Taula rodona

Els mètodes d'estimació de petites àrees a l'estadística oficial de Catalunya: entre la recerca i l'aplicació.

Àlex Costa, *Institut d'Estadística de Catalunya*
Maribel Garcia, *Institut d'Estadística de Catalunya*
Albert Satorra, *Universitat Pompeu Fabra*
Eva Ventura, *Universitat Pompeu Fabra*

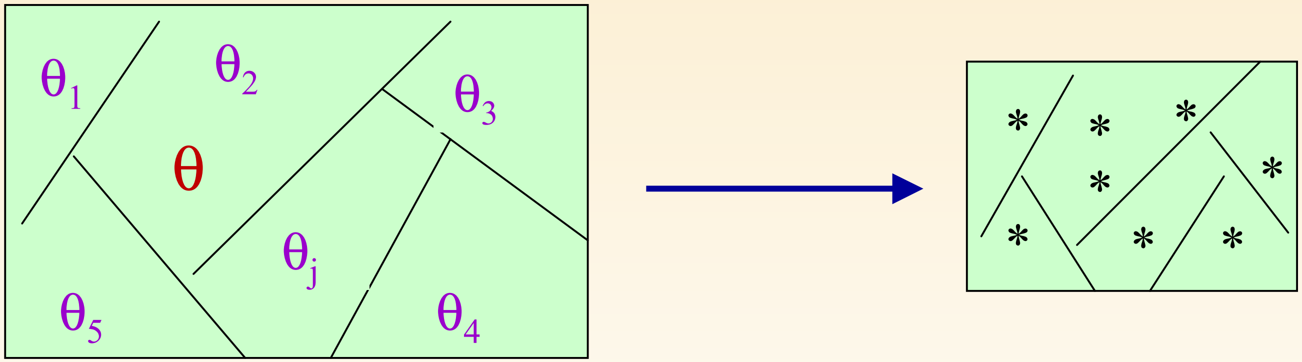
Dia de l'estadística *22 de juny de 2006*

idescat

Fonaments

Estimació de θ_1, ..., θ_j, ..., θ_M (estimació d'àrea petita) on θ_j és el paràmetre de l'àrea *j*.

Població Mostra



La mida mostral és insuficient per a l'estimació de θ_j

idescat

Fonaments

- Si un vol augmentar l'exactitud de l'estimació en un àrea específica
  - Pot augmentar la mida de mostra i utilitzar estimació directa
  - O pot recórrer a la teoria estadística per millorar la qualitat dels estimadors. Per exemple, pot utilitzar estimadors compostos
    - que combinen la informació de l'àrea d'interès amb la informació que prové d'àrees veïnes
    - o que utilitzen informació que prové de diverses fonts
    - o que utilitzen informació de diverses fonts i diverses àrees.

Estimadors compostos sense informació auxiliar

- Sigui $\hat{\theta}_j \sim N\left(\theta_j, \sigma_j^2\right), \quad j = 1,2,\ldots,J$ l'estimador directe de l'àrea j i $\hat{\theta}_* \sim N\left(\theta_*, \sigma_*^2\right)$ l'estimador indirecte (directe d'àrea total).

- Dos supòsits addicionals
  - si $\theta_j \sim N(\theta_*, b_j^2)$
  - i $\gamma_j = 0$
- aleshores obtenim

$$\boxed{\tilde{\theta}_{\mathrm{j}} = \pi_j \hat{\theta}_* + \left(1 - \pi_j\right) \hat{\theta}_j} \quad \text{amb} \quad \boxed{\pi_j = \frac{\sigma_j^2}{b_j^2 + \sigma_j^2}}$$

- Aquest és l'estimador compost òptim o teòric, amb un biaix, $b_j = E(\theta_j - \theta^*)$ i una variància, $\sigma^2_j = \text{var}(\hat{\theta}_j)$ coneguts

Dia de l'estadística 22 de juny de 2006

idescat

Estimadors compostos sense informació auxiliar

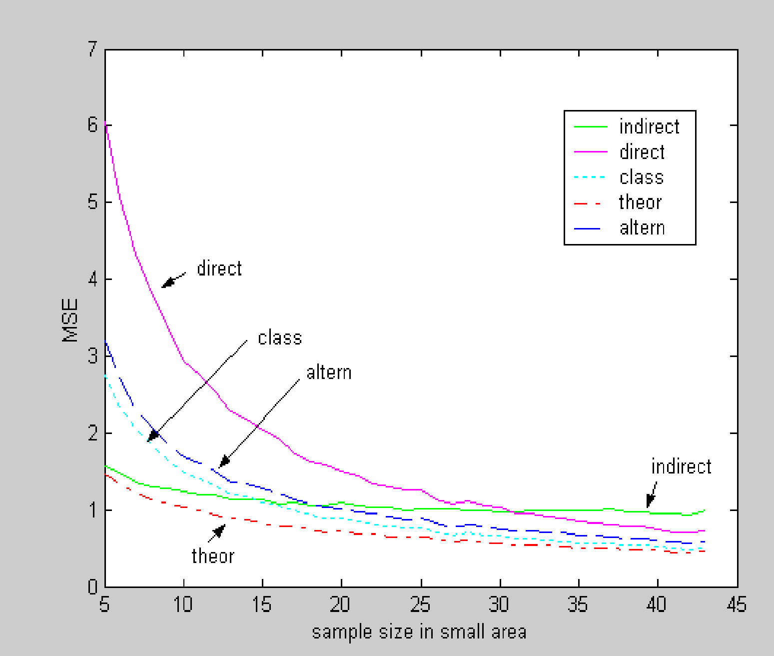
Estimador directe	$\hat{\theta}_j \sim \mathrm{N}(\theta_j, \sigma_j^2/n_j)$ j=1,2,...J (mitjana àrea j) $\theta_j \sim \mathrm{N}(\theta^*, b_j^2)$
Estimador indirecte	$\hat{\theta}_* \sim \mathrm{N}(\theta_*, \sigma_*^2)$ (mitjana area total)
Estimador compost	$\tilde{\theta}_j = \pi_j \hat{\theta}_* + \left(1-\pi_j\right)\hat{\theta}_j$
Compost teòric	$\pi_j = \dfrac{\sigma_j^2/n_j}{\sigma_j^2/n_j + b_j^2}$
Compost clàssic	$\pi_j = \dfrac{\bar{s}^2/n_j}{\bar{s}^2/n_j + b^2}$ amb $\bar{s}^2 = \dfrac{\sum_{j=1}^{J}(n_j-1)s_j^2}{(n-J)}$ i $b^2 = \dfrac{1}{J}\sum_{j=1}^{J}\left(\hat{\theta}_j - \hat{\theta}_*\right)^2$
Compost alternatiu	$\hat{\pi}_j = \dfrac{s_j^2/n_j}{\left(\hat{\theta}_j - \hat{\theta}_*\right)^2}$

idescat

Propietats dels estimadors

- **"An empirical evaluation of small area estimators"**, Alex Costa, Albert Satorra i Eva Ventura, *SORT (Statistics and Operations Research Transactions),* VOL. 27, 1, 113-135, 2003.
  - Realitzem un Monte Carlo en el context de poblacions artificials i reals per estudiar la distribució de l'EQM dels 5 estimadors anteriors.
    - Mida de mostra
    - Correlació intra-classe
    - Grau d'homogeneïtat de la variància de les àrees

Propietats dels estimadors



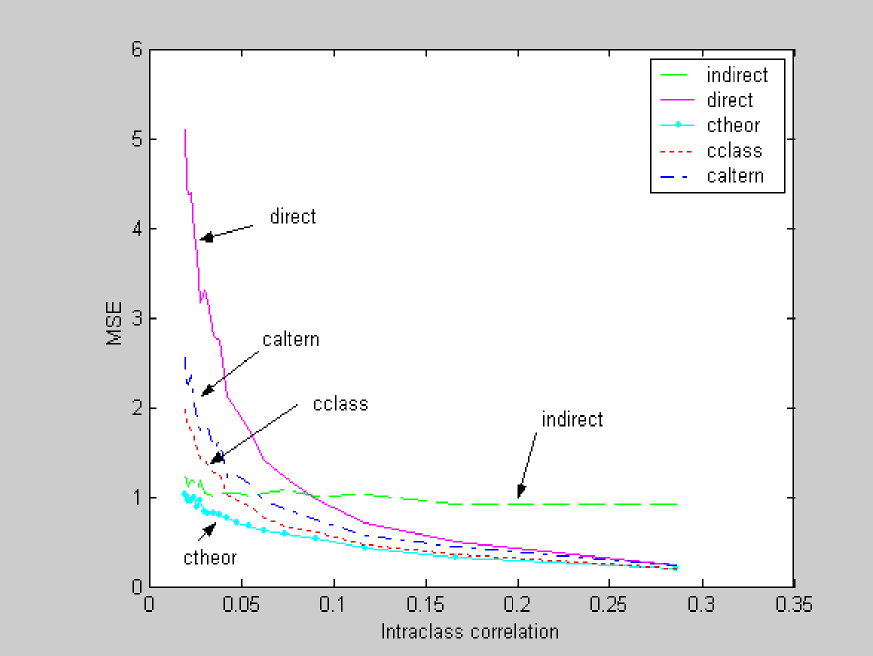


Figura 1: EQM dels estimadors com a funció de la mida mostral amb idèntica variació dins de l'àrea.

Figura 4: EQM dels estimadors com a funció de la correlació intra classe

Propietats dels estimadors

- Amb variàncies idèntiques en cada àrea els EQM dels estimadors convergeixen
  - segons augmentem la mida de mostra
  - Segons augmentem la correlació intra-classe
- L'indirecte pot ser un bon estimador només si la mida de mostra és molt petita. En aquest cas el directe és clarament el pitjor estimador.
- En la població real (*Censo laboral de empresas afiliadas a la Seguridad Social*) l'estimador compost (l'alternatiu) obté tanta precisió com un estimador directe amb una mida de mostra quatre vegades més gran.

Propietats dels estimadors

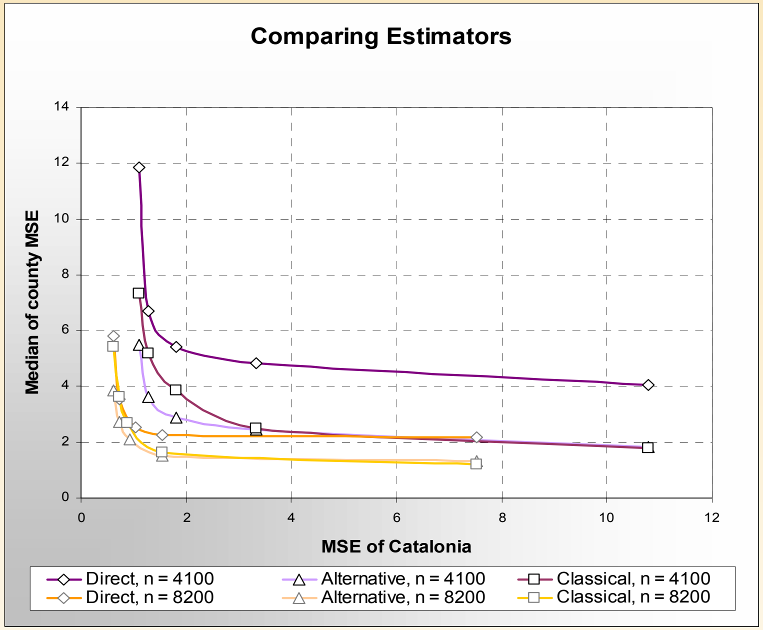
- "Using composite estimators to improve both domain and total area estimation", Alex Costa, Albert Satorra, Eva Ventura. *SORT (Statistics and Operations Research Transactions*), VOL. 28, 1, 69-86, 2004.

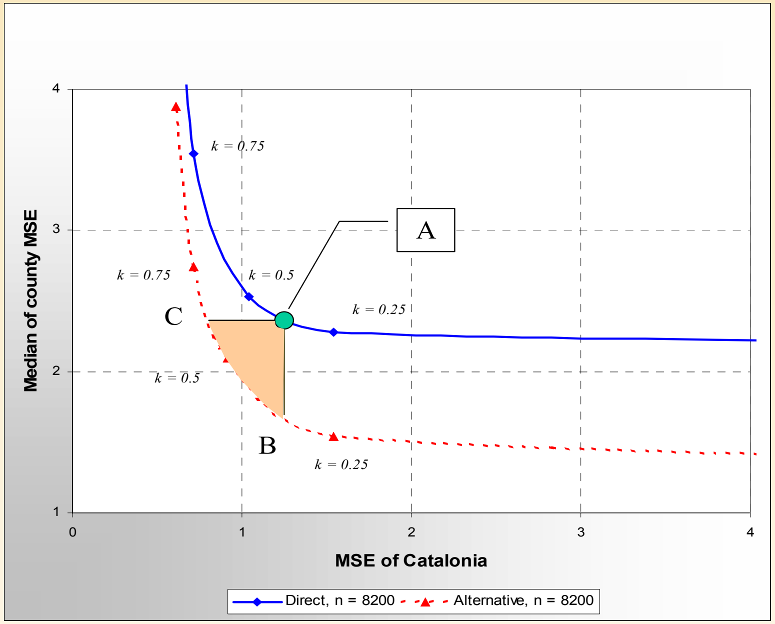
Proporcional	$\sum_{j=1}^{J} n_j = n$	$\frac{n_j}{n} = \frac{N_j}{N}$	for $j = 1, 2, \ldots, J$
Fixa	$\sum_{j=1}^{n} n_j = n$	$n_j = \frac{n}{J}$	for $j = 1, 2, \ldots, J$
Mixta	$\sum_{j=1}^{J} n_j = n$	$n_j = k \frac{N_j}{N} n + (1-k) \frac{n}{J}$	for $j = 1, 2, \ldots, J$

$k = \{0, 1/4, 1/2, 3/4, 1\}$

$n = \{2.050, 4.100, 8.200, 16.400\}$

Propietats dels estimadors





L'ús dels estimadors de petita àrea, en combinació amb un disseny mixt, ens permet obtenir millors estimadors tant per a les petites àrees com per a l'àrea total.

n= 8200, k = 35% aprox.

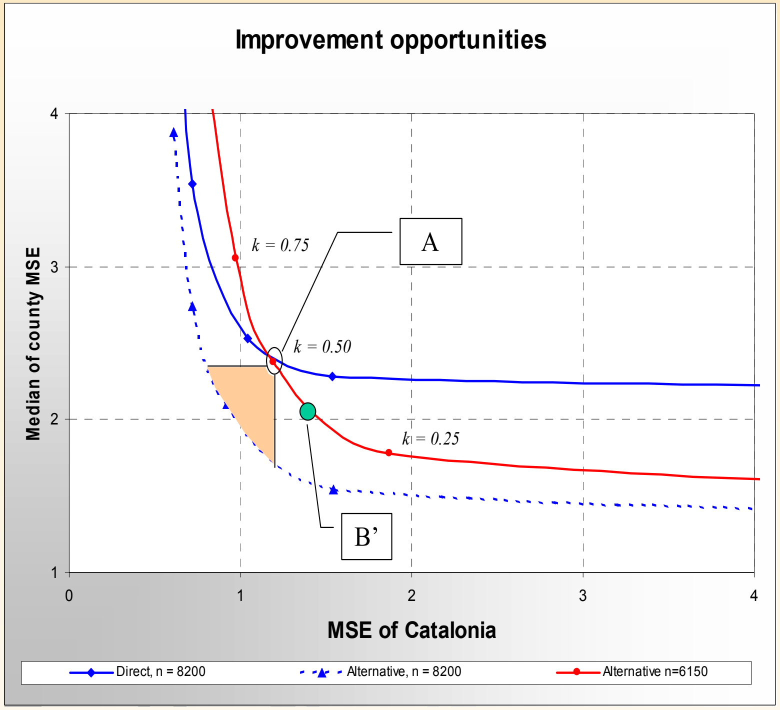
Amb l'estimador compost, millorem l'estimació.

També podem variar el disseny de la mostra i fer-lo més proporcional.

Dia de l'estadística 22 de juny de 2006

idescat

Propietats dels estimadors



- La figura ens mostra que podem fer servir l'estimador compost per reduir la mida total de la mostra, i per tant el seu cost. La combinació de EQM del punt A pot assolir-se de dues maneres alternatives.
  - O bé amb l'estimador directe i una mostra de 8200 observacions,
  - o bé amb el compost alternatiu i una mostra de 6150 (reducció del 25%), i un disseny més proporcional.