Modelos a nivel de área

Muestreo II

Licenciatura en Estadística

2023

Introducción

Muchas veces la información auxiliar a nivel de individuo es **limitada** o incluso nula.

Una alternativa atractiva es operar con covariables a nivel de dominio.

Ejemplo:

- Cantidad de escuelas por barrio.
- Cantidad de desempleados por departamento.
- Cantidad de pixeles clasificados como soja/maíz por segmento.

Los modelos de área son aquellos en donde se modela el parámetro de interés a nivel de dominio.

La información auxiliar se encuentra a nivel de área.

Modelo de Fay-Herriot

Dados D dominios y $\mu_1, \mu_2, ..., \mu_D$ parámetros de interés asociados a esos dominios, el modelo de **Fay-Herriot** busca modelar el comportamiento de estas cantidades haciendo uso de sus estimaciones directas y un conjunto de covariables:

$$x_d = (x_{d1}, x_{d2}, ..., x_{dp})', d = 1, ..., D$$

El modelo supone:

$$\hat{\mu}_d^{\text{DIR}} = {x_d}'\beta + u_d + \varepsilon_d$$

Donde:

$$u_d \sim \mathcal{N}(0, \sigma_u^2)$$

$$\varepsilon_d \sim \mathcal{N}(0, \psi_d)$$

Se cumple a su vez que u_d y ε_d son independientes.

El estimador de **Fay-Herriot** en un dominio, queda definido como la predicción del modelo ajustado en ese dominio.

$$\hat{\mu}_d^{FH} = {x_d}'\hat{\beta} + \hat{u}_d$$

Ajuste del modelo (BLUP)

Es fácil notar que el modelo supone una estructura de covarianza Block-diagonal en la población, en donde:

$$y_i = \hat{\mu}_d^{\text{DIR}}, \quad X_i = {x_d}', \quad Z_i = 1$$

$$u_i = u_d, \quad \varepsilon_i = \varepsilon_d, \quad \beta = (\beta_1, ..., \beta_p)'$$

$$G_i = \sigma_u^2, \quad R_i = \psi_d, \quad V_i = \psi_d + \sigma_u^2$$

Asumiendo los parámetros de varianza (σ_u^2, ψ_d) son conocidos y aplicando las fórmulas para estimar β y u de la diapositiva anterior, se llega al mejor predictor líneal insesgado.

BLUP

$$\tilde{\beta}(\sigma_u^2, \psi) = \left(\sum_{i=1}^{m} \frac{x_d {x_d}'}{\psi_d + \sigma_u^2} \right)^{-1} \left(\frac{{x_d}' \hat{\mu}_d^{\text{DIR}}}{\psi_d + \sigma_u^2} \right)$$

$$\tilde{u}_d(\sigma_u^2, \psi) = \gamma_d(\hat{\mu}_d^{\text{DIR}} - {x_d}' \tilde{\beta}), \quad \gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \psi_d}$$

$$\Rightarrow \tilde{\mu}_d^{\text{FH}} = {x_d}' \tilde{\beta} + \tilde{u}_d$$

Expresión alternativa

$$
\begin{aligned}
\tilde{\mu_d}^{FH} &= {x_d}'\tilde{\beta} + \tilde{u_d} \\
&= {x_d}'\tilde{\beta} + \gamma_d \left(\hat{\mu_d}^{DIR} - {x_d}'\tilde{\beta} \right) \\
&= {x_d}'\tilde{\beta} + \gamma_d \hat{\mu_d}^{DIR} - \gamma_d {x_d}'\tilde{\beta} \\
&= \gamma_d \hat{\mu_d}^{DIR} + (1 - \gamma_d)\, {x_d}'\tilde{\beta}
\end{aligned}
$$

- El estimador es una combinación convexa del estimador directo y el sintético.
- Asigna los pesos con base en la variabilidad del estimador directo y la homogeneidad de los dominios.

Sesgo

El estimador de Fay-Herriot es asintóticamente insesgado:

$$\begin{aligned}
\mathbb{B}(\tilde{\mu}^{FH}) &= \mu_d - \mathbb{E}(\tilde{\mu}^{FH}) \\
&= \mu_d - \mathbb{E}\left(x_d{}'\tilde{\beta} + \gamma_d\left(\hat{\mu_d}^{DIR} - x_d{}'\tilde{\beta}\right)\right) \\
&= \mu_d - x_d{}'\tilde{\beta} - \gamma_d\mathbb{E}\left(\hat{\mu_d}^{DIR}\right) + \gamma_d x_d{}'\tilde{\beta} \\
&= \mu_d(1-\gamma_d) - x_d{}'\tilde{\beta}(1-\gamma_d) \\
&= (1-\gamma_d)(\mu_d - x_d\tilde{\beta})
\end{aligned}$$

$$\mathbb{B}(\tilde{\mu}^{FH}) \to 0 \text{ cuando } n_d \to N_d$$

$\mathbb{ECM}$ del BLUP

$$\mathbb{ECM}(\tilde{\mu}_d^{\mathsf{FH}}) = g_{1d}(\sigma_u^2) + g_{2d}(\sigma_u^2) \quad (1)$$

Con:

$$g_{1d}(\sigma_u^2) = \gamma_d \psi_d \quad (2)$$

$$g_{2d}(\sigma_u^2) = (1-\gamma_d)^2 x_d' \left(\frac{\sum_{d=1}^{D} x_d x_d'}{\psi_d + \sigma_u^2} \right)^{-1} x_d \quad (3)$$

EBLUP

En la práctica no conocemos ni a σ_u^2 ni a ψ_d.

1. $\psi_d = \mathbb{VAR}(\mu_d^{\text{DIR}} | x_d{}'\beta + u_d)$, es la varianza del estimador directo. Utilizamos fórmulas cerradas, método del último conglomerado, bootstrap, jacknife, etc.. para obtener $\hat{\psi}_d$.
2. Obtenida $\hat{\psi}_d$, se incorpora al modelo y posteriormente se estima σ_u^2 con máxima verosimilitud o máxima verosimilitud restringida.

De ese modo se llega a:

$$\hat{\mu}_d^{FH-EBLUP} = \hat{\gamma}_d \hat{\mu}^{\text{DIR}} + (1 - \hat{\gamma}_d) x_d' \hat{\beta}$$

Con:

$$\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\psi}_d}$$

$$\hat{\beta}(\hat{\sigma}_u^2, \hat{\psi}) = \left(\sum_{i=1}^{m} \frac{x_d {x_d}'}{\hat{\psi}_d + \hat{\sigma}_u^2} \right)^{-1} \left(\frac{{x_d}' \hat{\mu}_d^{\text{DIR}}}{\hat{\psi}_d + \hat{\sigma}_u^2} \right)$$

ECM EBLUP

$$\mathbb{ECM}(\hat{\mu}_d^{\text{FH-EBLUP}}) = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + g_{3i}(\sigma_u^2)$$

Un estimador insesgado:

$$\widehat{\mathbb{ECM}}(\hat{\mu}_d^{\text{FH-EBLUP}}) = g_{1i}(\hat{\sigma_u}^2) + g_{2i}(\hat{\sigma}_u^2) + 2g_{3i}(\hat{\sigma}_u^2)$$

Con:

$$g_{3i}(\sigma_u^2) = \psi_d^2(\psi_d + \sigma_u^2)\bar{V}(\hat{\sigma}_u^2)$$