

Estimación en Áreas Pequeñas sobre Desnutrición Crónica Infantil en Ecuador

Instituto Nacional de Estadística y Censos (INEC)

Noviembre - 2018

La necesidad de SAE en Ecuador

- En Ecuador no se dispone de fuentes de información continuas para medir la Desnutrición Crónica Infantil (DCI).
- Se disponen como fuentes de información ECV 2006, ENSANUT 2012 y ECV 2014.
- Las muestras levantadas tanto en ENSANUT como ECV tienen únicamente representatividad provincial.
- Las estrategias para reducir la DCI generalmente necesitan de información con representatividad en área más pequeñas.
- En base a esta necesidad, se incorporó en la agenda de investigación del Instituto de Estadísticas de Ecuador un SAE para estimar DCI.

Estructura de la presentación:

1. Empirical Best Predictor bajo Regresión con Errores Anidados (NER)
2. Empirical Best Predictor bajo Regresiones Logísticas Mixtas
3. Modelo Fay-Herriot (Modelo de Área Lineal Mixto)

***Empirical Best Predictor* bajo Regresión con Errores Anidados**

Instituto Nacional de Estadística y Censos (INEC)

***Empirical Best Predictor* bajo Regresión con Errores Anidados**

- Se estima la regresión NER donde la variable dependiente es la talla para la edad, se tiene cuatro variables continuas y una variable dicotómica.
- Para la estimación de DCI se considera que esta es equivalente a un FGT(0) en relación a la curva de talla para la edad propuesta por OMS. Por lo cual se puede escribir la siguiente ecuación:

$$\hat{p}_d^{eb} = \frac{1}{N_d} \left\{ \sum_{j \in s_d} h_0(y_{dj}) + \sum_{j \in r_d} \hat{p}_{dj}^{eb} \right\} \quad (1)$$

Empirical Best Predictor bajo Regresión con Errores Anidados

- Respecto a la ecuación (1) s_d se referirá la información obtenida desde la encuesta mientras que r_d será información obtenida desde el censo excluyendo la encuesta:

$$\hat{\rho}_{dj}^{eb} = \frac{1}{N_d} \left\{ \sum_{j \in s_d + r_d} \rho_{dj}^{eb} \right\} \quad (2)$$

- $s_d + r_d$ representará el uso del Censo completo. De esta forma, para estimar \hat{p}^{eb}_{dj} se seguirá las ecuaciones:

$$\hat{p}_{dj}^{eb} = P(N(0, 1) < \alpha_{dj}) \quad (3)$$

$$\alpha_{dj} = v_{dj|s}^{-1/2} (T(z_0) - \mu_{dj|s}) \quad (4)$$

Empirical Best Predictor bajo Regresión con Errores Anidados

- Para estimar la ecuación (4) es necesario conocer $v_{dj|s}$ & $\mu_{dj|s}$, para su cálculo se utilizará:

$$\mu_{dj|s} = \begin{cases} \mathbf{x}_{dj}\beta + \gamma_d^w (\bar{y}_{ds}^w - \bar{\mathbf{x}}_{ds}^w \beta) & \text{si } n_d \neq 0, \\ \mathbf{x}_{dj}\beta & \text{si } n_d = 0, \end{cases} \quad (5)$$

- Dónde γ_d^w se define como:

$$\gamma_d^w = \frac{w_d \sigma_u^2}{w_d \sigma_u^2 + \sigma_e^2} \quad (6)$$

Empirical Best Predictor bajo Regresión con Errores Anidados

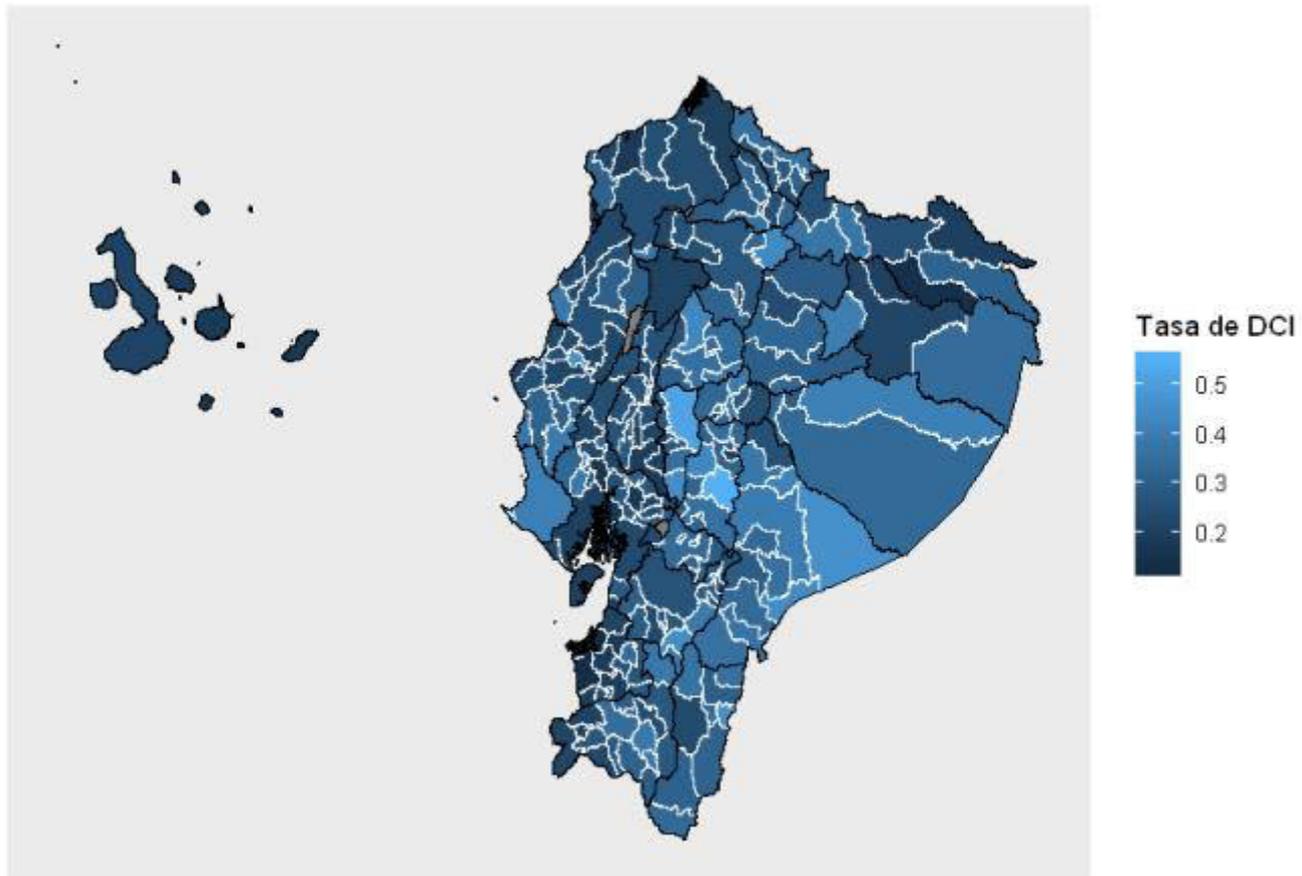
- Para estimar $v_{dj|s}$ se tiene la definición:

$$v_{dj|s} = \begin{cases} \sigma_u^2(1 - \gamma_d^w) + w_{dj}^{-1}\sigma_e^2 & \text{si } n_d \neq 0, \\ \sigma_u^2 + w_{dj}^{-1}\sigma_e^2 & \text{si } n_d = 0, \end{cases} \quad (7)$$

- A continuación se presentan las ecuaciones para estimar la ecuación (3) en el caso de $n = 0$. Para este caso la ecuación (4) vendría a ser:

$$\alpha_{dj} = v_{dj|s}^{-1/2}(T(z_0) - \mu_{dj|s}) = (\sigma_u^2 + w_{dj}^{-1}\sigma_e^2)^{-1/2}(T(z_0) - \mathbf{x}_{dj}\beta) \quad (8)$$

Mapa – EBP NER (cantón)



Empirical Best Predictor bajo Regresiones Logísticas Mixtas

Instituto Nacional de Estadística y Censos (INEC)

CÁLCULO DE LA DESNUTRICIÓN CRÓNICA EN ECUADOR UTILIZANDO MODELOS DE UNIDAD LOGÍSTICOS MIXTOS

El modelo de unidad logístico mixto consiste en la estimación de la probabilidad de ocurrencia de una variable dicotómica (dummy) explicada por una serie de variables auxiliares.

En nuestro caso, se utiliza el siguiente modelo:

- Modelo : y =desnutrición; x =categorías de la educación del jefe de hogar (3 categorías) más una variable cuantitativa (edad en meses de los niños menores a 5 años)

MODELO DE UNIDAD LOGÍSTICO MIXTO

- La variable objetivo y_{dj} denota la j ésima observación muestral del dominio d y su distribución, condicionada a los efectos aleatorios v_d es:

$$y_{dj} | v_d \sim \text{Bin}(m_{dj}, p_{dj}), \quad d = 1, \dots, D, \quad j = 1, \dots, n_d$$

- Se asume el parámetro:

$$\eta_{dj} = \log \frac{p_{dj}}{1 - p_{dj}} = \mathbf{x}_{dj} \boldsymbol{\beta} + \phi v_d$$

MODELO DE UNIDAD LOGÍSTICO MIXTO

- De igual forma se asume que las y_{dj} son independientes y condicionadas a \mathbf{v} , llegando a determinar que:

$$P_{dj} = \frac{\exp\{\mathbf{x}_{dj}\beta + \phi v_d\}}{1 + \exp\{\mathbf{x}_{dj}\beta + \phi v_d\}}$$

Predictores Empíricos Óptimos (EBP)

- El EBP de p_{dj} es $p_{dj}^{\wedge}(\theta^{\wedge})$ y se puede aproximar con el siguiente procedimiento:

1. Estimar $\hat{\theta} = (\hat{\beta}, \hat{\phi})$.

2. Para $s = 1, \dots, S$, generar $\nu_d^{(s)}$ i.i.d $N(0, 1)$ y $\nu_d^{(S+s)} = -\nu_d^{(s)}$

3. Calcular $\hat{p}_{dj}(\hat{\theta}) = \frac{\hat{N}_{dj}}{\hat{D}_{dj}}$,

Predictores Empíricos Óptimos (EBP)

- Dónde:

$$\hat{N}_{dj} = \frac{1}{2S} \sum_{s=1}^{2S} \left\{ \frac{\exp\{\mathbf{x}_{dj}\beta + \phi\nu_d^{(s)}\}}{1 + \exp\{\mathbf{x}_{dj}\beta + \phi\nu_d^{(s)}\}} \right.$$

$$\left. \cdot \exp \left\{ \hat{\phi}y_d\nu_d^{(s)} - \sum_{i=1}^{n_d} m_{di} \log \left[\exp \left\{ \mathbf{x}_{di}\hat{\beta} + \phi\nu_d^{(s)} \right\} \right] \right\} \right\},$$

$$\hat{D}_{dj} = \frac{1}{2S} \sum_{s=1}^{2S} \exp \left\{ \hat{\phi}y_d\nu_d^{(s)} - \sum_{i=1}^{n_d} m_{di} \log \left[\exp \left\{ \mathbf{x}_{di}\hat{\beta} + \phi\nu_d^{(s)} \right\} \right] \right\}$$

Predictores Empíricos Óptimos (EBP)

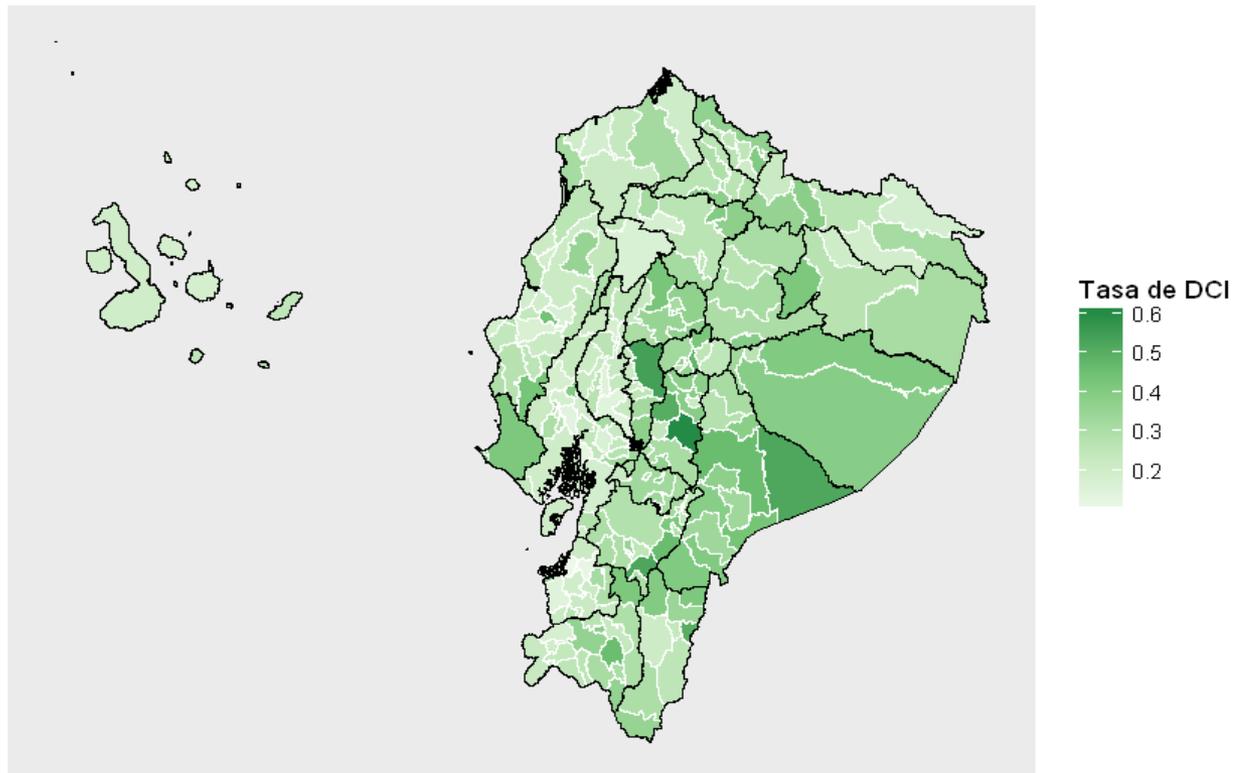
- En el caso de que $n_d=0$ para un dominio d y alguna de las variables auxiliares sea continua ; el EBP de μ_d es $\mu^*_d(\theta^*)$ y puede aproximarse de la siguiente forma:

1. Estimar $\hat{\theta} = (\hat{\beta}, \hat{\phi})$

2. Calcular $\hat{\mu}_d(\hat{\theta}) = \sum_{j=1}^{N_d} \hat{q}_{dj} y$

3. Se asume que $\hat{\phi} \nu_d^{(s)} = 0$

Mapa – EBP Logit (cantón)



Modelo Fay-Herriot (Modelo de Área Lineal Mixto)

Instituto Nacional de Estadística y Censos (INEC)

Modelo Fay-Herriot (F-H)

- En la primera etapa el **modelo muestral** da cuenta de como los estimadores directos de la muestra están relacionados con valores de la población y describe el error de muestreo:

$$y_d = \mu_d + e_d \quad d = 1, \dots, D,$$

- En la segunda etapa el **modelo de enlace** asume que las características de un área (dominio) μ_d , dependen linealmente de p variables auxiliares de área:

$$\mu_d = \chi_d \beta + u_d \quad d = 1, \dots, D,$$

- El modelo *F-H* enlaza el parámetro de interés Y_d (media o proporción) en un área μ_d a un vector de χ_d , variables auxiliares

$$y_d = \chi_d \beta + u_d + e_d \quad d = 1, \dots, D,$$

Datos usados en el modelo

Indicador director

Desnutrición crónica < 5 años

ECV-2014

Variables auxiliares

Desnutrición crónica < 5 años

Lactancia exclusiva en los primeros 6 meses

Dotación de Chispas

Dotación de Vitaminas

SIVAN-2016

Porcentaje de indígenas

Pobreza por NBI (hogares con menores de 5)

Agua sin conexión a red pública

Servicios higiénicos sin conexión a alcantarillado

Escolaridad

CPV-2010

Bajo peso al nacer

RNV-2014

Pobreza por consumo

ECV-2014

Función Generalizada de la Varianza

En la función generalizada de la varianza, se tomaron variables auxiliares del Registro de Nutrición 2016 y del Censo de Población 2010.

$$\log \widehat{\text{var}}_{\pi}(y_d) = b_0 + b_1 y_d + \varepsilon_d$$

$$\log \widehat{\text{var}}(dcro) = b_0 + b_1 nd + b_2 indigena + b_3 d_cronica + e$$

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.303e+00	1.924e-04	11970.112	< 2e-16	***
diry	4.003e-03	6.214e-04	6.442	8.12e-10	***
nd	2.715e-06	3.306e-06	0.821	0.4124	
aux\$indigena	-1.184e-03	4.914e-04	-2.409	0.0169	*
aux\$nbi_xpob_hu5	-5.328e-08	7.048e-08	-0.756	0.4506	
diry:nd	-2.417e-05	1.020e-05	-2.371	0.0187	*

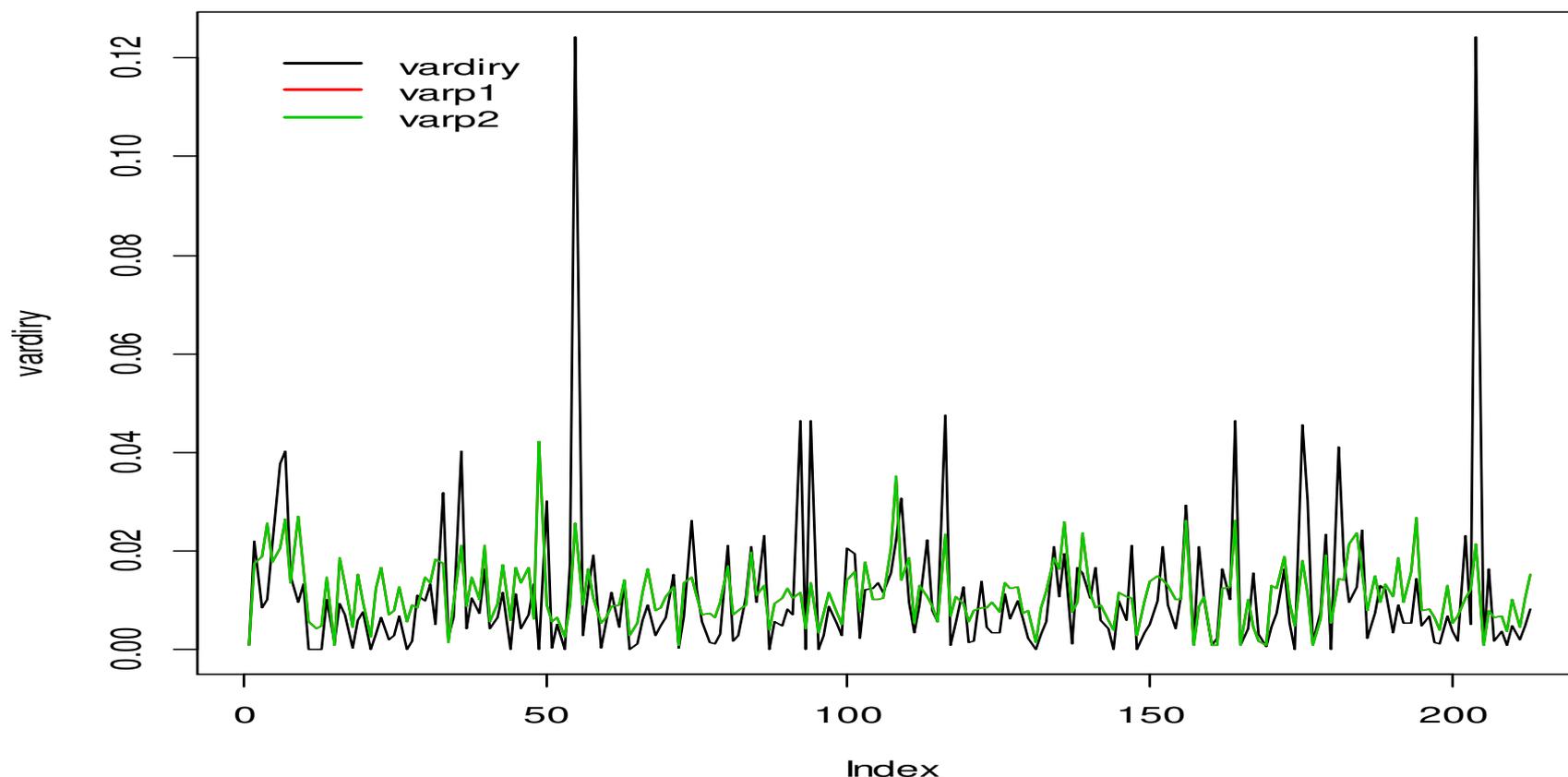
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001324 on 207 degrees of freedom
Multiple R-squared: 0.2309, Adjusted R-squared: 0.2123
F-statistic: 12.43 on 5 and 207 DF, p-value: 1.482e-10

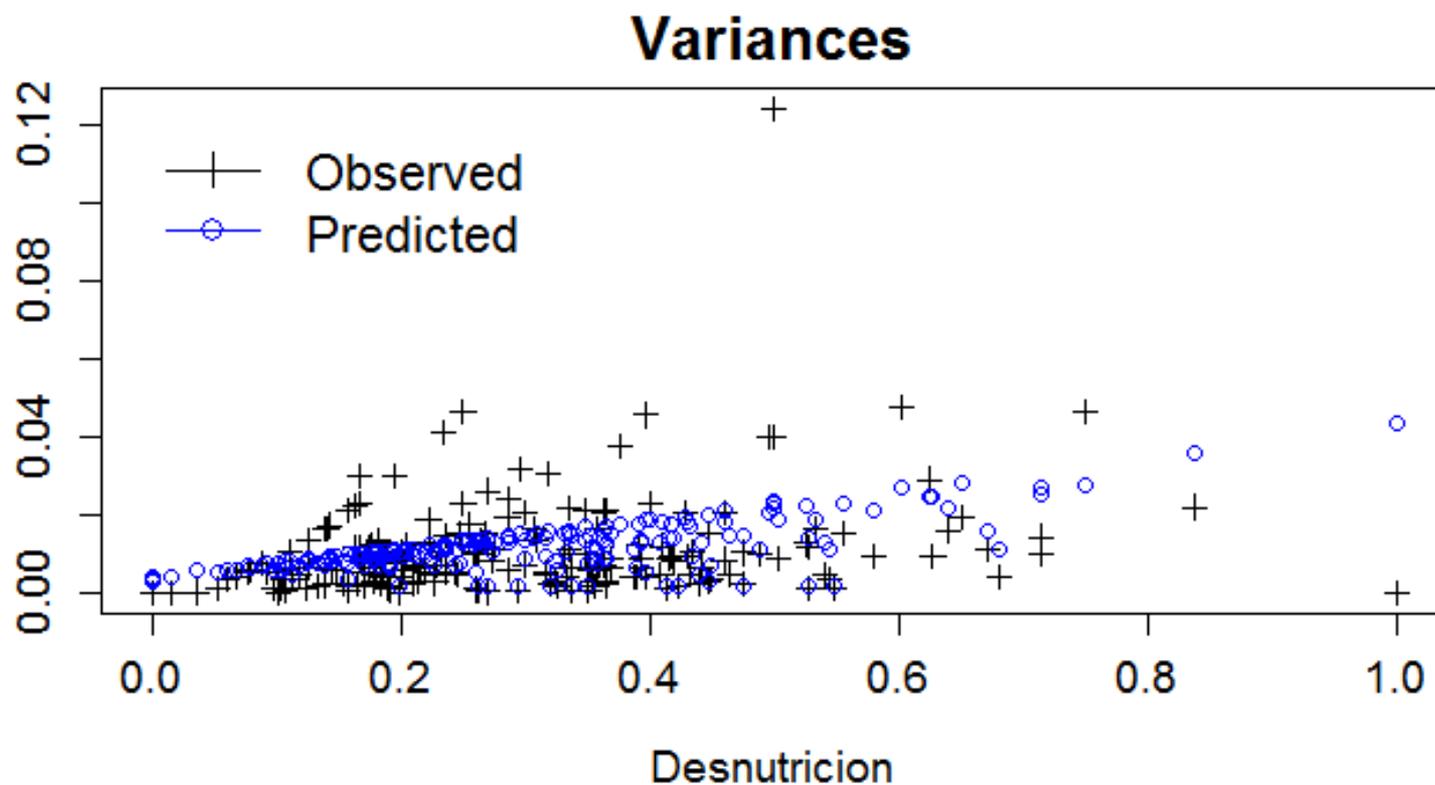
Modelo	AIC
Modelo 1	-2183.596
Modelo 2	-2201.788
Modelo 3	-2208.158
Modelo 4	-2211.763

Estimadores de la varianza GVF

Las varianzas $varp1$ y $varp2$ fueron ajustadas. Se imputó los valores negativos con el valor mínimo no negativo de los dominios.



Varianzas observadas y predichas



EBLUP – Fay-Herriot

- El EBLUP bajo el modelo Fay-Herriot, se define de la siguiente forma:

$$\hat{Y}_d^{FH} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_d^2} y_d + \frac{\sigma_d^2}{\sigma_u^2 + \sigma_d^2} x_d \hat{\beta}$$

- Si el estimador directo es preciso (n_d grande), por tanto:

$$\hat{Y}_d^{FH} \approx y_d = \hat{Y}_d^{dir} \text{ (estimador directo)}$$

- Si el estimador directo no es preciso (n_d pequeño), por tanto:

$$\hat{Y}_d^{FH} \approx x_d \hat{\beta} \text{ (estimador sintético)}$$

EBLUP – desnutrición crónica

EBLUP para desnutrición

$$\widehat{dcro} = b_0 + b_1 indigena + b_2 d_cronica + e$$

Modelos:

```
M1 <- eblupFH(diry ~ aux$indigena + aux$d_cronica, varp2)
M1$fit$estcoef
M1$fit$goodness[2]

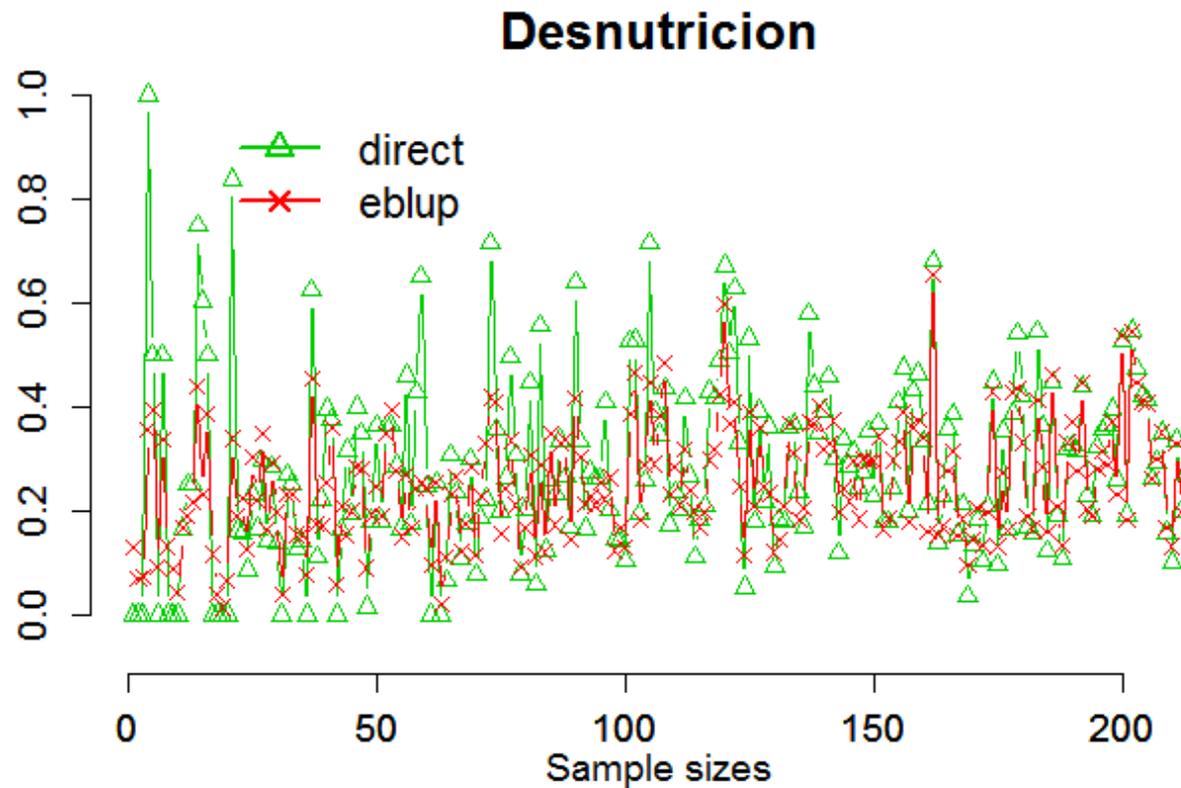
M2 <- eblupFH(diry ~ aux$indigena + aux$d_cronica + aux$nbi_xpob_hu5, varp2)
M2$fit$estcoef
M2$fit$goodness[2]

M3 <- eblupFH(diry ~ aux$indigena + aux$d_cronica + aux$vab, varp2)
M3$fit$estcoef
M3$fit$goodness[2]

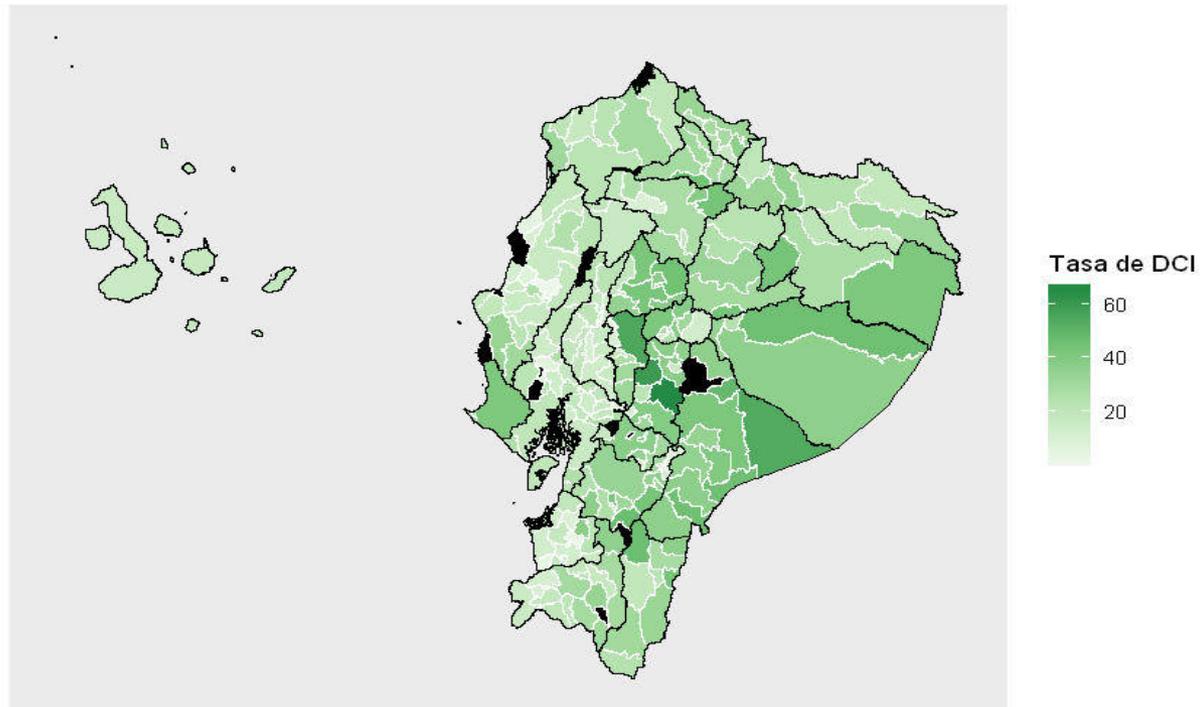
M4 <- eblupFH(diry ~ aux$indigena + aux$d_cronica + aux$vab.pc, varp2)
M4$fit$estcoef
M4$fit$goodness[2]
```

Modelo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
AIC	-240.3098	-238.5652	-238.4788	-238.3057

Dispersión de residuales F-H



Mapa – EBLUP F-H (cantón)



Bibliografía:

- FAY, R.E. Y HERRIOT, R.A. (1979), «Estimates of income for small places: An application of James-Stein procedures to census data». Journal of the American Statistical Association
- RAO, J.N.K. (2003), «Small area estimation». John Wiley
- PRASAD, N.G.N. Y RAO, J.N.K. (1990), «The estimation of the mean squared error of small-area estimators». Journal of the American Statistical Association
- Michael R Chernick and Robert A LaBudde. *An introduction to bootstrap methods with applications to R*. John Wiley & Sons, 2014.
- Tomás Hobza and Domingo Morales. Empirical best prediction under unit-level logit mixed models. *Journal of official statistics*, 32(3):661692, 2016.



www.ecuadorencifras.gob.ec



-  @ecuadorencifras
-  INEC/Ecuador
-  Inec
-  INECEcuador
-  INEC Ecuador