Small Area Estimation of Food Insecurity based on FAO's Food Insecurity Experience Scale

Clara Aida Khalil, Statistician, Food and Agriculture Organization of the United Nations
06/06/2013

Background

The **Food and Agriculture Organization of the United Nations** (FAO) is the Custodian Agency of 21 SDG Indicators and the Contributing Agency of an additional 5.

The Organization contributed to the **IAEG-SDGs Working Group on Data Disaggregation and the Task Force on SAE.**

The FAO has done considerable **methodological work on data disaggregation and SAE** for the SDGs: https://www.fao.org/sustainable-development-goals/news/detail-news/en/c/1539866/

Based on this experience, **technical support and trainings** are being provided to selected countries on data disaggregation and indirect estimation methods for FAO's SDG indicators based on survey data.

SDG Indicator 2.1.2

2 ZERO HUNGER

Target 2.1

By 2030, end hunger and ensure access by all people, in particular the poor and people in vulnerable situations, including infants, to safe, nutritious and sufficient food all year round.

- **Indicador 2.1.1:** Prevalence of Undernourishment.
- **Indicador 2.1.2:** Prevalence of Moderate and Severe Food Insecurity in the population, based on the Food Insecurity Experience Scale (FIES).

The two indicators provide complementary information on food access which is one of the fundamental components of food security.

What is Food Security?

Food security exists when *all people, at all times, have physical, social and economic access to sufficient safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life*.

Food security is a ***multi-dimensional concept****:*

The Food Insecurity Experience Scale

Now I would like to ask you some questions about food.

During the last **12 months**, was there a time when...

You or others in your household were **worried you would not have enough food to eat** because of lack of money or other resources?

You or others in your household were **unable to eat healthy and nutritious food** because of lack of money or other resources?

You or others in your household **ate only few kinds of food** because of lack of money or other resources?

You or others in your household **had to skip a meal** because of lack of money or other resources?

You or others in your household **ate less than you thought you should** because of lack of money or other resources?

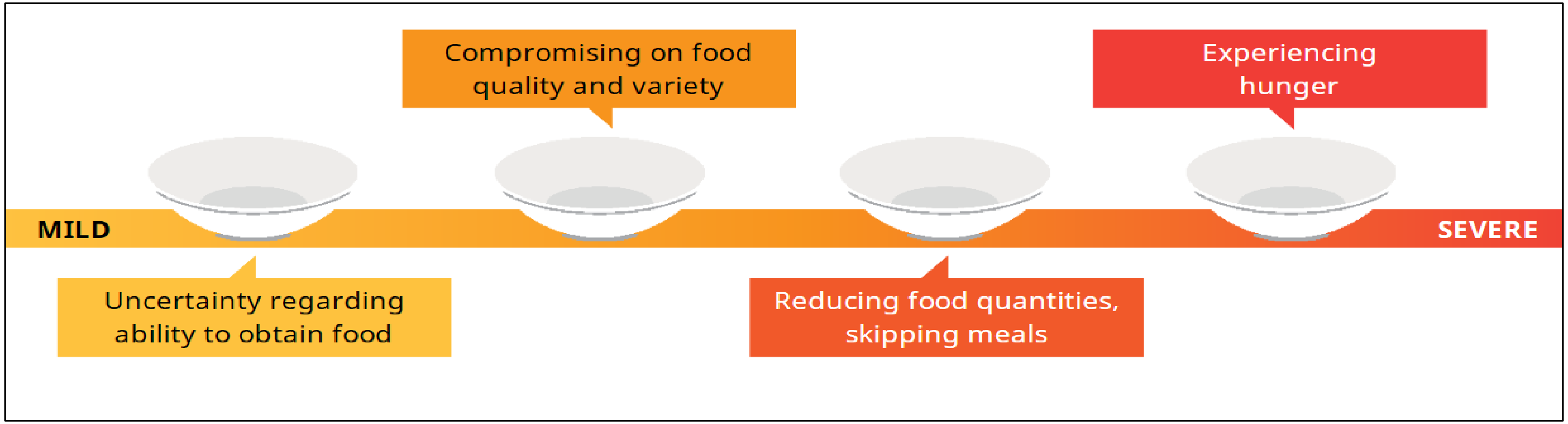
You or others in your household **run out of food** because of lack of money or other resources?

You or others in your household **were hungry but did not eat** because of lack of money or other resources?

You or others in your household **went without eating for a whole day** because of lack of money or other resources?

The Rasch Model to Estimate SDG Indicator 2.1.2

Fundamental assumption of the FIES: the severity of experienced food insecurity cannot be measured directly, but should be analyzed as a **latent trait** to be inferred from the observed data by applying the **Item Response Theory model**.



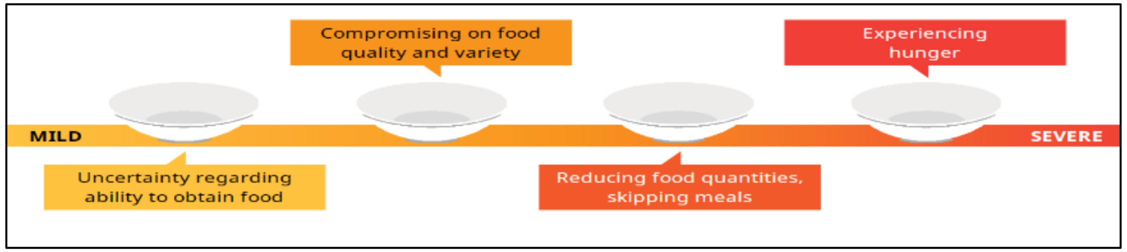
The Rasch Model (G. Rasch, 1960) or **single-parameter logistic IRT** model is used to estimate the **probability of being moderately or severely food insecure**.

The Rasch Model to Estimate SDG Indicator 2.1.2 (2)

Hypothesis of the Rasch Model:

I. The severity of the food insecurity condition of the respondent and that associated with each of the experiences can be located on the same one-dimensional scale.

II. Higher severity of the food insecurity condition of a respondent will increase the probability of reporting the occurrence of experiences associated with food insecurity (i.e. answering yes to questions in the FIES).



Model formulation: $Prob\left(X_{i,j}=1\right)=\frac{\exp\left(a_i-b_j\right)}{1+\exp\left(a_i-b_j\right)}$

Where:

- $Prob\left(X_{i,j}=1\right)$ is the probability of the $i-th$ individual to answer positively to item j;
- $\boldsymbol{a_i}$ **is the individual/household severity parameter**, i.e. the severity level of individual/household i;
- $\boldsymbol{b_j}$ **is the item severity parameter**, i.e. the severity level of the corresponding food insecurity experience associated to item j.

The Rasch Model to Estimate SDG Indicator 2.1.2 (3)

The Rasch Model can be read as follows: **the probability** that a household – with position on the severity scale determined by a_i - affirms an item with severity level b_j on the same scale **is a logistic function of the difference** $(\boldsymbol{a_i - b_j})$.

- The estimation of $Prob(X_{i,j} = 1)$ is based on the **maximum likelihood approach**. This process allows estimating both the individual/household severity parameters and the item severity parameters.
- Estimated probabilities are then averaged using **sampling weighs** to estimate the **Prevalence of Moderate and Severe Food Insecurity in the population** (SDG Indicator 2.1.2).
- **Important:** the standard error of SDG Indicator 2.1.2 is affected by two components: <u>the sampling error and the measurement error from the model</u>.

$$\boxed{SE_{tot} = \sqrt{(Sampling\ Error)^2 + (Measurement\ Error)^2}}$$

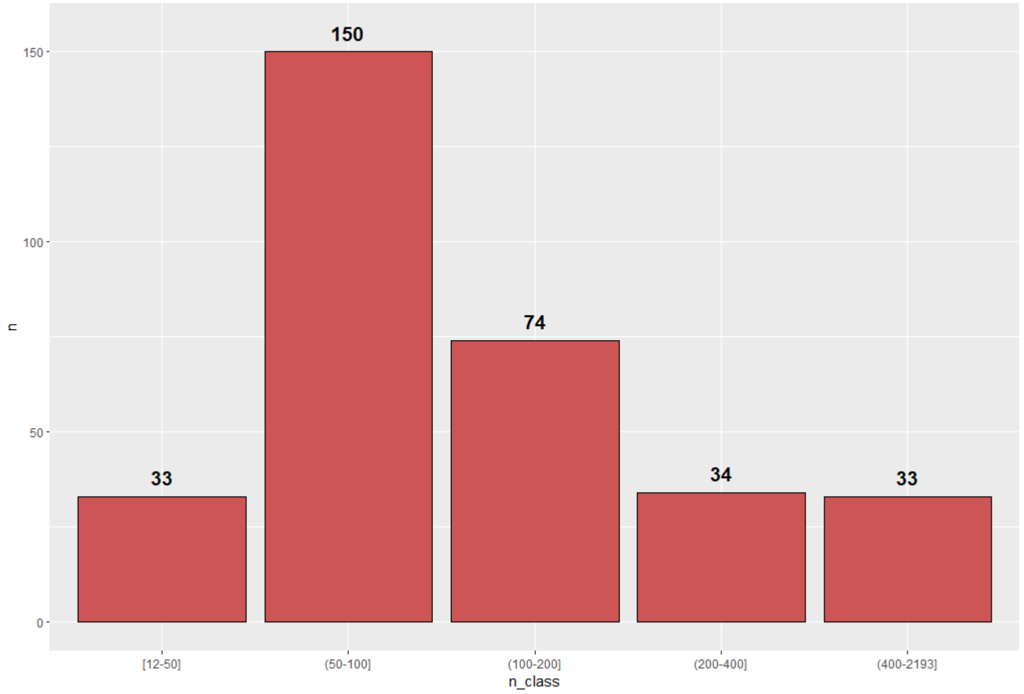
Estimation of SDG Indicator 2.1.2 in Chile with the CASEN

- The **FAO** is collaborating with the **Chilean Ministry of Social Development** and the **CEPAL** to produce **food insecurity maps based on SDG Indicator 2.1.2 at the level of Chile's municipalities (comunas)**.

- Direct estimation of Indicator 2.1.2 is performed with microdata collected through the **Encuesta de Caracterización Socioeconómica Nacional (CASEN) 2020**.

- **CASEN sampling design:** stratified two-stage design representative at the national and regional level, and for the two combinations national-urban and national-rural. The selected sample was not intended to be representative at the municipality (comuna) level. As a result, only 324 of the 346 Chilean comunas were included in the sample.

CASEN sampling size by comuna

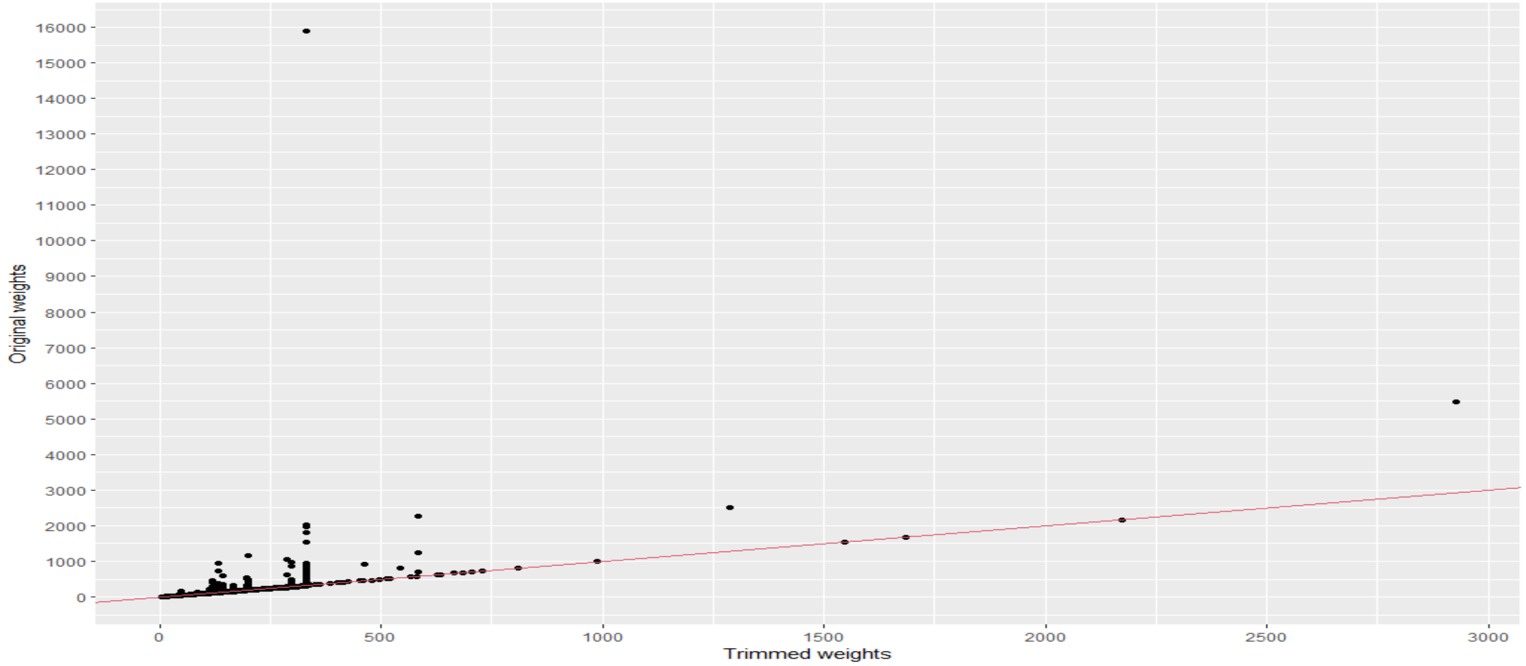


Trimming of sampling weights

- The sampling weights, induced by the original sampling design, may have extreme values due to weight adjustments (made for different reasons such as non-response). This can have an impact on the variance of the estimator in small areas.

- Therefore, **the sampling weights are trimmed using the Potter methodology (1993):** the optimal trimmed sampling weights are find by minimizing the following expression:

$$MSE(\hat{\theta}_{d,trimmed}) = Var(\hat{\theta}_{d,trimmed}) + (\hat{\theta}_{d,trimmed} - \hat{\theta}_d)^2$$



Small Area Estimation of SDG Indicator 2.1.2 in Chile

Selected SAE Method: We started considering an extension of the basic area-level SAE model – Fay-Herriot model (Fay and Herriot, 1979) – within the framework of the IRT modelling. Indeed, the uncertainty related to the implementation of the Rasch model needs to be accounted for in the MSE estimation of SAE. In this framework, **the direct estimator can be written as follows:**

$$\hat{\theta}_{d,dir} = \theta_d \quad + e_d + \nu_d$$

$$e_d \sim N\left(0, \sigma^2_{e_d}\right), \text{ and } \nu \sim N\left(0, \sigma^2_{\nu_d}\right)$$

Direct estimator with two independent sources of variation: one due to the sampling and one due to the Rasch model. The variances of this error are assumed to be known. The total variance of direct estimates can then be written as $\boldsymbol{\sigma^2_d} = \boldsymbol{\sigma^2_{e_d}} + \boldsymbol{\sigma^2_{\nu_d}}$

The parameter of interest, can be written as

$$\theta_d \quad = X_d^T \beta + u_d$$

$$u_d \sim N(0, \sigma^2_u)$$

Where β (regression parameter) and σ^2_u (variance of random effect) are unknown

Small Area Estimation of SDG Indicator 2.1.2 in Chile (2)

The combination of the sampling and linking models leads to the **area-level SAE model**:

$$\hat{\theta}_{d,dir} = X_d^T \beta_d + u_d + e_d + \nu_d$$

The area-level estimator is obtained considering the conditional expectation with respect to the random effect. Under the frequentist framework, the **Empirical Best Linear Unbiased Predictor (EBLUP)** can be expressed as a linear combination between the direct and synthetic estimator:

$$\hat{\theta}_{d,SAE} = \hat{\gamma}_d \hat{\theta}_{d,dir} + (1 - \hat{\gamma}_d) X_d^T \hat{\beta}_d$$

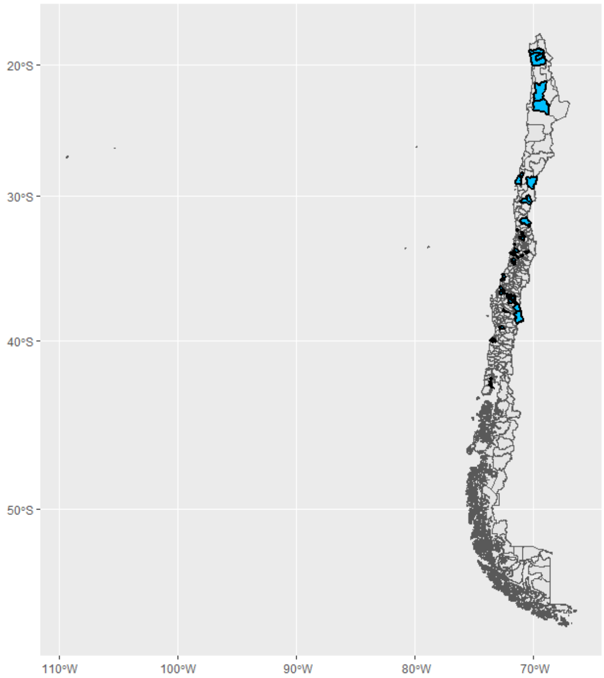
$$\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_d^2}$$

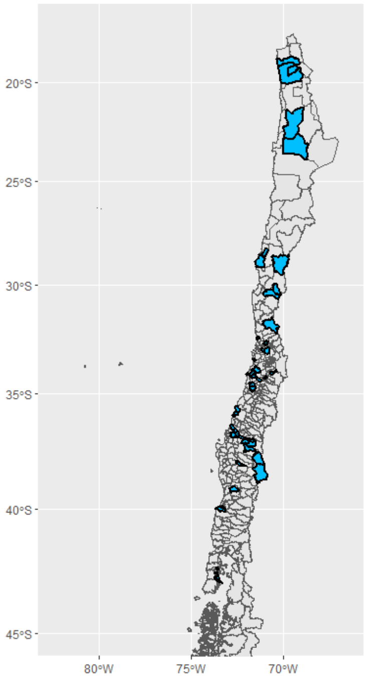
Criteria for the Selection of Comuna

Criteria to include direct estimates at municipality level in the SAE estimator:

- Degree of freedom (DF) ≥ 14 (DF=number of PSUs – number of strata); or
- $n \geq 50$ & DF>2 & $\rho \geq \min(\rho)$

where $\rho = (Deff - 1)/((n/npsu) - 1)$ and $\min(\rho) = 1 - \left[\frac{npsu}{npsu-1}\right]$





Smoothing of σ_d^2

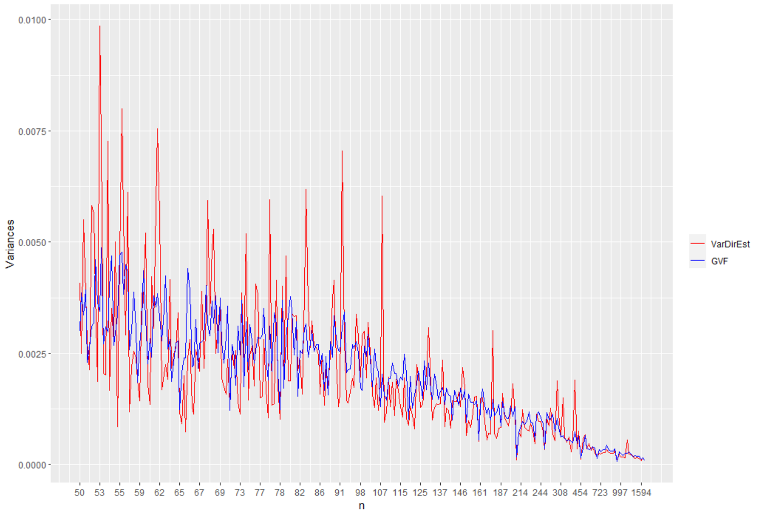
Need to apply smoothing techniques to stabilize the variance in small areas – e.g. method based on the **Generalized Variance Function** (GFV).

The GVF is a mathematical model describing the relationship between the variance of a survey estimator and its expectation. The following GVF was used in the exercise:

$$\log \hat{\sigma}_d^2 = z_d^T \alpha + e_d$$

where z_d are explanatory variables including the direct estimate, the sampling size, the interactions between the two, and their transformations. The bias of smoothed variances $\tilde{\sigma}_d^2$ was corrected applying a ratio adjustment with a factor equal to $\Delta = \frac{\sum \hat{\sigma}_d^2}{\sum \tilde{\sigma}_d^2}$. This resulted in:

$$\hat{\tilde{\sigma}}_d^2 = \exp(z_d^T \alpha)\, \Delta$$



Sources and Selection of Auxiliary Variables

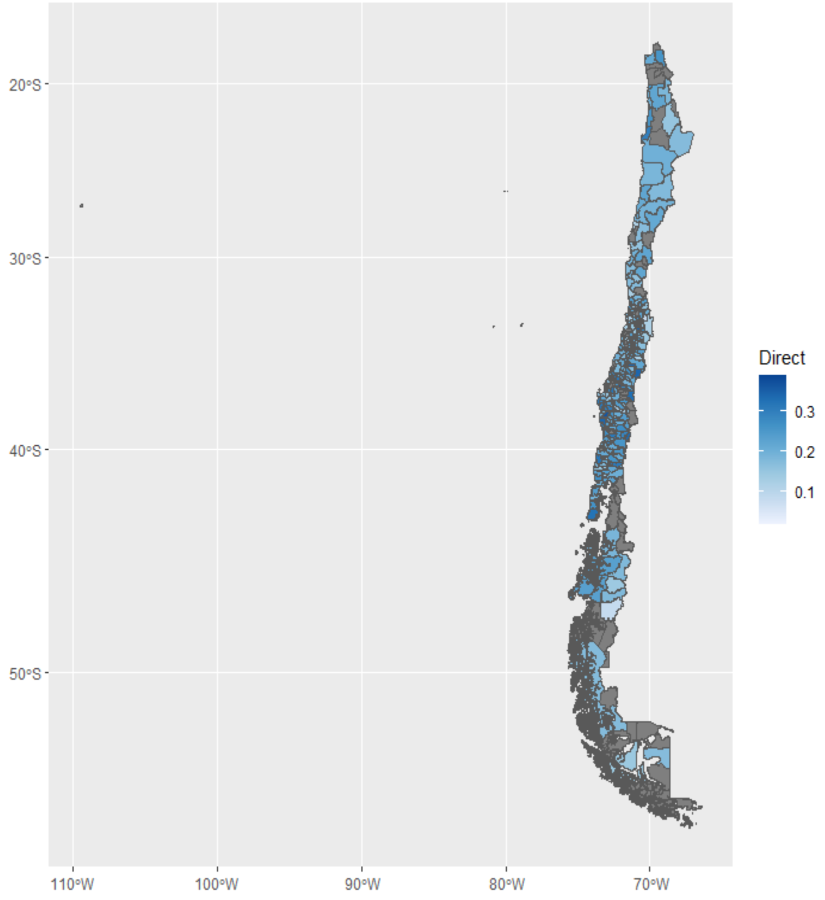
A list of 135 auxiliary variables were considered from: 1) The **Chilean Population and Housing Census** 2017; 2) A series of **Administrative registers**; 3) **Google Earth Engine** (night light intensity, crop coverage, human modification, distance from hospitals).

This initial list was reduced to avoid redundancies and multicollinearity. Finally, the set of variables to be used for the SAE model were selected with a **step-wise regression** using the **Akaike Information Criterion** (AIC) as selection criteria. This procedure lead to the selection of 14 auxiliary variables.

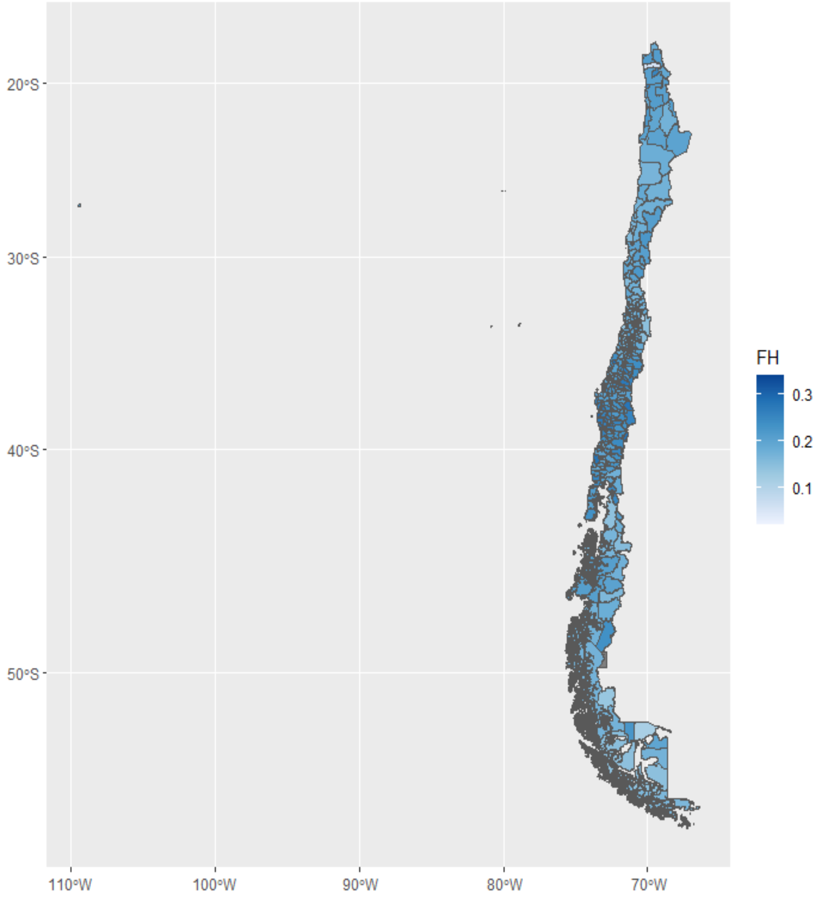
Variable Name	Description
prop_b50median_afc_2020	Percentage of formal wage earners with taxable income below 50% of median income
prop_fonasa_a_2019 prop_fonasa_b_2019 prop_fonasa_c_2019	Affiliation to FONASA (Public health system)
log_ing_municipales_permanentes_pc_2018	Municipal income: Own permanent income per capita
prop_obeso_sobrepeso_menores_2018_w prop_obeso_sobrepeso_menores_2018	Child nutritional status rate
prop_red_publica_2017 prop_rio_vertiente_estero_canal_2017 prop_camion_aljibe_2017	Proportion of dwellings by water source
prop_am_obeso_2018 prop_am_normal_2018 prop_am_bajo_peso_2018	Elderly nutritional status rate
prop_ism_afc_2020	Percentage of formal wage earners with taxable income equal to the minimum income

Preliminary Results

Direct Estimates

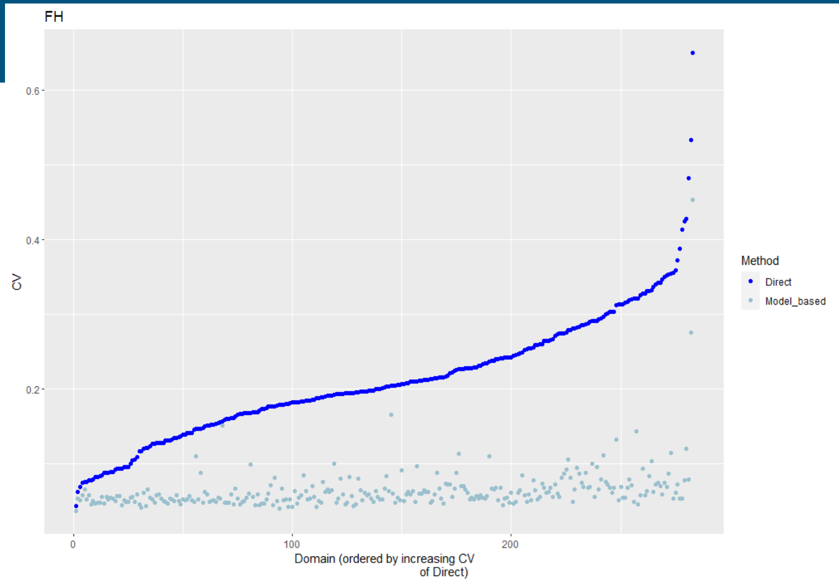


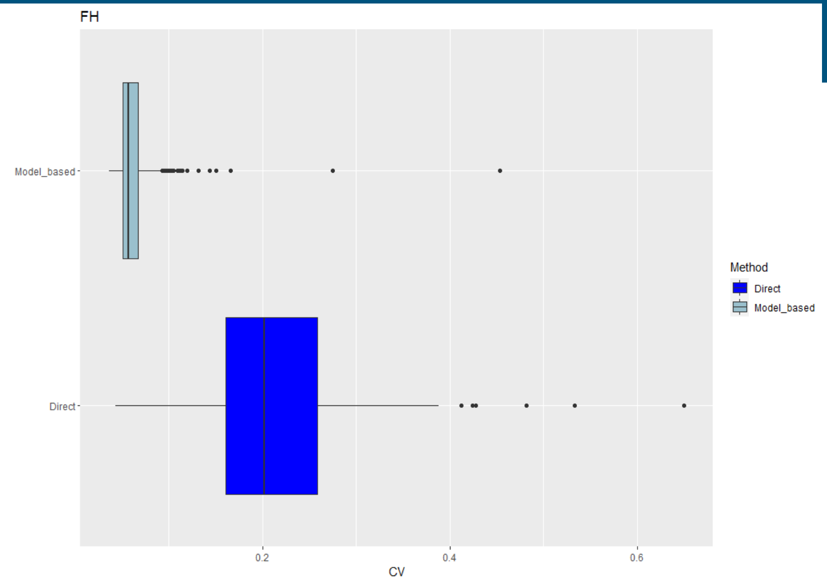
FH Estimates

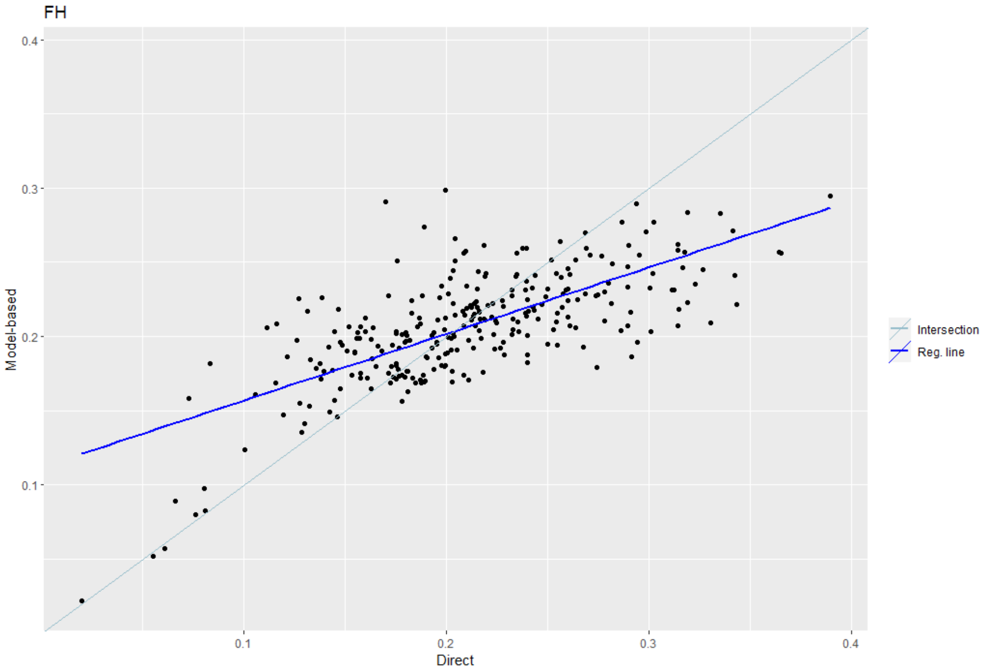


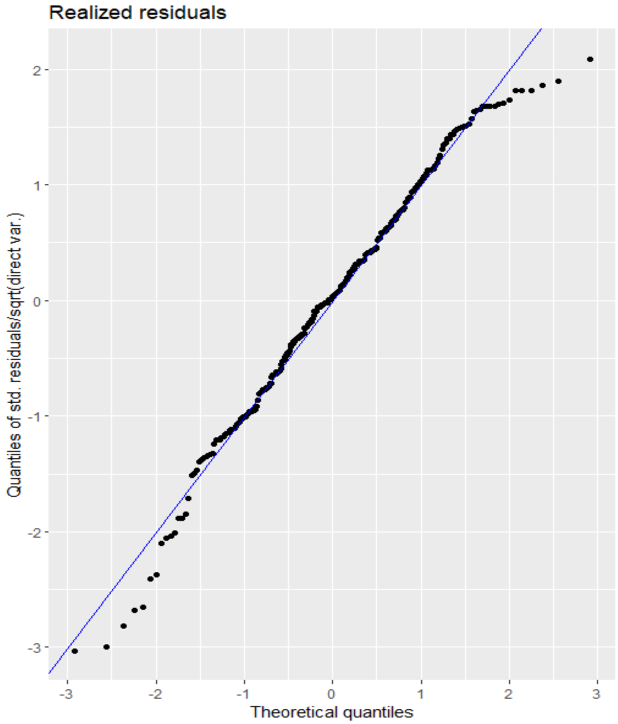
Preliminary Results(2)

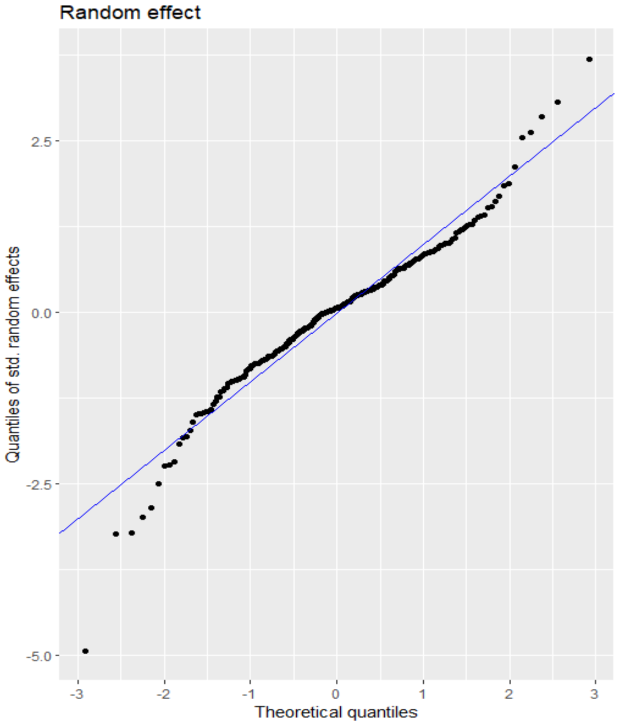
Model assessment and validation











Way Forward and Future Work

Discussed study still ongoing and multiple aspects need to be finetuned: 1) Benchmarking; 2) Assessing the performance of the model with suitable transformation of y; 3) Improve variance smoothing; 4) Compare results obtained adopting a Bayesian approach.

SUSTAINABLE DEVELOPMENT GOALS

The **FAO Regional Office for Latin America and the Caribbean** is now engaging with **Dominican Republic and a third country in the region** (yet to be identified) to implement similar approaches for SDG Indicator 2.1.2.

Looking ahead: The methodology developed thanks to the collaboration with these countries could be used to produce food insecurity maps in all countries implementing the FIES and having suitable sources of auxiliary information.

Thank you!

ClaraAida.Khalil@fao.org