



Herramienta de Monitoreo Nutricional: Mapeo de Resultados Nutricionales con Datos del Censo

Matteo Cellamare and Vit Bubak

16 Noviembre 2018

Que es un Mapeo?

Metodologia que provee una descripcion detallada de la distribucion espacial de un indicador dentro de un pais

La idea es combinar datos a nivel individual y de hogares de Encuestas y Censos con el objetivo de estimar indicadores de bienestar para un area geografica especifica tales como departamentos, distritos o comunas.

Para que se utiliza?

- Resalta variaciones geograficas
- Seleccion de intervenciones (i.e. asignacion de fondos publicos)
- Diseño de intervenciones (i.e. planes regionales)
- ...

Algunos ejemplos

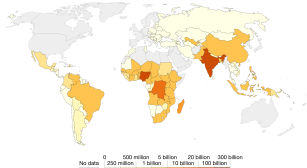
- **Sudafrica:** Mapas de pobreza, calidad sanitaria del agua y enfermedades.
- **Nicaragua:** Mapas de pobreza para guiar servicios de salud publica.
- **Guatemala:** Mapas de pobreza para guiar obras de infraestructura publica (network data)
- **Bangladesh:** Mapas de malnutricion infantil.
- Estimacion de Pobreza en **países europeos** y en **EEUU**
- ...

Algunos ejemplos visuales?

The poverty gap, in international-\$, 2013

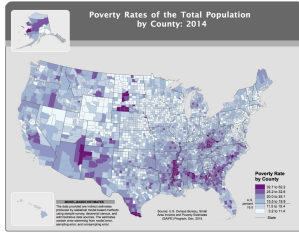
The poverty gap is the amount of money that would be theoretically needed to lift the incomes of all people in extreme poverty up to the international poverty line of \$1.90 a day. These estimates are expressed in international dollars using 2011 PPP conversion rates. This means that figures account for differences in price levels, as well as for inflation.

OurWorld in Data

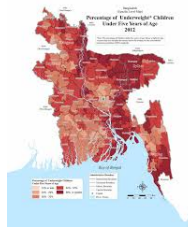
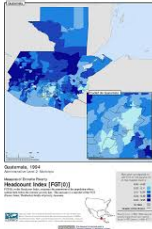
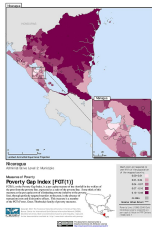
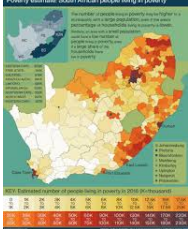


Source: Cost per year of closing the poverty gap in 2013-2014 - Pivovarov (World Bank) (2017). OurWorldinData.org/explorer/poverty/ - CC BY-SA
Note: The cost of closing the poverty gap does not take into account costs and inefficiencies from making the necessary transfers.

Poverty Rates of the Total Population by County: 2014



Poverty estimate: South African people living in poverty



Methodologia: Estimacion Directa

Para cada area d , un estimador directo usa solo observaciones de un area dado que hay sido encuestado.

π_{id} probabilidad de inclusion de unidad i en muestra s_d .

$w_{id} = 1/\pi_{id}$ pesos de muestreo (factor de expansion).

y_{id} indicador observado de unidad i en area d (HAZ, WAZ,...).

el estimador basico puede ser definido como:

$$\hat{Y}_j = \frac{1}{n_d} \sum_{i \in s_d} y_{id} w_{id}$$

con tamaño muestral n_d para el area d .

Propiedades

Ventajas

- No requieren supuestos en el modelo; es decir, son completamente no parametricos.
- Muy faciles de usar

Desventajas

- Se basa solo en el uso de datos muestrales.
- Muy ineficiente para areas de bajo n_d .
- Areas excluidas de la encuesta no pueden ser estimadas.

Metodologia: Modelo de Estimacion de Areas Pequeñas (SAE)

El metodo SAE asume un modelo lineal mixto para la variable y en la unidad de muestreo individual i , con efectos aleatorios u_j para clusters muestrales (distritos, departamentos, hogares,..).

$$y_{id} = x_{id}^T \beta + u_d + e_{id}, \quad i = 1, \dots, n_j, \quad d = 1, \dots, D$$

donde u_j son efectos aleatorios a nivel de area (variacion entre areas)

La estimacion de parametros al nivel de area son obtenidos via:

- Metodo ELL del Banco Mundial (*Elbers et al. 2003*)
- El Empirical Best Predictor (EBP) (*Molina & Rao 2009*)

Ambos metodos usan la tecnica del Bootstrap para obtener medidas de mayor precision.

Propiedades

Ventajas

- Basado en datos a nivel unitario y usan un mayor tamaño muestral para estimar el modelo
- Produce estimadores eficientes para dominios con bajo (o nulo) tamaño muestral.
- Pueden ser aplicados para estimar indicadores generales definidos en funcion a variable de respuesta y .

Desventajas

- Realiza supuestos en modelo. Verificar validez del modelo es crucial.
- Estimadores parametricos bootstrap del Error Cuadratico Medio (MSE) son computacionalmente intensivos.

Que necesitamos para implementar el SAE?

Identificar la variable de interes y

- **Microdatos de Encuestas:** Disponen y y variables de control x (relacionadas a y)
- **Microdatos de Censos:** Disponen variables de control x pero no y
- **Geographic Information Systems (GISs):** software que permite construir el mapa, es decir, asigna informacion de interes segun coordenadas geograficas.

Escenario Ideal y Desafios para este mapeo

- Acceso a microdatos del Censo 2012. **(Gracias DGEEC!)**
- Buen modelo que vincule y con x.... ✓.
- Metodologia de estimacion ✓.y visualizacion apropiada ✓ **No GISs, si software libre y gratuito R!**

Metodologia SAE

- 1 Definir variable dependiente y (i.e. Altura por edad (HAZ),...)
- 2 Usar datos de Encuesta para estimar el modelo
- 3 Combinar los parametros estimados β con variables x del Censo para crear una poblacion sintetica de valores y .
- 4 Usar estas predicciones para estimar parametros target y su precision asociada (i.e. Prevalencia de niños menores de 5 años con retraso de crecimiento o *stunted children*,...)
- 5 Mapeo de resultados

Monitoreo de Nutricion en Paraguay

A nivel de Departamento y Distrito

Medidas de nutrición

- Consumo energetico diario mediano por individuo (Kcal/capita/day)
- Prevalencia de niños menores de 5 años con retraso de crecimiento (stunted children under 5 years old)
- Prevalencia de niños menores de 5 años con insuficiencia de peso/sobrepeso.

Fuentes de datos:

Survey data

Encuesta de Ingresos y Gastos y de Condiciones de Vida 2011-2012

Census/Admin data

Censo Nacional de Poblacion y Viviendas 2012

Provistos por: *Dirección General de Estadística, Encuestas y Censos*

Software y GISs

Analisi y Aplicacion web desarrollada por autor con **RStudio** y **Shiny**

Consumo energetico diario mediano por individuo

Estimating food consumption patterns by reconciling food balance sheets and household budget surveys (FAO).

Metodologia

- 1 Colectar toda la informacion de alimentos usualmente comprados por cada hogar (seccion *Gastos fuera del hogar* y *Gastos del hogar* de Encuesta)
- 2 Calculo de consumo diario segun alimento identificado
- 3 Calculo de ingesta diaria de calorías de cada hogar usando **Tabla de composicion de alimento de Centroamerica** de la *INCAP*
- 4 Consumo energetico diario por individuo se calcula tras sumar las calorías diarias y dividir las por tamaño del hogar.

info importante: la unidad utilizada del Censo son los hogares, no los individuos.

Hogar compuesto por 3 personas

Tabla original

producto	frecuencia	cantidad
Milanesa de carne	semanal	500 gr
Arroz a granel	quincenal	2.00 kg
Banana carape	mensual	3.43 kg
⋮	⋮	⋮

Tabla INCAP

producto	Kcal _{100g}
Milanesa de carne	257.9
Arroz a granel	360.0
Banana carape	89.0
⋮	⋮

$$Kcal_{day} = \frac{(\text{cantidad/frecuencia})}{100} Kcal_{100g}$$

Tabla Final

producto	frecuencia	cantidad	Kcal _{day}	Kcal _{p/c}
Milanesa de carne	Diaria	71.4 gr	184.2	61.4
Arroz a granel	Diaria	133.3 gr	479.8	159.9
Banana carape	Diaria	122.5 gr	109.1	36.4
⋮	⋮	⋮	⋮	⋮
TOTAL			8761.6	2920.5

Indicador de Stunting (Prevalencia de niños menores de 5 años con retraso de crecimiento)

Global Database on Child Growth and Malnutrition (WHO)

- Unidad utilizada del Censo son los niños menores de 5 años.
- *Altura por edad Z-scores (HAZ)* como indicador de stunting

Prevalencia esperada de stunting del distrito d es:

$$E(S_d) = \frac{1}{B} \sum_b \left[\frac{\sum_{i \in d} (HAZ_{id}^b < -2)}{N_d} \right]$$

donde N_d es el numero de niños en distrito d y $b = 1, \dots, B$ son estimadores bootstrap.

La WHO Global Database on Child Growth and Malnutrition usa como Z-score umbral $< -2 sd$ para clasificar niños con baja altura por edad (stunting).

Indicadores de Malnutricion

Global Database on Child Growth and Malnutrition (WHO)

- *Peso por edad Z-scores (WAZ)* para malnutricion

Prevalencia esperada de niños con insuficiencia de peso del distrito d es:

$$E(U_d) = \frac{1}{B} \sum_b \left[\frac{\sum_{i \in d} (WAZ_{id}^b < -2)}{N_d} \right]$$

Prevalencia esperada de niños con sobrepeso del distrito d es:

$$E(U_d) = \frac{1}{B} \sum_b \left[\frac{\sum_{i \in d} (WAZ_{id}^b > 2)}{N_d} \right]$$

El WHO usa Z-score umbrales $< 2 sd$ y $> 2 sd$ para clasificar niños con insuficiencia de peso y sobrepeso , respectivamente.

Selección de variables de control

Datos de Encuesta

Variables al nivel de hogares:

- Consumo energetico per capita
- departamento y distrito donde vive
- Numero de miembros del hogar
- Proporción de mujeres
- Proporción de niños < 6 años de edad
- Proporción de niños entre 6 y 14 años de edad
- Proporción de adultos > 65 años de edad
- Mujer con grado de educación secundario (al menos)
- Conexión eléctrica
- Necesidad Básica Insatisfecha: Agua
- Necesidad Básica Insatisfecha: Vivienda

Variables del jefe de hogar:

- Edad
- Sexo
- Estado civil
- Años de educación
- Estado de empleo

Variables de Niños < 5 años de edad

- Peso por edad Z-score (WAZ)
- Altura por edad Z-score (WAZ)
- Edad
- Sexo

Datos del Censo

Mismas variables excepto las variables subrayadas.

Vamos al Web-App

disponible online muy pronto!!