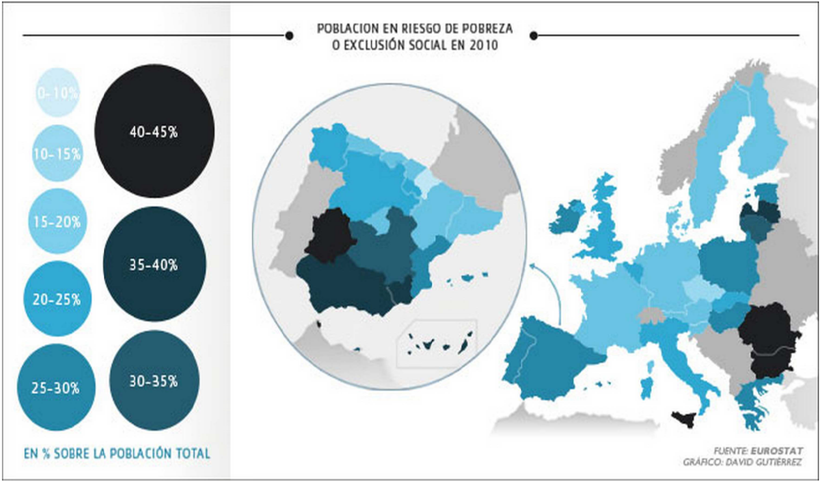
Estimación en áreas pequeñas: aplicación a la cartografía de la pobreza

Isabel Molina

Departamento de Estadística
Universidad Carlos III de Madrid

POBLACIÓN EN RIESGO DE POBREZA

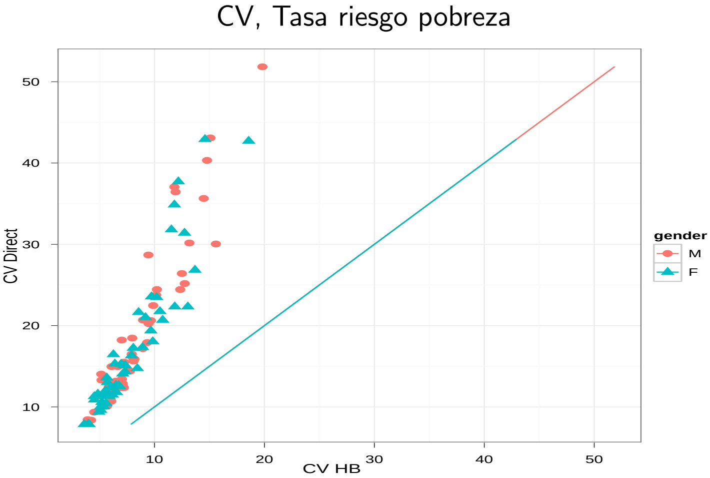


EJEMPLO: POBREZA EN ESPAÑA

- Datos: Encuesta de Condiciones de Vida, 2006.
- Tamaño muestral: $n = 34,389$ de $N = 43,162,384$.
- Parámetros: Tasas riesgo de pobreza para 52 provincias por género.
- Umbral pobreza $z = 0.6 \times$ Mediana(renta disp. equiv.): En 2006, $z = 6,557$ euros$\rightarrow$ approx. **20 %** en riesgo.

Provincia	Género	n_d	En riesgo	$\hat{CV}$ Dir.	$\hat{CV}$ EB	$\hat{CV}$ HB
Soria	M	**17**	**6**	**51.87**	**16.56**	**19.82**
Tarragona	V	129	**18**	**24.44**	**14.88**	**12.35**
Córdoba	M	230	73	13.05	6.24	6.93
Badajoz	V	472	175	8.38	3.48	4.24
Barcelona	M	1483	191	9.38	6.51	4.52

EJEMPLO: POBREZA EN ESPAÑA



INFERENCIA BAJO EL MODELO/DISEÑO

ELEMENTO	BAJO EL MODELO	BAJO EL DISEÑO
Población	$Y \sim P_\theta$	$U = \{1, \ldots, N\}$, $\mathcal{Y} = \{Y_1, \ldots, Y_N\}$
Muestra	$\mathbf{y} = (Y_1, \ldots, Y_n)$ Y_i i.i.d. como Y	$s = (i_1, \ldots, i_n) \subset U$, $\mathbf{y}_s = (Y_{i_1}, \ldots, Y_{i_n})$
Distrib. prob.	$\mathbf{P}_\theta(\mathbf{y})$	$P_\pi(s)$
Parámetro	θ (i.e., $\theta = E(Y)$)	$\delta = h(Y_1, \ldots, Y_N)$ (i.e., $\delta = \bar{Y}$)
Estimador	$\hat{\theta}(\mathbf{y})$	$\hat{\delta}(s)$

INFERENCIA BAJO EL DISEÑO

- U población **finita** de tamaño N.
- $Y_1, \ldots, Y_N$ valores de la variable de interés para los indiv. (fijos).
- **Parámetro/s** objetivo:

$$\delta = h(Y_1, \ldots, Y_N).$$

- Ejemplo: **media** poblacional

$$\bar{Y} = \frac{1}{N} \sum_{j=1}^{N} Y_j.$$

- s **muestra** de tamaño n extraída de la población U.
- $r = U - s$ **no-muestra** (tamaño $N - n$).

ESTIMADOR DE HORVITZ-THOMPSON

- π_j probabilidad de inclusión de unidad j en la muestra.
- $w_j = 1/\pi_j$ peso del muestreo (factor de elevación) de unidad j.
- Estimador **Horvitz-Thompson (HT)** de la media:

$$\hat{\bar{Y}} = \frac{1}{N} \sum_{j \in s} \frac{Y_j}{\pi_j} = \frac{1}{N} \sum_{j \in s} w_j y_j .$$

- Varianza bajo el diseño:

$$V_\pi(\hat{\bar{Y}}) = \frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} (\pi_{j,k} - \pi_j \pi_k) \frac{Y_j}{\pi_j} \frac{Y_k}{\pi_k},$$

$\pi_{j,k}$ probabilidades de inclusión de unidades j y k.

VARIANZA BAJO EL DISEÑO

- Estimador de la varianza insesgado bajo el diseño:

$$\hat{V}_\pi(\hat{\bar{Y}}) = \frac{1}{N^2} \sum_{j \in s} \sum_{k \in s} \frac{\pi_{j,k} - \pi_j \pi_k}{\pi_{j,k}} \frac{Y_j}{\pi_j} \frac{Y_k}{\pi_k},$$

✓ *Särndal, Swensson and Wretman (1992), ecuación (5.8.5)*

- Bajo la aproximación $\pi_{j,k} \cong \pi_j \pi_k$, $j \neq k$,

$$\hat{V}_\pi(\hat{\bar{Y}}) \cong \frac{1}{N^2} \sum_{j \in s} \left(\frac{1 - \pi_j}{\pi_j^2} \right) Y_j^2 = \sum_{j \in s} w_j (w_j - 1) Y_j^2.$$

OTROS PARÁMETROS: INDICADORES FGT

- E_j poder adquisitivo del indiv. j; e.g., renta disponible equivalente.
- z umbral de pobreza: En la UE, se usa

$$z = 0.6 \times \text{Mediana}(E_j).$$

- Familia de indicadores de pobreza FGT:

$$F_\alpha = \frac{1}{N} \sum_{j=1}^{N} \left(\frac{z - E_j}{z} \right)^\alpha I(E_j < z), \quad \alpha \geq 0.$$

- $\alpha = 0 \Rightarrow$ Incidencia de pobreza
- $\alpha = 1 \Rightarrow$ Brecha de pobreza

✓ *Foster, Greer & Thornbecke (1984), Econometrica*

ESTIMADOR HORVITZ-THOMPSON

- Indicador FGT como una media:

$$F_\alpha = \frac{1}{N}\sum_{j=1}^{N} F_{\alpha j}, \quad F_{\alpha j} = \left(\frac{z - E_j}{z}\right)^{\alpha} I(E_j < z).$$

- Estimador HT de F_α:

$$\hat{F}_\alpha = \frac{1}{N}\sum_{j\in s} w_j F_{\alpha j}.$$

- Estimador de varianza:

$$\hat{V}_\pi(\hat{F}_\alpha) = \frac{1}{N^2}\sum_{j\in s} w_j(w_j - 1)F_{\alpha j}^2.$$

ESTIMACIÓN EN DOMINIOS/ÁREAS

- U particionada en D dominios $U_1, \ldots, U_D$ de tamaños $N_1, \ldots, N_D$.
- s_d muestra de tamaño n_d de U_d.
- Tamaño muestral total: $n = \sum_{d=1}^{D} n_d$.
- $r_d = U_d - s_d$ no-muestra, de tamaño $N_d - n_d$.
- Y_{di} valor variable objetivo para ind. i-ésimo del dominio d.

ESTIMADORES DIRECTOS

- Parámetro:
$$\delta_d = h_d(Y_{d1}, \dots, Y_{dN_d}).$$
- Ejemplo: media del dominio d,
$$\bar{Y}_d = \frac{1}{N_d} \sum_{i=1}^{N_d} Y_{di}.$$
- Estimador **directo**: Se basa en los datos **dentro** del área; p.ej.,
$$\hat{\bar{Y}}_d^{DIR} = \frac{1}{N_d} \sum_{i \in s_d} w_{di} y_{di}.$$
- Estimador de la varianza: Bajo $\pi_{di,dj} \cong \pi_{di}\pi_{dj}$, $i \neq j$,
$$\hat{V}_\pi(\hat{\bar{Y}}_d^{DIR}) = \frac{1}{N_d} \sum_{i \in s_d} w_{di}(w_{di} - 1) Y_{di}^2.$$

ESTIMADORES INDIRECTOS

- **Toman prestada información** de las demás áreas asumiendo ciertas hipótesis de **homogeneidad** entre las áreas (**modelo** que usa información auxiliar, con parámetros **comunes** para todas las áreas).

MODELO FAY-HERRIOT (FH)

(i) Modelo que enlaza las áreas:

$$\delta_d = \mathbf{x}_d' \boldsymbol{\beta} + u_d, \;\; u_d \overset{iid}{\sim} (0, \sigma_u^2), \quad d = 1, \ldots, D$$

σ_u^2 desconocida

(ii) Modelo muestral:

$$\hat{\delta}_d^{DIR} = \delta_d + e_d, \;\; e_d \overset{ind}{\sim} (0, \psi_d), \quad d = 1, \ldots, D$$

u_d y e_d indep., ψ_d conocidas $\forall d$

- u_d efectos del área: representan la heterogeneidad entre áreas no explicada por las covariables.
- e_d error del estimador directo, $\psi_d = V_\pi(\hat{\delta}_d^{DIR} | \delta_d)$ varianza.

✓ *Fay & Herriot (1979), JASA*

MEJOR PREDICTOR LINEAL INSESGADO (BLUP)

- Modelo combinado: modelo lineal mixto con efectos aleatorios

$$\hat{\delta}_d^{DIR} = \mathbf{x}_d'\boldsymbol{\beta} + u_d + e_d, \quad d = 1, \dots, D.$$

- BLUP de $\delta_d = \mathbf{x}_d'\boldsymbol{\beta} + u_d$: minimiza el ECM entre estimadores lineales e insesgados.
- Se obtiene ajustando el modelo:

$$\tilde{\delta}_d^{BLUP} = \mathbf{x}_d'\tilde{\boldsymbol{\beta}} + \tilde{u}_d,$$

- $\tilde{\boldsymbol{\beta}}$ estimador de MMCC ponderados de $\boldsymbol{\beta}$.
- $\tilde{u}_d$ BLUP de u_d,

$$\tilde{u}_d = \tilde{u}_d(\sigma_u^2) = \gamma_d(\hat{\delta}_d^{DIR} - \mathbf{x}_d'\tilde{\boldsymbol{\beta}}), \quad \gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \psi_d}$$

BUENA PROPIEDAD DEL BLUP

- Estimador compuesto entre directo y sintético de regresión:

$$\tilde{\delta}_d^{BLUP} = \gamma_d \, \hat{\delta}_d^{DIR} + (1-\gamma_d)\mathbf{x}_d'\tilde{\boldsymbol{\beta}}, \quad \gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \psi_d}.$$

- Directo fiable ($\downarrow \psi_d$) o heterogeneidad no bien explicada por $\mathbf{x}_d$ ($\uparrow \sigma_u^2$): $\tilde{\delta}_d^{BLUP} \longrightarrow \hat{\delta}_d^{DIR}$.
- Directo no fiable o $\mathbf{x}_d$ explica bien la heterogeneidad: $\tilde{\delta}_d^{BLUP} \longrightarrow \mathbf{x}_d'\tilde{\boldsymbol{\beta}}$.

BLUP EMPÍRICO (EBLUP)

- $\tilde{\delta}_d^{BLUP}$ depende de σ_u^2, desconocida:

$$\tilde{\delta}_d^{BLUP} = \tilde{\delta}_d^{BLUP}(\sigma_{\mathbf{u}}^{\mathbf{2}})$$

- BLUP empírico (EBLUP): $\hat{\sigma}_u^2$ estimador consistente de σ_u^2,

$$\hat{\delta}_d^{EBLUP} = \tilde{\delta}_d^{BLUP}(\hat{\sigma}_{\mathbf{u}}^{\mathbf{2}}), \quad d = 1, \ldots, D.$$

- EBLUP insesgado para estimadores habituales de σ_u^2.
- Buenas aproximaciones del MSE del EBLUP bajo el modelo, con normalidad.

MODELO DE ERRORES ANIDADOS

- Modelo a nivel de individuo:

$$Y_{di} = \mathbf{x}'_{di}\beta + u_d + e_{di}, \quad i = 1, \dots, N_d, \; d = 1, \dots, D$$
$$u_d \overset{iid}{\sim} N(0, \sigma_u^2), \quad e_{di} \overset{iid}{\sim} N(0, \sigma_e^2)$$

- $\boldsymbol{\theta} = (\beta', \sigma_u^2, \sigma_e^2)'$ vector de parámetros desconocidos.

✓ *Battese, Harter & Fuller (1988), JASA*

MODELO DE ERRORES ANIDADOS

- Media de un área:

$$\bar{Y}_d = N_d^{-1}\left(\sum_{i\in s_d} Y_{di} + \sum_{i\in r_d} Y_{di}\right).$$

- BLUP bajo modelo con errores anidados:

$$\tilde{\bar{Y}}_d^{BLUP} = N_d^{-1}\left(\sum_{i\in s_d} Y_{di} + \sum_{i\in r_d} \tilde{Y}_{di}^{BLUP}\right).$$

- $\tilde{Y}_{di}^{BLUP} = \mathbf{x}_{di}'\tilde{\beta} + \tilde{u}_d$ valores predichos (BLUPs de Y_{di}, $i \in r_d$)
- $\tilde{\beta}$ estimador MMCC ponderados de β.
- $\tilde{u}_d$ BLUP de u_d.

MODELO DE ERRORES ANIDADOS

- BLUP para áreas con $n_d/N_d \approx 0$:

$$\tilde{\bar{Y}}_d^{BLUP} \approx \gamma_d \left\{ \bar{y}_d + (\bar{\mathbf{X}}_d - \bar{\mathbf{x}}_d)'\tilde{\beta} \right\} + (1-\gamma_d)\bar{\mathbf{X}}_d'\tilde{\beta}.$$

- Estimador **compuesto** entre los estimadores "survey regression" y sintético de regresión.
- Peso del estimador "survey regression":

$$\gamma_d = \sigma_u^2/(\sigma_u^2 + \sigma_e^2/n_d).$$

- Toma prestada información a través del estimador sintético de regresión de forma **automática**.

MÉTODO ELL

- Tradicionalmente usado por el Banco Mundial.
- Primer método para estimar parámetros generales.
- Modelo con errores anidados para $Y_{di} = \log(E_{di})$:

$$Y_{di} = \mathbf{x}'_{di}\boldsymbol{\beta} + u_d + e_{di}, \quad i = 1, \ldots, N_d,\ d = 1, \ldots, D,$$

- u_d efectos aleatorios de las u1e (conglomerados).
- Por simplicidad, aquí consideramos que las u1e son las áreas.
- Estimador ELL de $\delta_d = h_d(Y_{d1}, \ldots, Y_{dN_d})$:

$$\hat{\delta}_d^{ELL} = E[\delta_d] \text{ aproximado por bootstrap.}$$

✓ *Elbers, Lanjouw & Lanjouw (2003), Econometrica*

MÉTODO ELL

- Para $\delta_d = \bar{Y}_d$, aprox. igual al estimador sintético de regresión:

$$\hat{\bar{Y}}_d^{ELL} \approx \bar{\mathbf{X}}_d' \boldsymbol{\beta}$$

- Sesgo considerable (bajo el diseño) si las covariables no explican toda la heterogeneidad entre áreas.
- ECM bajo el modelo muy alto $\Rightarrow$ Puede ser mucho peor que los estimadores directos.

✓ *Molina & Rao (2010), CJS*

MÉTODO EB

- $\mathbf{y}_d = (Y_{d1}, \ldots, Y_{dN_d})' = (\mathbf{y}'_{ds}, \mathbf{y}'_{dr})'$
- $\mathbf{y}_{ds}$ parte observada (encuesta), $\mathbf{y}_{dr}$ parte no observada.
- Mejor predictor de $\delta_d = h_d(\mathbf{y}_d)$: Minimiza el ECM,

$$\tilde{\delta}_d^B(\boldsymbol{\theta}) = E_{\mathbf{y}_{dr}}[h_d(\mathbf{y}_d)|\mathbf{y}_{ds}; \boldsymbol{\theta}],$$

✓ *Molina & Rao (2010), CJS*

MÉTODO EB

- Modelo con errores anidados:

$$Y_{di} = \mathbf{x}_{di}'\beta + u_d + e_{di}, \quad i = 1, \ldots, N_d,\ d = 1, \ldots, D,$$

- $Y_{dj} = T(E_{dj})$ transformación biyectiva del poder adquisitivo con distrib. aproximadamente Normal; típicamente, $Y_{di} = \log(E_{di} + c)$, $c > 0$.
- u_d efectos aleatorios de las áreas.

✓ *Molina & Rao (2010), CJS*

MEJOR PREDICTOR: INDICADORES FGT

- Indicadores FGT en función de Y_{di}:

$$F_{\alpha d} = \frac{1}{N_d} \sum_{i=1}^{N_d} \underbrace{\left\{ \frac{z - T^{-1}(Y_{di})}{z} \right\}^{\alpha} I\left\{ T^{-1}(Y_{di}) < z \right\}}_{F_{\alpha,di}} = h_d(\mathbf{y}_d).$$

- Mejor predictor de $F_{\alpha d}$:

$$\tilde{F}_{\alpha d}^{B}(\boldsymbol{\theta}) = E_{\mathbf{y}_{dr}}[F_{\alpha d} | \mathbf{y}_{ds}; \boldsymbol{\theta}] = \frac{1}{N_d} \left\{ \sum_{i \in s_d} F_{\alpha,di} + \sum_{i \in r_d} \tilde{F}_{\alpha,di}^{B}(\boldsymbol{\theta}) \right\}.$$

- Mejor predictor de $F_{\alpha,di}$, $i \in r_d$:

$$\tilde{F}_{\alpha,di}^{B}(\boldsymbol{\theta}) = E(F_{\alpha,di} | \mathbf{y}_{ds}; \boldsymbol{\theta}).$$

✓ *Molina and Rao (2010), CJS*

MEJOR PREDICTOR EMPÍRICO

- $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma}_u^2, \hat{\sigma}_e^2)'$ estimador consistente de $\boldsymbol{\theta}$.
- Mejor predictor **empírico** (EB):

$$\hat{F}_{\alpha d}^{EB} = \tilde{F}_{\alpha d}^{B}(\hat{\boldsymbol{\theta}}).$$

- **Census** EB: Se predice también para la muestra.
- EB se puede calcular analíticamente para $\alpha = 0, 1$, donde $Y_{di} = \log(E_{di} + c)$.
- En general, se puede aproximar mediante **Monte Carlo**.
- ECM bajo el modelo se puede aproximar mediante **bootstrap**.

✓ *Molina & Rao (2010), CJS*

APROXIMACIÓN MONTE CARLO

- Generar L vectores no muestrales $\mathbf{y}_{dr}^{(\ell)}$, $\ell = 1, \ldots, L$ de la distribución condicionada $\mathbf{y}_{dr} | \mathbf{y}_{ds}$.
- Para cada elemento $Y_{di}^{(\ell)}$ de $\mathbf{y}_{dr}^{(\ell)}$, calcular $F_{\alpha,di}^{(\ell)} = h_\alpha(Y_{di}^{(\ell)})$, $\ell = 1, \ldots, L$ y promediar para las L poblaciones Monte Carlo,

$$E_{\mathbf{y}_{dr}}\left[F_{\alpha,di} | \mathbf{y}_{ds}\right] \cong \frac{1}{L} \sum_{\ell=1}^{L} F_{\alpha,di}^{(\ell)}.$$

✓ *Molina & Rao (2010), CJS*

BOOTSTRAP PARAMÉTRICO

(1) Ajuste del modelo: $\hat{\sigma}_u^2$, $\hat{\sigma}_e^2$, $\hat{\boldsymbol{\beta}}$ ✓ ML ✓ REML ✓ MM

(2) Generar efectos aleatorios bootstrap:

$$u_d^* \overset{iid}{\sim} N(0, \hat{\sigma}_u^2), \quad d = 1, \dots, D.$$

(3) Generar, independientemente de $u_1^*, \dots, u_D^*$, errores

$$e_{di}^* \overset{iid}{\sim} N(0, \hat{\sigma}_e^2), \quad i = 1, \dots, N_d, \; d = 1, \dots, D.$$

(4) Generar una población bootstrap a partir del modelo

$$Y_{di}^* = \mathbf{x}_{di}'\hat{\boldsymbol{\beta}} + u_d^* + e_{di}^*, \quad i = 1, \dots, N_d, \; d = 1, \dots, D.$$

✓ *González-Manteiga, Lombardía, Molina, Morales and Santamaría (2008), JSCS.*

BOOTSTRAP PARAMÉTRICO

(5) Calcular los parámetros objetivo para la población bootstrap

$$F_{\alpha d}^{*} = \frac{1}{N_d} \sum_{i=1}^{N_d} h_\alpha(Y_{di}^{*}), \quad d = 1, \ldots, D.$$

(6) Tomar la parte de la muestra y, con ella, ajustar el modelo: $\hat{\sigma}_u^{2*}$, $\hat{\sigma}_e^{2*}$, $\hat{\beta}^*$

(7) Calcular el estimador EB bootstrap: $\hat{F}_{\alpha d}^{EB*}$

(8) Repetir (2)–(7) B veces: $(F_{\alpha d}^{*(b)}, \hat{F}_{\alpha d}^{EB*(b)})$, $b = 1, \ldots, B$.

(9) ECM bootstrap:

$$\mathrm{mse}_B(\hat{F}_{\alpha d}^{EB}) = B^{-1} \sum_{b=1}^{B} \left(\hat{F}_{\alpha d}^{EB*(b)} - F_{\alpha d}^{*(b)} \right)^2$$

✓ *González-Manteiga, Lombardía, Molina, Morales and Santamaría (2008), JSCS.*

EXPERIMENTO BAJO DEL MODELO

- Simular $L = 10^4$ poblaciones a partir del **modelo con errores anidados**.
- Para cada población $\ell = 1, \ldots, L$, calcular los verdaderos valores de los indicadores FGT para cada área, $F_{\alpha d}^{(\ell)}$, para $\alpha = 0, 1$ y $d = 1, \ldots, D$.
- Extraer la muestra de cada población (MAS dentro de cada dominio) y calcular estimadores EB, directos y ELL.
- Aproximar los **verdaderos ECMs** de cada estimador $\hat{F}_{\alpha d}$:

$$\mathrm{MSE}(\hat{F}_{\alpha d}) = \frac{1}{L}\sum_{\ell=1}^{L}\left(\hat{F}_{\alpha d}^{(\ell)} - F_{\alpha d}^{(\ell)}\right)^2,\ \alpha = 0, 1,\ d = 1, \ldots, D.$$

EXPERIMENTO BAJO DEL MODELO

- **Tamaños:**

$$N = 20000, \quad D = 80$$
$$N_d = 250, \quad n_d = 50, \; d = 1, \ldots, D$$

- **Componentes de la varianza:**

$$\sigma_e^2 = (0.5)^2, \quad \sigma_u^2 = (0.15)^2$$

- **Variables auxiliares:** 2 binarias,

$$X_1 \in \{0, 1\}, \quad p_{1d} = 0.3 + 0.5d/80, \quad d = 1, \ldots, D.$$
$$X_2 \in \{0, 1\}, \quad p_{2d} = 0.2, \quad d = 1, \ldots, D.$$

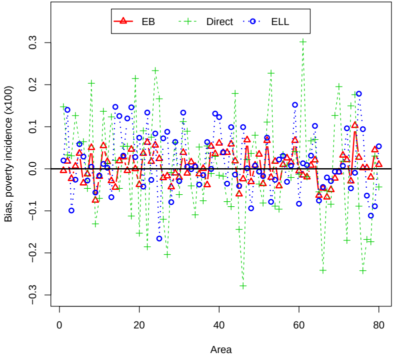
- **Coeficientes:**

$$\boldsymbol{\beta} = (3, 0.03, -0.04)'$$

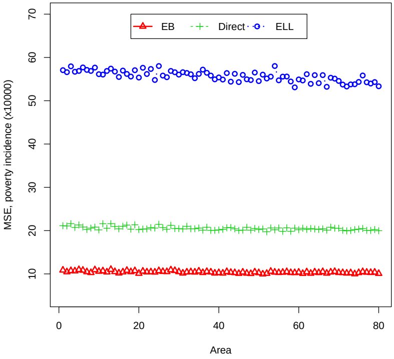
INCIDENCIA DE POBREZA

- EB mucho más eficiente que ELL y Directos.
- ELL peor que estimadores directos!

Bias (%)

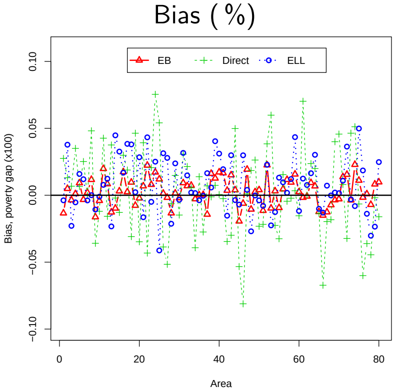


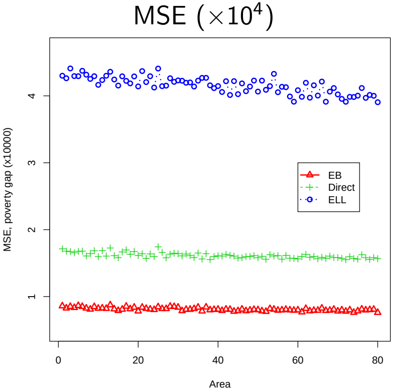
MSE ($\times 10^4$)



BRECHA DE POBREZA

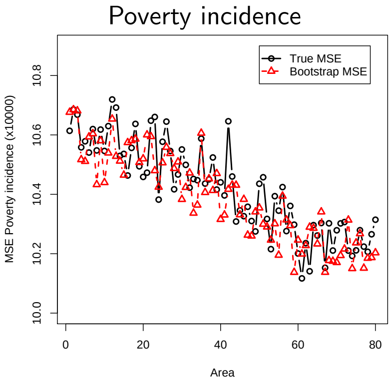
- Conclusiones similares.

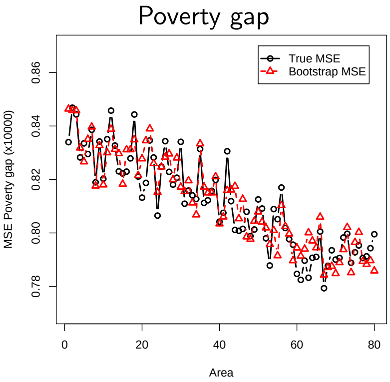




ECM BOOTSTRAP

- El ECM bootstrap persigue a los verdaderos valores.



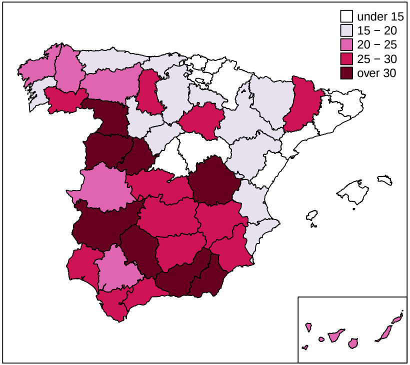


POBREZA EN ESPAÑA

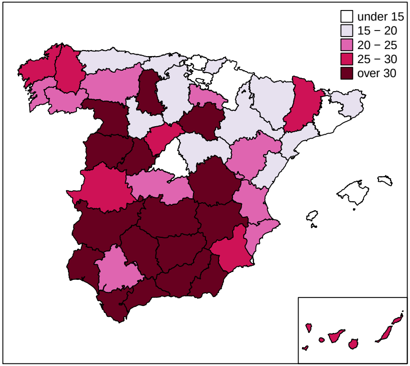
- **Datos:** Encuesta de Condiciones de Vida, 2006.
- **Objetivo:** Calcular estimadores EB y HB de las incidencias y brechas de pobreza por provincias y género.
- **Áreas:** $D = 52$ **provincias** para cada **género**. Modelo distinto para cada género.
- **Transformación:** Consideramos que el log de la renta trasladada $Y_{di} = T(E_{di}) = \log(E_{di} + k)$ sigue el modelo con errores anidados.
- **Variables auxiliares:** Indicadores de 5 grupos de **edad**, de tener **nacionalidad** española, de 3 niveles **educativos** y del estado **laboral** (desempleado, ocupado o inactivo).

RESULTADOS: INCIDENCIA DE POBREZA (%)

Hombres



Mujeres

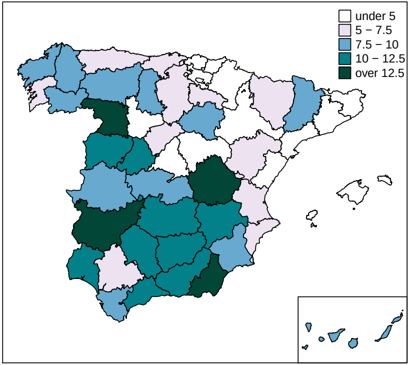


Inc.Pob.$\geq$ 30 %, Hombres: Almería, Granada, Córdoba, Badajoz, Ávila, Salamanca, Zamora, Cuenca.

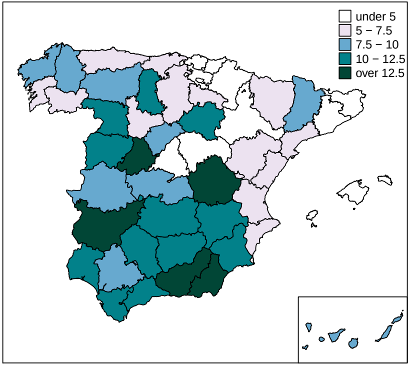
Mujeres: también Jaén, Albacete, Ciudad Real, Palencia, Soria.

RESULTADOS: BRECHA DE POBREZA (%)

Hombres



Mujeres



Brecha pob. $\geq$ 12.5 %, Hombres: Almería, Badajoz, Zamora, Cuenca.
Mujeres: Granada, Almería, Badajoz, Ávila, Cuenca.

EXTENSIONES

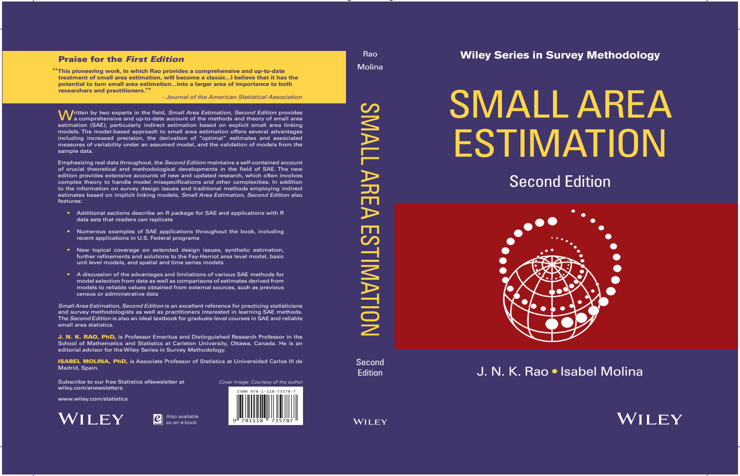
- EB no incorpora pesos del diseño→ Pseudo EB
  ✓ *Guadarrama, Molina & Rao (2018), CSDA*
- Dos niveles de agrupación → EB bajo un modelo con errores anidados a dos niveles.
  ✓ *Marhuenda, Molina, Morales & Rao (2017), JRSSA*
- Parámetro no lineal particular: Media del área $\bar{E}_d$ bajo un modelo con errores anidados para $Y = \log(E)$ → Estimador EB explícito y aproximación asintótica del ECM.
  ✓ *Molina & Martín (2018), AOS*
- EB original basado en normalidad para $Y = T(E)$ → Extensión a distribuciones asimétricas.
  ✓ *Graf, Marín & Molina (2018), Test*

SOFTWARE

El paquete de R `sae` contiene funciones para:

- Modelo FH: `eblupFH`, `mseFH`.
- Modelo FH espacial: `eblupSFH`, `mseSFH`, `pbmseSFH`, `npbmseSFH`.
- Modelo FH espacio-temporal: `eblupSTFH`, `pbmseSTFH`.
- Modelo de errores anidados: `eblupBHF`, `pbmseBHF`.
- Método EB: `ebBHF`, `pbmseebBHF`.
- Otros: `direct`, `pssynt`, `ssd`.
- Conjuntos de datos y ejemplos.

LIBRO



TBC

Praise for the *First Edition*

"This pioneering work, in which Rao provides a comprehensive and up-to-date treatment of small area estimation, will become a classic...I believe that it has the potential to turn small area estimation...into a larger area of importance to both researchers and practitioners."

- Journal of the American Statistical Association

Written by two experts in the field, *Small Area Estimation, Second Edition* provides a comprehensive and up-to-date account of the methods and theory of small area estimation (SAE), particularly indirect estimation based on explicit small area linking models. The model-based approach to small area estimation offers several advantages including increased precision, the derivation of "optimal" estimates and associated measures of variability under an assumed model, and the validation of models from the sample data.

Emphasizing real data throughout, the *Second Edition* maintains a self-contained account of crucial theoretical and methodological developments in the field of SAE. The new edition provides extensive accounts of new and updated research, which often involves complex theory to handle model misspecifications and other complexities. In addition to the information on survey design issues and traditional methods employing indirect estimates based on implicit linking models, *Small Area Estimation, Second Edition* also features:

- Additional sections describe an R package for SAE and applications with R data sets that readers can replicate
- Numerous examples of SAE applications throughout the book, including recent applications in U.S. Federal programs
- New topical coverage on extended design issues, synthetic estimation, further refinements and solutions to the Fay-Herriot area level model, basic unit level models, and spatial and time series models
- A discussion of the advantages and limitations of various SAE methods for model selection from data as well as comparisons of estimates derived from models to reliable values obtained from external sources, such as previous census or administrative data

Small Area Estimation, Second Edition is an excellent reference for practicing statisticians and survey methodologists as well as practitioners interested in learning SAE methods. The *Second Edition* is also an ideal textbook for graduate-level courses in SAE and reliable small area statistics.

J. N. K. RAO, PhD, is Professor Emeritus and Distinguished Research Professor in the School of Mathematics and Statistics at Carleton University, Ottawa, Canada. He is an editorial advisor for the Wiley Series in Survey Methodology.

ISABEL MOLINA, PhD, is Associate Professor of Statistics at Universidad Carlos III de Madrid, Spain.

Subscribe to our free Statistics eNewsletter at wiley.com/enewsletters

www.wiley.com/statistics

Cover Image: Courtesy of the author

ISBN 978-1-118-73578-7

9 781118 735787

WILEY

Also available as an e-book

Rao
Molina

SMALL AREA ESTIMATION

Second Edition

WILEY

Wiley Series in Survey Methodology

SMALL AREA ESTIMATION

Second Edition

J. N. K. Rao • Isabel Molina

WILEY

DOMO ARIGATO!!
ESKERRIK ASKO!!
EFCHARISTÓ POLÝ!!
GRAZIE MILLE!!
MERCI BEAUCOUP!!
¡¡MOITAS GRAZAS!!
¡¡MOLTES GRÀCIES!!
¡¡MUCHAS GRACIAS!!
THANK YOU VERY MUCH!!
VIELEN DANK!!
XIE XIE!!