFORMAZIO OSAGARRIA ERABILIZ

Elena BUENO

Lankidetzarako bekaduna EUSTATen

Ainhoa ZARRAGA

Lankidetzarako bekaduna EUSTATen

Anjeles IZTUETA

Metodologia arloko arduraduna EUSTATen

LABURPENA

Laginen egokitzapen- eta orokortze-fasearen helburua laginetako ondorioak populaziora zabaltzea da. Horretarako eskura dagoen informazio guztia erabiliko da, eta inkesta edo iturri ezberdinen emaitzak bateratu egingo dira.

Koaderno honetan laginen egokitzapen estatistikoan erabilitako metodologia deskribatzen da, informazio osagarriaren zenbatekoaren arabera bi metodologia-mota ezberdintzen direlarik:

- Informazio osagarri handieneko metodoak, bariante anitzeko antolamendu osagarria dagoenean erabiltzen direnak.

- Iterazio-prozedurak, erabil daitekeen informazio osagarria bariante bakarrekoa denean.

Sarreran hainbat egokitzapen-mota aurkezten da, dagoen informazio osagarriaren arabera bereizita. Ondorengo ataletan erabiliko den oinarrizko notazioa ere ematen da.

Bigarren atala informazio osagarri handiena dagoenean erabiltzen diren metodoei lotzen zaie: lehenbizi, metodo horiek erabiltzen diren egoerei dagozkien notazio zehatz eta xeheak aurkezten dira, zenbatesle horien adierazpen matematikoarekin batera. Ondoren, aipatutako zenbatesleen berezitasun estatistikoen analisia egiten da. Erabilitako metodologiaren funtzionamendua argitzeko, eta metodologia hori ondo ulertarazteko, adibide errazak ematen dira. Egokitzapen-metodoak ezartzeko erabilitako prozedura informatikoak labur aipatu ondoren, aplikazio-egoera egokienak adierazten dira.

Hirugarren atalean bi iterazio-prozedura aurkezten dira egokitzapen-metodo gisa, hain zuzen bariante bakarreko antolamendu osagarria dagoenean erabiltzen direnak: *Raking* deritzana eta RAS, haren aldaera bat, SPADen Redre prozeduran erabiltzen dena. Lehenbizi erabili behar den notazioa aurkezten da, eta gero iterazio-prozedura biak deskribatzen dira, kasu errazean emango liratekeen pausoak adieraziz. Ondoren, bi

metodoak adibide batzuei aplikatzen zaizkie, hauek ere simulatuak direlarik. Azkenik, bi metodologiak aplikatzeko erabili diren prozedura informatikoak izendatzen dira, ezarpen-baldintza aproposenak adieraziz.

Laugarren atalean ondorio txiki bat ematen da, eta koadernoan aztertutako bi egokitzapen-metodoak alderatzen dira.

HITZ GARRANTZITSUENAK: Lagin-egokitzapena, Bariante Anitzeko Antolamendu Osagarria, Bariante Bakarreko Antolamendu Osagarria, iterazio-prozedurak, *Raking*, pisuak, orokortzaileak, geruzapena, post-geruzapena, aldagai osagarri koalitatibo eta koantitatiboak.

Eskerrak: Jaime Garrido, Juan José Ortiz eta Yolanda Pérez-i.

Aurkibidea

1.
kapitulua

Sarrera

Sarrera eta helburuak

Koaderno honetan estatistika munduan oso arrunta den gai bati buruz jardungo dugu:

N tamainako populazio edo unibertsoa daukagu, eta horri buruz zenbatespen bat egin nahi dugu, hain zuzen ere Y ezaugarri jakin bat zein neurritan dagoen. Oro har, populazio osoari inkesta egitea ezinezkoa denez, n tamainako lagin bat hartzen da, eta Y ezaugarriak lagineko indibiduoengan duen jokabidea aztertzen da. Argi dago, ordea, lagineko elementuengan dagoen Y ezaugarriaren zenbatekoa ez dela populazio osoan dagoenaren berdina, eta, guztizkoaren zenbatespena lortu ahal izateko, orokortzaile batzuk erabiliko dira, laginaren emaitzak kontuan hartuta. Orokortze-lan horrek populazio osoan Y ezaugarria zenbatekoa denari buruz lehen zenbatespena emango digu.

Orokorpena beste zerbaiterako ere erabiltzen da, laginean erantzuten ez dutenek eragiten duten informazio-galera konpentsatzeko alegia. Azken konpentsazio-mota hau erroldetan nahiz lagin-inkestetan erabil daiteke, erantzuten dutenek profil espezifiko edo bereizturik ez badute behintzat.

Oro har, *helburu den* Y *aldagaiaz* gain, badugu populazioan eta laginean aurkitutako beste ezaugarrien berri, *aldagai osagarrien* berri, hain zuzen. Hori dela eta, askotan *helburu den aldagai* edo *menpeko aldagai* deituko diogu Y aldagaiari. Suposizio honi esker, lagineko indibiduoen hasierako orokorpen eta pisuak aldatu egingo ditugu, horrela aldagai osagarriak populazio-banaketa antolamendura egokitzeko.

Laginean erabilitako egokitzapen- eta orokortze-metodoen artean, batzuek informazio osagarri handiena erabiltzen dute, eta beste batzuek, berriz, informazio osagarri marjinala. Bi egoera horietan gaur egun EUSTATek erabiltzen dituen metodoak aztertuko ditugu: teorian ezagutzen diren metodoetatik erakunde honek erabiltzen dituenak aukeratu eta azalduko ditugu. Egokitzapen-prozedurak notazio egokiaren bidez deskribatuko ditugu, eta ondoren adibide errazari aplikatuko dizkiogu. Aplikatzeko erabil daitezkeen software-mota ezberdinak ere aurkeztuko ditugu, eta horiekin batera bakoitzaren abantaila eta eragozpenak ere bai.

Egokitzapen-metodoen sailkapena

Egokitzapen-prozedura ezberdinak daude, aldagai osagarrien eta eskura dagoen informazio osagarria zenbatekoa denaren arabera. (Ikus [1] eta [2])

Aldagai Osagarrien motak

Bi aldagai-multzo daude: (Ikus [10])

- *Aldagai Osagarri Koalitatiboak*: modalitate-kopuru finituak dituzten aldagaiak dira. Sexua eta lan-merkatuarekiko harremana (lana izatea, langabezian egotea, lanik ez izatea, etab.) multzo honetakoak dira.

- *Aldagai Osagarri Koantitatiboak:* aldagai jarraituak dira, tarte bateko balioak hartzen dituztenak, eta ondorioz, infinitu balio izan ditzaketenak. Aldagai-mota honen adibideak fakturazioaren guztizkoa, etab. dira.

Eskura dagoen informazioa

Ondorengo sailkapena eskura dagoen informazio osagarriaren arabera egin da. Inkesta guztietan populazioa gelaska edo geruzetan banatzen da, horretarako aldagai koalitatiboen bariante anitzeko gurutzaketa erabiliz. Bi kasu egon daitezke:

- Aldagai osagarriak *koalitatiboak* direnean, geruza bakoitzak dituen elementuak adierazten dira. *Informazio Osagarri Handiena* baldin badago, geruzapenaren gelaska bakoitzeko populazio-aleak ezagutzen dira. *Handiena ez den Informazio Osagarria* baldin badago, aldagai osagarrien bariante bakarreko antolamendu marjinala baino ez da ezagutzen. (Ikus [6] y [7]).

- Aldagai osagarriak *koantitatiboak* direnean, dagoen informazioa aldagai osagarrien antolamenduari buruzkoa da, lortutako geruzetan nola banatzen diren, alegia. *Informazio Osagarri Handiena* dagoenean, aldagai osagarrien populazioaren guztizkoa, batezbestekoa… izaten da geruza bakoitzeko, eta Handiena ez den Informazio Osagarria denean, edukitzen den informazioa aldagai osagarriaren populazaioaren guztizkoa, batezbestekoa… izaten da bariante bakarreko modalitate bakoitzean.

Zenbatespen-metodoak

Azterketako helburu diren estatistikoei buruzko hasierako zenbatesleetatik abiatzen da, informazio osagarria kontuan hartu gabe: (Ikus [15])

- *Horvitz-Thompson-en metodoa*

  Azterketako helburu diren estatistikoen zenbatesleak, indibiduo bakoitzak laginean sartzeko dituen probabilitateen alderantzizkoak pisu gisa hartuz lortzen dira. Indibiduoei emandako pisuak dira, beraz.

- *Horvitz-Thompson-en metodo zuzendua*

  Erantzunik eza zuzentzeko erabiltzen da. Hautespen-probabilitatea eta erantzunik ezaren tasaren arteko biderkaduraren alderantzia erabiltzen da pisu gisa. Erantzunik ezaren tasa geruza homogeneotan hartzen da. Kasu honetan haztaketa gelaskaka egiten da, eta gelaska bakoitzeko elementu guztiek hautapen-probilitate berdinak izan behar dituzte.

Informazio osagarria erabiltzen duten zenbatespen-metodoak *erregresio anizkoitzeko* eredu orokorrean sailka daitezke.

ERREGRESIO ANIZKOITZEKO EREDUA: helburu den Y aldagaiaren eta aldagai osagarrien arteko harremana dagoela onartzen du. Normalean onartzen den erregresio--eredua erregresio lineala izaten da. Erregresio-zenbatesle hau zenbatespen-prozedura orokorrenetakoa da, eta bertan aldagai osagarri bat baino gehiago erabil daitezke, aldagai horiek jarraituak nahiz nominalak izan daitezkeelarik.

Ondoren erregresio-ereduaren kasu zehatz batzuk izendatuko ditugu: (Ikus [2], [7], [8] eta [9])

- *Geruzapen eta post-geruzapen* zenbateslea, edo ARRAZOIAREN ohizko zenbateslea: aldagai osagarri dikotomikoak hartuta lortzen da. Aldagai horiek geruzapenean eta post-geruzapenaren lortutako geruza bakoitzari lotuta daude. Zenbatesle-mota hau Bariantzaren Analisiko eredura moldatutako erregresio--zenbateslearen parekoa da, beraz. Ondorioz sortzen diren orokortzaileak ere gelaskaka aztertzen dira, eta geruzapeneko gelaska bakoitzeko populazio eta laginaren zatidura dira, aleetan zenbatuak.

- *Arrazoiaren* zenbateslea: parametro independente nulua duen erregresio linealeko eredua hartu behar da, hau da, jatorritik pasatzen den erregresio-zuzena. Gelaskako orokortzailea lortzen da, eta gelaska horretako aldagai koantitatiboaren populazio-guztizkoa eta laginaren guztizkoaren arrazoia da.

- *Raking* zenbatesleak: iterazioz askatzen den metodoa da. Bi motako Raiking-ak daude:

    1. Informazio osagarria koalitatiboa da eta eskura ditugun datuak ale marjinalak dira. Lortzen diren gelaska-orokortzaileak populazio-ale marjinalen eta iterazioz lortu diren hurbileko marjinalen arteko zatidurak dira. Hau da Raking arruntena. Erregresio-eredu gisa, Arrazoiaren zenbatesle arruntaren kasu berezia da, eta bertan bariante bakarreko aldagai osagarriak hartzen dira.

    2. Informazio osagarria koantitatiboa da. Kasu honetan, eskura dauden datuak geruzapeneko aldagai koantitatiboen guztizko marjinalak dira. Erregresio--eredu gisa, Arrazoiaren zenbateslearen kasu berezia da.

- *MISTOAK*: badaude aurreko metodoen zenbait aldaera, aurrekoen konbinazio diren metodoak. Horietakoa da *Raking Ratio-ren zenbatesle aldatua* izenekoa.

Oinarrizko notazioa

Jo dezagun U populazioa eta populazio horren S lagina daukagula:

U={1,2,...,N}, k=1,...,N delarik

S={1,2,..,n}, k=1,...,n delarik

Helburu den Y aldagaia Nx1 bektore gisa definituko dugu:

$$\mathbf{Y}= \begin{pmatrix} Y(1) \\ \dots \\ Y(k) \\ \dots \\ Y(N) \end{pmatrix}$$

Zenbatetsi beharreko populazioaren guztizkoa honakoa da:

$$\mathrm{Y}=\sum_{k=1}^{N} Y(k)=\sum_{k=1}^{N} y_k \qquad (1)$$

Indibiduo bakoitzari hasierako pisua ematen zaio:

$z_k = {}^{1}\!/_{p_k}$ k=1,..,n delarik

p_k k=1,..,n delarik, probabilitate-laginketan k lagin-indibiduoak S-n duen hautespen-
-probabilitatea da.

Azkeneko pisua honela izendatuko dugu:

w_k k=1,...,n delarik.

HORBITZ-THOMSONen zenbateslea, populazioaren guztizkoan ageri den eta helburu den Y aldagaia kalkulatzeko erabiltzen dena: (Ikus [4] eta [8])

$$\hat{Y}_P = \sum_{k=1}^{n} z_k \bullet y_k \qquad (2)$$

Alboragabea da Y aldagaiaren populazioaren guztizkoarekiko, adib.,

$$E\left(\hat{Y}_P\right) = Y$$

Bere *bariantza* honakoa da:

$$Var\left(\hat{Y}_P\right) = \sum_{k,l=1}^{N} \left(p_{kl} - p_k p_l\right) \bullet \frac{y_k}{p_k} \bullet \frac{y_l}{p_l},$$

p_{kl} k eta l elementuak batera hartuta duten hautespen-probabilitatea delarik.

Bariantza horren zenbatespena honakoa da:

$$\hat{V}ar\left(\hat{Y}_P\right) = \sum_{k,l=1}^{n} \left(p_{kl} - p_k p_l\right) \bullet \frac{y_k}{p_k} \bullet \frac{y_l}{p_l}$$

Lagineko indibiduo guztiei hautespen-probabilitate berdina ematen baldin bazaie, hautespen-probabilitateak honakoak dira:

$$p_k = \frac{n}{N} \quad \forall k = 1,..,n$$ eta $$p_{kl} = \frac{n}{N} \bullet \frac{n-1}{N-1} \quad \forall k,l = 1,..,n$$

Ausazko laginketa sinplean baldintza hori bete egiten da, eta aurreko adierazpenek honako forma hartzen dute:

$$\hat{Y}_P = \sum_{k=1}^{n} \frac{N}{n} \bullet y_k \tag{3}$$

$$Var\left(\hat{Y}_P\right) = N^2 (1-f) * \frac{S^2}{n} \tag{4}$$

$$non\, S^2 = \frac{\sum_{k=1}^{N} \left(y_k - \overline{Y}\right)}{N-1} \quad \text{eta} \quad \overline{Y} = \sum_{k=1}^{N} \frac{y_k}{N}$$

$$\hat{V}ar\left(\hat{Y}_P\right) = N^2 (1-f) * \frac{s^2}{n} \tag{5}$$

$$non\, s^2 = \frac{\sum_{k=1}^{n} \left(y_k - \overline{y}\right)}{n-1} \quad \text{eta} \quad \overline{y} = \sum_{k=1}^{n} \frac{y_k}{n}$$

Aldagai osagarriei dagokien notazioa hurrengo ataletan ikusiko dugu, informazio osagarri horren arabera zer motatako arazo aurki ditzakegun zehaztuz.

2.
kapitulua

Informazio Osagarri Handieneko metodoak

Hemen aztertuko ditugun metodoak informazio osagarri HANDIENA dagoenean, hau da, geruzapenaren ondorioz sortutako gelaska bakoitzean dauden aldagai osagarrien zenbatekoa ezagutzen dugunean baino ez daitezke erabili.

Notazioa

Zenbait murrizketa proposatzen dira, laginaren-antolamendu haztatua eta populazioaren baterako antolamendua berdinak izan daitezen. Ondoren murrizketa horiek adierazteko notazioa aurkeztuko dugu, informazio osagarriaren arabera zein kasu den zehaztuz.

Aldagai Osagarri Koalitatiboak

Murrizketen adierazpen matrizialak honakoak dira: (Ikus [16])

$$\hat{X} \bullet \mathbf{w} = \mathbf{X} \bullet \mathbf{I}, \qquad (6)$$

non pisuen bektorea **w'**=(w_1 ,.., w_k ,.., w_n) eta **I** bektorea 1s-ren Nx1 bektorea baitira:

$$\mathbf{I} = \begin{pmatrix} 1 \\ .. \\ 1 \\ .. \\ 1 \end{pmatrix}$$

$\hat{X}$ eta **X** matrizeak (m+1)xn lagin-matrizea eta (m+1)xN populazio-matrizea dira. Matrize horien osagaiak deskribatzeko, zenbait aldagai deskribatuko ditugu:

L aldagai osagarri koalitatibo ditugu, bakoitza $L_1, L_2,..., L_l,..., L_L$ modalitateekin. L aldagaien bariante anitzeko gurutzaketaren bidez populazioaren geruzapena egiten da, eta $L_1 \bullet L_2 \bullet .. \bullet L_l \bullet .. \bullet L_L = m$ geruzak lortzen dira.

Geruza bakoitzaren aldagai identifikadore edo dikotomikoak definitzen dira:

$$X_{h_1..h_l..h_L}(k) = \begin{cases} 1, baldin\ k \in geruza(h_1,..,h_l,..h_L) \\ 0, bestela \end{cases}$$

$$\Rightarrow L_1 \bullet .. \bullet L_l \bullet .. \bullet L_L = m$$

m aldagai dikotomikoen kopurua delarik.

Errenkadako populazio-bektorea daukagu

$X_{h_1..h_l..h_L} = \left(X_{h_1..h_l..h_L}(1),..,X_{h_1..h_l..h_L}(k),..,X_{h_1..h_l..h_L}(N)\right)$ (1xN) dimentsiokoa

Eta errenkadako lagin-bektorea

$\hat{X}_{h_1..h_l..h_L} = \left(X_{h_1..h_l..h_L}(1),..,X_{h_1..h_l..h_L}(k),..,X_{h_1..h_l..h_L}(n)\right)$ (1xn) dimentsiokoa

Honakoa ziurtatzeko:

$$w_1 + .. + w_k + ..w_n = N$$

X_0 aldagaia identikoki 1 hartu eta honela adierazten dugu:

$X_0 = \left(X_0(1),..,X_0(k),..,X_0(N)\right) = (1,..,1,..,1)$ (1xN) dimentsioko errenkadako populazio-bektorea

$\hat{X}_0 = \left(X_0(1),..,X_0(k),..,X_0(n)\right) = (1,..,1,..,1)$ (1xn) dimentsioko errenkadako lagin--bektorea

X eta $\hat{X}$ matrizeak errenkadako bektore hauek osatzen dituzte:

$$\mathbf{X} = \begin{pmatrix} X_0 \\ X_1 \\ \dots \\ X_h \\ \dots \\ X_m \end{pmatrix} \quad \text{eta} \quad \hat{X} = \begin{pmatrix} \hat{X}_0 \\ \hat{X}_1 \\ \dots \\ \hat{X}_h \\ \dots \\ \hat{X}_m \end{pmatrix}$$ [1]

Matrize-adierazpena garatuz, honakoa lortuko dugu:

$$\begin{cases} \sum_{k=1}^{n} w_k \bullet X_0(k) = \sum_{k=1}^{N} X_0(k) \Leftrightarrow w_1 + .. + w_k + ..w_n = N \\ \sum_{k=1}^{n} w_k \bullet X_{h_1..h_l...h_L}(k) = \sum_{k=1}^{N} X_{h_1..h_l..h_L}(k), \forall 1 \le h_1 \le L_{1,..,} 1 \le h_l \le L_{l,..,} 1 \le h_L \le L_L \end{cases}$$

Aldagaiak geruzen identifikadore izanik, populazioari buruzko datuek honako forma hartzen dute:

[1] Aldagaien adierazleak indize bakarraren bidez eman dira, besterik gabe aldagaiak ordenatuz.

$$\begin{cases} \sum_{k=1}^{N} X_{h_1..h_l..h_L}(k) = N_{h_1..h_l..h_L} \\ \sum_{k=1}^{n} X_{h_1..h_l...h_L}(k) = n_{h_1..h_l..h_L} \end{cases} \tag{7}$$

Sistema askatu egingo dugu, gelaska-egokitzapena dela eta, beraz, geruza bereko indibiduo guztiek pisu berdina dutela, kontuan hartuz. Sistemaren ezezagunak $W_{h_1..h_l..h_L}$, gelaskei emandako orokortzaileak, izango dira. Sistema honakoa da:

$$\begin{cases} W_{h_1..h_l..h_L} \bullet n_{h_1..h_l..h_L} = N_{h_1..h_l..h_L} \quad \forall\ 1 \le h_1 \le L_1,..,1 \le h_l \le L_l,..,1 \le h_L \le L_L \\ \sum_{h_1..h_l..h_L} W_{h_1..h_l..h_L} \bullet n_{h_1..h_l..h_L} = N \end{cases} \tag{8}$$

Sistemak $L_1 \bullet L_2 \bullet \ldots \bullet L_l \bullet \ldots \bullet L_L = m$ ezezagunak eta $L_1 \bullet L_2 \bullet \ldots \bullet L_l \bullet \ldots \bullet L_L + 1$ ekuazioak ditu. Bertan, azkena aurrekoen konbinazio lineala da, sistema mxm hutsa beraz, zeinaren matrize elkartua $A \in (mxm)$ matrize diagonala baita. Matrize horren elementu diagonalak geruzapen-gelaska bakoitzeko lagin-aleak dira.

Aurrekoaren ondorioz, sistemak honakoa dauka:

Ebazpen bakarra $\Leftrightarrow n_{h_1..h_l..h_L} \neq 0\ \forall\, h_1..h_l..h_L$

Egoera honetan, sistemaren ebazpena hau da:

$$\hat{W}_{h_1..h_l..h_L} = \frac{N_{h_1..h_l..h_L}}{\hat{n}_{h_1..h_l..h_L}}, \tag{9}$$

$$\forall 1 \le h_1 \le L_1,..,1 \le h_l \le L_l,..,1 \le h_L \le L_L$$

Orokortzaileak erabiliz lortzen den zenbatespenari Geruzapen-zenbatespena deitzen zaio, eta ausazko laginketa geruzatua duela suposatzen da. Zenbatespena populazioari buruz genituen datuak egokituz egiten da, GERUZAPEN-egokitzapena deritzanaren bidez, alegia. Era honetako egokitzapen-mota berezia populazioaren geruzapena langinketaren ondoren, eta ez aurretik, egitea litzateke. Azken egokitzapen horretan lortzen den zenbatespenari post-geruzapeneko zenbatespena deitzen zaio; egokitzapen-metodoari, berriz, post-geruzapen metodoa. Batean nahiz bestean, gelasketako orokortzaileak beren populazio-ale eta lagin-ale kopuruen arteko zatidura dira. (Ikus [4], [6] eta [14])

Aldagai Osagarri Koantitatiboak

Kasu honetan, X aldagai osagarri koantitatibo bakarra dagoela suposatuko dugu. Populazioaren geruzpena laginketa-aurreko aldagaien bidez egiten da: aurregeruzapena deitzen zaio.

Murrizketek matrize-forma berdina hartzen dute, ondoren zehazten diren matrize berriak dituztelarik:

$$\left(\hat{X}^{*}\right)\bullet\ \mathbf{w} = \left(X^{*}\right)\bullet \mathbf{I}, \qquad (10)$$

zeinetan, lehen bezala, pisuen bektorea **w'**=(w_1 ,.., w_k ,.., w_n) eta **I** bektorea 1s-ren Nx1 bektorea baitira.

$$\mathbf{I}=\begin{pmatrix}1\\..\\1\\..\\1\end{pmatrix}$$

X^{*} eta $\hat{X}^{*}$ matrizeak definitzeko, aldagai osagarriak definituko ditugu:

$$X^{*}_{h_1..h_l..h_L}(k) = X_{h_1..h_l..h_L}(k)\bullet X(k),\ \forall 1\le h_1\le L_1,..,1\le h_l\le L_l,..,1\le h_L\le L_L,$$

$X_{h_1..h_l..h_L}(k)$ ondoriozko geruza bakoitzaren aldagai dikotomiko identifikatzailea, eta

$X(k)$ populazioaren elementu bakoitzeko aldagai koantitatiboaren balioa direlarik.

Halaber, honakoa definituko dugu:

$$X^{*}_{0}(k) = X_0(k)\bullet X(k) = X(k)$$

Honako bektoreak lortuko ditugu:

$X^{*}_{h_1..h_l..h_L} = \left(X^{*}_{h_1..h_l..h_L}(1),..,X^{*}_{h_1..h_l..h_L}(k),..,X^{*}_{h_1..h_l..h_L}(N)\right)$ (1xN) dimentsioko errenkadako bektorea

$\hat{X}^{*}_{h_1..h_l..h_L} = \left(\hat{X}^{*}_{h_1..h_l..h_L}(1),..,\hat{X}^{*}_{h_1..h_l..h_L}(k),..,\hat{X}^{*}_{h_1..h_l..h_L}(n)\right)$ (1xn) dimentsioko errenkadako lagin-bektorea.

Bektore horiek X^{*} eta $\hat{X}^{*}$ matrizeen errenkadako osagaiak dira. Honako forma hartzen dute:

$$X^* = \begin{pmatrix} X_0^* \\ X_1^* \\ \dots \\ X_h^* \\ \dots \\ X_m^* \end{pmatrix} \quad \textbf{eta} \quad \hat{X}^* = \begin{pmatrix} \hat{X}_0^* \\ \hat{X}_1^* \\ \dots \\ \hat{X}_h^* \\ \dots \\ \hat{X}_m^* \end{pmatrix}$$

Murrizketen matrize-adierazpena garatu eta honako sistema lortuko dugu:

$$\begin{cases} \sum_{k=1}^{n} w_k \bullet X^*_{h1..hl..hL}(k) = \sum_{k=1}^{N} X^*_{h1..hl..hL}(k) \quad \forall 1 \le h_1 \le L_1,..,1 \le h_l \le L_l,..,1 \le h_L \le L_L \\ \sum_{k=1}^{n} w_k \bullet X_0^*(k) = \sum_{k=1}^{N} X_0^*(k) \Leftrightarrow w_1 + .. + w_k + .. + w_n = \sum_{k=1}^{N} X(k) \end{cases}$$

$L_1 \bullet L_2 \bullet \dots \bullet L_l \bullet \dots \bullet L_L = m$ ezezagun eta $L_1 \bullet L_2 \bullet \dots \bullet L_l \bullet \dots \bullet L_L + 1$ ekuazioko sistema da. Bertan, azkena aurrekoen konbinazio lineala da, sistema mxm beraz, zeinetan matrize elkartua $A \in (mxm)$ matrize diagonala baita, elementu diagonalak dituena:

$$\sum_{k=1}^{n} X^*_{h1..hl..hL}(k)$$

eta *ez-nuluak* izango dira *baldin eta*:

geruza bakoitzeko X aldagai osagarria guztira zero ez bada. Hori *aldagaia >0 izan dadila* eskatuz lor daiteke

geruza bakoitzeko lagin-*aleak ez-nuluak* badira

Lehenengo murrizketak ez du inolako eraginik, aztertzen diren aldagai koantitatiboak gehienetan aldagai positiboak izaten baitira, hala nola aldagai ekonomikoak: fakturazioaren guztizkoa, etab.

Sistemaren matrizearen singulartasun eza egiaztatu ondoren, lortzen den ebazpen bakarra honakoa da:

$$\hat{W}^*_{h1..hl..hL} = \frac{\sum_{k=1}^{N} \hat{X}^*_{h1..hl..hL}(k)}{\sum_{k=1}^{\hat{n}} \hat{X}^*_{h1..hl..hL}(k)} \qquad (11)$$

Berriz ere, $\hat{W}^*_{h1..hl..hL}$ geruzapen-gelaskei dagozkien orokortzaileak dira, egiten ari garen egokitzapena gelaskakoa dela kontuan izanik.

Ondoriozko metodoari *ARRAZOIAREN (RATIOAREN) metodoa* deitzen zaio. Metodo horretan gelasken orokortze-faktoretzat hartzen direnak aldagai koantitatiboaren populazioaren guztizkoa eta laginaren guztizkoaren arteko zatidurak dira.

Ikusi dugunez, informazio osagarri koalitatiboa nahiz koantitatiboa dagoenean, haztapen-metodoak geruza bakoitzerako orokortzaile bakarra ematen digu, *baldin eta* geruza bakoitzean *lagin-elementuren bat* badago.

Batezbestekoaren zenbatesleearen ezaugarriak

Azter ditzagun informazio osagarri handiena dagoenean kontsideratu ditugun bi zenbatesleak, hau da, *post-geruzapen* zenbateslea eta *arrazoiaren* zenbateslea. Azterketa horretarako helburu den aldagaiaren *populazioaren batezbestekoa* hartuko dugu estatistikotzat.

Aldagai Osagarri Koalitatiboak

Populazioaren batezbestekoa aurkitzeko ausazko laginketa geruzatuan lortzen den zenbateslea honakoa da:

$$\bar{Y}_{st} = \sum_{h=1}^{m} W_h \bullet \bar{y}_h \text{, non } W_h = \frac{N_h}{N}, ^2 \text{ den} \qquad (12)$$

$\bar{Y}_h$ h geruzako populazioaren batezbestekoa delarik. Zenbatespena analogikoki lortzen da, laginketa-aurreko nahiz laginketa-ondorengo geruzapenean (post-geruzapenean).

Jo dezagun hasieran laginketa-aurreko geruzapena daukagula, eta, ondorioz, n_h balioak finkoak direla.

Zenbatesle hau kontsideratzeko $W_h = \frac{N_h}{N}$ ezagutu behar dugu, edo, beste era batera esanda, $N_h \ \forall h = 1,..,m$.

Datu horiek ezagutzen ditugunean, gelaska bakoitzeko Y-ren populazioaren batezbestekoa zenbatetsiko dugu:

$\bar{y}_h = \frac{y_h}{n_h}$, non y_h gelaska bakoitzeko helburu den Y aldagaiaren laginaren guztizkoa den.

Adierazpen hori zenbateslean ordezkatu eta honako zenbatespena lortuko dugu:

[2] Gurutzaketan lortutako geruzen zenbakuntza egiten da, azpindize bakarrekoa jada.

$$\hat{\bar{Y}}_{st} = \sum_{h=1}^{m} \frac{N_h}{N} \bullet \frac{y_h}{n_h} \qquad (13)$$

Hori da Y-ren populazio-batezbestekoaren geruzapen-zenbatespena. Hortik aurrera erraza da populazioaren guztizkoaren zenbatespena egitea:

$$\hat{Y}_{st} = \hat{\bar{Y}}_{st} \bullet N$$

Aurreko adierazpenean ordezkapena berriz eginez, honakoa lortuko dugu:

$$\frac{\hat{Y}_{st}}{N} = \sum_{h=1}^{m} \frac{N_h}{N} \bullet \frac{y_h}{n_h} \Leftrightarrow$$

$$\hat{Y}_{st} = \sum_{h=1}^{m} \frac{N_h}{n_h} \bullet y_h \qquad (14)$$

Azter dezagun orain populazioaren batezbestekorako geruzapen-zenbatespena hartzen denean sortzen den errorea:

$$Var(\hat{\bar{Y}}_{st}) = Var(\sum_{h=1}^{m} W_h \bullet \bar{y}_h) = \sum_{h=1}^{m} (W_h)^2 \bullet Var\left(\bar{y}_h\right) + 2 \bullet \sum_{h=1}^{m} \sum_{j>h}^{m} Cov(\bar{y}_h, \bar{y}_j)$$

eta errorea honakoa da:

$$\sqrt{Var(\hat{\bar{Y}}_{st})}$$

Ausazko laginketa geruzatua izanik, kobariantzen terminoa desagertu egiten da, laginketa independentea baita geruza bakoitzean. Bariantzaren adierazpenak honako forma hartzen du:

$$Var(\hat{\bar{Y}}_{st}) = Var(\sum_{h=1}^{m} W_h \bullet \bar{y}_h) = \sum_{h=1}^{m} (W_h)^2 \bullet Var(\bar{y}_h)$$

Geruza bakoitzeko laginketa ausazkoa eta sinplea da, eta bariantza honakoa da:

$$Var\left(\bar{y}_h\right) = \frac{S_h^{\,2}}{n_h} \bullet \frac{N_h - n_h}{N_h}$$

Beraz, bariantzaren azken adierazpena hau da:

$$Var(\hat{\bar{Y}}_{st}) = \sum_{h=1}^{m} \frac{W_h^{\,2} \bullet S_h^{\,2}}{n_h} - \frac{\sum_{h=1}^{m} W_h \bullet S_h^{\,2}}{N}$$

S_h h geruzako Y aldagaiaren benetako bariantza delarik.

Geruza bakoitzeko bariantza honela zenbatetsiko dugu:

$$\hat{V}ar\left(\bar{y}_h\right) = \frac{s_h^{\,2}}{n_h} \bullet \frac{N_h - n_h}{N_h}$$

h geruzako Y-ren lagin-bariantza s_h delarik. Eta ondoren populazioaren batezbestekorako geruzapen-zenbatesleraren bariantzaren zenbatespena lortuko dugu.

$$\hat{Var}(\hat{\bar{Y}}_{st}) = \sum_{h=1}^{m} \frac{W_h^2 \bullet s_h^2}{n_h} - \frac{1}{N} \bullet \sum_{h=1}^{m} W_h \bullet s_h^2 \qquad (15)$$

Errorearen zenbatespenaren adierazpenak argi uzten du zenbatespenaren errorea helburu den Y aldagaiaren arabera gertatzen dela geruza bakoitzean. Horregatik, *geruza horiek helburu den aldagaiarekiko homogeneoak* baldin badira, *guztizko bariantza txikia izango da,* n_h guztiak *behar bezain handiak* badira behintzat. Horrela errorea neurri batean kontrolatu ahal izango da.

Gelasketako helburu den aldagaiaren bariantza kontrolatu ahal izateko, helburuari zuzenen lotzen zaizkion geruzapen-aldagaiak har daitezke. Horrela, aldagaien bariante anitzeko gurutzaketaren bidez lortutako geruzak homogeneoak izango dira helburu den aldagaiarekiko.

Post-geruzapen zenbateslea berdintsu aztertuko dugu. Kasu honetan n_h balioak ausazkoak dira, laginari lotuta daude, eta $\hat{n}_h$ bezala adieraziko ditugu. Lagina hartu ondoren, $\hat{n}_h$ balioak, denak zerotik gorakoak, finkoak dira.

Populazioaren batezbestekotik lortzen dugun zenbatesleak geruzapen-zenbateslearen antzeko forma hartzen du:

$$\bar{Y}_W = \sum_{h=1}^{m} W_h \bullet \bar{Y}_h$$

Lagin bakoitzean, geruza bakoitzeko $\bar{y}_h = \dfrac{y_h}{\hat{n}_h}$ lagin-batezbestekoa hartu eta post--geruzapen zenbateslea lortuko dugu:

$$\hat{\bar{Y}}_W = \sum_{h=1}^{m} W_h \bullet \bar{y}_h = \sum_{h=1}^{m} W_h \bullet \frac{y_h}{\hat{n}_h} = \sum_{h=1}^{m} \frac{N_h}{N} \bullet \frac{y_h}{\hat{n}_h} = \frac{1}{N} \bullet \sum_{h=1}^{m} \frac{N_h}{\hat{n}_h} \bullet y_h$$

Lehen adierazi dugunez, laginketa egin ondoren $\hat{n}_h$ horiek finkoak dira. Ondorioz, zenbatesle honen bariantza-zenbatespena geruzapen-zenbateslearena bezalakoa da, eta adierazpen hau daukagu:

$$Var(\hat{\bar{Y}}_W) = Var(\sum_{h=1}^{m} W_h \bullet \bar{y}_h) = \sum_{h=1}^{m} (W_h)^2 \bullet Var(\bar{y}_h) = \sum_{h=1}^{m} \frac{W_h^2 \bullet S_h^2}{\hat{n}_h} - \frac{1}{N} \bullet \sum_{h=1}^{m} W_h \bullet S_h^2$$

Dena den, bariantza ez da finkoa, $\hat{n}_h$ ak ere ez baitira finkoak. Beraz, bariantzaren batezbesteko balioa aurkitu beharko dugu. Batezbesteko balioa aurkitzeko nahikoa da itxaropena hartzea:

$$E\left[Var(\hat{\bar{Y}}_W)\right] = E\left[\sum_{h=1}^{m}\frac{W_h^2 \bullet S_h^2}{\hat{n}_h} - \frac{1}{N}\bullet\sum_{h=1}^{m} W_h \bullet S_h^2\right] =$$

$$= \sum_{h=1}^{m} W_h^2 \bullet S_h^2 \bullet E\left(\frac{1}{\hat{n}_h}\right) - \frac{1}{N}\bullet\sum_{h=1}^{m} W_h \bullet S_h^2$$

Baina $E\left(\frac{1}{\hat{n}_h}\right) = \frac{1}{n \bullet W_h} + \frac{1-W_h}{n^2 \bullet W_h^2}$ izanik, aurreko adierazpena honela geratuko da:

$$E\left[Var(\hat{\bar{Y}}_W)\right] = \frac{1-f}{n}\sum_{h=1}^{m} W_h \bullet S_h^2 + \frac{1}{n^2}\bullet\sum_{h=1}^{m}(1-W_h)\bullet S_h^2$$

Lehenbiziko gaia $Var\left(\bar{Y}_{st}\right)$ geruzapen-zenbateslearen bariantzari dagokio, eta bigarrena, post-geruzapeneko zenbateslea hartzean bariantzak jasaten duen gehikuntza da. Bigarren gaia garatuz honako adierazpena lortuko dugu:

$$\frac{1}{n^2}\bullet\sum_{h=1}^{m}(1-W_h)\bullet S_h^2 = \frac{1}{n \bullet \bar{n}_h}\bullet \bar{S}_h^2 - \frac{1}{n^2}\bullet\sum_{h=1}^{m} W_h \bullet S_h^2$$

non $\bar{n}_h = \frac{n}{m}$ laginketa - unitateen batezbesteko zenbakia geruzako eta $\bar{S}_h^2$-ren batezbestekoa diren. Azkenik, post-geruzapeneko zenbateslea hartuta gertatzen den bariantzaren gehikuntza txikia izango da $\bar{n}_h$ a handi samarra den bitartean.

Batezbesteko bariantzaren zenbatespena honako da:

$$\hat{E}\left[Var(\hat{\bar{Y}}_W)\right] = \frac{1-f}{n}\sum_{h=1}^{m} W_h \bullet s_h^2 + \frac{1}{n^2}\bullet\sum_{h=1}^{m}(1-W_h)\bullet s_h^2$$

Post-geruzapeneko kasu honetan, gelasketan helburu den aldagaiaren bariantza kontrolatzeko, helburuarekiko korrelazio handia duten aldagai koalitatiboak har daitezke. Horrela, aldagaien bariante anitzeko gurutzaketaren bidez lortutako geruzak homogeneoak izango dira helburu den aldagaiarekiko. (Ikus [4] eta [9])

Beraz, geruzapen nahiz post-geruzapeneko zenbatespenean izaten den errorea zuzenki proportzionala da gelasketako aldagaiaren bariantzarekiko, eta alderantziz proportzionala geruzetako laginaren neurriarekiko.

Aldagai Osagarri Koantitatiboak

Ausazko laginketa geruzatuan bi arrazoi-zenbatespen mota daude:

1) Arrazoiaren zenbatespen banandua

2) Arrazoiaren zenbatespen konbinatua

Guk arrazoiaren zenbatespen banandua hartuko dugu. Zenbatesle hau erregresio-eredu jakin bati dagokio, zeinetan erregresio-zuzenaren maldak ez baitu berdina izan beharrik geruzapen-gelaska bakoitzarentzat. Arrazoiaren zenbatespen konbinatuan, berriz, gelaska guztiek erregresio-malda berdina dute. Koaderno honetan Arrazoiaren Zenbatesle Banandua aztertuko dugu.

Populazioaren batezbestekoari dagokion ***Arrazoiaren Zenbatesle Banandua*** honakoa da: (Ikus [4])

$$\bar{Y}_{Rs} = \sum_{h=1}^{m} W_h^* \bullet \bar{Y}_h \qquad (16)$$

$\bar{Y}_h$ h geruzako populazioaren batezbestekoa delarik:

$$\bar{Y}_h = \sum_{\substack{k \in \, populazio- \\ -geruza\, h}} \frac{y_k}{N_h}$$

eta W_h^* aipatutako geruzako X aldagaiaren populazioaren guztizkoa eta batezbestekoaren arteko zatidura delarik:

$$W_h^* = \frac{\sum_{k=1}^{N} X_h^*(k)}{\sum_{k=1}^{n} X_h^*(k)},$$

$X_h^*(k) = X_h(k) \bullet X(k)$ izanik

Ondoriozko zenbatespena honakoa da:

$$\hat{\bar{Y}}_{Rs} = \sum_{h=1}^{m} W_h^* \bullet \bar{y}_h$$

non $\qquad \bar{y}_h = \sum_{\substack{k \in \, laginketa- \\ -geurza\, h}} \frac{y_k}{n_h}$ den

Zenbateslearen bariantzari dagokionez, geruza guztietan laginaren neurria nahiko handia duen ausazko laginketa geruzatuan, zenbateslearen bariantza honakoa da:

$$Var(\hat{\bar{Y}}_{Rs}) = \sum_{h=1}^{m} \frac{N_h^2 \bullet (1 - f_h)}{n_h} \bullet \left(S_{yh}^2 + R_h^2 S_{Xh}^2 - 2 R_h r_h S_{yh} S_{Xh}\right) \qquad (17)$$

eta, geruzapeneko gelaska bakoitzean arrazoiaren zenbateslea erabiltzen denez, nahikoa da zenbateslearen bariantza hartzea gelaska bakoitzean, non ausazko laginketa sinplea izaten den:

$$Var\left(\hat{Y}_{Rh}\right)=\frac{1-f_h}{n_h}\bullet\left[\frac{\sum_{\substack{k=1\\k\in geruzah}}^{N_h}\left(y_k-R_h x_k\right)^2}{N_h-1}\right] \quad (18)$$

R_h h geruzako Y eta Xren arteko lagin-batezbestekoaren arrazoia delarik, hau da, h geruzan Y eta X aladagaien arteko erregresio-zuzenari dagokion malda:

$$R_h=\frac{\sum_{k=1}^{n_h} y_k}{\sum_{k=1}^{n_h} x_k}$$

Aldagai osagarri koalitatiboen kasuan bezala, honako ondorioa aterako dugu:

Arrazoi Bananduaren zenbateslearen bariantza *alderantziz proportzionala* da geruzapeneko gelasken laginaren neurriarekiko, eta *zuzenki proportzionala* gelaska bakoitzeko arrazoiaren zenbateslearen bariantzarekiko.

Adibideak

Adibide guztietan *probabilitate-laginketak* aztertuko ditugu, eta bertako elementu guztiek hautespen-probabilitate berdinak izango dituzte, hots,

$$z_k = N / n \; \forall k = 1,..,n$$

Datuak taulatan bilduta aztertuko ditugu, eta aldagai osagarriak koalitatiboak direnean *kontingentzi taula* izena emango diegu.

Aldagai Osagarri Koalitatiboak

Azter ditzagun helburu diren bi *aldagai*, bata *koalitatiboa* eta bestea *koantitatiboa.*

Y_1 populazio *ez-langabetuaren kopurua* da.

Y_2 kolektibo batek *guztira irabazten duen dirua* da.

Bi *aldagai* osagarri daude:

- A: *Sexua*. 2 kategoria ditu: *a=2*.

- B: *Heziketa-maila*: Oinarrizko Heziketa, Bigarren Heziketa edo Unibertsitate--ikasketak. Hiru kategoria ditu: *b=3*.

Neurriak: populazio osoa *N=20* eta lagina *n=12*.

Honakoa lortzen dugu: axb=6 gelaska: 6 aldagai $X_{hh'}$ $1 \leq h \leq 2, 1 \leq h' \leq 3$ eta X_0 aldagaia direlarik

Aldagai osagarriak.

Populazio osoaren datuak. ($N_{hh'}$) *(1) taula*

$\sum_{k=1}^{N} X_{hh'}(k)$	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA
Gizona	4	4	1	9
Emakumea	3	5	3	11
GUZTIRA	7	9	4	20

Aldagai osagarriak.

*Populazio-laginaren datuak***.** ($n_{hh'}$) *(2) taula*

$\sum_{k=1}^{n} X_{hh'}(k)$	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoal	Unibertsitate--mailakoak	GUZTIRA
Gizona	1	1	1	3
Emakumea	3	3	3	9
GUZTIRA	4	4	4	12

Helburu den Y_1 aldagaia.

Populazio-laginaren datuak. $(Y_{1,hh'})$ *(3) taula*

Populazio-geruza bakoitzeko *ez-langabetuen* kopuruaren taula.

$\sum_{k=1}^{n} Y_1(k)$	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA
Gizona	0	1	0	1
Emakumea	3	2	0	5
GUZTIRA	3	3	0	6

Helburu den Y_2 *aldagaia*.

Populazio-laginaren datuak. $\left(Y_{2,hh'}\right)$ *(4) taula*

Geruza bakoitzeko hileko diru-sarrera guztira.

$\sum_{k=1}^{n} Y_2(k)$	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	60000	90000	150000	300000
Emakumea	300000	375000	600000	1275000
GUZTIRA	360000	465000	750000	1575000

Aldagai *osagarrientzat* lortzen ditugun matrizeak honakoak dira:

$$\mathbf{X} = \begin{pmatrix} X_{00}' \\ X_{11}' \\ X_{12}' \\ \dots \\ X_{hh}' \\ \dots \\ X_{ab}' \end{pmatrix} = \begin{pmatrix} 1 & 1 & & 1 \\ X_{11}(1) & X_{11}(2) & & X_{11}(N) \\ X_{12}(1) & X_{12}(2) & & X_{12}(N) \\ \dots\dots\dots & & & \\ X_{hh'}(1) & X_{hh'}(2) & & X_{hh'}(N) \\ \dots\dots\dots & & & \\ X_{ab}(1) & X_{ab}(2) & & X_{ab}(N) \end{pmatrix}$$

7xN matrizea da.

Kasu orokorrean (a*b+1)xN matrizea da, a*b populazioaren geruzapeneko geruza-kopurua delarik.

X matrizearen konposizioa informazioa errolda-moduan bilduta dagoenean lortuko da.

$\hat{X}$ matrizea X matrizearen zenbatespena da, 7xn dimentsiokoa

(6) formula aplikatuz, honako murrizketak lortuko ditugu:

$$\begin{pmatrix} 1 & 1 & 1 \\ X_{11}(1) & X_{11}(k) & X_{11}(n) \\ \dots\dots\dots & & \\ X_{hh'}(1) & X_{hh'}(k) & X_{hh'}(n) \\ \dots\dots\dots & & \\ X_{23}(1) & X_{23}(k) & X_{23}(n) \end{pmatrix} * \begin{pmatrix} w_1 \\ \dots \\ w_k \\ \dots \\ w_n \end{pmatrix} = \begin{pmatrix} X_{00}(1) & X_{00}(k) & X_{00}(n) \\ X_{11}(1) & X_{11}(k) & X_{11}(n) \\ \dots\dots\dots & & \\ X_{hh'}(1) & X_{hh'}(k) & X_{hh'}(n) \\ \dots\dots\dots & & \\ X_{23}(1) & X_{23}(k) & X_{23}(n) \end{pmatrix} * \begin{pmatrix} w_1 \\ \dots \\ w_k \\ \dots \\ w_n \end{pmatrix} =$$

$$
= \begin{pmatrix} N \\ \sum_{k=1}^{N} X_{11}(k) \\ \ldots \\ \sum_{k=1}^{N} X_{hh'}(k) \\ \ldots \\ \sum_{k=1}^{N} X_{23}(k) \end{pmatrix} = \begin{pmatrix} 20 \\ 4 \\ 4 \\ 1 \\ 3 \\ 5 \\ 3 \end{pmatrix}
$$

Geruzak orokortzeko arazoa batera konponduko dugu bi kasuetan (helburu den aldagai koalitatibo zein koantitatiboaren kasuan). Konpontzen erraza da, nahikoa da (1) eta (2) tauletako dautei (9) formula aplikatzea.

Honako orokortzaileak lortuko dira: $\hat{w}_{hh'} = \frac{N_{hh'}}{\hat{n}_{hh'}}$

	IKASKETAK		
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak
Gizona	4=4/1	4=4/1	1=1/1
Emakumea	1=3/3	1,667=5/3	1=3/3

Ez-langabetuen taula egokitu hau lortuko dugu:

Orokortzaileak x (3) taula

	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	0	4	0	4
Emakumea	3	3	0	6
GUZTIRA	3	7	0	10

Guztizko *diru-sarreraren taula egokitua* honakoa da:

	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	240000	360000	150000	750000
Emakumea	300000	625125	600000	1525125
GUZTIRA	540000	985125	750000	2275125

Guztizkoaren zenbatespenak, EZ-LANGABETUA IZAN zein DIRU-SARRERA GUZTIRA aldagai koalitatiboentzat, honakoak dira:

$\hat{Y}_1$ =10 ez-langabetuen zenbatetsitako kopurua

$\hat{Y}_2$ =2275125 zenbatetsitako hileko diru-sarrera guztira

Beraz honakoa daukagu:

- 20 laguneko populazioan zenbatetsitako ez-langabetuak guztira 10 direla, eta populazioaren batezbestekoa 10/20=0, 5 dela.

- Halaber, populazioak hilero 2275125 pezeta jasotzen duela guztira, eta zenbatetsitako batezbestekoa 2275125/20=113756,25 pezeta dela.

Aldagai Osagarri Koantitatiboak

Azter ditzagun berriz ere helburu diren bi *aldagai*: bat *koalitatiboa* eta bestea *koantitatiboa.*

Y_1 populazio *ez-langabetuaren* guztizkoa da.

Y_2 hileko *aurrezkiak* dira.

Hileko guztizko aurrrezkiak aldagai osagarria X(k) da.

Populazioaren geruzapena bi aldagaien arabera egingo dugu:

- A: esparru landatar edo hiritarrean bizitzea. 2 kategoria ditu: *a=2***.**

- B: Lurralde Historikoak: Bizkaia, Araba ala Gipuzkoa. 3 kategoria ditu: *b=3.*

N=20 populazioaren neurritzat eta *n=12 laginaren* neurritzat hartuz gero, axb=6 gelaska lortuko ditugu: 6 aldagai $X_{hh'}$ $1 \leq h \leq 2, 1 \leq h' \leq 3$ eta X_0 aldagaia direlarik.

Aldagai osagarria.

Populazio osoaren datuak. $\left(\sum_{K=1}^{N} X^{*}_{hh'}(K) = X^{*}_{hh'} \right)$ *(5) taula*

Populazioaren hileko diru-sarrera geruzaka.

$\sum_{k=1}^{N} X^{*}_{hh'}(k)$	LURRALDE HISTORIKOA			
ESPARRUA	Bizkaia	Araba	Gipuzkoa	GUZTIRA
Landatarra	440000	460000	420000	1320000
Hiritarra	380000	420000	400000	1200000
GUZTIRA	820000	880000	820000	2520000

Aldagai osagarria.

Populazio-laginaren datuak. $\left(\sum_{K=1}^{n} X^{*}_{hh'}(K) = \hat{X}^{*}_{hh'} \right)$ *(6) taula*

Laginaren hileko diru-sarrera geruzaka.

$\sum_{k=1}^{n} X^{*}_{hh'}(k)$	PROBINTZIA			
ESPARRUA	Bizkaia	Araba	Gipuzkoa	GUZTIRA
Landatarra	75000	250000	275000	600000
Hiritarra	120000	220000	250000	590000
GUZTIRA	195000	470000	525000	1190000

Erabiliko dugun notazioa aurreko kasuetako berdina da, baina lehen ageri ziren $X_{hh'}$ aldagaien lekuan $X^{*}_{hh'}$ aldagaiak agertuko dira orain. Aldagai berri horiek atal honen 1) zenbakian definitu ditugu.

Helburu den Y_1 aldagaia.

$\left(Y_{1,hh'}\right)$ *Populazio-laginaren datuak.* *(7) taula*

Laginaren geruza bakoitzeko *ez-langabetuen kopurua*.

$Y_{1,hh'}$	LURRALDE HISTORIKOA			
ESPARRUA	Bizkaia	Araba	Gipuzkoa	GUZTIRA
Landatarra	0	1	0	1
Hiritarra	3	2	0	5
GUZTIRA	3	3	0	6

Helburu den Y_2 *aldagaia*.

$\left(Y_{2,hh'}\right)$ *populazio-laginaren datuak* *(8) taula*

Hileko aurrezkien guztizkoa geruzaka.

$Y_{2,hh'}$	LURRALDE HISTORIKOA			
ESPARRUA	Bizkaia	Araba	Gipuzkoa	GUZTIRA
Landatarra	20000	175000	200000	395000
Hiritarra	25000	175000	70000	270000
GUZTIRA	45000	350000	270000	665000

(10) formula murrizketetan aplikatuz, honakoa lortuko dugu:

$$\begin{pmatrix} 1 & 1 & 1 \\ X^*_{11}(1) & X^*_{11}(k) & X^*_{11}(n) \\ & \ldots\ldots\ldots & \\ X^*_{hh'}(1) & X^*_{hh'}(k) & X^*_{hh'}(n) \\ & \ldots\ldots\ldots & \\ X^*_{23}(1) & X^*_{23}(k) & X^*_{23}(n) \end{pmatrix} * \begin{pmatrix} w_1 \\ \ldots \\ w_k \\ .. \\ w_n \end{pmatrix} = \begin{pmatrix} X^*_{00}(1) & X^*_{00}(k) & X^*_{00}(n) \\ X^*_{11}(1) & X^*_{11}(k) & X^*_{11}(n) \\ & \ldots\ldots\ldots & \\ X^*_{hh'}(1) & X^*_{hh'}(k) & X^*_{hh'}(n) \\ & \ldots\ldots\ldots & \\ X^*_{23}(1) & X^*_{23}(k) & X^*_{23}(n) \end{pmatrix} * \begin{pmatrix} w_1 \\ \ldots \\ w_k \\ .. \\ w_n \end{pmatrix} =$$

$$\begin{pmatrix} \sum_{k=1}^{N} X(k) \\ \sum_{k=1}^{N} X_{11}^{*}(k) \\ \ldots \\ \sum_{k=1}^{N} X_{hh'}^{*}(k) \\ \ldots \\ \sum_{k=1}^{N} X_{23}^{*}(k) \end{pmatrix} = \begin{pmatrix} 2520000 \\ 440000 \\ 460000 \\ 420000 \\ 380000 \\ 420000 \\ 400000 \end{pmatrix}$$

Orokortzaile hauek lortzen dira:

$W_{hh'}^{*}$	LURRALDE HISTORIKOA		
ESPARRUA	Bizkaia	Araba	Gipuzkoa
Landatarra	5,867= 440000/75000	1,84= 460000/250000	1,527= 420000/275000
Hiritarra	3,167= 380000/120000	1,909= 420000/220000	1,6= 400000/250000

$W_{hh'}^{*} = \dfrac{X_{hh'}^{*}}{\hat{X}_{hh'}^{*}}$ direlarik, (11) formulan definitu ditugunez.

Honako *ez-langabetuen* taula *egokitua* lortuko dugu: *(7) taula x orokortzaileak*

$W_{hh'}^{*} \bullet Y_{1,hh'}$	LURRALDE HISTORIKOA			
ESPARRUA	Bizkaia	Araba	Gipuzkoa	GUZTIRA
Landatarra	0	2	0	2
Hiritarra	9	4	0	13
GUZTIRA	9	6	0	15

Honako *hileko diru-sarreraren guztizkoaren* taula *egokitua* lortuko dugu:

(8) taula x orokortzaileak

$W^{*}_{hh'} \bullet Y_{2,hh'}$	LURRALDE HISTORIKOA			
	Bizkaia	Araba	Gipuzkoa	GUZTIRA
Landatarra	117340	322000	305400	744740
Hiritarra	79175	334075	112000	525250
GUZTIRA	196515	656075	417400	1269990

Guztizkoaren zenbatespenak, ez-langabetua aldagai koalitatiboa zein *aurrezki* aldagai koantitatiboarentzat, honakoak dira:

$\hat{Y}_1$ =15 populazioaren zenbatetsitako ez-langabetuen kopurua da

$\hat{Y}_2$ =1269990 populazioak hilero guztira aurrezten duenaren zenbatespena da

Ondorioak honakoak dira beraz:

- 20 laguneko populazioan zenbatetsitako ez-langabetuen kopurua 15 da; beraz, populazio horretako ez-langabetuen batezbestekoa 15/20=0,75 da.

- Halaber, populazioak hilero guztira 1269990 pezeta aurrezten dituela zenbatesten da. Zenbatetsitako batezbestekoa hileko 1269990/20=63499,5 pezeta da.

Prozedura informatikoak

Informazio Osagarri Handiena dugunean, software ezberdina erabil daiteke laginen egokitzapena egiteko.

EXCEL

EXCEL kalukulu-orria erabiliz, erraza da orokortzaileak kalkulatzea. Halaber, WINDOWS Office-ren tresna da, eta horrek atzipena eta erabilera errazten ditu. Tresna hau egokitzapenerako erabiltzean aurkituko dugun zailtasuna orokortzaileak kalkulatzeko makro bat egin beharra da. Makro horrek datu-taulen neurriak aurretik ezarriak eduki beharko ditu. Beraz, metodoa ezingo da datu-base orokorrekin erabili.

SAS

Datu-baseak erabili eta tratatzeko erabiltzen den programazio-lengoaia da. Beraren bidez erraza da egokitzapen-orokortzaileak definitzea, baina lan-ingurunea es da Excel--ena bezain erosoa.

Beste zenbait software

Aipatuez gain, matrizeak erabilitzea posible egiten duten programazio-lengoaia guztiak egokiak dira prozedura honetarako. Horietakoak dira FORTRAN, PASCAL, C++,. …

Programaren helburua orokortzaileak definitzea izango da, eta horiek geruza bakoitzeko guztizko populazioa eta populazio-laginaren zatidurak emango ditu.

Aplikazio-egoera egokienak

Aipatutako guztia kontuan hartuta, aplikaziorako behar den egoera honakoa izango litzateke:

- Geruzapenaren gelaska-kopurua ez da oso handia izango, bestela laginaren gelaskako banaketa ez baita oso dentsoa izango, eta baliteke geruza batzuetan elementu-kopurua oso txikia izatea: 20-25 gutxienekotik beherako gelaska guztiek arazoak sortuko dituzte. Arazo horiek konpontzeko gelaska-kolapsora jo daiteke, zeinetan populazioari buruzko informazio galdu eta alborapena sortzen baita. Bariantza ere jaitsi egiten da, alderantziz proportzionala baita gelasken laginketa-neurriarekiko, eta horrela neurriak behar bezain handiak izatea lortuko dugu.

- Metodoak estaldura eza konpentsatzen du. Elementuen-laginean hautespen-probabilitateekin batera erantzun ez dutenen tasa erabiliz gero, erantzun eza eta estaldura eza konpentsatuko lukeen egokitzapena lortuko genuke.

3.
kapitulua

Informazio osagarri koalitatiboko iterazio-prozedurak

Kapitulu honetan bi iterazio-prozedura aurkeztuko ditugu, RAKING USUAL izeneakoa eta bere aldaera bat, REDRE prozeduran erabiltzen dena. Aldagai osagarri koalitatiboak ditugunean, eta aldagai osagarri horien bariante bakarreko antolamenduetan informazio handiena dagoenean erabiltzen dira biak. Geruzapena aldagai osagarrien bariante bakarreko gurutzaketaren bidez egiten da.

Notazioa

Laginaren bariante bakarreko antolamendu haztatua populazioaren berdina izan dadin, zenbait murrizketa egiten dira. Murrizketa horien adierazpen matriziala ez da aldatzen:

$$\hat{X} \bullet \mathbf{w} = \mathbf{X} \bullet \mathbf{I}$$

X eta $\hat{X}$ matrize osagarrien adierazpena honakoa da:

L aldagai osagarri koalitatibo ditugu, bakoitza $L_1, L_2, \ldots, L_l, \ldots, L_L$ modalitateekin. L aldagai horien bariante anitzeko gurutzaketaren bidez populazioa geruzatu egiten da, $L_1 \bullet L_2 \bullet \ldots \bullet L_l \bullet \ldots \bullet L_L$ geruzak lortuz. Modalitateen kopurua $L_1 + L_2 + \ldots + L_l + \ldots + L_L = m$ da guztira.

Kasu honetan, modalitateen aldagai identifikadore edo dikotomikoak hartuko ditugu:

$$X_{h_l}(k) = \begin{cases} 1, \text{baldin k - k } h_l \text{ modalitatea badu} \\ 0, \text{bestela} \end{cases} \qquad 1 \le h_l \le m$$

k=1,..,n laginean eta k=1,..,N populazioan direlarik.

$\Rightarrow L_1 + \ldots + L_l + \ldots + L_L = m$ aldagai dikotomikoak.

Populazioaren errenkadako bektoreak hartuko ditugu:

$X_{h_l} = \left(X_{h_l}(1), \ldots, X_{h_l}(k), \ldots, X_{h_l}(N)\right)$, (1xN) dimentsiokoak, $1 \le h_l \le m$ delarik

eta laginarenak

$\hat{X}_{h_l} = \left(\hat{X}_{h_l}(1), \ldots, \hat{X}_{h_l}(k), \ldots, \hat{X}_{h_l}(n)\right)$, (1xn) dimentsiokoak, $1 \le h_l \le m$ delarik

Honakoa ziurtatzeko:

$$w_1 + .. + w_k + .. w_n = N$$

X_0 aldagaia identikoki 1 hartu eta honela adierazten da:

$X_0 = (X_0(1),.., X_0(k),.., X_0(N)) = (1,..,1,..,1)$ (1xN) dimentsioko populazioaren errenkadako bektorea

$\hat{X}_0 = (X_0(1),.., X_0(k),.., X_0(n)) = (1,..,1,..,1)$ (1xn) dimentsioko laginaren errenkadako bektorea

X eta $\hat{X}$ matrizeak honako errenkadako bektoreek osatzen dituzte:

$$\mathbf{X} = \begin{pmatrix} X_0 \\ X_1 \\ \cdots \\ X_h \\ \cdots \\ X_m \end{pmatrix} \quad \textbf{eta} \quad \hat{X} = \begin{pmatrix} \hat{X}_0 \\ \hat{X}_1 \\ \cdots \\ \hat{X}_h \\ \cdots \\ \hat{X}_m \end{pmatrix}$$ #

Hots, (m+1)xN eta (m+1)xn dimentsioko bi matrize dira.

Matrize-adierazpena garatuz, honakoa lortuko dugu:

$$\begin{cases} \sum_{k=1}^{n} w_k \bullet X_0(k) = \sum_{k=1}^{N} X_0(k) \Leftrightarrow w_1 + .. + w_k + .. w_n = N \\ \sum_{k=1}^{n} w_k \bullet X_h(k) = \sum_{k=1}^{N} X_h(k), \forall 1 \le h \le m \end{cases}$$

Notazioa erraztu egingo dugu, soilik bi modalitate (*a* eta *b*) dituzten aldagai osagarriak hartuz. Modalitateak zenbakituz gero, eskura ditugun $N_{h.}$ y $N_{.h'}$ datu marjinalen arabera, honakoa daukagu:

$$\begin{cases} N_h = N_{h.} \quad 1 \le h \le a \quad \text{delarik} \\ N_{h'} = N_{.h'} \quad a+1 \le h' \le a+b \quad \text{delarik} \end{cases}$$ *a*+*b* =m delarik

$\rightarrow$ 1. aldagaiaren bariante bakarreko antolamendua

$\rightarrow$ 2. aldagaiaren bariante bakarreko antolamendua

eta era berean, bariante bakarreko antolamenduaren laginketa-datuentzako notazioa definituko dugu:

$$\begin{cases} n_h = n_{h.} \quad 1 \le h \le a \quad \text{delarik eta} \\ n_{h'} = n_{.h'} \quad a+1 \le h' \le a+b \quad \text{delarik} \end{cases}$$

Datu horien adierazpena, bere aldagai osagarriekin hartuz, honakoa daukagu:

Modalitateak 1etik m-ra zenbakitu ditugu.

$$\begin{cases} \sum_{k=1}^{N} X_h(k) = N_h \\ & 1 \le h \le \mathrm{m} \quad \text{delarik} \\ \sum_{k=1}^{n} X_h(k) = n_h \end{cases}$$

Sistema askatu egingo dugu

Egokitzapena gelaskaka egiten da, gelaska bakoitzeko indibiduo guztiek pisu bera dutelarik. Ezezagunak $W_{hh'}$ gelasketako orokortzaile bihurtzen dira, $1 \le h \le a,\ a+1 \le h' \le a+b$ delarik. Eta sistemak honako forma dauka:

$$\begin{cases} \sum_{h=1}^{a} \sum_{h'=a+1}^{a+b} W_{hh'} = N \\ \sum_{h'=a+1}^{a+b} W_{hh'} \bullet n_{hh'} = \sum_{k=1}^{N} X_h(k) = N_h = N_{h.}, \forall 1 \le h \le a \\ \sum_{h=1}^{a} W_{hh'} \bullet n_{hh'} = \sum_{k=1}^{N} X_{h'}(k) = N_{h'} = N_{.h'}, \forall a+1 \le h' \le a+b \end{cases} \tag{18}$$

Oro har, $L_1 \bullet L_2 \bullet \ldots \bullet L_l \bullet \ldots \bullet L_L$ ezezagunak eta $L_1 + L_2 + \ldots + L_l + \ldots + L_L + 1 = m+1$ ekuazioak dituen sistema lortzen da, zeinetan azkena aurrekoen konbinazio lineala baita. Beraz $L_1 \bullet L_2 \bullet \ldots \bullet L_l \bullet \ldots \bullet L_L$ koefiziente ezezagun eta $L_1 + L_2 + \ldots + L_l + \ldots + L_L = m$ ekuazio sistema baino ez ditugu. Bi aldagai osagarri besterik ez daudenean, *a*b* ezezagun eta *a+b* ekuazioko sistema bihurtzen da.

Gehienetan ekuazio baino ezezagun gehiago egoten dira. Beraz sistemak orokortzaile--kopuru infinitua izan dezake. Hori saihesteko funtzio batentzako minimizazio--irizpidea hartzen da, horrela emaitza bakarra lortu ahal izateko. Hasierako pisuetatik bukaerarako distantzia-bektorea hartuko da funtziotzat. Distantzia-funtzioaren arabera, emaitzak ezberdinak izango dira.

Prozeduraren deskribapena

Ondoren, Raking Usual zein Redreko haren aldaeran erabiltzen diren iterazio--prozedurak deskribatuko ditugu.

Prozesu bakoitzeko iterazio-formula deskribatu aurretik, informazioari dagokion egoera azalduko dugu, bi prozesuetan berdina dena:

Populazioari buruzko bariante anitzeko taula honakoa da:

A\B	1	2		h'		b	Guztira
1	W_{11}	W_{12}		$W_{1h'}$		W_{1b}	$W_{1.} = W_{1}$
2	W_{21}	W_{22}		$W_{2h'}$		W_{2b}	$W_{2.} = W_{2}$
h	W_{h1}	W_{h2}		$W_{hh'}$		W_{hb}	$W_{h'.} = W_{h}$
a	W_{a1}	W_{a2}		$W_{ah'}$		W_{ab}	$W_{a.} = W_{a}$
Guztira	$W_{.1} = W_{a+1}$	$W_{.2} = W_{a+2}$		$W_{.h'} = W_{a+h'}$		$W_{.b} = W_{a+b}$	1

Taula honetan ezagutzen ditugun datu bakarrak bariante bakarreko antolamenduari dagozkionak dira, gelaska grisetan daudenak, alegia.

Laginketa-datuei dagokienez, lortzen den taulak honako forma dauka:

A\B	1	2		h'		b	Guztira
1	q_{11}	q_{12}		$q_{1h'}$		q_{1b}	$q_{1.} = q_1$
2	q_{21}	q_{22}		$q_{2h'}$		q_{2b}	$q_{2.} = q_2$
h	q_{h1}	q_{h2}		$q_{hh'}$		q_{hb}	$q_{h.} = q_h$
a	q_{a1}	q_{a2}		$q_{ah'}$		q_{ab}	$q_{a.} = q_a$
Guztira	$q_{.1} = q_{a+1}$	$q_{.2} = q_{a+2}$		$q_{.h'} = q_{a+h'}$		$q_{.b} = q_{a+b}$	1

Kasu honetan baterako antolamendua ezagutzen da, bariante anitzekoa.

Kasu orokorra bi aldagai osagarriko egoerara murriztu dugu, eta suposizio horretatik bi iterazio-prozedurak azalduko ditugu. Horretarako taula irudikari sinpleei aplikatuz argituko ditugu bi iterazio-prozedurak.

Simulazioa

Populazioaren tamaina *20 da*, eta *laginak 10* ale *ditu*.

Populazioari buruz datu hauek ezagutzen ditugu $(W_{hh'})$:

A\B	1	2	3	Guztira
1	W_{11}	W_{12}	W_{13}	$W_{1.} = 0{,}75$
2	W_{21}	W_{22}	W_{23}	$W_{2.} = 0{,}25$
Guztira	$W_{.1} = 0{,}2$	$W_{.2} = 0{,}4$	$W_{.3} = 0{,}4$	1

Laginketa-datuen taula honakoa da $(q_{hh'})$:

A\B	1	2	3	Guztira
1	$q_{11} = 0{,}1$	$q_{12} = 0{,}2$	$q_{13} = 0{,}2$	$q_{1.} = 0{,}5$
2	$q_{21} = 0{,}2$	$q_{22} = 0{,}1$	$q_{23} = 0{,}2$	$q_{2.} = 0{,}5$
Guztira	$q_{.1} = 0{,}3$	$q_{.2} = 0{,}3$	$q_{.3} = 0{,}4$	1

Raking Usual

RAKING metodoa definitzen duen iterazio-algoritmoa honakoa da: (Ikus [4] eta [5])

$$w(k,m) = w(k,m-1) * \frac{W_{h_0}}{q_{h_0,m-1}} * \frac{W_{h'_0}}{q'_{h'_0,m-1}}$$

$$\begin{cases} q'_{h',0} = \sum_{h=1}^{a} q'_{hh',0} = \sum_{h=1}^{a} q_{hh',0} \bullet \frac{W_{h.}}{q_{h,0}} \\ q'_{h',m} = \sum_{h=1}^{a} q'_{hh',m} = \sum_{h=1}^{a} q_{hh',m} \bullet \frac{W_{h}}{q_{h,m}}, \ \mathrm{m} \geq 1 \ \ \text{delarik} \end{cases} \quad a+1 \leq h' \leq a+b \text{–rentzat}$$

eta k elementua (h_0, h'_0) gelaskan.

Ondoren iterazio-formula simulatutako adibideari aplikatuko diogu. Horretarako, lortutako pisuak bilduko dituzten taulak aurkeztuko ditugu, eta pisu horien arabera lortutako laginaren antolamendu haztatua ere bai.

PISUAK (0) =20/10:

2	2	2
2	2	2

0 $(q_{hh',0})$ pisuen arabera haztatutako laginketa-taula**:**

A\B	1	2	3	Guztira
1	0,2	0,4	0,4	1
2	0,4	0,2	0,4	1
Guztira	0,6	0,6	0,8	2

Iterazioaren lehen pausoa lagina lehen ezaugarri osagarriaren edo geruzapen ezaugarriaren arabera egokitzea da. Egokitzapenaren ondorioz lortzen diren orokorpen--faktoreak honakoak dira:

0,75	0,75	0,75
0,25	0,25	0,25

$(q'_{hh',0})$ *portzentaien laginketa-taula haztatua* lortuko dugu:

A\B	1	2		Guztira
1	0,15	0,3	0,3	0,75
2	0,1	0,05	0,1	0,25
Guztira	0,25	0,35	0,4	1

Lehen iterazioko bigarren pausoan laginaren antolamendua bigarren geruzapen--aldagaira egokitzen da. Iterazioaren bigarren pausoko laginaren orokortze-faktoreak hauek dira:

0,8	1,143	1
0,8	1,143	1

Orokortzaile berriak erabiliz lagina *orokortzean* $(q_{hh',1})$ *portzentaien laginketa-taula* hau lortzen da:

A\B	1	2		Guztira
1	0,12	0,343	0,3	0,763
2	0,08	0,057	0,1	0,237
Guztira	0,2	0,4	0,4	1

Laginaren taula *desegokitu egin da orain lehen geruzapen-aldagaiarekiko*. Beraz prozeduraren *bigarren iteraziora* joko dugu:

Bigarren iterazioaren lehen pausoa. Egokitzapenaren ondorioz lortzen diren *orokortze--faktoreak* honakoak dira:

0,983	0,983	0,983
1,055	1,055	1,055

Eta $(q'_{hh',1})$ *portzentaien laginketa-taula haztatua* lortuko dugu:

A\B	1	2		Guztira
1	0,118	0,337	0,295	0,75
2	0,084	0,06	0,106	0,25
Guztira	0,202	0,397	0,401	1

Bigarren iterazioaren bigarren pausoa. Faktoreak hauek dira:

0,99	1,008	0,998
0,99	1,008	0,998

Orokortzaile berri horien bidez lagina *orokortzean* lortzen den $\left(q_{hh',2}\right)$ *portzentaien laginketa-taula* honakoa da:

A\B	1	2		Guztira
1	0,117	0,34	0,294	0,751
2	0,083	0,06	0,106	0,249
Guztira	0,2	0,4	0,4	1

Laginketa-portzentaia haztatuek populazioaren portzentaia ezagunekiko sortzen duten errorea 0,0035 ingurukoa da, eta, soilik 3 digituekin lanean ari garenez, perdoi hori nahikoa da egokitzapenerako. Beraz, egokitzapena bukatutzat ematen da, bi iterazioekin, eta horietako bakoitzari dagozkion bi pausoekin.

Lortzen den taula *populaziora egokitutako laginketa-portzentaiena* da. Geruzapen--gelaska bakoitzean zenbatetsitako *populazio-aleen* kopurua lortzeko, nahikoa da taula eta N=20 populazioaren tamaina biderkatzea. Zenbatetsitako aleen honako taula lortuko dugu:

A\B	1	2		Guztira
1	2	7	6	15
2	2	1	2	5
Guztira	4	8	8	20

Argi dago laginaren antolamendua desegokitu egin dela lehen geruzapen--aldagaiarekiko, bigarrenera egokituta. Ondorioz, prozedura berriz ere iteratu egingo da, prozesua taula haztatu berriarekin hasiko delarik, Raking-aren lehen iterazioan lortutakoarekin, hain zuzen.

Iterazio-formula honako prozesuan lortzen da:

- Lehenengo aldagai osagarriaren bariante bakarreko antolamendua hartzen da. h lerroko gelaska guztiak $\frac{W_{h.}}{q_{h.,0}} = \frac{W_h}{q_{h,0}}$ faktorearen bidez haztatzen dira, laginaren antolamendu marjinal osagarri haztatua populaziora egokitzeko. Lehen pauso honetan lortzen den laginaren antolamendua honela adierazten da: $q'_{hh',0}$.

- Lortutako laginaren antolamendua egokitzen da, bigarren aldagai osagarriaren antolamendu marjinalera egokitzeko. Horretarako $\dfrac{W_{\cdot h'}}{q'_{\cdot h',0}} = \dfrac{W_{h'}}{q'_{h',0}}$ orokortze--faktorea hartzen da.

Aurreko pausoekin iterazioa osatzen da, eta ondoren, laginaren antolamendua $\dfrac{W_{h\cdot}}{q_{h\cdot,0}} \bullet \dfrac{W_{\cdot h'}}{q'_{\cdot h',0}} = \dfrac{W_h}{q_{h,0}} \bullet \dfrac{W_{h'}}{q'_{h',0}}$ faktorearen bidez orokortzen da.

Ondorioz sortzen den antolamendua bigarren aldagai osagarriaren bariante bakarreko antolamendura zuzenean egokituta dago, baina bigarren pausoa dela eta, desegokitu egin da lehen aldagai osagarriaren bariante bakarreko antolamenduarekiko.

Edozein m iteraziorako, lehenbiziko egokitzapenean $\dfrac{W_{h\cdot}}{q_{h\cdot,m-1}} = \dfrac{W_h}{q_{h,m-1}}$ faktoreak lortuko ditugu, eta bigarren pausoan $\dfrac{W_{\cdot h'}}{q'_{\cdot h',m-1}} = \dfrac{W_{h'}}{q'_{h',m-1}}$ faktoreak. Ondorioz, azken pisuak honakoak dira:

$$w(k,m) = w(k,m-1) \bullet \frac{W_h}{q_{h',m-1}} \bullet \frac{W_{h'}}{q'_{h',m-1}} \quad (h,h') \text{ geruzean } k \text{ delarik}$$

Prozedura *iterazio-kopuru finkora* heldu arte iteratzen jarraitzen da, edota bariante bakarreko populazioaren antolamenduarekin *konbergentzia* egon arte.

Prozesu iteratiboa murrizketa-sistema hartu eta *bi sistematan* banatuz lortzen da:

1. Lehenbiziko sistema, non *lehen aldagai osagarriari* dagozkion murrizketak egiten diren, ondoriozko pisuek honako forma lortzen dutelarik:

$$\boxed{w'(k,m-1) = w(k,m-1) * c_h}$$

Orduan a ezezagun eta a ekuazioko sistema sortzen da, matrize elkartu

lortutako laginketa-aleen kopurua direlarik. Sistema honek ebazpen bakarra dauka, baldin elementu diagonalak nuluak ez badira. Iterazio bakoitzeko lehen pausoa sistema

2. Bigarren pausoa ana

$$\boxed{w(k,m) = w'(k,m-1) * c_{h'}}$$

eta orduan b ekuazio eta b murrizketako sistema eratzen da, $c_{h'}$ ezezagunak dituena.

Sistema honek matrize diagonala dauka matrize elkartutzat, elementu diagonalak bigarren aldagai osagarriaren modalitateentzat lortutako laginketa-ale marjinalen kopurua direlarik. Elementu horiek nuluak ez badira, sistemak *soluzio bakarra* dauka. Ondorioz, iterazioaren bigarren pausoan antolamendu haztatua lortzen da, eta, bi aldagai baino ez daudenez, iteraziorako azken pisuak ere bai.

Aldagai osagarri gehiago daudenean, sistema gehiago sortzen dira iterazio bakoitzean, aldagaien kopurua adina.

RAKING iterazio-prozeduraren bidez lortzen den ebazpena *Karratuen Minimoen Egokitzapena* izeneko metodoaren bidez lortzen den berdina da: hartzen den sistemak murrizketa batzuk ditu, laginaren antolamendu haztatuaren populazioari buruzko informazio osagarriaren finkotasunari dagozkionak. Horrez gain, helburu den funtzioa dauka, aipatutako laginaren antolamendua zehatza den ala ez neurtzen duena, honako adierazpena duen funtzioa distantzia gisa hartuz: $D(x,y) = \frac{(x-y)^2}{x}$.

Horrela, distantzia-funtzioa modalitate bakoitzean bariante bakarreko laginketa--antolamenduen eta populazio-antolamenduen artean dagoen distantziaren batura da, eta minimizatu behar den distantzia-funtzioa lortzen da.

Murrizketak dituen aldagaiek osatutako helburu den funtzioaren minimizazioaren arazoa konpontzeko *Lagrange-ren biderkatzaileen* metodoa erabiltzen da. Biderkatzaileen bidez egiten den askapenak Raking-aren soluzio berdina ematen du, baina azken prozedura hau azkarragoa da.

Raking bidez guretzat interes berezia duen gelaska baterako egokitzapena nahiko azkar egin dezakegu, gainerako gelaskak konpresioz ondoko gelaska bihurtuz.

Ikusten dugunez, Raking-a modalitateen antolamendu marjinala aldagai osagarrietan egokitzen duen iterazio-prozedura da.

Azpimarratu beharra dago geruzapeneko gelaskaren batek lagin-elementurik ez baldin badauka $(n_{hh'} = 0)$, Raking-ean gutxi gorabeherako populazio-aleen kopurua ematen zaiola, $\tilde{N}_{hh'} = 0$ eta $\tilde{W}_{hh'} = 0$, hurrenez hurren.

Raking-ak zenbait erregulartasun-egoeratan bat egiten du. Konbergentzia-kasu horietan, $N_{hh'}$ eta $\tilde{N}_{hh'}$ zenbatesleak asintotikoki alboragabeak dira, antolamendu arruntekoak eta bariantza txikienekoak. ZAAO (Zenbatesle Asintotikoki Arrunt Onenak) zenbatesleak dira, beraz.

Redre

RAKING USUAL metodoaren aldaera da. Iterazio-formula honek definitzen du:

$$w(k,m) = w(k,m-1) * \left(\sum_{h=1}^{a} \sum_{h'=1}^{b} X_{hh'}(k) * \left(\frac{\left(\sum_{k=1}^{N} X_h(k)\right) \Big/ N}{\left(\sum_{k=1}^{n} X_h(k) * w(k,m-1)\right) \Big/ n} + \frac{\left(\sum_{k=1}^{N} X_{h'}(k)\right) \Big/ N}{\left(\sum_{k=1}^{n} X_{h'}(k) * w(k,m-1)\right) \Big/ n} \right) \right) \Big/ 2$$

- $w(k,m)$ m-garren iterazioan k elementuari emandako pisua delarik,

$$w(k,0) = 1$$

- $\left(\sum_{k=1}^{N} X_h(k)\right) \Big/ N$ formulak h modalitateko populazio-elementuen portzentaia adierazten du, modalitate hori lehen aldagai osagarriarenetako bat delarik. Lagin nahiz populazioko ale-kopuruaren izenak hartuz gero, portzentaiek honako forma dute:

$$\begin{cases} N_h \big/ N & 1 \leq h \leq a \quad \text{delarik} \\ N_{h'} \big/ N & a+1 \leq h' \leq a+b \quad \text{delarik} \end{cases}$$

Notazioa are eta gehiago errazteko, populazio-portzentaiak honela adieraziko ditugu:

$$W_h = \frac{\left(\sum_{k=1}^{N} X_h(k)\right)}{N} = \frac{N_h}{N} \; eta \; W_{h'} = \frac{\left(\sum_{k=1}^{N} X_{h'}(k)\right)}{N} = \frac{N_{h'}}{N}$$

- $\left(\sum_{k=1}^{n} X_{h'}(k) * w(k,m-1)\right) \Big/ n$ *eta* $\left(\sum_{k=1}^{n} X_h(k) * w(k,m-1)\right) \Big/ n$ m iterazioan lortutako pisuen arabera lortutako laginketa-portzentaia haztatuak dira. Lehen eta bigarren aldagai osagarrien modalitateei dagozkien portzentaia marjinalak dira. Adierazpenak errazteko honako notazioa erabiliko dugu:

$$q_{h,\,m-1} = \left(\sum_{k=1}^{n} X_h(k) * w(k,m-1)\right) \Big/ n \quad eta$$

$$q_{h',\,m-1} = \left(\sum_{k=1}^{n} X_{h'}(k) * w(k,m-1)\right) \Big/ n$$

m=1en kasuan, laginketa-portzentaia haztatuak honakoak dira:

$$\begin{cases} \left(n_h \big/ n \right) & 1 \le h \le a \quad delarik\ eta \\ \left(n_{h'} \big/ n \right) & a+1 \le h \le a+b \quad delarik \end{cases},$$

Adierazpen horiek $w(k,0)=1$ zuzenean hartuta lortzen dira.

- 2: aldagai osagarrien kopurua adierazten du.

- $\sum_{h=1}^{a} \sum_{h'=1}^{b} X_{hh'}(k)$ iterazio bakoitzean lortutako pisuan eragina izango duten faktoreak aurkitzeko balio duen iragazkia da, $X_{hh'}(k)$ $1 \le h \le a$ eta $1 \le h' \le b$ delarik, gelaska bakoitzari lotzen zaizkion aldagai dikotomikoak direlarik. Hau da, probleman erabili dugun notazioaren arabera, $X_{hh'}(k) = X_h(k) * X_{h'}(k)$.

Orain arte esandakoa kontuan hartuta, REDREren iterazio-formula honakoa da:

$$w(k,m) = w(k,m-1) * \left(\sum_{h=1}^{a} \sum_{h'=1}^{b} X_{hh'}(k) * \left(\frac{\dfrac{W_h}{q_{h,\, m-1}} + \dfrac{W_{h'}}{q_{h',\, m-1}}}{2} \right) \right)$$

Aurreko iterazio-formula orokortuz, hau da, aldagai osagarri gehiagoren kasu orokorrera eramanez, honakoa lortuko dugu:

$$w(k,m) = w(k,m-1) * \left(\frac{batura\left(\dfrac{W_h}{q_{h,\, m-1}} \right)}{aldagai\ osagarrien\ kop.} \right)$$ baturak k elementua duten laginaren modalitate guztientzat, aldagai osagarri guztientzat, balio duelarik.

Pisu horiek $\frac{N}{n}$-z biderkatzen dira gero EUSTATen, $\sum_{k=1}^{n} w(k,m) = N$ lortzeko. Hori, $w(k,0) = \frac{N}{n}$ hasierako pisu gisa hartzearen parekoa da, hau da, indibiduo bakoitzak laginerako hautatua izateko dituen probabilitateen alderantziak hartzea bezala, hots, probabilitate-laginketetan lagineko indibiduo guztientzat H-T hasierako zenbateslea $z_k = \frac{n}{N}$ hautespen-probabilitate berdinaz hartzea bezala.

Gelaskaka egiten den egokitzapen iteratiboa da, eta bertan gelaska bateko ale guztiei pisu berdina ematen zaie.

Azter dezagun gure adibide simulatuari aplikatuz REDREk nola jarduten duen:

Lehen iterazioa

Orokortzaileak

=(0,75/0,5+0,2/0,3)/2 =**1,083**	=(0,75/0,5+0,4/0,3)/2 =**1,417**	=(0,75/0,5+0,4/0,4)/2 =**1,25**
=(0,25/0,5+0,2/0,3)/2 =**0,583**	=(0,25/0,5+0,4/0,3)/2 =**0,917**	=(0,25/0,5+0,4/0,4)/2 =**0,75**

Populaziora egokitutako laginketa-taula portzentaiatan:

A\B	1	2	3	Guztira
1	0,108	0,283	0,25	0,641
2	0,117	0,092	0,15	0,359
Guztira	0,225	0,375	0,4	1

Bigarren iterazioa

Orokortzaileak

=(0,75/0,641+0,2/0,225)/2 =**1,029**	=(0,75/0,641+0,4/0,375)/2 =**1,118**	=(0,75/0,641+0,4/0,4)/2 =**1,085**
=(0,25/0,359+0,2/0,225)/2 =**0,793**	=(0,25/0,359+0,4/0,375)/2 =**0,882**	=(0,25/0,359+0,4/0,4)/2 =**0,848**

Populaziora egokitutako laginketa-taula portzentaiatan:

A\B	1	2	3	TOTAL
1	0,111	0,316	0,271	0,698
2	0,093	0,081	0,127	0,301
TOTAL	0,204	0,397	0,398	0,999

Hirugarren iterazioa

Orokortzaileak

=(0,75/0,698+0,2/0,204)/2 =**1,027**	=(0,75/0,698+0,4/0,397)/2 =**1,041**	=(0,75/0,698+0,4/0,398)/2 =**1,04**
=(0,25/0,301+0,2/0,204)/2 =**0,905**	=(0,25/0,301+0,4/0,397)/2 =**0,919**	=(0,25/0,301+0,4/0,398)/2 =**0,918**

Populaziora egokitutako laginketa-taula portzentaiatan:

A\B	1	2	3	Guztira
1	0,114	0,329	0,282	0,725
2	0,084	0,074	0,117	0,275
Guztira	0,198	0,403	0,399	1

Laugarren iterazioa

Orokortzaileak

=(0,75/0,725+0,2/0,198)/2 =**1,022**	=(0,75/0,725+0,4/0,403)/2 =**1,014**	=(0,75/0,725+0,4/0,399)/2 =**1,018**
=(0,25/0,275+0,2/0,198)/2 =**0,96**	=(0,25/0,275+0,4/0,403)/2 =**0,951**	=(0,25/0,275+0,4/0,399)/2 =**0,956**

Populaziora egokitutako laginketa-taula portzentaiatan:

A\B	1	2	3	Guztira
1	0,117	0,334	0,287	0,738
2	0,081	0,07	0,112	0,263
Guztira	0,198	0,404	0,399	1,001

Bosgarren iterazioa

Orokortzaileak

=(0,75/0,738+0,2/0,198)/2 =**1,013**	=(0,75/0,738+0,4/0,404)/2 =**1,003**	=(0,75/0,738+0,4/0,399)/2 =**1,009**
=(0,25/0,263+0,2/0,198)/2 =**0,98**	=(0,25/0,263+0,4/0,404)/2 =**0,97**	=(0,25/0,263+0,4/0,399)/2 =**0,977**

Populaziora egokitutako laginketa-taula portzentaiatan:

A\B	1	2	3	Guztira
1	0,119	0,335	0,29	0,744
2	0,079	0,068	0,109	0,256
Guztira	0,198	0,403	0,399	1

Seigarren iterazioa

Orokortzaileak

=(0,75/0,744+0,2/0,198)/2 =**1,009**	=(0,75/0,744+0,4/0,403)/2 =**1**	=(0,75/0,744+0,4/0,399)/2 =**1,005**
=(0,25/0,256+0,2/0,198)/2 =**0,993**	=(0,25/0,256+0,4/0,403)/2 =**0,985**	=(0,25/0,256+0,4/0,399)/2 =**0,99**

Populaziora egokitutako laginketa-taula portzentaiatan:

A\B	1	2	3	TOTAL
1	0,12	0,335	0,291	0,746
2	0,078	0,067	0,108	0,253
TOTAL	0,198	0,402	0,399	0,999

Zazpigarren iterazioa

Orokortzaileak

=(0,75/0,746+0,2/0,198)/2 =**1,008**	=(0,75/0,746+0,4/0,402)/2 =**1**	=(0,75/0,746+0,4/0,399)/2 =**1,004**
=(0,25/0,253+0,2/0,198)/2 =**0,999**	=(0,25/0,253+0,4/0,402)/2 =**0,992**	=(0,25/0,253+0,4/0,398)/2 =**0,995**

Populaziora egokitutako laginketa-taula portzentaiatan:

A\B	1	2	3	TOTAL
1	0,121	0,335	0,292	0,748
2	0,078	0,066	0,107	0,251
TOTAL	0,199	0,401	0,399	0,999

Antolamendua jada egokituta gelditu da, *0,002ko* perdoiarekin, zazpi iteraziotan.

Geruzapeneko aldagaien bariante bakarreko antolamendura egokitu ondoren, zenbatetsitako ale-kopuruaren taula honakoa da:

A\B	1	2	3	Guztira
1	2	7	6	15
2	2	1	2	5
Guztira	4	8	8	20

Gure simulatutako adibideari dagokion taula haztatu berdina lortzen da metodo bat zein bestea erabiliz.

RAS deritzan *egokitzapen-metodoaren* helburua iterazio bakoitzeko pausoetan pisu--aldaketa gutxiago egitea da. Horretarako pisuen aldaketa-faktorea elementu bakoitzari dagokion modalitateetara egokitzean lortu diren faktoreen batezbestekoa bezala hartzen da. Horrela, askatu beharreko sistema bakarra hartzen du, $a+b$ ezezagun eta ekuazioko
a, ondorioz lortzen diren pisuek honako forma hartzen dutelarik:

$$w(k,m) = w(k,m-1) * \sum_{h,h'} \left(X_h(k) * c_h\right) * \left(X_{h'}(k) * c_{h'}\right)$$

$$1 \leq h \leq a \text{ eta } a+1 \leq h' \leq a+b+1 \quad \text{delarik}$$

c_h eta $c_{h'}$ faktoreak proposatutako murrizketa-sistema askatuz lortzen dira. Elementu bakoitzaren azken pisua, berriz, hasierako pisua c_h eta $c_{h'}$ faktoreen batezbestekoarekin orokortuz.

Adibideak

Helburu diren bi aldagai hartuko ditugu, egon daitezkeen bi motatakoak: koalitatiboa eta koantitatiboa.

Y_1 populazioaren elementu *ez-langabetuen* kopurua da.

Y_2 populazioaren *hileko fakturazioaren guztizkoa* da.

Aldagai osagarriak bi dira:

- A: *Sexua.* 2 kategoria ditu, *a=2*.
- B: Heziketa-maila: Oinarrizkoak, Bigarren mailakoak ala Unibertsitate-mailakoak. 3 kategoria ditu, *b=3*.

N=20 populazioaren tamaina eta *n=12 laginaren tamaina* dira.

Probabilitate-laginketa hartuko dugu beti, eta bertako elementu guztiek hautespen--probabilitate berdinak izango dituzte, 12/20.

Aldagai osagarriak

Populazio osoaren datuak ($N_{hh'}$) *(1) taula*

$\sum_{k=1}^{N} X_{hh'}(k)$	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	?	?	?	9
Emakumea	?	?	?	11
GUZTIRA	7	9	4	20

Populazio-laginaren datuak ($n_{hh'}$) *(2) taula*

$\sum_{k=1}^{n} X_{hh'}(k)$	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	1	1	1	3
Emakumea	3	3	3	9
GUZTIRA	4	4	4	12

Helburu den Y_1 aldagaia.

Populazio-laginaren datuak. $\left(Y_{1,hh'}\right)$ *(3) taula*

Populazio-geruza bakoitzeko *ez-langabetuen* laginketa-taula.

$\sum_{k=1}^{n} Y_1(k)$	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	0	1	0	1
Emakumea	3	2	0	5
GUZTIRA	3	3	0	6

Helburu den Y_2 aldagaia.

Populazio-laginaren datuak. $(Y_{2,hh'})$ *(4) taula*

Hileko fakturazioaren guztizkoa geruzaka.

$\sum_{k=1}^{n} Y_2(k)$	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA
Gizona	75000	250000	275000	600000
Emakumea	120000	220000	250000	590000
GUZTIRA	195000	470000	525000	1190000

Aldagi osagarriei buruzko datuak portzentaiatan aurkeztuko ditugu:

Aldagai osagarriak. Populazio osoaren datuak. $(W_{hh'})$

	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA
Gizona	?	?	?	0,45
Emakumea	?	?	?	0,55
GUZTIRA	0,35	0,45	0,2	1

Aldagai osagarriak. Populazio-laginaren datuak. $(q_{hh'})$

	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA
Gizona	0,0833	0,0833	0,0833	0,25
Emakumea	0,25	0,25	0,25	0,75
GUZTIRA	0,333	0,333	0,333	1

Bi problemak berdin askatzen dira: aldagai koalitatibo eta koantitatiboarekin egokitzapen berdina lortuko dugu. Erabil dezagun Redre lehenbizi eta Raking Usual ondoren.

Raking Usual

Problema Raking metodoaren bidez askatuko dugu, Excel kalkulu-orriak erabiliz.

Iterazioekin hasiko gara, bi egokitzapen-aldagai baino ez daudenez, bakoitzak 2 pauso izango dituela gogoan izanez:

1. Antolamendua populazioaren antolamendu marjinalaren lehen aldagaiaren modalitateetara egokitzen da.

2. Bigarren fasean, laginaren antolamendu haztatuaren bidez antolamendua bigarren ezaugarriaren modalitateetara egokitzen da, eta horrela iterazioa osatu egingo da.

Lehenbizi *laginaren portzentaia-taula* hartuko dugu $(q_{hh'})$:

	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	0,0833	0,0833	0,0833	0,25
Emakumea	0,25	0,25	0,25	0,75
GUZTIRA	0,333	0,333	0,333	1

Aleei ematen zaien lehen pisua *Horvitz-Thompson*ena da, N/n alegia. Pisu horiek haztatzean lortzen den taulari *taula zabaldua* deitzen zaio $(q_{hh',0})$:

	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	0,138	0,138	0,138	0,415
Emakumea	0,417	0,417	0,417	1,25
GUZTIRA	0,555	0,555	0,555	1,665

Lehen iterazioko lehen pausoa: *ikasketak* osagaiaren antolamenduaren egokitzapena bariante bakarreko antolamenduarekiko

	IKASKETAK			$(q_{hh',0})$
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	0,138	0,138	0,138	0,415
Emakumea	0,417	0,417	0,417	1,25
GUZTIRA	0,555	0,555	0,555	1,665
	0,35	0,45	0,2	
Orokortzaileak	0,631	0,811	0,36	

Laginketa-taula aurkitutako orokortzaileen bidez haztatu eta *bigarren pausora* pasako gara: *sexua* aldagaiaren bariante bakarreko antolamendua egokitzera.

	IKASKETAK					
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA		orokortzaileak
Gizona	0,087	0,112	0,05	0,249	0,45	1,807
Emakumea	0,263	0,338	0,15	0,751	0,55	0,732
GUZTIRA	0,35	0,45	0,2	1		

Azken taula honakoa da:

	IKASKETAK			$(q''_{hh',0} = q_{hh',1})$
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA
Gizona	0,157	0,202	0,09	0,449
Emakumea	0,193	0,247	0,11	0,55
GUZTIRA	0,35	0,449	0,2	0,999

Iterazio bakarraren ondoren, Ikasketak eta Sexua aldagaien bariante bakarreko antolamendua egokituta gelditu da.

Haztatutako lagin-kontingentziaren taula honakoa da:

	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA
Gizona	2	2	1	5
Emakumea	2	3	1	6
GUZTIRA	4	5	2	11

Halaber, lortu dugun populazio-kontingentziaren taula haztatua honakoa da:

	IKASKETAK			
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate--mailakoak	GUZTIRA
Gizona	3	4	2	9
Emakumea	4	5	2	11
GUZTIRA	7	9	4	20

Lagin-ale bakoitzaren pisu-taula hau lortzen da:

	IKASKETAK		
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak
Gizona	3	4	2
Emakumea	1,33333333	1,66666667	0,66666667

Helburu den aldagai bakoitzarentzat gelaskaka zenbatetsitako guztizkoen taulak honakoak dira:

	IKASKETAK			$\left(\tilde{Y}_{1,\,hh'}\right)$
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	0	4	0	4
Emakumea	4	3	0	7
GUZTIRA	4	7	0	11

	IKASKETAK			$\left(\tilde{Y}_{2,\,hh'}\right)$
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	225000	1000000	550000	1775000
Emakumea	160000	366666,67	166666,67	693333,34
GUZTIRA	385000	1366666,67	716666,67	2468333,34

Beraz, honakoa lortzen dugu:

$\hat{Y}_1$ =11 ez-langabetu

$\hat{Y}_2$ =2468333,34 zenbatetsitako fakturazioaren guztizkoa

Halaber, ez-langabetuen batezbesteko zenbatetsia 0,55 dela daukagu, eta analogikoki, fakturazioren batezbesteko zenbatetsia 123416,66 dela.

3 hamarrenekin lan egin dugu, beraz perdoia 0,005 dela uste dugu. Perdoi horrekin, metodoaren konbergentzia lehen iterazioan lortzen dela ikusten dugu.

Redre

Laginaren elementuentzako datu-baseak hartuz eta populazioaren bariante bakarreko antolamenduen arabera, SPADN paketeko REDRE prozeduraren bidez datuak erabiliko dira. Aldagai osagarriei dagozkien ondoriozko taulak honakoak dira:

Laginketa-taula egokitua:

	IKASKETAK			$\left(\tilde{n}_{hh'}\right)$
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	2	3	1	6
Emakumea	2	3	1	6
GUZTIRA	4	6	2	12

Laginketa-taula egokitua = aurreko taula * 20/12

	IKASKETAK			$\left(\tilde{N}_{hh'}\right)$
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	4	5	2	11
Emakumea	3	4	2	9
GUZTIRA	7	9	4	20

Lagin-elementu bakoitzaren pisu-taula, dagokion gelaskaren arabera:

	IKASKETAK		
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak
Gizona	4	5	2
Emakumea	1	1,333	0,667

Horrela populazioaren antolamendu marjinalera egokitutako laginaren antolamendu haztatua lortu dugu jada. Antolamendua ez da zehatza, 0,001ko perdoia erabili baitugu, eta gehienez 10 iterazio.

Populaziora egokitutako *ez-langabetua* aldagaiaren taula haztatua:

	IKASKETAK			$\left(\tilde{Y}_{1,hh'}\right)$
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	0	5	0	5
Emakumea	3	3	0	6
GUZTIRA	3	8	0	11

Populaziora egokitutako *fakturazioaren guztizkoa* aldagaiaren guztizkoen taula:

	IKASKETAK			$(\tilde{Y}_{2,hh'})$
SEXUA	Oinarrizkoak	Bigarren mailakoak	Unibertsitate- -mailakoak	GUZTIRA
Gizona	300000	1250000	550000	2100000
Emakumea	120000	293260	166750	580010
GUZTIRA	420000	1543260	716750	2680010

Helburu diren bi aldagaientzat ondoriozko honako zenbatespenak lortzen dira:

$\hat{Y}_1$=11 ez-langabetuak

$\hat{Y}_2$=2680010 zenbatetsitako fakturazioaren guztizkoa

Halaber, ez-langabetuen batezbesteko zenbatetsia 0,55 da, eta analogikoki, fakturazioren batezbesteko zenbatetsia 134000,5.

Prozedura informatikoak

Raking Usual

Raking-a erabiltzeko ezagutzen diren software informatikoak honakoak dira:

CALMAR, *Institut National de la Statistique et des Etudes Economiques,* INSEErena. S.A.S.n programatutako makroa da, eta Raking-a aldagai koalitatibo zein koantitatiboekin erabiltzeko balio du. Software honen bidez laginak batera egokitu daitezke indibiduo-maila desberdinetara.

BASCULA, *Netherlands Central Bureau of Statistics*, NCBSrena. Software honek Raking arrunta erabiltzen du, beraz aldagai *koalitatiboetarako* soilik balio du.

Saiakuntza moduan *Euskal Estatistika Erakundeak,* EUSTATek, FORTRANen garatutako programa. Helburua *aldagai koalitatibo nahiz koantitatiboen* laginak egokitzea da, *datuak eguneratzea* eta *indibiduo-maila ezberdinetara egokitzea* ere bai.

Redre

Kapitulu honetan aztertu dugun RAKINGaren aldaera erabiltzeko SPADN programako prozedura informatiko bat dago, REDRE izenekoa. SPADN WINDOWS pakete informatikoa da, eta bere programetako bat REDRE da. Programa honen bidez laginak egokitu daitezke, baina aldagai *koalitatiboak* dituztenak soilik.

Aplikazio-egoera egokienak

Aldagai osagarrien *bariante bakarreko populazioaren antolamendua* ezagutzea beharrezkoa da.

Iterazio-prozedurek gelaska asko dituzten geruzapenean sortzen den arazo bat saihesten dute, gelaska txikiena. Horrelakoetan, gelaska askotako ale kopurua zerotik gertu dago, beraz *orokortzailearen aldakuntza* handia dagokio. Horrez gain, geruzapenean aldaketa txikiak izaten dira, eta gelaska bakoitzean laginaren elementu asko ager daitezke. Gelaska horiei zenbatesle handia ematen zaie, hots *zenbateslea ezegonkorra da*. Ezegonkortasuna saihesteko, bariante anitzeko geruzapena ez da erabiltzen, aldagai osagarrien bariante bakarreko antolamendua baino ez da hartzen, eta egokitzapena kapitulu honetan aipatu diren iterazio-prozeduraren baten bidez egiten da. Horrela zenbatesle egonkorra lortzen da lagin-geruzapenarekiko: gelaska txikien arazoa desagertu egiten da.

4.
kapitulua

Ondorioak eta proposamenak

Koaderno honetan laginen egokitzapen estatistikoaz jardun dugu, bi zenbatesle-multzo handien arabera:

Informazio Osagarri Handieneko metodoak. Gelaskaz gelaska orokortzaileak lortzen dira. Metodo hauek bi egoeratan hartu eta aztertu ditugu: *geruzapen-* eta *post--geruzapen*-egoeretan*.* Informazio osagarria *koantitatiboa* denean, orokortzaileak gelaska bakoitzeko populazio-ale eta lagin-aleen zatidura dira. Informazio osagarri *koantitatiboa* dagoenerako *Arrazoiaren Metodoa* (*Ratioarena*) aurkeztu dugu, zeinetan orokortzaileak lortutako gelasketako aldagai osagarriaren populazioa-guztikoa eta laginaren guztizkoaren zatidura diren.

Egokitzapen-metodoak erabiltzean askotan gelaska hutsen edo oso txikien arazoa sortzen da: gelasken gutxieneko laginketa-tamaina 20-25ekoa da. Arazo hori konpontzeko, gutxieneko tamainara iristen ez diren *gelaskak kolapsatu* egiten dira.

Iterazio-prozedurak, Bariante Bakarreko Antolamendu Osagarriak eskura daudenean soilik erabiltzen direnak. Egokitzapen-metodo hau aldagai osagarri koalitatiboekin baino ez da erabili koaderno honetan. Egokitzapen-metodo gisa Raking Usual deritzana eta bere aldaera bat erabili dira. Software informatikoetan, berriz, honakoak erabili dira: CALMAR, BASCULA eta EUSTATek saiakuntza moduan FORTRANen egindako programa, Raking-a aplikatzeko balio duena, eta REDRE, SPADNen barruan, Raking-aren aldakuntza aplikatzeko.

Gelaska oso txikiak edo oso handiak batera daudenean beste aukera *METODO MISTOA erabiltzea da:* (Ikus 13])

Lehenbizi lagin-elementu *asko* dituzten gelaskak hartzen dira, eta horiekin *geruzapen eta post-geruzapen* metodoak erabiltzen dira. Egokitu ondoren, gelaska horiek taulatik atera egiten dira.

Ondoren *RAKING*aren aldaera, *Raking Bornatua* izenekoa, erabiltzen da gainerako *gelasketan*: egokitzapen-metodo honek gelaska bakoitzeko orokortzaileen aldakuntza bornatzen du hasierako orokortzaileekiko. Prozedura honako da:

- Lagin-elementu gutxi dituzten gelasketan dagokien $\tilde{W}_{hh'}$ orokortzailearen zenbatespena bornatzen da. Orokortzaileak ez du hasierakoarekiko ezberdinegia izan behar. Zenbatespena egin ondoren, gelaskako populazio-elementuen guztizkoaren zenbatespena hartzen da, $\tilde{N}_{hh'} = \tilde{W}_{hh'} \bullet n_{hh'}$ besterik egin gabe. Gelaska egokituak tauletatik ateratzen dira.

- Gainerako behakuntzak RAKINGaren bidez egokitzen dira, hots, bariante bakarreko antolamenduetara egokitzen dira.

Metodo mistoari buruz zenbait gauza kontuan hartu behar dira:

Gelaska handien eta gehiegi aldatzen den haztaketa- edo orokortze-faktoreen definizioa arbitrarioa da.

Arrazoiaren zenbatespena gelaska handietan aplikatu ondoren, zergatik ez dira gelaskak banan-banan egokitu, gainerako gelaskak kolapsatuz?

Raking Bornatuan konbergentziarik ba al dago? Derrigor ezarri behar den baldintza, prozedura gehiegitan ez aplikatzea da.

Kontu horiek guztiak aztertzen eta hobetzen ari dira oraindik. Hori dela eta, oraindik ez dugu metodoa aplikatzeko softwareen berririk.

Ikusi dugunez, bariante anitzeko antolamendu osagarria edukita ere, egokitzapenerako aukera eta aldaera ugari daude.

5.
kapitulua

Bibliografia

[1] Alexander, C.H.

A Class of Methods for Using Person Controls in Household Weighting

Survey Methodology. 1987ko abendua 13. alea 2. zk.

[2] Bethlehem, J.G.& Keller, W.J.

Linear Weighting of Sample Survey Data

Journal of Official Statistics. 1987 3. alea 2. zk.

[3] Bishop, Y. M. M. & Fienberg, S.E.

Incomplete Two-Dimensional Contingency Tables

Biometrics. 1969ko martxoa 119-128. orr.

[4] Cochran W. G.

Técnicas de muestreo. Bigarren argitaraldia 1981. Ed. Continental

[5] Copeland, K. R. , Pertzmeier, F.K. & Hoy, C.E.

An Alternative Method of Controlling Current Population Survey Estimates to Population Counts

Survey Methodology. 1987ko abendua 13. alea 2. zk.

[6] Deming, W.E. & Stephan, F.F.

On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known

Annals of Mathematical Statistics. 1940, 11. zk. 427-444. orr.

[7] Dupont, F.

Alternative Adjustments Where There Are Several Levels of Auxiliary Information

Survey Methodology. 1995eko abendua, 21. alea 2. zk. 125-135. orr.

[8] Fuller, W.A., Loughin, M.M.& Baker, H.D.

Regression weighting in the Presence of Nonresponse

Survey Methodology. 1994ko ekaina 20. alea 1. zk.

[9] Kalton, G.

Compensating for Missing Survey Data

Research Report Series. University of Michigan

[10] Ku, H. H. & Kullback, S.

Loglinear Models in Contingency Table Analysis

The American Statiscian. 1974ko azaroa 28. alea 4. zk.

[11] Little & Rubin

Adjustment Cells

Statistical Analysis with Missing Data. Ed. Wiley 1987

[12] Niyorvenga, T.

Nonparametric Estimation of Response Probabilities and Sampling Theory

Survey Methodology. 20. alea 2. zk.

[13] Oh, H.L. & Scheuren, F.

Modified Raking Ratio Estimation

Survey Methodology. 1987ko abendua, 13. alea 2. zk. 209-219. orr.

[14] Rao, P.S.R.S.

Ratio Estimation with Subsampling the Nonrespondents

Survey Methodology. 1986 12. alea 2. zk.

[15] Särndal, C.E.

A regression Approach to Estimation in the Presence of Nonresponse

Survey Methodology. 1986 12. alea 2. zk.

[16] Sitter, R. R. & Skinner, D.J.

Multi-way Stratification by Linear Programming

Survey Methodology. 1994ko ekaina 20. alea 1. zk.

[17] Stephan, F.F.

An Iterative Method of Adjusting Sample Frequency Tables when Expected Marginal Totals are Known

Annals of Mathematical Statistics. 13. alea, 166-178. orr.