

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

Escuela de Posgrado



Un modelo Fay-Herriot espacial para la predicción del porcentaje de niños con anemia y riesgo de retraso del crecimiento en distritos no encuestados y en distritos con pocas observaciones disponibles

Tesis para obtener el grado académico de Magíster en Estadística que
presenta:

Marcial Eduardo Haro Abanto

Asesora:

Anna Sikov

Lima, 2022

Declaración jurada de autenticidad

Yo, Anna Sikov, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis titulada *Un modelo Fay-Herriot espacial para la predicción del porcentaje de niños con anemia y riesgo de retraso del crecimiento en distritos no encuestados y en distritos con pocas observaciones disponibles* del autor Eduardo Haro Abanto, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 14%, lo que está dentro del límite establecido. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 29/11/2022.
- He revisado con detalle dicho reporte y confirmo que cada una de las coincidencias detectadas no constituyen plagio alguno.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar

y

fecha:

Lima, 29 de noviembre de 2022

Apellidos y nombres de la asesora: <u>Sikov, Anna</u>	
DNI: 49064306	Firma: 
ORCID: 0000000198695952	

Resumen

La presente tesis tiene como objetivo realizar predicciones del nivel de prevalencia de anemia y riesgo de retraso del crecimiento en niños menores de 5 años, en distritos encuestados con observaciones faltantes y distritos no encuestados con niveles de pobreza mayor a 55% del país, con datos de la Encuesta Demográfica de Salud Familiar (ENDES) para el año 2019 y el censo nacional del 2017. Para alcanzar este objetivo se hace uso del modelo Fay-Herriot y su variante espacial. Este modelo permite obtener estimaciones fiables en dominios en los cuales el diseño muestral no tiene un nivel de inferencia aceptable.

El uso de variables auxiliares obtenidas del censo a nivel de distrito permite implementar el modelo Fay-Herriot, obteniendo predicciones con un error de estimación menor a las estimaciones muestrales. En el caso de la variante espacial del modelo Fay-Herriot, este incluye correlaciones entre los efectos aleatorios de áreas vecinas. Además, se estima el error cuadrático medio de las predicciones obtenidas mediante los métodos de bootstrap paramétrico y no paramétrico.

Los resultados muestran que los distritos de los departamentos de Puno, Cusco y Huancavelica tienen indicadores de anemia en niños menores de 5 años que superan el 48.5% en tanto que los distritos de los departamentos de Puno, Huancavelica y Tacna tienen indicadores de riesgo de retraso del crecimiento en niños menores de 5 años por encima del 45.5%.

Palabras clave: Modelo Fay-Herriot, anemia infantil, riesgo de retraso del crecimiento infantil, estimación en áreas pequeñas, ENDES.

Abstract

The objective of this thesis is to make predictions of the level of prevalence of anemia and risk of stunted growth in children under 5 years of age, in surveyed districts with missing observations and non-surveyed districts with poverty levels greater than 55% of the country, with data from the Family Health Demographic Survey (ENDES) for the year 2019 and the national census of 2017. To achieve this objective, the Fay-Herriot model and its spatial variant are used. This model makes it possible to obtain reliable estimates in domains in which the sample design does not have an acceptable level of inference.

The use of auxiliary variables obtained from the census at the district level makes it possible to implement the Fay-Herriot model, obtaining predictions with an estimation error smaller than the sample estimates. In the case of the spatial variant of the Fay-Herriot model, it includes correlations between the random effects of neighboring areas. In addition, the mean square error of the predictions obtained using the parametric and non-parametric bootstrap methods is estimated.

The results show that the districts of the departments of Puno, Cusco and Huancavelica have indicators of anemia in children under 5 years of age that exceed 48.5%, while the districts of the departments of Puno, Huancavelica and Tacna have indicators of risk of stunted growth in children under 5 years of age above 45.5%.

Keywords: Fay-Herriot model, anemia in children, risk of stunted growth in children, small area estimation, ENDES.

Índice

1. Introducción	5
1.1. Consideraciones preliminares	5
1.2. Revisión de literatura	6
1.3. Objetivos	9
1.4. Organización del trabajo	9
2. Estimadores muestrales	10
2.1. Estimadores basados en el diseño	10
2.1.1. Estimadores directos	10
2.2. Estimadores basados en el modelo	11
2.2.1. Estimadores sintéticos	11
2.2.2. Estimadores compuestos	11
2.3. Modelos de áreas pequeñas	11
3. Modelos lineales mixtos	12
3.1. Modelo lineal mixto	12
3.2. Modelo lineal mixto para la estimación en áreas pequeñas	13
4. Modelamiento de datos espaciales	15
4.1. Matrices de pesos espaciales	15
4.1.1. Matrices espaciales basadas en distancias de centroide	15
4.1.2. Matrices espaciales basadas en límites	16
4.2. Autocorrelación espacial	16
4.2.1. Estadístico I de Moran	16
4.2.2. Estadístico C de Geary	17
4.3. Modelos de regresión espacial	17
4.3.1. Modelo autorregresivo simultáneo	17
4.3.2. Modelo condicional autorregresivo	18
5. Modelo Fay-Herriot	19
5.1. Modelo Fay-Herriot	19
5.2. Modelo Fay-Herriot espacial	20
5.3. Estimador BLUP y EBLUP	20
5.4. Estimación de σ_v^2	22
5.4.1. Método de momentos	22
5.4.2. Métodos de máxima verosimilitud	22
5.5. Aproximación analítica para el error cuadrático medio del estimador EBLUP y SEBLUP	23
5.5.1. Modelo Fay-Herriot	23
5.5.2. Modelo Fay-Herriot espacial	25
5.6. Estimación del error cuadrático medio del estimador EBLUP y SEBLUP mediante bootstrap	26
5.6.1. Modelo Fay-Herriot	26
5.6.2. Modelo Fay-Herriot espacial	27
6. Aplicaciones	30
6.1. Prevalencia en anemia en niños menores de 5 años	30
6.1.1. Estimación del estimador Horvitz-Thompson	30
6.1.2. Modelo Fay-Herriot	30
6.1.3. Índice de Moran y Geary	32
6.1.4. Modelo Fay-Herriot espacial	32
6.1.5. Predicción del porcentaje de niños menores a 5 años con anemia	34
6.1.6. Estimación del error cuadrático medio de las predicciones del porcentaje de niños menores a 5 años con anemia	37
6.1.7. Mapas de prevalencia de anemia: Modelo Fay-Herriot	38
6.1.8. Mapas de prevalencia de anemia: Modelo Fay-Herriot espacial	40
6.2. Riesgo de retraso del crecimiento en niños menores de 5 años	42
6.2.1. Estimación del estimador Horvitz-Thompson	42

6.2.2.	Modelo Fay-Herriot	42
6.2.3.	Índice de Moran y Geary	44
6.2.4.	Modelo Fay-Herriot espacial	44
6.2.5.	Predicción del porcentaje de niños menores a 5 años con riesgo de retraso del crecimiento	45
6.2.6.	Estimación del error cuadrático medio de las predicciones del porcentaje de niños menores a 5 años con riesgo de retraso del crecimiento	48
6.2.7.	Mapas de riesgo de retraso del crecimiento: Modelo Fay-Herriot	49
6.2.8.	Mapas del riesgo de retraso del crecimiento: Modelo Fay-Herriot espacial	51
7.	Conclusiones	53
7.1.	Conclusiones	53
7.2.	Sugerencias para investigaciones futuras	53
8.	Referencias	54
9.	Anexos	59
9.1.	Anexo A	59
9.1.1.	Prevalencia de anemia en niños menores a 5 años	59
9.1.2.	Riesgo de retraso del crecimiento de niños menores a 5 años	60
9.2.	Anexo B	61
9.2.1.	Prevalencia de Anemia:Estimaciones	61
9.2.2.	Riesgo de retraso del crecimiento:Estimaciones	79



1. Introducción

1.1. Consideraciones preliminares

En la actualidad es cada vez más importante la generación de estadísticas de alta calidad para la toma de decisiones tanto en el sector público como privado. La mayor parte de estas estadísticas provienen de encuestas que tienen representatividad en áreas como departamentos o regiones lo cual limita el uso de estas a niveles más desagregados como provincias o distritos. Dada la importancia de tener estadísticas más precisas para áreas con menor extensión es que surgen los modelos de estimación en áreas pequeñas, modelos que permiten combinar información de censos y encuestas para obtener información precisa de subconjuntos de la población en dominios no planificados en el diseño muestral.

Según Rao & Molina (2015) la demanda por estimaciones confiables en áreas pequeñas se ha venido incrementando debido, entre otras cosas, a su creciente uso en la formulación de políticas y programas en actores como Estados Unidos, la Unión Europea y en el sector privado debido a que gran parte de las decisiones comerciales dependen en gran medida de las condiciones de una localidad o distrito, en lugar de áreas más agregadas como una región o departamento.

Entidades gubernamentales como la Oficina Europea de Estadística, más conocida como Eurostat, y la Oficina del Censo de los Estados Unidos utilizan este tipo de metodología para la publicación de estadísticas oficiales¹². En esa misma línea, investigaciones como la de Romero (2018) en la que se estima la tasa de embarazos adolescentes en Colombia a nivel municipal demuestra las ganancias en términos de precisión gracias al uso de estimaciones en áreas pequeñas, frente a las realizadas bajo el enfoque del diseño muestral. Garrudo et al. (2018) en su estudio sobre prevalencia de anemia en la zona oriental de Cuba estiman bajo el enfoque de áreas pequeñas, un modelo multinivel que permite obtener estimaciones de niveles de anemia en niños menores de dos años a nivel municipal y además obtener qué factores de riesgos son los que conllevan a desarrollar esta enfermedad.

Para el caso peruano, el uso de técnicas para la estimación en áreas pequeñas aún no es de uso extendido. Publicaciones recientes como el mapa de pobreza provincial y distrital en el Perú realizado por el Instituto Nacional de Estadística e Informática (INEI)³ y el trabajo realizado por Escobal y Armas (2015) para la estimación del ingreso neto agropecuario son algunos ejemplos del creciente interés por desarrollar este tipo de metodología.

Sin embargo, aún existen una variedad de indicadores sociales que no tienen una cobertura adecuada lo cual dificulta una mejor implementación de políticas por parte de diferentes actores del sector público y privado. Los indicadores de porcentaje de niños con anemia y retraso del crecimiento son ejemplo de ello.

Según la Organización Mundial de la Salud (OMS), la anemia en la niñez es un trastorno en la capacidad de transporte de oxígeno en la sangre lo cual dificulta satisfacer las necesidades del organismo teniendo efectos inmediatos como fatiga, debilidad y retraso de crecimiento.⁴ A su vez, el retraso del crecimiento en la niñez, producto de la prevalencia de desnutrición crónica, se define como una estatura inferior en más de dos desviaciones típicas a la mediana de los patrones de crecimiento infantil de la OMS teniendo efectos a largo plazo como mala salud, disminución de la capacidad cognitiva entre otros. El riesgo de retraso del crecimiento amplía esta definición considerando a los niños vulnerables a desarrollar esta enfermedad, los cuales se encuentran por debajo de 1.28 desviaciones estándar de la distribución del indicador Talla/Desviación estándar.⁵

Para el caso de Perú, según el documento técnico *"Plan Nacional para la Reducción y Control de la Anemia Materno Infantil y la Desnutrición Crónica Infantil en el Perú, 2017-2021"* desarrollado por el Ministerio de Salud (MINSA) la anemia afecta al 43.6% de los niños y niñas de 6 a 36 meses de edad, siendo más usual entre los niños de 6 a 18 meses, grupo en donde 6 de cada 10 niños presenta esta enfermedad. Por otro lado, el retraso del crecimiento producto de la prevalencia de desnutrición crónica, afecta al 13.1% de niños menores a 5 años a nivel nacional.

¹Ver https://ec.europa.eu/eurostat/cros/content/small-area-estimation-0_en

²Ver <https://www.census.gov/topics/research/stat-research/expertise/small-area-est.html>

³Ver https://www.inei.gob.pe/media/MenuRecursivo/boletines/mapa-de-pobreza-25022020_ponencia.pdf

⁴Para mayor información acerca del concepto de anemia, ver <https://www.who.int/health-topics/anaemia>

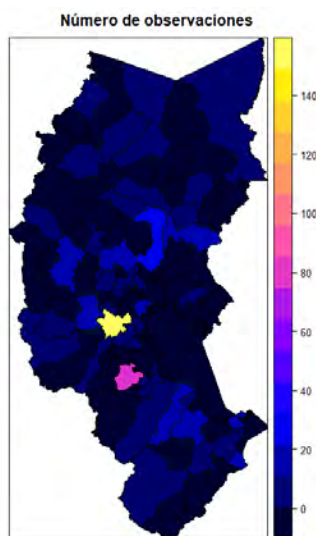
⁵Para mayor información acerca de los efectos de la desnutrición, ver <https://www.who.int/health-topics/malnutrition>

Sin embargo, los indicadores señalados anteriormente, recogidos en la Encuesta Demográfica de Salud Familiar (ENDES), solo tienen inferencia a nivel nacional, nacional urbana, nacional rural y por región natural lo cual limita el análisis de estrategias para la reducción de los niveles de anemia y riesgo de retraso del crecimiento en niños a nivel provincial o distrital.⁶

Esta investigación tratará de dar solución a esta falta de información, para ello se utilizará la ENDES para el año 2019 la cual recoge datos de anemia y retraso del crecimiento en niños y niñas menores de 5 años, y el censo nacional realizado el año 2017. Cabe resaltar que en esta encuesta hay algunas provincias o distritos que no fueron incluidos en la muestra e inclusive algunos no cuentan con suficientes observaciones para realizar ejercicios de inferencia.

Como ejemplo de ello tenemos al departamento de Puno. En la figura 1 se muestra que solo el 41.4% de distritos tiene observaciones suficientes para estimar datos de prevalencia de anemia, siendo los distritos de Juliaca y Puno, aquellos que presentan un mayor número de encuestados (más de 70 observaciones) en tanto que el resto de distritos como Lampa o Ilave, tienen en promedio menos de 15 observaciones.

Figura 1: Puno: Número de observaciones recogidas en la ENDES 2019



Para ello se propone la construcción del modelo de Fay-Herriot espacial desarrollado por Rao y Molina (2015) basado en la metodología de estimación en áreas pequeñas y la aplicación de modelos lineales mixtos. Además se estimará el error de las predicciones obtenidas mediante el método de remuestreo bootstrap.

1.2. Revisión de literatura

Los modelos de estimación en áreas pequeñas (*small area estimation*) son modelos que permiten obtener estimaciones robustas de subgrupos de la población como es el caso de regiones geográficas pequeñas (provincia, distrito, etc) o grupos socio-demográficos (población por estrato socio-económico, población por raza, etc) . En estos subgrupos de la población es probable que la estimación directa de algún estadístico, basado solo en los datos del subgrupo, no sea confiable dado que no se dispone con el número suficiente de observaciones para realizar inferencia estadística. Es en este contexto que el uso de modelos de estimación en áreas pequeñas logran obtener mayor importancia.

Gosh y Rao (1994) recalcan la importancia de estos modelos debido a que reducen los errores de estimación producidos por el incorrecto uso de estimadores directos para la estimación en áreas pequeñas. Pfeffermann (2013) realiza una clasificación de los modelos que utilizan esta metodología. Por un lado se tiene los métodos basados en el diseño (*design based*) y los basados en el modelo (*model-based*). Los métodos basados en el diseño construyen un modelo para los estimadores pero la varianza, el sesgo y otras propiedades del estimador son

⁶Para mayor información acerca del nivel de inferencia, ver <https://proyectos.inei.gob.pe/endes/>

derivadas de la aleatorización del diseño muestral, en cambio los métodos basados en el modelo incorporan y vinculan la información de áreas vecinas con modelos adecuados para los estimadores directos de las áreas pequeñas. Un rasgo común de ambos métodos residen en el uso de información auxiliar proveniente de datos administrativos o censales con la finalidad de tener un mejor poder predictivo para las estimaciones en áreas pequeñas.

Entre los métodos basados en el modelo se encuentran el modelo de nivel de área (*area level model*) desarrollado por Fay y Herriot (1979), el modelo de nivel de unidad con error anidado (*Nested error unit level model*) propuesto por Battese, Harter y Fuller (1988) y el modelo mixto logístico (*Mixed logistic model*) generalizado por MacGibbon y Tomberlin (1989). Todos estas propuestas se basan en los supuestos del modelo lineal mixto que permite incorporar variabilidad entre las observaciones a través de efectos fijos o aleatorios. Los efectos fijos asumen que existe una relación constante entre la variable de interés y las variables explicativas en tanto que los efectos aleatorios asume que estas relaciones constantes varían entre observaciones. En particular, el modelo lineal mixto con efectos aleatorios es sumamente utilizado en estimaciones en áreas pequeñas debido a que permite capturar la variabilidad entre observaciones de manera más precisa y permite mejorar las estimaciones derivadas de encuestas mediante la incorporación de datos de covariables de otras fuentes. Uno de los modelos más utilizados es el modelo Fay-Herriot, es en este tipo de modelo que, para realizar la estimación de los efectos aleatorios, es necesario tener información acerca de la estructura de la varianza de estos efectos. Conocida la estructura, se puede obtener las predicciones de este modelo derivando el mejor predictor lineal insesgado (BLUP por sus siglas en inglés), el cual minimiza el error cuadrático medio entre todos los estimadores lineales insesgados.

Segun Sugawara y Kubokawa (2020) el estimador BLUP desarrollado en los modelos de áreas pequeñas se caracteriza por agrupar y reducir las estimaciones directas provenientes de encuestas, provocando un efecto de contracción (*shrinkage effect*), donde las estimaciones directas se reducen a la media de la regresión.

Consideremos el siguiente modelo Fay-Herriot para estimar la media θ_i basados en estadísticas a nivel de área en los estimadores directos y_i para $i = 1, \dots, m$, también considerados como estimadores directos y \mathbf{x}_i el cual es un vector de covariables

$$y_i = \theta_i + \epsilon_i, \theta_i = \mathbf{x}_i^t \boldsymbol{\beta} + v_i, i = 1, \dots, m$$

donde $\boldsymbol{\beta}$ es un vector de coeficientes de regresión, ϵ_i es un error de muestreo y v_i son efectos aleatorios, $\epsilon_i \sim N(0, \delta_i)$ y $v_i \sim N(0, \tau)$. Donde δ_i es la varianza de y_i dado θ_i , la cual es conocida y τ es la varianza del efecto aleatorio, la cual es desconocida.

De la ecuación anterior se deriva el mejor predictor lineal para θ_i , el cual se obtiene considerando la función de pérdida del error cuadrático medio

$$E[\theta_i | y_i] = \psi y_i + (1 - \psi) \mathbf{x}_i^t \boldsymbol{\beta},$$

donde $\psi_i = \frac{\tau}{\tau + \delta_i}$ es conocido como el coeficiente de contracción (*shrinkage coefficient*) el cual contiene la estructura de la varianza de los efectos aleatorios y de los errores de muestreo. Si estimamos el vector de coeficientes de regresión mediante el método de Mínimos Cuadrados Generalizados (MCG), podemos reemplazar $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}_{MCG}$ obteniendo así el estimador BLUP⁷

$$\tilde{\theta}_i = \psi_i y_i + (1 - \psi_i) \mathbf{x}_i^t \hat{\boldsymbol{\beta}}_{MCG}$$

donde $\tilde{\theta}_i$ es el estimador BLUP del parámetro de interés θ_i , y_i es el estimador directo del parámetro de interés θ_i para todo $i = 1, \dots, m$ ⁸, y $\hat{\boldsymbol{\beta}}_{MCG}$ es el estimador de la regresión estimado mediante MCG.

Es así que el efecto de contracción del estimador BLUP mencionado líneas arriba va a depender del valor que tome ψ_i , el cual va a reducir la influencia del estimador directo, el cual es impreciso, hacia el estimador derivado del modelo. Cabe resaltar que en la práctica, la varianza de los efectos aleatorios (τ) no es conocida y

⁷En la subsección 4.3 del documento se realiza un desarrollo más a detalle acerca de la derivación del estimador BLUP en modelos de áreas pequeñas

⁸Nótese que el estimador y_i tiene una varianza muy alta debido a que proviene de una muestra pequeña, con lo cual puede obtenerse estimaciones imprecisas

por lo tanto se estima mediante diversos métodos como el método de momentos o máxima verosimilitud (Harmenting et al.2020). Luego de ser estimada y reemplazada en ψ_i se obtiene el mejor predictor lineal insesgado empírico (EBLUP por sus siglas en inglés).

En particular, el modelo Fay-Herriot es un modelo lineal mixto que estima un parámetro de interés basado en estadísticas a nivel de área. Además incorpora errores de muestreo y efectos aleatorios que permiten capturar la heterogeneidad entre áreas de manera más precisa .

Entre algunas de las aplicaciones de este modelo se encuentran el trabajo desarrollado por Benavent y Morales (2016) en el cual estiman la proporción y brechas de pobreza a nivel provincial en España utilizando la Encuesta de Condiciones de Vida para los años 2005 y 2006, Xie, Raghunathan y Lepkowski(2007) estiman la proporción de personas con sobrepeso en Estados Unidos utilizando datos del sistema de vigilancia de factores de riesgo conductual (BRFSS por sus siglas en ingles) del año 2003 y datos censales del año 2000 y por último, Ubaidillah, Notodiputro, Kurnia y Mangku (2019) estiman el consumo y gasto per cápita en los distritos de Indonesia utilizando el modelo de Fay-Herriot multivariado.

Sin embargo, el modelo básico de Fay-Herriot asume efectos aleatorios distribuidos independientemente en cada área lo cual puede no ser un supuesto que se cumpla en la realidad. Puede existir ciertas correlaciones entre áreas vecinas que pueden enriquecer aún más el análisis realizado con la introducción de efectos aleatorios. Usualmente, las predicciones de la variable de interés en áreas vecinas son similares debido a covariables que no han sido medidas y pueden estar asociadas al resultado de interés (Wu et al., 2021) y además, las variables a considerar en un modelo pueden variar en el espacio considerando así una posible interacción espacial con la variable de interés (Pratessi y Salvati, 2008).

La incorporación de un término de autocorrelación espacial y una matriz de contigüidad espacial permite incluir una interacción espacial entre la variable de interés y las covariables en un modelo, facilitando el modelamiento de la dependencia espacial entre áreas vecinas. Al igual que en el modelo Fay-Herriot estándar, en la variante espacial se deriva un mejor predictor lineal insesgado espacial (SBLUP por sus siglas en inglés). También no se tiene información acerca de la varianza de los efectos aleatorios y del parámetro de autocorrelación espacial por lo que se debe utilizar métodos de estimación como el de máxima verosimilitud o máxima verosimilitud restringido. Luego de ser estimados estos parámetros y reemplazados se obtiene el mejor predictor lineal insesgado empírico espacial (SEBLUP por sus siglas en inglés)⁹

Entre algunas de las aplicaciones de esta variante del modelo Fay-Herriot se encuentra el trabajo desarrollado por Porter,Holan,Wikle y Cressie(2014) en donde incorporaran un término de autocorrelación espacial al modelo Fay-Herriot, con la finalidad de analizar cambios relativos en las tasas de porcentaje de hogares de habla hispana en la región este de Estados Unidos, Wawrowski (2016) desarrolla un modelo Fay-Herriot espacial con la finalidad de estimar las tasas de pobreza a nivel 1 de las unidades administrativas locales en Polonia.

Por otro lado,las predicciones provenientes de estos modelos deben ser evaluadas mediante un criterio de precisión como es el Error Cuadrático Medio (ECM) o el coeficiente de variación (CV).El indicador más utilizado en modelos de áreas pequeñas es el ECM, para el cual existen diversos métodos para realizar su estimación dependiendo de la estimación de la varianza del efecto aleatorio como la derivación analítica propuestas por Datta y Lahiri(2000), Prasad y Rao (1990), Rao y Molina (2015) o utilizando estrategias de remuestreo como jacknife o bootstrap. El método más utilizado es el bootstrap debido a que permite estimar estimadores complejos donde una solución analítica no es posible o es muy difícil de implementar computacionalmente. Cabe resaltar que este método permite relajar el supuesto de asumir una forma paramétrica para la distribución de la población en estudio y además, permite obtener intervalos de confianza más precisos que los obtenidos utilizando la varianza muestral.Dada la versatilidad que tiene este algoritmo, Rao y Molina (2015) proponen una estimación bootstrap para el estimador BLUP derivado del modelo Fay-Herriot. Para el caso del modelo Fay-Herriot espacial, se tiene derivaciones analíticas como la propuesta por Singh et al. (2005) y la basada en bootstrap paramétrico y no paramétrico desarrollada por Molina et al. (2009).

El presente trabajo se enfoca en el desarrollo de las estimaciones en áreas pequeñas desarrolladas a partir del modelo básico propuesta por Fay y Herriot (1979) y su variante espacial desarrollado ampliamente por Rao y Molina (2015). Además se estimará por medio del método de remuestreo bootstrap el error cuadrático medio

⁹En la subsección 4.3 del documento se realiza un desarrollo más a detalle acerca de la derivación del estimador SBLUP y SEBLUP en modelos de áreas pequeñas

de las predicciones para el modelo espacial.

1.3. Objetivos

El objetivo general de la tesis es realizar predicciones del nivel de prevalencia de anemia y riesgo de retraso del crecimiento en niños menores a 5 años, en distritos encuestados con observaciones faltantes y distritos no encuestados con niveles de pobreza mayor a 55% a partir de los datos provenientes de los módulos 64 y 74 de la Encuesta Demográfica de Salud Familiar (ENDES) para el año 2019 y el censo nacional del año 2017. De manera específica:

- Revisar la literatura acerca de las diferentes técnicas de estimación en áreas pequeñas teniendo un particular énfasis en el modelo de Fay-Herriot espacial y su aplicación en la estimación de estadísticas oficiales.
- Implementar el modelo de Fay-Herriot y su variante espacial mediante métodos de inferencia clásica.
- Obtener en base a la ENDES las predicciones de los porcentajes de niños anémicos y con riesgo de retraso del crecimiento en distrito encuestados con pocas observaciones y en distritos no encuestados mediante la estimación del modelo Fay-Herriot y su variante espacial.
- Comparar las predicciones de los porcentajes de niños anémicos y con riesgo de retraso del crecimiento obtenidas por el modelo de Fay-Herriot y su variante espacial.
- Estimar el error cuadrático medio de las predicciones de los porcentajes de niños anémicos y con riesgo de retraso del crecimiento para el modelo Fay-Herriot mediante su aproximación analítica. Para el caso del modelo Fay-Herriot espacial, la estimación del error cuadrático medio se realiza mediante los métodos de bootstrap paramétrico y no paramétrico.
- Comparar los resultados del error cuadrático medio de las predicciones obtenidas por el modelo de Fay-Herriot y su variante espacial.

1.4. Organización del trabajo

La tesis está estructurada de la siguiente manera. En el capítulo 2, se realiza una breve revisión de los diversos estimadores muestrales recogidos en la literatura de análisis de muestras. En el capítulo 3 se presenta modelos lineales mixtos y su importancia en la estimación en áreas pequeñas. En el capítulo 4 se presenta un breve revisión del modelamiento de datos espaciales y los modelos de regresión espacial. En el capítulo 5 se presenta el modelo básico de Fay-Herriot y su variante espacial, además se detalla la derivación de los estimadores BLUP y EBLUP utilizados en los modelos lineales mixtos así como los diferentes métodos existentes para la estimación de la varianza del efecto aleatorio del modelo el cual es clave para la estimación del modelo Fay-Herriot. Finalmente, se presenta la derivación analítica y bootstrap del error cuadrático medio del estimador EBLUP. En el capítulo 6 se obtiene las predicciones de los indicadores de prevalencia de anemia y riesgo de retraso del crecimiento derivados del modelo Fay Herriot y su variante espacial, además de comparar las predicciones obtenidas y las estimaciones del error cuadrático medio de estas predicciones obtenidas mediante la aproximación analítica para el modelo Fay-Herriot y mediante bootstrap paramétrico y no paramétrico para el modelo Fay-Herriot espacial. Finalmente, en el capítulo 7, se detalla las conclusiones del estudio y sugerencias para investigaciones futuras.

2. Estimadores muestrales

Como lo señalamos en la sección de revisión de literatura, la estimación en áreas pequeñas combina la información muestral con datos a nivel administrativo para obtener estadísticos más precisos a nivel de dominio, con lo cual va a ser determinante conocer el diseño muestral de la encuesta a utilizar para luego realizar ejercicios de inferencia estadística adecuados. Para ello es determinante conocer los diferentes paradigmas en la inferencia estadística. De acuerdo a Dorfman (1993), existe dos principales enfoques para la realización de inferencia estadística, el paradigma clásico basado en el diseño (*design-based*) propuesto por Neyman (1934), el cual se sustenta en la realización de inferencia estadística a través de la obtención de repetidas muestras de una población finita de acuerdo al diseño muestral previamente definido. Los estimadores (media, proporción, entre otros) son obtenidos teniendo en cuenta los pesos de las observaciones en la muestra haciendo uso de los pesos muestrales para lograr una inferencia a nivel poblacional; sin embargo, este tipo de estimadores tienen limitaciones al momento de realizar inferencia en muestras que son pequeñas o cuando se tiene errores muestrales (Oficina Europea de Estadística, s.f). Entre este tipo de estimadores se encuentran los directos e indirectos.

En resumen, para realizar inferencia de algún parámetro de interés como el total poblacional ($Y = \sum_n y_j$), donde $\sum_n y_j$ denota la sumatoria de todos los elementos y_j de la población, se asume que todos los elementos j pertenecen a una muestra seleccionada s ($j \in s$) y que pueden ser observados. Bajo el enfoque de diseño, un estimador \hat{Y} de Y es insesgado si se cumple la siguiente condición propuesto por Rao y Molina (2015)

$$E(\hat{Y}) = \sum p(s)(\hat{Y})_s = Y$$

donde $p(s)$ es la probabilidad de selección de la muestra¹⁰ y $(\hat{Y})_s$ es el valor de \hat{Y} para la muestra s con varianza $E[\hat{Y} - E(\hat{Y})]^2$. Cabe resaltar que, en el contexto de la estimación en áreas pequeñas, los estimadores directos son imprecisos debido a los grandes errores de estimación derivados de utilizar muestras pequeñas.

Por otro lado, se tiene el enfoque basado en el modelo (*based-model*) desarrollado ampliamente por Royall (1970) en donde el diseño muestral juega rol menos predominante y en lugar de ello se utiliza información auxiliar disponible para modelar la relación entre la variable de interés y los datos auxiliares disponibles a través de un modelo de regresión. Finalmente, se tiene los modelos en áreas pequeñas que, además de incorporar información auxiliar, permiten introducir heterogeneidad entre áreas a través de efectos aleatorios.

2.1. Estimadores basados en el diseño

2.1.1. Estimadores directos

Los estimadores directos basados en el diseño (*design-based*) consideran las propiedades del diseño muestral para realizar la estimación de algún parámetro de interés como el total o la media. El estimador Horvitz y Thompson es el estimador basado en el diseño más conocido, el cual se basa solamente en las probabilidades de inclusión en la muestra.

Para entender con más detalle la formulación de los estimadores directos analicemos el siguiente ejemplo:

Consideramos una población finita U de tamaño N , donde y_i , $i \in U$ son los valores de la variable de interés en la población. A través de una muestra $k \in U$ tomada de la población U deseamos obtener un estadístico que se aproxime al valor del total poblacional Y . El estimador de Horvitz-Thompson para el total ($\hat{\tau}_{HT}$) y su varianza, desarrollado por Horvitz Thompson (1952), se define de la siguiente manera, siguiendo la especificación propuesta por Valdivieso (2020)

$$\hat{\tau}_{HT} = \sum_{i=1}^n \frac{\hat{\tau}_i}{\pi_i} \delta_i \quad y \quad V(\hat{\tau}_{HT}) = \sum_{i=1}^n (1 - \pi_i) \frac{\tau_i^2}{\pi_i} + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^N (\pi_{ij} - \pi_i \pi_j) \frac{\tau_i}{\pi_i} \frac{\tau_j}{\pi_j} + \sum_{i=1}^n \frac{V(\hat{\tau}_i)}{\pi_i}$$

siendo $\hat{\tau}_i = \sum_{i \in k} y_i$ un estimador insesgado del total poblacional τ_i para la i -ésima unidad, π_i es la probabilidad de inclusión de la unidad i en la muestra, π_{ij} es la probabilidad de inclusión de tanto la unidad i como la j y δ_i es una variable indicadora que toma el valor de 1 si la observación es seleccionada en la muestra y 0 caso contrario.

¹⁰Esta probabilidad depende del diseño muestral a utilizar, ya sea un muestreo aleatorio simple o diseño complejos como muestreos multietápico estratificados

Cabe resaltar que el estimador de Horvitz-Thompson es un estimador insesgado del total poblacional τ_i , tal como se muestra en la siguiente ecuación

$$E(\hat{\tau}_{HT}) = \sum_{i=1}^n E\left(\frac{\hat{\tau}_i}{\pi_i}\right)E(\delta_i) = \sum_{i=1}^n \frac{\tau_i}{\pi_i} \pi_i = \sum_{i=1}^n \tau_i = \tau$$

2.2. Estimadores basados en el modelo

Por otro lado, Van den Brakel y Bethlehem (2008) señalan que los estimadores basados en el modelo llamados también estimadores indirectos son sumamente útiles cuando el tamaño de muestra de algunos dominios o subpoblaciones son pequeños por lo que se utiliza información auxiliar de otros dominios para mejorar la precisión de las estimaciones. Entre los tipos de estimadores indirectos se encuentran los estimadores sintéticos y estimadores compuestos.

2.2.1. Estimadores sintéticos

Los estimadores sintéticos se caracterizan por derivar un estimador indirecto para un área pequeña bajo el supuesto de que las áreas pequeñas tienen las mismas características que el área para el cual se tiene información auxiliar. Entre los estimadores sintéticos se encuentran los derivados del modelo sin información auxiliar (*Area level auxiliary information*), modelo con información auxiliar a nivel de área (*area level auxiliary information*) y modelo con información auxiliar a nivel de unidad (*unit level auxiliary information*). Como ejemplo, definiremos el modelo con información auxiliar a nivel de área definido por Rao y Molina (2015).

Sea \tilde{Y}_i los estimadores muestrales de los totales de las áreas Y_i y $x_{i1}, x_{i2}, \dots, x_{ip}$ son las p variables que contienen la información auxiliar para m de n áreas ($i = 1, \dots, m$). Dada la información, se define un modelo de regresión lineal para las m áreas de la muestra

$$\tilde{Y}_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, m$$

Los coeficientes de regresión estimados ($\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$) facilitan la obtención de predictores sintéticos de regresión para todas las n áreas, los cuales están dados por

$$\hat{Y}_i = \mathbf{x}_i^T \hat{\mathbf{B}}, \quad i = 1, \dots, n$$

donde \hat{Y}_i es el estimador sintético para la i -ésima área, $\hat{\mathbf{B}}$ son las estimaciones de los coeficientes de regresión obtenidos mediante mínimos cuadrados ponderados y la matriz \mathbf{x}_i^T representa la matriz de diseño que contiene las p variables auxiliares.

2.2.2. Estimadores compuestos

Rao y Molina (2015) sostiene que los estimadores compuestos se derivan de un promedio ponderado entre un estimador directo y un estimador sintético. Esta ponderación permite obtener una potencial solución al problema de sesgo proveniente de los estimadores sintéticos y el de alta varianza procedente de los estimadores directos.

$$\hat{Y}_i^{compuesto} = \alpha_i \hat{Y}_i^{directo} + (1 - \alpha_i) \hat{Y}_i^{sintetico}$$

donde α es un parámetro que toma valores en el intervalo abierto $(0,1)$ el cual permite ponderar las estimaciones de los estimadores directo y sintético. Pfefferman (2002) propone definir $\alpha_i = f_i = (n_i/N_i)$, así el estimador directo tiene una mayor ponderación a medida que la fracción muestral f_i se incrementa. EUSTAT (2005) señala que cuando la muestra no es representativa de la población, el estimador indirecto es el que cobra mayor relevancia.

2.3. Modelos de áreas pequeñas

Los modelos de áreas pequeñas permiten introducir mayor información que la recogida por variables auxiliares. En este tipo de modelos se puede modelar la variabilidad entre áreas mediante efectos aleatorios, también se puede realizar pruebas de diagnóstico como análisis de residuos, selección de variables, detección de casos atípicos, entre otros (Rao y Molina, 2015). Tanto los modelos lineales mixtos, modelos lineales generalizados o

modelos no lineales (regresión logística, entre otros) son utilizados con la finalidad de mejorar la precisión de las estimaciones en áreas pequeñas. Cabe resaltar que los modelos de áreas pequeñas permiten manejar variables temporales o de carácter espacial. Recientes avances computacionales facilitan la estimación de este tipo de modelos con mayor complejidad.

Para el caso de esta investigación, solo se analizará el caso de los modelos lineales mixtos aplicados a la estimación de áreas pequeñas¹¹. En la siguiente sección se analizará con mayor detalle el funcionamiento de este tipo de modelos.

3. Modelos lineales mixtos

3.1. Modelo lineal mixto

Los modelos lineales mixtos son extensiones del modelo de regresión lineal clásico en el cual se incluye tanto efectos fijos o aleatorios para capturar la heterogeneidad de los datos y así poder modelarlos a través de una matriz de covarianzas (Gałecki y Burzykowski, 2013). Este tipo de modelos se utiliza usualmente en estudios longitudinales con distintos niveles de análisis, como es el caso de las áreas de investigación educativa (rendimiento escolar) o investigación psicológica (conducta, bienestar) entre otras en donde se presume la existencia de variabilidad o correlación a nivel de individuo o a un nivel más agregado en ciertos casos en donde la unidad de análisis es una región o país.

Para el caso de datos con un solo nivel de agrupamiento con n grupos indexados por i ($i = 1, 2, \dots, N$) y n_i observaciones por grupo indexados por j ($j = 1, 2, \dots, n_i$), podemos formular el siguiente modelo lineal mixto siguiendo lo especificado por Gałecki y Burzykowski (2013) :

$$\begin{aligned} \mathbf{y}_i &= \mathbf{z}_i \boldsymbol{\beta} + \mathbf{b}_i \mathbf{v}_i + \boldsymbol{\epsilon}_i, i = 1, 2, \dots, n \\ \mathbf{v}_i &\sim N(\mathbf{0}, \mathbf{D}), \\ \boldsymbol{\epsilon}_i &\sim N(\mathbf{0}, \mathbf{R}_i), \\ \text{Cov}(\mathbf{v}_i, \boldsymbol{\epsilon}_i) &= 0 \end{aligned}$$

donde

$$\mathbf{y}_i \equiv \begin{pmatrix} y_{i1} \\ \vdots \\ y_{ij} \\ \vdots \\ y_{in_i} \end{pmatrix}, \mathbf{z}_i \equiv \begin{pmatrix} z_{i1}^{(1)} & z_{i1}^{(2)} & \dots & z_{i1}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ z_{in_i}^{(1)} & z_{in_i}^{(2)} & \dots & z_{in_i}^{(p)} \end{pmatrix}, \boldsymbol{\epsilon}_i \equiv \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{ij} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}$$

$$\mathbf{b}_i \equiv \begin{pmatrix} b_{i1}^{(1)} & b_{i1}^{(2)} & \dots & b_{i1}^{(q)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{in_i}^{(1)} & b_{in_i}^{(2)} & \dots & b_{in_i}^{(q)} \end{pmatrix}, \mathbf{v}_i \equiv \begin{pmatrix} v_{i1} \\ \vdots \\ v_{iq} \end{pmatrix},$$

\mathbf{y}_i es un vector de la variable respuesta, \mathbf{z}_i es una matriz de variables explicativas para el i -ésimo grupo, $\boldsymbol{\beta}$ es un vector de efectos fijos y $\boldsymbol{\epsilon}_i$ es un vector de errores con distribución normal multivariada, en tanto que la matriz \mathbf{b}_i es una matriz de covariables con los correspondientes efectos aleatorios \mathbf{v}_i .

También definimos que

$$\begin{aligned} \mathbf{D} &= \sigma^2 \mathbf{D}, \\ \mathbf{R}_i &= \sigma^2 \mathbf{R}_i, \mathbf{R}_i = \boldsymbol{\Lambda}_i \mathbf{C}_i \boldsymbol{\Lambda}_i \end{aligned}$$

¹¹Para mayor información acerca de los modelos no lineales y modelos lineales generalizados aplicados a la estimación de áreas pequeñas, ver Rao y Molina (2015)

donde σ^2 es un parámetro de varianza desconocido, \mathbf{D} y \mathbf{R}_i son matrices de varianza y covarianza, $\mathbf{\Lambda}_i$ es una matriz diagonal con elementos no negativos, \mathbf{C}_i es una matriz de correlación. Definiendo la matriz $\mathbf{\Lambda}_i$ permitimos la heterocedasticidad de las observaciones entre los grupos i y la matriz \mathbf{C}_i modela la correlación entre las observaciones dentro de cada grupo.

La forma matricial de un modelo lineal mixto está definida como:

$$\begin{aligned}\mathbf{Y} &= \mathbf{z}^T \boldsymbol{\beta} + \mathbf{b}\mathbf{v} + \boldsymbol{\epsilon} \\ \mathbf{v} &\sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_D), \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{R}), \\ \text{Cov}(\mathbf{v}, \boldsymbol{\epsilon}) &= 0\end{aligned}$$

donde

$$\mathbf{Z} \equiv \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix}, \mathbf{b} \equiv \begin{bmatrix} \mathbf{b}_1 & 0 & \dots & 0 \\ 0 & \mathbf{Z}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{Z}_N \end{bmatrix}$$

$$\boldsymbol{\Sigma}_D \equiv \begin{bmatrix} \mathbf{D} & 0 & \dots & 0 \\ 0 & \mathbf{D} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{D} \end{bmatrix} = I_n \otimes \mathbf{D}, \mathbf{R} \equiv \begin{bmatrix} \mathbf{R}_1 & 0 & \dots & 0 \\ 0 & \mathbf{R}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{R}_N \end{bmatrix}$$

donde \mathbf{Y} es un vector que contiene los $n = \sum_{n=1}^N n_i$ valores observados de la variable dependiente, \mathbf{z} es una matriz de $n \times p$ covariables, $\boldsymbol{\beta}$ es un vector de coeficientes, \mathbf{b} es una matriz de diseño de tamaño q , \mathbf{v} es un vector que contiene los Nq efectos aleatorios y $\boldsymbol{\epsilon}$ es un vector que contiene los n errores aleatorios.

3.2. Modelo lineal mixto para la estimación en áreas pequeñas

El uso de modelos mixtos para la estimación en áreas pequeñas es de uso extendido dado que, como lo comentamos en la sección de revisión de literatura, permite incorporar información auxiliar de censos o bases administrativas para la estimación de parámetros. En particular, la incorporación de efectos aleatorios a nivel de área a este tipo de modelos permite capturar la heterogeneidad no observable entre áreas la cual no es explicada por las variables del modelo. Cabe resaltar que estos efectos no necesariamente reflejan variación geográfica entre áreas sino que también pueden interpretarse como efectos de variación estacional o efectos entre unidades de muestreo como personas.

Usualmente en este tipo de modelos la variable dependiente toma valores continuos; sin embargo puede darse el caso de variables binarias o de conteo en donde el uso de modelos lineales mixtos no es apropiado. Para este tipo de datos, existen los modelos de regresión logística con efectos aleatorios en donde los métodos de estimación son el método empírico de Bayes (Empirical Bayes en inglés) y el jerárquico bayesiano (Hierarchical Bayes en inglés) (Rao y Molina, 2015). Recientes avances en el campo de la estimación en áreas pequeñas como la incorporación de las técnicas de selección de variables en base a criterios como el Akaike o Kullback (Marhuenda et al. (2014)) o la implementación computacional de modelos que incorporan dependencia espacial (Molina y Marhuenda, 2015 y Kreutzmann et al., 2019) ha permitido mejorar las estimaciones obtenidas por este tipo de modelo. Entre los modelos de estimación en áreas pequeñas más utilizados se encuentran el modelo a nivel de área y el modelo a nivel de unidad.

Siguiendo la definición propuesta por Rao y Molina (2015), el modelo a nivel de área para n áreas pequeñas esta definido por la siguiente ecuación:

$$\begin{aligned}\theta_i &= g(\bar{Y}_i), \theta_i = \mathbf{z}_i^T \boldsymbol{\beta} + b_i v_i, \quad i = 1, 2, \dots, n, \\ v_i &\sim N(0, \sigma_v^2)\end{aligned}$$

donde

$$\mathbf{z}_i = (z_{1i}, z_{2i}, \dots, z_{pi})^T \text{ y } \boldsymbol{\beta} = (\beta_1, \beta_i, \dots, \beta_p)^T$$

\bar{Y}_i es la media poblacional de la variable de interés en el área pequeña i , θ_i es la media a estimar para el área i mediante el modelo, \mathbf{z}_i es un vector de covariables relacionadas al área i provenientes de datos auxiliares, $\boldsymbol{\beta}$ es un vector de coeficientes de regresión y v_i representa los efectos aleatorios de áreas específicas que se asumen independientes e idénticamente distribuidas con media cero y varianza σ_v^2 .

Para realizar inferencias acerca de la media del área pequeña \bar{Y}_i , se asume que se tiene información acerca de los estimadores directos \hat{Y}_i , por lo que se asume la siguiente condición de igualdad.

$$\hat{\theta}_i = g(\hat{Y}_i) = \theta_i + e_i, \quad i = 1, \dots, n$$

donde e_i son los errores de muestreo.

Por otro lado, el modelo a nivel de unidad desarrollado por Battese et al. (1988) asume que los datos auxiliares están disponibles para cada individuo de la población j y área i $\mathbf{z}_{ij} = (z_{ij1}, z_{ij2}, \dots, z_{ijp})$ con lo cual se puede asumir que la media poblacional $\bar{\mathbf{Z}}_i$ es conocida y además, la variable dependiente y_{ij} se encuentra relacionado con la variable \mathbf{z}_{ij} mediante la siguiente ecuación:

$$\begin{aligned} y_{ij} &= \mathbf{z}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, \quad j = 1, 2, \dots, N_i, i = 1, 2, \dots, n, \\ e_{ij} &\sim N(0, \sigma_e^2) \\ v_i &\sim N(0, \sigma_v^2) \end{aligned}$$

donde y_{ij} es la variable de interés, \mathbf{z}_{ij} es un vector de covariables que contiene información de datos auxiliares, v_i son efectos aleatorios a nivel de área que se asumen independientes e idénticamente distribuidas con media cero y varianza σ_v^2 y e_{ij} son errores aleatorios independientes de v_i .

Si el tamaño de la población N_i es grande, entonces la media de la variable de interés en el área pequeña se aproxima a la siguiente expresión

$$\mu_i = \bar{Y}_i = \bar{\mathbf{Z}}_i^T \boldsymbol{\beta} + v_i$$

donde $\bar{Y}_i = \sum_{j=1}^{N_i} y_{ij}$ y $\bar{\mathbf{Z}}_i^T = \sum_{j=1}^{N_i} \mathbf{z}_{ij}$

4. Modelamiento de datos espaciales

Según Goodchild (2008), el análisis de datos espaciales comprende la búsqueda de patrones de comportamiento dentro de algún área o dominio geográfico. La importancia de incorporar el espacio se debe a que en muchos procesos de recolección de datos puede manifestarse un factor de variación local o dependencia entre observaciones en un área específica lo cual sugiere el uso de conceptos y técnicas estadísticas distintas a las convencionales. Uno de los conceptos más importantes en la literatura de estadística espacial es la matriz de pesos espaciales.

4.1. Matrices de pesos espaciales

La dependencia espacial entre observaciones va a estar definida por la estructura de la matriz de pesos espaciales \mathbf{W} . Esta matriz resume las relaciones espaciales entre áreas vecinas, donde $\mathbf{W}=(w_{ir})$ refleja la influencia espacial del área r sobre la i . Valores más altos para w_{ir} indica altos niveles de proximidad. Estas relaciones son descritas mediante la siguiente matriz de pesos no negativa:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots \\ \vdots & \ddots & \vdots \\ w_{n1} & \dots & w_{nn} \end{bmatrix}$$

Para conocer más a detalle los tipos de matrices espaciales utilizadas en el análisis espacial primero se debe considerar como se representa un punto en una área geográfica y como se relaciona con el cálculo de distancias.

4.1.1. Matrices espaciales basadas en distancias de centroide

De acuerdo a Smith (2020), para representar la distancia entre las áreas $L_i, i = 1, 2, \dots, n$ es necesario obtener un punto central c que naturalmente se encuentre cercano a todos los puntos $x_i \in L_i$. La opción más utilizada es el cálculo de un centroide espacial el cual minimiza el promedio al cuadrado de la distancia de todos los puntos x_i que pertenecen al área L_i . Formalmente el centroide c es calculado mediante la siguiente expresión:

$$\min_c \int_L ||x - c|| dx$$

Este centroide es calculado en diferentes áreas con la finalidad de obtener el peso de cada área. Entre las matrices más utilizadas que se basan en el concepto de centroide se encuentran las siguientes:

- Matriz de pesos de K-vecinos más cercanos: Considera distancias de centroide (d_{ir}) para cada área i hacia todas las unidades $r \neq i$ clasificados de la siguiente manera: $d_{ir(1)} \leq d_{ir(2)} \leq \dots \leq d_{ir(m-1)}$ con el conjunto $N_k(i)$ que contiene las k unidades más cercanas a las áreas i . Por cada k vecinos dado, la matriz de pesos espaciales tiene la siguiente forma:

$$w_{ir} = \begin{cases} 1, & \text{si } r \in N_k(i) \\ 0, & \text{en otros casos} \end{cases}$$

El peso espacial w_{ir} toma el valor de uno si la región r es uno de los k vecinos más cercanos del centroide de la región i

- Matriz de pesos de distancia radial: Considera la distancia radial d_{ir} como un criterio de influencia espacial teniendo en cuenta una distancia umbral d donde más allá de este límite no existe influencia espacial entre áreas, dado estas consideraciones la matriz de pesos espaciales tiene la siguiente forma:

$$w_{ir} = \begin{cases} 1, & \text{si } 0 \leq d_{ir} \leq d \\ 0, & d_{ir} > d \end{cases}$$

Cabe resaltar que los pesos w_{ir} son reescalados mediante una proceso de estandarización por filas, con lo cual las los pesos estandarizados toman la siguiente forma $w_{ir}^* = w_{ir} / \sum_j w_{ir}$ cumpliéndose la siguiente condición $\sum_j w_{ir}^* = 1$.

4.1.2. Matrices espaciales basadas en límites

Este tipo de matrices consideran al límite compartido entre unidades espaciales como un factor determinante al momento de medir la dependencia espacial. Entre las matrices más utilizadas se encuentran las siguientes:

- Matriz de pesos de contigüidad espacial: Considera si los pares de unidad de área comparten un límite o no. Si definimos al conjunto de puntos limítrofes del área L_i como $lim(i)$ entonces los componentes de la matriz están definidos por:

$$w_{ir} = \begin{cases} 1, & lim(i) \cap lim(r) \neq \emptyset, \\ 0, & lim(i) \cap lim(r) = \emptyset, \end{cases}$$

- Matriz de pesos con límites compartidos: Si se define ℓ_{ir} como la longitud total del límite $lim(i)$ que es compartido con otras unidades espaciales, entonces la fracción de esta longitud compartida con cualquier área j esta dada por ℓ_{ir}/ℓ_i . Estas fracciones se consideran como ponderaciones de los límites compartidos en la matriz de pesos

$$w_{ir} = \begin{cases} \frac{\ell_{ir}}{\ell_i}, & \ell_{ir} > 0 \\ 0, & \ell_{ir} = 0 \end{cases}$$

Al igual que las matrices de pesos basadas en distancias de centroide, los pesos espaciales w_{ir} son normalizados con la finalidad de eliminar algún factor de escala extraño. Según Chi y Zhu (2019), la correcta selección de una matriz de ponderación espacial adecuada va a depender del tipo de modelo espacial a utilizar y además esto asegurará obtener una mejor medida de la correlación espacial presente en los datos.

4.2. Autocorrelación espacial

Una manera de determinar si existe dependencia espacial entre áreas vecinas es a través de un parámetro de autocorrelación espacial. De acuerdo a Bivand et al. (2008), la autocorrelación espacial es una medida de similitud de una variable de interés entre dos áreas que se encuentran espacialmente cercanas. Una autocorrelación positiva muestra que áreas vecinas presenta valores cercanos con lo cual existe un agrupamiento de estas áreas en el espacio, en el caso de autocorrelación negativa estos valores tiene valores diferentes por lo que existe una dispersión entre áreas vecinas. Entre los estadísticos más comunes para medir la autocorrelación espacial se encuentran el estadístico I de Moran y el estadístico C de Geary (Waller y Gotaway, 2004)

4.2.1. Estadístico I de Moran

Consideremos a n como el total de áreas en un espacio geográfico, y_i a la variable de interés en el área i para $i = 1, 2, \dots, n$. Waller y Gotaway (2004) definen el estadístico I de Moran de la siguiente manera:

$$I = \left(\frac{1}{s^2} \right) \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}, \quad i \neq j,$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

donde: w_{ij} refleja una estructura espacial de correlación entre las áreas i e j de acuerdo a la matriz de pesos espaciales a utilizar, \bar{y} es el promedio de la variable de interés para todas las áreas. El estadístico I de Moran toma usualmente valores entre -1 y 1, cuando los valores para y_i suelen ser similares entre áreas, el valor para I es positivo. Por el contrario, si estos valores tiene valores diferentes el valor para I es negativo.

Cuando no existe correlación entre valores cercanos, el valor esperado para I esta dado por la siguiente expresión, aproximandose a cero en tanto n se incrementa.

$$E(I) = -\frac{1}{n-1}$$

Valores menores que $-\frac{1}{n+1}$ para I reflejan una autocorrelación espacial negativa en tanto que valores mayores a $-\frac{1}{n+1}$ corresponde a una autocorrelación espacial positiva (Cerdá-Hernández y Sikov, 2021). Bajo la hipótesis

nula de ausencia de autocorrelación espacial, los momentos del estadístico de Moran son los siguientes:

$$E(I) = -\frac{1}{n-1}$$

$$V(I) = \frac{(3S_0^2 + S_1n^2 - nS_2)}{S_0(n+1)(n-1)} - E^2(I),$$

donde $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$, $S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})$ y $S_2 = \sum_{j=1}^n (w_{j.} + w_{.j})^2$, siendo $w_{j.} = \sum_{i=1}^n w_{ij}$.

Según Gómez, Cid y Paz (2012), debido a que la distribución probabilística del estadístico de Moran para muestras finitas es desconocida, se utiliza su distribución asintótica normal

$$\sqrt{n}[I - E(I)] \sim N(0, V(I))$$

$$I \sim N(E(I), V(I))$$

El uso de esta aproximación asintótica permitirá estimar el valor-p para definir el aceptar o rechazar la hipótesis nula de ausencia de autocorrelación espacial (Cerdá-Hernández y Sikov, 2021).

4.2.2. Estadístico C de Geary

Waller y Gotaway (2004) definen el estadístico C de Geary mediante la siguiente expresión:

$$C = \left(\frac{n-1}{2 \sum_{j=1}^n (y_i - \bar{y})^2} \right) \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - y_j)^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}, \quad i \neq j,$$

donde: w_{ij} refleja una estructura espacial de correlación entre las áreas i e j de acuerdo a la matriz de pesos espaciales a utilizar, \bar{y} es el promedio de la variable de interés para todas las áreas. Un valor de C cercano a 0 demuestra la existencia de autocorrelación espacial positiva en tanto que valores mayores a 1 denotan la existencia de autocorrelación espacial negativa. Bajo la hipótesis nula de ausencia de autocorrelación, los momentos del estadístico de Moran son los siguientes:

$$E(C) = 1$$

$$V(C) = \frac{(2S_1 + S_2)(n-1) - 4S_0^2}{2(n+1)S_0}$$

donde $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$, $S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})$ y $S_2 = \sum_{j=1}^n (w_{j.} + w_{.j})^2$, siendo $w_{j.} = \sum_{i=1}^n w_{ij}$.

Al igual que el estadístico de Moran, se desconoce la distribución probabilística del estadístico de Geary, por ello se utiliza la aproximación asintótica normal el cual es la siguiente:

$$\sqrt{n}[C - 1] \sim N(0, V(C))$$

$$C \sim N(E(C), V(C))$$

Al igual que el estadístico de Moran, el uso de esta aproximación asintótica permitirá estimar el valor-p para definir el rechazo o aceptación de la hipótesis nula de ausencia de autocorrelación espacial.

4.3. Modelos de regresión espacial

El modelo de regresión espacial incorpora la estructura espacial de los datos en un modelo de regresión lineal, a través de un término de error. Según Bivand, Pebesma, Gómez-Rubio y Pebesma (2008) y Chi y Zhu (2019) entre los modelos más utilizados se encuentran el modelo autorregresivo simultáneo (SAR por sus siglas en inglés) y el modelo autorregresivo condicional (CAR por sus siglas en inglés).

4.3.1. Modelo autorregresivo simultáneo

El modelo autorregresivo simultáneo utiliza una regresión sobre los valores de las otras áreas para establecer la dependencia espacial, definiendo una estructura para los términos de error (e_i) que dependen de la relación entre áreas vecinas. Bivand, Pebesma, Gómez-Rubio y Pebesma (2008) definen el término de error para la i -ésima

área de la siguiente manera:

$$e_i = \sum_{i=1}^n b_{ij}e_i + \epsilon_i, \quad i = 1, 2, \dots, n \quad \text{áreas}, \quad (1)$$

donde ϵ_i representa los errores residuales que se asumen independientes e idénticamente distribuidos con media cero y matriz de covarianzas (Σ_ϵ) de elementos $\sigma_{\epsilon_i}^2 \sim N(0, \sigma^2)$. Los valores b_{ij} representan la dependencia espacial entre las áreas y b_{ii} toma valores de cero para que no se tenga problemas de autodependencia. Con lo cual, el modelo SAR se define de la siguiente manera:

$$Y_i = \mathbf{z}^\top \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, n \quad \text{áreas}$$

donde Y_i representa la variable respuesta en el área i , \mathbf{z} es un vector de covariables y e_i es un término de error.

Considerando lo sugerido por Bivand, Pebesma, Gómez-Rubio y Pebesma (2008), la especificación matricial del modelo SAR está definida por la siguiente expresión:

$$\mathbf{Y} = \mathbf{Z}^\top \boldsymbol{\beta} + \mathbf{e}$$

donde \mathbf{Y} es un vector de la variable de respuesta de tamaño $n \times 1$, \mathbf{z} es una matriz de variables explicativas de tamaño $n \times p$, $\boldsymbol{\beta}$ es un vector de coeficientes de regresión y \mathbf{e} es un vector de términos de errores.

Si definimos de (1), la matriz de errores \mathbf{e} como $\mathbf{e} = \mathbf{B}\mathbf{Y} + \boldsymbol{\epsilon}$, el modelo se puede reformular de la siguiente manera:

$$\mathbf{Y} = \mathbf{Z}^\top \boldsymbol{\beta} + \mathbf{B}(\mathbf{Y} - \mathbf{Z}^\top \boldsymbol{\beta}) + \boldsymbol{\epsilon}$$

Por lo tanto, el modelo descrito anteriormente puede ser formulado de la siguiente manera:

$$(\mathbf{I} - \mathbf{B})(\mathbf{Y} - \mathbf{Z}^\top \boldsymbol{\beta}) = \boldsymbol{\epsilon}$$

siendo \mathbf{B} la matriz que contiene los parámetros de dependencia espacial b_{ij} e \mathbf{I} es una matriz de identidad. Cabe resaltar que la matriz $(\mathbf{I} - \mathbf{B})$ debe ser invertible para obtener las estimaciones del modelo. Es resumen, los modelos SAR incluyen variables rezagadas espacialmente de las covariables \mathbf{Z} y de la variable dependiente \mathbf{Y} .

Finalmente, \mathbf{Y} se distribuye como una normal multivariada con media $E[\mathbf{Y}] = \mathbf{Z}^\top \boldsymbol{\beta}$ y matriz de covarianzas $\text{Var}[\mathbf{Y}] = (\mathbf{I} - \mathbf{B})^{-1} \Sigma_\epsilon (\mathbf{I} - \mathbf{B}^\top)^{-1}$.

4.3.2. Modelo condicional autorregresivo

Según Schabenberger y Gotway (2005), el modelo condicional autorregresivo se basa en la distribución condicional de los términos de error. En este caso, la distribución de e_i condicionado a e_{-i} (vector que contiene todos los errores aleatorios menos e_{-i}). En lugar del vector e_{-i} , solo las áreas vecinas a i son utilizados, los cuales son representados por el vector $e_{j \sim i}$. Finalmente, definimos la distribución condicional del término de error para e_i :

$$e_i | e_{j \sim i} \sim N \left(\frac{\sum_{j \neq i} c_{ij} e_j}{\sum_{j \neq i} e_{ij}}, \frac{\sigma_{e_i}^2}{\sum_{j \neq i} c_{ij}} \right)$$

donde c_{ij} son los parámetros de dependencia espacial similares a los especificados en el modelo SAR (b_{ij}).

Teniendo en cuenta la distribución condicional de los términos de error, podemos definir el modelo condicional autorregresivo de la siguiente manera

$$Y_i = \mathbf{z}^\top \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, n \quad \text{áreas}$$

con

$$e_i | e_{j \sim i} \sim N \left(\frac{\sum_{j \neq i} c_{ij} e_j}{\sum_{j \neq i} e_{ij}}, \frac{\sigma_{e_i}^2}{\sum_{j \neq i} c_{ij}} \right)$$

5. Modelo Fay-Herriot

5.1. Modelo Fay-Herriot

El modelo Fay-Herriot es una variante del modelo lineal mixto, desarrollado por Fay y Herriot (1979) que incluye efectos aleatorios y covariables con información a nivel de área. Para ello consideremos una población dividida en n áreas pequeñas donde θ_i representa un parámetro de interés a estimar (ej. total o la media) en el área i -ésima. Además, se asume que θ_i se relaciona con un vector de p variables auxiliares que dependen de las características del área en mención \mathbf{z}_i mediante un modelo lineal mixto

$$\theta_i = \mathbf{z}_i^\top \boldsymbol{\beta} + v_i, \quad i = 1, \dots, n \quad (2)$$

donde $\mathbf{z}_i = (z_{1i}, z_{2i}, \dots, z_{pi})^\top$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ es el vector de $p \times 1$ coeficientes de regresión y v_i representa a los efectos aleatorios del área, que se asumen distribuidos normalmente, además de ser independientes e idénticamente distribuidos (iid) con

$$E(v_i) = 0, \quad V(v_i) = \sigma_v^2 \geq 0$$

Por otra parte, se asume que se tiene información disponible acerca del estimador directo del área (Y_i) para θ_i para todas las n áreas. Este estimador se define en la siguiente ecuación

$$Y_i = \theta_i + e_i, \quad i = 1, \dots, n \quad (3)$$

donde los errores de muestreo e_i son independientes con

$$E(e_i|\theta_i) = 0, \quad V(e_i|\theta_i) = \Psi_i,$$

y los Ψ_i son conocidas. Nótese que el estimador directo Y_i tiene un alto nivel de varianza en algunas de las áreas debido a que el número de observaciones es pequeño en tanto que en áreas con un tamaño de muestra suficientemente grande, sucede todo lo contrario.

Igualando la ecuación (2) y (3) obtenemos el modelo Fay-Herriot definido por Molina, Salvati y Pratesi (2009):

$$\begin{aligned} Y_i &= \mathbf{z}_i^\top \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, n \quad (4) \\ v_i &\sim N(0, \sigma_v^2) \\ e_i &\sim N(0, \Psi_i) \\ \text{cov}(v_i, e_i) &= 0 \end{aligned}$$

El modelo planteado en la ecuación (4) puede ser reescrito de forma matricial de la siguiente manera

$$\begin{aligned} \mathbf{Y} &= \mathbf{Z}\boldsymbol{\beta} + \mathbf{v} + \mathbf{e} \quad (5) \\ \mathbf{v} &\sim N(\mathbf{0}, \mathbf{G}) \\ \mathbf{e} &\sim N(\mathbf{0}, \mathbf{R}) \end{aligned}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \quad \mathbf{Z} = \begin{bmatrix} z_{11} & z_{21} & \dots & z_{p1} \\ z_{12} & z_{22} & \dots & z_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{1n} & \dots & \dots & z_{pn} \end{bmatrix}_{n \times p},$$

$$\mathbf{G} = \sigma_v^2 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}, \quad \mathbf{R} = \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \Psi_n \end{bmatrix}_{n \times n}$$

donde $\boldsymbol{\beta}$ es una matriz de coeficientes de tamaño $n \times p$ y $\mathbf{e} = (e_1, \dots, e_n)^\top$. Asimismo, \mathbf{v} y \mathbf{e} son distribuidas independientemente con $\mathbf{v} \sim N(\mathbf{0}, \mathbf{G})$ y $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, para $\mathbf{G} = \sigma_v^2 \mathbf{I}$ y $\mathbf{R} = \text{diag}\{\Psi_1, \Psi_2, \dots, \Psi_n\}$.

Se asume la normalidad del estimador Y_i y además el parámetro de varianza Ψ_i es conocido. Para el caso específico de la varianza de los efectos aleatorios (σ_v^2), no se conoce su distribución por lo que se debe realizar una estimación para poder obtener los coeficientes $\boldsymbol{\beta}$ del modelo Fay-Herriot.¹²

¹²La estimación del parámetro σ_v^2 se verá con mas detalle en la sección 5.4

5.2. Modelo Fay-Herriot espacial

El modelo Fay-Herriot desarrollado en la subsección 5.1 asume efectos de área independientes e idénticamente distribuidos (v_i) omitiendo la posible correlación existente entre dominios vecinos. El modelo Fay-Herriot espacial incorpora correlaciones entre los efectos aleatorios (v_i) que dependen de la proximidad geográfica entre áreas vecinas, como ejemplo de estos casos se tiene a las tasas de pobreza o mortalidad. Pratesi y Salvati (2008) desarrolla un modelo Fay-Herriot espacial en donde los efectos aleatorios a nivel de área v_i siguen un proceso autoregresivo de orden 1.

Para ello consideremos un proceso autoregresivo de orden conocido como SAR(1) en el contexto de estimación en áreas pequeñas de acuerdo a lo desarrollado por Singh et al. (2005), el cual modela los efectos aleatorios v_i descritos en la ecuación (5) del modelo Fay-Herriot de la siguiente manera

$$\mathbf{v} = \phi \mathbf{W} \mathbf{v} + \mathbf{u}, \mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_n) \quad (6)$$

Además, si se asume que $(\mathbf{I}_n - \phi \mathbf{W}) \mathbf{v}$ es una matriz no singular y por lo tanto tiene inversa, se tiene la siguiente expresión:

$$\mathbf{v} = (\mathbf{I}_n - \phi \mathbf{W})^{-1} \mathbf{u} \sim N(\mathbf{0}, \mathbf{G}(\delta))$$

donde $\mathbf{G}(\delta) = \sigma_u^2 [(\mathbf{I}_n - \phi \mathbf{W})(\mathbf{I}_n - \phi \mathbf{W})^T]^{-1}$ y $\delta = (\phi, \sigma_u^2)$.

Tomando la ecuación (5) descrita en la subsección anterior y agregando la dependencia espacial en los efectos aleatorios desarrollado en la ecuación (6) obtenemos el modelo Fay-Herriot espacial definido por Molina, Salvati y Pratesi (2009), el cual se define de la siguiente manera:

$$\mathbf{Y} = \mathbf{Z} \boldsymbol{\beta} + \mathbf{v} + \mathbf{e} \quad (7)$$

$$\mathbf{v} = (\mathbf{I}_n - \phi \mathbf{W})^{-1} \mathbf{u}$$

$$\mathbf{v} \sim N(\mathbf{0}, \mathbf{G}(\delta))$$

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$$

donde \mathbf{R} se define como en la ecuación (5) y además se asume independencia entre la matriz de efectos aleatorios y la matriz de errores de muestreo ($\mathbf{v} \perp \mathbf{e}$).

Por otro lado, la matriz \mathbf{W} representa una matriz de pesos espaciales donde los elementos de esta matriz w_{ir} con $w_{ir} = 0$ dependen de la distancia entre los centroides de las áreas pequeñas o el límite común entre ellas y ϕ es un parámetro que describe la dependencia espacial entre los efectos aleatorios asociados a áreas vecinas y toma valores entre -1 y 1.

Una elección para \mathbf{W} es una matriz de pesos espaciales binaria $\mathbf{W} = (\tilde{w}_{iR})$, en donde $\tilde{w}_{ir} = 1$ si $r \neq i$ y $r \in A_i$ y $\tilde{w}_{ir} = 0$ si $A_i \not\subseteq 1, \dots, n$ es nuevamente el conjunto de áreas vecinas del área i . Otra variación respecto a la matriz $\mathbf{W} = (\tilde{w}_{ir})$ consiste en tomar $\tilde{w}_{ir} = (\tilde{w}_{ir} / \sum_{k=1}^n \tilde{w}_{ik})$, esto permite obtener valores para \tilde{w}_{ir} entre 0 y 1 lo cual facilita la interpretación de los pesos de los valores vecinos.

Finalmente, Molina, Salvati y Pratesi (2009) definen la matriz de covarianzas para \mathbf{Y} está dada por:

$$\mathbf{V}(\delta) = \mathbf{R} + \mathbf{G}(\delta)$$

5.3. Estimador BLUP y EBLUP

Para obtener los estimadores de los parámetros del modelo Fay-Herriot y de su variante espacial es necesario derivar un estimador que tenga las propiedades deseadas de insesgamiento y de menor varianza similares al estimador BLUE derivado del teorema de Gauss-Markov. En el caso de los modelos lineales mixtos se obtiene el mejor predictor lineal insesgado (BLUP por sus siglas en inglés), el cual minimiza el error cuadrático medio del parámetro de interés (Sahai y Ojeda, 2007).

Bajo el modelo definido en la ecuación (4) estamos interesados en predecir la media del área pequeña θ_i para

el cual se define un estimador el cual es un promedio ponderado del estimador directo Y_i y el estimador de regresión $\mathbf{z}_i^T \boldsymbol{\beta}$ (Ghosh y Rao, 1994)

$$\hat{\theta}_i = \gamma_i Y_i + (1 - \gamma_i) \mathbf{z}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, n \quad (8)$$

donde $\gamma_i = \sigma_v^2 / (\Psi_i + \sigma_v^2)$ es un factor de ponderación, y $0 < \gamma_i < 1$

Si se asume que σ_v^2 y Ψ_i son conocidos, $\boldsymbol{\beta}$ se obtiene mediante el estimador de mínimos cuadrados ponderados $\tilde{\boldsymbol{\beta}}$.

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\sigma_v^2) = \left[\sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^T / (\Psi_i + \sigma_v^2) \right]^{-1} \left[\sum_{i=1}^m \mathbf{z}_i Y_i / (\Psi_i + \sigma_v^2) \right] \quad (9)$$

$\tilde{\boldsymbol{\beta}}$ es el mejor estimador lineal insesgado (BLUE) de $\boldsymbol{\beta}$ según Ghosh y Rao (1994).

Por lo tanto, si se reemplaza el estimador descrito en (9) en la ecuación (8), se deriva el estimador BLUP

$$\hat{\theta}_i^{BLUP} = \gamma_i Y_i + (1 - \gamma_i) \mathbf{z}_i^T \tilde{\boldsymbol{\beta}}, \quad i = 1, \dots, n \quad (10)$$

El estimador BLUP obtenido en la ecuación (10) depende del componente de la varianza σ_v^2 el cual es desconocido en la mayoría de casos. Existen diversos métodos para estimar σ_v^2 entre los más conocidos se encuentran el método de máxima verosimilitud y el método de máxima verosimilitud restringida. Reemplazando σ_v^2 por su valor estimado $\hat{\sigma}_v^2$ obtenemos el estimador EBLUP $\hat{\theta}_i^{EBLUP}$ (You, 2021)

$$\hat{\theta}_i^{EBLUP} = \hat{\gamma}_i Y_i + (1 - \hat{\gamma}_i) \mathbf{z}_i^T \hat{\boldsymbol{\beta}}, \quad i = 1, \dots, n \quad (11)$$

donde

$$\hat{\gamma}_i = \hat{\sigma}_v^2 / (\Psi_i + \hat{\sigma}_v^2), \quad \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\sigma}_v^2) = \left[\sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^T / (\Psi_i + \hat{\sigma}_v^2) \right]^{-1} \left[\sum_{i=1}^m \mathbf{z}_i Y_i / (\Psi_i + \hat{\sigma}_v^2) \right]$$

Por otro lado, hemos asumido que la ecuación (4) se cumple para todas las áreas pequeñas tanto para las áreas en las que se tiene información muestral ($i = 1, \dots, m$) como para las áreas que no ($R = m + 1, \dots, n$); sin embargo, para las áreas no muestreadas el estimador directo Y_R no es posible de obtener por lo que se utiliza un estimador sintético de regresión basado en las covariables observadas de las áreas no muestreadas (\mathbf{z}_R) y sus respectivos efectos aleatorios v_R (Rao y Molina, 2015). La ecuación está definida por

$$\hat{\theta}_R = \mathbf{z}_R^T \boldsymbol{\beta}_R + v_R, \quad R = m + 1, \dots, n \quad (12)$$

Reemplazando en la ecuación (12) $\boldsymbol{\beta}_R$ por $\hat{\boldsymbol{\beta}}$, el cual es estimado a partir de la información obtenida de las áreas que tienen información muestral $\{(Y_i, \mathbf{z}_i) : i = 1, \dots, m\}$ se obtiene el estimador para las áreas no muestreadas

$$\hat{\theta}_R^{RS} = \mathbf{z}_R^T \hat{\boldsymbol{\beta}} \quad (13)$$

Para el caso del modelo Fay-Herriot espacial, Molina, Salvati y Pratesi (2009) desarrollan el mejor predictor espacial lineal insesgado (SBLUP por sus siglas en inglés) para $\hat{\theta}_i$, el cual está dado por

$$\hat{\theta}_i^{SBLUP}(\boldsymbol{\delta}) = \mathbf{Z}^T \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) + \mathbf{b}_i^T \mathbf{G}(\boldsymbol{\delta}) \mathbf{V}^{-1}(\boldsymbol{\delta}) [\hat{\boldsymbol{\theta}} - \mathbf{Z}_i^T \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta})], \quad (14)$$

donde $\tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = (\mathbf{Z}^T \mathbf{V}^{-1}(\boldsymbol{\delta}) \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{V}^{-1}(\boldsymbol{\delta}) \mathbf{Y}$ y $\mathbf{b}_i^T = (0, \dots, 0, 1, 0, \dots, 0)$ es un vector de tamaño $1 \times n$ con el valor de uno en la i -ésima posición. Este estimador espacial SBLUP depende de los componentes del vector de varianzas desconocido $\boldsymbol{\delta} = (\phi, \sigma_u^2)$.

El estimador $\hat{\theta}_i^{SEBLUP}(\boldsymbol{\delta})$ se obtiene reemplazando el estimador $\boldsymbol{\delta}$ en la ecuación (14) por $\hat{\boldsymbol{\delta}} = (\hat{\phi}, \hat{\sigma}_u^2)$. Asumiendo normalidad de los errores y efectos aleatorios se puede estimar $\hat{\boldsymbol{\delta}} = (\hat{\phi}, \hat{\sigma}_u^2)$ mediante el método de máxima verosimilitud o el enfoque restringido de máxima verosimilitud.

5.4. Estimación de σ_v^2

Nótese que para la estimación del estimador EBLUP es necesario realizar la estimación de σ_v^2 . Diferentes métodos han sido propuestos en la literatura como el uso de un método iterativo o el uso del estimador de momentos como lo sugieren Prasad y Rao (1990); sin embargo, puede darse el caso que las estimaciones para σ_v^2 resulten negativas con lo cual se han planteado diversas soluciones como las desarrolladas por Li y Lahiri (2010) y Yishimori y Lahiri (2014).

5.4.1. Método de momentos

El estimador de método de momentos ($\hat{\sigma}_{MM}^2$) definido por Rao y Molina (2015) se obtiene a través de la siguiente expresión

$$E \left[\sum_{i=1}^n (Y_i - \mathbf{z}_i^\top \tilde{\boldsymbol{\beta}})^2 / (\Psi_i + \sigma_v^2) \right] = E[a(\sigma_v^2)] = n - p$$

donde $\tilde{\boldsymbol{\beta}}$ está definido por la ecuación (10). $\hat{\sigma}_{MM}^2$ se obtiene resolviendo la siguiente ecuación mediante métodos de iteración

$$a(\sigma_v^2) = n - p$$

Uno de estos métodos es el algoritmo de Fisher-scoring desarrollado por Rao y Molina (2015), el cual está dado por la siguiente expresión

$$\sigma_v^{2(a+1)} = \sigma_v^{2(a)} + [\mathbf{I}(\sigma_v^{2(a)})]^{-1} s(\tilde{\boldsymbol{\beta}}^{(a)}, \sigma_v^{2(a)}) \quad (15)$$

donde

$$\mathbf{I}(\sigma_v^2) = \frac{1}{2} \sum_{i=1}^n \frac{1}{(\sigma_v^2 + \Psi_i)^2}$$

y

$$s(\tilde{\boldsymbol{\beta}}, \sigma_v^2) = -\frac{1}{2} \sum_{i=1}^n \frac{1}{(\sigma_v^2 + \Psi_i)} + \frac{1}{2} \sum_{i=1}^n \frac{(Y_i - \mathbf{z}_i^\top \tilde{\boldsymbol{\beta}})^2}{(\sigma_v^2 + \Psi_i)^2}$$

El estimador se obtiene de $\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0)$, donde $\tilde{\sigma}_v^2$ es la solución obtenida de la ecuación (15)

5.4.2. Métodos de máxima verosimilitud

Uno de ellos es el método de máxima verosimilitud (MV) aplicado al modelo Fay-Herriot, Avila-Valdez et al. (2020) definen la función de log-verosimilitud de la siguiente forma

$$\ell_{MV}(\sigma_v^2, \boldsymbol{\beta}; \mathbf{Y}) = c - \frac{1}{2} \log(|\mathbf{V}|) - \frac{1}{2} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}), \quad (16)$$

donde c es una constante que es independiente de σ_v^2 y $\boldsymbol{\beta}$. Diferenciando la ecuación (16) respecto a σ_v^2 y $\boldsymbol{\beta}$ obtenemos lo siguiente:

$$\frac{\partial \ell_{MV}(\sigma_v^2, \boldsymbol{\beta}; \mathbf{Y})}{\partial \boldsymbol{\beta}} = \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Y} - \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z} \boldsymbol{\beta} \quad (17)$$

$$\frac{\partial \ell_{MV}(\sigma_v^2, \boldsymbol{\beta}; \mathbf{Y})}{\partial \sigma_v^2} = \frac{1}{2} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^\top \mathbf{V}^{-2} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}) - \frac{1}{2} \text{tr}(\mathbf{V}^{-1}) \quad (18)$$

Igualando a cero las ecuaciones (17) y (18) obtenemos los estimadores de máxima verosimilitud para $\boldsymbol{\beta}$ ($\hat{\boldsymbol{\beta}}$) y σ_v^2 ($\hat{\sigma}_{vML}^2$). Cabe resaltar que en el caso que las estimaciones para la varianza den resultados negativos, éstos son ajustados a cero mediante la siguiente condición $\hat{\sigma}_v^2 = \max(\sigma_v^2, 0)$. Si además reemplazamos $\boldsymbol{\beta}$ por $\tilde{\boldsymbol{\beta}}$ en la ecuación (16) obtenemos la función de log-verosimilitud perfil (*profile log-likelihood* en inglés)

$$\ell_{PML}(\sigma_v^2, \mathbf{Y}) = c - \frac{1}{2} \log(|\mathbf{V}|) - \frac{1}{2} \mathbf{Y}^\top \mathbf{P} \mathbf{Y} \quad (19)$$

donde $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{Z} (\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{V}^{-1}$. Igualando a cero la ecuación (19) y resolviendo la ecuación respecto a σ_v^2 obtenemos el mismo estimador de máxima verosimilitud para la varianza. Datta y Lahiri (2000) obtienen el sesgo y la varianza asintótica para $\hat{\sigma}_{vML}^2$.

Por otro lado, Datta y Lahiri (2010) demuestran que el estimador de máxima verosimilitud tiende a obtener resultados con mayor sesgo a comparación del método restringido de máxima verosimilitud. Este método mantiene las mismas características de consistencia, eficiencia y normalidad asintótica que el estimador de máxima verosimilitud. Una característica particular del estimador restringido es que considera los grados de libertad en la estimación de efectos fijos. Datta y Lahiri (2000) demuestran que la varianza asintótica para $\hat{\sigma}_v^2$ derivada del enfoque restringido de máxima verosimilitud es similar al obtenido mediante máxima verosimilitud¹³

Pese a todas las características que tienen los estimadores mencionados anteriormente, estos pueden producir estimaciones de varianza negativa que se supone deben ser estrictamente positivas (Li y Lahiri, 2010). El problema se debe a que las estimaciones de varianza negativa son ajustados a cero $\hat{\sigma}_v^2 = \max(\sigma_v^2, 0)$ ocasionando que el factor de ponderación γ_i tenga valor cero y por consiguiente no existe una ponderación para el estimador directo Y_i lo cual conlleva a un problema de sobreajuste en las estimaciones derivadas del modelo Fay-Herriot. Cabe resaltar que este tipo de problema se presenta cuando el número de áreas pequeñas a estimar es pequeño (Rao y Molina, 2015).

Debido a este problema, Li y Lahiri (2010) y Yoshimori y Lahiri (2014) proponen ajustes a los métodos de verosimilitud enfocándose en la derivación de la función de verosimilitud. Li y Lahiri (2010) define una función de la función de verosimilitud como el producto de la función de verosimilitud estándar (MV) y σ_v^2 obteniendo así una función de log-verosimilitud ajustada (LMV) $\ell_{LMV}(\sigma_v^2; \mathbf{Y}) = \ell(\sigma_v^2; \mathbf{Y}) + \log(\sigma_v^2)$ donde $\ell(\sigma_v^2; \mathbf{Y})$ corresponde a la función de log-verosimilitud perfil (Profile likelihood o la función restringida de máxima verosimilitud (Restricted maximum likelihood). La maximización de esta función de verosimilitud nos permite obtener los estimadores de máxima verosimilitud Li-Lahiri ($\hat{\sigma}_{v_{LMV}}^2$) y el de máxima verosimilitud restringido Li-Lahiri ($\hat{\sigma}_{v_{LMVR}}^2$).

Yoshimori y Lahiri (2014) define la función de verosimilitud ajustada como el producto de una función $h(\sigma_v^2)$ y la función de verosimilitud estándar. La función de verosimilitud ajustada esta definida como

$$\ell_{YMV}(\sigma_v^2; \mathbf{Y}) = \ell(\sigma_v^2; \mathbf{Y}) + \log(h(\sigma_v^2))$$

donde $\ell(\sigma_v^2; \mathbf{Y})$ corresponde a la función de log-verosimilitud perfil o la función restringida de máxima verosimilitud, $h(\sigma_v^2) = (\tan^{-1}(\text{traza}(\mathbf{I}_n - \boldsymbol{\gamma})))^{1/n}$ con $\boldsymbol{\gamma} = \text{diag}\{\gamma_1, \gamma_2, \dots, \gamma_n\}$ y γ_i definida en la ecuación (8).

La maximización de esta función de verosimilitud nos permite obtener los estimadores de máxima verosimilitud Yoshimori-Lahiri ($\hat{\sigma}_{v_{YMV}}^2$) y el de máxima verosimilitud restringido Yoshimori-Lahiri ($\hat{\sigma}_{v_{YMV R}}^2$). Estos estimadores son estrictamente positivos, sus varianzas asintóticas son idénticas a las obtenidas por máxima verosimilitud. El estimador $\hat{\sigma}_{v_{YMV R}}^2$ es asintóticamente insesgado para σ_v^2 (Yoshimori y Lahiri, 2014).

Para el caso del modelo Fay-Herriot espacial, el vector de parámetros $\boldsymbol{\delta} = (\phi, \sigma_u^2)$ se estima considerando la función de log-verosimilitud definida en la ecuación (16) mediante el método de máxima verosimilitud o el método restringido de máxima verosimilitud (Singh et al., 2005).

El método a utilizar en esta investigación será el método restringido de máxima verosimilitud (REML por sus siglas en inglés) debido a las propiedades estadísticas mencionadas al inicio de esta subsección. El paquete estadístico desarrollado por Kreutzmann et al. (2019) permite implementar computacionalmente el método restringido de máxima verosimilitud tanto para el modelo Fay-Herriot como para su variante espacial.

5.5. Aproximación analítica para el error cuadrático medio del estimador EBLUP y SEBLUP

5.5.1. Modelo Fay-Herriot

Rao y Molina (2015) definen el error cuadrático medio del estimador EBLUP de θ_i a partir de la siguiente expresión:

$$ECM(\hat{\theta}_i^{EBLUP}) = E(\hat{\theta}_i^{EBLUP} - \theta_i)^2 \quad (20)$$

$$ECM(\hat{\theta}_i^{EBLUP}) = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2, \quad i = 1, \dots, n$$

¹³Para mayor detalle acerca de la derivación matemática del estimador mediante el enfoque restringido de máxima verosimilitud, ver Bucyibaruta (2014) y Ávila-Valdez et al. (2020)

donde

$$g_{1i}(\sigma_v^2) = \sigma_v^2 \Psi_i / (\Psi_i + \sigma_v^2) \quad , \quad g_{2i}(\sigma_v^2) = (1 - \gamma_i)^2 \mathbf{z}_i^T \left[\sum_{j=1}^m \mathbf{z}_j \mathbf{z}_j^T / (\Psi_i + \sigma_v^2) \right]^{-1} \mathbf{z}_i \quad (21)$$

Según Avila-Valdez et al. (2020), el término $g_{1i}(\sigma_v^2)$ captura la incertidumbre del estimador definido en la ecuación (8) producto de la presencia de los parámetros de varianza del efecto aleatorio y de los errores de muestreo, y el término $g_{2i}(\sigma_v^2)$ refleja la variabilidad procedente de la estimación de β .

El último término de la ecuación (20) $E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2$, el cual captura la incertidumbre debido a la estimación de σ_v^2 , se aproxima mediante la siguiente expresión (Li y Lahiri, 2010)

$$g_{3i}(\sigma_v^2) = E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2 = \Psi_i^2 (\Psi_i + \sigma_v^2)^{-3} \bar{V}(\hat{\sigma}_v^2) \quad (22)$$

y $\bar{V}(\hat{\sigma}_v^2)$ es la varianza asintótica del estimador $\hat{\sigma}_v^2$ para σ_v^2 . Rao y Molina (2015) definen la varianza asintótica para los estimadores de máxima verosimilitud (MV) y máxima verosimilitud restringido (MVR), de la siguiente manera:

$$\bar{V}(\hat{\sigma}_{MV}^2) = \bar{V}(\hat{\sigma}_{MVR}^2) = \bar{V}(\hat{\sigma}_v^2) = 2 \left[\sum_{i=1}^n 1/(\sigma_v^2 + \Psi_i)^2 \right]^{-1} \quad (23)$$

Por lo tanto, se realiza una aproximación de segundo orden para $ECM(\hat{\theta}_i^{EBLUP})$ en la ecuación (20) la cual esta definida por la siguiente expresión:

$$ECM(\hat{\theta}_i^{EBLUP}) = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2) \quad (24)$$

Nótese que $ECM(\hat{\theta}_i^{EBLUP})$ depende de σ_v^2 , el cual es desconocido, por lo tanto no se puede utilizar para conocer la incertidumbre del estimador EBLUP. Debido a ello, se debe realizar una estimación del estimador $ECM(\hat{\theta}_i^{EBLUP})$, Datta y Lahiri (2000) definen este estimador mediante la siguiente expresión

$$\widehat{ECM}(\hat{\theta}_i^{EBLUP}) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2) - \hat{\gamma}_i^2 \widehat{Sesgo}(\hat{\sigma}_v^2)$$

donde $g_{1i}(\hat{\sigma}_v^2)$ y $g_{2i}(\hat{\sigma}_v^2)$ se definen en las ecuaciones (21), y $g_{3i}(\hat{\sigma}_v^2)$ se detalla en la ecuación (22) cuando σ_v^2 es reemplazado por su estimador ($\hat{\sigma}_v^2$) y $\widehat{Sesgo}(\hat{\sigma}_v^2)$ es el estimador de $Sesgo(\hat{\sigma}_v^2)$ el cual se determina de la siguiente manera según Rao y Molina (2015)

$$Sesgo(\hat{\sigma}_v^2) = -[2\mathbf{I}(\sigma_v^2)]^{-1} traza \left\{ \left[\sum_{i=1}^n (\Psi + \sigma_v^2)^{-1} \mathbf{z}_i \mathbf{z}_i^T \right]^{-1} \times \left[\sum_{i=1}^n (\Psi + \sigma_v^2)^{-2} \mathbf{z}_i \mathbf{z}_i^T \right] \right\} \quad (25)$$

donde $\mathbf{I}(\sigma_v^2)$ es la matriz de información de Fisher definida en la ecuación (15).

Asimismo, Rao y Molina (2015) sostienen que para el caso de las áreas que no cuentan con información muestral $R = m + 1, \dots, n$, el estimador del error cuadrático medio del estimador sintético $\hat{\theta}_R^{RS}$ presentado en la ecuación (20) está dado por

$$ECM(\hat{\theta}_R^{RS}) = E(\mathbf{z}_R^T \hat{\beta} - \theta_R)^2 = E[\mathbf{z}_R^T (\hat{\beta} - \beta) - v_R]^2$$

donde $\hat{\theta}_R^{RS}$ se encuentra definido en la ecuación (13). Notar que $\hat{\beta}$ es independiente de los efectos aleatorios v_R dado que $\hat{\beta}$ es calculado de la muestra $\{(Y_i, \mathbf{z}_i) : i = 1, \dots, n\}$ con lo cual se obtiene la siguiente expresión

$$ECM(\hat{\theta}_R^{RS}) = \sigma_v^2 + \mathbf{z}_R^T \left[\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T / (\Psi_i + \sigma_v^2) \right]^{-1} \mathbf{z}_R + \|V(\hat{\beta}) - V(\tilde{\beta})\|$$

$$ECM(\hat{\theta}_R^{RS}) = \sigma_v^2 + h_R(\sigma_v^2) + \|V(\hat{\beta}) - V(\tilde{\beta})\|$$

Rao y Molina (2015) definen la estimación del error cuadrático medio mediante la siguiente expresión:

$$\widehat{ECM}(\hat{\theta}_R^{RS}) = \hat{\sigma}_v^2 + h_R(\hat{\sigma}_v^2) + \widehat{Sesgo}(\hat{\sigma}_v^2)$$

donde $\widehat{Sesgo}(\hat{\sigma}_v^2)$ es el estimador de la ecuación definida en (23)

5.5.2. Modelo Fay-Herriot espacial

Para el caso de la variante espacial del modelo Fay-Herriot, Molina, Salvati y Pratesi (2009) definen el error cuadrático medio para el estimador SEBLUP estimación del error cuadrático medio se descompone de la siguiente manera:

$$\begin{aligned} ECM(\hat{\theta}_i^{SEBLUP}(\hat{\delta})) &= ECM(\hat{\theta}_i^{SEBLUP}(\delta)) + E([\hat{\theta}_i^{SEBLUP}(\hat{\delta}) - \hat{\theta}_i^{SEBLUP}(\delta)]^2) \\ ECM(\hat{\theta}_i^{SEBLUP}(\hat{\delta})) &= g_{1i}(\delta) + g_{2i}(\delta) + g_{3i}(\delta), \quad i = 1, \dots, n \end{aligned} \quad (26)$$

donde $g_{1i}(\delta)$ refleja la incertidumbre debido a la predicción de los efectos aleatorios, $g_{2i}(\delta)$ mide el error derivado de la estimación de β y el último término refleja la incertidumbre derivada de la estimación de los componentes de la varianza $\delta = (\phi, \sigma_u^2)$. Los dos primeros términos pueden ser calculados de forma exacta debido a la propiedad de linealidad del estimador SEBLUP $\hat{\theta}_i^{SEBLUP}(\delta)$ respecto al vector θ los cuales son:

$$g_{1i}(\delta) = \mathbf{b}_i^\top [\mathbf{G}(\delta) - \mathbf{G}(\delta)\mathbf{V}^{-1}(\delta)\mathbf{G}(\delta)]\mathbf{b}_i \quad (27)$$

$$g_{2i}(\delta) = \mathbf{b}_i^\top [\mathbf{I}_n - \mathbf{G}(\delta)\mathbf{V}^{-1}(\delta)]\mathbf{Z}(\mathbf{Z}^\top\mathbf{V}^{-1}(\delta)\mathbf{Z})^{-1}\mathbf{Z}^\top[\mathbf{I}_n - \mathbf{V}^{-1}(\delta)\mathbf{G}(\delta)]\mathbf{b}_i \quad (28)$$

Para el último término de la ecuación, $g_{3i}(\delta)$, no es posible derivar una forma exacta; sin embargo, si se toma en cuenta que bajo el modelo Fay-Herriot con efectos aleatorios independientes, Prasad y Rao (1990) obtienen el término $g_{3i}(\sigma_v^2)$ a través de una linealización de Taylor. La aplicación de este método conlleva a que se obtenga la siguiente expresión:

$$g_{3i}^{PR}(\delta) = \text{traza} \left\{ \mathbf{L}_i(\delta)\mathbf{V}^{-1}(\delta)\mathbf{L}_i^\top(\delta)\mathbf{I}^{-1}(\delta) \right\}$$

donde $\mathbf{I}^{-1}(\delta)$ es matriz de información de Fisher y

$$\mathbf{L}_i(\delta) = \begin{bmatrix} \mathbf{b}_i^\top (\mathbf{C}^{-1}\mathbf{V}^{-1} - \sigma_v^2\mathbf{C}^{-1}\mathbf{V}^{-1}\mathbf{C}^{-1}\mathbf{V}^{-1}) \\ \mathbf{b}_i^\top (\mathbf{A}\mathbf{V}^{-1} - \sigma_v^2\mathbf{C}^{-1}\mathbf{V}^{-1}\mathbf{A}\mathbf{V}^{-1}) \end{bmatrix}$$

con \mathbf{b}_i definido en la ecuación (13), $\mathbf{C} = (\mathbf{I}_n - \phi\mathbf{W})$ y $\mathbf{A} = \sigma_v^2\mathbf{C}^{-1}(\mathbf{W} + \mathbf{W}^\top - 2\phi\mathbf{W}^\top\mathbf{W})\mathbf{C}^{-1}$.

Por lo tanto, el error cuadrático medio se puede aproximar a la siguiente expresión

$$ECM(\hat{\theta}_i^{SEBLUP}(\delta)) = g_{1i}(\delta) + g_{2i}(\delta) + g_{3i}^{PR}(\delta) \quad (29)$$

Como lo sugiere Zimmerman y Cressie (1992), cuando δ se estima por el método de máxima verosimilitud restringido, el estimador del error cuadrático medio de la ecuación (26) se aproxima al estimador derivado por Prasad y Rao (1990). Molina, Salvati y Pratesi (2009) definen este estimador de la siguiente manera:

$$ECM^{PR}[(\hat{\theta}_i^{SEBLUP}(\hat{\delta}))] = g_{1i}(\hat{\delta}) + g_{2i}(\hat{\delta}) + 2g_{3i}^{PR}(\hat{\delta})$$

Si se estima $\hat{\delta} = (\phi, \sigma_u^2)$ por máxima verosimilitud, el estimador insesgado del ECM está definido por la siguiente ecuación (Molina, Salvati y Pratesi, 2009)

$$ECM_{MV}^{PR}[(\hat{\theta}_i^{SEBLUP}(\hat{\delta}))] = g_{1i}(\hat{\delta}) + g_{2i}(\hat{\delta}) + 2g_{3i}(\hat{\delta}) - \text{Sesgo}_{MV}^\top(\hat{\delta})\nabla g_{1i}(\hat{\delta}),$$

donde $\nabla g_{1i}(\hat{\delta}) = \partial g_{1i}(\delta)/\partial \delta$ es la gradiente de $g_{1i}(\delta)$ y $\text{Sesgo}_{MV}^\top(\hat{\delta})$ es el sesgo del estimador de máxima verosimilitud para $\hat{\delta}$. Este sesgo es igual a $\text{Sesgo}_{MV}^\top(\hat{\delta}) = \mathbf{I}(\sigma_v^2)^{-1}\mathbf{h}(\hat{\delta})/2$ con $\mathbf{h}(\hat{\delta}) = (h_1(\hat{\delta}), h_2(\hat{\delta}))^\top$ y

$$h_k(\delta) = \text{traza} \left\{ (\mathbf{Z}^\top\mathbf{V}^{-1}(\delta)\mathbf{Z})^{-1} \frac{\partial \ell(\mathbf{Z}^\top\mathbf{V}^{-1}(\delta)\mathbf{Z})}{\partial \delta_k} \right\}, \quad k = 1, 2$$

Singh et al. (2005) desarrolla un estimador derivado del método de máxima verosimilitud restringido, el cual es desarrollado por Molina, Salvati y Pratesi (2009) y es el siguiente:

$$ECM_{MVR}^{SSK}[(\hat{\theta}_i^{SEBLUP}(\hat{\delta}))] = g_{1i}(\hat{\delta}) + g_{2i}(\hat{\delta}) + 2g_{3i}^{PR}(\hat{\delta}) - g_{4i}(\hat{\delta})$$

donde

$$g_{4i}(\hat{\delta}) = \frac{1}{2} \left[\sum_{k=1}^2 \sum_{\ell=1}^2 \mathbf{b}_i^\top \mathbf{R} \mathbf{V}^{-1}(\delta) \frac{\partial^2 \mathbf{V}(\delta)}{\partial \delta_k \partial \delta_\ell} \mathbf{V}^{-1}(\delta) \mathbf{R} \Gamma_{k\ell}^{-1}(\delta) \mathbf{b}_i \right]$$

5.6. Estimación del error cuadrático medio del estimador EBLUP y SEBLUP mediante bootstrap

5.6.1. Modelo Fay-Herriot

Las técnicas de remuestreo como el bootstrap pueden ser utilizadas para obtener el error cuadrático medio del estimador EBLUP, $\hat{\theta}_i^{EBLUP}$ u otros estimadores complejos $\hat{\rho}_i^{EBLUP}$, de $\rho_i = h(\theta_i)$ donde $h(\cdot)$ es una función monótona no lineal definida por Slud y Maiti (2006), la cual puede ser una función logarítmica o exponencial. Un ejemplo de este tipo de estimadores para la media \bar{Y}_i es \hat{Y}_i^{*EBLUP} el cual es un estimador de la media ajustado por un factor de sesgo definido por Rao y Molina (2015) de la siguiente manera:

$$\hat{Y}_i^{*EBLUP} = \hat{F}_i \hat{Y}_i^{EBLUP}$$

$$\hat{F}_i = E[(h(\mathbf{z}_i^T \boldsymbol{\beta} + v_i)] / E[h(\mathbf{z}_i^T \boldsymbol{\beta} + \gamma_i(v_i + e_i))]$$

\hat{F}_i se obtiene cuando $(\boldsymbol{\beta}, \sigma_v^2) = (\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$

Con lo cual, la generalización de este método de bootstrap a estimadores complejos permite estimar el error cuadrático medio del estimador EBLUP donde la aproximación analítica no es fácilmente aplicable (Rao y Molina, 2015)

Rao y Molina (2015) definen los siguientes pasos para la estimación bootstrap del error cuadrático medio

- (i) Asumiendo normalidad para v_i y e_i y $\hat{\sigma}^2 > 0$, se genera un conjunto de datos bootstrap $(\hat{\theta}_{i*}^{EBLUP}, \mathbf{z}_i)$ independientes para cada área $i = 1, \dots, n$, de la siguiente manera:
 - (i)1. Se genera θ_{i*}^{BLUP} a partir de la distribución $N(\mathbf{z}_i^\top \hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$ para cada área $i = 1, \dots, n$ y definimos $\rho_i^* = h(\theta_{i*}^{BLUP})$
 - (i)2. Se genera $\hat{\theta}_{i*}^{EBLUP}$ a partir de la distribución $N(\theta_{i*}^{BLUP}, \Psi_i)$.
 - (i)3. El método utilizado para estimar $\hat{\rho}_i^{EBLUP}$, ya sea por máxima verosimilitud o máxima verosimilitud restringido, es aplicado a la muestra bootstrap $(\hat{\theta}_{i*}^{EBLUP}, \mathbf{z}_i); i = 1, \dots, n$ para obtener $\hat{\rho}_{i*}^{EBLUP}$
 - (i)4. Repetir los pasos anteriores B veces con la finalidad de obtener B estimaciones bootstrap $\hat{\rho}_{i*}^{EBLUP}(1), \dots, \hat{\rho}_{i*}^{EBLUP}(B)$ y los valores bootstrap para ρ , $\rho_{i*}(1), \dots, \rho_{i*}(B)$

De acuerdo a Rao y Molina (2015), el estimador teórico bootstrap para ECM ($\hat{\rho}_i^{EBLUP}$) esta dado por $ECM_B(\hat{\rho}_i^{EBLUP}) = E(\hat{\rho}_{i*}^{EBLUP} - \rho_{i*}^{BLUP})^2$, donde E representa el valor esperado bootstrap. Para poder estimar este término se realiza una aproximación por Monte Carlo utilizando B replicas bootstrap.

$$ECM_B(\hat{\rho}_i) = B^{-1} \sum_{b=1}^B [\hat{\rho}_{i*}^{EBLUP}(b) - \rho_{i*}^{BLUP}(b)]^2$$

Para el caso especial de $\rho = \theta_i$ siguiendo la aproximación de segundo orden desarrollada en la ecuación (23), Rao y Molina (2015) sostienen que el modelo Fay-Herriot bootstrap es similar al modelo Fay-Herriot derivado en la ecuación (4) considerando $(\boldsymbol{\beta}, \sigma_v^2) = (\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$, con lo cual obtenemos la siguiente expresión:

$$ECM_B(\hat{\theta}_i) \approx g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + g_{3i}(\hat{\sigma}_v^2)$$

donde

$$g_{1i}(\hat{\sigma}_v^2) = \hat{\sigma}_v^2 \hat{\Psi}_i / (\hat{\Psi}_i + \hat{\sigma}_v^2) \quad , \quad g_{2i}(\hat{\sigma}_v^2) = (1 - \hat{\gamma}_i)^2 \hat{z}_i^T \left[\sum_{i=1}^m \hat{z}_i \hat{z}_i^T / (\hat{\Psi}_i + \hat{\sigma}_v^2) \right]^{-1} \hat{z}_i$$

y $g_{3i}(\hat{\sigma}_v^2) = \hat{\Psi}_i^2 (\hat{\Psi}_i + \hat{\sigma}_v^2)^{-3} \hat{V}(\hat{\sigma}_v^2)$, siendo $\hat{V}(\hat{\sigma}_v^2)$ el estimador asintótico de la varianza definido en la ecuación (22)

5.6.2. Modelo Fay-Herriot espacial

Para el caso del modelo Fay-Herriot espacial se tiene 2 alternativas para la estimación del error cuadrático medio mediante bootstrap. El primero es el denominado bootstrap paramétrico debido a que considera que los datos provienen de una distribución conocida con parámetros desconocidos (ej. Poisson, negativa binomial, entre otros), estimando primero los parámetros a partir de los datos y luego se usa las distribuciones estimada para realizar ejercicios de simulación con las muestras. Por otro lado se encuentra el método bootstrap no paramétrico, el cual los datos aleatorios se crean a partir de muestras con reemplazo de los datos.¹⁴

5.6.2.1 Bootstrap paramétrico

La expresión desarrollada en las ecuación (24) tiene soluciones exactas para $g_{1i}(\boldsymbol{\delta})$ y $g_{2i}(\boldsymbol{\delta})$ pero no para $g_{3i}(\boldsymbol{\delta})$, el cual representa la incertidumbre del estimador SEBLUP debido a los parámetros de varianza $\boldsymbol{\delta} = (\sigma_v^2, \phi)$. El estimador bootstrap para $g_{3i}(\boldsymbol{\delta})$ a desarrollar en esta subsección puede ser incorporado con los estimadores analíticos de g_{1i} y g_{2i} expuestos en las ecuaciones (25) y (26) obteniendo así todos los componentes del estimador del error cuadrático medio $ECM(\hat{\theta}_i^{SEBLUP}(\hat{\boldsymbol{\delta}}))$.

Molina, Salvati y Pratesi (2009) definen los siguientes pasos para realizar el método de bootstrap paramétrico:

1. Se estima el modelo Fay-Herriot espacial desarrollado en la ecuación (7), con los datos para $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$, obteniendo estimaciones para $\hat{\boldsymbol{\delta}} = (\hat{\sigma}_v^2, \hat{\phi})^\top$ y $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\delta}})$
2. Se genera un vector t_1^* cuyos elementos son n copias independientes de una distribución $N(0, 1)$. Luego se construye los vectores bootstrap $\mathbf{u}^* = \hat{\sigma}_v t_1^*$ y $\mathbf{v}^* = (\mathbf{I}_n - \hat{\phi} \mathbf{W})^{-1} \mathbf{u}^*$, y se calcula el remuestreo de la expresión $\boldsymbol{\theta}^* = \mathbf{Z} \hat{\boldsymbol{\beta}} + \mathbf{v}^*$, considerando a $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\delta}}$ como valores verdaderos de los parámetros $\boldsymbol{\beta}$ y $\boldsymbol{\delta}$.
3. Se genera un vector t_2^* con n copias independientes de las generadas para t_1^* y se construye un vector de errores aleatorios $\mathbf{e}^* = \Psi^{1/2} t_2^*$.
4. Se construye una base de datos bootstrap del modelo, $\mathbf{Y}^* = \boldsymbol{\theta}^* + \mathbf{e}^* = \mathbf{Z} \hat{\boldsymbol{\beta}} + \mathbf{v}^* + \mathbf{e}^*$
5. Considerando a $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\delta}}$ como valores verdaderos de los parámetros $\boldsymbol{\beta}$ y $\boldsymbol{\delta}$, estimamos el modelo de la ecuación (9) a la base de datos bootstrap \mathbf{Y}^* , obteniendo estimaciones del "verdadero" parámetro $\hat{\boldsymbol{\beta}}^*$ y $\hat{\boldsymbol{\delta}}$ basado en la base de datos bootstrap \mathbf{Y}^* . Primero se calcula el estimador para $\hat{\boldsymbol{\beta}}$ teniendo en cuenta lo estimado para $\hat{\boldsymbol{\delta}}$

$$\tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\delta}}) = (\mathbf{Z}^\top \mathbf{V}^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{V}^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{Y}^*$$

Luego se estima $\hat{\boldsymbol{\beta}}^*$ basado en \mathbf{Y}^* , y finalmente el estimador de $\hat{\boldsymbol{\beta}}$ es calculado teniendo en cuenta $\hat{\boldsymbol{\beta}}^*$, teniendo como resultado al estimador $\tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\delta}}^*)$. Estos estimadores son llamados estimadores bootstrap.

6. Se calcula el estimador SBLUP bootstrap de la base de datos bootstrap \mathbf{Y}^* y considerando $\hat{\boldsymbol{\delta}}$ como el verdadero valor para $\boldsymbol{\delta}$,

$$\hat{\theta}_i^{*SBLUP}(\hat{\boldsymbol{\delta}}) = \mathbf{z}_i^\top \tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\delta}}) + \mathbf{b}_i^\top \mathbf{G}(\hat{\boldsymbol{\delta}}) \mathbf{V}(\hat{\boldsymbol{\delta}})^{-1} [\mathbf{Y}^* - \mathbf{Z} \tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\delta}})]$$

Se calcula también el estimador SEBLUP bootstrap utilizando $\hat{\boldsymbol{\delta}}^*$ en lugar del verdadero valor $\hat{\boldsymbol{\delta}}$

$$\hat{\theta}_i^{*SEBLUP}(\hat{\boldsymbol{\delta}}^*) = \mathbf{z}_i^\top \tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\delta}}^*) + \mathbf{b}_i^\top \mathbf{G}(\hat{\boldsymbol{\delta}}^*) \mathbf{V}(\hat{\boldsymbol{\delta}}^*)^{-1} [\mathbf{Y}^* - \mathbf{Z} \tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\delta}}^*)]$$

7. Se repite los pasos del 2 al 6 B veces. En la b -ésima replicación, obtenemos el estimador SBLUP bootstrap para el área i $\hat{\theta}_i^{*SBLUP(b)}$, $\hat{\boldsymbol{\delta}}^{*(b)}$ la estimación bootstrap de $\boldsymbol{\delta}$ y el estimador SEBLUP bootstrap $\hat{\theta}_i^{*(b)}(\hat{\boldsymbol{\delta}}^{*(b)})$ para el área i .

¹⁴Para mayor detalle acerca de la derivación del método bootstrap paramétrico y no paramétrico del modelo Fay-Herriot espacial, ver Molina, Salvati y Pratesi (2009)

8. El estimador bootstrap para $g_{3i}(\boldsymbol{\delta})$ está dada por la siguiente expresión

$$g_{3i}^{BP}(\hat{\boldsymbol{\delta}}) = B^{-1} \sum_{b=1}^B \left[\hat{\theta}_i^{*SEBLUP(b)}(\hat{\boldsymbol{\delta}}^{*(b)}) - \hat{\theta}_i^{*SEBLUP(b)}(\hat{\boldsymbol{\delta}}) \right]^2$$

Adicionalmente, el estimador bootstrap naive del ECM esta dado por

$$ECM^{BPna}[\hat{\theta}^{SEBLUP}(\hat{\boldsymbol{\delta}})] = B^{-1} \sum_{b=1}^B \left[\hat{\theta}_i^{*(b)}(\hat{\boldsymbol{\delta}}^{*SEBLUP(b)}) - \theta_i^{*(b)} \right]^2 \quad (30)$$

También se puede obtener otra estimación del ECM agregando las estimaciones bootstrap para $g_{3i}^{PB}(\hat{\boldsymbol{\delta}})$ y también las correspondientes a $g_{1i}(\hat{\boldsymbol{\delta}})$ y $g_{2i}(\hat{\boldsymbol{\delta}})$. Luego se incluye una estimación bootstrap con corrección de sesgo para $g_{1i}(\hat{\boldsymbol{\delta}}) + g_{2i}(\hat{\boldsymbol{\delta}})$ tal como lo propone Pfeffermann y Tiller (2006), obteniendo la siguiente expresión:

$$ECM^{BPcs}[\hat{\theta}^{SEBLUP}(\hat{\boldsymbol{\delta}})] = 2[g_{1i}(\hat{\boldsymbol{\delta}}) + g_{2i}(\hat{\boldsymbol{\delta}})] - B^{-1} \sum_{b=1}^B \left[g_{1i}(\hat{\boldsymbol{\delta}}^{*(b)}) + g_{2i}(\hat{\boldsymbol{\delta}}^{*(b)}) \right] + g_{3i}^{PB}(\hat{\boldsymbol{\delta}}) \quad (31)$$

donde $B^{-1} \sum_{b=1}^B \left[g_{1i}(\hat{\boldsymbol{\delta}}^{*(b)}) + g_{2i}(\hat{\boldsymbol{\delta}}^{*(b)}) \right]$ es la corrección de sesgo bootstrap para la expresión $g_{1i}(\hat{\boldsymbol{\delta}}) + g_{2i}(\hat{\boldsymbol{\delta}})$ definida en las ecuaciones (26) y (27).

5.6.2.2 Bootstrap no paramétrico

De acuerdo a Molina, Salvati y Pratesi (2009), en el método de bootstrap no paramétrico las estimaciones bootstrap de los efectos aleatorios $(u_1^*, \dots, u_m^*)^\top$ y los errores aleatorios $(e_1^*, \dots, e_m^*)^\top$ se obtienen a partir del remuestreo de la distribución empírica de la predicción de los errores aleatorios $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_m)^\top$ y los residuos $\hat{\mathbf{r}} = \hat{\mathbf{Y}} - \mathbf{Z}\hat{\boldsymbol{\beta}} - \hat{\mathbf{v}} = (\hat{r}_1, \dots, \hat{r}_n)^\top$. Este método nos permite prescindir de los supuestos de la distribución de los errores y efectos aleatorios, por consiguiente se espera que el método sea más robusto a la presencia de distribuciones no normales.

Bajo la versión matricial del modelo Fay-Herriot espacial descrito en la ecuación (7), tenemos que los estimadores BLUP para \mathbf{v} y \mathbf{u} son:

$$\tilde{\mathbf{v}}(\boldsymbol{\delta}) = \mathbf{G}(\boldsymbol{\delta})\mathbf{V}^{-1}(\boldsymbol{\delta})[\mathbf{Y} - \mathbf{Z}\tilde{\boldsymbol{\beta}}(\boldsymbol{\delta})], \quad \tilde{\mathbf{u}}(\boldsymbol{\delta}) = (\mathbf{I} - \phi\mathbf{W})\tilde{\mathbf{v}}(\boldsymbol{\delta})$$

y la matriz de varianzas y covarianzas de $\tilde{\mathbf{u}}(\boldsymbol{\delta})$ está dada por

$$\mathbf{V}_u(\boldsymbol{\delta}) = (\mathbf{I} - \phi\mathbf{W})\mathbf{G}(\boldsymbol{\delta})\mathbf{P}(\boldsymbol{\delta})\mathbf{G}(\boldsymbol{\delta})(\mathbf{I} - \phi\mathbf{W}^\top)$$

donde

$$\mathbf{P}(\boldsymbol{\delta}) = \mathbf{V}^{-1}(\boldsymbol{\delta}) - \mathbf{V}^{-1}(\boldsymbol{\delta})\mathbf{Z}(\mathbf{Z}^\top\mathbf{V}^{-1}(\boldsymbol{\delta})\mathbf{Z}^\top)^{-1}\mathbf{Z}^\top\mathbf{V}^{-1}(\boldsymbol{\delta})$$

Además, consideramos el vector de residuos

$$\tilde{\mathbf{r}}(\boldsymbol{\delta}) = \mathbf{Y} - \mathbf{Z}\tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) - \tilde{\mathbf{v}}(\boldsymbol{\delta})$$

Por otro lado, la matriz de covarianzas de $\tilde{\mathbf{r}}(\boldsymbol{\delta})$ es

$$\mathbf{V}_r(\boldsymbol{\delta}) = \boldsymbol{\Psi}\mathbf{P}(\boldsymbol{\delta})\boldsymbol{\Psi}$$

las matrices de covarianzas $\mathbf{V}_u(\boldsymbol{\delta})$ y $\mathbf{V}_r(\boldsymbol{\delta})$ no son diagonales, por lo tanto los elementos de los vectores $\tilde{\mathbf{u}}(\boldsymbol{\delta})$ y $\tilde{\mathbf{r}}(\boldsymbol{\delta})$ están correlacionados. Debido a que los métodos de remuestreo a partir de una distribución empírica funcionan bien bajo el supuesto de *iid*, la estandarización de estas matrices es un paso crucial para obtener elementos de varianza unitaria y no correlacionados. El proceso de estandarización se realiza a las matrices de

covarianzas estimadas $\hat{\mathbf{V}}_u = \hat{\mathbf{V}}_u(\hat{\boldsymbol{\delta}})$ y $\hat{\mathbf{V}}_r = \hat{\mathbf{V}}_r(\hat{\boldsymbol{\delta}})$ mediante el método de descomposición espectral.¹⁵

El procedimiento para realizar la estimación del método bootstrap no paramétrico se obtiene reemplazando los pasos 2 y 3 del bootstrap paramétrico por los nuevos pasos 2' y 3' descritos a continuación:

- 2'. Con las estimaciones para $\hat{\boldsymbol{\delta}} = (\hat{\phi}, \hat{\sigma}_u^2)y$ $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\delta}})$ obtenidos en el paso 1, se calcula las predicciones para \mathbf{v} y \mathbf{u} de la siguiente manera

$$\hat{\mathbf{v}} = \mathbf{G}(\hat{\boldsymbol{\delta}})\mathbf{V}\hat{\boldsymbol{\delta}}^{-1}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}), \quad \hat{\mathbf{u}} = (\mathbf{I} - \hat{\phi}\mathbf{W})\hat{\mathbf{v}} = (\hat{u}_1, \dots, \hat{u}_n)^\top$$

Luego se toma en cuenta la estimación para $\hat{\mathbf{u}}^S$ derivada de la descomposición espectral $\hat{\mathbf{u}}^S = \hat{\mathbf{V}}_u^{-1/2}\hat{\mathbf{u}}$ donde $\hat{\mathbf{V}}_u^{1/2}$ es la raíz cuadrada de la inversa generalizada de $\hat{\mathbf{V}}_u$ obtenida por descomposición espectral. Finalmente es conveniente reescalar los elementos \hat{u}_i^S de modo que se tenga una media muestral exactamente igual a cero y una varianza muestral $\hat{\sigma}_u^2$. Esto se logra mediante la transformación

$$\hat{u}_i^{SS} = \frac{\hat{\sigma}_u (\hat{u}_i^S - n^{-1} \sum_{j=1}^n \hat{u}_j^S)}{\sqrt{n^{-1} \sum_{d=1}^n (\hat{u}_d^S - n^{-1} \sum_{j=1}^n \hat{u}_j^S)^2}}, \quad i = 1, 2, \dots, n$$

Construimos el vector $\mathbf{u}^* = (u_1^*, \dots, u_n^*)^\top$, cuyos elementos son obtenidos mediante muestreo simple con reemplazamiento de tamaño n del conjunto $\hat{u}_1^{SS}, \dots, \hat{u}_n^{SS}$. Luego se obtiene $\mathbf{v}^* = (\mathbf{I} - \hat{\phi}\mathbf{W})^{-1} \mathbf{u}^*$ y se calcula la estimación bootstrap para el parámetro de interés $\boldsymbol{\theta}^* = \mathbf{Z}\hat{\boldsymbol{\beta}} + \mathbf{v}^* = (\theta_1^*, \dots, \theta_n^*)^\top$.

- 3'. Se calcula el vector de residuos $\hat{\mathbf{r}} = \mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}} - \hat{\mathbf{v}} = (\hat{r}_1, \dots, \hat{r}_n)^\top$. Estandarizamos los residuos a $\hat{\mathbf{r}}^S = \hat{\mathbf{V}}_r^{-1/2}\hat{\mathbf{r}} = (\hat{r}_1^S, \dots, \hat{r}_n^S)^\top$, donde $\hat{\mathbf{V}}_r = \boldsymbol{\Psi}\mathbf{P}(\hat{\boldsymbol{\delta}})\boldsymbol{\Psi}$ es la matriz de covarianzas estimadas y $\hat{\mathbf{V}}_r^{-1/2}$ es la raíz cuadrada de la inversa generalizada derivada de la descomposición espectral. Estandarizando los valores de la matriz tenemos lo siguiente:

$$\hat{r}_i^{SS} = \frac{\hat{r}_i^S - m^{-1} \sum_{j=1}^n \hat{r}_j^S}{\sqrt{m^{-1} \sum_{d=1}^n (\hat{r}_d^S - m^{-1} \sum_{j=1}^n \hat{r}_j^S)^2}}, \quad i = 1, 2, \dots, n$$

Construimos $\mathbf{r}^* = (r_1^*, \dots, r_n^*)^\top$, los cuales son obtenidos mediante muestreo simple con reemplazamiento de tamaño n del conjunto $\hat{r}_1^{SS}, \dots, \hat{r}_n^{SS}$. Luego se obtiene $\mathbf{e}^* = (e_1^*, \dots, e_n^*)^\top$, donde $e_i^* = \Psi_i^{1/2} r_i^*$, $i = 1, 2, \dots, n$. Este procedimiento, al igual que en el bootstrap paramétrico, genera estimadores análogos a las ecuaciones, las cuales se denotan como $ECM^{BPna}[\hat{\boldsymbol{\theta}}^{SEBLUP}(\hat{\boldsymbol{\delta}})]$ y $ECM^{BPcs}[\hat{\boldsymbol{\theta}}^{SEBLUP}(\hat{\boldsymbol{\delta}})]$ descritos en las ecuaciones (29) y (30).

¹⁵Para mayor detalle acerca del proceso de estandarización mediante el método de descomposición espectral, ver Molina, Salvati y Pratesi (2009).

6. Aplicaciones

En esta sección realizaremos dos aplicaciones del modelo Fay-Herriot y su variante espacial para el contexto peruano. La primera aplicación esta dirigida a predecir los niveles de prevalencia de anemia en niños menores de 5 años en aquellos distritos que tienen niveles de pobreza mayores al 55 % y la segunda aplicación se centra en calcular los niveles de riesgo de retraso del crecimiento en niños menores de 5 años en aquellos distritos que tienen niveles de pobreza mayores al 55 %. Para obtener estas estimaciones se utiliza los datos de la Encuesta Demográfica y de Salud Familiar (ENDES) para el año 2019 y del censo nacional realizado el 2017, ambas desarrolladas por el INEI. Además se utilizará para la estimación y predicción de los modelos Fay-Herriot el paquete estadístico *emdi*¹⁶, el cual se encuentra disponible para el software R.

6.1. Prevalencia en anemia en niños menores de 5 años

La anemia es una enfermedad que se caracteriza por tener una disminución en la concentración de la hemoglobina en la sangre. La OMS ha establecido los rangos para la detección de esta enfermedad en niños menores de 5 años. Si el nivel de hemoglobina se encuentra por debajo de 11g/L, el niño presenta anemia. Cabe resaltar que esta enfermedad presenta diversas intensidades, tal es así que si la hemoglobina se encuentra en el rango de 10 a 10.9 mg/l, se considera anemia leve; si se está en el rango de 7 a 9.99 mg/L anemia moderada y si es menos de 7 mg/L la anemia es severa. Cabe resaltar que para propósitos de esta investigación, se utilizará la definición general de anemia. Las consecuencias de esta enfermedad se reflejan en la reducción de la respuesta inmunológica, irritabilidad, palidez, entre otros problemas de salud, por lo que es de suma importancia tener información acerca de este indicador para realizar una mejor detección de esta enfermedad.

La estimación del indicador de prevalencia de anemia en niños menores de 5 años se realiza con los datos provenientes de la ENDES. En el módulo 74 de la encuesta, se encuentra información acerca del peso, talla y presencia de anemia en niños. Específicamente, la variable HC57 del módulo mencionado anteriormente permite calcular la prevalencia de anemia en niños de 6 a 59 meses de edad a nivel departamental y nacional bajo el criterio establecido por la OMS.

6.1.1. Estimación del estimador Horvitz-Thompson

Como primer paso para la estimación del modelo Fay-Herriot y posterior predicción de los valores de prevalencia de anemia tanto en distritos encuestados como no encuestados es necesario obtener información de los estimadores directos. Para ello se realiza la estimación de la prevalencia de anemia mediante el estimador directo Horvitz-Thompson y la varianza de este estimador en aquellos distritos con niveles de pobreza mayor 55 % que cuenten con información muestral proveniente de la ENDES 2019¹⁷.

6.1.2. Modelo Fay-Herriot

Luego de obtener las estimaciones directas de los indicadores de prevalencia de anemia se realiza la estimación del modelo Fay-Herriot para obtener las predicciones de prevalencia de anemia tanto en distritos encuestados como no encuestados y además, se estimó un modelo de regresión lineal para realizar una comparación entre ambos modelos. Para ello se toma como variable dependiente las estimaciones directas de los indicadores de prevalencia de anemia por distrito así como las estimaciones de las varianzas. Debido a que se tiene un conjunto de potenciales variables predictoras que explicarían el comportamiento de este indicador, se realiza un proceso de selección de variables para obtener un modelo de regresión parsimonioso.

6.1.2.1 Selección de variables

La variable dependiente a analizar es la prevalencia de anemia en niños menores de 5 años a nivel distrital calculada con los datos de la ENDES 2019. La información auxiliar de las variables explicativas proviene del censo nacional realizado el 2017, el cual contiene más de 30 variables con información acerca del acceso a servicios públicos (agua, desagüe, electricidad), acceso a seguro de salud (Essalud, SIS, entre otros), condiciones de la vivienda (material del techo de la vivienda, material del piso, entre otros), uso de combustible para cocina y altitud del distrito. Cabe resaltar que el objetivo de esta aplicación es obtener predicciones confiables de los indicadores de prevalencia de anemia.

¹⁶Para mayor información acerca del paquete *emdi*, ver Kreutzmann et al. (2019).

¹⁷Las estimaciones del estimador directo Horvitz-Thompson se realizaron a aquellos distritos que tienen como mínimo 2 o más observaciones y además que cuenten con observaciones suficientes para el cálculo del indicador de prevalencia de anemia.

La decisión del mejor modelo se tomó en base a la comparación de modelos provenientes de los criterios de información de Akaike, Bayesiano y Kullback (AIC,BIC,KIC) y sus versiones bootstrap y corrección de sesgos (AICc,AICb1,AICb2, entre otros) desarrollados por Marhuenda et al.(2014) para el modelo Fay-Herriot e implementadas computacionalmente por Kreutzmann et al. (2019).El uso de estos criterios permitirá obtener un modelo con un menor número de variables y con niveles de significancia aceptables.

El modelo final fue escogido en base al criterio de información AICc el cual desarrolla un modelo con 8 variables explicativas con niveles de significancia mayores a 10% . Cabe resaltar que el modelo final se estimó mediante el enfoque restringido de máxima verosimilitud.

Entre las variables explicativas del modelo se encuentran las siguientes:

- *Altitud*: Corresponde a la altitud del distrito. Áreas con mayor altitud sobre el nivel del mar incrementan la prevalencia de anemia. Estudios como el de Cornejo-Roselló et al.(2020) encuentran una relación positiva entre la altitud geográfica y niveles de anemia. Este particular fenómeno se debe a que, en lugares con mayor altitud, el aporte de oxígeno hacia los tejidos es menor lo cual disminuye la presión barométrica del niño reduciendo así las provisiones de ferritina (proteína de la sangre que contiene hierro).
- *Paredes de cemento*: Corresponde al porcentaje de viviendas con paredes de cemento. El material de construcción de una vivienda es una variable importante para medir la calidad de la vivienda. El uso de materiales precarios (adobe, madera, entre otros) para la construcción de la vivienda es uno de los factores de riesgo que enfrenta el niño debido al diseño y estructuras no adecuadas, hacinamiento, ventilación e iluminación inapropiada entre otros problemas. Es por ello, que el uso de materiales nobles como cemento o ladrillo permite mejorar la calidad de la vivienda y así mismo el bienestar del niño.
- *Acceso a desagüe*: Corresponde al porcentaje de la población que tiene acceso a desagüe mediante dentro de la vivienda. El acceso a saneamiento adecuado previene las deficiencias de higiene en la vivienda y evita las infecciones diarreicas que agudizan el problema de la anemia, es por ello que tiene un efecto negativo. El acceso a saneamiento básico es sumamente importante tal como lo refiere el Ministerio de Salud.¹⁸
- *Acceso a SIS*: Corresponde al porcentaje de la población con acceso al Seguro Integral de Salud (SIS).Un mayor acceso de la población al Sistema Integrado de Salud (SIS) reduce los indicadores de prevalencia de anemia. El tener acceso a un seguro de salud, permite que los niños puedan tener un adecuado control de su estado de salud y detectar posibles casos de anemia.
- *No tiene seguro*: Corresponde al porcentaje de la población que no cuenta con algún seguro médico. Una mayor población que no cuente con algún seguro médico incrementa los indicadores de prevalencia de anemia. El no estar asegurado puede incrementar el riesgo de que el niño no pueda acceder a sus controles pediátricos, perjudicando la salud del niño.
- *Castellano*: Corresponde al porcentaje de la población que tiene el idioma castellano como lengua con el que aprendió hablar. Los resultados de la regresión muestran que la población que tiene como lengua materna el castellano reduce la prevalencia de anemia. Este resultado muestra que la población que habla en su mayoría castellano tiene un mejor acceso a la información o tiene un mayor conocimiento de como prevenir la anemia.

¹⁸Ver MINSA(2016): <http://bvs.minsa.gob.pe/local/MINSA/3514.pdf>

Cuadro 1: Análisis de regresión

Variables	(Reg.Lineal)	(Fay-Herriot)
	Prev. de anemia	Prev. de anemia
Intercepto	0.767 (0.146)	0.795**** (0.139)
Altitud	0.019* (0.069)	0.021** (0.025)
Paredes de cemento	-0.204*** (0.072)	-0.218*** (0.071)
Acceso a desagüe	-0.130** (0.059)	-0.134** (0.058)
Acceso a SIS	-0.305** (0.137)	-0.350*** (0.131)
No tiene seguro	0.141 (0.183)	0.164 (0.176)
Castellano	-0.216**** (0.034)	-0.251**** (0.035)
Observaciones	234	234
Dominios en la muestra	234	234
Dominios fuera de la muestra	0	384
R^2	0.29	
R^2 ajustado	0.27	
AIC	-182.204	-188.658
BIC	-154.561	-161.015

Errores estándar en paréntesis

**** $p < 0.001$, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

6.1.3. Índice de Moran y Geary

La existencia de autocorrelación espacial entre los indicadores de prevalencia de anemia de los distritos analizados puede generar sesgos en la predicción, es por ello que se debe evaluar la existencia de correlación espacial mediante el índice de Moran y Geary. Para ello se plantea el criterio de vecinos más cercanos. Los resultados de los índices de Moran y Geary para se muestran en el cuadro 2.

Cuadro 2: Vecinos más cercanos-Test de autocorrelación espacial

Vecinos	I_{Moran}	p-value	I_{Geary}	p-value
3	0.17	1.71e-04	0.79	5.1e-05
4	0.18	8.92e-06	0.8	5.16e-06
5	0.21	9.3e-09	0.77	1e-08
6	0.21	5.2e-10	0.75	6.7e-11

Los resultados sugieren la presencia de autocorrelación espacial en los indicadores de prevalencia de anemia en los distritos analizados es por ello que se estimará la variante espacial del modelo Fay-Herriot considerando una matriz de autocorrelación espacial de 6 vecinos más cercanos.

6.1.4. Modelo Fay-Herriot espacial

Para la estimación del modelo Fay-Herriot espacial se consideró la matriz de autocorrelación espacial de 6 vecinos más cercanos y el modelo espacial auto-regresivo de orden uno SAR (1) desarrollado por Petrucci y Salvati (2006) e implementado computacionalmente por Kreutzmann et al. (2019). Además se estimó un modelo de regresión que incorpora la dependencia espacial en los errores. Los resultados de las regresiones se muestran en el cuadro 4. Al igual que el modelo Fay-Herriot, la decisión del mejor modelo se tomó en base a

la comparación de criterios de información de Akaike y Bayesiano (AIC y BIC). El modelo final fue escogido en base al criterio de información BIC el cual desarrolla un modelo con 4 variables explicativas con niveles de significancia mayores a 10% . Cabe resaltar que el modelo final se estimó mediante el enfoque restringido de máxima verosimilitud y considerando los métodos de bootstrap paramétrico y no paramétrico. Entre las variables explicativas del modelo se encuentran las siguientes:

- *Paredes de cemento*: Corresponde al porcentaje de viviendas con paredes de cemento. El material de construcción de una vivienda es una variable importante para medir la calidad de la vivienda. El uso de materiales precarios (adobe, madera, entre otros) para la construcción de la vivienda es uno de los factores de riesgo que enfrenta el niño debido al diseño y estructuras no adecuadas, hacinamiento, ventilación e iluminación inapropiada entre otros problemas. Es por ello, que el uso de materiales nobles como cemento o ladrillo permite mejorar la calidad de las familias.
- *Acceso a desagüe*: Corresponde al porcentaje de viviendas con acceso a desagüe. Un mayor acceso a saneamiento permite mejorar la calidad de vida de las personas y por consiguiente mejorar las condiciones de salud de los niños.
- *Acceso a SIS*: Corresponde al porcentaje de la población con acceso al Seguro Integral de Salud (SIS). Un mayor acceso de la población al Sistema Integrado de Salud (SIS) reduce los indicadores de prevalencia de anemia. El tener acceso a un seguro de salud, permite que los niños puedan tener un adecuado control de su estado de salud y detectar posibles casos de anemia.
- *Castellano*: Corresponde al porcentaje de la población que tiene el idioma castellano como lengua con el que aprendió hablar. Los resultados de la regresión muestran que la población que tiene como lengua materna el castellano reduce la prevalencia de anemia. Este resultado muestra que la población que habla en su mayoría castellano tiene un mejor acceso a la información o tiene un mayor conocimiento de como prevenir la anemia.
- *Combustible de biomasa*: Corresponde al porcentaje de los hogares que utiliza combustibles de biomasa como bosta o estiércol para la cocción de sus alimentos. Cabe resaltar que este tipo de combustible es considerado un combustible contaminante que puede tener efectos en la salud y desarrollo de niños. El resultado de la regresión muestra que un mayor uso de este combustible en hogares, generaría una mayor prevalencia de anemia. Mishra y Retherford (2007) en su estudio sobre el efecto de combustibles provenientes de biomasa (madera, estiércol, residuos de cultivos) ,y los niveles de anemia y retraso del crecimiento en niños, encuentran una relación positiva entre estas variables. El uso de este combustible genera altos niveles de contaminación en la vivienda , lo cual conlleva a un impacto directo en la salud de los niños.

Cuadro 3: Análisis de regresión

Variables	(Reg. espacial)	(Fay-Herriot espacial)
	Prev. de anemia	Prev. de anemia
Intercepto	0.905**** (0.057)	0.772**** (0.058)
Paredes de cemento	-0.229**** (0.082)	-0.221*** (0.077)
Acceso a desagüe	-0.149** (0.059)	-0.073 (0.056)
Acceso a SIS	-0.448**** (0.101)	-0.287*** (0.092)
Castellano	-0.187**** (0.040)	-0.221**** (0.041)
Combustible de biomasa	0.108** (0.071)	0.196** (0.086)
Observaciones	234	234
Dominios en la muestra	234	234
Dominios fuera de la muestra	0	384
<i>AIC</i>	-157.712	-200.984
<i>BIC</i>	-131.326	-173.341

Errores estándar en paréntesis

****p<0.001 ,*** p<0.01, ** p<0.05,* p<0.1

6.1.5. Predicción del porcentaje de niños menores a 5 años con anemia

En el cuadro 4 se presenta las predicciones de los niveles de anemia a nivel departamental del modelo Fay-Herriot, indicando los distritos con mayor y menor nivel de anemia. Cabe resaltar que las estimaciones a nivel departamental resultan del promedio de los resultados de los distritos con niveles de pobreza mayor al 55%. Los resultados a nivel departamental muestran que el departamento con mayor nivel de anemia es Puno con el 60.5% de la población de niños menores a 5 años con problemas de anemia, seguido de Cusco (48.4%) y Huancavelica (45.3%). En el departamento de Puno, el distrito con mayor nivel de anemia es Anapia con el 74.2% de la población de niños menores a 5 años con este problema de salud, en tanto que en el distrito con menor nivel de anemia es Cuyocuyo (51.5%). Por otro lado, el departamento con menor nivel de prevalencia de anemia es San Martín (24.1%), siendo los distritos de Awajún (40.5%) y Morales (12.8%) los que tienen una mayor y menor prevalencia de anemia respectivamente.

Cuadro 4: Predicciones de indicadores de prevalencia de anemia

Departamento	Estim.FH (%)	Distritos	Estim.FH (%)
Amazonas	27.0	El Cenepa	51.9
		Santo Tomas	15.0
Ancash	37.1	Marca	55.6
		Chimbote	16.8
Apurimac	40.8	Circa	49.8
		Abancay	24.1
Arequipa	34.7	Yanaquihua	52.9
		Miraflores	19.5
Ayacucho	43.3	Santa Ana de Huaycahuacho	54.7
		Huanta	22.8
Cajamarca	26.4	Chugur	39.3
		San Miguel	12.6
Callao	25.9	Ventanilla	25.9
Cusco	48.4	Alto Pichigua	60.8
		Kimbiri	27.5
Huancavelica	45.3	Andabamba	55.5
		Castrovirreyna	35.7
Huánuco	38.4	Colpas	38.4
		Pillco Marca	13.8
Junín	39.4	Carhuacallanga	69.6
		Chanchamayo	17.8
La Libertad	32.9	Pacanga	35.3
		Pueblo Nuevo	30.6
Lima	32.9	San Antonio	55.5
		San Miguel	19.3
Loreto	31.8	Saquena	32.2
		Jenaro Herrera	31.4
Pasco	34.9	Huayllay	48.1
		Paucartambo	24.6
Piura	29.3	Huarmaca	42.9
		Salitral	21.4
Puno	60.5	Anapia	74.2
		Cuyocuyo	51.5
San Martín	24.1	Awajún	40.5
		Morales	12.8
Tacna	43.9	Susapaya	43.9
Ucayali	44.8	Purús	54.0
		Neshuya	35.8

Por otro lado, en el cuadro 5 se presenta las predicciones de los niveles de anemia a nivel departamental del modelo Fay-Herriot espacial indicando los distritos con mayor y menor nivel de anemia. Cabe resaltar que las estimaciones a nivel departamental resultan del promedio de los resultados de los distritos con niveles de pobreza mayor al 55%. Los resultados a nivel departamental muestran que el departamento con mayor nivel de anemia es Puno con el 65.4% de la población de niños menores a 5 años con problemas de anemia, seguido de Cusco (50.6%) y Huancavelica (48.8%). En el departamento de Puno, el distrito con mayor nivel de anemia es Huata con el 75.6% de la población de niños menores a 5 años con este problema de salud en tanto que en el distrito con menor nivel de anemia es Cuyocuyo (51.6%). Por otro lado, el departamento con menor nivel de prevalencia de anemia es San Martín (25.3%), siendo los distritos de Awajún (43.9%) y Posic (13.8%) los que tienen una mayor y menor prevalencia de anemia respectivamente.

Cuadro 5: Predicciones de indicadores de prevalencia de anemia

Departamento	Estim.FH esp. (%)	Distritos	Estim.FH esp. (%)
Amazonas	27.2	El Cenepa	52.7
		Santo Tomas	16.3
Ancash	37.7	Marca	53.3
		Chimbote	17.3
Apurimac	41.5	Tumay Huaraca	54.7
		Abancay	24.9
Arequipa	33.8	Chilcaymarca	57.1
		Miraflores	19.3
Ayacucho	44.2	Coronel Castañeda	54.5
		Huanta	26.6
Cajamarca	26.5	Sexi	35.9
		San Miguel	11.7
Callao	26.3	Ventanilla	26.3
Cusco	50.6	Alto Pichigua	71.7
		Kimbiri	31.6
Huancavelica	48.8	Huayllahuara	61.1
		Pampas	39.3
Huánuco	34.6	Pampamarca	49.0
		Chanchamayo	12.9
Junín	39.2	San Juan de Jarpa	65.0
		Chanchamayo	12.9
La Libertad	32.5	Pacanga	33.0
		Pueblo Nuevo	31.9
Lima	32.5	Huancaya	49.5
		El Agustino	19.1
Loreto	31.7	Saquena	33.3
		Jenaro Herrera	30.1
Pasco	34.9	Huayllay	44.7
		Paucartambo	27.6
Piura	30.6	Huarmaca	42.1
		Buenos Aires	25.7
Puno	65.4	Huata	75.6
		Cuyocuyo	51.6
San Martín	25.3	Awajún	43.9
		Posic	13.8
Tacna	43.1	Susapaya	43.1
Ucayali	42.7	Purús	54.8
		Neshuya	31.3

A modo de resumen, en las figuras 2,3 y 4 se muestran las estimaciones del estimador directo Horvitz-Thompson y los estimadores EBLUP derivados del modelo Fay-Herriot y su variante espacial para los departamentos con mayor nivel de anemia (Puno), nivel promedio (Huánuco) y menor nivel respectivamente (San Martín). Cada uno de los gráficos esta ordenado por distritos de menor a mayor tamaño de observaciones. En las figuras se observa que la trayectoria de los estimadores EBLUP tienen una correlación muy fuerte con el estimador directo, siendo este último más volátil a medida que se reduce el tamaño de las observaciones.¹⁹

¹⁹Para mayor detalle acerca de los resultados para el resto de distritos de los departamentos analizados en esta investigación, ver Anexo B

Figura 2: Puno: Estimaciones

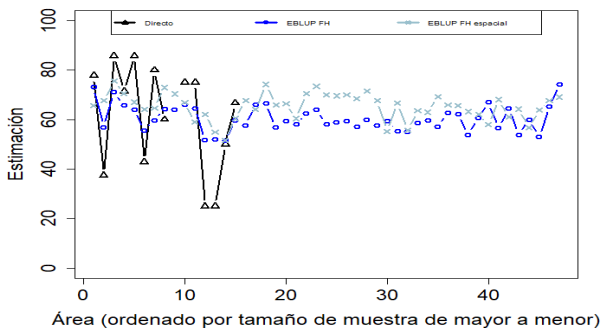


Figura 3: Huánuco: Estimaciones

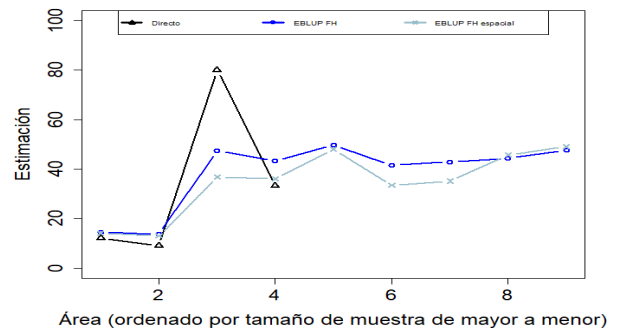
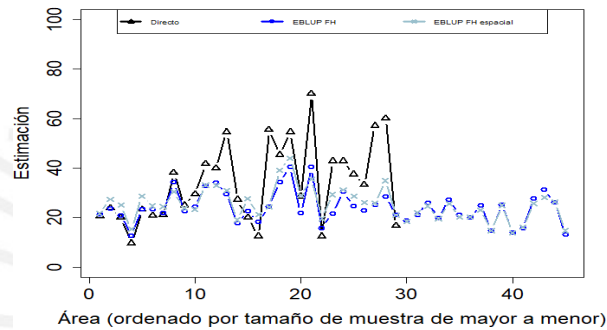


Figura 4: San Martín: Estimaciones



6.1.6. Estimación del error cuadrático medio de las predicciones del porcentaje de niños menores a 5 años con anemia

En esta sección, se realizará una comparación de las estimaciones del error cuadrático medio entre el modelo Fay-Herriot y su variante espacial, para seleccionar qué tipo de modelo es el mejor para la predicción de indicadores de prevalencia de anemia. Para ello se utilizará el error cuadrático medio como medida de precisión del modelo. En las figuras 5, 6 y 7 se muestran las estimaciones del error cuadrático medio para los modelos mencionados anteriormente en sus variantes bootstrap para el modelo espacial. El valor promedio del error cuadrático medio del modelo Fay-Herriot es de 0.0077 en tanto que para el modelo Fay-Herriot espacial bootstrap paramétrico con corrección de sesgo es de 0.0041, modelo Fay-Herriot espacial bootstrap no paramétrico; 0.0041 . El método de bootstrap que tiene un menor valor promedio es el paramétrico con corrección de sesgo. Además, se observa que el modelo Fay-Herriot espacial tiene menores niveles de error cuadrático medio y por consiguiente una mejor precisión en sus estimaciones respecto al modelo Fay-Herriot estándar. Debido a esta mejora en las predicciones, se elige el modelo Fay-Herriot espacial que utiliza el método de bootstrap paramétrico con corrección de sesgo para predecir los indicadores de prevalencia de anemia en niños menores de 5 años.

Figura 5: Error Cuadrático Medio: Distritos con menos de 10 observaciones

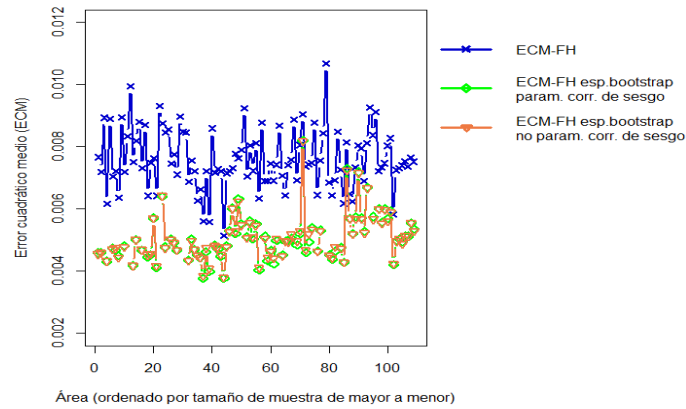


Figura 6: Error Cuadrático Medio: Distritos con 10 a 20 observaciones

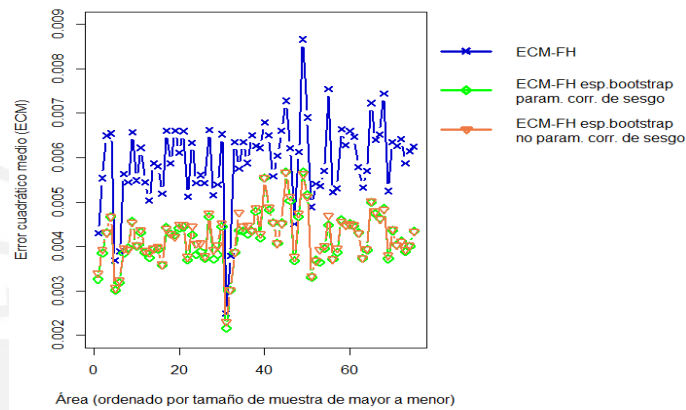
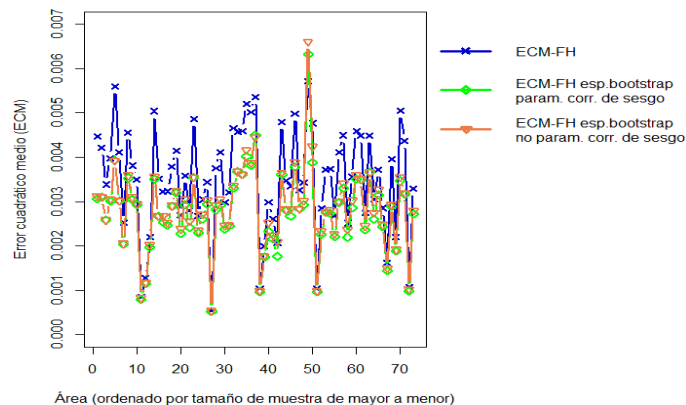


Figura 7: Error Cuadrático Medio: Distritos con más de 20 observaciones

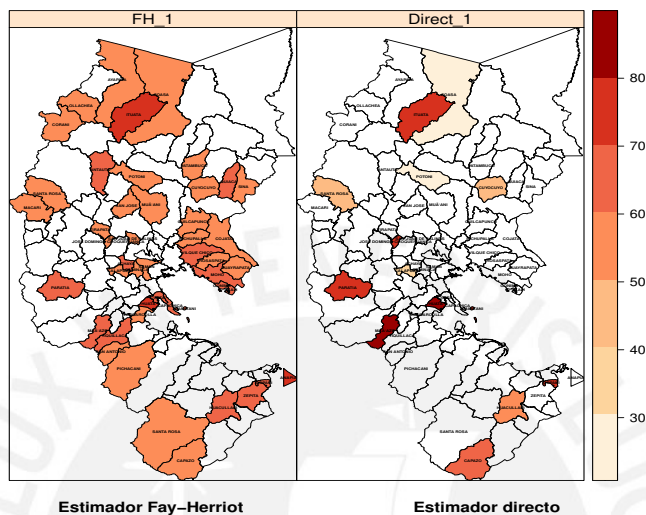


6.1.7. Mapas de prevalencia de anemia: Modelo Fay-Herriot

En esta sección analizamos las estimaciones derivadas del estimador directo y EBLUP del modelo Fay-Herriot mediante el uso de mapas para detectar diferencias en niveles de anemia entre distritos. En el caso del departamento de Puno, según las estimaciones derivadas del modelo Fay-Herriot (EBLUP), los tres distritos con

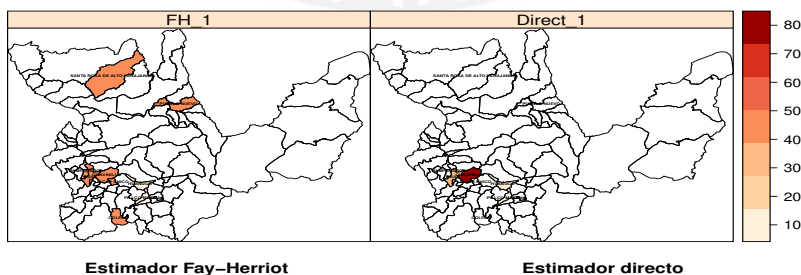
mayor prevalencia de anemia son Anapia (74.2%) seguido de Ituata (73.1%) y Huata (71.0%) y entre aquellos distritos con menores niveles de anemia se encuentran Cuyocuyo (51.5%), Potoni (51.8%) y Coasa (51.9%). Cabe resaltar que según la clasificación de prevalencia de anemia como problema de salud pública establecido por la OMS, si la prevalencia de anemia se encuentra entre 5% a 19.9% se considera un problema de salud pública leve, problema de salud pública moderado; entre 20% a 39% y problema de salud pública grave si tiene niveles mayores al 40%. Teniendo en cuenta estos criterios, se tiene que la totalidad de distritos con niveles de pobreza mayor al 55% en el departamento de Puno tienen problemas de salud pública grave que deberían ser atendidos.

Figura 8: Puno: Prevalencia de anemia por distritos



En el caso del departamento de Huánuco, según las estimaciones derivadas del modelo Fay-Herriot (EBLUP), los tres distritos con mayor prevalencia de anemia son Colpas (49.8%), Pampamarca (47.7%) seguido de Chavinillo (47.4%) y entre aquellos distritos con menores niveles de anemia se encuentran Pillco Marca (13.8%), Huánuco (14.5%) y Pueblo Nuevo (41.5%). Estos resultados muestran que el 77.8% de los distritos con niveles de pobreza mayor al 55% en el departamento de Huánuco tienen niveles de anemia en niños menores a 5 años por encima del 40% con lo cual se tiene una gran parte de distritos con problemas de salud pública grave.

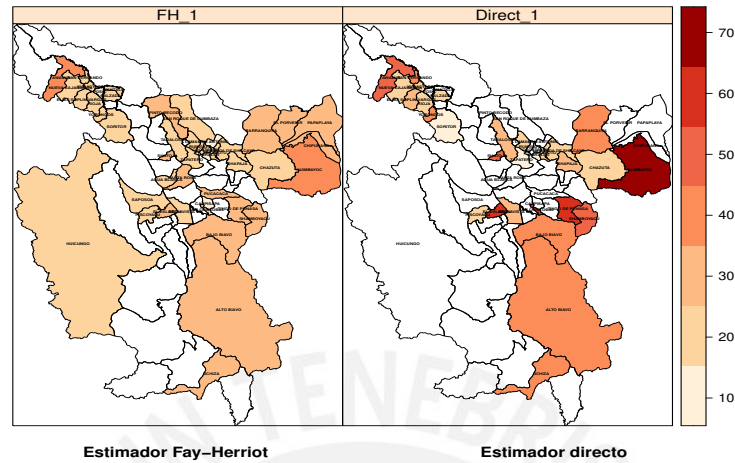
Figura 9: Huánuco: Prevalencia de anemia por distritos



Finalmente, en el caso del departamento de San Martín, según las estimaciones derivadas del modelo Fay-Herriot (EBLUP), los tres distritos con mayor prevalencia de anemia son Awajún (40.5%) seguido de Huimayoc (40.5%) y Shamboyacu (34.3%), y entre aquellos distritos con menores niveles de anemia se encuentran Morales (12.8%),

San Antonio (13.2%) y Posic (13.9%). Estos resultados muestran que el 17.8% de los distritos con niveles de pobreza mayor al 55% en el departamento de San Martín tienen niveles de anemia en niños menores a 5 años por encima del 30% con lo cual se tiene una minoría de distritos con problemas de salud pública moderado.

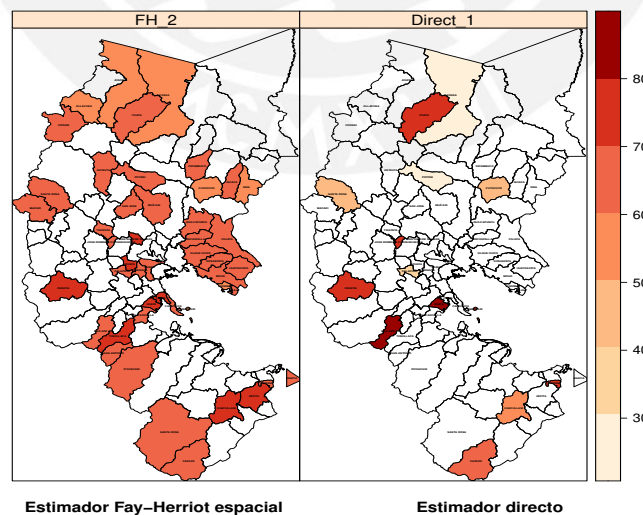
Figura 10: San Martín: Prevalencia de anemia por distritos



6.1.8. Mapas de prevalencia de anemia: Modelo Fay-Herriot espacial

Por último, analizamos las estimaciones derivadas del estimador directo y EBLUP del modelo Fay-Herriot espacial mediante el uso de mapas para detectar diferencias en niveles de anemia entre distritos. En el caso del departamento de Puno, los tres distritos con mayor prevalencia de anemia son Huata (75.6%) seguido de Coata (74.2%) y Achaya (73.4%) y entre aquellos distritos con menores niveles de anemia se encuentran Cuyocuyo (51.6%), Coasa (54.9%) y Ayapata (55.2%). Estos resultados muestran que la totalidad de distritos con niveles de pobreza mayor al 55% en el departamento de Puno tienen niveles de anemia en niños menores a 5 años por encima del 40% con lo cual se tiene una minoría de distritos con problemas de salud pública moderado.

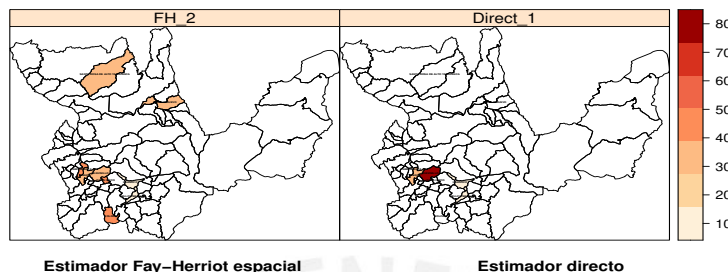
Figura 11: Puno: Prevalencia de anemia por distritos



En el caso del departamento de Huánuco, según las estimaciones derivadas del modelo Fay-Herriot espacial, los tres distritos con mayor prevalencia de anemia son Pampamarca (49.0%) seguido de Colpas (48.0%) y San Cristobal de Jacas Chico (45.8%) y entre aquellos distritos con menores niveles de anemia se encuentran Pillco Marca (13.0%), Huánuco (14.2%) y Pueblo Nuevo (33.4%). Estos resultados muestran que el 33.3% de los

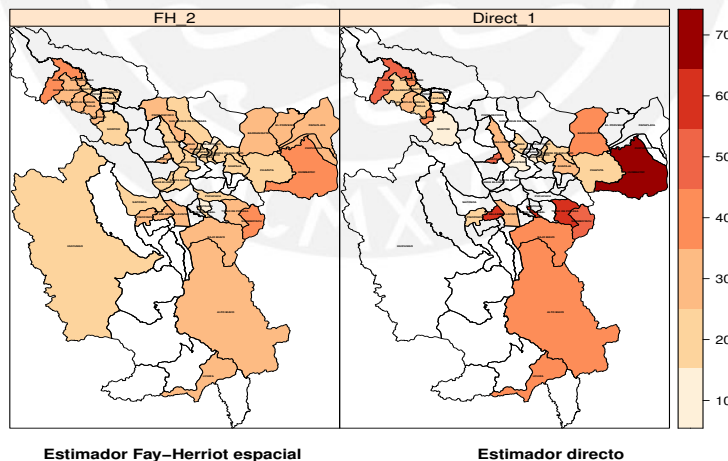
distritos con niveles de pobreza mayor al 55 % en el departamento de Huánuco tienen niveles de anemia en niños menores a 5 años por encima del 40 % con lo cual se tiene una gran parte de distritos con problemas de salud pública grave.

Figura 12: Huánuco: Prevalencia de anemia por distritos



Finalmente, en el caso del departamento de San Martín, según las estimaciones derivadas del modelo Fay-Herriot espacial, los tres distritos con mayor prevalencia de anemia son Awajún (43.9 %) seguido de Shamboyacu (39.2 %) y Huimbayoc (35.5 %), y entre aquellos distritos con menores niveles de anemia se encuentran Posic (13.8 %), Caspisapa (14.8 %) y San Antonio (14.8 %). Estos resultados muestran que el 20.0 % de los distritos con niveles de pobreza mayor al 55 % en el departamento de San Martín tienen niveles de anemia en niños menores a 5 años por encima del 30 % con lo cual se tiene una minoría de distritos con problemas de salud pública moderado.

Figura 13: San Martín: Prevalencia de anemia por distritos



6.2. Riesgo de retraso del crecimiento en niños menores de 5 años

El retraso del crecimiento es una consecuencia que experimentan los niños debido a una mala nutrición, entre otros factores. Este retraso pone en riesgo el desarrollo cognitivo y psicomotor del niño, aumentando su vulnerabilidad a diversas enfermedades. Según los patrones de crecimiento infantil propuestos por la OMS, el retraso de crecimiento se mide con el indicador "Talla para la edad". Aquellos niños que presenten una relación talla para la edad inferior a dos desviaciones estándar de las puntuaciones Z calculadas para la distribución de tallas para niños de la misma edad padecen de retraso del crecimiento para su edad. Las puntuaciones Z o puntuaciones de desviación estándar se utilizan para describir qué tan lejos se encuentra una medición respecto de la mediana (promedio).²⁰.

Para propósitos de esta investigación, se realizará la estimación del indicador de riesgo de retraso de crecimiento en niños menores de 5 años. Este indicador nos permitirá conocer la población infantil vulnerable a desarrollar retraso del crecimiento. Este indicador considera a los niños en riesgo de crecimiento a aquellos que se encuentran -1.28 desviaciones estándar por debajo del resto de la distribución del indicador Talla/Desviación estándar.

El cálculo de este indicador se realiza con los datos provenientes de la ENDES. En el módulo 74 de la encuesta, se encuentra información acerca del peso, talla y presencia de anemia en niños. Específicamente, la variable HC5 del módulo mencionado anteriormente permite calcular el riesgo de retraso del crecimiento en niños de 6 a 59 meses de edad a nivel departamental y nacional considerando la relación Talla/Desviación estándar del niño.

6.2.1. Estimación del estimador Horvitz-Thompson

Como primer paso para la estimación del modelo Fay-Herriot y posterior predicción de los valores de riesgo de retraso del crecimiento tanto en distritos encuestados como no encuestados es necesario obtener información de los estimadores directos. Para ello se realiza la estimación del indicador mencionado anteriormente mediante el estimador directo Horvitz-Thompson y la varianza en aquellos distritos con niveles de pobreza mayor 55 % que cuenten con información muestral proveniente de la ENDES 2019²¹.

6.2.2. Modelo Fay-Herriot

Luego de obtener las estimaciones directas de los indicadores de riesgo de retraso en el crecimiento se realiza la estimación del modelo Fay-Herriot para obtener las predicciones de riesgo de retraso en el crecimiento tanto en distritos encuestados como no encuestados y además, se estimó un modelo de regresión lineal para realizar una comparación entre ambos modelos. Para ello se toma como variable dependiente las estimaciones directas de los indicadores de riesgo de retraso en el crecimiento por distrito así como las estimaciones de las varianzas. Debido a que se tiene un conjunto de potenciales variables predictoras que explicarían el comportamiento de este indicador, se realiza un proceso de selección de variables para obtener un modelo de regresión parsimonioso.

6.2.2.1 Selección de variables

La variable dependiente a analizar es el retraso del crecimiento en niños menores de 5 años a nivel distrital. La información auxiliar de las variables explicativas proviene del censo nacional realizado el 2017 el cual contiene más de 30 variables con información acerca del acceso a servicios públicos (agua, desagüe, electricidad), acceso a seguro de salud (Essalud, SIS, entre otros), condiciones de la vivienda (material del techo de la vivienda, material del piso, entre otros), uso de combustible para cocina y altitud del distrito.

Entre las variables explicativas del modelo se encuentran las siguientes:

- *Altitud*: Corresponde a la altitud del distrito en miles de metros sobre el nivel del mar. Los niños que se encuentran en distritos con mayor altura no desarrollarían su potencial de crecimiento debido a los bajos niveles de oxígeno registrado en zonas con mayor altitud. En su mayoría, el riesgo de retraso del crecimiento se desarrolla durante el periodo previo e inmediato al nacimiento del niño, con lo cual resulta

²⁰Para mayor información acerca de los patrones de crecimiento infantil desarrollados por la OMS y el cálculo de las puntuaciones Z, ver https://www.who.int/childgrowth/training/module_c_interpreting_indicators.pdf

²¹Las estimaciones del estimador directo Horvitz-Thompson se realizaron a aquellos distritos que tienen como mínimo 2 o más observaciones y además que cuenten con observaciones suficientes para el cálculo del indicador de riesgo de retraso.

de suma importancia realizar una mayor control en mujeres embarazadas con la finalidad de controlar los efectos de la altitud en el feto (Baye Hirvonen(2020))

- *Acceso a SIS*: Corresponde al porcentaje de la población con acceso al Seguro Integral de Salud (SIS). Un mayor acceso de la población a un seguro de salud permite reducir los indicadores de riesgo de retraso del crecimiento en niños a través de controles nutricionales y consultas médicas para diagnosticar el problema de retraso a tiempo.
- *Acceso a electricidad*: Corresponde al porcentaje de viviendas con acceso a electricidad. Viviendas con mayor acceso a electricidad permite que se mejore el estado nutricional del niño, a través de mejores equipos de iluminación y calefacción, uso de cocina, entre otros. A su vez, el acceso a electricidad permite mejorar la difusión de información acerca del cuidado infantil a través de medios de comunicación masiva como la televisión o la radio (Fujii, Shonchoy, Xu (2018)).
- *Hogares con acceso a una refrigeradora o congeladora*: Corresponde al porcentaje de hogares con acceso a una refrigeradora o congeladora. Hogares que cuenten con refrigeradoras o congeladoras pueden mejorar su dieta alimenticia y comprar alimentos que requieran mayor refrigeración. Estudios como el de Martínez, Murguía, Rejas y Winters (2021) demuestran que la refrigeración de alimentos juegan un rol importante en la reducción de niveles de desnutrición en niños y por consiguiente en los niveles de retraso de crecimiento, a través de un mayor gasto en alimentos. Los autores además encuentran evidencia que en aquellos hogares que cuentan con un refrigerador, los niños son más altos para su edad después de dos años.
- *Analfabetismo*: Corresponde al porcentaje de la población que no sabe leer y escribir. El saber leer y escribir permite comprender mejor los problemas de salud de los niños por parte de los padres y así comunicar mejor los síntomas que el niño tiene. Según Khattak, Iqbal Ghazanfar (2017) el nivel educativo de los padres es uno de los principales factores para explicar el nivel de desnutrición en niño que a su vez trae un mayor riesgo de padecer retraso del crecimiento.
- *Combustible a gas*: Corresponde al porcentaje de hogares que utilizan combustible a gas para la cocción de sus alimentos. Hogares que utilicen combustibles limpios como el gas, reducen sus niveles de contaminación en el hogar y por consiguiente brindan un mejor ambiente para el desarrollo del niño.

Cuadro 6: Análisis de regresión

Variables	(Reg.Lineal)	(Fay-Herriot)
	Riesgo en el ret. de crecimiento	Riesgo en el ret. de crecimiento
Intercepto	0.595**** (0.107)	0.647**** (0.107)
Altitud	0.017 (0.010)	0.018* (0.009)
Acceso a SIS	-0.216** (0.1)	-0.327**** (0.1)
Acceso a electricidad	-0.178** (0.072)	-0.205**** (0.075)
Hogar con acces. a refrig.	-0.322**** (0.088)	-0.297**** (0.079)
Analfabetismo	0.911*** (0.341)	1.172*** (0.382)
Combustible a gas	-0.123 (0.078)	-0.153 (0.081)
Observaciones	235	235
Dominios en la muestra	235	235
Dominios fuera de la muestra	0	383
R^2	0.37	
R^2 ajustado	0.35	
AIC	-180.479	-180.379
BIC	-152.803	-152.703

Errores estándar en paréntesis

****p<0.001 ,*** p<0.01, ** p<0.05,* p<0.1

6.2.3. Índice de Moran y Geary

La existencia de autocorrelación espacial entre los indicadores de riesgo del retraso de crecimiento de los distritos analizados puede generar sesgos en la predicción, es por ello que se debe evaluar la existencia de correlación espacial mediante el índice de Moran y Geary . Para ello se plantea el criterio de vecinos más cercanos. Los resultados de los índices de Moran y Geary para se muestran en el cuadro 7.

Cuadro 7: Vecinos más cercanos-Test de autocorrelación espacial

Vecinos	I_{Moran}	p-value	I_{Geary}	p-value
3	0.137	0.002	0.843	0.001
4	0.148	0	0.823	0
5	0.156	0	0.842	0
6	0.131	0	0.876	0

6.2.4. Modelo Fay-Herriot espacial

Para la estimación del modelo Fay-Herriot espacial se consideró la matriz de autocorrelación espacial de 6 vecinos más cercanos y el modelo espacial auto-regresivo de orden uno SAR (1) desarrollado por Petrucci y Salvati (2006) e implementado computacionalmente por Kreutzmann et al. (2019). Además se estimó un modelo de regresión que incorpora la dependencia espacial en los errores. Los resultados de las regresiones se muestran en el cuadro 8. El modelo final fue escogido en base a las variables seleccionadas en el modelo Fay-Herriot y considerando los métodos de bootstrap paramétrico y no paramétrico. Entre las variables explicativas del modelo se encuentran las siguientes:

- *Acceso a SIS*: Corresponde al porcentaje de la población con acceso al Seguro Integral de Salud (SIS). Un mayor acceso de la población a un seguro de salud permite reducir los indicadores de riesgo de retraso del crecimiento en niños a través de mejores controles nutricionales y consultas médicas para diagnosticar el problema a tiempo.
- *Acceso a electricidad*: Corresponde al porcentaje de viviendas con acceso a electricidad. Viviendas con mayor acceso a electricidad permite que se mejore el estado nutricional del niño, a través de mejores equipos de iluminación y calefacción, uso de cocina, entre otros. A su vez, el acceso a electricidad permite mejorar la difusión de información acerca del cuidado infantil a través de medios de comunicación masiva como la televisión o la radio (Fujii, Shonchoy, Xu (2018)).
- *Hogares con acceso a una refrigeradora o congeladora*: Corresponde al porcentaje de hogares con acceso a una refrigeradora o congeladora. Hogares que cuenten con refrigeradoras o congeladoras pueden mejorar su dieta alimenticia y comprar alimentos que requieran mayor refrigeración. Estudios como el de Martínez, Murguía, Rejas y Winters (2021) demuestran que la refrigeración de alimentos juegan un rol importante en la reducción de niveles de desnutrición en niños y por consiguiente en los niveles de retraso de crecimiento, a través de un mayor gasto en alimentos. Los autores además encuentran evidencia que en aquellos hogares que cuentan con un refrigerador, los niños son más altos para su edad después de dos años.
- *Analfabetismo*: Corresponde al porcentaje de la población que no sabe leer y escribir. El saber leer y escribir permite comprender mejor los problemas de salud de los niños por parte de los padres y así comunicar mejor los síntomas que el niño tiene. Según Khattak, Iqbal Ghazanfar (2017) el nivel educativo de los padres es uno de los principales factores para explicar el nivel de desnutrición en niño que a su vez trae un mayor riesgo de padecer retraso del crecimiento.
- *Combustible a gas*: Corresponde al porcentaje de hogares que utilizan combustible a gas para la cocción de sus alimentos. Hogares que utilicen combustibles limpios como el gas, reducen sus niveles de contaminación en el hogar y por consiguiente brindan un mejor ambiente para el desarrollo del niño.

Cuadro 8: Análisis de regresión

Variables	(Reg. espacial)	(Fay-Herriot espacial)
	Riesgo en el ret. de crecimiento	Riesgo en el ret. de crecimiento
Intercepto	0.645**** (0.099)	0.741 **** (0.102)
Acceso a SIS	-0.227*** (0.099)	-0.344*** (0.106)
Acceso a electricidad	-0.175*** (0.071)	-0.212*** (0.078)
Hogar con acces. a refrig.	-0.331**** (0.086)	-0.322**** (0.085)
Analfabetismo	0.763*** (0.332)	1.058*** (0.391)
Combustible a gas	-0.135* (0.077)	-0.164* (0.059)
Observaciones	235	235
Dominios en la muestra	235	235
Dominios fuera de la muestra	0	383
AIC	-178.916	-178.341
BIC	-151.239	-150.664

Errores estándar en paréntesis

****p<0.001, *** p<0.01, ** p<0.05, * p<0.1

6.2.5. Predicción del porcentaje de niños menores a 5 años con riesgo de retraso del crecimiento

En el cuadro 9 se presenta las predicciones de los niveles de retraso de crecimiento a nivel departamental del modelo Fay-Herriot, indicando los distritos con mayor y menor nivel de riesgo de retraso. Cabe resaltar que

las estimaciones a nivel departamental resultan del promedio de los resultados de los distritos con niveles de pobreza mayor al 55 %. Los resultados a nivel departamental muestran que el departamento con mayor nivel de riesgo de retraso es Puno con el 50.4 % de la población de niños menores a 5 años con riesgo de tener retraso del crecimiento, seguido de Huancavelica (47.0 %) y Huánuco (46.5 %). En el departamento de Puno, el distrito de Amantani tiene el 67.7 % de la población de niños menores a 5 años con este problema de salud. Por otro lado, el departamento con menor nivel de prevalencia de retraso es Callao (17.9 %), siendo el distrito de Ventanilla el que tiene el 17.9 % de la población de niños menores a 5 años con riesgo de desarrollar retraso del crecimiento.

Cuadro 9: Predicciones de indicadores de riesgo de retraso del crecimiento

Departamento	Estim.FH	Distritos	Estim.FH
Amazonas	38.8	El Cenepa	73.0
		La Peca	12.4
Ancash	41.5	Amashca	64.3
		Chimbote	12.4
Apurímac	42.8	Pampachiri	63.1
		Abancay	17.8
Arequipa	29.0	Chilcaymarca	29.0
		Miraflores	9.4
Ayacucho	40.2	Coronel Castañeda	52.9
		San Juan Bautista	15.8
Cajamarca	44.2	Encañada	72.9
		Chilete	21.6
Callao	17.9	Ventanilla	17.9
Cusco	40.8	Alto Pichigua	60.9
		Santa Ana	8.2
Huancavelica	47.0	Palca	60.3
		Pampas	33.7
Huánuco	46.5	Pampamarca	59.6
		Huánuco	18.0
Junín	40.3	Rio Tambo	68.4
		Chanchamayo	19.7
La Libertad	24.9	Pacanga	27.3
		Pueblo Nuevo	22.5
Lima	29.4	Huancapón	53.7
		Chaclacayo	9.9
Loreto	43.3	Jenaro Herrera	43.6
		Saquena	43.1
Pasco	40.8	Huayllay	67.6
		Paucartambo	21.0
Piura	40.0	Sondorillo	66.2
		Chulucanas	22.7
Puno	50.4	Amantani	67.7
		Patambuco	39.4
San Martín	31.3	Shamboyacu	62.9
		Tarapoto	11.2
Tacna	46.1	Susapaya	46.1
Ucayali	44.5	Yurua	77.3
		Irazola	22.4

Por otro lado, en el cuadro 10 se presenta las predicciones de los niveles de riesgo del retraso del crecimiento a nivel departamental del modelo Fay-Herriot espacial indicando los distritos con mayor y menor nivel de riesgo de retraso. Cabe resaltar que las estimaciones a nivel departamental resultan del promedio de los resultados de los distritos con niveles de pobreza mayor al 55 %. Los resultados a nivel departamental muestran que el departamento con mayor nivel de riesgo del retraso del crecimiento es Puno con el 48.2 % de la población de niños menores a 5 años con problemas de riesgo de retraso de crecimiento, seguido de Huancavelica (46.2 %) y Tacna (45.8 %). En el departamento de Puno, el distrito con mayor nivel de retraso de crecimiento es Amantani

con el 63.9% de la población de niños menores a 5 años con este problema de salud, en tanto que en el distrito con menor nivel de retraso es Patambuco (37.3%). Por otro lado, el departamento con menor riesgo de retraso del crecimiento es Callao (17.9%), teniendo así al 17.9% de la población de niños menores a 5 años con riesgo de retraso de crecimiento.

Cuadro 10: Predicciones de indicadores de riesgo de retraso del crecimiento

Departamento	Estim.FH esp.	Distritos	Estim.FH esp.
Amazonas	39.3	Amashca	74.1
		La Peca	12.6
Ancash	42.8	Amashca	66.4
		Chimbote	12.7
Apurímac	44.2	Pampachiri	62.4
		Abancay	18.1
Arequipa	30.7	Chilcaymarca	50.3
		Miraflores	9.9
Ayacucho	39.7	Coronel Castañeda	55.9
		San Juan Bautista	14.7
Cajamarca	45.1	Encañada	74.7
		Chilete	24.9
Callao	17.9	Ventanilla	17.9
Cusco	40.0	Alto Pichigua	59.7
		Santa Ana	8.1
Huancavelica	46.2	Palca	58.3
		Pampas	33.3
Huánuco	45.0	Pampamarca	57.5
		Huánuco	17.9
Junín	40.3	Río Tambo	66.8
		Chanchamayo	19.7
La Libertad	24.5	Pacanga	26.4
		Pueblo Nuevo	22.7
Lima	29.6	Huancapón	54.2
		Chaclacayo	8.6
Loreto	42.9	Saquena	43.6
		Jenaro Herrera	42.3
Pasco	40.8	Huayllay	65.8
		Paucartambo	21.9
Piura	38.9	Sondorillo	63.4
		Chulucanas	22.4
Puno	48.2	Amantani	63.9
		Patambuco	37.3
San Martín	30.5	Shamboyacu	63.3
		Tarapoto	11.1
Tacna	45.8	Susapaya	45.8
Ucayali	43.7	Yurúa	75.3
		Irazola	22.3

A modo de resumen, en las figuras 14,15 y 16 se muestran las estimaciones del estimador directo Horvitz-Thompson y los estimadores EBLUP derivados del modelo Fay-Herriot y su variante espacial para los departamentos con mayor nivel de riesgo de retraso del crecimiento (Puno), nivel promedio (Piura) y menor nivel respectivamente luego de Callao y La Libertad (Lima)²². Cada uno de los gráficos esta ordenado por tamaño de muestra de mayor a menor. En las figuras se observa que la trayectoria de los estimadores EBLUP y EBLUP espacial tienen una correlación muy fuerte con el estimador directo, siendo este último más volátil a medida que se reduce el tamaño de la muestra.²³

²²Se decidió realizar el gráfico de las estimaciones de estos departamentos debido a que tanto Callao y La Libertad no tienen suficientes observaciones para realizar un análisis visual

²³Para mayor detalle acerca de los resultados para el resto de distritos de los departamentos analizados en esta investigación ,

Figura 14: Puno: Estimaciones

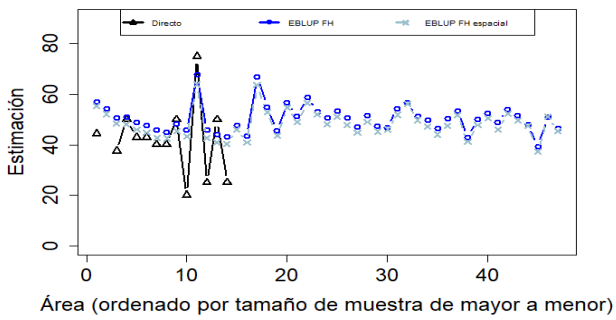


Figura 15: Piura: Estimaciones

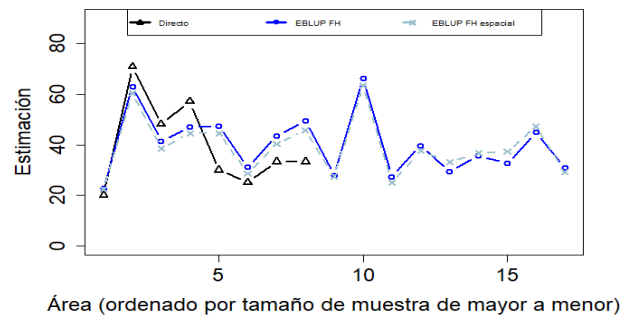
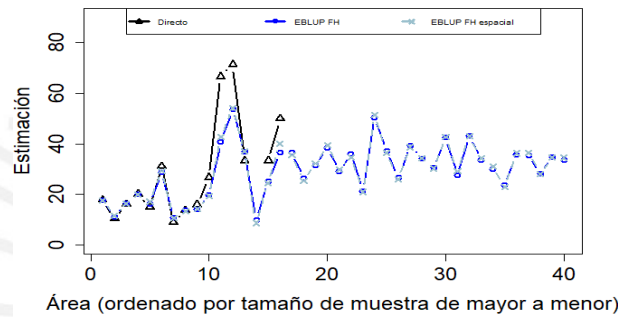


Figura 16: Lima: Estimaciones



6.2.6. Estimación del error cuadrático medio de las predicciones del porcentaje de niños menores a 5 años con riesgo de retraso del crecimiento

En esta sección, se realizará una comparación de las estimaciones del error cuadrático medio entre el modelo Fay-Herriot y su variante espacial, para seleccionar qué tipo de modelo es el mejor para la predicción de indicadores de riesgo de retraso del crecimiento. Para ello se utilizará el error cuadrático medio como medida de precisión del modelo. En las figuras 17, 18 y 19 se muestran las estimaciones del error cuadrático medio para los modelos mencionados anteriormente en sus variantes bootstrap para el modelo espacial. El valor promedio del error cuadrático medio del modelo Fay-Herriot es de 0.009 en tanto que para el modelo Fay-Herriot espacial bootstrap paramétrico con corrección de sesgo es de 0.0063 y modelo Fay-Herriot espacial bootstrap no paramétrico con corrección de sesgo; 0.0062. El método de bootstrap que tiene un menor valor promedio es el no paramétrico con corrección de sesgo. Además, se observa que el modelo Fay-Herriot espacial tiene menores niveles de error cuadrático medio y por consiguiente una mejor precisión en sus estimaciones respecto al modelo Fay-Herriot estándar. Debido a esta mejora en las predicciones, se elige el modelo Fay-Herriot espacial que utiliza el método de bootstrap no paramétrico para predecir los indicadores de riesgo de retraso del crecimiento en niños menores de 5 años.

ver Anexo B

Figura 17: Error Cuadrático Medio: Distritos con menos de 10 observaciones

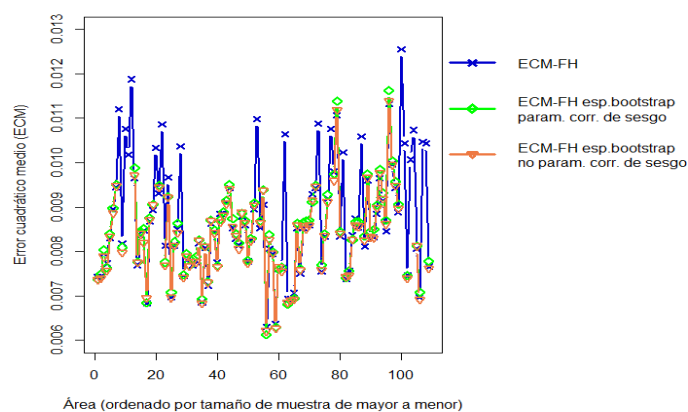


Figura 18: Error Cuadrático Medio: Distritos con 10 a 20 observaciones

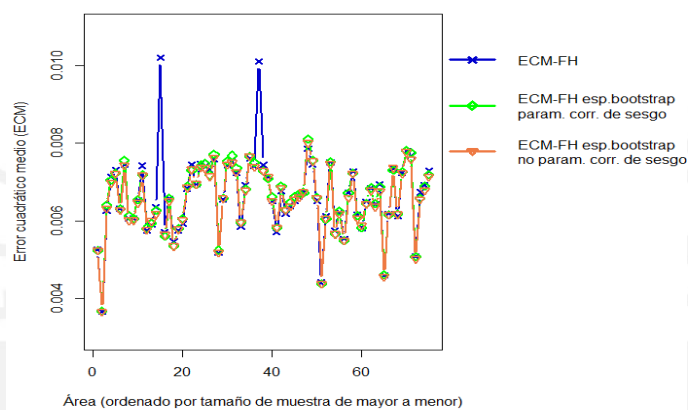
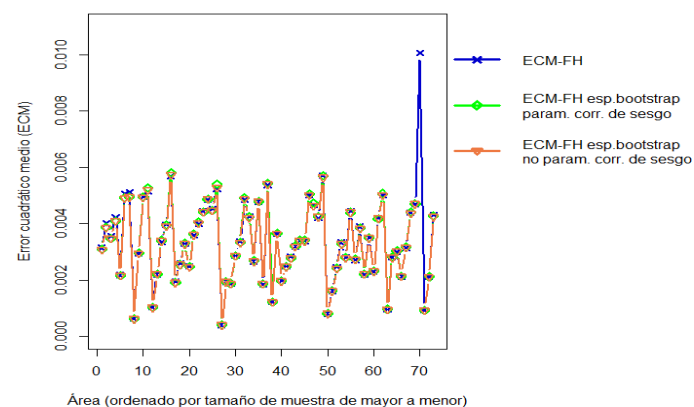


Figura 19: Error Cuadrático Medio: Distritos con más de 20 observaciones

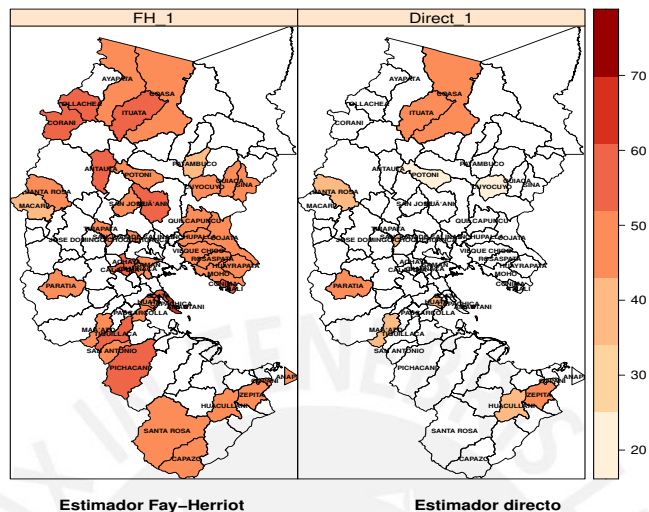


6.2.7. Mapas de riesgo de retraso del crecimiento: Modelo Fay-Herriot

En esta sección analizamos las estimaciones derivadas del estimador directo y EBLUP del modelo Fay-Herriot mediante el uso de mapas para detectar diferencias en niveles de riesgo de retraso del crecimiento entre distritos. En el caso del departamento de Puno, según las estimaciones derivadas del modelo Fay-Herriot (EBLUP),

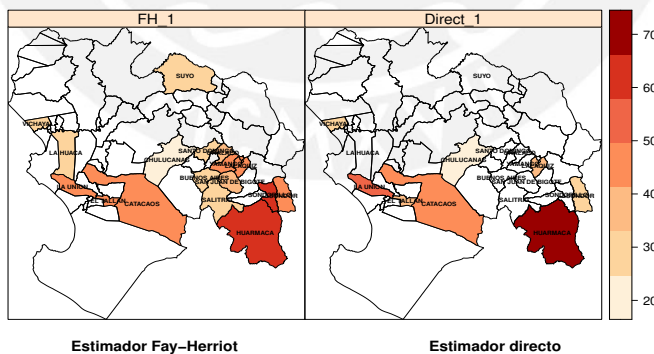
los tres distritos con mayores indicadores de riesgo de retraso del crecimiento son Amantani (67.7%) seguido de Capachica (67.1%) y Tiquillaca (58.9%) y entre aquellos distritos con menores niveles de riesgo de retraso se encuentran Patambuco (39.4%), Macari (42.9%) y Cuyocuyo (43.1%). Cabe resaltar que el 100% de los distritos analizados para Puno cuentan con niveles de riesgo de retraso del crecimiento mayores al 35% con lo cual muestra un problema de salud pública por atender.

Figura 20: Puno: Indicador de riesgo de retraso del crecimiento por distritos



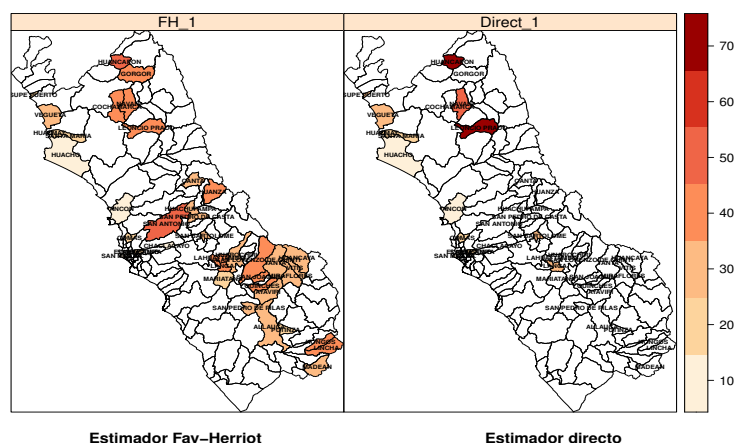
En el caso del departamento de Piura, según las estimaciones derivadas del modelo Fay-Herriot (EBLUP), los tres distritos con mayores indicadores de riesgo de retraso del crecimiento son Sondorillo (66.2%) seguido de Huarmaca (63.1%) y Lalaquiz (49.4%) y entre aquellos distritos con menores niveles de riesgo de retraso se encuentran Chulucanas (22.7%), Buenos Aires (27.3%) y Suyo (27.9%). Cabe resaltar que el 76.5% de los distritos analizados para Piura cuentan con niveles de riesgo de retraso del crecimiento mayores al 30%.

Figura 21: Piura: Indicador de riesgo de retraso de crecimiento por distritos



Finalmente, en el caso del departamento de Lima, según las estimaciones derivadas del modelo Fay-Herriot (EBLUP), los tres distritos con mayores indicadores de riesgo de retraso del crecimiento son Huancapón (53.7%) seguido de San Antonio (51.4%) y Lincha (43.2%) y entre aquellos distritos con menores niveles de riesgo de retraso se encuentran Chaclacayo (9.9%), Ancón (10.6%) y Huacho (10.9%). Cabe resaltar que el 50% de los distritos analizados para Lima cuentan con niveles de riesgo retraso de crecimiento mayores al 30%.

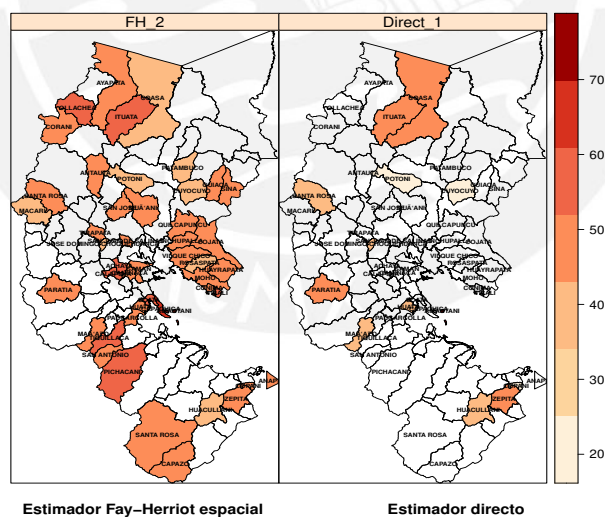
Figura 22: Lima: Indicador de riesgo de retraso del crecimiento por distritos



6.2.8. Mapas del riesgo de retraso del crecimiento: Modelo Fay-Herriot espacial

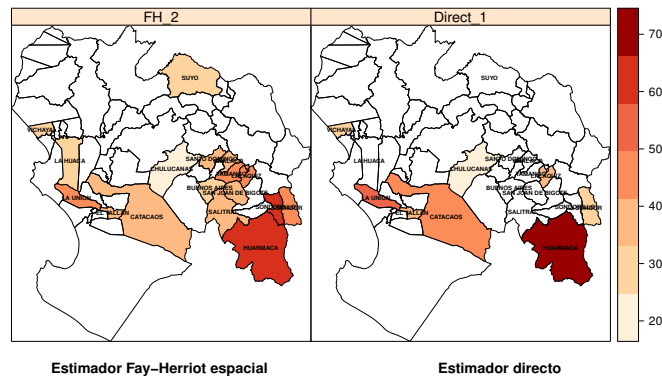
En esta sección analizamos las estimaciones derivadas del estimador directo y EBLUP del modelo Fay-Herriot espacial mediante el uso de mapas para detectar diferencias en niveles de riesgo de retraso del crecimiento entre distritos. En el caso del departamento de Puno, según las estimaciones derivadas del modelo Fay-Herriot espacial (EBLUP espacial), los tres distritos con mayores indicadores de riesgo de retraso del crecimiento son Amantani (63.9%) seguido de Capachica (63.8%) y Tiquillaca (56.5%) y entre aquellos distritos con menores niveles de anemia se encuentran Patambuco (37.3%), Cuyocuyo (40.3%) y Rosaspata (40.8%). Cabe resaltar que el 100% de los distritos analizados para Puno cuentan con niveles de retraso de crecimiento mayores al 35%.

Figura 23: Puno: Indicador de riesgo de retraso del crecimiento por distritos



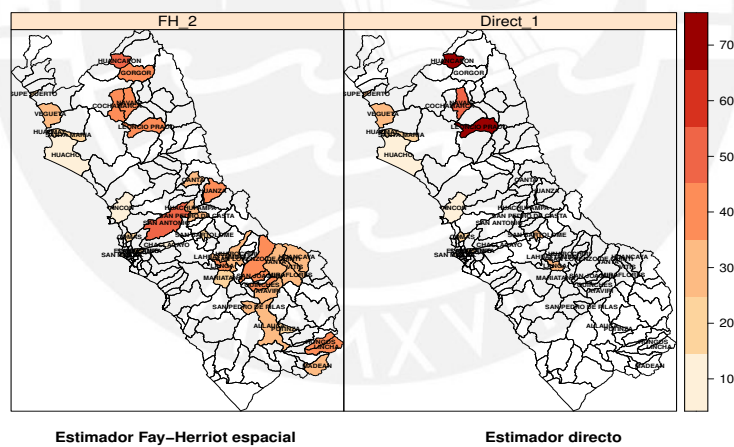
En el caso del departamento de Piura, según las estimaciones derivadas del modelo Fay-Herriot espacial (EBLUP espacial), los tres distritos con mayores indicadores de riesgo de retraso del crecimiento son Sondorillo (63.4%) seguido de Huarmaca (60.4%) y Yamango (47.3%) y entre aquellos distritos con menores niveles de retraso se encuentran Chulucanas (22.4%), Buenos Aires (24.9%) y Suyo (27.5%). Cabe resaltar que el 70.6% de los distritos analizados para Piura cuentan con niveles de retraso de crecimiento mayores al 30%.

Figura 24: Piura: Indicador de riesgo de retraso del crecimiento por distritos



Finalmente, en el caso del departamento de Lima, según las estimaciones derivadas del modelo Fay-Herriot espacial (EBLUP espacial), los tres distritos con mayores indicadores de riesgo de retraso del crecimiento son Huancapón (54.2%) seguido de San Antonio (51.4%) y Lincha (42.9%) y entre aquellos distritos con menores niveles de riesgo de retraso se encuentran Chaclacayo (8.6%), Ancón (10.5%) y Huacho (11.3%). Cabe resaltar que el 52.5% ninguno de los distritos analizados para Lima cuentan con niveles de retraso de crecimiento mayores al 30%.

Figura 25: Lima: Indicador de riesgo de retraso del crecimiento por distritos



7. Conclusiones

7.1. Conclusiones

Esta investigación ha permitido obtener predicciones de principales indicadores de salud en niños menores a 5 años como son la prevalencia de anemia y riesgo de retraso del crecimiento en distritos de todo el país con niveles de pobreza mayor al 55% . Para ello, se hace uso de modelos estadísticos, como es el caso del modelo Fay-Herriot y su variante espacial, que permiten afrontar el problema de falta de información en distritos con reducido tamaño de muestra . Estos modelos permiten obtener predicciones de estos indicadores a nivel distrital con un mayor nivel de precisión, combinando datos de una encuesta como la ENDES y datos poblacionales como el censo de población del 2017.

Los resultados obtenidos para las predicciones de anemia muestran un grave problema de salud pública en la gran parte de distritos pobres del país, como es el caso de los distritos de los departamentos de Puno, Cusco y Huancavelica que tienen indicadores de anemia por encima del 48.5% ²⁴. Entre las variables que explicarían los altos niveles de anemia se encuentran la altitud, el no tener acceso a seguro de salud y utilizar combustibles de biomasa. Por otro lado, aquellas variables que reducirían los niveles de anemia están ligadas a la mejora de las condiciones de la vivienda como el uso de materiales resistentes como cemento, acceso a servicios públicos como desagüe, acceso a servicios de salud como el SIS y el tener como lengua materna el castellano.

Respecto a los resultados obtenidos para las predicciones de riesgo de retraso de crecimiento muestran un problema preocupante de salud pública , como es el caso de los distritos de los departamentos de Puno, Huancavelica y Tacna que tienen indicadores de riesgo de retraso de crecimiento por encima del 45.5%. Entre las variables que explicarían los altos niveles de riesgo de retraso de crecimiento se encuentran la altitud y el analfabetismo. Por otro lado, aquellas variables que reducirían los niveles de riesgo de retraso están ligadas al acceso de servicios públicos como acceso a electricidad , acceso a seguro de salud como el SIS, uso de electrodomésticos refrigerantes y uso de combustibles no contaminantes como es el caso del gas.

Es por ello que es necesario mejorar las condiciones de vida en estos distritos para reducir el impacto negativo que tiene tanto la anemia como el riesgo de retraso del crecimiento en niños menores de 5 años.

7.2. Sugerencias para investigaciones futuras

- Realizar la predicción de los indicadores de prevalencia de anemia y riesgo de retraso del crecimiento en el resto de distritos del país con la finalidad de obtener un mayor nivel de desagregación a nivel distrital, provincial y departamental.
- Considerar otros tipos de matrices de pesos espaciales para la estimación de la variante espacial del modelo Fay-Herriot. Estas nuevas matrices permitirán obtener diversos criterios de contigüidad entre distritos.
- Considerar otras variantes de modelos predictivos para áreas pequeñas como son el caso bayesiano del modelo Fay-Herriot, modelos de regresión con errores anidados, entre otros. ²⁵

²⁴Cabe resaltar que la prevalencia de anemia a nivel nacional en niños menores a 5 años para el 2019 fue de 29.5%

²⁵Para mayor información acerca de los más recientes desarrollos en el campo de la estimación en áreas pequeñas, ver Sugawara y Kubokawa (2020)

8. Referencias

- Avila-Valdez, J. L., Huerta, M., Leiva, V., Riquelme, M., Trujillo, L. (2020). The Fay-Herriot model in small area estimation: EM algorithm and application to official data. *REVSTAT–Statistical Journal*, 18(5), 613-635.
- Battese, G. E., Harter, R. M., y Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401), 28-36.
- Baye, K., Hirvonen, K. (2020). Evaluation of linear growth at higher altitudes. *JAMA pediatrics*, 174(10), 977-984.
- Battese, G. E., Harter, R. M., Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401), 28-36.
- Benavent, R., y Morales, D. (2016). Multivariate Fay–Herriot models for small area estimation. *Computational Statistics Data Analysis*, 94, 372-390.
- Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., Pebesma, E. J. (2008). *Applied spatial data analysis with R* (Vol. 747248717, pp. 237-268). New York: Springer.
- Borooah, V. K. (2004). The role of maternal literacy in reducing the risk of child malnutrition in India. *Journal of Quantitative Economics*, 2(2), 186-202.
- Bucyibaruta, G. (2014). *Statistical Models for Small Area Estimation* (Doctoral dissertation, Centro de Investigación en Matemáticas).
- Cerdá-Hernández y Sikov (2021). *Introducción a la estadística espacial para ingeniería teoría y aplicaciones con R*. Universidad Nacional de Ingeniería.
- Cornejo-Roselló, I., Rojas-Pauca, S., Alvarez-Cervantes, G., Bernabé-Ortiz, J. C., Gallegos, A., Mercado-Mamani, S. Toia-Larsen, M. (2020). Childhood anemia in populations residing at different geographical altitudes of Arequipa, Peru: A descriptive and retrospective study. *Medwave*, 20(7), e8004-e8004.
- Cossio-Bolaños, M. A., de Arruda, M., Álvarez, V. N., Alonso, J. L. (2011). Efectos de la altitud sobre el crecimiento físico en niños y adolescentes. *Revista Andaluza de Medicina del Deporte*, 4(2), 71-76.
- Chi, G., y Zhu, J. (2019). *Spatial regression models for the social sciences*. *SAGE publications*.
- Cho, M. J., Eltinge, J. L., Gershunskaya, J., Huff, L. (2002). Evaluation of generalized variance function estimators for the US Current Employment Survey. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 534-539).
- Datta, G. S. (2009). Model-based approach to small area estimation. *Handbook of statistics*, 29, 251-288.
- Datta, G. S., Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 613-627.
- Dorfman, A. H. (1993). A comparison of design-based and model-based estimators of the finite population distribution function. *Australian Journal of Statistics*, 35(1), 29-41.
- Escobal, J. y Armas, C. (2015). El uso de encuestas y censos agropecuarios para desarrollar una tipología de la pequeña y mediana agricultura familiar en el Perú (pp. 15-86). *MISC*.

- Eurostat (2014) Handbook on Methodology of Modern Business Statistics. Methodologies and Working papers.
- EUSTAT (2005). Cálculo de coeficientes de variación para diferentes estimadores directos e indirectos utilizados en las encuestas económicas de Eustat.
- 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*, 74, 269–277.
- Fujii, T., Shonchoy, A. S., Xu, S. (2018). Impact of electrification on children’s nutritional status in rural Bangladesh. *World Development*, 102, 315-330
- Garrudo, A., Montero, M., Pita, G., y Díaz, M. E. (2018). Estimación de la prevalencia de anemia en áreas pequeñas: un enfoque multinivel. *Investigación Operacional*, 38(1), 19-29.
- Gałecki, A., Burzykowski, T. (2013). Linear mixed-effects model. In *Linear mixed-effects models using R* (pp. 245-273). Springer, New York, NY.
- Ghosh, M., y Rao, J. N. (1994). Small area estimation: an appraisal. *Statistical science*, 9(1), 55-76.
- Goodchild M. (2008) Data Analysis, Spatial. In: Shekhar S., Xiong H. (eds) *Encyclopedia of GIS*. Springer, Boston, MA.
- Gómez, M. H., Cid, J. C., Paz, J. A. (2012). Introduction to Spatial Econometrics: An application to the study of fertility in Argentina using R.
- Gomez Arias, G. L. del C. (2018). Efectos del Seguro Integral de Salud sobre el estado de salud : aplicación de regresión discontinua. (Tesis de maestría, Universidad del Pacífico, Lima, Perú). Recuperado de <http://repositorio.up.edu.pe/handle/11354/2196>
- Harmening, S., Kreutzmann, A., Pannier, S., Salvati, N., Schmid, T. (2020). A Framework for Producing Small Area Estimates Based on Area-Level Models in R.
- Henderson, C. (1950). Estimation of genetic parameters. *The Annals of Mathematical Statistics*, 21, 309–310.
- Hidiroglou, M. A., You, Y. (2016). Comparison of unit level and area level small area estimators. *Survey Methodology*, 42(1), 41-61.
- Horvitz, D. G., Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260), 663-685.
- INEI (2018). Línea de base de los principales indicadores disponibles de los Objetivos de Desarrollo Sostenible (ODS). Recuperado de https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1578/libro.pdf
- James, W. y Stein, C. (1992). Estimation with quadratic loss. In *Breakthroughs in statistics* (pp. 443-460). Springer, New York, NY.
- Khattak, U. K., Iqbal, S. P., Ghazanfar, H. (2017). The role of parents’ literacy in malnutrition of children under the age of five years in a semi-urban community of Pakistan: a case-control study. *Cureus*, 9(6).

Kreutzmann A, Pannier S, Rojas-Perilla N, Schmid T, Templ M, Tzavidis N (2019). “The R Package emdi for Estimating and Mapping Regionally Disaggregated Indicators.” *Journal of Statistical Software*, 91(7), 1–33. doi: 10.18637/jss.v091.i07.

Kwami, C. S., Godfrey, S., Gavilan, H., Lakhanpaul, M., Parikh, P. (2019). Water, sanitation, and hygiene: linkages with stunting in rural Ethiopia. *International journal of environmental research and public health*, 16(20), 3793.

MacGibbon, B., y Tomberlin, T. J. (1987). Small area estimates of proportions via empirical Bayes techniques. *Faculty of Commerce and Administration, Concordia University*.

Marhuenda, Y., Morales, D., del Carmen Pardo, M. (2014). Information criteria for Fay–Herriot model selection. *Computational Statistics Data Analysis*, 70, 268-280.

Miltiadou M (2020). “Measuring and Reporting Reliability of Labour Force Survey and Annual Population Survey Estimates Force Survey and Annual Population Survey Estimates.” UK Office for National Statistics. Recuperado de <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/measuringandreportingreliabilityoflabourforcesurveyandannualpopulationsurveyestimates>.

Marhuenda, Y., Morales, D., del Carmen Pardo, M. (2014). Information criteria for Fay–Herriot model selection. *Computational Statistics Data Analysis*, 70, 268-280.

Martinez, S., Murguia, J. M., Rejas, B., Winters, S. (2021). Refrigeration and child growth: What is the connection?. *Maternal child nutrition*, 17(2), e13083.

Ministerio de Salud del Perú (2016). Plan Nacional Para la Reducción de la Desnutrición Infantil y la Prevención de la Anemia en el País: 2014-2016. Lima, Perú.

Ministerio de Salud del Perú (2017) Documento Técnico. Plan Nacional para la Reducción y Control de la Anemia Materno Infantil y la Desnutrición Crónica Infantil en el Perú: 2017-2021.

Molefe, W. B. (2011). Sample design for small area estimation.

Molina, I., Salvati, N., Pratesi, M. (2009). Bootstrap for estimating the MSE of the Spatial EBLUP. *Computational Statistics*, 24(3), 441-458.

Molina, I., Marhuenda, Y. (2015). sae: An R Package for Small Area Estimation. *R J.*, 7(1), 81.

Oficina Europea de Estadística. (s.f) Module 1. Methods of statistical inference: design-based, model-based and algorithm-based. Comisión Europea.

Pajuelo Ramírez, J., Miranda Cuadros, M., Bernui Leo, I. (2017). Asociación entre altitud de residencia y malnutrición en niños peruanos menores de cinco años. *Acta Médica Peruana*, 34(4), 259-265.

Pfeffermann, D. (2002). Small area estimation-new developments and directions. *International Statistical Review*, 70(1), 125-143.

Pfeffermann, D., Tiller, R. (2005). Bootstrap Approximation to Prediction MSE for State–Space Models with Estimated Parameters. *Journal of Time Series Analysis*, 26(6), 893-916

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28(1), 40-68.

- Pratesi, M., Salvati, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical methods and applications*, 17(1), 113-141.
- Porter, A. T., Holan, S. H., Wikle, C. K., y Cressie, N. (2014). Spatial Fay–Herriot models for small area estimation with functional covariates. *Spatial Statistics*, 10, 27–42.
- 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*, 85, 163-171.
- Rao, J.N.K y Molina, I (2015). Small Area Estimation. Second edition, *John Wiley and Sons, Inc. Hoboken*.
- Romero Romero, J.J. (2018). *Un modelo Fay-Herriot espacial para estimar la tasa de embarazo adolescente en Colombia*.
- Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57(2), 377-387.
- Sahai, H., Ojeda, M. M. (2007). *Analysis of Variance for Random Models, Volume 2: Unbalanced Data: Theory, Methods, Applications, and Data Analysis*. Springer Science Business Media.
- Schabenberger, O., Gotway, C. A. (2005). *Statistical methods for spatial data analysis* (No. LC-0419). Chapman Hall/CRC.
- Singh, B. B., Shukla, G. K., Kundu, D. (2005). Spatio-temporal models in small area estimation. *Survey Methodology*, 31(2), 183.
- Smith, Tony E. (2020). Notebook for spatial data analysis: Part III. Areal Data Analysis. *Spatial Data Analysis with GIS*. University of Pennsylvania. Recuperado de https://www.seas.upenn.edu/~ese502/NOTEBOOK/Part_III/2_Modeling_the_Spatial_Structure_of_Areal_Units.pdf
- Srinivasan, S. (2008). *Spatial Regression Models*.
- Sugasawa, S., y Kubokawa, T. (2020). Small area estimation with mixed models: a review. *Japanese Journal of Statistics and Data Science*, 3(2), 693-720.
- Tasnim, T., Dasvarma, G., Mwanri, L. (2017). Housing conditions contribute to underweight in children: An example from rural villages in Southeast Sulawesi, Indonesia. *Journal of Preventive Medicine and Public Health*, 50(5), 328.
- Ubaidillah, A., Notodiputro, K. A., Kurnia, A., y Mangku, I. W. (2019). Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia. *Journal of Applied Statistics*
- Van den Brakel, J. A., Bethlehem, J. (2008). *Model-based estimation for official statistics*. Statistics Netherlands.
- Valdivieso Serrano, L. (2020). *Notas de Técnicas de Muestreo*.
- Verbeke, G. y Molenberghs, G. (2000). A model for Longitudinal Data. Linear mixed models for longitudinal data, 19-29. *Springer Series in Statistics*
- Waller, L. A., Gotway, C. A. (2004). *Applied spatial statistics for public health data*. John Wiley Sons.

Wawrowski, Ł. (2016). The spatial fay-herriot model in poverty estimation. *Folia Oeconomica Stetinensia*, 16(2), 191-202.

Wu, Yunhan, Zehang Richard Li, Benjamin K. Mayala, Houjie Wang, Peter Gao, Johnny Paige, Geir-Arne Fuglstad, Caitlin Moe, Jessica Godwin, Rose E. Donohue, Trevor N. Croft, and Jon Wakefield, 2021. Spatial Modeling for Subnational Administrative Level 2 Small-Area Estimation. DHS Spatial Analysis Reports No. 21. Rockville, Maryland, USA: ICF.

Xie, D., Raghunathan, T. E., y Lepkowski, J. M. (2007). Estimation of the proportion of overweight individuals in small areas—a robust extension of the Fay–Herriot model. *Statistics in medicine*, 26(13), 2699-2715.

You, Y. (2021). Small area estimation using Fay-Herriot area level model with sampling variance smoothing and modeling. *Survey Methodology*, 47(2), 361-371.



9. Anexos

9.1. Anexo A

9.1.1. Prevalencia de anemia en niños menores a 5 años

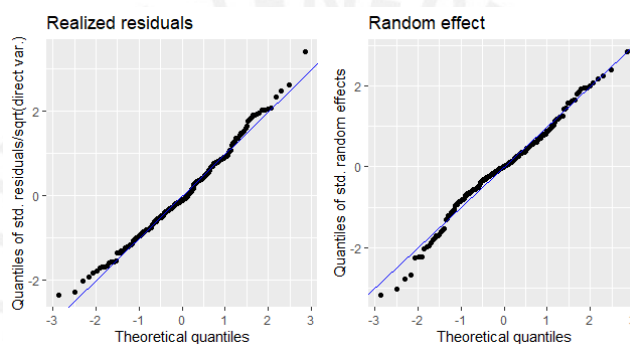
9.1.1.1 Modelo Fay-Herriot: Resultados prueba de diagnóstico

Los resultados de la prueba de Shapiro-Wilk que se muestran en el cuadro 12 indican que la normalidad de los errores no se rechaza. El gráfico de residuos del modelo presente en la figura 41 confirma lo mencionado anteriormente.

Cuadro 11: Pruebas de diagnóstico para residuos

VARIABLES	Asimetría	Curtosis	Shapiro-Wilk	p-value
Residuos estandarizados	0.35789	3.10546	0.99	0.12
Efectos aleatorios	-0.24983	3.75231	0.98420	0.01054

Figura 26: Gráfico cuantil-cuantil



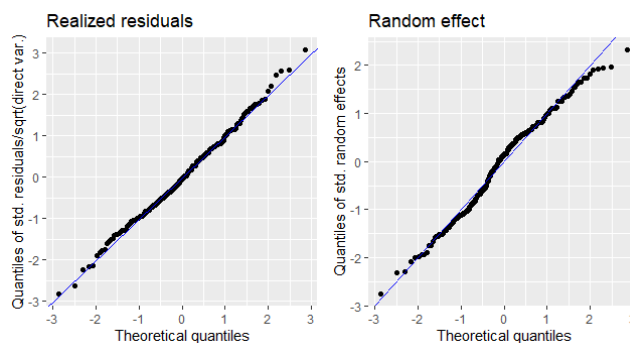
9.1.1.2 Modelo Fay-Herriot espacial: Resultados prueba de diagnóstico

Los resultados de la prueba de Shapiro-Wilk que se muestran en el cuadro 13 indican que la normalidad de los errores no se rechaza. El gráfico de residuos del modelo que se presenta en la figura 18 confirma lo mencionado anteriormente.

Cuadro 12: Pruebas de diagnóstico para residuos

VARIABLES	Asimetría	Curtosis	Shapiro-Wilk	p-value
Residuos estandarizados	0.16094	3.09928	0.99589	0.79360
Efectos aleatorios	-0.22611	2.40416	0.98447	0.01172

Figura 27: Gráfico cuantil-cuantil



9.1.2. Riesgo de retraso del crecimiento de niños menores a 5 años

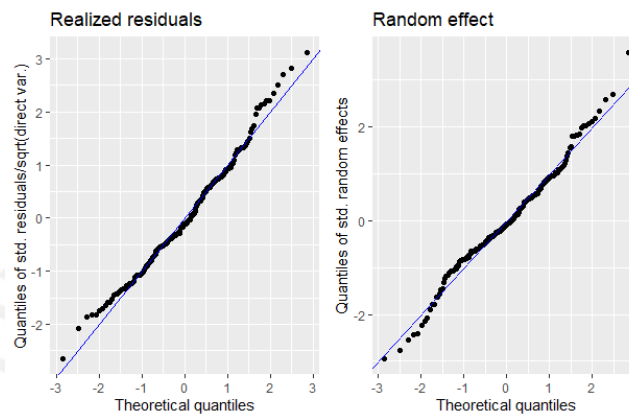
9.1.2.1 Modelo Fay-Herriot: Resultados prueba de diagnóstico

Los resultados de la prueba de Shapiro-Wilk que se muestran en el cuadro 12 indican que la normalidad de los errores no se rechaza. El gráfico de residuos del modelo presente en la figura 41 confirma lo mencionado anteriormente.

Cuadro 13: Pruebas de diagnóstico para residuos

VARIABLES	Asimetría	Curtosis	Shapiro-Wilk	p-value
Residuos estandarizados	0.417	3.168	0.986	0.02
Efectos aleatorios	0.149	3.947	0.984	0.01

Figura 28: Gráfico cuantil-cuantil



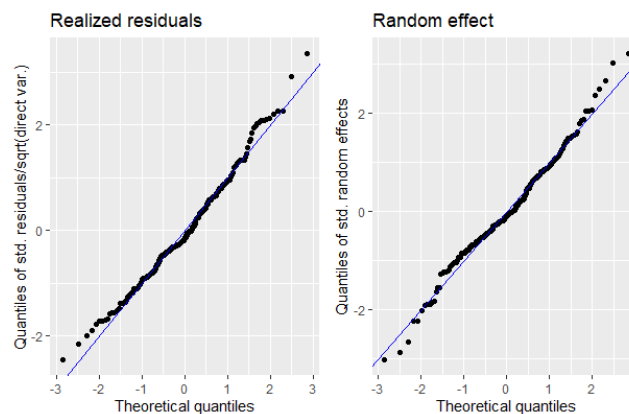
9.1.2.2 Modelo Fay-Herriot espacial: Resultados prueba de diagnóstico

Los resultados de la prueba de Shapiro-Wilk que se muestran en el cuadro 13 indican que la normalidad de los errores no se rechaza. El gráfico de residuos del modelo que se presenta en la figura 18 confirma lo mencionado anteriormente.

Cuadro 14: Pruebas de diagnóstico para residuos

VARIABLES	Asimetría	Curtosis	Shapiro-Wilk	p-value
Residuos estandarizados	0.411	3.163	0.987	0.03
Efectos aleatorios	0.158	3.771	0.987	0.03

Figura 29: Gráfico cuantil-cuantil



9.2. Anexo B

9.2.1. Prevalencia de Anemia:Estimaciones

Cuadro 15: Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
Amazonas										
1	010104	CHETO		16.56	21.40		0.0094			
2	010105	CHILQUIN		23.99	26.62		0.0086			
3	010106	CHUQUIBAMBA	17.65	23.57	23.04	0.0082	0.0043	0.0033	0.0034	17
4	010109	LA JALCA		22.55	25.34		0.0088			
5	010111	LEVANTO	40.00	27.13	21.51	0.0558	0.0077	0.0046	0.0045	5
6	010112	MAGDALENA	28.57	21.70	19.43	0.0310	0.0072	0.0046	0.0046	9
7	010113	MARISCAL CASTILLA		25.09	27.84		0.0092			
8	010117	QUINJALCA		40.47	35.37		0.0089			2
9	010119	SAN ISIDRO DE MAINO	14.29	18.56	17.64	0.0174	0.0062	0.0043	0.0043	8
10	010120	SOLOCO		24.09	25.89		0.0089			3
11	010121	SONCHE		24.19	27.02		0.0089			
12	010202	ARAMANGO	29.17	25.95	27.00	0.0088	0.0045	0.0031	0.0031	25
13	010205	IMAZA	52.94	49.43	47.82	0.0075	0.0042	0.0031	0.0031	41
14	010206	LA PECA	16.00	18.29	20.28	0.0053	0.0034	0.0026	0.0026	26
15	010302	CHISQUILLA		25.35	28.03		0.0092			
16	010304	COROSHA		23.17	26.01		0.0086			
17	010308	RECTA		25.84	28.21		0.0086			
18	010401	NIEVA	50.00	49.58	50.29	0.0067	0.0040	0.0030	0.0030	43
19	010402	EL CENEPA	50.00	51.91	52.69	0.0130	0.0056	0.0039	0.0039	24
20	010403	RIO SANTIAGO	54.29	46.65	45.27	0.0072	0.0041	0.0030	0.0030	40
21	010503	COCABAMBA		28.84	29.50		0.0085			
22	010504	COLCAMAR		24.97	26.19		0.0086			
23	010505	CONILA	18.18	21.90	20.11	0.0139	0.0055	0.0038	0.0039	11
24	010507	LONGUITA	40.00	27.46	22.13	0.0236	0.0065	0.0043	0.0043	10
25	010508	LONYA CHICO	42.86	28.50	24.50	0.0362	0.0070	0.0047	0.0047	9
26	010510	LUYA VIEJO		21.38	24.88		0.0089			
27	010511	MARIA		25.69	26.33		0.0089			
28	010514	PISUQUIA	25.00	27.00	23.65	0.0264	0.0065	0.0047	0.0047	10
29	010515	PROVIDENCIA		25.51	27.99		0.0087			
30	010516	SAN CRISTOBAL		18.55	23.38		0.0092			
31	010517	SAN FRANCISCO DEL YESO	42.86	27.19	22.65	0.0367	0.0072	0.0047	0.0047	8

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
32	010518	SAN JERONIMO	16.67	24.71	23.43	0.0239	0.0064	0.0045	0.0044	6
33	010520	SANTA CATALINA	15.00	18.81	19.22	0.0061	0.0037	0.0030	0.0031	20
34	010521	SANTO TOMAS	8.33	15.02	16.32	0.0066	0.0039	0.0032	0.0032	19
35	010609	OMIA	18.18	26.48	24.46	0.0147	0.0056	0.0038	0.0039	12
36	010612	VISTA ALEGRE		32.19	32.04		0.0088			
37	010702	CAJARURO	22.00	22.38	22.66	0.0034	0.0025	0.0020	0.0021	54
Ancash										
38	020102	COCHABAMBA		47.76	45.61		0.0089			
39	020103	COLCABAMBA		41.06	39.00		0.0085			
40	020104	HUANCHAY		37.45	35.12		0.0084			
41	020105	INDEPENDENCIA	27.27	30.16	28.84	0.0094	0.0046	0.0035	0.0036	26
42	020106	JANGAS		41.08	41.21		0.0089			
43	020107	LA LIBERTAD		44.72	43.90		0.0087			
44	020108	OLLEROS		41.39	41.76		0.0087			
45	020109	PAMPAS GRANDE		37.80	34.93		0.0085			
46	020110	PARIACOTO		39.77	39.24		0.0085			
47	020111	PIRA		49.48	48.20		0.0087			
48	020112	TARICA	31.25	37.15	36.19	0.0139	0.0054	0.0040	0.0039	18
49	020301	LLAMELLIN		48.34	45.57		0.0089			
50	020304	CHINGAS		45.63	43.51		0.0085			
51	020305	MIRGAS		50.70	49.40		0.0090			
52	020306	SAN JUAN DE RONTOY		44.86	44.48		0.0088			
53	020401	CHACAS		46.17	45.33		0.0088			
54	020402	ACOHACA		45.65	45.25		0.0089			4
55	020501	CHIQUIAN		32.39	35.27		0.0092			
56	020505	CAJACAY		28.59	32.44		0.0087			
57	020507	COLQUIOC		27.98	29.95		0.0090			
58	020508	HUALLANCA		36.61	39.01		0.0087			
59	020509	HUASTA		30.17	30.62		0.0094			
60	020510	HUAYLLACAYAN		43.54	40.63		0.0086			
61	020511	LA PRIMAVERA		48.19	43.37		0.0085			
62	020512	MANGAS		48.95	44.90		0.0086			
63	020513	PACLLON		31.95	32.98		0.0093			
64	020514	SAN MIGUEL DE CORPANQUI		42.28	39.17		0.0086			
65	020515	TICLOS		44.06	42.70		0.0084			
66	020601	CARHUAZ	20.00	29.64	30.14	0.0065	0.0038	0.0030	0.0031	27
67	020602	ACOPAMPA		34.52	36.25		0.0087			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
68	020603	AMASHCA	60.00	48.80	43.21	0.0239	0.0066	0.0046	0.0045	12
69	020604	ANTA		41.35	40.41		0.0087			
70	020606	MARCARA		43.18	41.54		0.0087			
71	020607	PARIAHUANCA		37.91	40.37		0.0089			
72	020608	SAN MIGUEL DE ACO	62.50	50.56	44.81	0.0323	0.0072	0.0048	0.0048	8
73	020703	YAUYA		51.97	49.17		0.0090			
74	020801	CASMA	33.33	33.50	32.29	0.0058	0.0035	0.0029	0.0029	40
75	020802	BUENA VISTA ALTA		38.50	35.57		0.0083			5
76	020803	COMANDANTE NOEL		32.20	29.04		0.0084			
77	020804	YAUTAN		39.10	36.53		0.0085			
78	020901	CORONGO		30.51	32.52		0.0090			
79	020902	ACO		40.22	37.41		0.0088			
80	020904	CUSCA		39.04	34.96		0.0090			
81	020905	LA PAMPA		36.62	35.45		0.0087			
82	020906	YANAC		24.68	26.51		0.0089			
83	020907	YUPAN		31.86	32.93		0.0099			
84	021003	CAJAY		45.77	45.59		0.0090			
85	021004	CHAVIN DE HUANTAR	38.89	41.94	39.98	0.0137	0.0055	0.0040	0.0040	18
86	021005	HUACACHI		46.77	46.37		0.0091			
87	021402	ACAS		46.38	41.24		0.0088			
88	021404	CARHUAPAMPA		35.99	33.67		0.0093			
89	021506	LACABAMBA		18.52	26.73		0.0095			
90	021509	PAMPAS		18.58	26.96		0.0093			
91	021510	SANTA ROSA		22.62	28.87		0.0095			
92	021511	TAUCA		18.90	26.98		0.0094			
93	021602	HUAYLLAN		44.12	48.20		0.0093			
94	021701	RECUAY		29.84	36.55		0.0091			
95	021703	COTAPARACO		25.63	32.83		0.0095			
96	021705	LLACLIN		38.23	42.68		0.0090			
97	021706	MARCA		55.56	53.29		0.0099			5
98	021709	TAPACOCHA		36.49	42.48		0.0093			
99	021801	CHIMBOTE	16.67	16.79	17.27	0.0009	0.0008	0.0008	0.0008	170
100	021804	MACATE		22.37	28.54		0.0090			
101	021808	SANTA	57.14	25.64	23.70	0.0407	0.0075	0.0042	0.0042	8
102	021901	SIHUAS		22.23	29.83		0.0096			
103	021903	ALFONSO UGARTE		24.62	31.40		0.0091			
104	021905	CHINGALPO	33.33	22.80	26.98	0.0175	0.0062	0.0043	0.0044	12

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
105	021908	RAGASH		18.73	26.02		0.0090			
106	022001	YUNGAY		40.13	42.61		0.0089			
107	022006	RANRAHIRCA		40.68	43.53		0.0089			
108	022007	SHUPLUY		39.52	45.04		0.0092			
109	022008	YANAMA		40.59	45.74		0.0091			
Apurimac										
110	030101	ABANCAY	23.77	24.12	24.86	0.0015	0.0013	0.0011	0.0012	129
111	030103	CIRCA	66.67	49.79	42.01	0.1081	0.0082	0.0050	0.0050	3
112	030104	CURAHUASI	38.89	41.25	36.37	0.0138	0.0054	0.0039	0.0039	19
113	030105	HUANIPACA		43.92	45.19		0.0088			5
114	030106	LAMBRAMA	20.00	41.05	40.30	0.0389	0.0073	0.0047	0.0047	6
115	030107	PICHIRHUA		46.65	46.03		0.0087			
116	030108	SAN PEDRO DE CACHORA		40.48	43.33		0.0087			4
117	030109	TAMBURCO		30.13	32.07		0.0086			
118	030201	ANDAHUAYLAS	34.62	34.17	31.45	0.0029	0.0022	0.0020	0.0020	87
119	030202	ANDARAPA	14.29	37.36	38.60	0.0201	0.0064	0.0045	0.0045	8
120	030204	HUANCARAMA	35.00	39.48	33.76	0.0113	0.0050	0.0035	0.0036	22
121	030206	HUAYANA		45.97	47.24		0.0088			
122	030207	KISHUARA	57.14	47.96	39.19	0.0404	0.0075	0.0045	0.0046	7
123	030208	PACOBAMBA		45.18	46.73		0.0089			
124	030209	PACUCHA	23.53	37.60	35.70	0.0109	0.0050	0.0037	0.0039	19
125	030210	PAMPACHIRI	50.00	48.23	47.78	0.0486	0.0076	0.0057	0.0057	7
126	030213	SAN JERONIMO	36.59	37.81	33.13	0.0057	0.0035	0.0027	0.0027	43
127	030215	SANTA MARIA DE CHICMO	25.64	31.01	32.41	0.0048	0.0032	0.0025	0.0027	40
128	030217	TUMAY HUARACA		43.22	54.66		0.0092			
129	030218	TURPO	33.33	39.64	37.31	0.0150	0.0059	0.0039	0.0039	15
130	030219	KAQUIABAMBA		48.09	48.48		0.0090			
131	030220	JOSE MARIA ARGUEDAS		48.45	48.01		0.0089			
132	030302	EL ORO	25.00	36.80	34.39	0.0220	0.0064	0.0041	0.0042	9
133	030402	CAPAYA		37.60	42.96		0.0091			
134	030406	COTARUSE		40.03	52.59		0.0093			3
135	030408	JUSTO APU SAHUARAURA		41.01	46.31		0.0091			
136	030410	POCOHUANCA		39.13	47.07		0.0092			
137	030606	ONGOY		44.34	47.52		0.0093			
138	030611	LOS CHANKAS		42.92	46.58		0.0091			
Arequipa										
139	040106	CHIGUATA	18.18	28.10	24.36	0.0142	0.0058	0.0039	0.0040	12

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
140	040107	JACOBO HUNTER	25.00	25.31	23.01	0.0048	0.0032	0.0025	0.0025	47
141	040110	MIRAFLORES	14.29	19.46	19.26	0.0061	0.0038	0.0029	0.0029	25
142	040111	MOLLEBAYA		35.85	30.87		0.0115			
143	040114	POLOBAYA		38.31	43.99		0.0098			
144	040115	QUEQUEÑA	37.50	36.19	27.97	0.0329	0.0088	0.0064	0.0064	8
145	040402	ANDAGUA		32.48	39.31		0.0093			
146	040403	AYO		36.09	38.47		0.0101			
147	040405	CHILCAYMARCA		47.63	57.06		0.0118			
148	040410	PAMPACOLCA	66.67	30.20	27.87	0.1088	0.0085	0.0048	0.0047	3
149	040607	SALAMANCA		34.62	35.96		0.0086			
150	040608	YANAQUIHUA		52.93	44.31		0.0094			
151	040702	COCACHACRA	16.67	29.72	31.99	0.0124	0.0052	0.0036	0.0036	16
152	040704	ISLAY		36.09	32.29		0.0089			
153	040705	MEJIA		37.93	30.76		0.0096			
Ayacucho										
154	050110	SAN JUAN BAUTISTA	39.39	35.15	33.73	0.0074	0.0041	0.0032	0.0032	37
155	050401	HUANTA	17.95	22.82	26.58	0.0038	0.0027	0.0023	0.0024	40
156	050408	LLOCHEGUA		37.46	36.98		0.0086			7
157	050502	ANCO		42.61	41.16		0.0090			
158	050611	LUCANAS		44.26	45.21		0.0086			
159	050618	SAN PEDRO DE PALCO		35.87	39.00		0.0084			
160	050620	SANTA ANA DE HUAYCAHUACHO		54.69	52.57		0.0090			
161	050621	SANTA LUCIA		45.91	40.73		0.0088			
162	050702	CHUMPI	50.00	43.79	42.83	0.0262	0.0066	0.0044	0.0044	10
163	050703	CORONEL CASTAÑEDA		51.02	54.51		0.0117			
164	050706	PUYUSCA	60.00	46.70	52.13	0.0582	0.0075	0.0050	0.0050	6
165	050707	SAN FRANCISCO DE RAVACAYCO	25.00	46.35	52.42	0.0568	0.0077	0.0049	0.0049	6
166	050901	QUEROBAMBA	16.67	37.80	40.04	0.0269	0.0071	0.0047	0.0047	7
167	051001	HUANCAPI		44.16	45.33		0.0096			
168	051007	COLCA		45.90	44.47		0.0088			
169	051012	VILCANCHOS		48.17	54.08		0.0090			7
170	051104	CONCEPCION		52.69	49.18		0.0090			
Cajamarca										
171	060102	ASUNCION		27.78	28.82		0.0085			3
172	060103	CHETILLA		30.39	30.89		0.0085			
173	060104	COSPAN		30.57	30.40		0.0085			
174	060105	ENCAÑADA	20.00	26.48	22.81	0.0177	0.0059	0.0043	0.0043	10

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
175	060106	JESUS		27.27	27.68		0.0085			7
176	060107	LLACANORA		31.98	30.36		0.0085			
177	060109	MAGDALENA	40.00	27.65	19.71	0.0263	0.0066	0.0043	0.0042	10
178	060110	MATARA		30.20	29.24		0.0084			
179	060111	NAMORA		27.42	27.55		0.0085			
180	060112	SAN JUAN	28.57	27.92	22.57	0.0334	0.0069	0.0043	0.0043	7
181	060201	CAJABAMBA	35.90	32.80	29.94	0.0060	0.0036	0.0029	0.0031	39
182	060202	CACHACHI	18.42	24.24	25.04	0.0040	0.0028	0.0024	0.0026	42
183	060203	CONDEBAMBA	40.00	30.97	27.98	0.0598	0.0076	0.0050	0.0050	6
184	060204	SITACOCHA	27.27	25.92	25.50	0.0196	0.0061	0.0044	0.0045	12
185	060301	CELENDIN		29.20	28.71		0.0085			
186	060302	CHUMUCH		26.58	28.23		0.0085			
187	060303	CORTEGANA	20.00	23.90	22.40	0.0397	0.0072	0.0047	0.0047	6
188	060304	HUASMIN	25.00	26.78	22.22	0.0266	0.0066	0.0044	0.0045	11
189	060305	JORGE CHAVEZ		29.19	29.81		0.0085			
190	060307	MIGUEL IGLESIAS		26.72	27.36		0.0086			
191	060308	OXAMARCA	22.22	26.16	23.70	0.0213	0.0063	0.0045	0.0045	9
192	060309	SOROCHUCO	16.67	26.89	22.13	0.0275	0.0066	0.0044	0.0044	6
193	060310	SUCRE		28.47	28.70		0.0085			
194	060311	UTCO		29.20	29.48		0.0086			
195	060312	LA LIBERTAD DE PALLAN		26.63	27.16		0.0086			
196	060402	ANGUIA		27.45	28.20		0.0086			
197	060403	CHADIN		25.60	26.44		0.0085			
198	060404	CHIGUIRIP		26.87	27.70		0.0086			
199	060405	CHIMBAN	11.11	17.99	17.59	0.0118	0.0051	0.0037	0.0038	11
200	060406	CHOROPAMPA		29.49	29.38		0.0085			
201	060407	COCHABAMBA		23.37	26.70		0.0086			
202	060408	CONCHAN		26.16	27.19		0.0085			
203	060409	HUAMBOS		26.69	28.10		0.0085			
204	060410	LAJAS	36.36	29.35	24.17	0.0107	0.0049	0.0035	0.0035	26
205	060411	LLAMA		29.93	30.41		0.0085			
206	060412	MIRACOSTA		31.62	30.79		0.0085			
207	060413	PACCHA		24.86	26.69		0.0085			
208	060414	PION	12.50	17.30	16.53	0.0146	0.0056	0.0038	0.0038	9
209	060416	SAN JUAN DE LICUPIS		30.04	29.33		0.0085			
210	060417	TACABAMBA		26.03	27.74		0.0084			
211	060418	TOCMOCHE		25.48	29.26		0.0087			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
212	060419	CHALAMARCA	66.67	34.52	25.86	0.0441	0.0072	0.0046	0.0047	6
213	060501	CONTUMAZA		29.24	30.33		0.0087			
214	060502	CHILETE		20.99	26.96		0.0090			
215	060503	CUPISNIQUE		28.66	30.07		0.0086			
216	060504	GUZMANGO		26.71	27.84		0.0085			
217	060505	SAN BENITO		27.63	29.61		0.0087			
218	060506	SANTA CRUZ DE TOLED		27.39	28.37		0.0085			
219	060507	TANTARICA		28.09	28.83		0.0086			
220	060601	CUTERVO	30.00	24.26	19.55	0.0233	0.0063	0.0043	0.0044	10
221	060602	CALLAYUC	26.67	25.42	22.73	0.0137	0.0054	0.0038	0.0041	18
222	060603	CHOROS		20.44	27.06		0.0089			
223	060604	CUJILLO		21.56	26.34		0.0087			
224	060605	LA RAMADA		23.24	26.97		0.0086			
225	060606	PIMPINGOS		23.77	27.29		0.0086			
226	060607	QUEROCOTILLO		25.81	28.06		0.0085			
227	060608	SAN ANDRES DE CUTERVO	12.50	17.75	15.72	0.0152	0.0056	0.0040	0.0041	8
228	060609	SAN JUAN DE CUTERVO		22.50	26.05		0.0086			
229	060610	SAN LUIS DE LUCMA	12.50	19.78	18.56	0.0151	0.0056	0.0039	0.0041	10
230	060611	SANTA CRUZ		25.22	28.35		0.0087			
231	060612	SANTO DOMINGO DE LA CAPILLA		23.25	26.40		0.0086			6
232	060613	SANTO TOMAS	9.09	13.57	13.82	0.0038	0.0027	0.0023	0.0023	25
233	060614	SOCOTA		21.74	25.32		0.0085			
234	060615	TORIBIO CASANOVA		22.01	27.20		0.0088			
235	060701	BAMBAMARCA	18.18	21.76	20.89	0.0046	0.0030	0.0026	0.0027	35
236	060702	CHUGUR		39.32	34.38		0.0089			
237	060703	HUALGAYOC	16.67	30.65	25.32	0.0125	0.0054	0.0037	0.0038	12
238	060804	COLASAY		26.50	28.39		0.0085			
239	060805	HUABAL		24.00	27.15		0.0086			
240	060806	LAS PIRIAS		24.41	27.34		0.0085			
241	060807	POMAHUACA		25.84	29.50		0.0086			
242	060809	SALLIQUE	25.00	24.42	19.65	0.0265	0.0066	0.0047	0.0047	10
243	060810	SAN FELIPE		28.34	29.60		0.0085			
244	060901	SAN IGNACIO	11.11	17.97	18.88	0.0123	0.0052	0.0037	0.0039	11
245	060902	CHIRINOS		27.83	28.71		0.0086			
246	060903	HUARANGO		21.60	26.88		0.0087			
247	060904	LA COIPA		28.40	29.96		0.0086			
248	060905	NAMBALLE		20.53	25.98		0.0087			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
249	060906	SAN JOSE DE LOURDES	36.84	28.41	26.25	0.0128	0.0054	0.0038	0.0040	20
250	060907	TABACONAS	16.00	20.04	18.72	0.0055	0.0034	0.0029	0.0030	27
251	061002	CHANCAY		26.14	27.23		0.0086			
252	061004	GREGORIO PITA		28.61	28.93		0.0086			
253	061005	ICHOCAN		29.29	30.17		0.0086			
254	061006	JOSE MANUEL QUIROZ	40.00	31.81	28.98	0.0258	0.0065	0.0044	0.0045	11
255	061007	JOSE SABOGAL	57.14	32.93	28.35	0.0406	0.0072	0.0048	0.0048	7
256	061101	SAN MIGUEL	5.88	12.61	11.74	0.0034	0.0025	0.0022	0.0023	19
257	061102	BOLIVAR		22.06	27.56		0.0087			
258	061103	CALQUIS	33.33	30.34	18.09	0.0437	0.0073	0.0047	0.0048	6
259	061104	CATILLUC		30.76	29.32		0.0084			
260	061105	EL PRADO		29.79	29.52		0.0085			
261	061106	LA FLORIDA		26.78	30.18		0.0088			
262	061107	LLAPA		29.05	28.91		0.0085			
263	061108	NANCHOC		23.17	28.76		0.0089			
264	061109	NIEPOS	33.33	28.92	19.45	0.0434	0.0072	0.0045	0.0045	7
265	061110	SAN GREGORIO		29.09	29.74		0.0087			
266	061111	SAN SILVESTRE DE COCHAN		30.22	29.12		0.0086			
267	061112	TONGOD		27.83	27.90		0.0084			
268	061113	UNION AGUA BLANCA	8.33	17.90	14.76	0.0066	0.0038	0.0030	0.0030	12
269	061201	SAN PABLO	11.11	20.75	16.26	0.0122	0.0051	0.0038	0.0038	9
270	061202	SAN BERNARDINO		25.12	28.38		0.0087			
271	061203	SAN LUIS		22.97	27.26		0.0086			
272	061204	TUMBADEN		29.43	28.89		0.0086			
273	061302	ANDABAMBA		28.59	29.30		0.0085			
274	061303	CATACHE		30.62	32.51		0.0089			
275	061304	CHANCAYBAÑOS		24.59	27.74		0.0086			
276	061305	LA ESPERANZA		26.08	28.69		0.0086			
277	061306	NINABAMBA		29.43	30.39		0.0086			
278	061307	PULAN		28.19	29.40		0.0085			
279	061308	SAUCEPAMPA		24.84	27.40		0.0087			
280	061309	SEXI		35.81	35.85		0.0098			
281	061310	UTICYACU		29.26	30.25		0.0087			
282	061311	YAUYUCAN		26.66	27.83		0.0084			
Callao										
283	070106	VENTANILLA	25.89	25.91	26.25	0.0006	0.0005	0.0005	0.0005	355

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
Cusco										
284	080101	CUSCO	42.50	42.25	42.28	0.0062	0.0038	0.0028	0.0029	43
285	080102	CCORCA		50.13	48.18		0.0091			
286	080103	POROY	36.36	40.76	43.30	0.0221	0.0064	0.0039	0.0039	12
287	080104	SAN JERONIMO	37.50	38.72	39.24	0.0075	0.0041	0.0029	0.0030	34
288	080105	SAN SEBASTIAN	40.00	39.87	39.82	0.0044	0.0030	0.0024	0.0024	57
289	080106	SANTIAGO	56.86	49.95	46.49	0.0049	0.0032	0.0024	0.0025	54
290	080108	WANCHAQ	32.00	33.21	33.84	0.0090	0.0047	0.0033	0.0034	28
291	080404	LARES	71.43	50.20	52.42	0.0336	0.0072	0.0048	0.0048	8
292	080406	SAN SALVADOR		46.29	48.29		0.0086			
293	080601	SICUANI	59.26	52.20	54.22	0.0092	0.0046	0.0037	0.0037	28
294	080602	CHECACUPE		47.27	53.71		0.0088			
295	080604	MARANGANI		59.90	64.44		0.0089			
296	080605	PITUMARCA		48.81	54.43		0.0088			
297	080608	TINTA	80.00	55.82	63.96	0.0395	0.0072	0.0053	0.0053	5
298	080701	SANTO TOMAS		50.63	54.56		0.0087			
299	080703	CHAMACA		54.62	66.19		0.0088			
300	080705	LIVITACA		55.99	66.99		0.0090			
301	080708	VELILLE	71.43	58.45	65.35	0.0336	0.0073	0.0060	0.0060	7
302	080801	ESPINAR	53.57	50.61	51.38	0.0091	0.0046	0.0036	0.0036	30
303	080802	CONDOROMA		55.11	61.26		0.0088			
304	080805	PALLPATA		53.43	60.08		0.0087			
305	080807	SUYCKUTAMBO		57.83	66.33		0.0089			
306	080808	ALTO PICHIGUA		60.81	71.72		0.0090			
307	080901	SANTA ANA	27.78	39.76	39.46	0.0116	0.0052	0.0040	0.0042	21
308	080907	KIMBIRI	23.08	27.51	31.60	0.0147	0.0058	0.0044	0.0048	14
309	080908	SANTA TERESA	30.00	44.04	44.43	0.0228	0.0064	0.0043	0.0044	11
310	080909	VILCABAMBA	50.00	45.72	45.97	0.0828	0.0078	0.0052	0.0053	4
311	080911	INKAWASI		55.14	52.10		0.0090			
312	080913	VILLA KINTIARINA		44.20	41.76		0.0086			
313	080914	MEGANTONI		49.11	43.79		0.0099			
314	081001	PARURO		51.69	49.64		0.0088			
315	081007	PACCARITAMBO		50.99	49.05		0.0090			
316	081102	CAICAY	75.00	54.15	56.70	0.0163	0.0059	0.0043	0.0045	12
317	081201	URCOS		42.82	42.43		0.0086			
318	081203	CAMANTI	22.22	33.89	33.62	0.0208	0.0065	0.0043	0.0044	10
319	081204	CCARHUAYO		50.45	53.71		0.0088			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
320	081205	CCATCA	72.73	57.14	58.46	0.0197	0.0063	0.0048	0.0049	12
321	081206	CUSIPATA	72.73	53.00	58.05	0.0192	0.0062	0.0042	0.0043	11
322	081207	HUARO		47.07	47.28		0.0088			
323	081208	LUCRE		37.75	40.04		0.0088			
324	081209	MARCAPATA		49.61	53.42		0.0087			
325	081210	OCONGATE	50.00	49.87	57.88	0.0276	0.0068	0.0055	0.0055	11
326	081211	OROPESA	47.83	43.53	41.36	0.0111	0.0050	0.0038	0.0039	26
327	081212	QUIQUIJANA		48.54	51.05		0.0088			
328	081301	URUBAMBA		49.18	44.24		0.0090			
329	081307	YUCAY		46.23	41.97		0.0093			
Huancavelica										
330	090107	HUAYLLAHUARA		51.57	61.13		0.0088			
331	090114	PALCA	60.00	45.91	59.23	0.0574	0.0076	0.0063	0.0062	6
332	090201	ACOBAMBA	54.55	47.81	47.18	0.0245	0.0065	0.0048	0.0049	13
333	090202	ANDABAMBA	64.71	55.50	52.33	0.0135	0.0056	0.0045	0.0045	19
334	090203	ANTA		51.78	55.95		0.0094			
335	090204	CAJA	25.00	46.75	49.02	0.0605	0.0079	0.0055	0.0055	4
336	090205	MARCAS		45.83	44.79		0.0087			
337	090401	CASTROVIRREYNA		35.66	40.00		0.0087			
338	090406	COCAS		38.67	50.69		0.0086			
339	090407	HUACHOS		36.34	40.49		0.0086			
340	090411	SANTA ANA		49.27	57.52		0.0092			1
341	090505	LA MERCED		54.93	53.20		0.0090			
342	090611	SAN FRANCISCO DE SANGAYAICO	57.14	44.39	45.78	0.0367	0.0071	0.0051	0.0051	7
343	090701	PAMPAS	36.84	37.01	39.33	0.0127	0.0054	0.0045	0.0044	21
344	090702	ACOSTAMBO		41.59	41.16		0.0087			
345	090711	PAZOS	66.67	41.18	43.11	0.1104	0.0081	0.0056	0.0055	6
Huánuco										
346	100101	HUANUCO	12.09	14.51	14.15	0.0012	0.0010	0.0010	0.0010	107
347	100111	PILLCO MARCA	9.09	13.75	13.02	0.0026	0.0020	0.0017	0.0018	35
348	100203	COLPAS		49.78	48.01		0.0092			
349	100609	PUEBLO NUEVO		41.52	33.44		0.0089			
350	100705	SANTA ROSA DE ALTO YANAJANCA		42.93	35.17		0.0089			
351	101101	CHAVINILLO	80.00	47.40	36.82	0.0395	0.0072	0.0050	0.0051	6
352	101105	JACAS CHICO		44.41	45.76		0.0086			
353	101106	OBAS	33.33	43.36	36.07	0.1099	0.0081	0.0055	0.0054	4
354	101107	PAMPAMARCA		47.74	49.04		0.0089			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
Junín										
355	120101	HUANCAYO	36.36	32.37	35.57	0.0043	0.0030	0.0023	0.0025	62
356	120104	CARHUACALLANGA		69.62	61.48		0.0105			
357	120105	CHACAPAMPA		46.97	52.78		0.0086			
358	120106	CHICCHE		43.14	52.19		0.0091			
359	120107	CHILCA	47.14	42.78	40.92	0.0036	0.0026	0.0022	0.0022	78
360	120108	CHONGOS ALTO		48.53	50.20		0.0088			
361	120112	COLCA		43.55	45.04		0.0084			
362	120114	EL TAMBO	42.39	39.04	38.18	0.0027	0.0021	0.0018	0.0021	100
363	120116	HUACRAPUQUIO		37.00	35.73		0.0085			
364	120117	HUALHUAS		35.24	29.48		0.0085			
365	120119	HUANCAN		29.82	30.07		0.0090			
366	120120	HUASICANCHA		50.12	61.23		0.0086			
367	120121	HUAYUCACHI	85.71	48.07	47.65	0.0202	0.0063	0.0040	0.0041	7
368	120122	INGENIO		40.11	36.45		0.0089			
369	120124	PARIAHUANCA		40.47	41.48		0.0088			3
370	120125	PILCOMAYO	52.00	39.91	38.06	0.0103	0.0048	0.0036	0.0036	27
371	120126	PUCARA		41.15	36.82		0.0084			
372	120127	QUICHUAY		36.33	33.95		0.0086			
373	120128	QUILCAS	37.50	37.96	42.96	0.0327	0.0069	0.0051	0.0050	8
374	120130	SAN JERONIMO DE TUNAN		30.37	27.69		0.0084			
375	120132	SAÑO	28.57	34.09	40.86	0.0335	0.0069	0.0043	0.0044	9
376	120134	SICAYA	50.00	32.83	32.88	0.0498	0.0075	0.0047	0.0046	6
377	120201	CONCEPCION	28.57	31.12	34.54	0.0338	0.0069	0.0042	0.0044	7
378	120202	ACO		27.86	28.36		0.0088			
379	120204	CHAMBARA		38.46	38.05		0.0086			
380	120205	COCHAS		36.67	34.63		0.0091			
381	120206	COMAS		39.57	37.43		0.0085			
382	120208	MANZANARES		34.22	31.84		0.0086			
383	120209	MARISCAL CASTILLA		34.62	31.17		0.0088			
384	120210	MATAHUASI	42.86	42.57	42.25	0.0402	0.0074	0.0050	0.0050	8
385	120211	MITO		34.84	31.84		0.0084			
386	120212	NUEVE DE JULIO		40.59	33.93		0.0087			
387	120213	ORCOTUNA		29.81	27.67		0.0086			
388	120301	CHANCHAMAYO	13.64	17.78	12.98	0.0056	0.0035	0.0028	0.0028	23
389	120303	PICHANAQUI	35.71	37.48	25.51	0.0176	0.0060	0.0041	0.0041	14
390	120305	SAN RAMON		25.21	22.08		0.0087			6

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
391	120306	VITOC		34.80	31.41		0.0086			
392	120402	ACOLLA		41.61	43.06		0.0089			
393	120403	APATA	70.00	51.42	47.62	0.0227	0.0066	0.0045	0.0045	12
394	120408	HUAMALI		41.41	36.94		0.0089			
395	120409	HUARIPAMPA		33.05	33.72		0.0091			
396	120412	JULCAN		36.18	34.24		0.0086			
397	120414	LLOCLLAPAMPA	70.00	50.90	54.77	0.0206	0.0073	0.0057	0.0057	11
398	120415	MARCO		39.00	43.55		0.0088			
399	120417	MASMA CHICCHE		35.27	41.66		0.0087			
400	120419	MONOBAMBA		35.22	30.57		0.0086			
401	120420	MUQUI		34.87	32.52		0.0084			
402	120422	PACA		40.71	41.74		0.0087			
403	120423	PACCHA		34.67	42.58		0.0085			
404	120424	PANCAN		39.64	34.77		0.0088			
405	120425	PARCO		35.41	38.03		0.0085			
406	120426	POMACANCHA		34.27	46.24		0.0085			
407	120427	RICRAN		38.57	43.68		0.0085			
408	120429	SAN PEDRO DE CHUNAN		49.88	44.88		0.0090			
409	120430	SAUSA	20.00	27.33	19.36	0.0393	0.0071	0.0045	0.0045	6
410	120433	YAULI		41.96	39.41		0.0088			
411	120501	JUNIN	25.00	31.41	30.84	0.0265	0.0064	0.0049	0.0050	9
412	120503	ONDORES		44.63	55.13		0.0087			
413	120504	ULCUMAYO		44.00	43.01		0.0085			
414	120601	SATIPO	30.00	31.11	32.54	0.0053	0.0034	0.0027	0.0028	45
415	120602	COVIRIALI	33.33	37.51	37.83	0.0440	0.0074	0.0050	0.0049	6
416	120603	LLAYLLA		44.39	38.40		0.0086			
417	120605	PAMPA HERMOSA		38.38	34.29		0.0085			
418	120608	RIO TAMBO	60.87	57.68	56.53	0.0108	0.0050	0.0038	0.0039	26
419	120609	VIZCATAN DEL ENE	46.94	52.71	53.05	0.0047	0.0033	0.0028	0.0028	52
420	120701	TARMA	50.00	42.52	39.60	0.0055	0.0034	0.0029	0.0030	50
421	120702	ACOBAMBA	33.33	38.85	39.62	0.0441	0.0076	0.0050	0.0052	7
422	120703	HUARICOLCA	63.64	41.16	44.71	0.0216	0.0062	0.0050	0.0051	15
423	120705	LA UNION		50.98	43.41		0.0096			
424	120706	PALCA		32.35	30.00		0.0089			6
425	120707	PALCAMAYO	28.57	32.45	36.44	0.0331	0.0069	0.0048	0.0050	8
426	120708	SAN PEDRO DE CAJAS		40.53	43.08		0.0084			
427	120709	TAPO	14.29	30.24	35.94	0.0091	0.0045	0.0037	0.0038	15

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
428	120801	LA OROYA	50.00	36.39	35.35	0.0830	0.0081	0.0052	0.0053	4
429	120802	CHACAPALPA	50.00	51.56	62.08	0.0782	0.0090	0.0082	0.0082	4
430	120803	HUAY-HUAY		39.35	42.23		0.0116			
431	120804	MARCAPOMACOCHA		47.39	43.89		0.0098			
432	120805	MOROCOCHA		31.60	29.10		0.0109			
433	120806	PACCHA		36.97	34.51		0.0087			
434	120807	SANTA BARBARA DE CARHUACAYAN		48.59	52.31		0.0099			
435	120808	SANTA ROSA DE SACCO	33.33	33.47	35.08	0.0198	0.0061	0.0047	0.0047	13
436	120809	SUITUCANCHA		41.97	49.56		0.0087			
437	120810	YAULI	55.56	40.52	42.69	0.0300	0.0087	0.0057	0.0056	12
438	120901	CHUPACA	42.86	38.90	34.92	0.0406	0.0074	0.0046	0.0047	7
439	120902	AHUAC	50.00	43.07	40.29	0.0494	0.0074	0.0049	0.0052	8
440	120903	CHONGOS BAJO		37.52	34.11		0.0084			
441	120904	HUACHAC		40.79	35.15		0.0088			
442	120906	SAN JUAN DE ISCOS		47.80	42.57		0.0091			
443	120907	SAN JUAN DE JARPA	44.44	49.08	65.03	0.0133	0.0057	0.0063	0.0066	21
La Libertad										
444	130402	PACANGA	50.00	35.33	33.04	0.0498	0.0075	0.0054	0.0053	6
445	130403	PUEBLO NUEVO	37.50	30.62	31.92	0.0332	0.0069	0.0052	0.0051	10
Lima										
446	150102	ANCON	31.82	33.88	34.89	0.0103	0.0048	0.0039	0.0042	23
447	150107	CHACLACAYO		25.26	22.39		0.0088			4
448	150110	COMAS	24.84	24.86	24.04	0.0012	0.0010	0.0010	0.0010	170
449	150111	EL AGUSTINO	18.42	20.56	19.21	0.0041	0.0029	0.0023	0.0023	47
450	150136	SAN MIGUEL	13.64	19.29	19.85	0.0056	0.0037	0.0028	0.0028	22
451	150137	SANTA ANITA	27.27	27.64	23.95	0.0062	0.0037	0.0027	0.0028	35
452	150205	SUPE PUERTO	15.38	20.38	22.37	0.0107	0.0049	0.0033	0.0033	15
453	150303	GORGOR		34.90	38.05		0.0085			
454	150304	HUANCAPON	83.33	50.84	47.67	0.0250	0.0064	0.0046	0.0046	7
455	150401	CANTA		31.16	29.79		0.0088			
456	150707	HUACHUPAMPA		42.55	40.70		0.0093			
457	150708	HUANZA		43.16	45.05		0.0098			
458	150709	HUAROCHIRI		28.70	30.14		0.0091			
459	150710	LAHUAYTAMBO		38.75	37.71		0.0092			
460	150711	LANGA	33.33	35.98	34.50	0.0403	0.0076	0.0053	0.0053	6
461	150713	MARIATANA		28.78	26.34		0.0086			
462	150716	SAN ANTONIO		55.53	45.20		0.0099			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
463	150717	SAN BARTOLOME		31.09	28.31		0.0084			3
464	150721	SAN LORENZO DE QUINTI		35.76	33.35		0.0091			
465	150724	SAN PEDRO DE CASTA		30.57	30.29		0.0089			
466	150801	HUACHO	20.45	23.61	24.97	0.0037	0.0027	0.0022	0.0023	49
467	150805	HUALMAY	38.89	31.66	29.02	0.0139	0.0054	0.0037	0.0037	19
468	150807	LEONCIO PRADO	11.11	26.51	33.75	0.0116	0.0054	0.0036	0.0039	10
469	150810	SANTA MARIA	33.33	34.25	33.53	0.0076	0.0041	0.0030	0.0030	34
470	150812	VEGUETA	39.29	39.61	38.22	0.0087	0.0045	0.0033	0.0034	29
471	150904	COCHAMARCA		31.24	29.06		0.0087			
472	150905	NAVAN		44.96	41.58		0.0107			2
473	151003	ALLAUCA		42.72	36.10		0.0090			
474	151004	AYAVIRI		28.93	30.11		0.0092			
475	151012	HONGOS		30.13	33.52		0.0087			
476	151014	HUANCAYA		51.05	49.45		0.0093			
477	151019	LINCHA		37.81	34.57		0.0087			
478	151020	MADEAN		30.98	32.33		0.0084			
479	151021	MIRAFLORES		26.01	30.08		0.0094			
480	151023	PUTINZA		28.33	24.97		0.0084			
481	151024	QUINCHES		30.87	30.32		0.0087			
482	151026	SAN JOAQUIN		29.68	29.64		0.0092			
483	151027	SAN PEDRO DE PILAS		24.15	26.57		0.0091			
484	151028	TANTA		36.70	46.38		0.0101			
485	151033	VITIS		30.59	32.69		0.0094			
Loreto										
486	160507	SAQUENA	37.50	32.15	33.33	0.0328	0.0068	0.0045	0.0045	8
487	160510	JENARO HERRERA		31.40	30.06		0.0084			
Pasco										
488	190103	HUARIACA	28.57	28.93	33.08	0.0154	0.0057	0.0040	0.0040	14
489	190104	HUAYLLAY	70.00	48.06	44.68	0.0230	0.0076	0.0045	0.0047	10
490	190106	PALLANCHACRA	25.00	30.32	36.66	0.0255	0.0064	0.0044	0.0044	8
491	190107	PAUCARTAMBO	22.45	24.59	27.58	0.0034	0.0025	0.0022	0.0024	52
492	190109	SIMON BOLIVAR	55.81	47.49	44.04	0.0056	0.0036	0.0029	0.0030	50
493	190110	TICLACAYAN	42.86	37.30	43.28	0.0399	0.0069	0.0046	0.0048	8
494	190303	HUANCABAMBA	21.43	27.30	29.74	0.0127	0.0052	0.0037	0.0037	14
Piura										
495	200105	CATACAOS	36.00	32.55	33.62	0.0096	0.0046	0.0035	0.0036	27
496	200108	EL TALLAN		28.56	27.24		0.0085			6

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
497	200110	LA UNION	21.43	28.16	29.44	0.0129	0.0053	0.0039	0.0040	14
498	200210	SUYO		32.29	30.17		0.0085			
499	200304	HUARMACA	57.14	42.98	42.10	0.0090	0.0045	0.0035	0.0035	33
500	200305	LALAQUIZ	33.33	29.88	35.98	0.0436	0.0073	0.0048	0.0047	6
501	200307	SONDOR	33.33	31.06	36.82	0.0274	0.0066	0.0046	0.0046	10
502	200308	SONDORILLO		32.68	30.10		0.0087			
503	200401	CHULUCANAS	27.45	27.41	30.36	0.0040	0.0027	0.0024	0.0024	55
504	200402	BUENOS AIRES		28.11	25.73		0.0085			
505	200403	CHALACO		31.95	29.75		0.0085			
506	200406	SALITRAL		21.40	26.43		0.0088			
507	200407	SAN JUAN DE BIGOTE		24.25	27.20		0.0085			
508	200409	SANTO DOMINGO		27.73	31.61		0.0090			
509	200410	YAMANGO		26.39	29.68		0.0087			
510	200505	LA HUACA		30.94	27.51		0.0087			
511	200507	VICHAYAL	14.29	22.40	26.46	0.0202	0.0062	0.0043	0.0043	8
Puno										
512	210103	AMANTANI	75.00	64.59	58.94	0.0615	0.0081	0.0073	0.0072	4
513	210105	CAPACHICA		66.10	64.16		0.0094			
514	210107	COATA		66.56	74.22		0.0095			
515	210108	HUATA	85.71	71.03	75.55	0.0199	0.0065	0.0057	0.0057	8
516	210109	MAÑAZO	85.71	63.94	67.06	0.0201	0.0062	0.0052	0.0052	7
517	210110	PAUCARCOLLA		56.94	65.98		0.0089			
518	210111	PICHACANI		59.37	66.41		0.0090			
519	210113	SAN ANTONIO		58.14	60.42		0.0092			
520	210114	TIQUILLACA		62.45	70.45		0.0091			
521	210202	ACHAYA		63.96	73.42		0.0093			
522	210205	CAMINACA		58.23	69.95		0.0092			
523	210207	JOSE DOMINGO CHOQUEHUANCA	80.00	59.77	64.70	0.0394	0.0073	0.0057	0.0057	5
524	210208	MUÑANI		58.83	69.64		0.0091			
525	210209	POTONI	25.00	51.75	62.13	0.0618	0.0080	0.0072	0.0072	4
526	210210	SAMAN		59.46	69.94		0.0092			
527	210212	SAN JOSE		57.12	68.42		0.0090			
528	210213	SAN JUAN DE SALINAS		60.00	71.48		0.0092			
529	210215	TIRAPATA		57.51	67.78		0.0089			
530	210303	AYAPATA		59.41	55.16		0.0090			
531	210304	COASA	25.00	51.96	54.89	0.0621	0.0079	0.0057	0.0057	4
532	210305	CORANI		55.33	66.59		0.0091			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
533	210307	ITUATA	77.78	73.11	65.76	0.0212	0.0069	0.0053	0.0052	9
534	210308	OLLACHEA		55.04	55.75		0.0087			
535	210403	HUACULLANI	60.00	64.37	72.94	0.0594	0.0081	0.0067	0.0067	5
536	210407	ZEPITA		64.03	70.40		0.0093			5
537	210502	CAPAZO	66.67	59.76	60.41	0.1052	0.0084	0.0057	0.0057	3
538	210504	SANTA ROSA		58.68	63.55		0.0089			
539	210602	COJATA		59.66	62.96		0.0091			
540	210604	INCHUPALLA		57.20	69.18		0.0092			
541	210606	ROSASPATA		57.74	67.63		0.0091			2
542	210608	VILQUE CHICO		62.64	65.92		0.0092			
543	210703	CALAPUJA	37.50	56.98	67.71	0.0318	0.0072	0.0060	0.0060	9
544	210707	PARATIA	71.43	65.77	70.55	0.0329	0.0073	0.0056	0.0055	8
545	210802	ANTAUTA		62.25	65.65		0.0101			
546	210805	MACARI		53.82	63.23		0.0089			
547	210808	SANTA ROSA	42.86	55.65	64.10	0.0402	0.0075	0.0060	0.0060	7
548	210901	MOHO		60.72	62.16		0.0090			
549	210902	CONIMA		66.95	58.07		0.0093			
550	210903	HUAYRAPATA		56.63	68.09		0.0090			
551	210904	TILALI		64.45	61.17		0.0092			
552	211004	QUILCAPUNCU		53.88	64.26		0.0089			
553	211005	SINA		60.05	56.76		0.0091			
554	211202	CUYOCUYO	50.00	51.52	51.62	0.0825	0.0080	0.0057	0.0056	4
555	211204	PATAMBUCO		53.00	63.79		0.0091			
556	211206	QUIACA		65.35	67.69		0.0093			
557	211302	ANAPIA		74.16	69.11		0.0110			
558	211303	COPANI	75.00	66.08	66.86	0.0617	0.0083	0.0059	0.0059	5
San Martín										
559	220102	CALZADA	21.05	21.85	24.45	0.0088	0.0045	0.0037	0.0037	24
560	220105	SORITOR	12.50	15.87	19.63	0.0156	0.0058	0.0042	0.0042	9
561	220106	YANTALO		18.90	18.45		0.0089			
562	220201	BELLAVISTA	33.33	22.84	26.17	0.0443	0.0073	0.0050	0.0050	7
563	220202	ALTO BIAVO	41.67	32.82	33.13	0.0218	0.0063	0.0045	0.0045	14
564	220203	BAJO BIAVO	40.00	34.12	33.01	0.0265	0.0066	0.0045	0.0045	13
565	220302	AGUA BLANCA		21.26	21.89		0.0085			
566	220304	SANTA ROSA		25.92	24.72		0.0089			
567	220305	SHATOJA	54.55	29.60	30.90	0.0240	0.0065	0.0045	0.0045	12
568	220401	SAPOSOA		19.90	19.60		0.0086			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
569	220403	EL ESLABON	57.14	25.18	25.84	0.0394	0.0073	0.0050	0.0050	7
570	220404	PISCOYACU	20.00	22.64	27.66	0.0174	0.0058	0.0043	0.0043	11
571	220501	LAMAS	20.00	20.77	25.12	0.0046	0.0031	0.0026	0.0027	37
572	220503	BARRANQUITA	42.86	30.48	31.23	0.0404	0.0074	0.0049	0.0049	8
573	220506	PINTO RECODO		27.25	25.55		0.0086			
574	220507	RUMISAPA		21.17	20.28		0.0090			
575	220508	SAN ROQUE DE CUMBAZA		20.11	20.18		0.0090			
576	220510	TABALOSOS	29.41	24.37	23.37	0.0128	0.0053	0.0037	0.0037	17
577	220511	ZAPATERO	12.50	18.45	21.32	0.0154	0.0057	0.0039	0.0039	11
578	220603	HUICUNGO		24.90	23.14		0.0085			
579	220703	CASPISAPA		14.75	14.75		0.0092			
580	220705	PUCACACA		25.21	24.98		0.0085			
581	220706	SAN CRISTOBAL	55.56	24.34	24.33	0.0287	0.0072	0.0050	0.0050	11
582	220708	SHAMBOYACU	45.45	34.33	39.19	0.0245	0.0064	0.0047	0.0048	11
583	220709	TINGO DE PONASA	60.00	28.58	34.91	0.0592	0.0075	0.0051	0.0051	6
584	220801	RIOJA	23.33	23.53	28.73	0.0061	0.0037	0.0031	0.0033	31
585	220802	AWAJUN	54.55	40.53	43.90	0.0245	0.0065	0.0046	0.0046	11
586	220803	ELIAS SOPLIN VARGAS	28.57	21.90	28.97	0.0339	0.0074	0.0048	0.0048	11
587	220804	NUEVA CAJAMARCA	23.91	23.79	27.41	0.0040	0.0029	0.0024	0.0025	52
588	220806	POSIC		13.90	13.84		0.0096			
589	220807	SAN FERNANDO	42.86	21.66	29.29	0.0400	0.0074	0.0051	0.0051	9
590	220808	YORONGOS	37.50	24.71	28.61	0.0324	0.0076	0.0055	0.0055	8
591	220809	YURACYACU	16.67	21.10	20.98	0.0273	0.0075	0.0054	0.0053	6
592	220901	TARAPOTO	20.48	21.29	21.66	0.0020	0.0016	0.0014	0.0015	93
593	220903	CACATACHI		15.68	16.08		0.0089			
594	220904	CHAZUTA	25.00	22.75	23.77	0.0123	0.0052	0.0037	0.0038	18
595	220905	CHIPURANA		27.85	25.73		0.0087			
596	220906	EL PORVENIR		31.45	28.23		0.0087			
597	220907	HUIMBAYOC	70.00	40.51	35.49	0.0230	0.0063	0.0044	0.0044	10
598	220909	LA BANDA DE SHILCAYO	20.83	23.41	24.70	0.0071	0.0040	0.0029	0.0029	28
599	220910	MORALES	9.68	12.84	15.36	0.0029	0.0022	0.0019	0.0019	37
600	220911	PAPAPLAYA		26.28	26.35		0.0086			
601	220912	SAN ANTONIO		13.15	14.77		0.0092			
602	220914	SHAPAJA	27.27	17.88	19.12	0.0185	0.0063	0.0040	0.0040	12
603	221005	UCHIZA	38.10	34.29	30.43	0.0117	0.0051	0.0035	0.0036	22
Tacna										
604	230406	SUSAPAYA		43.93	43.11		0.0087			

Cuadro 15 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
Ucayali										
605	250102	CAMPOVERDE	50.00	43.65	41.39	0.0085	0.0044	0.0032	0.0032	34
606	250104	MASISEA	41.67	49.59	48.61	0.0219	0.0064	0.0041	0.0041	12
607	250105	YARINACOCHA	42.50	41.96	41.37	0.0012	0.0011	0.0010	0.0010	216
608	250106	NUEVA REQUENA		37.81	33.23		0.0086			
609	250202	SEPAHUA		45.82	41.94		0.0085			
610	250203	TAHUANIA	40.00	48.75	50.12	0.0169	0.0059	0.0039	0.0039	15
611	250204	YURUA		52.11	48.91		0.0091			
612	250302	IRAZOLA	41.30	40.94	40.82	0.0052	0.0033	0.0027	0.0028	55
613	250303	CURIMANA	46.67	45.11	43.31	0.0175	0.0061	0.0040	0.0040	16
614	250304	NESHUYA		35.79	31.26		0.0085			
615	250305	ALEXANDER VON HUMBOLDT		42.85	37.01		0.0091			
616	250401	PURUS	61.54	54.04	54.77	0.0190	0.0062	0.0043	0.0043	14

9.2.2. Riesgo de retraso del crecimiento:Estimaciones

Cuadro 16: Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
Amazonas										
1	010104	CHETO		23.15	24.53		0.0114			
2	010105	CHILIQUN		38.48	38.19		0.0107			
3	010106	CHUQUIBAMBA	58.82	50.49	50.59	0.0137	0.0063	0.0064	0.0063	17
4	010109	LA JALCA		35.60	35.03		0.0103			
5	010111	LEVANTO	60.00	36.22	36.68	0.0558	0.0089	0.0090	0.0089	5
6	010112	MAGDALENA	33.33	31.52	33.92	0.0246	0.0074	0.0074	0.0074	9
7	010113	MARISCAL CASTILLA		32.92	33.77		0.0103			
8	010117	QUINJALCA		53.74	53.84		0.0112			2
9	010119	SAN ISIDRO DE MAINO	37.50	27.51	27.89	0.0279	0.0080	0.0080	0.0079	8
10	010120	SOLOCO	66.67	34.83	35.79	0.1073	0.0094	0.0095	0.0095	3
11	010121	SONCHE		37.34	38.42		0.0104			
12	010202	ARAMANGO	33.33	36.70	36.23	0.0094	0.0051	0.0049	0.0049	25
13	010205	IMAZA	29.27	37.15	37.75	0.0051	0.0036	0.0035	0.0035	41
14	010206	LA PECA	7.69	12.42	12.60	0.0027	0.0022	0.0022	0.0022	26
15	010302	CHISQUILLA		34.88	36.92		0.0110			
16	010304	COROSHA		37.37	38.01		0.0103			
17	010308	RECTA		35.35	37.41		0.0127			
18	010401	NIEVA	58.54	56.56	56.51	0.0060	0.0040	0.0039	0.0039	43
19	010402	EL CENEPA	77.27	73.02	74.09	0.0082	0.0051	0.0049	0.0049	24
20	010403	RIO SANTIAGO	56.41	55.89	57.26	0.0064	0.0042	0.0041	0.0041	40
21	010503	COCABAMBA		45.97	46.45		0.0106			
22	010504	COLCAMAR		32.58	30.79		0.0112			
23	010505	CONILA	45.45	37.83	36.94	0.0231	0.0073	0.0072	0.0072	11
24	010507	LONGUITA	80.00	54.33	53.31	0.0158	0.0063	0.0063	0.0063	10
25	010508	LONYA CHICO	33.33	34.32	37.18	0.0237	0.0075	0.0075	0.0074	9
26	010510	LUYA VIEJO		40.38	39.35		0.0104			
27	010511	MARIA		29.60	27.84		0.0106			
28	010514	PISUQUIA	40.00	42.86	43.13	0.0261	0.0074	0.0076	0.0074	10
29	010515	PROVIDENCIA		39.59	40.09		0.0112			
30	010516	SAN CRISTOBAL		35.52	35.03		0.0104			
31	010517	SAN FRANCISCO DEL YESO	37.50	36.62	37.11	0.0297	0.0077	0.0076	0.0076	8
32	010518	SAN JERONIMO	33.33	37.89	39.59	0.0382	0.0083	0.0084	0.0083	6

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
33	010520	SANTA CATALINA	30.00	31.56	32.18	0.0101	0.0053	0.0052	0.0052	20
34	010521	SANTO TOMAS	11.11	18.46	19.91	0.0054	0.0037	0.0037	0.0037	19
35	010609	OMIA	41.67	42.42	43.46	0.0218	0.0071	0.0070	0.0070	12
36	010612	VISTA ALEGRE		58.09	60.46		0.0112			
37	010702	CAJARURO	38.89	37.31	37.09	0.0044	0.0031	0.0031	0.0031	54
Ancash										
38	020102	COCHABAMBA		46.39	44.69		0.0105			
39	020103	COLCABAMBA		40.18	41.09		0.0114			
40	020104	HUANCHAY		42.70	42.48		0.0117			
41	020105	INDEPENDENCIA	44.00	34.65	35.76	0.0102	0.0051	0.0053	0.0052	26
42	020106	JANGAS		45.44	44.77		0.0102			
43	020107	LA LIBERTAD		47.65	47.19		0.0104			
44	020108	OLLEROS		47.91	47.32		0.0104			
45	020109	PAMPAS GRANDE		40.86	41.43		0.0129			
46	020110	PARIACOTO		38.41	37.64		0.0102			
47	020111	PIRA		48.67	48.09		0.0104			
48	020112	TARICA	55.56	50.35	50.86	0.0141	0.0061	0.0061	0.0060	18
49	020301	LLAMELLIN		43.20	41.83		0.0102			
50	020304	CHINGAS		42.65	41.77		0.0102			
51	020305	MIRGAS		55.91	54.40		0.0105			
52	020306	SAN JUAN DE RONTROY		40.35	39.66		0.0104			
53	020401	CHACAS		48.79	48.87		0.0106			
54	020402	ACOHACA	25.00	48.37	48.71	0.0614	0.0097	0.0099	0.0097	4
55	020501	CHIQUIAN		32.69	34.34		0.0105			
56	020505	CAJACAY		40.96	44.21		0.0116			
57	020507	COLQUIOC		24.25	26.97		0.0109			
58	020508	HUALLANCA		38.01	39.11		0.0102			
59	020509	HUASTA		42.48	41.84		0.0110			
60	020510	HUAYLLACAYAN		43.57	43.18		0.0111			
61	020511	LA PRIMAVERA		47.35	47.28		0.0122			
62	020512	MANGAS		50.82	50.57		0.0110			
63	020513	PACLLON		35.54	34.81		0.0111			
64	020514	SAN MIGUEL DE CORPANQUI		43.10	44.80		0.0131			
65	020515	TICLLOS		44.35	44.52		0.0119			
66	020601	CARHUAZ	55.56	50.11	50.85	0.0093	0.0050	0.0050	0.0049	27
67	020602	ACOPAMPA		42.17	42.96		0.0101			
68	020603	AMASHCA	75.00	64.27	66.35	0.0150	0.0065	0.0065	0.0065	12

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
69	020604	ANTA		42.48	40.78		0.0102			
70	020606	MARCARA		50.10	47.74		0.0106			
71	020607	PARIAHUANCA		48.41	48.68		0.0103			
72	020608	SAN MIGUEL DE ACO	37.50	54.38	57.06	0.0323	0.0082	0.0081	0.0080	8
73	020703	YAUYA		55.58	53.54		0.0105			
74	020801	CASMA	20.00	22.17	20.89	0.0041	0.0029	0.0030	0.0029	40
75	020802	BUENA VISTA ALTA		40.25	38.65		0.0102			5
76	020803	COMANDANTE NOEL		26.48	26.32		0.0102			
77	020804	YAUTAN		36.15	36.35		0.0100			
78	020901	CORONGO		40.41	40.09		0.0102			
79	020902	ACO		47.57	46.28		0.0104			
80	020904	CUSCA		48.80	48.39		0.0111			
81	020905	LA PAMPA		37.05	36.21		0.0103			
82	020906	YANAC		38.96	37.77		0.0106			
83	020907	YUPAN		37.78	39.34		0.0128			
84	021003	CAJAY		56.49	55.10		0.0105			
85	021004	CHAVIN DE HUANTAR	61.11	56.56	57.31	0.0137	0.0061	0.0061	0.0060	18
86	021005	HUACACHI		47.89	47.01		0.0106			
87	021402	ACAS		41.70	43.82		0.0126			
88	021404	CARHUAPAMPA		38.33	39.00		0.0115			
89	021506	LACABAMBA		44.37	48.81		0.0119			
90	021509	PAMPAS		45.33	50.37		0.0112			
91	021510	SANTA ROSA		26.17	31.10		0.0117			
92	021511	TAUCA		33.02	37.82		0.0121			
93	021602	HUAYLLAN		51.16	54.77		0.0113			
94	021701	RECUAY		30.78	35.54		0.0108			
95	021703	COTAPARACO		29.00	34.72		0.0133			
96	021705	LLACLIN		32.68	37.53		0.0110			
97	021706	MARCA		58.42	63.60		0.0119			5
98	021709	TAPACOCHA		37.50	42.56		0.0115			
99	021801	CHIMBOTE	12.65	12.44	12.66	0.0007	0.0006	0.0006	0.0006	170
100	021804	MACATE		28.53	33.69		0.0130			
101	021808	SANTA		13.93	18.06		0.0108			8
102	021901	SIHUAS		25.78	30.29		0.0107			
103	021903	ALFONSO UGARTE		39.45	44.40		0.0122			
104	021905	CHINGALPO	41.67	42.12	44.31	0.0192	0.0074	0.0072	0.0072	12
105	021908	RAGASH		45.61	49.72		0.0111			

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
106	022001	YUNGAY		37.72	41.94		0.0106			
107	022006	RANRAHIRCA		45.65	50.07		0.0107			
108	022007	SHUPLUY		38.93	42.58		0.0108			
109	022008	YANAMA		42.37	46.26		0.0108			
Apurímac										
110	030101	ABANCAY	17.97	17.76	18.08	0.0011	0.0010	0.0010	0.0010	129
111	030103	CIRCA	33.33	47.15	46.32	0.1081	0.0093	0.0095	0.0094	3
112	030104	CURAHUASI	36.84	37.00	36.33	0.0127	0.0058	0.0058	0.0058	19
113	030105	HUANIPACA	60.00	49.37	48.68	0.0585	0.0089	0.0091	0.0090	5
114	030106	LAMBRAMA	33.33	47.72	49.02	0.0430	0.0087	0.0088	0.0087	6
115	030107	PICHIRHUA		45.50	45.48		0.0103			
116	030108	SAN PEDRO DE CACHORA		42.98	44.96		0.0103			4
117	030109	TAMBURCO		21.92	23.31		0.0102			
118	030201	ANDAHUAYLAS	38.82	37.13	37.26	0.0028	0.0022	0.0022	0.0022	87
119	030202	ANDARAPA	50.00	47.22	47.67	0.0351	0.0084	0.0084	0.0084	8
120	030204	HUANCARAMA	40.00	43.35	43.04	0.0120	0.0057	0.0058	0.0057	22
121	030206	HUAYANA		48.55	49.12		0.0104			
122	030207	KISHUARA	42.86	48.57	49.05	0.0404	0.0085	0.0085	0.0082	7
123	030208	PACOBAMBA		44.68	46.62		0.0106			
124	030209	PACUCHA	42.11	48.24	48.94	0.0131	0.0059	0.0059	0.0060	19
125	030210	PAMPACHIRI	85.71	63.12	62.37	0.0197	0.0068	0.0068	0.0070	7
126	030213	SAN JERONIMO	30.23	33.57	33.27	0.0049	0.0034	0.0034	0.0034	43
127	030215	SANTA MARIA DE CHICMO	45.00	45.36	46.22	0.0060	0.0040	0.0039	0.0039	40
128	030217	TUMAY HUARACA		50.35	54.54		0.0109			
129	030218	TURPO	33.33	38.86	40.92	0.0150	0.0063	0.0063	0.0062	15
130	030219	KAQUIABAMBA		53.85	53.49		0.0107			
131	030220	JOSE MARIA ARGUEDAS		52.26	53.12		0.0102			
132	030302	EL ORO	33.33	36.50	37.12	0.0222	0.0077	0.0078	0.0078	9
133	030402	CAPAYA		35.19	40.09		0.0108			
134	030406	COTARUSE		38.50	43.45		0.0109			3
135	030408	JUSTO APU SAHUARAURA		43.23	47.59		0.0106			
136	030410	POCOHUANCA		34.50	39.13		0.0109			
137	030606	ONGOY		44.18	48.53		0.0107			
138	030611	LOS CHANKAS		44.14	48.94		0.0110			
Arequipa										
139	040106	CHIGUATA	16.67	26.00	26.63	0.0120	0.0057	0.0056	0.0056	12
140	040107	JACOBO HUNTER	11.36	11.81	12.14	0.0023	0.0019	0.0019	0.0019	47

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
141	040110	MIRAFLORES	8.33	9.42	9.86	0.0033	0.0026	0.0025	0.0026	25
142	040111	MOLLEBAYA		31.66	36.61		0.0112			
143	040114	POLOBAYA		33.56	38.13		0.0108			
144	040115	QUEQUEÑA	25.00	31.80	33.75	0.0263	0.0081	0.0078	0.0077	8
145	040402	ANDAGUA		32.70	37.94		0.0114			
146	040403	AYO		27.77	33.44		0.0125			
147	040405	CHILCAYMARCA		45.15	50.29		0.0124			
148	040410	PAMPACOLCA	33.33	30.48	30.84	0.1088	0.0097	0.0092	0.0092	3
149	040607	SALAMANCA		45.10	42.43		0.0112			
150	040608	YANAQUIHUA		42.46	43.21		0.0107			
151	040702	COCACHACRA		27.40	27.18		0.0102			16
152	040704	ISLAY		23.11	22.38		0.0103			
153	040705	MEJIA		17.23	15.53		0.0107			
Ayacucho										
154	050110	SAN JUAN BAUTISTA	13.51	15.77	14.66	0.0032	0.0025	0.0025	0.0025	37
155	050401	HUANTA	25.00	26.36	25.99	0.0048	0.0033	0.0033	0.0033	40
156	050408	LLOCHEGUA	14.29	22.48	22.42	0.0203	0.0070	0.0071	0.0070	7
157	050502	ANCO		39.91	37.87		0.0104			
158	050611	LUCANAS		46.21	44.58		0.0101			
159	050618	SAN PEDRO DE PALCO		51.78	51.23		0.0101			
160	050620	SANTA ANA DE HUAYCAHUACHO		46.60	48.38		0.0107			
161	050621	SANTA LUCIA		34.34	34.91		0.0103			
162	050702	CHUMPI	20.00	30.98	29.87	0.0168	0.0066	0.0066	0.0065	10
163	050703	CORONEL CASTAÑEDA		52.90	55.87		0.0115			
164	050706	PUYUSCA		38.85	37.86		0.0104			6
165	050707	SAN FRANCISCO DE RAVACAYCO	16.67	45.56	43.41	0.0240	0.0074	0.0075	0.0074	6
166	050901	QUEROBAMBA	57.14	43.34	43.58	0.0393	0.0081	0.0082	0.0081	7
167	051001	HUANCAPI		42.77	42.10		0.0101			
168	051007	COLCA		47.03	45.61		0.0101			
169	051012	VILCANCHOS	66.67	46.18	45.14	0.0430	0.0085	0.0086	0.0084	7
170	051104	CONCEPCION		52.60	50.49		0.0103			
Cajamarca										
171	060102	ASUNCION	66.67	47.31	55.89	0.1107	0.0094	0.0095	0.0094	3
172	060103	CHETILLA		59.38	57.41		0.0119			
173	060104	COSPAN		45.14	45.52		0.0101			
174	060105	ENCAÑADA	90.00	72.99	74.67	0.0099	0.0052	0.0052	0.0052	10
175	060106	JESUS	57.14	44.49	51.50	0.0406	0.0082	0.0083	0.0083	7

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
176	060107	LLACANORA		45.87	46.14		0.0106			
177	060109	MAGDALENA	80.00	54.02	57.97	0.0175	0.0065	0.0066	0.0066	10
178	060110	MATARA		45.24	44.83		0.0102			
179	060111	NAMORA		51.08	50.33		0.0101			
180	060112	SAN JUAN	85.71	60.01	65.46	0.0201	0.0069	0.0069	0.0068	7
181	060201	CAJABAMBA	46.15	42.49	44.17	0.0065	0.0040	0.0041	0.0040	39
182	060202	CACHACHI	66.67	63.79	65.75	0.0053	0.0036	0.0036	0.0036	42
183	060203	CONDEBAMBA	50.00	49.95	53.04	0.0498	0.0086	0.0087	0.0087	6
184	060204	SITACOCHA	72.73	59.08	60.25	0.0196	0.0068	0.0069	0.0068	12
185	060301	CELENDIN		39.19	39.42		0.0100			
186	060302	CHUMUCH		48.58	49.05		0.0103			
187	060303	CORTEGANA	66.67	55.09	55.23	0.0441	0.0084	0.0085	0.0084	6
188	060304	HUASMIN	45.45	50.48	52.98	0.0245	0.0075	0.0074	0.0074	11
189	060305	JORGE CHAVEZ		39.86	40.78		0.0116			
190	060307	MIGUEL IGLESIAS		47.66	46.93		0.0104			
191	060308	OXAMARCA	55.56	50.95	53.83	0.0304	0.0078	0.0079	0.0079	9
192	060309	SOROCHUCO	83.33	62.98	64.13	0.0275	0.0078	0.0077	0.0076	6
193	060310	SUCRE		41.54	41.96		0.0104			
194	060311	UTCO		49.29	49.96		0.0102			
195	060312	LA LIBERTAD DE PALLAN		49.42	48.63		0.0104			
196	060402	ANGUIA		47.25	46.74		0.0107			
197	060403	CHADIN		45.91	45.99		0.0106			
198	060404	CHIGUIRIP		48.37	47.67		0.0109			
199	060405	CHIMBAN	30.00	39.15	39.20	0.0223	0.0074	0.0075	0.0074	11
200	060406	CHOROPAMPA		49.13	48.36		0.0105			
201	060407	COCHABAMBA		42.58	43.84		0.0102			
202	060408	CONCHAN		43.40	43.14		0.0103			
203	060409	HUAMBOS		45.60	45.94		0.0101			
204	060410	LAJAS	26.92	35.53	34.45	0.0076	0.0045	0.0045	0.0045	26
205	060411	LLAMA		30.84	32.04		0.0104			
206	060412	MIRACOSTA		43.91	43.12		0.0104			
207	060413	PACCHA		42.84	43.40		0.0103			
208	060414	PION	44.44	37.55	35.78	0.0285	0.0078	0.0077	0.0077	9
209	060416	SAN JUAN DE LICUPIS		40.47	40.45		0.0118			
210	060417	TACABAMBA		49.49	49.76		0.0105			
211	060418	TOCMOCHE		41.65	44.32		0.0108			
212	060419	CHALAMARCA	33.33	47.34	44.91	0.0441	0.0089	0.0087	0.0087	6

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
213	060501	CONTUMAZA		41.91	42.48		0.0104			
214	060502	CHILETE		21.62	24.96		0.0105			
215	060503	CUPISNIQUE		41.57	43.93		0.0116			
216	060504	GUZMANGO		48.68	49.84		0.0118			
217	060505	SAN BENITO		37.03	40.04		0.0112			
218	060506	SANTA CRUZ DE TOLED		49.96	51.41		0.0115			
219	060507	TANTARICA		37.49	37.58		0.0106			
220	060601	CUTERVO	60.00	47.56	43.93	0.0266	0.0074	0.0075	0.0074	10
221	060602	CALLAYUC	29.41	43.76	44.39	0.0127	0.0059	0.0060	0.0060	18
222	060603	CHOROS		42.55	46.48		0.0106			
223	060604	CUJILLO		32.90	35.01		0.0108			
224	060605	LA RAMADA		45.23	47.06		0.0102			
225	060606	PIMPINGOS		41.34	42.86		0.0103			
226	060607	QUEROCOTILLO		45.93	46.77		0.0101			
227	060608	SAN ANDRES DE CUTERVO	25.00	34.85	33.81	0.0261	0.0078	0.0079	0.0078	8
228	060609	SAN JUAN DE CUTERVO		34.24	35.21		0.0105			
229	060610	SAN LUIS DE LUCMA	40.00	45.56	44.80	0.0255	0.0075	0.0077	0.0075	10
230	060611	SANTA CRUZ		45.84	47.72		0.0104			
231	060612	SANTO DOMINGO DE LA CAPILLA	50.00	44.85	43.49	0.0492	0.0089	0.0089	0.0088	6
232	060613	SANTO TOMAS	45.83	42.78	40.93	0.0103	0.0052	0.0054	0.0053	25
233	060614	SOCOTA		36.52	37.91		0.0106			
234	060615	TORIBIO CASANOVA		33.74	34.52		0.0111			
235	060701	BAMBAMARCA	45.71	48.50	47.75	0.0072	0.0044	0.0044	0.0044	35
236	060702	CHUGUR		45.77	46.19		0.0114			
237	060703	HUALGAYOC	50.00	54.20	51.24	0.0225	0.0075	0.0073	0.0073	12
238	060804	COLASAY		35.93	36.97		0.0101			
239	060805	HUABAL		45.34	46.95		0.0108			
240	060806	LAS PIRIAS		42.93	44.45		0.0102			
241	060807	POMAHUACA		44.16	46.28		0.0108			
242	060809	SALLIQUE	80.00	60.02	60.11	0.0176	0.0072	0.0074	0.0073	10
243	060810	SAN FELIPE		43.33	44.39		0.0108			
244	060901	SAN IGNACIO	54.55	39.74	39.72	0.0247	0.0073	0.0072	0.0071	11
245	060902	CHIRINOS		38.08	39.41		0.0101			
246	060903	HUARANGO		35.26	38.45		0.0105			
247	060904	LA COIPA		37.58	39.34		0.0102			
248	060905	NAMBALLE		33.20	36.66		0.0109			
249	060906	SAN JOSE DE LOURDES	30.00	33.77	34.65	0.0109	0.0055	0.0054	0.0053	20

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
250	060907	TABACONAS	34.62	41.21	42.70	0.0089	0.0049	0.0049	0.0049	27
251	061002	CHANCAY		39.14	38.51		0.0103			
252	061004	GREGORIO PITA		48.99	48.44		0.0106			
253	061005	ICHOCAN		35.84	35.47		0.0102			
254	061006	JOSE MANUEL QUIROZ	36.36	42.16	44.42	0.0223	0.0076	0.0077	0.0076	11
255	061007	JOSE SABOGAL	71.43	60.95	62.04	0.0338	0.0081	0.0081	0.0081	7
256	061101	SAN MIGUEL	36.84	39.01	39.53	0.0127	0.0058	0.0058	0.0058	19
257	061102	BOLIVAR		27.66	31.29		0.0117			
258	061103	CALQUIS	66.67	47.78	47.81	0.0437	0.0091	0.0091	0.0091	6
259	061104	CATILLUC		51.08	51.20		0.0102			
260	061105	EL PRADO		40.06	39.91		0.0106			
261	061106	LA FLORIDA		35.04	38.54		0.0111			
262	061107	LLAPA		45.20	44.47		0.0103			
263	061108	NANCHOC		33.96	38.82		0.0115			
264	061109	NIEPOS	14.29	34.87	35.56	0.0199	0.0072	0.0073	0.0073	7
265	061110	SAN GREGORIO		33.86	35.96		0.0114			
266	061111	SAN SILVESTRE DE COCHAN		44.58	44.39		0.0105			
267	061112	TONGOD		41.01	41.49		0.0110			
268	061113	UNION AGUA BLANCA	33.33	40.75	40.91	0.0191	0.0069	0.0069	0.0069	12
269	061201	SAN PABLO	55.56	46.29	48.25	0.0306	0.0077	0.0078	0.0078	9
270	061202	SAN BERNARDINO		49.35	51.12		0.0104			
271	061203	SAN LUIS		39.87	42.17		0.0106			
272	061204	TUMBADEN		57.20	56.93		0.0107			
273	061302	ANDABAMBA		45.67	45.58		0.0109			
274	061303	CATACHE		46.40	49.53		0.0107			
275	061304	CHANCAYBAÑOS		46.38	47.94		0.0105			
276	061305	LA ESPERANZA		46.12	47.24		0.0108			
277	061306	NINABAMBA		39.77	40.55		0.0108			
278	061307	PULAN		38.82	40.17		0.0111			
279	061308	SAUCEPAMPA		33.71	34.80		0.0106			
280	061309	SEXI		47.76	50.31		0.0133			
281	061310	UTICYACU		45.23	45.27		0.0107			
282	061311	YAUYUCAN		42.12	42.15		0.0103			
Callao										
283	070106	VENTANILLA	18.36	17.87	17.98	0.0004	0.0004	0.0004	0.0004	355
Cusco										
284	080101	CUSCO	19.51	19.77	19.50	0.0039	0.0029	0.0028	0.0029	43

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
285	080102	CCORCA		54.75	52.45		0.0107			
286	080103	POROY	58.33	41.05	37.18	0.0210	0.0069	0.0068	0.0068	12
287	080104	SAN JERONIMO	20.59	20.57	20.52	0.0049	0.0034	0.0033	0.0033	34
288	080105	SAN SEBASTIAN	15.79	16.76	16.68	0.0024	0.0019	0.0019	0.0019	57
289	080106	SANTIAGO	13.46	15.02	14.63	0.0023	0.0019	0.0019	0.0019	54
290	080108	WANCHAQ	25.00	20.78	20.51	0.0069	0.0043	0.0043	0.0042	28
291	080404	LARES	62.50	54.39	53.42	0.0330	0.0085	0.0087	0.0085	8
292	080406	SAN SALVADOR		42.04	41.08		0.0105			
293	080601	SICUANI	10.71	17.29	17.06	0.0035	0.0027	0.0027	0.0027	28
294	080602	CHECACUPE		44.97	45.82		0.0100			
295	080604	MARANGANI		46.90	45.95		0.0102			
296	080605	PITUMARCA		49.84	48.88		0.0107			
297	080608	TINTA	20.00	37.40	34.90	0.0395	0.0081	0.0082	0.0081	5
298	080701	SANTO TOMAS		49.24	47.71		0.0102			
299	080703	CHAMACA		51.52	50.03		0.0103			
300	080705	LIVITACA		58.64	56.97		0.0105			
301	080708	VELILLE	28.57	45.33	47.30	0.0336	0.0083	0.0084	0.0084	7
302	080801	ESPINAR	53.33	47.34	47.67	0.0085	0.0048	0.0049	0.0049	30
303	080802	CONDOROMA		46.32	45.78		0.0115			
304	080805	PALLPATA		54.18	53.04		0.0109			
305	080807	SUYCKUTAMBO		57.17	54.90		0.0106			
306	080808	ALTO PICHIGUA		60.95	59.70		0.0109			
307	080901	SANTA ANA	4.76	8.15	8.06	0.0022	0.0018	0.0019	0.0019	21
308	080907	KIMBIRI	21.43	24.97	24.79	0.0128	0.0058	0.0060	0.0060	14
309	080908	SANTA TERESA		34.48	32.59		0.0101			11
310	080909	VILCABAMBA	25.00	37.59	37.78	0.0621	0.0088	0.0089	0.0089	4
311	080911	INKAWASI		45.59	46.97		0.0109			
312	080913	VILLA KINTIARINA		47.57	46.05		0.0102			
313	080914	MEGANTONI		54.63	53.85		0.0110			
314	081001	PARURO		40.78	39.02		0.0101			
315	081007	PACCARITAMBO		52.13	49.56		0.0108			
316	081102	CAICAY	50.00	51.30	51.30	0.0217	0.0076	0.0077	0.0077	12
317	081201	URCOS		34.18	34.39		0.0100			
318	081203	CAMANTI	20.00	31.46	29.86	0.0170	0.0065	0.0066	0.0065	10
319	081204	CCARHUAYO		50.49	50.90		0.0101			
320	081205	CCATCA	58.33	57.02	53.29	0.0219	0.0074	0.0075	0.0074	12
321	081206	CUSIPATA	54.55	47.32	49.86	0.0240	0.0074	0.0073	0.0073	11

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
322	081207	HUARO		43.03	41.63		0.0103			
323	081208	LUCRE		27.09	30.43		0.0103			
324	081209	MARCAPATA		49.84	49.11		0.0101			
325	081210	OCONGATE	36.36	43.82	41.58	0.0230	0.0071	0.0071	0.0071	11
326	081211	OROPESA	34.62	29.54	27.29	0.0088	0.0048	0.0048	0.0048	26
327	081212	QUIQUIJANA		54.34	52.73		0.0112			
328	081301	URUBAMBA		32.30	31.23		0.0101			
329	081307	YUCAY		27.53	26.09		0.0103			
Huancavelica										
330	090107	HUAYLLAHUARA		50.74	51.52		0.0107			
331	090114	PALCA	83.33	60.34	58.25	0.0263	0.0078	0.0078	0.0077	6
332	090201	ACOBAMBA	38.46	39.34	39.43	0.0194	0.0068	0.0069	0.0068	13
333	090202	ANDABAMBA	63.16	56.41	54.81	0.0122	0.0057	0.0058	0.0058	19
334	090203	ANTA		53.65	50.10		0.0112			
335	090204	CAJA	75.00	49.05	48.51	0.0605	0.0090	0.0091	0.0090	4
336	090205	MARCAS		47.15	45.00		0.0103			
337	090401	CASTROVIRREYNA		40.64	40.45		0.0103			
338	090406	COCAS		43.91	43.04		0.0107			
339	090407	HUACHOS		35.22	35.68		0.0127			
340	090411	SANTA ANA		54.52	55.50		0.0110			1
341	090505	LA MERCED		42.23	41.71		0.0108			
342	090611	SAN FRANCISCO DE SANGAYAICO	57.14	50.10	49.63	0.0367	0.0086	0.0087	0.0086	7
343	090701	PAMPAS	33.33	33.68	33.33	0.0109	0.0054	0.0054	0.0054	21
344	090702	ACOSTAMBO		40.73	38.84		0.0104			
345	090711	PAZOS	80.00	54.64	52.95	0.0396	0.0083	0.0083	0.0082	6
Huánuco										
346	100101	HUANUCO	17.31	18.04	17.98	0.0014	0.0012	0.0012	0.0012	107
347	100111	PILLCO MARCA	25.71	24.29	24.14	0.0056	0.0037	0.0036	0.0036	35
348	100203	COLPAS		56.29	52.35		0.0109			
349	100609	PUEBLO NUEVO		49.61	46.40		0.0105			
350	100705	SANTA ROSA DE ALTO YANAJANCA		50.60	48.68		0.0108			
351	101101	CHAVINILLO	66.67	59.02	57.24	0.0438	0.0085	0.0087	0.0086	6
352	101105	JACAS CHICO		49.71	49.75		0.0102			
353	101106	OBAS	75.00	51.30	51.13	0.0616	0.0091	0.0094	0.0094	4
354	101107	PAMPAMARCA		59.63	57.47		0.0105			
Junín										
355	120101	HUANCAYO	34.43	29.51	31.42	0.0037	0.0028	0.0028	0.0028	62

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
356	120104	CARHUACALLANGA		60.92	62.56		0.0118			
357	120105	CHACAPAMPA		52.60	53.18		0.0102			
358	120106	CHICCHE		43.84	48.07		0.0106			
359	120107	CHILCA	42.86	39.17	39.51	0.0032	0.0025	0.0025	0.0025	78
360	120108	CHONGOS ALTO		46.39	45.94		0.0103			
361	120112	COLCA		49.08	48.64		0.0104			
362	120114	EL TAMBO	38.38	35.06	35.40	0.0024	0.0020	0.0020	0.0020	100
363	120116	HUACRAPUQUIO		36.27	35.54		0.0104			
364	120117	HUALHUAS		35.68	35.23		0.0103			
365	120119	HUANCAN		27.48	31.74		0.0109			
366	120120	HUASICANCHA		55.82	56.70		0.0107			
367	120121	HUAYUCACHI		31.85	35.79		0.0106			7
368	120122	INGENIO		45.99	45.83		0.0109			
369	120124	PARIAHUANCA		51.13	54.51		0.0109			3
370	120125	PILCOMAYO	36.00	31.35	29.64	0.0095	0.0050	0.0051	0.0050	27
371	120126	PUCARA		38.79	37.45		0.0102			
372	120127	QUICHUAY		43.23	43.18		0.0104			
373	120128	QUILCAS	62.50	45.42	44.80	0.0327	0.0079	0.0080	0.0079	8
374	120130	SAN JERONIMO DE TUNAN		29.00	29.42		0.0103			
375	120132	SAÑO	12.50	28.31	29.79	0.0154	0.0063	0.0061	0.0062	9
376	120134	SICAYA	33.33	31.29	29.88	0.0443	0.0085	0.0086	0.0086	6
377	120201	CONCEPCION	14.29	24.78	26.17	0.0203	0.0069	0.0068	0.0069	7
378	120202	ACO		32.93	31.84		0.0107			
379	120204	CHAMBARA		41.65	41.25		0.0102			
380	120205	COCHAS		49.72	50.71		0.0114			
381	120206	COMAS		53.22	51.80		0.0104			
382	120208	MANZANARES		35.25	35.32		0.0106			
383	120209	MARISCAL CASTILLA		48.86	48.14		0.0104			
384	120210	MATAHUASI	12.50	27.59	27.31	0.0154	0.0064	0.0063	0.0063	8
385	120211	MITO		32.36	32.37		0.0106			
386	120212	NUEVE DE JULIO		42.70	42.31		0.0107			
387	120213	ORCOTUNA		30.39	29.58		0.0102			
388	120301	CHANCHAMAYO	18.18	19.72	19.69	0.0070	0.0042	0.0042	0.0042	23
389	120303	PICHANAQUI	35.71	39.03	39.14	0.0176	0.0065	0.0066	0.0065	14
390	120305	SAN RAMON	16.67	23.11	23.90	0.0277	0.0075	0.0076	0.0076	6
391	120306	VITOC		34.31	34.81		0.0102			
392	120402	ACOLLA		36.93	36.25		0.0103			

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
393	120403	APATA	25.00	38.63	38.46	0.0165	0.0067	0.0067	0.0067	12
394	120408	HUAMALI		37.44	37.19		0.0107			
395	120409	HUARIPAMPA		36.47	37.50		0.0107			
396	120412	JULCAN		36.32	35.28		0.0106			
397	120414	LLOCLLAPAMPA	30.00	40.88	40.75	0.0206	0.0074	0.0076	0.0075	11
398	120415	MARCO		41.30	42.19		0.0107			
399	120417	MASMA CHICCHE		45.27	45.40		0.0112			
400	120419	MONOBAMBA		36.04	35.57		0.0107			
401	120420	MUQUI		35.85	35.92		0.0112			
402	120422	PACA		39.25	38.15		0.0104			
403	120423	PACCHA		37.73	37.99		0.0104			
404	120424	PANCAN		37.61	37.93		0.0106			
405	120425	PARCO		32.45	32.13		0.0106			
406	120426	POMACANCHA		46.61	46.14		0.0101			
407	120427	RICRAN		44.15	43.28		0.0102			
408	120429	SAN PEDRO DE CHUNAN		53.55	53.05		0.0113			
409	120430	SAUSA	33.33	35.18	34.15	0.0435	0.0085	0.0086	0.0086	6
410	120433	YAULI		41.28	40.83		0.0109			
411	120501	JUNIN	44.44	39.54	39.34	0.0305	0.0081	0.0084	0.0083	9
412	120503	ONDORES		49.57	49.50		0.0107			
413	120504	ULCUMAYO		49.93	47.90		0.0104			
414	120601	SATIPO	31.82	32.16	31.64	0.0050	0.0034	0.0034	0.0034	45
415	120602	COVIRIALI	33.33	44.64	44.57	0.0440	0.0086	0.0087	0.0085	6
416	120603	LLAYLLA		51.81	51.19		0.0108			
417	120605	PAMPA HERMOSA		55.81	54.64		0.0104			
418	120608	RIO TAMBO	73.08	68.38	66.75	0.0078	0.0047	0.0047	0.0046	26
419	120609	VIZCATAN DEL ENE	42.00	43.44	43.14	0.0045	0.0032	0.0032	0.0032	52
420	120701	TARMA	40.82	35.87	35.75	0.0050	0.0034	0.0034	0.0034	50
421	120702	ACOBAMBA	14.29	26.42	27.80	0.0202	0.0069	0.0069	0.0069	7
422	120703	HUARICOLCA	66.67	48.55	47.22	0.0144	0.0062	0.0063	0.0063	15
423	120705	LA UNION		39.12	38.64		0.0106			
424	120706	PALCA	50.00	36.97	36.40	0.0493	0.0086	0.0087	0.0086	6
425	120707	PALCAMAYO	25.00	30.77	30.91	0.0260	0.0076	0.0076	0.0076	8
426	120708	SAN PEDRO DE CAJAS		40.31	39.86		0.0103			
427	120709	TAPO	40.00	41.95	42.14	0.0165	0.0064	0.0065	0.0064	15
428	120801	LA OROYA	75.00	35.87	36.25	0.0623	0.0093	0.0091	0.0091	4
429	120802	CHACAPALPA	75.00	58.14	58.48	0.0587	0.0094	0.0095	0.0095	4

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
430	120803	HUAY-HUAY		40.77	41.78		0.0116			
431	120804	MARCAPOMACOCHA		47.66	47.38		0.0117			
432	120805	MOROCOCHA		35.35	35.69		0.0120			
433	120806	PACCHA		39.00	39.60		0.0111			
434	120807	SANTA BARBARA DE CARHUACAYAN		48.14	49.18		0.0118			
435	120808	SANTA ROSA DE SACCO	25.00	27.35	27.85	0.0167	0.0066	0.0067	0.0066	13
436	120809	SUITUCANCHA		47.99	48.48		0.0113			
437	120810	YAULI	50.00	42.19	43.25	0.0219	0.0079	0.0081	0.0080	12
438	120901	CHUPACA	14.29	28.00	29.74	0.0203	0.0071	0.0069	0.0069	7
439	120902	AHUAC	25.00	36.90	37.00	0.0264	0.0076	0.0076	0.0075	8
440	120903	CHONGOS BAJO		34.73	34.27		0.0102			
441	120904	HUACHAC		37.50	37.11		0.0104			
442	120906	SAN JUAN DE ISCOS		46.02	45.97		0.0109			
443	120907	SAN JUAN DE JARPA	47.62	53.71	54.76	0.0112	0.0057	0.0057	0.0056	21
La Libertad										
444	130402	PACANGA	16.67	27.28	26.42	0.0277	0.0075	0.0077	0.0076	6
445	130403	PUEBLO NUEVO	20.00	22.49	22.65	0.0176	0.0065	0.0066	0.0066	10
Lima										
446	150102	ANCON	8.70	10.57	10.51	0.0036	0.0027	0.0027	0.0027	23
447	150107	CHACLACAYO		9.94	8.59		0.0108			4
448	150110	COMAS	17.75	17.60	17.58	0.0009	0.0008	0.0008	0.0008	170
449	150111	EL AGUSTINO	15.91	16.39	16.27	0.0031	0.0025	0.0024	0.0024	47
450	150136	SAN MIGUEL	13.64	13.74	13.37	0.0056	0.0039	0.0038	0.0038	22
451	150137	SANTA ANITA	20.00	19.97	19.81	0.0047	0.0033	0.0033	0.0033	35
452	150205	SUPE PUERTO	26.67	19.67	19.30	0.0138	0.0061	0.0060	0.0061	15
453	150303	GORGOR		36.47	35.60		0.0107			
454	150304	HUANCAPON	71.43	53.69	54.24	0.0300	0.0083	0.0084	0.0083	7
455	150401	CANTA		26.24	25.20		0.0107			
456	150707	HUACHUPAMPA		31.44	32.25		0.0113			
457	150708	HUANZA		38.46	39.43		0.0119			
458	150709	HUAROCHIRI		28.99	29.71		0.0115			
459	150710	LAHUAYTAMBO		35.91	34.71		0.0106			
460	150711	LANGA	33.33	36.77	37.12	0.0403	0.0092	0.0093	0.0091	6
461	150713	MARIATANA		21.33	20.93		0.0112			
462	150716	SAN ANTONIO		50.41	51.38		0.0130			
463	150717	SAN BARTOLOME	33.33	25.02	24.49	0.1065	0.0098	0.0097	0.0096	3
464	150721	SAN LORENZO DE QUINTI		37.09	36.39		0.0112			

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
465	150724	SAN PEDRO DE CASTA		26.60	25.97		0.0118			
466	150801	HUACHO	10.20	10.97	11.34	0.0019	0.0016	0.0016	0.0016	49
467	150805	HUALMAY	15.79	13.97	14.09	0.0073	0.0044	0.0044	0.0044	19
468	150807	LEONCIO PRADO	66.67	40.80	42.63	0.0261	0.0075	0.0075	0.0075	10
469	150810	SANTA MARIA	14.71	16.26	17.01	0.0038	0.0028	0.0028	0.0028	34
470	150812	VEGUETA	31.03	28.60	29.24	0.0076	0.0044	0.0044	0.0044	29
471	150904	COCHAMARCA		39.18	38.66		0.0130			
472	150905	NAVAN	50.00	36.57	39.96	0.2418	0.0111	0.0114	0.0112	2
473	151003	ALLAUCA		34.14	34.32		0.0108			
474	151004	AYAVIRI		30.55	30.22		0.0118			
475	151012	HONGOS		42.46	42.72		0.0104			
476	151014	HUANCAYA		27.56	29.19		0.0114			
477	151019	LINCHA		43.20	42.92		0.0114			
478	151020	MADEAN		33.63	34.31		0.0107			
479	151021	MIRAFLORES		30.00	31.10		0.0129			
480	151023	PUTINZA		23.66	22.94		0.0112			
481	151024	QUINCHES		35.54	36.24		0.0130			
482	151026	SAN JOAQUIN		35.44	36.36		0.0142			
483	151027	SAN PEDRO DE PILAS		28.12	27.90		0.0105			
484	151028	TANTA		34.77	34.57		0.0110			
485	151033	VITIS		33.61	34.43		0.0149			
Loreto										
486	160507	SAQUENA	50.00	43.09	43.62	0.0350	0.0083	0.0084	0.0084	8
487	160510	JENARO HERRERA		43.57	42.27		0.0108			
Pasco										
488	190103	HUARIACA	21.43	27.93	28.97	0.0127	0.0057	0.0057	0.0057	14
489	190104	HUAYLLAY	90.00	67.62	65.76	0.0099	0.0055	0.0055	0.0055	10
490	190106	PALLANCHACRA		43.72	43.39		0.0102			8
491	190107	PAUCARTAMBO	17.65	21.52	21.92	0.0027	0.0022	0.0022	0.0022	52
492	190109	SIMON BOLIVAR	51.02	47.19	46.34	0.0050	0.0035	0.0035	0.0035	50
493	190110	TICLACAYAN	25.00	41.19	42.71	0.0261	0.0074	0.0074	0.0074	8
494	190303	HUANCABAMBA	28.57	36.12	36.18	0.0154	0.0062	0.0062	0.0062	14
Piura										
495	200105	CATACAOS	48.15	41.49	38.44	0.0096	0.0050	0.0051	0.0050	27
496	200108	EL TALLAN	33.33	43.56	40.30	0.0440	0.0084	0.0083	0.0083	6
497	200110	LA UNION	57.14	46.95	44.59	0.0188	0.0067	0.0067	0.0066	14
498	200210	SUYO		27.95	27.52		0.0104			

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
499	200304	HUARMACA	70.97	63.11	60.41	0.0068	0.0042	0.0042	0.0042	33
500	200305	LALAQUIZ	33.33	49.38	45.71	0.0436	0.0088	0.0087	0.0086	6
501	200307	SONDOR	30.00	47.35	44.59	0.0230	0.0073	0.0072	0.0072	10
502	200308	SONDORILLO		66.22	63.38		0.0112			
503	200401	CHULUCANAS	20.00	22.74	22.36	0.0029	0.0023	0.0023	0.0023	55
504	200402	BUENOS AIRES		27.30	24.94		0.0103			
505	200403	CHALACO		39.64	37.63		0.0102			
506	200406	SALITRAL		29.32	33.12		0.0106			
507	200407	SAN JUAN DE BIGOTE		35.69	36.63		0.0103			
508	200409	SANTO DOMINGO		32.53	37.19		0.0111			
509	200410	YAMANGO		44.93	47.32		0.0103			
510	200505	LA HUACA		30.75	29.09		0.0102			
511	200507	VICHAYAL	25.00	31.02	28.63	0.0265	0.0076	0.0076	0.0075	8
Puno										
512	210103	AMANTANI	75.00	67.73	63.88	0.0615	0.0113	0.0116	0.0114	4
513	210105	CAPACHICA		67.05	63.80		0.0125			
514	210107	COATA		54.99	52.97		0.0104			
515	210108	HUATA	37.50	50.80	48.36	0.0325	0.0081	0.0083	0.0082	8
516	210109	MAÑAZO	42.86	48.86	46.04	0.0401	0.0083	0.0084	0.0083	7
517	210110	PAUCARCOLLA		45.47	43.72		0.0103			
518	210111	PICHACANI		56.72	54.68		0.0105			
519	210113	SAN ANTONIO		51.41	49.04		0.0122			
520	210114	TIQUILLACA		58.86	56.47		0.0106			
521	210202	ACHAYA		53.17	51.99		0.0111			
522	210205	CAMINACA		50.74	48.08		0.0103			
523	210207	JOSE DOMINGO CHOQUEHUANCA	40.00	46.00	42.76	0.0592	0.0088	0.0090	0.0090	5
524	210208	MUÑANI		53.29	51.18		0.0104			
525	210209	POTONI	25.00	45.81	42.71	0.0618	0.0100	0.0100	0.0099	4
526	210210	SAMAN		50.54	47.77		0.0104			
527	210212	SAN JOSE		47.15	44.84		0.0104			
528	210213	SAN JUAN DE SALINAS		51.52	49.10		0.0102			
529	210215	TIRAPATA		47.31	45.29		0.0102			
530	210303	AYAPATA		46.88	46.16		0.0105			
531	210304	COASA	50.00	44.16	40.99	0.0827	0.0094	0.0096	0.0095	4
532	210305	CORANI		54.43	51.91		0.0103			
533	210307	ITUATA	44.44	56.90	55.42	0.0303	0.0085	0.0087	0.0086	9
534	210308	OLLACHEA		56.62	56.21		0.0105			

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
535	210403	HUACULLANI	40.00	45.00	42.54	0.0594	0.0096	0.0099	0.0098	5
536	210407	ZEPITA	50.00	48.41	45.36	0.0831	0.0092	0.0093	0.0092	5
537	210502	CAPAZO		47.76	46.11		0.0126			3
538	210504	SANTA ROSA		51.25	49.78		0.0110			
539	210602	COJATA		49.90	47.34		0.0119			
540	210604	INCHUPALLA		46.55	44.08		0.0103			
541	210606	ROSASPATA		43.50	40.82		0.0104			2
542	210608	VILQUE CHICO		50.33	47.42		0.0112			
543	210703	CALAPUJA		54.14	52.18		0.0106			9
544	210707	PARATIA	50.00	51.02	48.21	0.0344	0.0096	0.0097	0.0097	8
545	210802	ANTAUTA		53.29	51.83		0.0110			
546	210805	MACARI		42.88	41.14		0.0108			
547	210808	SANTA ROSA	42.86	47.56	44.90	0.0402	0.0083	0.0085	0.0083	7
548	210901	MOHO		50.00	47.91		0.0103			
549	210902	CONIMA		52.37	50.52		0.0105			
550	210903	HUAYRAPATA		48.73	46.12		0.0104			
551	210904	TILALI		54.08	52.28		0.0104			
552	211004	QUILCAPUNCU		51.44	49.52		0.0103			
553	211005	SINA		47.99	47.61		0.0103			
554	211202	CUYOCUYO	25.00	43.14	40.27	0.0619	0.0089	0.0091	0.0090	4
555	211204	PATAMBUCO		39.37	37.29		0.0114			
556	211206	QUIACA		50.89	51.18		0.0105			
557	211302	ANAPIA		46.48	45.33		0.0109			
558	211303	COPANI	20.00	45.87	43.26	0.0394	0.0085	0.0087	0.0086	5
San Martín										
559	220102	CALZADA	29.17	29.48	28.41	0.0085	0.0047	0.0047	0.0047	24
560	220105	SORITOR	25.00	30.62	29.71	0.0267	0.0075	0.0074	0.0074	9
561	220106	YANTALO		26.89	26.16		0.0102			
562	220201	BELLAVISTA	14.29	17.30	16.36	0.0203	0.0070	0.0071	0.0069	7
563	220202	ALTO BIAVO	35.71	39.35	37.54	0.0174	0.0065	0.0064	0.0064	14
564	220203	BAJO BIAVO	61.54	51.89	50.48	0.0196	0.0068	0.0068	0.0068	13
565	220302	AGUA BLANCA		33.63	33.17		0.0104			
566	220304	SANTA ROSA		40.89	40.07		0.0103			
567	220305	SHATOJA	25.00	33.17	32.28	0.0164	0.0064	0.0064	0.0064	12
568	220401	SAPOSOA		28.41	26.95		0.0105			
569	220403	EL ESLABON		33.50	31.73		0.0105			7
570	220404	PISCOYACU	9.09	19.73	20.18	0.0081	0.0046	0.0046	0.0045	11

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
571	220501	LAMAS	18.92	20.60	20.02	0.0041	0.0030	0.0030	0.0030	37
572	220503	BARRANQUITA		35.94	33.96		0.0107			8
573	220506	PINTO RECODO		42.32	40.43		0.0102			
574	220507	RUMISAPA		23.97	21.53		0.0108			
575	220508	SAN ROQUE DE CUMBAZA		31.45	29.24		0.0106			
576	220510	TABALOSOS	29.41	33.44	31.42	0.0128	0.0058	0.0058	0.0059	17
577	220511	ZAPATERO	18.18	27.08	25.83	0.0146	0.0062	0.0061	0.0062	11
578	220603	HUICUNGO		33.04	31.34		0.0102			
579	220703	CASPISAPA		19.36	17.90		0.0108			
580	220705	PUCACACA		24.03	23.80		0.0105			
581	220706	SAN CRISTOBAL	45.45	28.98	26.75	0.0227	0.0073	0.0073	0.0074	11
582	220708	SHAMBOYACU	81.82	62.93	63.31	0.0147	0.0061	0.0062	0.0062	11
583	220709	TINGO DE PONASA		31.20	30.18		0.0104			6
584	220801	RIOJA	16.13	18.61	19.58	0.0045	0.0031	0.0032	0.0032	31
585	220802	AWAJUN	54.55	45.64	46.81	0.0245	0.0072	0.0073	0.0072	11
586	220803	ELIAS SOPLIN VARGAS	44.44	36.65	37.78	0.0307	0.0078	0.0078	0.0078	11
587	220804	NUEVA CAJAMARCA	26.92	27.50	28.80	0.0038	0.0028	0.0028	0.0028	52
588	220806	POSIC		41.27	40.50		0.0100			
589	220807	SAN FERNANDO		33.98	34.01		0.0101			9
590	220808	YORONGOS	50.00	36.03	36.68	0.0345	0.0081	0.0081	0.0081	8
591	220809	YURACYACU	16.67	22.62	24.27	0.0273	0.0077	0.0078	0.0076	6
592	220901	TARAPOTO	10.87	11.22	11.08	0.0010	0.0010	0.0010	0.0010	93
593	220903	CACATACHI		16.73	15.50		0.0104			
594	220904	CHAZUTA	44.44	39.17	38.98	0.0143	0.0061	0.0061	0.0061	18
595	220905	CHIPURANA		38.89	36.09		0.0105			
596	220906	EL PORVENIR		33.87	32.85		0.0107			
597	220907	HUIMBAYOC	40.00	43.49	42.88	0.0263	0.0077	0.0078	0.0076	10
598	220909	LA BANDA DE SHILCAYO	25.93	21.17	20.33	0.0073	0.0044	0.0044	0.0044	28
599	220910	MORALES	10.81	12.16	11.96	0.0026	0.0021	0.0021	0.0021	37
600	220911	PAPAPLAYA		46.07	44.37		0.0102			
601	220912	SAN ANTONIO		24.59	23.08		0.0106			
602	220914	SHAPAJA	27.27	27.26	26.88	0.0185	0.0069	0.0069	0.0068	12
603	221005	UCHIZA		32.79	30.51		0.0101			22
Tacna										
604	230406	SUSAPAYA		46.10	45.77		0.0108			
Ucayali										
605	250102	CAMPOVERDE	41.18	36.48	35.53	0.0072	0.0043	0.0043	0.0042	34

Cuadro 16 Perú: Estimaciones

N°	Ubigeo	Nombre	Dir.	FH	FH esp.	ECM-Dir.	ECM-FH	ECM param.	ECM no param.	n
606	250104	MASISEA	75.00	70.28	69.09	0.0169	0.0073	0.0072	0.0071	12
607	250105	YARINACOCHA	32.56	31.21	30.95	0.0010	0.0009	0.0009	0.0009	216
608	250106	NUEVA REQUENA		34.16	32.99		0.0102			
609	250202	SEPAHUA		49.39	48.44		0.0103			
610	250203	TAHUANIA	66.67	64.89	63.92	0.0157	0.0067	0.0066	0.0065	15
611	250204	YURUA		77.33	75.30		0.0136			
612	250302	IRAZOLA	18.18	22.40	22.32	0.0026	0.0021	0.0021	0.0021	55
613	250303	CURIMANA	18.75	27.92	28.10	0.0100	0.0051	0.0051	0.0050	16
614	250304	NESHUYA		28.45	26.98		0.0101			
615	250305	ALEXANDER VON HUMBOLDT		36.53	36.28		0.0101			
616	250401	PURUS	42.86	54.69	54.64	0.0181	0.0069	0.0069	0.0068	14

