

Estimación de áreas pequeñas y mapeo de la pobreza en Uganda

Dr Marco Pomati



Profile

Background



Proyectos de investigación recientes:

Medición de la privación material

Malnutrición en el África subsahariana

Asequibilidad de la vivienda y privación de vivienda



Enseñanza:

Métodos cuantitativos para científicos sociales

Política social



De qué trata esta charla



- Compartir los hallazgos de un ejemplo empírico (Estimación de áreas pequeñas de la pobreza multidimensional en Uganda)
- Guiarlo a través de (algunos de) los aspectos prácticos de hacer estimación de áreas pequeñas
- Comprender el flujo de trabajo de la estimación de áreas pequeñas para evitar la dependencia excesiva de los programas de estimación de áreas pequeñas preempaquetado (es decir, ¿qué hay debajo del capó?)
- La presentación asume la comprensión básica del modelado de regresión lineal
-

De qué no se trata esta charla



- No es una clase en profundidad sobre las matemáticas detrás de la estimación de áreas pequeñas
- No es una serie de simulaciones con algunas conclusiones sobre qué estimador SAE es mejor
- No es una clase sobre el modelado jerárquico bayesiano

Contorno

- Estimación de áreas pequeñas en teoría
- Estimación de áreas pequeñas en la práctica
- Ejemplo aplicado

Estimación de área pequeña

- Un conjunto de métodos destinados a medir la pobreza a nivel local
- Importante para:
 - Asignación de fondos
 - Evaluación de políticas ("Nadie se queda atrás")
 - Zonas pobres dentro de las regiones más ricas
 - Influencia del lugar en los resultados individuales

Estimaciones directas de la pobreza local

- Margen de error de proporción $1.96 \times \sqrt{\frac{\text{Proporción} (1-\text{Proporción})}{\text{número de encuestados válidos}}}$
- Número de casos necesarios para alcanzar un cierto margen de error para una proporción determinada:

$$n = \frac{1.96^2 \times \text{Proportion} \times (1-\text{Proportion})}{\text{MoE}^2}$$

- MoE de estimaciones de áreas pequeñas demasiado altas
- ¡Muchas áreas pequeñas a menudo no están cubiertas en a
-



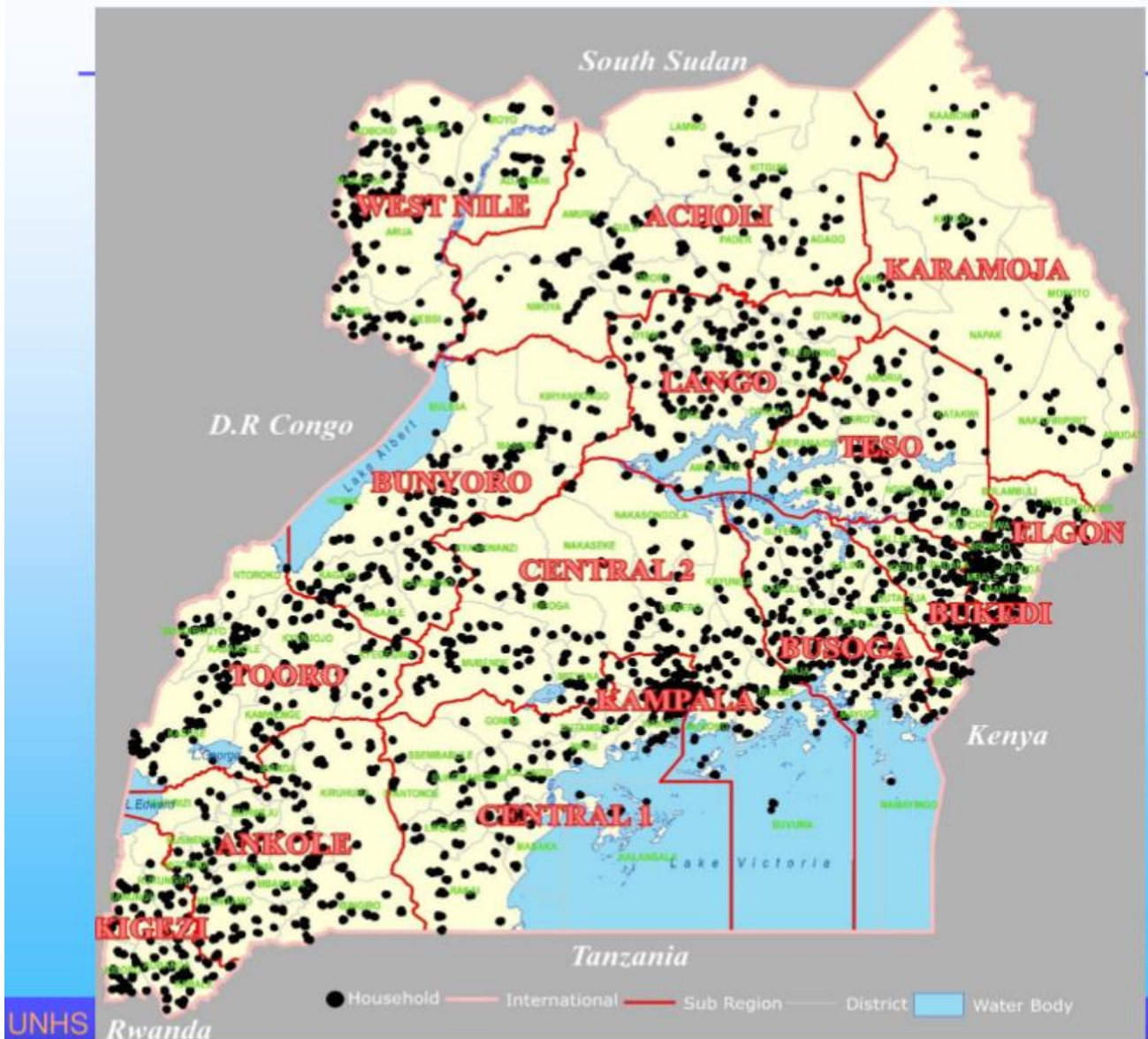
Ejemplo: Uganda

➤ Ejemplo: Ejemplo de límites administrativos de Uganda

- **Región** > subregión > distrito > condado > subcondado > parroquia > área de enumeración
- ¿Podemos producir estimaciones de pobreza a nivel de subcondado utilizando la Encuesta Nacional de Hogares?



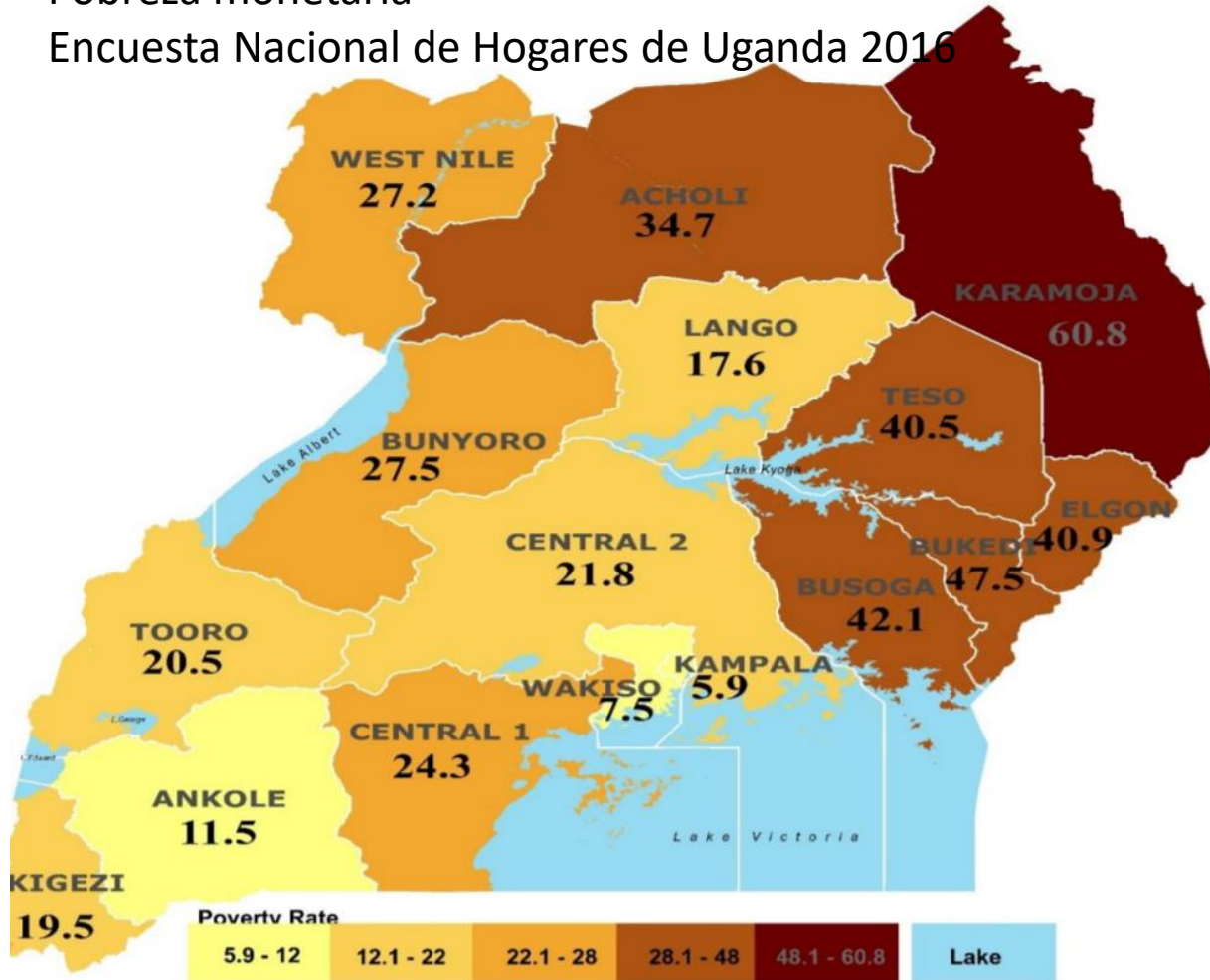
Distribution of 2016/2017 UNHS sampled households



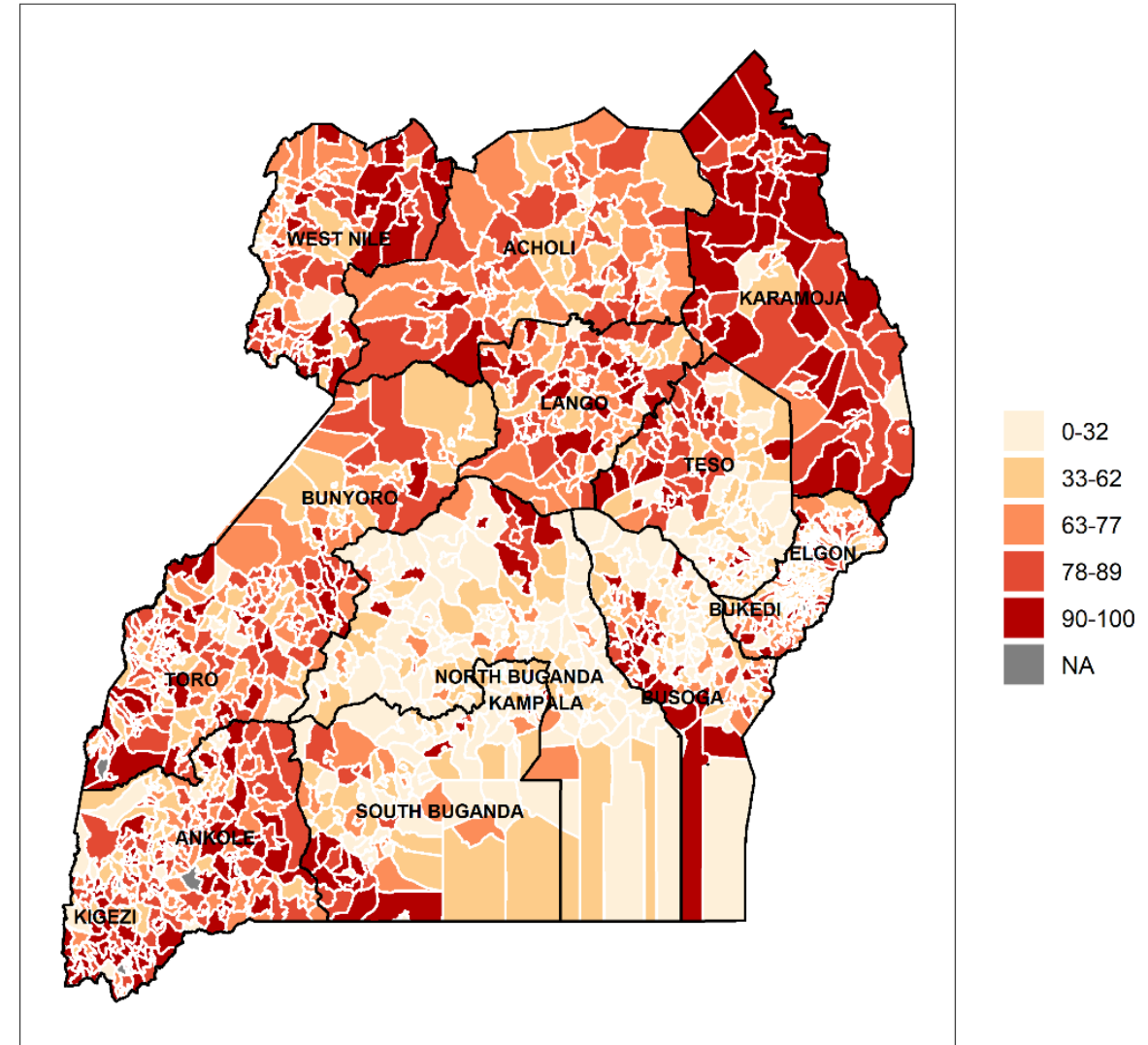
Poverty Estimates

Pobreza monetaria

Encuesta Nacional de Hogares de Uganda 2016



% of households experiencing wall deprivation
(Subcounty Census Estimates)



Estimación de áreas pequeñas: intuición básica

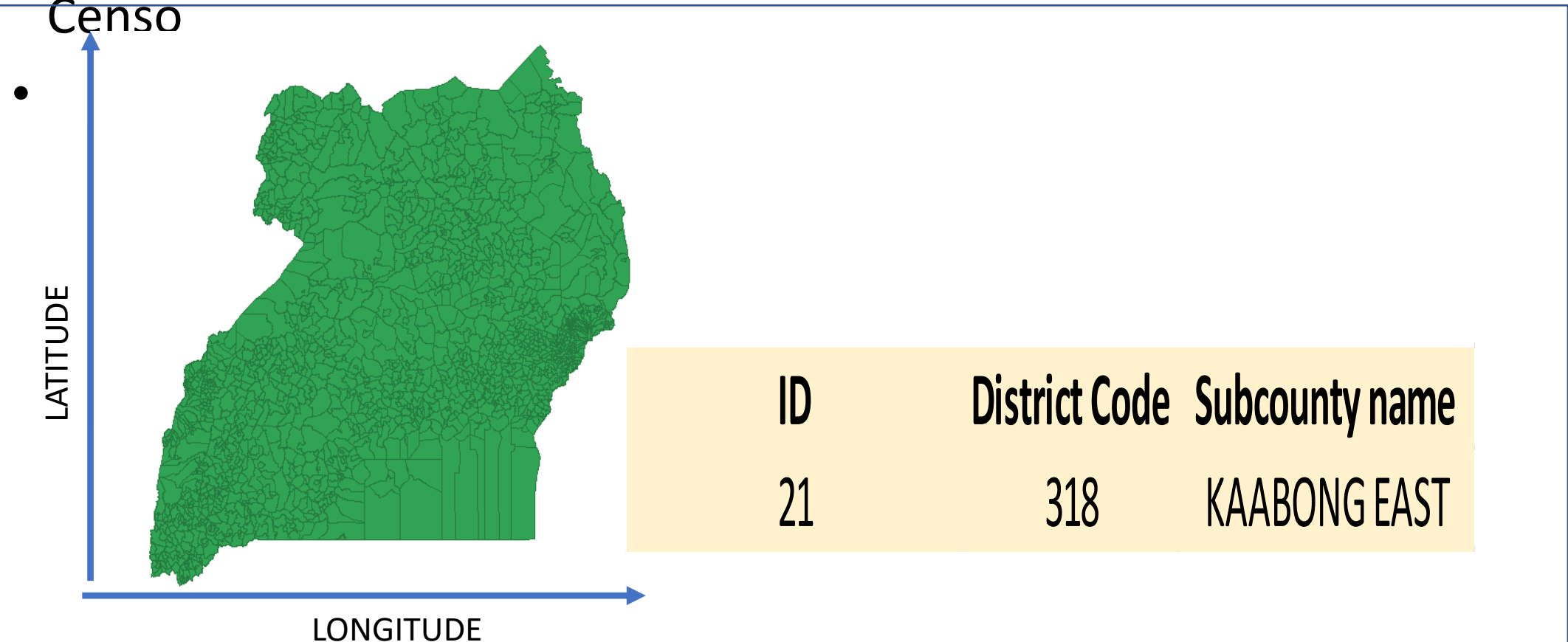
- Datos de la encuesta: Más preguntas, más frecuentes, menos casos
- Datos del censo: Más casos, menos preguntas, menos frecuentes

SAE: Combinar fortalezas para obtener estimaciones de áreas pequeñas

El tercer conjunto de datos

- Su área pequeña Shapefile/s. Tan importante como la encuesta y el

Censo



See IPUMS repository <https://international.ipums.org/international/gis.shtml>

Estimadores indirectos

- Modelización estadística

por ejemplo, modelado de regresión multinivel (frecuentista o bayesiano)

- Microsimulación especial

Implica la creación de micro-población sintética o la reponderación utilizando puntos de referencia a nivel de área censal

Fuerza de endeudamiento (modelado a nivel individual)

Ejemplo de Marco de regresión simple

$$Y_i = \alpha_0 + \beta_1 X1_i + \beta_2 X2_i + e_i$$

$$\text{Income}_i = \alpha_0 + \beta_1 \text{Secondary Education}_i + \beta_2 \text{Urban}_i + e_i$$

$$\text{Income}_i = 50 + 150_1 \text{Secondary Education}_i + 3 \text{Urban}_i + e_i$$

	Income	Has Secondary Education	Lives in Urban area	Survey model predicted Income	Residual
Survey Respondent 1	213	Yes	Yes	50+150+3=203	10
Survey Respondent 2	10	No	No	50	-40

Fuerza de endeudamiento (modelado a nivel individual)

SURVEY

$$\text{Income}_i = 50 + 150_1 \text{ Secondary Education}_i + 3 \text{ Urban}_i$$

Survey data model estimates

	Income	Has Secondary Education	Lives in Urban area	Survey model predicted Income	Residual
Survey Respondent 1	213	Yes	Yes	$50+150+3=203$	10
Survey Respondent 2	10	No	No	50	-40

CENSUS

	Income	Has Secondary Education	Lives in Urban area	Survey model predicted Income
Census Respondent 1	Not collected	Yes	Yes	$50+150+3=203$
Census Respondent 2	Not collected	No	No	50

Modelos SAE de pobreza

Metodologías más extendidas en la pobreza SAE:

- Fay-Harriot
- ELL/Banco Mundial
- Bayes empírico
- Bayes jerárquicos

Fuerza de endeudamiento- Modelado multinivel

Ejemplo de Marco de regresión simple

$$\text{Income}_i = \alpha_0 + \beta_1 X_{1i} + e_i \quad \text{Survey data model (example with just one independent variable)}$$

Marco multinivel

$$\text{Income}_{ij} = \alpha_{0j} + \beta_{1j} X_{1ij} + \varepsilon_{ij}$$

$$\alpha_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + \mu_{1j}$$

Borrowing strength (individual level modelling)

$$\text{Income}_{ij} = \alpha_j + \beta_1 X_{1ij} + \varepsilon_{ij}$$

$$\alpha_j = \gamma_{00} + \gamma_{01} W_j + \mu_j$$

SURVEY

CENSUS

	Income	Has Secondary Education	Lives in Urban area	Survey model predicted Income
Census Respondent 1 in <u>Busaana</u>	Not collected	Yes	Yes	50+150+3+ μ_{BUSAANA}
Census Respondent 2 in <u>Busaana</u>	Not collected	No	No	50+ μ_{BUSAANA}

Estimación de áreas pequeñas en la práctica

Opciones de pre-modelado: ¿Qué área pequeña?

- Elija sus niveles objetivo (su "Área pequeña").

Ejemplo de Uganda 2016:

Región > **subregión** > **distrito** > **condado** > sub-county > parroquia > área de enumeración

Comprobaciones previas a la modelización: Comparación de datos de encuestas y censos

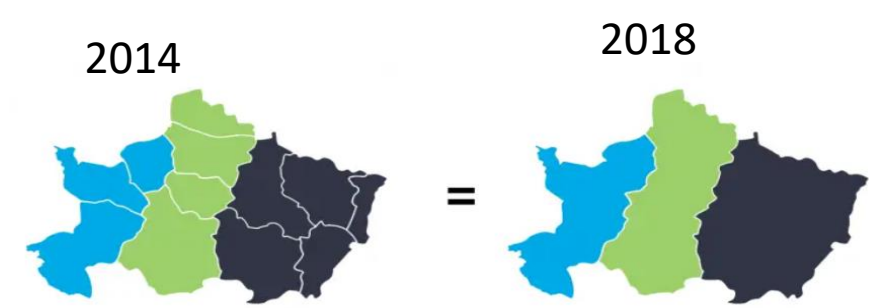
1) Comprobar

- Compruebe que la distribución de sus variables dependientes e independientes coincida en muestra y Censo
- ¿Cómo derivan el tamaño del hogar?
- ¿Cómo se selecciona el jefe de hogar?

2) Inform variable selection and possible post-stratification

Do differences between Census and survey reflect genuine differences or sampling error and/or measurement error and/or nonresponse error?

Opciones de pre-modelado: límites



- 1) Identificar una estructura de anidación consistente y relevante de los límites geográficos. Elija su nivel objetivo (su "Área pequeña")
- 2) Establezca que esto es consistente tanto en la encuesta, los datos del Censo como en el shapefile.
- 3) Investigar los cambios en los límites entre el Censo y la encuesta
- 4) Establezca una estrategia para actualizar los límites del Censo y la Encuesta y crear variables geográficas y shapefiles actualizados.
- 5) Cree un ID único en su shapefile (SIG) y luego náleelo con su encuesta y datos del Censo

5) Cree un ID único en su shapefile (SIG) y luego náleelo con su encuesta y datos del Censo

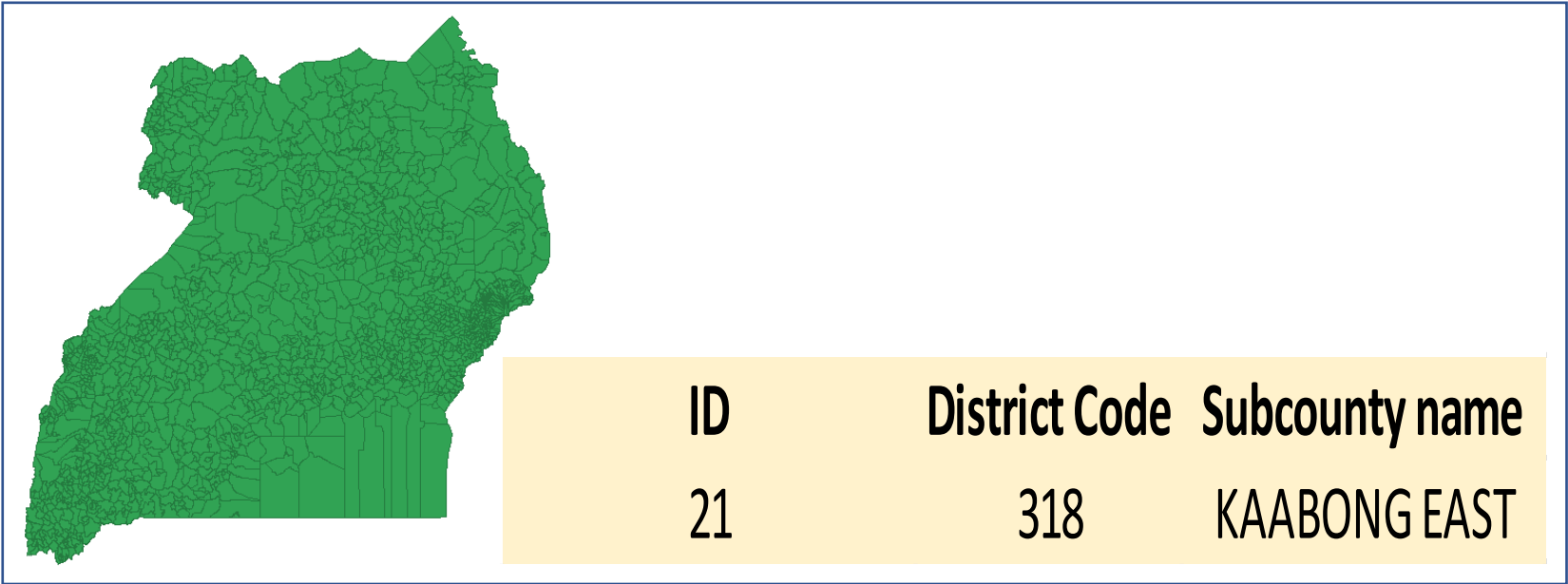
SURVEY

District Code	Subcounty name	Shapefile subcounty ID
318	KAABONG EAST	21

Censo

District Code	Subcounty name	Shapefile subcounty ID
318	KAABONG EAST	21

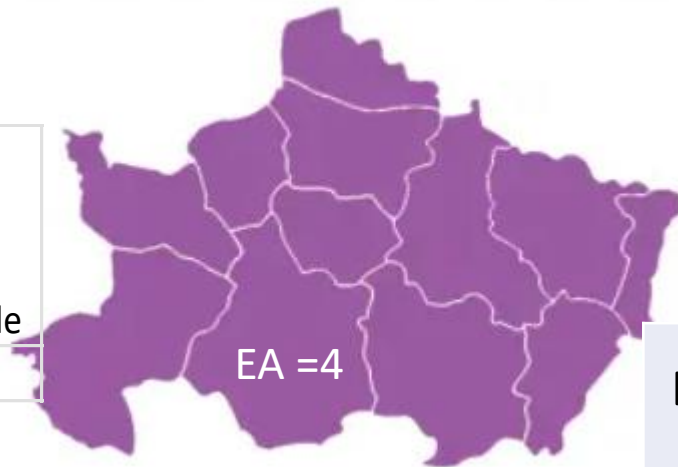
SHAPEFILE



5) Si la coincidencia no está completa, intente coincidir en el área de nivel más bajo disponible en shapefiles y datasets (por ejemplo, áreas de enumeración).

ENCUESTA

District Code	Subcounty name	EA name	EA code
318	NA	KALODEKE	4



shapefile

CENSO

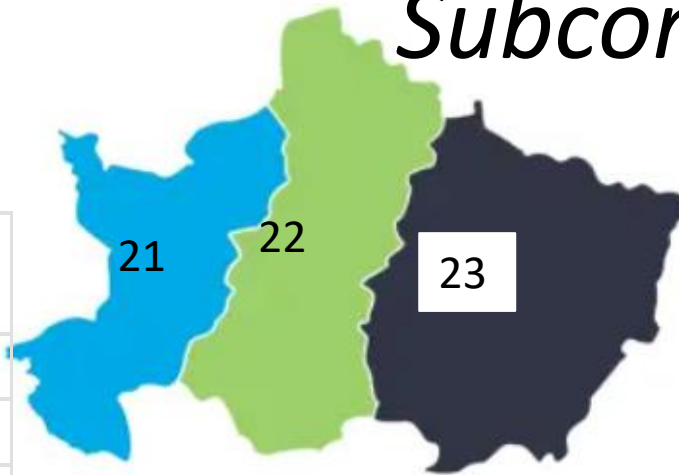
District Code	Subcounty name	EA name	Subcounty name
318	NA	KALODEKE	4

5) Si la coincidencia no está completa... perseverar y eventualmente triangular, inspeccionar manualmente, etc....

Subcondados de Shapefile

ENCUES

District Code	EA name	EA code	Shapefile subcounty ID
318	KALODEKE	4	21
318	SOKODU B	8	22
318	KOBUIN_a	11	?



CENSO

District Code	EA name	Shapefile subcounty ID
318	KALODEKE	21
318	SOKODU North	22
318	SOKODU South	22
318	KOBUIN	23

Opciones de modelado: Explore y elija el modelo de datos de la encuesta final

Diferentes especificaciones de coeficientes

Covariables a nivel Individual, Agregado e Individual + Agregado

Diferentes especificaciones de error aleatorio

¿Intercepciones aleatorias? ¿Pendientes aleatorias? ¿A qué niveles?

Evaluar el ajuste del modelo

Por ejemplo, regresión logística a nivel individual: Pseudo R², Sensibilidad, Sensibilidad, AIC, BIC

Por ejemplo, regresión logística multinivel: arriba + WAIC, Loo (Vehtari et al, 2016)

Reproducir estimaciones directas

¿Puede el modelo de encuesta reproducir tasas de pobreza precisas a nivel regional (utilizando datos de encuestas)?

Validación

- Después de haber utilizado su modelo de encuesta para predecir la pobreza en los casos del Censo y producido estimaciones de áreas pequeñas, todavía hay algunas verificaciones que debe llevar a cabo
- ¿Puede reproducir las estimaciones de pobreza de la encuesta directa utilizando los valores predichos del Censo?
-
- Variación de sus estimaciones (véase Molina y Rao, 2010)
-
- Deje una variable fuera de su conjunto de variables independientes para validar sus estimaciones de área pequeña. Lo ideal es utilizar datos auxiliares.
-

Resumen de un flujo de trabajo SAE simple

➤ Pre-modelling choices:

- ¿Qué área pequeña?
- Comparación de datos de encuestas y censos
- Límites

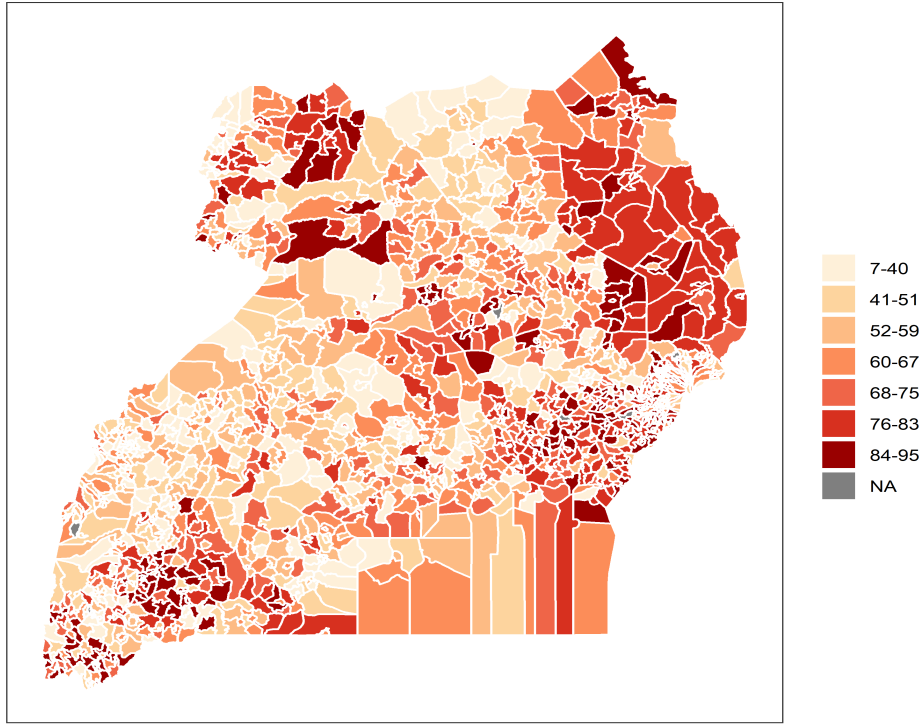
➤ Opciones de modelado: Explore, compare y elija el modelo de datos final de la encuesta

➤ Aplicar el modelo de encuesta al Censo (predecir el estado de pobreza de los encuestados del Censo) y producir estimaciones de áreas pequeñas

➤ Validar



% of children in multidimensional poverty (Subcounty SAE Estimates)



Ejemplo aplicado

Estimaciones de la pobreza multidimensional en Uganda en Uganda
2016

Pobreza multidimensional en Uganda

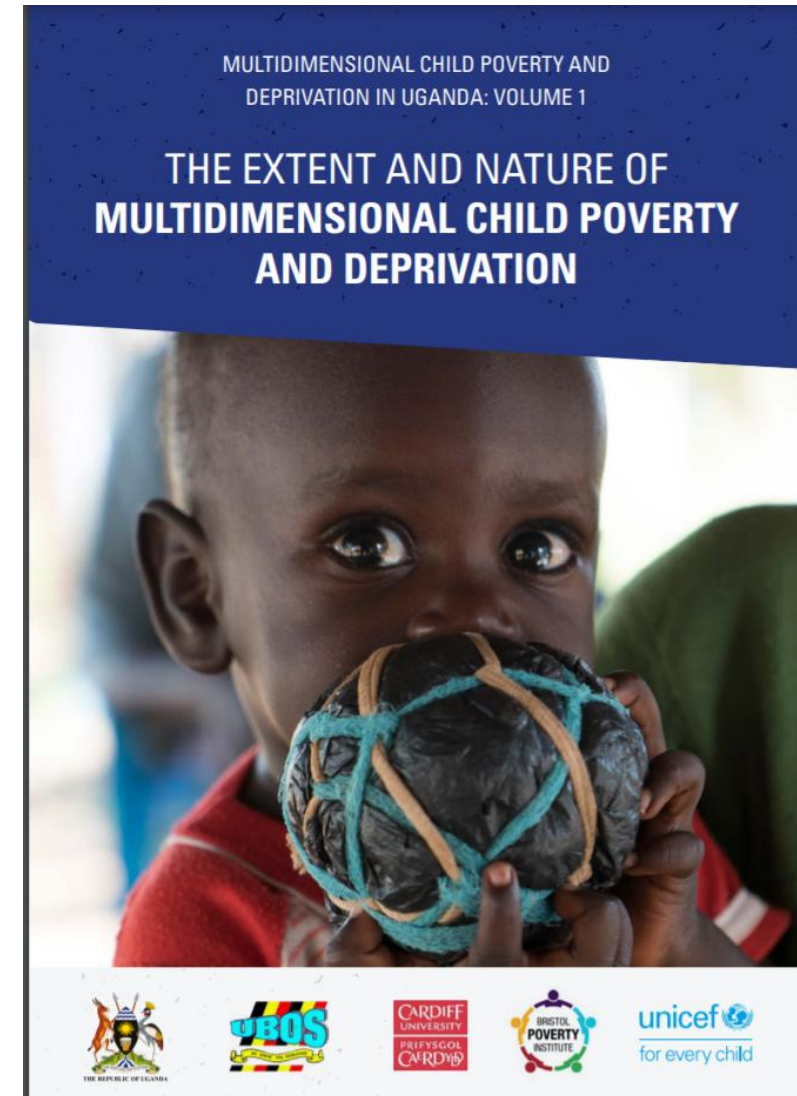
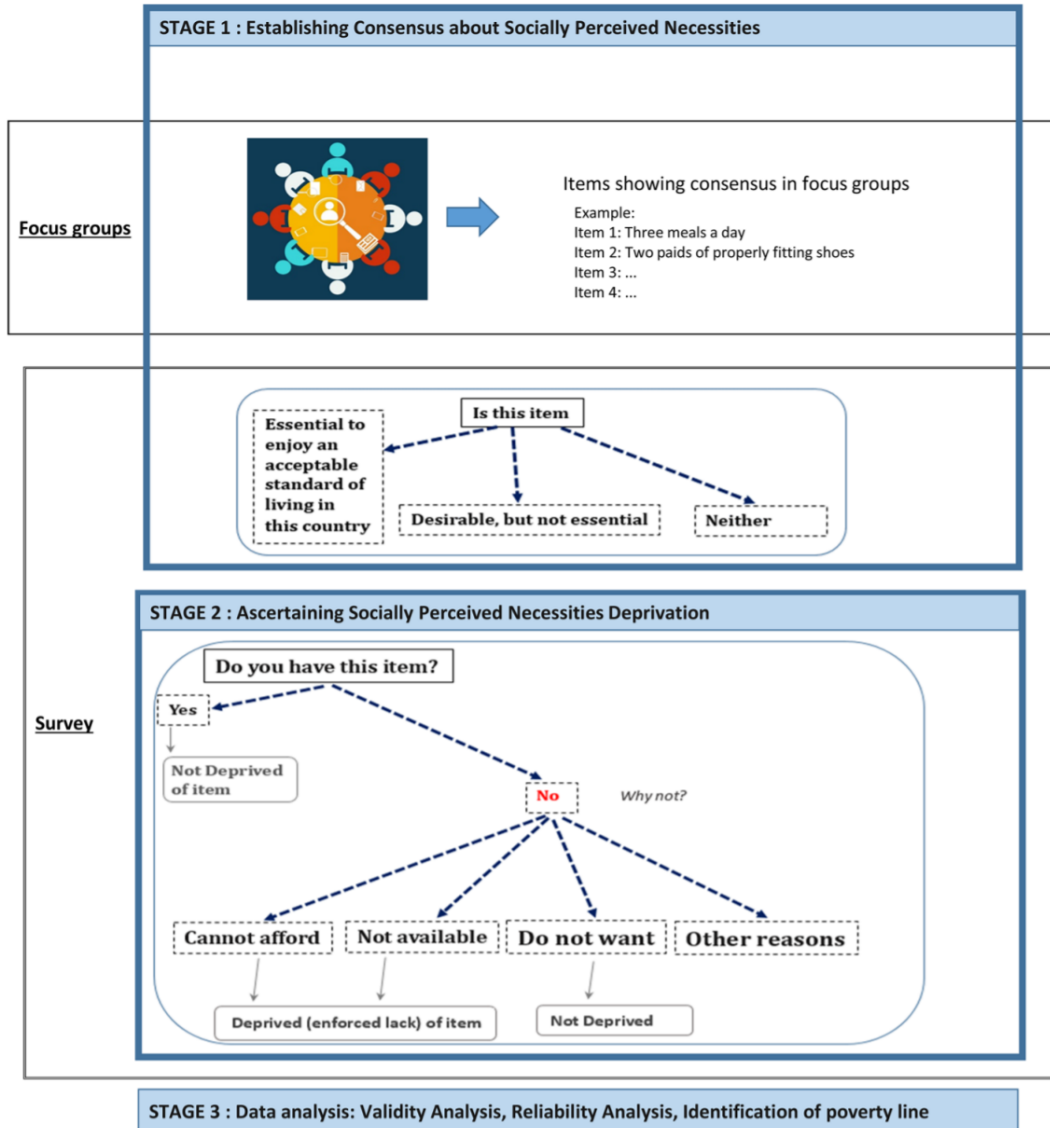


Table 2.5 Final child deprivation index components

<i>Relevant age range</i>	<i>Item</i>
0-17	A visit to the health facility when ill and all prescribed medication
0-17	Three meals a day
0-17	Two sets of clothing
0-17	Toiletries to be able to wash everyday
6-17	All fees, uniforms of correct size, and equipment
0-17	Own blanket
0-17	Own bed
0-17	Two pairs of properly fitting shoes
11-17	Own room for children over 10 of different sexes
3-17	Books at home for their age
0-17	Some new clothes
6-17	To be able to participate in school trips
6-17	Bus/taxi fare or other transport
6-17	A desk and chair for homework
6-17	Presents for children once a year on special occasions
3-12	Educational toys and games
0-17	To be able to make regular savings for emergencies
0-17	Repair a leaking roof for main living quarters
0-17	Repair or replace any worn out furniture
0-17	Replace broken pots and pans for cooking
0-17	Take children to a medical facility when sick
0-17	Pay school fees for children

Opciones de pre-modelado

Encuesta Nacional de Hogares de Uganda 2016/17 ☒ Censo de Uganda 2014

Opciones de pre-modelado:

¿Qué área pequeña? Subcondado a nivel nacional y parroquias de Kampala

- **Comparación de datos de encuestas y censos**
- - Post-estratificación aplicada a UNHS 2016 (subapresentación de los hogares rurales en la encuesta)
- - Definición inconsistente del tamaño del hogar y la cabeza del hogar en dos conjuntos de datos (variables marcadas/excluidas)
- **Límites:** Región > subregión > distrito > condado > subcondado > área de enumeración de > parroquiales. Varios cambios administrativos, muchos días dedicados a armonizar los límites de shapefile, Censo y encuesta.

Opciones de modelado: Explore, compare y elija el modelo de datos final de la encuesta

- Comenzó con modelos anidados básicos de regresión logística individual
-
- El conjunto candidato de variables independientes se basa en la literatura previa y el asesoramiento de UBOS
-
- Condujo al conjunto final de efectos fijos

Logistic regression models predicting individual-level poverty status. Log-odds	
	Model 1
Urban	-0.15 *
clothes deprivation	0.29 **
shoes deprivation	1.33 ***
roof deprivation	0.29 ***
wall deprivation	0.44 ***
Sanitation type (Flush toilet)	
Latrine	1.59 ***
Covered pit latrine	2.55 ***
Covered pit latrine with a slab	3.02 ***
Covered pit latrine without a slab	2.60 ***
Uncovered pit latrine with a slab	3.29 ***
Uncovered pit latrine without a slab	2.13 *
No facility	3.39 ***
Other	4.28 ***
tv deprivation	1.59 ***
Improved water	0.16 *
Number of children	0.27 ***
Overcrowding	0.52 ***
bicycle deprivation	0.74 ***
Household head working in subsistence agriculture	
Household head Illiterate	
N	15646
Nagelkerke R2	0.32
Specificity	0.75
Sensitivity	0.75

Opciones de modelado: Explore, compare y elija el modelo de datos final de la encuesta

Luego pasamos al modelado multinivel utilizando modelos de regresión logística de Bayes jerárquicos

- Modelo 1 sin incluir intercepciones a nivel de subregión.
- Modelo 1 + cabeza de familia que trabaja en actividades agrícolas de subsistencia + cabeza de familia que trabaja en actividades agrícolas de subsistencia (variables previamente marcadas como problemáticas)
- Modelo 1 + intercepciones de subregión.

Modelo 1 + intercepciones aleatorias a nivel de distrito

Modelo 1 + pendientes aleatorias a nivel de distrito

El modelo final de la encuesta

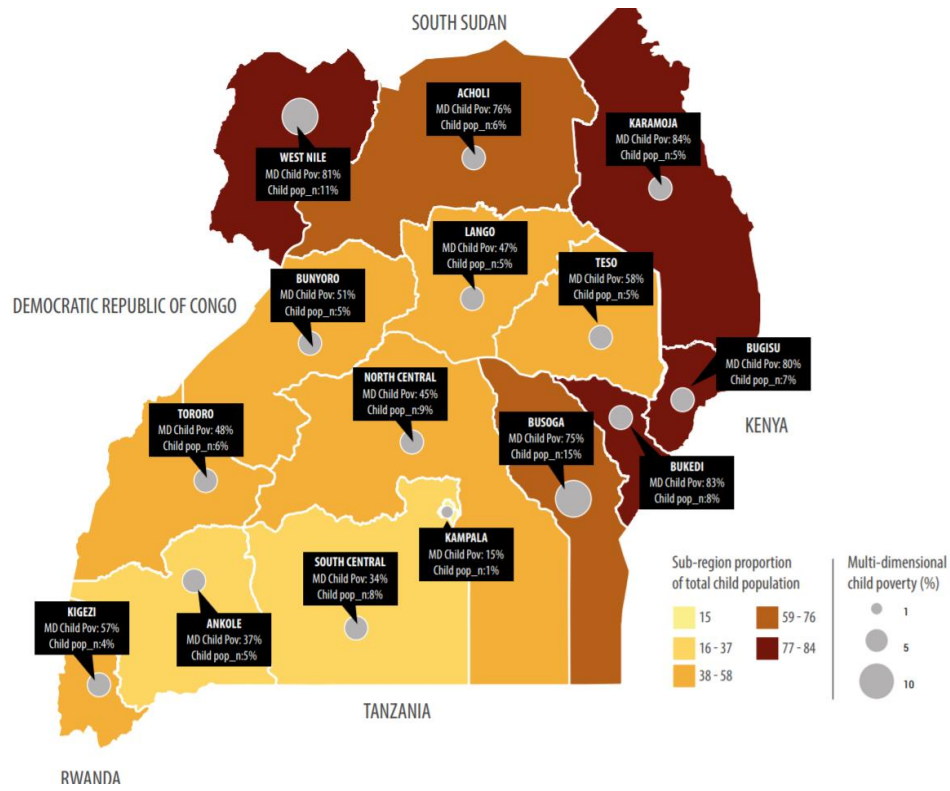
- Modelo 1 + intercepciones aleatorias a nivel de distrito

- Ajuste estadístico: la estadística CIAA (criterio de información ampliamente aplicable) de ajuste y Loo (validación cruzada de salida única para modelos bayesianos ajustados).
- Capacidad para reproducir las estimaciones puntuales subregionales, es decir, si el modelo reproduce los datos observados (estimaciones de diseño de la pobreza multidimensional).
-

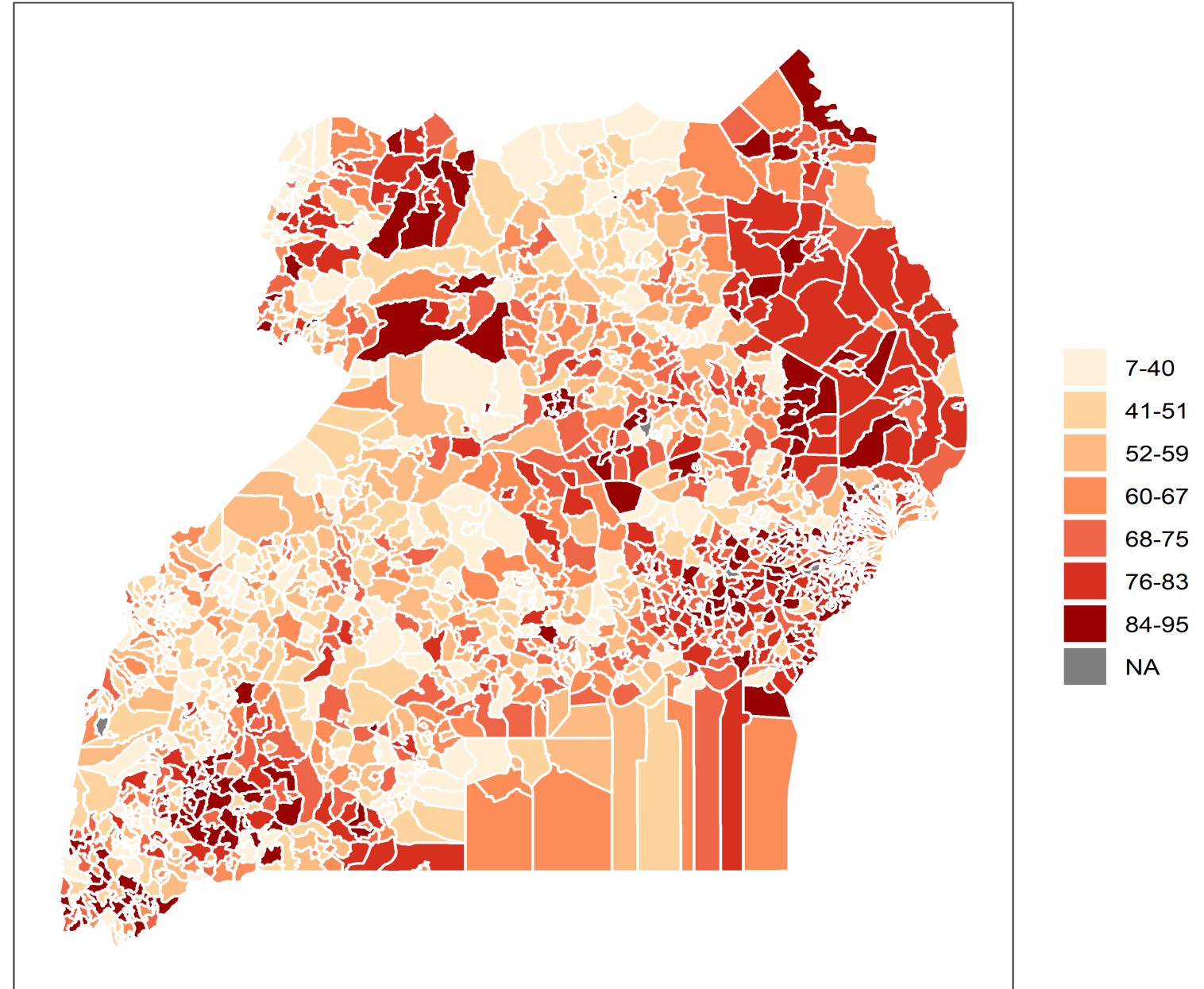
Validación

% Poor (multidimensional poverty)

SubRegion	Survey direct estimate	Survey: Model prediction	Census: Model Prediction
Kampala	9 [6-11]	9	8
Central 1	27 [24-29]	27	27
Central 2	38 [34-41]	38	39
Busoga	59 [56-63]	60	59
Bukedi	74 [69-78]	74	74
Bugishu	63 [59-68]	63	64
Teso	53 [49-57]	53	52
Karamoja	75 [70-80]	75	75
Lango	42 [39-46]	42	42
Acholi	61 [57-65]	61	59
Westnile	70 [66-74]	70	72
Bunyoro	41 [37-45]	41	41
Tooro	41 [37-45]	41	39
Ankole	29 [25-34]	29	31
Kigezi	50 [44-55]	50	52





% of children in multidimensional poverty (Subcounty SAE Estimates)



Conclusiones para Uganda 2016 (de próxima aparición)

- Las zonas del norte, en particular del noreste, tienden a tener tasas de pobreza multidimensional muy altas (por encima del 60%).
-
- Sin embargo, hay focos de alta pobreza en las subregiones que no parecen tener tasas de pobreza muy altas a nivel de subregión.
-
- Las áreas con los niveles más altos de necesidad también son aquellas con los niveles más bajos de infraestructura.
-

Conclusión

- Dedique tiempo a asegurarse de que los shapefiles sean compatibles entre los  y  los datos
-
- Utilice software estadístico que pueda manejar fácilmente varios tipos de conjuntos de datos al mismo tiempo (por ejemplo, R, Python)
-
- Evite confiar en funciones SAE preempaquetadas (por ejemplo, intente comprender el proceso usted mismo)
- Importancia de la información auxiliar para la validación cruzada
- Importancia de la pequeña superficie relevante
-

Reconocimientos

- Nada de este trabajo hubiera sido posible sin la contribución de Héctor E. Nájera Catalán, Shailen Nandy, Diego Angemi, David Gordon, Sheila Depio, Gemma Ahaibwe, Ibrahim Kasirye, James Muwonge, Vincent Ssenono, Stephen Baryahirwa, Baylon Twesigye, Sebnem Eroglu-Hawksworth, Eldin Fahmy, Acomo Oloya, Arthur Muteesasira, Sarah Kabaija, Vincent Ssenono, Beatrice Winnie Nyemera and Stephen Baryahirwa and others...



Bristol Poverty Institute
University of Bristol