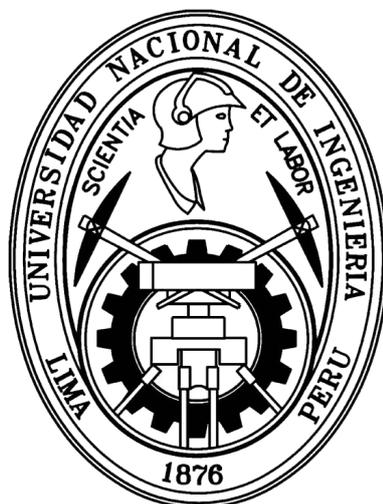


UNIVERSIDAD NACIONAL DE INGENIERIA

FACULTAD DE INGENIERIA ECONOMICA, ESTADISTICAS Y CIENCIAS SOCIALES



TESIS

**“DETERMINACION DEL MEJOR ESTIMADOR DE LA POBREZA
MONETARIA ENTRE CINCO METODOS DE AREA PEQUEÑA (SAE)
MEDIANTE EXPERIMENTOS DE MONTECARLO Y REMUESTREO DE
JACKKNIFE Y BOOSTRAP EN EL PERIODO 2012-2013”**

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO ESTADÍSTICO

ELABORADO POR:

MICHEL STEVE SUAREZ INOCENTE

ASESORA

MAG. MAGEN INFANTE ROJAS

LIMA – PERÚ

2016

RESUMEN

El propósito de este trabajo es estimar y comparar la precisión del indicador de pobreza monetaria en base a 5 metodologías de áreas menores o también conocido como SAE (Small Area Estimation) en los 24 departamentos y en la Provincia Constitucional del Callao para el periodo de años 2012 y 2013, utilizando los datos de la Encuesta Nacional de Hogares 2012- 2013 (ENAH0 2012-2013) y del Empadronamiento Distrital de Población y Vivienda 2012-2013 (SISFOH 2012-2013), estos dos instrumentos se pueden descargar de la web del Instituto Nacional de Estadística e Informática (INEI) en el módulo de microdatos (INEI, Microdatos INEI).

El indicador de pobreza monetaria se evaluó en base a 500 simulaciones de muestras tipo ENAH0 2012-2013 por experimentos de Montecarlo obtenidas a partir de la base censal SISFOH 2012-2013 y para las 5 metodologías SAE. Posteriormente se evaluaron los indicadores Sesgo Relativo y el Error Cuadrático Relativo para estimar la precisión relativa de las diferentes metodologías, los cuales permitieron conocer que la metodología SAE de Estimador Sintético posee mejores características de un estimador preciso.

Además se presentó para la metodología SAE con mejor característica de precisión relativa 2 formas para calcular su coeficiente de variación mediante remuestreo, utilizando Jackknife y Bootstrap, posteriormente se le comparó con el método estándar y el resultado fue que el estimador sintético de variable edad-sexo es el que posee mejores características de un estimador preciso para la pobreza monetaria.

Los métodos en cuestión permitieron calcular estimados de pobreza monetaria con mejores indicadores de precisión para los 24 departamentos y la Provincia Constitucional del Callao en el periodo de años 2012-2013.

Palabras claves: pobreza monetaria, precisión, sesgo, varianza, simulación, Montecarlo, sae, áreas menores, remuestreo

INDICE GENERAL

RESUMEN	1
INDICE GENERAL	2
INDICE DE CUADROS	4
INDICE DE GRÁFICOS	5
INTRODUCCIÓN	6
CAPITULO I	7
REALIDAD PROBLEMÁTICA	7
1.1 DESCRIPCIÓN DEL CONTEXTO DEL PROBLEMA.....	7
1.2 FORMULACIÓN DEL PROBLEMA GENERAL Y ESPECIFICO	14
1.3 FORMULACIÓN DE LOS OBJETIVOS, GENERAL Y ESPECÍFICOS	14
1.4 JUSTIFICACIÓN, IMPORTANCIA, ALCANCES Y LIMITACIONES DE LA INVESTIGACIÓN	15
CAPITULO II	16
MARCO TEORICO	16
2.1 ANTECEDENTES	16
2.2 BASES TEÓRICAS: GENERALES Y ESPECIALIZADAS	22
2.2.1 EXPERIMENTOS DE MONTECARLO.....	22
2.2.2 INTRODUCCION A LA ESTIMACION EN AREAS PEQUEÑAS (SAE)	25
ESTIMADORES DIRECTOS: ESTIMADOR DE HORVITZ – THOMPSON	27
ESTIMADORES DIRECTOS: ESTIMADORES POSTESTRATIFICADO.....	29
ESTIMADORES INDIRECTOS: ESTIMADOR SINTETICO	30
ESTIMADOR INDIRECTO: ESTIMADOR COMPUESTO	31
ESTIMADOR BASADO EN UN MODELO LOGIT	33
2.2.3 INDICADORES DE EVALUACION	34
SESGO RELATIVO (SRAM)	34
ERROR CUADRÁTICO RELATIVO (RECMRM).....	36
2.2.4 REMUESTREO	38
REMUESTREO POR JACKKNIFE.....	38
REMUESTREO POR BOOTSTRAP	44
2.3 FORMULACIÓN DE HIPÓTESIS.....	48
2.4 VARIABLES Y OPERACIONALIZACIÓN.....	49
CAPITULO III	51

MARCO METODOLOGICO	51
3.1 TIPO, NIVEL Y DISEÑO DE INVESTIGACIÓN.....	51
3.2 POBLACIÓN, MUESTRA Y TAMAÑO DE MUESTRA.....	51
3.3 TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS.....	60
3.4 TÉCNICAS DE ANÁLISIS E INTERPRETACIÓN DE DATOS.....	64
CAPITULO IV.....	65
RESULTADOS.....	65
CAPITULO V.....	93
CONCLUSIONES Y RECOMENDACIONES	93
REFERENCIAS BIBLIOGRÁFICAS.....	94

INDICE DE CUADROS

Cuadro 1: Resultados del indicador SRAM a nivel nacional y SRA a nivel departamental.....	35
Cuadro 2: Resultados del indicador RECMRM a nivel nacional y RECMR a nivel departamental.....	37
Cuadro 3: Operacionalización para las variables de ENAHO	49
Cuadro 4: Operacionalización de Variables para el SISFOH.....	50
Cuadro 5: Población, Viviendas y Hogares por departamentos en la base del SISFOH	52
Cuadro 6: Población, Viviendas y Hogares con Información de Gastos por departamentos en la base del SISFOH	53
Cuadro 7: Población, Viviendas y Hogares por departamentos en la base de ENAHO.....	54
Cuadro 8: Población, Viviendas y Hogares con Información de Gastos por departamentos en la base de ENAHO 2012-2013.....	55
Cuadro 9: Total de conglomerados y hogares por departamento en la base del SISFOH	56
Cuadro 10: Total de conglomerados y hogares por departamento en la base de ENAHO	57
Cuadro 11: Cantidad de viviendas a extraer por conglomerado para la simulación de la encuesta	58
Cuadro 12: Cantidad de hogares seleccionados para cada simulación de muestra por departamento	59
Cuadro 13: Pobreza Monetaria por departamentos con la base ENAHO 2012-2013.....	67
Cuadro 14: Pobreza Monetaria por departamentos con base del SISFOH 2012-2013	68
Cuadro 15: Tipos de estimadores SAE analizados	70
Cuadro 16: Indicadores de Evaluación del Estimador EST 1 por departamentos y total.....	71
Cuadro 17: Indicadores de Evaluación del Estimador EST 2 por departamentos y total.....	72
Cuadro 18: Indicadores de Evaluación del Estimador EST 3 por departamentos y total.....	73
Cuadro 19: Indicadores de Evaluación del Estimador EST 4 por departamentos y total.....	74
Cuadro 20: Indicadores de Evaluación del Estimador EST 5 por departamentos y total.....	75
Cuadro 21: Indicadores de Evaluación del Estimador EST 6 por departamentos y total.....	76
Cuadro 22: Indicadores de Evaluación del Estimador EST 7 por departamentos y total.....	77
Cuadro 23: Indicadores de Evaluación del Estimador EST 8 por departamentos y total.....	78
Cuadro 24: Indicadores de Evaluación del Estimador EST 9 por departamentos y total.....	79
Cuadro 25: Indicadores de Evaluación del Estimador EST 10 por departamentos y total.....	80
Cuadro 26: Análisis descriptivo del Sesgo Relativo por Tipo de Estimador	81
Cuadro 27: Estimados de Pobreza Monetaria con las metodologías Directa y Sintética por Departamentos. 83	
Cuadro 28: Pobreza Monetaria, Errores Típicos y Coeficiente de Variación por departamento.....	85
Cuadro 29: Pobreza Monetaria, Error Típico Jackknife y Coeficiente de Variación del Estimador Sintético EST 4 por departamentos.....	86
Cuadro 30: Pobreza Monetaria, Error Típico Bootstrap y Coeficiente de Variación del Estimador Sintético EST 4 por departamentos.....	87
Cuadro 31: Pobreza Monetaria, Error Típico Jackknife y Coeficiente de Variación del Estimador Sintético EST 5 por departamentos.....	88
Cuadro 32: Pobreza Monetaria, Error Típico Bootstrap y Coeficiente de Variación del Estimador Sintético EST 5 por departamentos.....	89
Cuadro 33: Coeficientes de Variación de los Estimadores Directos y Sintéticos por departamentos	90

INDICE DE GRÁFICOS

Gráfico 1: Importancia de la Focalización.....	7
Gráfico 2: Definición INEI de Pobre Monetario.....	8
Gráfico 3: Evolución de la Pobreza Monetaria 2014.....	10
Gráfico 4: Mapa de Pobreza Provincial y Distrital 2013.....	10
Gráfico 5: Coeficiente de Variación a nivel de Departamentos.....	11
Gráfico 6: Coeficiente de Variación a nivel de Grupos de Departamentos.....	12
Gráfico 7: Sesgo y Varianza en Modelos de Estimación.....	13
Gráfico 8: Experimentos de Montecarlo con Simulación de Muestras.....	24
Gráfico 9: Explicación de la lógica tras el Jackknife original.....	41
Gráfico 10: Resumen del remuestreo por Jackknife.....	43
Gráfico 11: Resumen del remuestreo por bootstrap.....	46
Gráfico 12: Ficha Socioeconómica Unica del SISFOH - Carátula.....	60
Gráfico 13: Ficha de Características de la vivienda del SISFOH.....	61
Gráfico 14: Ficha de Características de la Población del SISFOH.....	61
Gráfico 15: Carátula de la Encuesta Nacional de Hogares ENAHO.....	62
Gráfico 16: Ficha de Características de la Vivienda y del Hogar de ENAHO.....	63
Gráfico 17: Ficha de Características de los Miembros del Hogar.....	63
Gráfico 18: Resumen del Procedimiento para el Análisis de Datos.....	64
Gráfico 19: Convergencia en Población Total.....	65
Gráfico 20: Convergencia en Desviación Poblacional Total.....	66
Gráfico 21: Intervalo de Confianza al 95% de la Pobreza Monetaria de la ENAHO y SISFOH por departamentos.....	69
Gráfico 22: Resultados de los indicadores SRAM, RECMRM por tipo de estimador para la Población Total.....	82
Gráfico 23: Estimados de Pobreza Monetaria por departamentos.....	84
Gráfico 24: Coeficientes de Variación de los estimadores Directos y Sintéticos por departamentos.....	91
Gráfico 25: Coeficientes de Variación de Distritos, Provincias y Departamentos.....	92

INTRODUCCIÓN

La presente tesis es una investigación que tiene por objetivo una revisión y comparación de metodologías de estimación en áreas menores (SAE) para estimar de manera más precisa y con metodologías estadísticas actuales el indicador de pobreza monetaria en el país en el periodo de tiempo 2012-2013.

La Encuesta Nacional de Hogares (ENAH) es el instrumento principal para obtener estimaciones de la pobreza monetaria en el Perú. Actualmente el Instituto Nacional de Estadística e Informática (INEI) publica estimaciones de pobreza monetaria a nivel de grupos de departamentos (INEI, 2015), ya que dicha encuesta no tiene inferencia en áreas de menores niveles geográficos. Es sabido también que actualmente el Perú vive un proceso de descentralización y por ende demanda una información sociodemográfica más precisa, a razón de ello por ejemplo el INEI produce los documentos de Mapas de Pobreza (INEI C. C., 2015) en donde se obtiene estimaciones de pobreza monetaria hasta nivel de distritos por demanda de los organismos del estado y programas sociales.

Este trabajo presenta los siguientes capítulos:

En el capítulo I se presenta la realidad problemática, en donde se especifican los problemas, objetivos, justificaciones, alcances y limitaciones sobre la estimación del indicador de pobreza monetaria en áreas menores.

En el capítulo II se aborda el marco teórico, donde se presentan los antecedentes de investigaciones que iniciaron las metodologías de áreas menores, así como también las bases teóricas que permitieron elaborar los resultados, por último se presentan las hipótesis de investigación y la operacionalización de todas las variables involucradas en la presente investigación.

En el capítulo III se aborda el marco metodológico, cuya estructura menciona el tipo, nivel y diseño de investigación, luego se desarrolla la población, muestra y tamaño de muestra y por último las técnicas de recolección de datos y de análisis e interpretación de datos.

En el capítulo IV se ofrece los resultados de la investigación en forma de cuadros, gráficos y sus respectivos comentarios.

En el capítulo V se presentan las conclusiones y recomendaciones a las cuales se llega en la presente tesis.

CAPITULO I

REALIDAD PROBLEMÁTICA

1.1 Descripción del contexto del problema

En los países de América Latina se observa en la actualidad interés, cada vez mayor, por contar con información geográfica más precisa sobre los aspectos sociales, económicos y demográficos, tal es el caso de la pobreza, necesaria para la programación y/o ejecución de las estrategias de desarrollo, basadas en la focalización espacial del gasto público y de las inversiones en los sectores sociales (ODM, 2005).

Gráfico 1: Importancia de la Focalización



Fuente: Propia

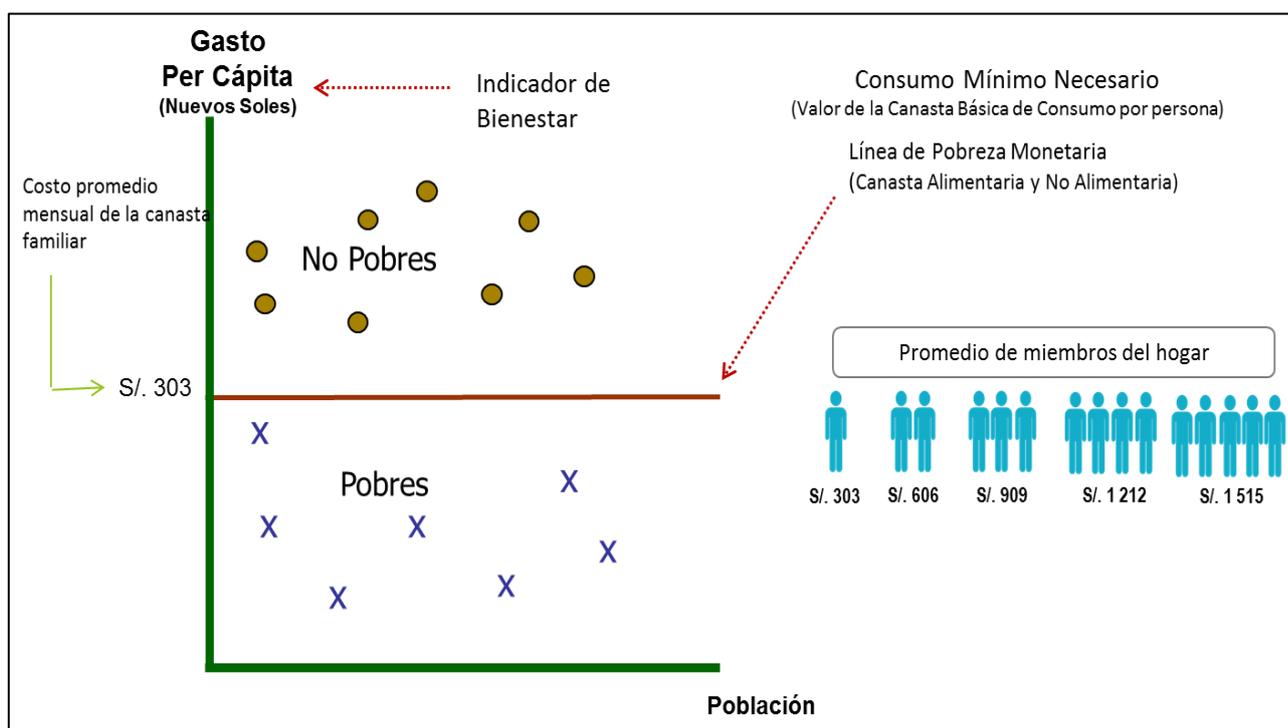
Elaboración: Propia

¿Qué se debe entender por Pobreza Monetaria?

La definición de pobreza monetaria que se dará en la presente investigación es la del INEI:

“Se considera pobre monetario a todas las personas residentes en hogares particulares, cuyo gasto per cápita valorizado monetariamente, no supera el umbral de la línea de pobreza”

Gráfico 2: Definición INEI de Pobre Monetario



Fuente: Presentación de pobreza monetaria del INEI

Elaboración: INEI

¿Es necesaria la información para áreas desagregadas geográficamente?

Específicamente, en el caso peruano, el gobierno actual ha declarado al "proceso de descentralización" como la acción prioritaria dentro de la política de desarrollo y combate a la pobreza (21, 2014). Desde esta óptica, las políticas y programas destinados a atender las necesidades básicas de nuestra población, cuya principal característica es su composición sociodemográfica heterogénea con grandes desequilibrios regionales, dará mayor responsabilidad a los gobiernos regionales y

locales. Esto propiciará una mayor demanda de información para las divisiones político-administrativas menores como las provincias, distritos y aún centros poblados, con la finalidad de asignar adecuadamente los recursos provenientes del presupuesto de la nación. Sin embargo, un problema fundamental en la programación de este tipo de actividades y la respectiva estimación de los recursos en niveles geográficos desagregados como los mencionados, es la carencia de buena información.

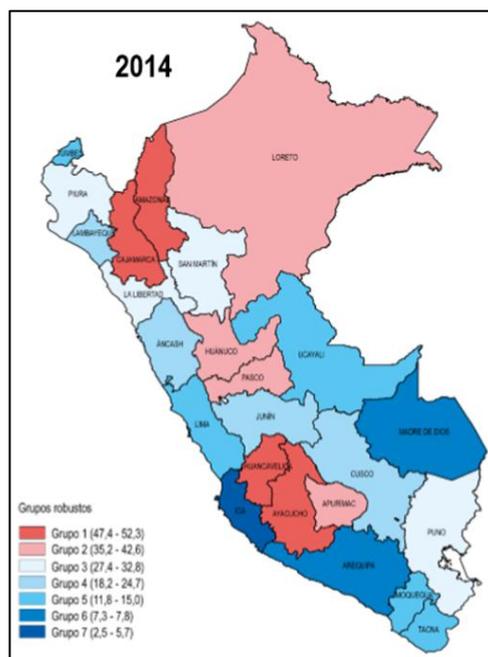
¿Hay información en el país que permita estimaciones de pobreza?

Las fuentes de datos que tradicionalmente brindan información para áreas pequeñas son el Registro Civil, los censos de población y los registros sectoriales. El primero facilita información sobre los hechos vitales; los Censos de Población y Vivienda dan información sobre las características generales de la población hasta el nivel de distritos; y los registros sectoriales como los de Salud, Educación y otros recaban información sobre la cobertura de sus propios programas, dejando de lado las realizadas por el sector privado. Hay que anotar que los registros sectoriales adolecen de una falta de publicación sistematizada de sus datos en forma oportuna, lo cual disminuye las posibilidades de su utilización. Asimismo, estos registros tienen como finalidad aspectos programáticos más que estadísticos.

Una fuente principal de información en temas de pobreza la constituye la encuesta por muestreo ENAHO realizadas por el INEI. Sin embargo, problemas técnicos y financieros limitan el nivel de injerencia a grandes regiones, por lo que no es factible obtener directamente datos desagregados para áreas pequeñas, como lo exige la programación, ejecución y evaluación de los servicios descentralizados.

Uno de los documentos más importante que el INEI anualmente elabora sobre temas de pobreza es el de “Evolución de la Pobreza Monetaria” (ver gráfico 3) en donde las estimaciones de pobreza monetaria se calculan tomando en cuenta el diseño de la encuesta. Otro documento especializado de pobreza es el llamado “Mapas de Pobreza” (ver gráfico 4), el último por ejemplo es el documento “Mapa de pobreza provincial y distrital 2013” que el INEI presentó en setiembre del 2015 y en donde se presentan los estimados de pobreza monetaria a nivel provincial y distrital.

Gráfico 3: Evolución de la Pobreza Monetaria 2014

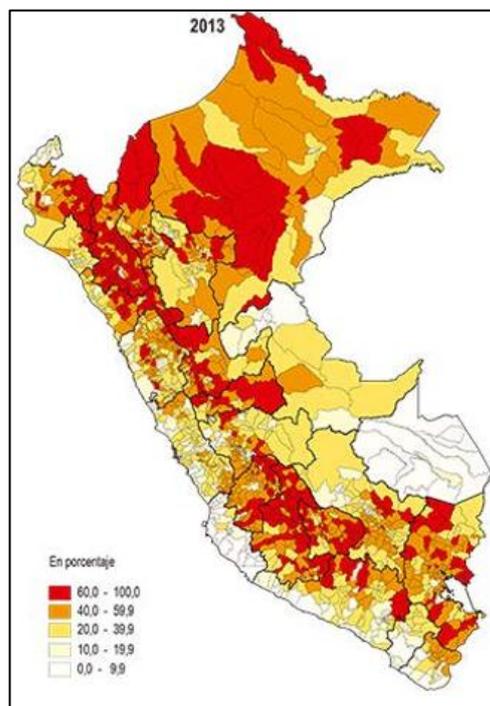


Inferencia a grandes regiones!!!!

Fuente: INEI

Elaboración: INEI

Gráfico 4: Mapa de Pobreza Provincial y Distrital 2013



Necesidad de Información censal continua!!!!

Fuente: INEI

Elaboración: INEI

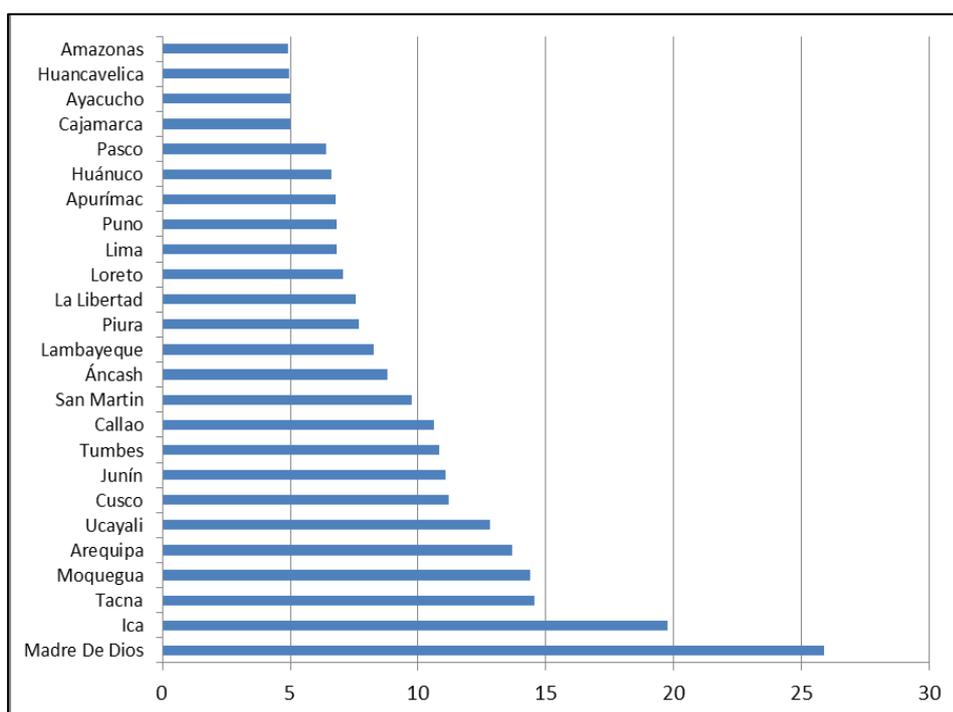
¿Cómo se evalúa la precisión de una estimación de pobreza?

Coeficiente de Variación (CV) permite al usuario evaluar la precisión del estimador en términos relativos y comparar niveles de precisión para estimaciones de diferentes parámetros en diferentes poblaciones

Según el INEI un coeficiente de variación (CV) menor al 5% se dan para precisiones muy buenas, CV del 5% al 10% para precisiones buenas, de 10% a 15% para precisiones aceptables y CV del más del 15% se darán para precisiones referenciales o no confiables.

Por ejemplo si analizásemos las gráficas 5 y 6 y teniendo en cuenta el CV, se decidiría por presentar las estimaciones de pobreza monetaria a nivel de grupos de departamentos pues ahí presentan un mayor número de departamentos con precisión muy buena.

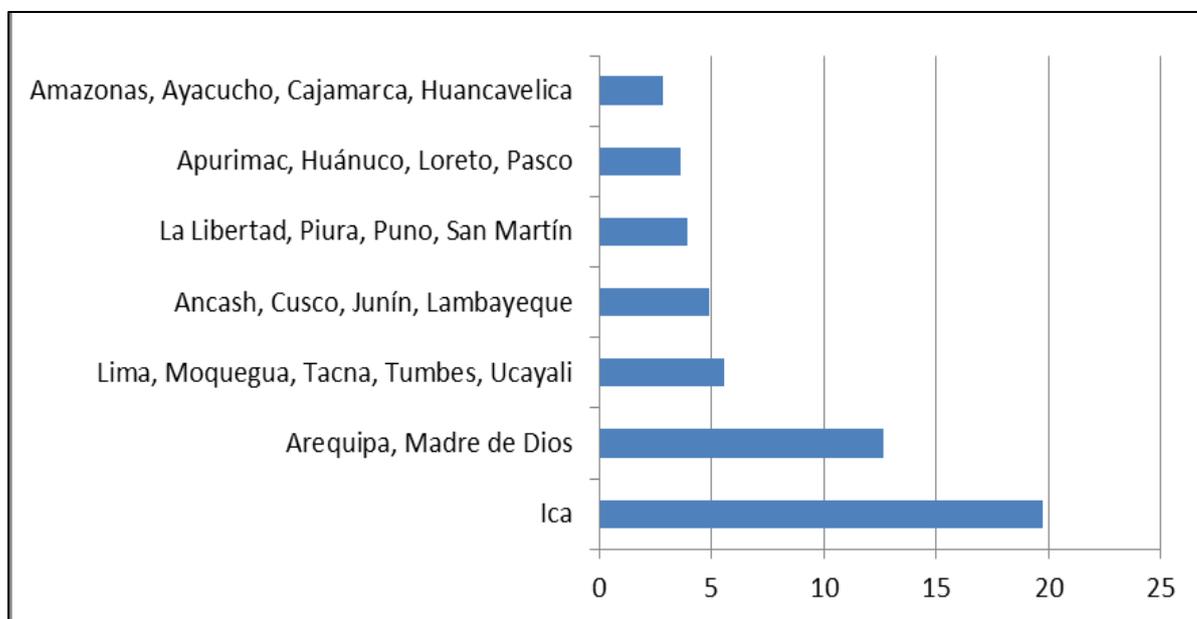
Gráfico 5: Coeficiente de Variación a nivel de Departamentos



Fuente: ENAHO 2014

Elaboración: Propia

Gráfico 6: Coeficiente de Variación a nivel de Grupos de Departamentos



Fuente: ENAHO 2014

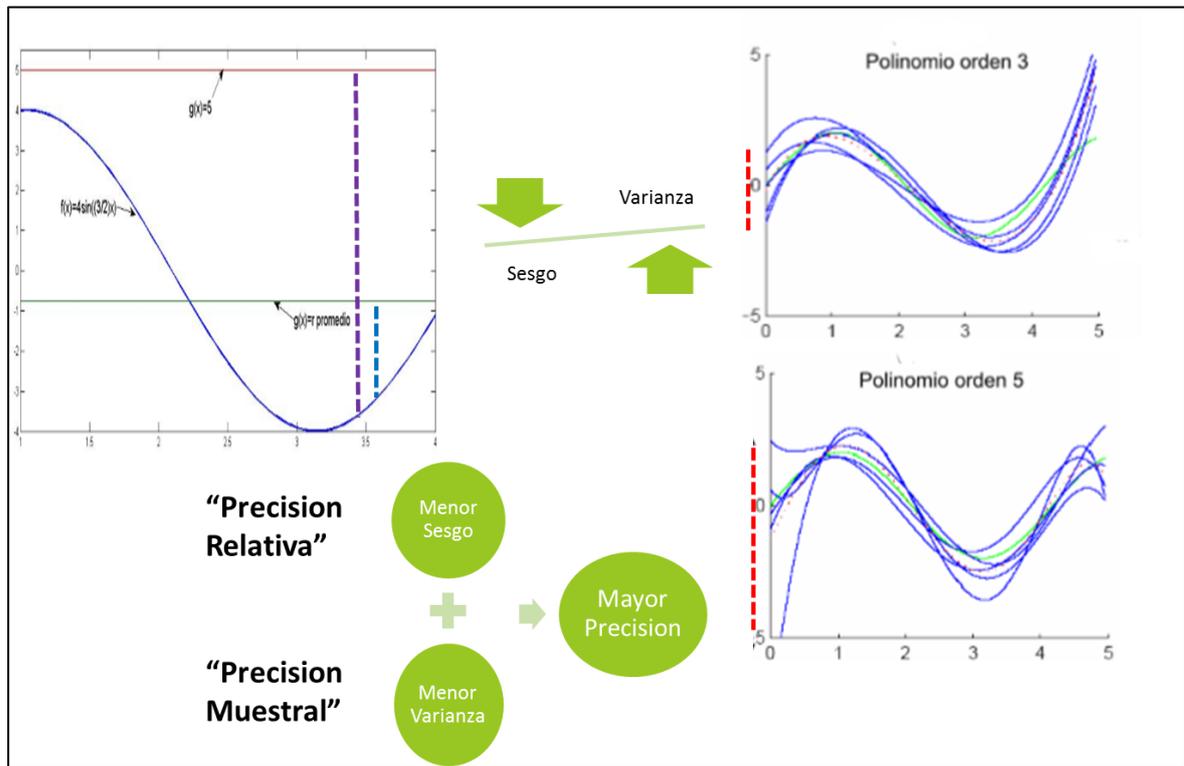
Elaboración: Propia

¿Qué se necesita para obtener buenas estimaciones de pobreza?

Todos los métodos de estimación de pobreza tienen en el fondo un modelo del cual se ayudan para obtener la respectiva estimación. Ahora en el aprendizaje de maquina se busca obtener modelos que ofrezcan dos características esenciales: ajuste de datos y generalización, el primero de ellos tiene que ver con el sesgo, es decir, con la diferencia entre el valor estimado y el valor real, en otras palabras cuanto menor sesgo obtengamos mejor será nuestro modelo pues se estaría prediciendo valores cercanos al verdadero valor. El segundo está relacionado con la varianza, es decir, la variabilidad o complejidad de los resultados y cuanto menor sea dicha varianza mejores estimaciones de pobreza monetaria obtenemos.

Como se presenta en la gráfica 7 una mayor precisión necesita un menor sesgo y menor varianza para ello se necesita analizar conjuntamente la Precisión Relativa y la Precisión Muestral.

Gráfico 7: Sesgo y Varianza en Modelos de Estimación



Fuente: Propia

Elaboración: Propia

Como síntesis del problema se puede expresar que hay demanda creciente por información más precisa sobre la pobreza monetaria, y ésta puede mejorar significativamente si se implementan nuevas formas de estimación como son las Estimaciones en Áreas Menores o Small Area Estimation (SAE).

1.2 Formulación del problema general y específico

Problema General

- Es necesario obtener el mejor estimador para la pobreza monetaria entre los cinco métodos de Área Pequeña (SAE) mediante experimentos de Montecarlo y Remuestreo de Jackknife y Bootstrap

Problemas Específicos

- Es necesario obtener el estimador que posea la mejor precisión relativa en la pobreza monetaria utilizando los cinco métodos de "área pequeña" (SAE)
- Es necesario obtener el estimador de mayor precisión muestral de la pobreza monetaria entre el método del mejor estimador de área pequeña (SAE) y el método estándar vigente

1.3 Formulación de los Objetivos, General y Específicos

Objetivo General

- Determinar el mejor estimador con cinco métodos de área pequeña (SAE) utilizando experimentos de Montecarlo para estimar el indicador de pobreza monetaria y compararlo con el método estándar utilizando remuestreo de Jackknife y Bootstrap.

Objetivos Específicos

- Determinar el mejor estimador entre cinco métodos de área pequeña (SAE), para obtener la mejor precisión relativa, utilizando experimentos de Montecarlo para estimar el indicador de pobreza monetaria
- Determinar el estimador de pobreza monetaria de mejor precisión muestral comparando el "mejor estimador de área pequeña" (SAE) con el método estándar vigente, utilizando remuestreo de Jackknife y Bootstrap

1.4 Justificación, Importancia y limitaciones de la investigación

La justificación: El presente trabajo servirá para obtener mejor precisión del indicador de pobreza monetaria al comparar diferentes metodologías SAE y al haber usado técnicas estadísticas de Experimentos de Montecarlo y técnicas de remuestreo como el Bootstrap y Jackknife.

La Importancia: El presente trabajo sentará las bases a seguir para que las diferentes instituciones que no solo manejen estimaciones de pobreza monetaria puedan implementar y así poder mejorar la precisión de diferentes estimaciones.

Limitación: La metodología pierde soporte de precisión para áreas muy desagregadas es por ello que se plantea estimar los indicadores a nivel de departamentos y los resultados son válidos para el periodo de años 2012 y 2013.

CAPITULO II

MARCO TEORICO

2.1 Antecedentes

Los métodos SAE (Small Area Estimation) se hicieron populares con los trabajos de Fay and Herriot (1979), Battese, Harter y Fuller (1988) y Ghosh y Rao (1994), y se presenta un resumen de dichas investigaciones a continuación:

a) Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data.

(Fay & Herriot, 1979)

En este artículo presentó una adaptación del estimador de James-Stein (estimador de promedio sesgado) para obtener estimaciones en la muestra de ingresos para áreas pequeñas (con población menor de 1 000 habitantes por área) del censo de población y vivienda 1970. La adaptación incorpora un modelo de regresión en el contexto de varianzas desiguales. Se presenta evidencia de que las estimaciones resultantes tienen un menor error promedio que la estimación en la muestra o algún otro procedimiento alternativo. Las nuevas estimaciones para estas áreas pequeñas ahora forman la base para la estimación actualizada del ingreso per cápita en la oficina del censo del Programa General de la repartición en Estados Unidos.

b) An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data.

(Battese, Harter, & Fuller, 1988)

El conocimiento de la zona con diferentes cultivos es importante para el Departamento de agricultura de Estados Unidos. Encuestas por muestreo han sido diseñadas para estimar áreas de grandes regiones, como informes de cultivo de los distritos, Estados y los Estados Unidos como un todo. Predicciones de cosecha para áreas pequeñas tales como condados no se ha intentado debido a la falta de información disponible de encuestas agrícolas de estas áreas. El uso de datos satelitales en

asociación con las observaciones de la encuesta de nivel de la finca ha sido objeto de una investigación en el referido artículo y que considera lo siguiente:

(a) Datos en 12 condados de Iowa, obtenidos de la encuesta enumerativa de junio de 1978 del Departamento de agricultura de Estados Unidos

(b) Datos obtenidos de satélites del Observatorio de tierra (LANDSAT) durante la temporada de crecimiento de 1978.

Se hace hincapié para predecir el área de maíz y soya en estos condados. Se especifica un modelo de regresión lineal para la relación entre hectáreas de maíz y soya en segmentos de muestra en la encuesta de junio enumerativa y la correspondiente determinación de satélite para las áreas con maíz y soya. Un estimador de la varianza del error fue construido, incluyendo las condiciones derivadas de la estimación de los parámetros del modelo. Se presentan las predicciones del promedio de hectáreas de maíz y la soya por el segmento de los 12 condados de Iowa. Los errores estándar de las predicciones son comparados con los de predictores tradicionales. El predictor sugerido para el área de Condado de cultivo promedio por segmento tiene un error de estándar considerablemente menor que el predictor de regresión tradicional en la encuesta.

**c) Small Area Estimation: An Appraisal.
(Ghosh & Rao, 1994)**

Las estimaciones en áreas pequeñas están convirtiéndose en importantes técnicas en el muestreo de encuestas debido a una creciente demanda de estadísticas confiables para áreas pequeñas de los sectores público y privado. Ahora es ampliamente reconocido que en la encuesta las estimaciones directas para áreas pequeñas son propensas a errores inaceptablemente grandes debido a la pequeñez de tamaño de las muestras en las áreas de producción. Esto hace necesario "pedir prestado la fuerza" de áreas relacionadas para encontrar estimaciones más precisas de una zona determinada o, simultáneamente, a varias áreas. Esto ha llevado al desarrollo de métodos alternativos tales como sintético, muestra tamaño dependiente, mejor estimador lineal empírico, Bayes empírico y estimación Bayes jerárquico. El presente artículo es en gran parte una valoración de algunos de estos métodos. También se evalúa el desempeño de estos métodos, utilizando algunos datos sintéticos que se asemejan a una población de negocios. Los estimadores "mejor estimador lineal empírico" como "Bayes empírico" y "jerárquico", para la mayoría de los casos, parecen tener una clara ventaja sobre otros métodos.

Hasta el momento se ha expuesto la importancia de la metodología de áreas menores o comúnmente llamadas SAE. A continuación se presentan 3 casos de investigaciones que utilizan el enfoque SAE para estimar el indicador de pobreza y son los siguientes:

**a) Non – parametric bootstrap mean squared error estimation for M-quantile estimators of small area averages, quantiles and poverty indicators.
(Marchetti, Tzavidis, & Pratesi, 2012)**

En el citado artículo hace hincapié en la estimación de indicadores de pobreza y de cuantiles para áreas pequeñas con modelos robustos, por ejemplo, el modelo de área pequeña M-cuantil. En paralelo al punto de estimación, la estimación del Error de cuadrado medio (MSE) es una tarea igualmente crucial y desafiante. Sin embargo, mientras que la estimación analítica de MSE para promedios de área pequeña es posible, estimación de MSE analítica para indicadores de cuantiles y la pobreza es difícil. Además, una de las principales críticas del estimador MSE analítica para las estimaciones de M-quantile de promedios de área pequeña es que puede ser inestable cuando el área específica muestra tamaños pequeños. Un marco bootstrap no paramétrico para la estimación de la MSE del promedio de área pequeña, indicadores de cuantiles y la pobreza estimados con el M-cuantil se proponen. Se pone énfasis en propiedades de orden de estimadores MSE cuyos resultados sugieren que el estimador bootstrap del MSE es más estable que los estimadores MSE analíticas correspondientes. La investigación fue evaluada en una serie de estudios de simulación bajo diferentes supuestos paramétricos para los términos de error de modelo y diferentes escenarios de la zona y tamaños de la población. Por último, se presentan los resultados de la aplicación del estimador propuesto MSE a los datos de los ingresos reales del estudio europeo de ingresos y las condiciones de vida (EU-SILC) en Italia y se proporciona información sobre la disponibilidad de funciones de R que se puede utilizar para la aplicación de los procedimientos de estimación propuesta en la práctica.

**b) Small area estimation with spatio-temporal Fay- Herriot models.
(Marhuenda, Molina, & Morales, 2013)**

La estimación de área pequeña es estudiada bajo un modelo espacio temporal fay-herriot que es un montaje de modelo basado en máxima verosimilitud. Se propone un procedimiento de bootstrap paramétrico para la estimación del error cuadrado medio de los estimadores de área pequeña. El modelo espacio-temporal es comparado con modelos más simples a través de experimentos de

simulación, análisis de la ganancia en la eficiencia alcanzada por el uso del modelo más complejo. También se evalúa el desempeño del estimador bootstrap paramétrico del error cuadrado medio. Se lleva a cabo una aplicación con datos españoles EU-SILC para obtener estimaciones de indicadores de pobreza por provincias españolas en 2008, haciendo uso de datos de encuestas de años 2004-2008.

c) Multivariate Fay–Herriot models for small area estimation.

(Benavent & Morales, 2016)

Se introducen modelos multivariantes de Fay – Herriot para estimar indicadores de área pequeña. Entre los procedimientos disponibles para el montaje de modelos lineales mixtos, se emplea la probabilidad máxima residual (REML). El mejor predictor empírico (EBLUP) del vector de medios del área se deriva. Se da una aproximación a la matriz de los errores de predicción cruzado cuadrado medio (MSE) y se proponen cuatro estimadores MSE. El primer estimador MSE es una versión de la aproximación de MSE. Los estimadores MSE restantes combinan bootstrap paramétrico con los términos analíticos de la aproximación de MSE. Se realizan varios experimentos de simulación para evaluar el comportamiento de la EBLUP multivariado y para comparar los estimadores MSE. La metodología desarrollada y el área de software aplican a datos españoles del 2005 y 2006 para estudios de las condiciones de vida. El objetivo de la aplicación es la estimación de proporciones de la pobreza y las brechas a nivel de la provincia.

Una de las principales publicaciones en términos de pobreza y áreas pequeñas que la mayoría de oficinas estadísticas elabora son los denominados Mapas de Pobreza. A continuación se presenta 2 artículos al respecto:

**a) Poverty projection using a small area estimation method: Evidence from Vietman.
(Viet, 2011)**

Para el seguimiento y evaluación de la pobreza, se necesita las estimaciones de la pobreza en los niveles de desagregación diferente. La predicción de la tendencia de la pobreza es también de interés para los responsables políticos, así como investigadores. El artículo presenta un método que se basa en una estimación de área pequeña de Elbers (2003) a un mapa de la pobreza de proyecciones en el

futuro. Este método se aplica para proyectar un mapa de la pobreza en zonas rurales de Vietnam para el año 2008 usando datos rurales del 2006, censo agrícola y pesca y datos de encuestas de hogares de Vietnam del 2004 y 2006.

**b) Can census data alone signal heterogeneity in the estimation of poverty maps?
(Tarozi, 2011)**

Metodologías ahora comúnmente utilizadas para la construcción de mapas de pobreza asumen un alto grado de homogeneidad dentro de áreas geográficas en la relación entre el ingreso y sus predictores. Sin embargo, la mano de obra local y los mercados de arrendamiento y otras diferencias locales son susceptibles de generar heterogeneidad en tales relaciones, al menos hasta cierto punto. El propósito de este artículo es argumentar que útil sería si hay evidencia indirecta sobre el grado de heterogeneidad del área. Tal evidencia indirecta es proporcionada por indicadores no monetarios, tales como alfabetización, propiedad de activos o acceso a saneamiento, que habitualmente incluyen en censos. Estos indicadores pueden utilizarse para realizar ejercicios de validación para medir el grado de heterogeneidad en su distribución condicional sobre predictores análogas a las utilizadas en cartografía de la pobreza. Sostenemos que los mismos factores que son susceptibles de generar heterogeneidad del área de cartografía de la pobreza son también susceptibles de generar heterogeneidad en este tipo de ejercicios de validación. Construimos un modelo muy simple para ilustrar este punto formalmente. Por último, se evalúa empíricamente el argumento con datos de México. En nuestras ilustraciones empíricas, el funcionamiento de las metodologías de imputación para construir mapas de indicadores generalmente factibles con datos del censo solo es de tipo informativo ya que la eficacia de esas metodologías puede producir inferencia correcta en cartografía de la pobreza.

A continuación se presenta la documentación de pobreza y el último mapa de pobreza para el caso del Perú

Evolución de la pobreza monetaria 2009-2014

(Instituto Nacional de Estadística e Informática, 2015)

El Informe Técnico presenta la información de pobreza monetaria de la serie 2009 al 2014 calcula con la encuesta de hogares ENAHO, desagregada por área de residencia, región natural y dominio

geográfico, así como la evolución del gasto y del ingreso real y nominal, desagregado por principales rubros y por deciles. Se incluyen indicadores sobre la evolución del grado de la desigualdad del gasto y del ingreso. Asimismo, se presenta la evolución de las líneas de pobreza total y extrema y la incidencia de pobreza monetaria total y extrema. Se incluyen indicadores de brecha y severidad de la pobreza monetaria, así como de las principales características de la población y de los hogares en condición de pobreza (lengua materna, nivel de educación, acceso a servicios de salud, participación en la actividad económica, jefatura del hogar, infraestructura de la vivienda, acceso a los principales servicios básicos, entre otros indicadores). Cabe destacar que las estimaciones de pobreza se hicieron con el procedimiento tradicional o también conocido como estimación directa y se presentaron a nivel de grupos de departamentos.

Mapa de pobreza Provincial y Distrital 2013

(Instituto Nacional de Estadística e Informática;, 2015)

Este documento contiene indicadores de pobreza que fueron elaborados con una metodología que combina datos del Empadronamiento Distrital de Población y Vivienda 2012-2013 (SISFOH), la Encuesta Nacional de Hogares 2012-2013 y otras fuentes de datos, y constituyen herramientas para la priorización de los distritos más pobres del país y la implementación de políticas sociales.

2.2 Bases teóricas: Generales y Especializadas

2.2.1 EXPERIMENTOS DE MONTECARLO

Definición: Los experimentos de Monte Carlo son un tipo de algoritmos computacionales que utilizan muestreo aleatorio repetido

- Los métodos de Monte Carlo se utilizan para simular sistemas matemáticos
- Se tienden a utilizar cuando calcular resultados exactos es imposible o muy costoso
- En este trabajo de investigación se va a utilizar para generar muestras de personas y estudiar el comportamiento de las estimaciones de pobreza monetaria cuando las muestras no son infinitas

¿Por qué se llaman Monte Carlo?

- Este nombre se usó por físicos nucleares americanos en los años 40
- Querían resolver un problema de difusión de radiación, pero no sabían hacerlo analíticamente, así que decidieron resolverlo utilizando la aleatoriedad
- Al proyecto lo llamaron “Monte Carlo” en referencia al casino en Mónaco donde el tío de uno de ellos se gastaba el dinero
- Desde entonces se ha utilizado en todos los campos de la ciencia
- Para los economistas, las técnicas de Monte Carlo son importantes para resolver ciertos problemas matemáticos que surgen en muchos modelos económicos (integración, optimización), en juegos y en estadística aplicada

Monte Carlo en Estadística Aplicada

- Las técnicas de Monte Carlo se utilizan frecuentemente en dos contextos en la Estadística Aplicada
 - Para comparar y contrastar estadísticas para muestras pequeñas
 - Para estudiar cómo el comportamiento en muestras pequeñas se transforma en comportamiento en muestras grandes
- Las técnicas de Monte Carlo son útiles
 - Frecuentemente sus resultados son más precisos que los resultados asintóticos
 - Pero no son tan difíciles de obtener como los test exactos

Monte Carlo vs. Simulación

- Una simulación es una representación ficticia de la realidad mediante ordenador
- Un estudio de Monte Carlo es una técnica que puede ser usada para resolver un problema matemático/estadístico
- Una simulación Monte Carlo utiliza el muestreo repetido de datos simulados para determinar las propiedades de algún fenómeno

Un sencillo ejercicio de simulación

1. Con la ayuda de un ordenador se extrae un valor pseudo-aleatorio de la uniforme $[0, 1]$
2. Si el valor es menor o igual a 0.5 designa el resultado del experimento como “cara”
3. Si el valor es mayor que 0.5 designa el resultado como “cruz”

Esta es una simulación de lanzar una moneda al aire

Un sencillo estudio de Monte Carlo

1. Extrae dos valores pseudo-aleatorios de la uniforme $[0, 1]$
2. Si el punto identificado por esos valores se encuentran dentro de la figura, designa el resultado del experimento como “éxito”, si no, “fracaso”
3. Se repite los pasos 1 y 2 muchas veces
4. La proporción de éxitos nos da el área de la figura

Esto es usar técnicas de Monte Carlo para computar una integral compleja

Usando Monte Carlo en Estadística Aplicada (Ver gráfico 8)

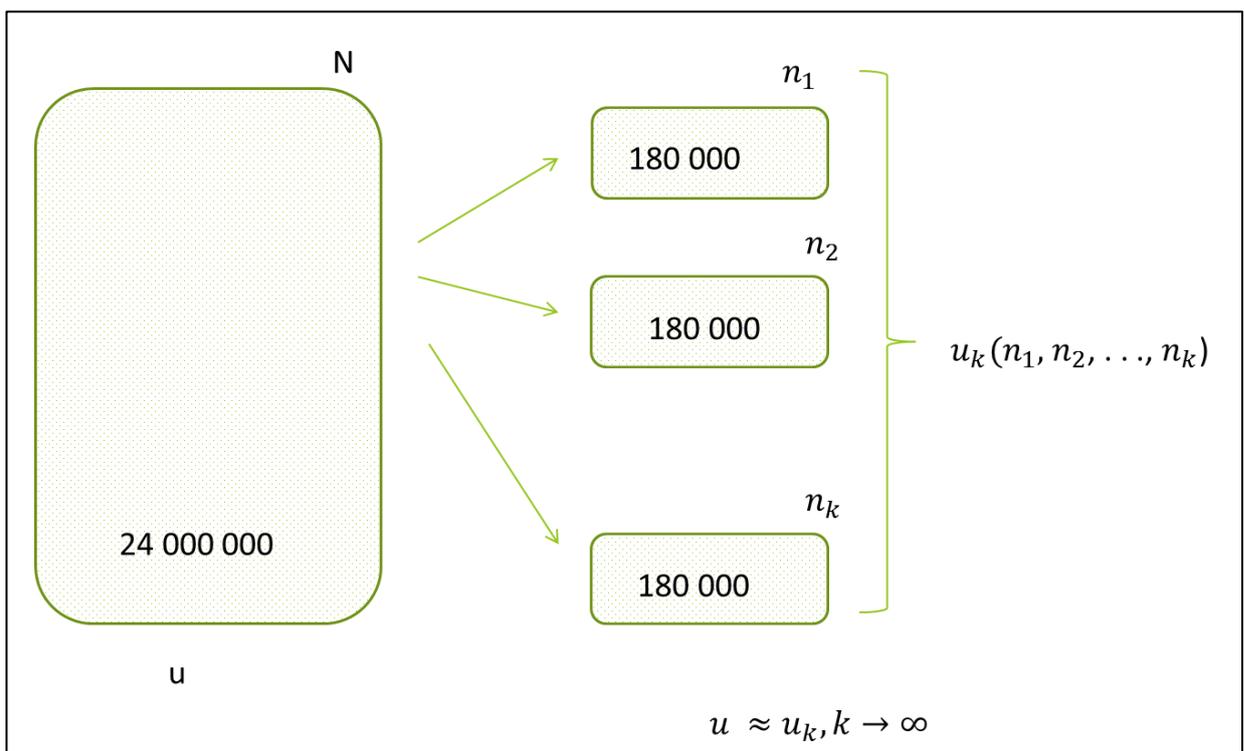
- Suponemos un modelo estadístico / econométrico y lo simulamos muchas veces
- De cada población simulada extraemos una muestra de tamaño n y computamos un estimador
- Observamos las propiedades estadísticas del estimador en todas las simulaciones, “estimamos” las propiedades del estimador cuando la muestra está fija en n
- Incrementando n y realizando todo de nuevo, “estimamos” cómo el estimador se comporta al aumentar la muestra

Resumen

- Las propiedades de muestras grandes nos dicen cómo se comporta un estimador conforme la muestra se hace arbitrariamente grande

- Las propiedades exactas de los estimadores para un tamaño muestral fijo son difíciles de obtener y cuando se obtienen es frecuentemente con supuestos muy restrictivos
- Las simulaciones de Monte Carlo son útiles en la Estadística Aplicada. Con ellas podemos estudiar:
 - Las propiedades en muestras pequeñas de los estimadores
 - Como estas propiedades se convierten en propiedades de muestras grandes

Gráfico 8: Experimentos de Montecarlo con Simulación de Muestras



Fuente: Propia

Elaboración: Propia

2.2.2 INTRODUCCION A LA ESTIMACION EN AREAS PEQUEÑAS (SAE)

Los métodos SAE (Small Area Estimation) se hicieron populares con los trabajos de Fay and Herriot (1979), Battesse, Harter y Fuller (1988) y Ghosh y Rao (1994).

Actualmente son muy populares en USA, Canada, Australia y el Norte de los países europeos

Se denomina “Estimación por áreas pequeñas (SAE)” a la metodología estadística específicamente diseñada para estimar en dominios donde el tamaño muestral es tan pequeño que los estimadores clásicos (estimadores directos) no proporcionan estimaciones precisas.

En el diseño clásico la estimación directa utiliza solamente la muestra contenida en las áreas pequeñas.

Los métodos SAE utilizan sin embargo información procedente de otras áreas y de otro tipo de información auxiliar procedente por ejemplo de registros administrativos. Además utilizan también modelos de forma explícita o implícita para obtener las estimaciones.

Métodos de estimación en poblaciones finitas

Métodos basados en el diseño

Estimadores Directos:

- Estimador Directo de Horvitz-Thompson
- Estimador Post-estratificados

Estimadores Indirectos:

- Estimador Sintético
- Estimador Compuesto

Métodos basados en modelos

- Estimador basado en un modelo logit

METODOS BASADOS EN EL DISEÑO

- No se especifican hipótesis acerca de la distribución de la población. Solamente necesitamos saber la probabilidad con la que se selecciona la muestra s .
- La variable de interés Y es una constante, y las inferencias asociadas están basadas en la distribución de probabilidad introducida en el diseño muestral.
- Los estimadores directos usan solamente información procedente de las observaciones en el dominio de interés mientras que los estimadores indirectos toman información fuera del propio dominio.
- En la teoría basada en el diseño la insesgadez y la consistencia bajo el diseño son propiedades deseables y obtenidas por la mayoría de los estimadores.
- Un estimador \hat{Y} es insesgado bajo el diseño si $E[\hat{Y}] = Y$ y es consistente bajo el diseño si es insesgado y su varianza tiende a cero a medida que aumenta el tamaño muestral.
- Los estimadores directos más conocidos son el llamado estimador directo y el postestratificado, mientras que los indirectos más conocidos son el sintético y los compuestos.
- Sea U la población (y_1, \dots, y_n) y $p(s)$ la probabilidad conocida de extraer la muestra s .
- En un muestreo aleatorio simple sin reemplazamiento, hay $\binom{N}{n}$ muestras diferentes de tamaño n . La probabilidad de seleccionar una muestra es $p(s) = 1/\binom{N}{n}$.
- En un muestreo aleatorio simple con reemplazamiento, hay N^n muestras. La probabilidad de seleccionar una muestra es $p(s) = 1/N^n$.
- En la práctica no necesitamos calcular $p(s)$, sino la probabilidad π_i de que y_i esté en la muestra. Le llamamos **probabilidad de inclusión**.

En el muestreo aleatorio

- Sea π_i la probabilidad de inclusión de la i ésima unidad en la muestra. Entonces
- $\pi_i = n/N$
- $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$
- Los métodos basados en el diseño utilizan los llamados pesos muestrales $w_i = 1/\pi_i$ (definidos como el inverso de las probabilidades de inclusión) para definir los estimadores.

ESTIMADOR DIRECTO DE HORVITZ – THOMPSON

- El estimador de Horvitz-Thompson estimator (1952) es el estimador directo más conocido
- El estimador Horvitz-Thompson del total de una variable viene dado por:

$$\hat{Y}_d = \sum_{i \in n_d} \frac{y_i}{\pi_i} = \sum_{i \in n_d} w_i y_i$$

- La varianza de \hat{Y}_d viene dada por:

$$var_{HT}(\hat{Y}_d) = \sum_{i=1}^N \sum_{j=1}^N (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j}.$$

- El estimador de Horvitz – Thompson estimator no utiliza información auxiliar, y sus pesos son los inversos de las probabilidades de inclusión.
- Es un estimador insesgado.

En muestreo aleatorio simple

$$\hat{t}_{HT} = \sum_{i=1}^n \frac{N}{n} y_i = N\bar{y}$$

Cuando $\pi_{ij} > 0 \forall i, j$, un estimador insesgado es:

$$var_{HT}(\hat{Y}_\pi) = \sum_{i \in \nu} \sum_{j \in \nu} (\pi_{ij} - \pi_i \pi_j) / \pi_{ij} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j}$$

- Cuando $n \geq 1$ the Horvitz -Thompson la varianza requiere un conocimiento completo de todos los π_{ij} , y es complicado de calcular para $n \geq 2$.
- Los estimadores pueden ser negativos por lo que es común utilizar métodos “computer-intensive”

$$\hat{y}_d^{directo} = \frac{\sum_{j=1}^{n_d} w_j y_j}{\sum_{j=1}^{n_d} w_j}$$

Dónde:

$\hat{y}_d^{directo}$ es el estimador de la variable y por área pequeña d

$y_j = 1$ (pobre) $y_j = 0$ (no pobre)

n_d tamaño muestral en la zona d

d es el área pequeña (ejemplo distrito, provincia, departamento, etc.) $d = 1, \dots, D$

w_j es el peso, dado por el inverso de la probabilidad de inclusión o pertenencia, es decir,

$w_j = N_h/n_h, \quad j = 1, \dots, n_d$ donde

N_h es la población total del estrato h

n_h es el número total de personas del estrato h

Es decir, todas las personas pertenecientes al mismo estrato tienen el mismo peso

ESTIMADOR DIRECTO POSTESTRATIFICADO

El estimador postestratificado es un estimador directo, pero puede considerarse como estimador asistido en un modelo lineal donde la variable explicativa es el grupo indicador de la variable

$$\hat{y}_d^{post} = \sum_g \frac{\sum_{j=1}^{n_{dg}} w_j y_j}{\sum_{j=1}^{n_{dg}} w_j}$$

Dónde:

\hat{y}_d^{post} es el estimador de la variable y por área pequeña d

n_{dg} Número de individuos en la muestra del grupo g en el área d

d Es el área pequeña (ejemplo distrito, provincia, departamento, etc.) $d = 1, \dots, D$

g es el grupo definido por las variables auxiliares

ESTIMADOR INDIRECTO SINTETICO

- Este término se utilizó en primer lugar por “The U.S. National Center of Health Statistics”
- Gonzalez (1973) describe los estimadores sintéticos del siguiente modo: “Una estimación objetiva se obtiene a partir de una encuesta por muestreo de un área grande. Cuando se utiliza esta estimación para obtener estimaciones para las subáreas bajo el supuesto de que las áreas pequeñas tienen las mismas características que el área grande , identificamos estas estimaciones como las estimaciones sintéticas”
- Habitualmente son estimadores sesgados
- Es un estimador indirecto que se utiliza cuando se puede suponer que las áreas pequeñas tienen las mismas características que las áreas grandes con respecto a la variable de interés. Cuando esto no sucede los estimadores sintéticos suelen ser sesgados
- Viene dado por

$$\hat{y}_d^{sint} = \sum_g \hat{y}_g$$

Dónde:

\hat{y}_d^{sint} es el estimador de la variable y por área pequeña d

d es el área pequeña (ejemplo distrito, provincia, departamento, etc.) $d = 1, \dots, D$

g es el grupo definido por las variables auxiliares

\hat{y}_g es la media en g calculada con el estimador directo anterior

ESTIMADOR INDIRECTO COMPUESTO

- Una forma de equilibrar el sesgo potencial del estimador sintético y la inestabilidad del estimador directo es tomar una media ponderada de ambos, que se llama estimador compuesto
- Los estimadores compuestos típicamente se definen como una combinación lineal entre un estimador directo y un estimador indirecto
- De este modo se equilibra la varianza del estimador directo y el sesgo del indirecto.
- En este caso presentamos los estimadores compuestos dependientes del tamaño muestral (Drew et al., 1982).

Están definidos como:

$$\hat{y}_d^{Comp} = \gamma_d \hat{y}_d^{post} + (1 - \gamma_d) \hat{y}_d^{sint}$$

Donde:

\hat{y}_d^{Comp} es el estimador del total de la variable y por área pequeña d

d es el área pequeña (comarca por sexo)

γ_d verifica $0 \leq \gamma_d \leq 1$ y viene dada por

$$\gamma_d = \begin{cases} 1 & \text{si } \hat{N}_d \geq \alpha N_d \\ \frac{\hat{N}_d}{\alpha N_d} & \text{en otro caso} \end{cases}$$

Siendo $\hat{N}_d = \sum_j w_j$ el total poblacional estimado en cada área pequeña d y α es un parámetro

Evaluaremos el estimador compuesto para distintos valores de $\alpha = 2/3, 1, 1.5$ y 2 . γ_d controla la contribución del estimador sintético. Cuando el tamaño muestral aumenta γ_d está próximo \hat{Y}^{comp} es similar al estimador postestratificado, en caso contrario se da más peso al estimador sintético.

ESTIMADORES BASADOS EN MODELOS

- Los métodos basados en modelos utilizan modelos estadísticos para predecir los valores no muestreados
- Los más populares son los modelos lineales mixtos, que utilizan efectos aleatorios para explicar la variabilidad entre áreas que no puede ser explicada por las variables auxiliares
- Las relaciones entre las variables aleatorias se expresan a través de un modelo para la distribución de probabilidad conjunta, y las predicciones se realizan en base a dicho modelo
- La teoría basada en los modelos se llama teoría de la predicción y considera que y_1, \dots, y_N son realizaciones de las variables aleatorias Y_1, \dots, Y_N
- Sea s el conjunto de observaciones muestrales y r las observaciones no muestreadas, el total de observaciones puede expresarse como

$$T = \sum_{i \in s} y_i + \sum_{i \in r} y_i$$

- Para estimar T predecimos el valor $\sum_{i \in r} y_i$ de las variables no observadas $\sum_{i \in r} y_i$
- Si \hat{Y}_i es el predictor de la i -ésima observación, entonces un estimador de T viene dado por

$$\hat{T} = \sum_{i \in s} y_i + \sum_{i \in r} \hat{Y}_i$$

- El error de estimación viene dado por

$$\hat{T} - T = \sum_{i \in r} \hat{Y}_i - \sum_{i \in r} Y_i$$

Es decir la diferencia entre la predicción de la parte no observada y su valor real

- Los predictores del total no muestreado son combinaciones lineales de y_i' s. Constituyen los llamados Mejores Estimadores Lineales Insesgados de Varianza Mínima (BLUE) (Valliant, Dorfman, and Royall, 2000)
- Notemos que insesgadez no significa $E_M(\hat{T}) = T$, ya que T en la filosofía del modelo es una v.a.

ESTIMADOR BASADO EN UN MODELO LOGIT

Supongamos que disponemos de variables binarias y_{id} que indican si el individuo i del área d está en condición de pobre (1) o no (0).

El modelo es el siguiente:

$$\log\left(\frac{p_{id}}{1 - p_{id}}\right) = x_{id}^T \beta$$

Dónde:

p_{id} es la probabilidad de que el individuo i de área d esté pobre

$x_{id} = (x_{id,1}, x_{id,2}, \dots, x_{id,p})^T$ es el vector de p covariables. El estimador del total en la comarca d es:

$$\hat{P}_d = \sum_{i=1}^{N_d} \frac{e^{x_{id}^T \hat{\beta}}}{1 + e^{x_{id}^T \hat{\beta}}}$$

2.2.3 INDICADORES DE EVALUACION

Acuracidad: De las muchas definiciones de acuracidad en la presente investigación se tomará como acuracidad a la proximidad al “valor verdadero” y para cuantificar dicha proximidad se evaluarán los siguientes indicadores:

a) **Media del sesgo relativo en valor absoluto(SRAM):** Definido de la siguiente manera:

$$SRA_d(\hat{y}) = \frac{1}{K} \left| \sum_{k=1}^K \frac{\hat{y}_d(k) - Y_d}{Y_d} \right| 100$$

$$SRAM(\hat{y}) = \frac{1}{D} \sum_d SRA_d(\hat{y})$$

Dónde:

D representa la cantidad de simulaciones

d representa la agrupación o dominio estimado

Y_d representa el verdadero valor de la pobreza monetaria dentro de la agrupación

$\hat{y}_d(k)$ representa el valor estimado de pobreza monetaria para la simulación k -ésima.

Ejemplo del indicador SRAM:

En el cuadro 1 los valores de los parámetros son:

$D = 6$ simulaciones

$d = 1,2,3, \dots, 25$, es decir, son los departamentos

$Y_d =$ son los valores de la columna pobreza referencial

$\hat{y}_d(k) =$ son los valores de la pobreza simulada por departamento

Cuadro 1: Resultados del indicador SRAM a nivel nacional y SRA a nivel departamental

Departamentos	Pobreza Simulación 1	Pobreza Simulación 2	Pobreza Simulación 3	Pobreza Simulación 4	Pobreza Simulación 5	Pobreza Simulación 6	Pobreza Referencial	SRA	SRAM
Amazonas	0.46	0.46	0.47	0.47	0.47	0.47	0.49	4.3	22.2
Áncash	0.24	0.25	0.25	0.25	0.26	0.24	0.27	6.4	22.2
Apurímac	0.50	0.51	0.50	0.51	0.51	0.50	0.49	3.8	22.2
Arequipa	0.15	0.16	0.15	0.15	0.16	0.16	0.12	24.8	22.2
Ayacucho	0.54	0.52	0.53	0.53	0.53	0.55	0.51	3.6	22.2
Cajamarca	0.54	0.56	0.56	0.55	0.55	0.55	0.56	1.7	22.2
Callao	0.17	0.19	0.18	0.18	0.18	0.18	0.19	6.9	22.2
Cusco	0.24	0.24	0.23	0.25	0.23	0.25	0.24	1.1	22.2
Huancavelica	0.48	0.47	0.48	0.47	0.45	0.45	0.48	3.9	22.2
Huánuco	0.45	0.44	0.44	0.44	0.46	0.43	0.45	0.7	22.2
Ica	0.06	0.06	0.06	0.06	0.07	0.06	0.07	11.0	22.2
Junín	0.22	0.23	0.22	0.23	0.23	0.23	0.23	0.0	22.2
La Libertad	0.29	0.30	0.29	0.29	0.29	0.30	0.31	6.0	22.2
Lambayeque	0.22	0.21	0.22	0.21	0.22	0.22	0.23	6.1	22.2
Lima	0.15	0.15	0.15	0.15	0.14	0.15	0.15	1.0	22.2
Loreto	0.36	0.37	0.36	0.37	0.36	0.36	0.38	2.9	22.2
Madre De Dios	0.04	0.04	0.04	0.04	0.03	0.04	0.04	5.9	22.2
Moquegua	0.11	0.12	0.13	0.11	0.11	0.13	0.12	0.2	22.2
Pasco	0.47	0.45	0.46	0.47	0.46	0.46	0.48	3.3	22.2
Piura	0.33	0.32	0.32	0.33	0.33	0.33	0.37	11.5	22.2
Puno	0.38	0.38	0.37	0.36	0.37	0.38	0.38	0.5	22.2
San Martín	0.28	0.29	0.28	0.29	0.28	0.29	0.31	8.0	22.2
Tacna	0.16	0.15	0.16	0.16	0.15	0.16	0.16	3.3	22.2
Tumbes	0.11	0.12	0.11	0.11	0.12	0.11	0.13	14.0	22.2
Ucayali	0.13	0.13	0.13	0.13	0.13	0.13	0.13	2.0	22.2

Fuente: Bases muestrales simuladas del SISFOH 2012-2013

Elaboración: Propia

Observación:

- El indicador SRA mide para cada departamento (dominio) los sesgos relativos con respecto a un valor referencial. Para el caso del departamento de Arequipa posee un SRA=24.8 más alto de todos los departamentos y esto es porque el valor referencial de pobreza es 0.12 y los valores simulados oscilan entre 0.15 y 0.16. Un análisis similar se da para el departamento de Junín con un SRA=0.0 (uno de los más pequeños), valor referencial es 0.23 y valores de simulación que oscilan entre 0.22 y 0.23.
- El indicador SRAM es un indicador en conjunto de los sesgos relativos. Dicho indicador será utilizado para comparar entre métodos de estimación de pobreza monetaria.

b) **Media de la raíz cuadrada del error cuadrático medio relativo(RECMRM):** Definido de la siguiente manera:

$$RECMR_a(\hat{y}) = \left(\frac{1}{K} \sum_{k=1}^K \left(\frac{\hat{y}_a(k) - Y_a}{Y_a} \right)^2 \right)^{\frac{1}{2}} 100$$

$$RECMRM(\hat{y}) = \frac{1}{D} \sum_d RECMR_a(\hat{y})$$

Dónde:

D representa la cantidad de simulaciones

d representa la agrupación o dominio estimado

Y_a representa el verdadero valor de la pobreza monetaria dentro de la agrupación

$\hat{y}_a(k)$ representa el valor estimado de pobreza monetaria para la simulación k -ésima.

Ejemplo del indicador RECMRM:

En el cuadro 2 los valores de los parámetros son:

$D = 6$ simulaciones

$d = 1,2,3, \dots, 25$, es decir, son los departamentos

$Y_a =$ son los valores de la columna pobreza referencial

$\hat{y}_a(k) =$ son los valores de la pobreza simulada por departamento

Cuadro 2: Resultados del indicador RECMRM a nivel nacional y RECMR a nivel departamental

Departamentos	Pobreza Simulación 1	Pobreza Simulación 2	Pobreza Simulación 3	Pobreza Simulación 4	Pobreza Simulación 5	Pobreza Simulación 6	Pobreza Referencial	RECMR	RECMRM
Amazonas	0.46	0.46	0.47	0.47	0.47	0.47	0.49	4.4	27.4
Áncash	0.24	0.25	0.25	0.25	0.26	0.24	0.27	6.8	27.4
Apurímac	0.50	0.51	0.50	0.51	0.51	0.50	0.49	4.0	27.4
Arequipa	0.15	0.16	0.15	0.15	0.16	0.16	0.12	25.2	27.4
Ayacucho	0.54	0.52	0.53	0.53	0.53	0.55	0.51	4.0	27.4
Cajamarca	0.54	0.56	0.56	0.55	0.55	0.55	0.56	1.9	27.4
Callao	0.17	0.19	0.18	0.18	0.18	0.18	0.19	7.6	27.4
Cusco	0.24	0.24	0.23	0.25	0.23	0.25	0.24	3.3	27.4
Huancavelica	0.48	0.47	0.48	0.47	0.45	0.45	0.48	4.7	27.4
Huánuco	0.45	0.44	0.44	0.44	0.46	0.43	0.45	2.2	27.4
Ica	0.06	0.06	0.06	0.06	0.07	0.06	0.07	12.2	27.4
Junín	0.22	0.23	0.22	0.23	0.23	0.23	0.23	1.4	27.4
La Libertad	0.29	0.30	0.29	0.29	0.29	0.30	0.31	6.3	27.4
Lambayeque	0.22	0.21	0.22	0.21	0.22	0.22	0.23	6.5	27.4
Lima	0.15	0.15	0.15	0.15	0.14	0.15	0.15	2.2	27.4
Loreto	0.36	0.37	0.36	0.37	0.36	0.36	0.38	3.4	27.4
Madre De Dios	0.04	0.04	0.04	0.04	0.03	0.04	0.04	15.1	27.4
Moquegua	0.11	0.12	0.13	0.11	0.11	0.13	0.12	7.5	27.4
Pasco	0.47	0.45	0.46	0.47	0.46	0.46	0.48	3.7	27.4
Piura	0.33	0.32	0.32	0.33	0.33	0.33	0.37	11.5	27.4
Puno	0.38	0.38	0.37	0.36	0.37	0.38	0.38	1.6	27.4
San Martín	0.28	0.29	0.28	0.29	0.28	0.29	0.31	8.2	27.4
Tacna	0.16	0.15	0.16	0.16	0.15	0.16	0.16	4.2	27.4
Tumbes	0.11	0.12	0.11	0.11	0.12	0.11	0.13	14.4	27.4
Ucayali	0.13	0.13	0.13	0.13	0.13	0.13	0.13	2.4	27.4

Fuente: Bases muestrales simuladas del SISFOH 2012-2013

Elaboración: Propia

Observación:

- El indicador RECMR mide para cada departamento (dominio) las desviaciones relativas con respecto a un valor referencial. Para el caso del departamento de Arequipa posee un RECMR=25.2 más alto de todos los departamentos y esto es porque el valor referencial de pobreza es 0.12 y los valores simulados oscilan entre 0.15 y 0.16. Un análisis similar se da para el departamento de Puno con un RECMR=1.6 (uno de los más pequeños), valor referencial es 0.38 y valores de simulación que oscilan entre 0.37 y 0.38.
- El indicador RECMRM es un indicador en conjunto de las desviaciones relativas. Dicho indicador será utilizado para comparar entre métodos de estimación de pobreza monetaria.

2.2.4 REMUESTREO

Definición: se denomina remuestreo (resampling en inglés) a una variedad de métodos que permiten estimar la precisión de muestras estadísticas (medianas, varianzas, percentiles) y entre los más importantes técnicas tenemos el remuestreo de Jackknife y el remuestreo de Bootstrap.

METODO DE ESTIMACION DE LA VARIANZA MEDIANTE REMUESTREO POR JACKKNIFE

Hasta hace algunos años, el único método para cálculo de varianza en muestreos complejos que aplicaban muchas instituciones era el de grupos aleatorios dependientes, o bien la linealización. El creciente desarrollo de los equipos de cómputo ha facilitado la exploración de métodos de intenso trabajo computacional. Tal es el caso del jackknife, el cual, cada vez está siendo adoptado por más analistas de encuestas en el mundo, en parte porque es más simple de aplicar que el bootstrap, además de tener cierto apoyo teórico y de que los estudios empíricos revelan, en general, un buen comportamiento.

Caracterización del Estimador

La técnica a la que se ha denominado por jackknife partió de un estimador del coeficiente de correlación serial, con corrección del sesgo, el cual fue creado por Quenouille (1949). Años más tarde, Tukey (1958), retornó el trabajo de Quenouille y propuso un estimador de la varianza y el sesgo de una estadística. De manera general, el jackknife es un método no-paramétrico basado en remuestreo. Sus inicios se dieron en el contexto de poblaciones infinitas aunque más tarde fue aplicado al caso de poblaciones finitas.

Es posible revisar la lógica tras la idea de Quenouille para estimar el sesgo, y posteriormente ver cómo fue aprovechada por Tukey para conseguir un estimador de la varianza. En primer lugar, Quenouille (1949) advierte que para una estadística t_n , basada en una muestra de tamaño n , que se pueda expandir a través de una serie de Taylor, que sea consistente, con cumulantes finitos, se cumple que:

$$E(t_n - \theta) = \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots$$

Igualmente, si se calculara la misma estadística, pero con $n-1$ elementos, se tendría que:

$$E(t_{n-1} - \theta) = \frac{a_1}{n-1} + \frac{a_2}{(n-1)^2} + \frac{a_3}{(n-1)^3} + \dots$$

Es claro que bajo los supuestos indicados, el sesgo disminuye, a medida que n crece. Efron (1982), da una explicación gráfica al planteamiento de Quenouille. (Para entenderla mejor, conviene ver la Gráfica xx). Sea E_n la esperanza del estimador t_n , basado en un tamaño de muestra n , entonces la grafica de E_n vs $\frac{1}{n}$ estaría dada por una curva creciente y cóncava, que parte del punto $(0, \theta)$, el cual correspondería al caso $n = \infty$. Aproximando tal curva con una linea, se tendria que:

$$\frac{E_n - E_\infty}{E_{n-1} - E_\infty} = \frac{1/n}{1/(n-1) - 1/n}$$

Se entiende por E_∞ , el verdadero valor del parámetro, o sea θ . De ahí que el sesgo del estimador basado en n elementos, está dado por $E_n - E_\infty$, lo cual, resolviendo la relación anterior, resulta en

$$\text{Sesgo} = E_n - E_\infty = (n-1)(E_{n-1} - E_n)$$

De la misma relación se obtiene

$$\theta = E_\infty = nE_n - (n-1)E_{n-1}$$

La propuesta de Quenouille es obtener, a partir de la muestra de tamaño n , un estimador θ , basado en los n elementos, junto con n estimadores basados en $(n-1)$ elementos cada uno, denominados por, $\hat{\theta}_{(.)}$. Así pues, se calcula el promedio de los últimos:

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)}$$

El estimador del sesgo está dado por:

$$\widehat{\text{Sesgo}} \equiv (n-1)(\hat{\theta}_{(.)} - \hat{\theta})$$

También se deduce un nuevo estimador con corrección del sesgo:

$$\tilde{\theta} = \hat{\theta} - \widehat{\text{Sesgo}} \equiv n\hat{\theta} - (n-1)\hat{\theta}_{(.)}$$

Se puede verificar que si bien $\hat{\theta}$ tiene un sesgo del orden de $1/n$, el sesgo del nuevo estimador es del orden $1/n^2$

Basado en el planteamiento de Quenouille, Tukey (1958), sugirió calcular n *pseudovalores*, denominados $\tilde{\theta}_i$, los cuales se construyen a partir del estimador que se consigue al dejar el dato i fuera de la muestra. Como se ve a continuación el pseudovalor de Tukey corresponde directamente al estimador con corrección del sesgo de Quenouille.

$$\tilde{\theta}_i = \hat{\theta} + (n - 1)(\hat{\theta} - \hat{\theta}_{(i)}) = n\hat{\theta} - (n - 1)\hat{\theta}_{(i)}$$

El estimador jackknife de θ , estaría dado por:

$$\tilde{\theta}_J = \frac{1}{n} \sum_{i=1}^n \tilde{\theta}_i \dots \dots (\alpha)$$

La estimación de la varianza se basa en las desviaciones de los pseudovalores con respecto a su media, de la siguiente forma:

$$\widehat{VAR}_{J1} = \frac{1}{n(n-1)} \sum_{\alpha=1}^n (\tilde{\theta}_i - \tilde{\theta}_J)^2$$

Es más común asociar el jackknife directamente a los estimadores $\hat{\theta}_{(i)}$ ($i = 1, \dots, n$), quizás porque es más conocida la aplicación del método a problemas de regresión, donde no se involucran pseudovalores como los que propone Tukey para un problema de muestreo. Así pues, se puede ver que la anterior fórmula es equivalente a:

$$\widehat{VAR}_{J1} = \frac{(n-1)}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2$$

Por lo tanto, es factible obtener el estimador de la varianza jackknife, sin necesidad de calcular los pseudovalores de Tukey. Por otra parte, en la práctica, \widehat{VAR}_{J1} no sólo se utiliza para estimar la varianza de $\tilde{\theta}$ sino también la de $\hat{\theta}$. De hecho, en ocasiones, también se usa el estimador:

$$\widehat{VAR}_{J2} = \frac{1}{n(n-1)} \sum_{i=1}^n (\tilde{\theta}_{(i)} - \hat{\theta})^2$$

1. Se calcula un estimador $\hat{\theta}$ de θ basado en toda la muestra y respetando el diseño.
2. La muestra es particionada en k grupos con m observaciones cada uno. Es decir, se supone n múltiplo de k ($n = mk$).
3. Se calculan k estimadores, con la misma forma de $\hat{\theta}$ pero omitiendo en cada ocasión uno de los grupos. Es decir, se obtienen k estimaciones con $m(k - 1)$ unidades a las que se denominan mediante $\hat{\theta}_{(i)}$ (donde $i = 1, \dots, k$).
4. Se calcula, para cada grupo, su pseudovalor denominado por: $\tilde{\theta}_i$, de la siguiente forma.

$$\tilde{\theta}_i = k\hat{\theta} - (k - 1)\hat{\theta}_{(i)}$$

5. Se obtiene el estimador jackknife promediando los pseudovalores, como en (α)

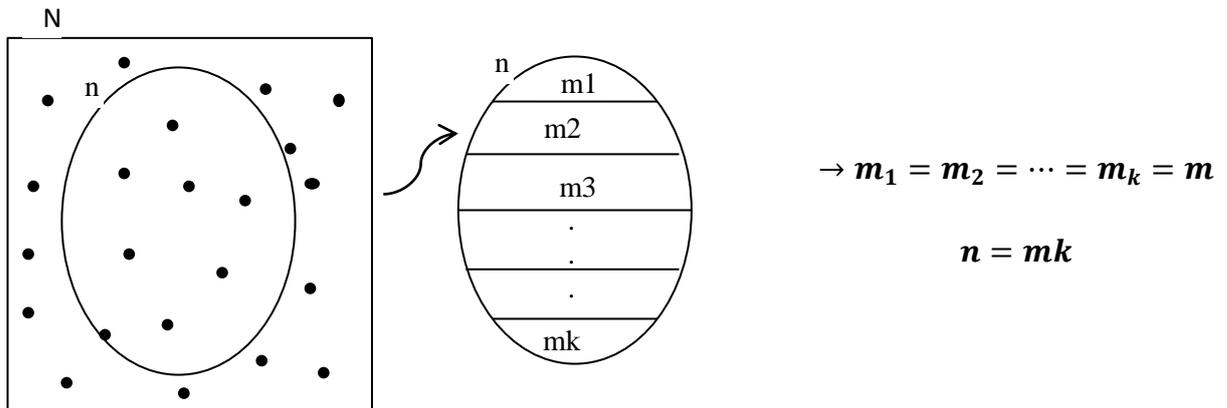
$$\tilde{\theta} = \frac{\sum_{i=1}^k \tilde{\theta}_i}{k}$$

Así pues:

$$\begin{aligned} \tilde{\theta}_j &= k\hat{\theta} - (k - 1) \frac{\sum_{(i)=1}^k \hat{\theta}_{(i)}}{k} \\ &= \sum_{(i)=1}^k \hat{\theta}_{(i)} \left(\frac{1}{k} - 1 \right) + \sum \hat{\theta} \\ &= \frac{1}{k} \sum_{(i)=1}^k \hat{\theta}_i - \sum_{i=1}^k (\hat{\theta}_{(i)} - \hat{\theta}) \end{aligned}$$

Vemos que $\tilde{\theta}_j$ no es simplemente el promedio de las estimaciones de k grupos, sino que se considera un término que aporta información sobre la desviación simple de las estimaciones de cada grupo respecto al estimador general.

Gráfico 10: Resumen del remuestreo por Jackknife



Fuente: Propia

Elaboración: Propia

Se define:

$\hat{\theta}_{(1)}$ considera a m_2, m_3, \dots, m_k elementos

$\hat{\theta}_{(k)}$ considera a m_1, m_2, \dots, m_{k-1} elementos

Un estimador podría ser:

$$\hat{\theta}_{(k)} = \frac{\sum_{i \in R} y_i}{mk - m}$$

Los pseudovalores se definen:

$$\tilde{\theta}_i = k\hat{\theta} - (k-1)\hat{\theta}_{(i)}$$

$$\tilde{\theta}_1 = k\hat{\theta} - (k-1)\hat{\theta}_{(1)}$$

$$\tilde{\theta}_k = k\hat{\theta} - (k-1)\hat{\theta}_{(k)}$$

$$\tilde{\theta}_J = \frac{\sum_{i=1}^k \tilde{\theta}_i}{k} = k\hat{\theta} - (k-1) \frac{\sum_{(i)=1}^k \hat{\theta}_{(i)}}{k}$$

$$V(\tilde{\theta}_J) = \frac{1}{n(n-1)} \sum_{i=1}^n (\tilde{\theta}_i - \tilde{\theta}_J)^2$$

METODO DE ESTIMACION DE LA VARIANZA MEDIANTE REMUESTREO POR BOOTSTRAP

Caracterización del Estimador

Efron (1979), dio a conocer una técnica para estimar la distribución acumulada de probabilidad, F , de una variable aleatoria. Al presentar el *bootstrap*, como llamó al método, resaltó su relación con el ya conocido *jackknife*, además de arguir mayor generalidad. Demostró que el *jackknife* es una aproximación lineal del *bootstrap*, lo cual, sin la referencia cronológica, podría pensarse que ocurriera al contrario (que primero se desarrollara el *bootstrap* y luego se aproximara linealmente). Cabe resaltar que aunque el *bootstrap* se concibió en el marco de una población infinita, Efron (1979, 1982) consideró brevemente el caso de población finita. Posteriormente, autores como Rao y Wu (1985, 1988) y Kovar, Rao y Wu (1988), incursionaron en la aplicación del *bootstrap* en problemas de muestreo, para obtener estimadores de la varianza e intervalos de confianza.

El problema más elemental se basa en la observación de una muestra independiente de tamaño n de una variable aleatoria X , cuya distribución de probabilidad F es desconocida. Se tiene,

$$X_i \sim F \text{ donde, } X_i \text{ i. i. d., } i = 1, 2, \dots, n.$$

Por otra parte, se observa $x = (x_1, x_2, \dots, x_n)$, la realización de $X = (X_1, X_2, \dots, X_n)$.

Ahora bien, si se tiene la variable aleatoria $R(X, F)$, que depende de X y F desconocida, se desearía estimar la distribución muestral de R con base en la observación x . Para tal efecto Efron (1979) propone la consecución de los siguientes pasos, que describen el *bootstrap* en su forma más simple:

1. Se construye la distribución empírica, \hat{F} , atribuyendo una masa de $1/n$ a cada observación x_1, x_2, \dots, x_n
2. Se considera

$$X_i^* \sim \hat{F} \text{ donde, } X_i^* \text{ i. i. d., } i = 1, 2, \dots, n$$

A partir de \hat{F} se obtiene la muestra "*bootstrap*" de tamaño n , $x^* = (x_1^*, x_2^*, \dots, x_n^*)$; lo que equivaldría a obtener dicha muestra, de (x_1, x_2, \dots, x_n) con reposición. Con base en x^* se obtiene $R^*(X^*, \hat{F})$, el valor "*bootstrap*" de R .

3. Se lleva a cabo el inciso 2 repetidas veces dejando \hat{F} fija, para así aproximar la distribución muestral de $R(X, F)$ mediante la distribución de las $R^*(X^*, \hat{F})$,

Es claro que si lo que se deseara fuera aproximar la distribución de X_i , el estimador máximo verosímil, no paramétrico, de F es \hat{F} , o sea, la distribución empírica; ante lo cual, no tendría sentido intentar afinar \hat{F} . Es importante hacer énfasis en el objetivo del *bootstrap*, porque si no es bien comprendido, pudiera parecer que es una solución redundante a un problema que ya sabemos resolver.

Cuando se obtiene la muestra (x_1, x_2, \dots, x_n) , se tienen n observaciones de X , y como ya mencionamos, podemos estimar $F_x(X) = F$. Sin embargo, si nos interesa $R(X, F) = t(X)$, sólo contamos con una observación, es decir, se tiene una muestra de tamaño $n = 1$, de $t(X)$. Obviamente, nos vemos imposibilitados para estimar $F_{T(X)}(T(X))$, pues además, $F_x(X)$ es desconocida.

La solución de Efron valora a \hat{F} como máximo verosímil; pues parte de la similitud que espera de, esta distribución con F , para simular una muestra de $R(X, F) = t(X)$, representada por los valores $R^*(x, \hat{F}) = t^*(x)$; de lo que se sigue, estimar la distribución de R , con la de R^* , a la cual se denomina distribución *bootstrap*. Queda claro entonces que el paso clave radica en cómo se calcule la distribución *bootstrap*. Para resolver este problema fundamental Efron (1979), propone tres alternativas, que se describen a continuación:

1. Cálculo teórico directo. (Obviamente, este método no siempre es posible aplicarlo).
2. Aproximación Monte Carlo. Se generan repetidas observaciones de X^* , conservando \hat{F} fija, se calculan los valores correspondientes de $R^*(X_1^*, \hat{F})$, $R^*(X_2^*, \hat{F})$, ..., $R^*(X_n^*, \hat{F})$, y el histograma de estos últimos, lo que se tomara como la aproximación de la distribución *bootstrap*.
3. Utilizar algún método de linealización por series de Taylor para obtener la media y varianza de la distribución *bootstrap* de R^* .

Aplicación del bootstrap en muestreo

Supongamos que se tiene una población finita de la que interesa conocer el parámetro θ , que es una función del vector de medias de p variables medidas en cada elemento de la población, $\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)^T$. Se considera el estimador $\hat{\theta} = g(\bar{X})^5$, que aun siendo una función no-lineal de las observaciones, puede ser calculado, pero no así su varianza. En el caso de un muestreo aleatorio simple, de una población finita, el "bootstrap ingenuo" se lleva a cabo mediante los siguientes pasos:

1. Partiendo de la muestra observada, se selecciona una muestra aleatoria simple con reemplazo, cuyos valores se denominan $\{x_i\}_1^n$. Se calculan $\bar{x}^* = \frac{1}{n} \sum_1^n x_i^*$, $\hat{\theta}^* = g(\bar{x}^*)$,
2. Se repite el paso anterior una gran cantidad de veces B , y cada estimación de θ se denomina $\hat{\theta}^{*1}$, $\hat{\theta}^{*2}$, ..., $\hat{\theta}^{*B}$. Se obtiene entonces la media de estas estimaciones:

$$\hat{\theta}_a^* = \sum_{b=1}^B \hat{\theta}^{*b} / B$$

3. El estimador de la varianza de $\hat{\theta} = g(\bar{x})$ está dado por:

$$v_b(a) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*b} - \hat{\theta}_a^*)^2$$

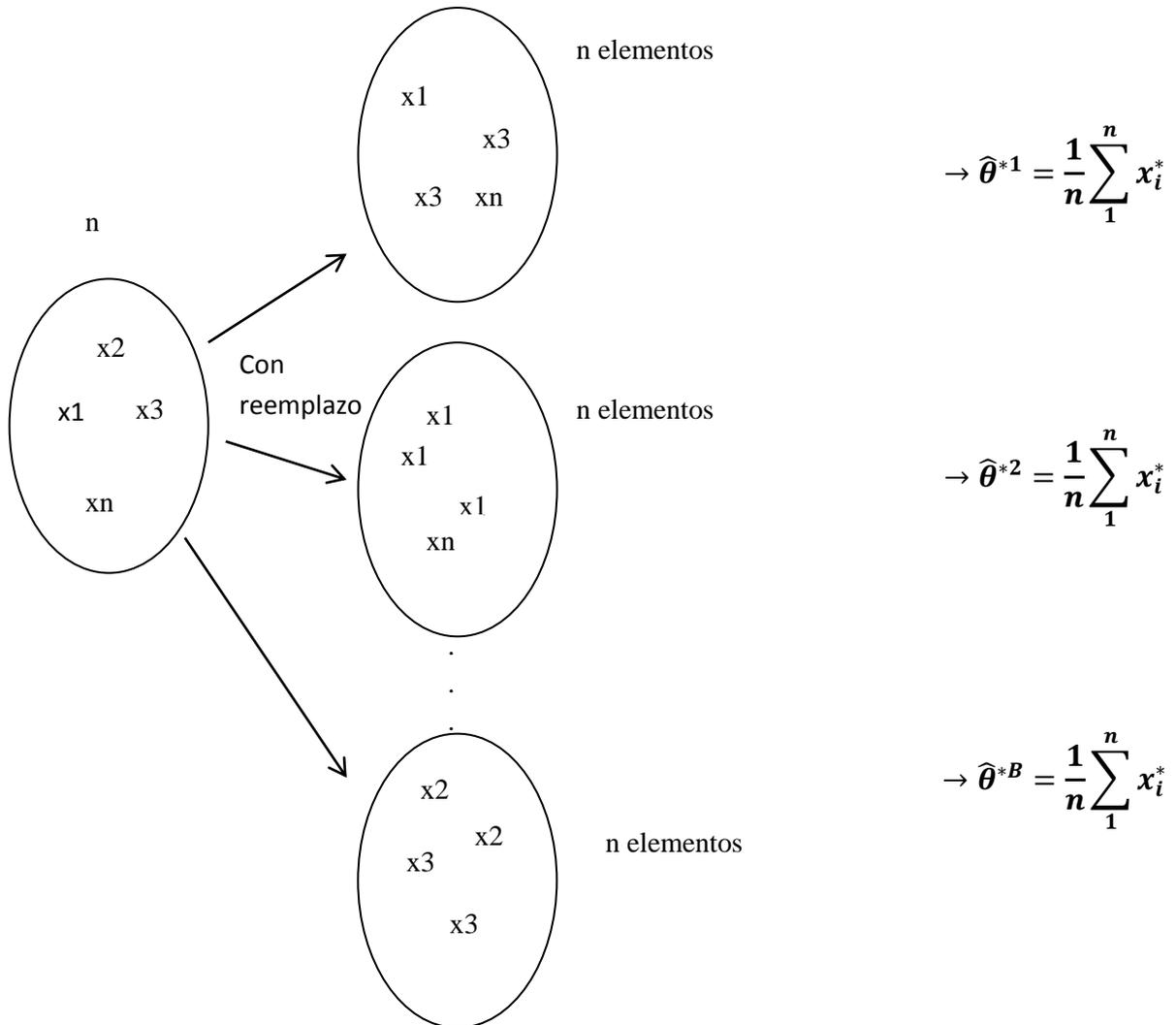
Se puede apreciar que el estimador de la varianza es la aproximación Monte Carlo a

$$Var(\hat{\theta}^*) = E(\hat{\theta}^* - E_*\hat{\theta}^*)^2,$$

Donde $E_*\hat{\theta}^*$ es la esperanza bajo la distribución bootstrap.

Es de esperar que si Rao y Wu (1988), llamaron a la metodología anterior "ingenua", ellos hayan presentado un cambio en la técnica, con una justificación adecuada.

Gráfico 11: Resumen del remuestreo por bootstrap



Fuente: Propia

Elaboración: Propia

Entonces el estimador bootstrap sería:

$$\hat{\theta}_a^* = \frac{\sum_{b=1}^B \hat{\theta}^{*b}}{B}$$

$$Var(\hat{\theta}_a^*) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*b} - \hat{\theta}_a^*)^2$$

2.3 Formulación de Hipótesis

Hipótesis General

- El estimador Sintético posee mejores características para estimar de manera más precisa el indicador de pobreza monetaria

Hipótesis Específica

- El estimador Sintético posee mejores indicadores de precisión relativa comparado con los cinco métodos de áreas pequeñas (SAE) para la pobreza monetaria
- El estimador Sintético posee mejor precisión muestral comparado con el método vigente para la pobreza monetaria

2.4 Variables y Operacionalización

Pobreza Monetaria y su definición Conceptual:

La pobreza es una condición en la cual una o más personas tienen un nivel de gasto inferior al mínimo socialmente aceptado que en este caso vendría a ser una línea de pobreza.

En el siguiente cuadro se muestra a más detalle las variables de análisis en la presente investigación:

Para el caso de la encuesta ENAHO:

Cuadro 3: Operacionalización para las variables de ENAHO

Clasificación	Variable	Definición Conceptual	Tipo	Escala	Definición Operacional	Indicador
Dependiente o respuesta	Pobre	Condición de una persona que posee un nivel de gasto inferior a un valor referencial.	Cualitativo	Nominal	0: No pobre 1: Pobre	Indicador del porcentaje de personas pobres en un dominio
Independiente o explicativa	Gpcm	Gasto percapita mensual de consumo del hogar	Cuantitativo	Razón, continuo	$gpcm = (\text{Gasto total bruto}) / (12 * \text{Miembros por hogar})$	Indicador del gasto percapita del consumo del hogar en un dominio
	Edad	Edad en años de la población	Cuantitativo	Razón, discreto	Valor observado en la variable p208a	Indicador de la edad promedio de las personas en un dominio
	Sexo	Sexo de la población	Cualitativo	Nominal	Valor observado en la variable p207: 1: Sexo masculino 2: Sexo femenino	Indicador de la cantidad de personas con determinado sexo en un dominio
	Edse	Agrupación de variables Edad y Sexo	Cualitativo	Nominal	Valor categorizado después de la agrupación de las variables edad y sexo: 1: 0-16 años y de sexo masculino 2: 0-16 años y de sexo femenino 3: 17-65 años y de sexo masculino 4: 17-65 años y de sexo femenino 5: 66 a más años y de sexo masculino 6: 66 a más años y de sexo femenino	Indicador de la cantidad de personas con determinada categoría de la variable edse en un dominio
	estrat	Estratosocioeconómico multivariado de los conglomerados	Cualitativo	Nominal	Valor observado en la variable estratosocioeconomico_multi: 1: Estrato socioeconómico mas alto 2: Estrato socioeconómico que le sigue al estrato 1 3: Estrato socioeconómico que le sigue al estrato 2 4: Estrato socioeconómico que le sigue al estrato 3 5: Estrato socioeconómico que le sigue al estrato 4 6: Estrato socioeconómico mas bajo	Indicador de la cantidad de personas con determinado estratosocioeconomico multi en un dominio

Fuente: ENAHO 2012-2013

Elaboración: Propia

Para el caso del censo SISFOH:

Cuadro 4: Operacionalización de Variables para el SISFOH

Clasificación	Variable	Definición Conceptual	Tipo	Escala	Definición Operacional	Indicador
Dependiente o respuesta	Pobre	Condición de una persona que posee un nivel de gasto inferior a un valor referencial.	Cualitativo	Nominal	0: No pobre 1: Pobre	Indicador del porcentaje de personas pobres en un dominio
Independiente o explicativa	Gpcm	Gasto percapita mensual de consumo del hogar	Cuantitativo	Razón, continuo	Se calcula en base a los valores: exp(YHAT_001), exp(YHAT_002), ... , exp(YHAT_100)	Indicador del gasto percapita del consumo del hogar en un dominio
	Edad	Edad en años de la población	Cuantitativo	Razón, discreto	Valor observado en la variable edad01	Indicador de la edad promedio de las personas en un dominio
	Sexo	Sexo de la población	Cualitativo	Nominal	Valor observado en la variable sexo01 1: Sexo masculino 2: Sexo femenino	Indicador de la cantidad de personas con determinado sexo en un dominio
	Edse	Agrupación de variables Edad y Sexo	Cualitativo	Nominal	Valor categorizado después de la agrupación de las variables edad y sexo: 1: 0-16 años y de sexo masculino 2: 0-16 años y de sexo femenino 3: 17-65 años y de sexo masculino 4: 17-65 años y de sexo femenino 5: 66 a más años y de sexo masculino 6: 66 a más años y de sexo femenino	Indicador de la cantidad de personas con determinada categoría de la variable edse en un dominio
	estrat	Estratosocioeconómico multivariado de los conglomerados	Cualitativo	Nominal	Valor observado en la variable estratosocioeconomico_multi: 1: Estrato socioeconómico mas alto 2: Estrato socioeconómico que le sigue al estrato 1 3: Estrato socioeconómico que le sigue al estrato 2 4: Estrato socioeconómico que le sigue al estrato 3 5: Estrato socioeconómico que le sigue al estrato 4 6: Estrato socioeconómico mas bajo	Indicador de la cantidad de personas con determinado estratosocioeconomico multi en un dominio

Fuente: SISFOH 2012-2013

Elaboración: Propia

CAPITULO III

MARCO METODOLOGICO

3.1 Tipo, Nivel y Diseño de investigación

El presente trabajo de investigación es de tipo aplicada, de alcance o nivel explicativo y diseño de tipo no experimental y transversal.

Aplicada: proporcionará mejoras en precisión a las estimaciones de pobreza monetaria

Descriptivo: debido a que se van a medir el indicador de pobreza monetaria en un instrumento censal (SISFOH) en agrupaciones departamentales y por distintos métodos de estimación.

No experimental: se observará el fenómeno de la pobreza monetaria tal como se da en su contexto natural y recopilado de las bases de datos de la ENAHO y del SISFOH

Transversal: se van a describir, analizar y comparar la pobreza monetaria solo para el periodo de años 2012-2013.

3.2 Población, Muestra y Tamaño de muestra

Sobre la Población:

La población de referencia ha de ser considerado sobre todos los registros de personas de la base de datos del SISFOH 2012-2013 que a nivel nacional llega a un total ponderado de **30,475,144** personas y a nivel de registro se tiene un total de **24,009,026** personas como se puede mostrar en el siguiente cuadro a nivel total y departamentos:

Cuadro 5: Población, Viviendas y Hogares por departamentos en la base del SISFOH

	Poblacion Ponderada 1/	Poblacion Registros	Viviendas Registros	Hogares Registros
Total	30,475,144	24,009,026	8,135,770	8,336,891
Departamentos				
Amazonas	419,404	348,444	116,170	117,529
Áncash	1,135,962	944,700	319,022	330,538
Apurímac	454,324	351,314	154,996	156,433
Arequipa	1,259,162	1,048,542	444,335	455,969
Ayacucho	673,609	524,400	239,351	243,098
Cajamarca	1,519,764	1,247,670	380,645	386,565
Callao	982,800	792,221	235,614	242,659
Cusco	1,300,609	942,239	343,275	345,780
Huancavelica	487,472	350,494	167,789	169,317
Huánuco	847,714	630,684	220,014	223,564
Ica	771,507	696,863	227,294	237,278
Junín	1,331,253	1,022,627	355,326	361,136
La Libertad	1,814,276	1,431,375	436,760	452,846
Lambayeque	1,239,882	1,020,488	305,253	314,275
Lima	9,540,996	7,362,581	2,379,174	2,448,257
Loreto	1,018,160	783,516	189,232	194,997
Madre De Dios	130,876	96,956	35,994	36,312
Moquegua	176,736	146,238	70,782	71,802
Pasco	299,807	205,913	74,047	75,170
Piura	1,814,622	1,582,669	447,308	462,372
Puno	1,389,684	970,599	481,557	485,283
San Martín	818,061	664,106	213,512	218,855
Tacna	333,276	255,357	115,565	120,769
Tumbes	231,480	194,652	70,033	71,051
Ucayali	483,708	394,378	112,722	115,036

Nota: 1/ Se excluyen las personas que son trabajadores del hogar, pensionistas y viviendas con más de 5 hogares

Fuente: Base de Datos del SISFOH

Elaboración: Propia

Observación: Debido a que se va a estimar el indicador de pobreza monetaria, entonces ello implicó que se tendría que manejar información de gastos de los hogares, por lo que del total de personas que se mostró en el cuadro anterior se hizo el filtro de usar hogares que informaron sobre sus gastos y de las siguientes restricciones:

- Se excluyeron las personas que estuvieron laborando como trabajadores del hogar
- Se excluyeron las personas que recibieron remuneración como pensionistas

- Se excluyeron las viviendas de más de cinco hogares para hacerlas compatible con la base de datos de la ENAHO 2012-2013

Por lo que nuestra población de referencia en la base de datos del SISFOH 2012-2013 registra **23,985,989** personas y la población ponderada llega a **30,473,524** personas que viven en **6,407,209** viviendas con **6,607,032** hogares en total como se muestra en el siguiente cuadro:

Cuadro 6: Población, Viviendas y Hogares con Información de Gastos por departamentos en la base del SISFOH

	Poblacion Ponderada 1/	Poblacion Registros	Viviendas Registros	Hogares Registros
Total	30,473,524	23,985,989	6,407,209	6,607,032
Departamentos				
Amazonas	419,360	348,261	90,292	91,636
Áncash	1,135,943	944,004	243,010	254,403
Apurímac	454,289	351,215	103,441	104,862
Arequipa	1,259,052	1,047,431	303,457	314,996
Ayacucho	673,533	524,167	155,900	159,615
Cajamarca	1,519,672	1,246,885	331,565	337,396
Callao	982,791	791,893	205,808	212,847
Cusco	1,300,355	941,417	272,239	274,719
Huancavelica	487,443	350,406	108,805	110,310
Huánuco	847,619	630,320	173,316	176,826
Ica	771,483	696,273	184,760	194,627
Junín	1,331,166	1,022,277	278,270	284,042
La Libertad	1,814,226	1,430,093	352,390	368,381
Lambayeque	1,239,871	1,019,336	248,343	257,296
Lima	9,540,836	7,351,872	1,941,551	2,010,546
Loreto	1,018,140	782,293	160,273	165,909
Madre De Dios	130,861	96,884	28,570	28,888
Moquegua	176,732	146,178	48,113	49,124
Pasco	299,798	205,866	56,106	57,224
Piura	1,814,573	1,581,946	378,977	394,010
Puno	1,389,401	969,996	340,538	344,244
San Martín	817,964	663,378	176,547	181,787
Tacna	333,247	254,943	79,562	84,686
Tumbes	231,473	194,597	51,868	52,881
Ucayali	483,694	394,058	93,508	95,777

Nota: 1/ Se excluyen las personas que son trabajadores del hogar, pensionistas, viviendas con más de 5 hogares y hogares que no registraron información de sus gastos

Fuente: Base de Datos del SISFOH

Elaboración: Propia

Sobre la Muestra:

La muestra de referencia ha de ser considerado sobre todos los registros de personas de la base de datos del ENAHO 2012-2013 que a nivel nacional llega a un total ponderado de **30,464,836** personas y a nivel de registro se tiene un total de **196,985** personas como se puede mostrar en el siguiente cuadro a nivel total y departamentos:

Cuadro 7: Población, Viviendas y Hogares por departamentos en la base de ENAHO

	Poblacion Ponderada 1/	Poblacion Registros	Viviendas Registros	Hogares Registros
Total	30,464,836	196,985	62,290	63,206
Departamentos				
Amazonas	419,404	7,712	2,247	2,273
Áncash	1,135,962	8,021	2,308	2,343
Apurímac	454,324	5,631	1,876	1,893
Arequipa	1,259,162	7,744	2,682	2,733
Ayacucho	672,924	6,764	2,374	2,419
Cajamarca	1,512,440	8,186	2,364	2,394
Callao	982,800	6,228	1,884	1,912
Cusco	1,300,609	7,584	2,618	2,658
Huancavelica	486,827	6,227	2,114	2,136
Huánuco	847,714	7,587	2,447	2,475
Ica	771,507	8,451	2,541	2,568
Junín	1,331,253	8,151	2,721	2,746
La Libertad	1,812,622	8,147	2,420	2,464
Lambayeque	1,239,882	8,844	2,323	2,342
Lima	9,540,996	24,284	7,979	8,088
Loreto	1,018,160	10,611	2,466	2,525
Madre De Dios	130,876	4,487	1,571	1,596
Moquegua	176,736	4,859	1,989	2,033
Pasco	299,807	5,365	1,888	1,907
Piura	1,814,622	9,331	2,516	2,550
Puno	1,389,684	6,369	2,676	2,722
San Martín	818,061	7,841	2,546	2,587
Tacna	333,276	4,878	2,009	2,051
Tumbes	231,480	5,407	1,596	1,621
Ucayali	483,708	8,276	2,135	2,170

Nota: 1/ Se excluyen a las personas que no son miembros del hogar y los hogares paneles

Fuente: Base de Datos del SISFOH

Elaboración: Propia

Observación: Debido a que se va a estimar el indicador de pobreza monetaria, entonces ello implicó que se tendría que manejar información de gastos de los hogares, por lo que del total de personas que se mostró en el cuadro anterior se hizo el filtro de usar hogares que informaron sobre sus gastos y de las siguientes restricciones:

- Se excluyeron las personas que no son miembros del hogar
- Se excluyeron las personas de los hogares paneles

Por lo que nuestra población de referencia en la base de datos del ENAHO 2012-2013 registra **186,119** personas y la población ponderada llega a **30,464,836** personas que viven en **48,077** viviendas con **48,310** hogares en total como se muestra en el siguiente cuadro:

Cuadro 8: Población, Viviendas y Hogares con Información de Gastos por departamentos en la base de ENAHO 2012-2013

	Poblacion Ponderada 1/	Poblacion Registros	Viviendas Registros	Hogares Registros
Total	30,464,836	186,119	48,077	48,310
Departamentos				
Amazonas	419,404	7,218	1,763	1,775
Áncash	1,135,962	7,532	1,904	1,911
Apurímac	454,324	5,379	1,441	1,446
Arequipa	1,259,162	7,308	2,055	2,066
Ayacucho	672,924	6,436	1,743	1,749
Cajamarca	1,512,440	7,867	1,963	1,972
Callao	982,800	6,003	1,523	1,530
Cusco	1,300,609	7,195	1,838	1,844
Huancavelica	486,827	5,892	1,575	1,579
Huánuco	847,714	7,139	1,805	1,808
Ica	771,507	8,060	2,174	2,186
Junín	1,331,253	7,737	2,121	2,130
La Libertad	1,812,622	7,782	1,983	1,991
Lambayeque	1,239,882	8,359	2,004	2,012
Lima	9,540,996	23,462	6,107	6,136
Loreto	1,018,160	9,501	1,965	1,974
Madre De Dios	130,876	4,037	1,087	1,102
Moquegua	176,736	4,582	1,403	1,410
Pasco	299,807	5,153	1,350	1,359
Piura	1,814,622	8,831	2,133	2,144
Puno	1,389,684	6,025	1,710	1,715
San Martín	818,061	7,303	1,892	1,905
Tacna	333,276	4,713	1,459	1,465
Tumbes	231,480	5,035	1,351	1,358
Ucayali	483,708	7,570	1,728	1,743

Nota: 1/ Se excluyen a las personas que no son miembros del hogar, hogares paneles y hogares que no registraron información de gastos

Fuente: Base de Datos del SISFOH

Elaboración: Propia

Sobre el tamaño de Muestra:

Como se explicará a más detalle en el punto 3.4 las simulaciones de muestra serán tomadas de la base del SISFOH 2012-2013 aproximadamente bajo el diseño de la ENAHO 2012-2013 y para propósitos de simplificación se utilizarán los conglomerados (campo nconglome1_18 en las 2 bases de datos) ya seleccionados en la base de ENAHO 2012-2013 y se sacarán aleatoriamente por cada conglomerado un número determinado de viviendas. Para determinar el número de viviendas que se tienen que elegir por conglomerado se tienen que observar primero la cantidad de conglomerados por departamento que se dan en los siguientes cuadros:

Cuadro 9: Total de conglomerados y hogares por departamento en la base del SISFOH

	Total de Conglomerados 1/	Cantidad de Hogares
Total	47,280	6,607,032
Departamentos		
Amazonas	801	91,636
Áncash	2,043	254,403
Apurímac	784	104,862
Arequipa	2,518	314,996
Ayacucho	1,339	159,615
Cajamarca	2,054	337,396
Callao	1,384	212,847
Cusco	2,033	274,719
Huancavelica	776	110,310
Huánuco	1,206	176,826
Ica	1,416	194,627
Junín	2,183	284,042
La Libertad	2,529	368,381
Lambayeque	1,614	257,296
Lima	14,033	2,010,546
Loreto	1,212	165,909
Madre De Dios	216	28,888
Moquegua	435	49,124
Pasco	507	57,224
Piura	2,626	394,010
Puno	2,501	344,244
San Martín	1,250	181,787
Tacna	690	84,686
Tumbes	385	52,881
Ucayali	745	95,777

Nota: 1/ Los conglomerados fueron tomados del campo nconglome1_18

Fuente: Base de Datos del SISFOH

Elaboración: Propia

Cuadro 10: Total de conglomerados y hogares por departamento en la base de ENAHO

	Total de Conglomerados 1/	Cantidad de Hogares
Total	5,848	48,310
Departamentos		
Amazonas	202	1,775
Áncash	218	1,911
Apurímac	155	1,446
Arequipa	262	2,066
Ayacucho	200	1,749
Cajamarca	221	1,972
Callao	200	1,530
Cusco	219	1,844
Huancavelica	163	1,579
Huánuco	206	1,808
Ica	267	2,186
Junín	250	2,130
La Libertad	257	1,991
Lambayeque	243	2,012
Lima	861	6,136
Loreto	226	1,974
Madre De Dios	109	1,102
Moquegua	173	1,410
Pasco	154	1,359
Piura	269	2,144
Puno	209	1,715
San Martín	232	1,905
Tacna	187	1,465
Tumbes	159	1,358
Ucayali	206	1,743

Nota: 1/ Los conglomerados fueron tomados del campo nconglome1_18

Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Entonces de los cuadros 8 y 9 se determina que los tamaños de muestra que se debieran sacar de la base del SISFOH 2012-2013 contendrán 5848 conglomerados y la cantidad de viviendas a extraer por cada conglomerado lo dará el cociente entre el número de viviendas de ENAHO y número de viviendas del SISFOH para un determinado conglomerado. Así por ejemplo en el cuadro siguiente el código de

conglomerado 3 posee 6 viviendas en la base de ENAHO y que representa aproximadamente un 3% del total de viviendas que es 193 viviendas en la base del SISFOH. El procedimiento consistirá en seleccionar aleatoriamente un porcentaje de viviendas dentro de cada conglomerado y para los 5848 conglomerados ya fijados.

Cuadro 11: Cantidad de viviendas a extraer por conglomerado para la simulación de la encuesta

Número de conglomerado	Código de conglomerado (nconglome1_18)	Número de viviendas ENAHO	Número de viviendas SISFOH	(Viviendas ENAHO)/ (viviendas SISFOH) * 100	1/
1	1	5	117	4	
2	3	6	193	3	
3	4	11	151	7	
4	5	6	173	3	
5	6	6	202	3	
6	7	12	177	7	
7	9	5	166	3	
8	20	16	109	15	
9	22	7	191	4	
10	23	8	152	5	
.	
.	
.	
5845	48026	6	92	7	
5846	48075	8	129	6	
5847	48078	8	163	5	
5848	48080	8	103	8	

Nota: 1/ se aproximó el cociente al valor entero más cercano

Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Luego de haber hecho las 500 simulaciones de encuestas a partir de la base del SISFOH se obtendrán las bases con las cuales se estimarán los diferentes modelos SAE. La siguiente tabla resumen muestra la distribución de hogares:

Cuadro 12: Cantidad de hogares seleccionados para cada simulación de muestra por departamento

	Simulacion 1	Simulacion 2	...	Simulacion 500
Total	47,268	47,268	...	47,268
Departamentos				
Amazonas	1,762	1,762	...	1,762
Áncash	1,867	1,867	...	1,867
Apurímac	1,444	1,444	...	1,444
Arequipa	2,026	2,026	...	2,026
Ayacucho	1,713	1,713	...	1,713
Cajamarca	1,972	1,972	...	1,972
Callao	1,459	1,459	...	1,459
Cusco	1,852	1,852	...	1,852
Huancavelica	1,552	1,552	...	1,552
Huánuco	1,802	1,802	...	1,802
Ica	2,104	2,104	...	2,104
Junín	2,106	2,106	...	2,106
La Libertad	1,960	1,960	...	1,960
Lambayeque	1,912	1,912	...	1,912
Lima	5,952	5,952	...	5,952
Loreto	1,869	1,869	...	1,869
Madre De Dios	1,071	1,071	...	1,071
Moquegua	1,374	1,374	...	1,374
Pasco	1,348	1,348	...	1,348
Piura	2,082	2,082	...	2,082
Puno	1,715	1,715	...	1,715
San Martín	1,858	1,858	...	1,858
Tacna	1,441	1,441	...	1,441
Tumbes	1,329	1,329	...	1,329
Ucayali	1,698	1,698	...	1,698

Fuente: Base de Datos del SISFOH 2012-2013

Elaboración: Propia

3.3 Técnicas e Instrumentos de recolección de datos

La información será tomada de 2 instrumentos: el primero será el SISFOH_2012-2013 (Empadronamiento Distrital de Población y Vivienda 2012-2013) que es una base de tipo censal donde se tiene información de los hogares a nivel nacional que utilizan los programas sociales, el segundo es la ENAHO_2012-2013 (Encuesta Nacional de Hogares 2012-2013) que es una base de datos tipo encuesta donde reúne la información consistenciada de la ENAHO 2012 y ENAHO 2013. Dichos instrumentos anteriores fueron elaborados por el INEI para el documento de Mapa de Pobreza 2013.

Sobre el SISFOH 2012-2013

El Empadronamiento Distrital de Población y Vivienda 2012-2013 (SISFOH) se llevó a cabo entre febrero del 2012 y setiembre de 2013. En total, se empadronaron a 24 009 026 millones de personas, siendo esta una población menor que la censada por el Censo de Población y Vivienda de 2007 (que contabilizó 27 412 157 millones de personas). A continuación se presentan algunas fichas que se utilizan para recojo de información:

Gráfico 12: Ficha Socioeconómica Unica del SISFOH - Carátula

I. LOCALIZACIÓN DE LA VIVIENDA		II. ENTREVISTA Y SUPERVISIÓN	
UBICACIÓN GEOGRÁFICA 1. DEPARTAMENTO 2. PROVINCIA 3. DISTRITO 4. CENTRO POBLADO 5. NÚCLEO URBANO (NO APLICABLE PARA CENTROS POBLADOS RURALES)		EMPADRONADOR(A) VISITA FECHA (dd-mm) RESULTADO VISITA (*) PRIMERA D D M M SEGUNDA D D M M TERCERA D D M M	
UBICACIÓN CENSAL 6. CONG. N° 7. ZONA N° 8. MANZANA N° DE 9. N° FRENTE DE 10. VIVIENDA N° 11. ¿CUANTOS HOGARES HABITAN EN ESTA VIVIENDA? 12. HOGAR N°		JEFE(A) DE BRIGADA VISITA FECHA (dd-mm) RESULTADO VISITA (*) PRIMERA D D M M SEGUNDA D D M M TERCERA D D M M	
13. APELLIDOS Y NOMBRES DEL INFORMANTE 14. DIRECCIÓN DE LA VIVIENDA TIPO DE VÍA: 1 ○ AVENIDA 2 ○ JIRÓN 3 ○ CALLE 4 ○ PASAJE 5 ○ CARRETERA 6 ○ OTRO NOMBRE DE LA VÍA (EN LOS CENTROS POBLADOS RURALES, INDICAR ALGUNA REFERENCIA DE UBICACIÓN) N° DE LA PUERTA BLOCK PISO INTERIOR MANZANA LOTE KM. TELEFONO DEL DOMICILIO		REVISOR(A) VISITA FECHA (dd-mm) RESULTADO REVISIÓN (*) PRIMERA D D M M 1 2 SEGUNDA D D M M 1 2 TERCERA D D M M 1 2	

Fuente: SISFOH

Elaboración: SISFOH

Gráfico 13: Ficha de Características de la vivienda del SISFOH

Pag. 02

III. CARACTERÍSTICAS DE LA VIVIENDA			
<p>1. TIPO DE VIVIENDA (Observe la vivienda y rellene sólo el óvalo que corresponda)</p> <p>1 <input type="radio"/> Casa independiente</p> <p>2 <input type="radio"/> Departamento en edificio</p> <p>3 <input type="radio"/> Vivienda en quinta</p> <p>4 <input type="radio"/> Vivienda en casa de vecindad (Callejón, solar o corralón)</p> <p>5 <input type="radio"/> Chozas o cabaña</p> <p>6 <input type="radio"/> Vivienda improvisada</p> <p>7 <input type="radio"/> Local no destinado para habitación humana</p> <p>8 <input type="radio"/> Otro (Especifique) _____</p> <p>_____</p>	<p>4. EL MATERIAL PREDOMINANTE EN LOS TECHOS ES: (Lea cada alternativa y rellene sólo el óvalo que corresponda)</p> <p>1 <input type="radio"/> Concreto armado?</p> <p>2 <input type="radio"/> Madera?</p> <p>3 <input type="radio"/> Tejas?</p> <p>4 <input type="radio"/> Planchas de calamina, fibra de cemento o similares?</p> <p>5 <input type="radio"/> Caña o estera con torta de barro?</p> <p>6 <input type="radio"/> Estera?</p> <p>7 <input type="radio"/> Paja, hojas de palmera?</p> <p>8 <input type="radio"/> Otro (Especifique) _____</p> <p>_____</p>	<p>7. EL ABASTECIMIENTO DE AGUA EN SU VIVIENDA PROCEDE DE: (Lea cada alternativa y rellene sólo el óvalo que corresponda)</p> <p>1 <input type="radio"/> Red pública dentro de la vivienda?</p> <p>2 <input type="radio"/> Red pública fuera de la vivienda, pero dentro del edificio?</p> <p>3 <input type="radio"/> Pilón de uso público?</p> <p>4 <input type="radio"/> Camión - cisterna u otro similar?</p> <p>5 <input type="radio"/> Pozo?</p> <p>6 <input type="radio"/> Río, acequia, manantial o similar?</p> <p>7 <input type="radio"/> Otro (Especifique) _____</p> <p>_____</p>	<p>3. ¿ SU HOGAR TIENE: (Lea cada alternativa y rellene sólo los óvalos que correspondan)</p> <p>1 <input type="radio"/> Equipo de sonido?</p> <p>2 <input type="radio"/> Televisor a color?</p> <p>3 <input type="radio"/> DVD?</p> <p>4 <input type="radio"/> Licuadora?</p> <p>5 <input type="radio"/> Refrigeradora / congeladora?</p> <p>6 <input type="radio"/> Cocina a gas?</p> <p>7 <input type="radio"/> Teléfono fijo?</p> <p>8 <input type="radio"/> Plancha eléctrica?</p> <p>9 <input type="radio"/> Lavadora?</p> <p>10 <input type="radio"/> Computadora?</p> <p>11 <input type="radio"/> Horno microondas?</p> <p>12 <input type="radio"/> Internet?</p> <p>13 <input type="radio"/> Cable?</p> <p>14 <input type="radio"/> Celular?</p> <p>15 <input type="radio"/> No tiene ninguno</p>
<p>2. SU VIVIENDA ES: (Lea cada alternativa y rellene sólo el óvalo que corresponda)</p> <p>1 <input type="radio"/> Alquilada?</p> <p>2 <input type="radio"/> Propia, pagándose a plazos?</p> <p>3 <input type="radio"/> Propia totalmente pagada?</p> <p>4 <input type="radio"/> Propia por inversión?</p> <p>5 <input type="radio"/> Cedida por el centro de trabajo?</p> <p>6 <input type="radio"/> Cedida por otro hogar o institución?</p> <p>7 <input type="radio"/> Otro (Especifique) _____</p> <p>_____</p>	<p>5. EL MATERIAL PREDOMINANTE EN LOS PISOS ES: (Lea cada alternativa y rellene sólo el óvalo que corresponda)</p> <p>1 <input type="radio"/> Parquet o madera pulida?</p> <p>2 <input type="radio"/> Láminas asfálticas, vinílicos o similares?</p> <p>3 <input type="radio"/> Losetas, terrazos o similares?</p> <p>4 <input type="radio"/> Madera (Entablados)?</p> <p>5 <input type="radio"/> Cemento?</p> <p>6 <input type="radio"/> Tierra?</p> <p>7 <input type="radio"/> Otro (Especifique) _____</p> <p>_____</p>	<p>8. EL SERVICIO HIGIÉNICO (Water, letrina, etc.) QUE TIENE SU VIVIENDA ESTA CONECTADO A: (Lea cada alternativa y rellene sólo el óvalo que corresponda)</p> <p>1 <input type="radio"/> Red pública dentro de la vivienda?</p> <p>2 <input type="radio"/> Red pública fuera de la vivienda, pero dentro del edificio?</p> <p>3 <input type="radio"/> Pozo séptico?</p> <p>4 <input type="radio"/> Pozo ciego o negro / letrina?</p> <p>5 <input type="radio"/> Río, acequia o canal?</p> <p>6 <input type="radio"/> NO TIENE</p>	<p>4. ¿CUÁL ES EL NÚMERO DE SU SUMINISTRO DE LUZ (1) O AGUA (2)?</p> <p>1 <input type="radio"/> Luz 2 <input type="radio"/> Agua 3 <input type="radio"/> NO TIENE</p> <p>_____</p>
IV. DATOS DEL HOGAR			
<p>1. SIN CONTAR BAÑO, COCINA, PASADIZOS, NI GARAJE ¿CUÁNTAS HABITACIONES OCUPA ESTE HOGAR? Anota la respuesta en los recuadros _____</p>			
<p>2. ¿CUÁL ES EL COMBUSTIBLE QUE MÁS SE UTILIZA EN EL HOGAR PARA COCINAR? (Lea cada alternativa y rellene sólo el óvalo que corresponda)</p> <p>1 <input type="radio"/> Electricidad?</p> <p>2 <input type="radio"/> Gas?</p> <p>3 <input type="radio"/> Keroseno?</p> <p>4 <input type="radio"/> Carbón?</p> <p>5 <input type="radio"/> Leña?</p> <p>6 <input type="radio"/> Bosta o estiércol?</p> <p>7 <input type="radio"/> Otro (Especifique) _____</p> <p>_____</p>			
<p>9. ¿CUÁNTAS HORAS DEMORAN EN LLEGAR DESDE SU VIVIENDA A LA CAPITAL DISTRITAL? Anota de 01 a 24 horas en el recadro _____ 1 <input type="radio"/> Más de 24 horas (Si es menor a una hora anote 00) 2 <input type="radio"/> Vive en la capital distrital</p>			
<p>5. ¿CUÁNTAS PERSONAS VIVEN PERMANENTEMENTE EN ESTE HOGAR? TOTAL = HOMBRES + MUJERES</p> <p>_____ = _____ + _____</p> <p>Incluya: - A las personas que viven permanentemente en este hogar. - A las personas que viven permanentemente en este hogar pero se encuentran ausentes temporalmente por trabajo, negocios, vacaciones, internados en un establecimiento de salud u otros motivos por un periodo igual o menor a seis (6) meses. No incluya: - A las personas que viven permanentemente en otro hogar. - A las personas que viven permanentemente en instituciones tales como hospitales psiquiátricos, instituciones de protección a menores, asilos, cárceles, conventos, etc. - A las personas que están ausentes del hogar por un periodo mayor a seis (6) meses.</p>			

Fuente: SISFOH

Elaboración: SISFOH

Gráfico 14: Ficha de Características de la Población del SISFOH

Pag. 03

V. CARACTERÍSTICAS DE LA POBLACIÓN															
NÚMERO DE ORDEN (01)	<p>Apellidos (En el primer renglón registre el apellido paterno y en el segundo el apellido materno, según DNI) Nombres (En el tercer renglón registre los nombres separándolos por un casillero, según DNI) Fecha de Nacimiento (dd-mm-aaaa) / Edad (Años cumplidos. Marque meses sólo para menores de 01 año) Tipo de documento: Seleccionar sólo una alternativa / Número de documento (escribir el número de documento. En el caso de la Partida de Nacimiento, poner sólo el número de CUI, de lo contrario no completar)</p>	Parentesco con el jefe de hogar	N° de núcleo familiar (0) es sin núcleo	Parentesco con el Jefe del Núcleo Familiar	Sexo (Si es M indicar si es gestante)	Estado Civil (12 y más años)	¿Qué tipo de seguro de salud tiene?	¿Sabe leer y escribir?	VII. EDUCACIÓN (3 años y más)		VIII. OCUPACIÓN		¿Presenta algún tipo de discapacidad?	¿De qué programa social es beneficiario en la actualidad?	
									Nivel educativo (1) ó (2) pasa a 17	Último año o grado de estudios aprobado	¿En el último mes era un...? (Para 6 años y más De (6) a (10) pase a 19	Sector en el que se desempeña (Para 6 años y más)			¿Presenta algún tipo de discapacidad?
		(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	
01	APPELLIDO PATERNO	1	2	3	4	5	6	7	1	2	3	4	5	6	7
	APPELLIDO MATERNO	8	9	10	11	12	13	14	1	2	3	4	5	6	7
	NOMBRES	15	16	17	18	19	20	21	1	2	3	4	5	6	7
	Fecha de Nacimiento	22	23	24	25	26	27	28	29	30	31	32	33	34	35
	Edad(años)	36	37	38	39	40	41	42	43	44	45	46	47	48	49
	Meses	50	51	52	53	54	55	56	57	58	59	60	61	62	63
	Tipo de Documento	64	65	66	67	68	69	70	71	72	73	74	75	76	77
02	APPELLIDO PATERNO	1	2	3	4	5	6	7	1	2	3	4	5	6	7
	APPELLIDO MATERNO	8	9	10	11	12	13	14	1	2	3	4	5	6	7
	NOMBRES	15	16	17	18	19	20	21	1	2	3	4	5	6	7
	Fecha de Nacimiento	22	23	24	25	26	27	28	29	30	31	32	33	34	35
	Edad(años)	36	37	38	39	40	41	42	43	44	45	46	47	48	49
	Meses	50	51	52	53	54	55	56	57	58	59	60	61	62	63
	Tipo de Documento	64	65	66	67	68	69	70	71	72	73	74	75	76	77

Fuente: SISFOH

Elaboración: SISFOH

Sobre la ENAHO 2012-2013

La Encuesta Nacional de Hogares (ENAHO) es una encuesta de derecho que se ejecuta a nivel nacional, durante todo el año. Dado que el SISFOH fue realizado durante los años 2012 y 2013, se compiló la base de datos anuales de los años 2012 y 2013, excluyéndose uno de los hogares que hacen parte del panel en ambos años. Se obtuvo una muestra total de 6 112 mil conglomerados, equivalente a 47 479 mil viviendas y 48 310 mil hogares. Gracias al mayor número de casos resultante de haber reunido dos años, se obtuvo una base de datos que permite tener estimados más robustos de los modelos econométricos desagregados geográficamente. A continuación se presenta algunas fichas que se utilizan para el recojo de información:

Gráfico 15: Carátula de la Encuesta Nacional de Hogares ENAHO

CONGLOMERADO N°		N° DE SELECCIÓN DE LA VIVIENDA	TIPO DE SELECCIÓN	¿Esta es una vivienda de reemplazo?	Anota el N° de Selección de la vivienda que reemplaza	ENAHO.01	
				SI 1 NO 2		Contiene: Características de la Vivienda, del Hogar y de los Miembros del Hogar, Gastos del Hogar, Programas Sociales, Participación Ciudadana y Medición del Oloro en los hogares.	
						CUESTIONARIO N°	Cuestionario Adicional
							1

UBICACIÓN GEOGRÁFICA				UBICACIÓN MUESTRAL			
1. DEPARTAMENTO				5. ZONA N°			
2. PROVINCIA				6. MANZANA N°			
3. DISTRITO				7. AER N°			
4. CENTRO POBLADO				8. VIVIENDA N°			

9. DIRECCIÓN DE LA VIVIENDA							
Nombre de la Calle, Av., Jr., Carretera, etc.	N°	INT.	PISO	MZ.	LOTE	KM.	TELÉFONO

10. TOTAL DE HOGARES QUE OCUPAN LA VIVIENDA		11. HOGAR N°
Sr(a): Si HOGAR es la persona o grupo de personas que se alimentan de una misma olla y atienden en común otras necesidades básicas.		
¿Cuántos hogares ocupan esta vivienda?		

12. NOMBRES Y APELLIDOS DEL JEFE(A) DEL HOGAR	

13. ENTREVISTA Y SUPERVISIÓN										
VISITA	ENCUESTADOR(A)				SUPERVISOR(A) LOCAL					
	FECHA	HORA		PRÓXIMA VISITA		RESULTADO DE LA VISITA (*)	FECHA	HORA		RESULTADO DE LA VISITA (*)
		DE	A	FECHA	HORA			DE	A	
Primera										
Segunda										
Tercera										
Cuarta										
Quinta										
Sexta										

14. RESULTADO FINAL DE LA ENCUESTA		(*) CÓDIGOS DE RESULTADO		
FECHA		1. COMPLETA	4. AUSENTE	7. OTRO
RESULTADO		2. INCOMPLETA	5. VIVIENDA DESOCUPADA	(Especifique)
		3. RECHAZO	6. NO SE INICIÓ LA ENTREVISTA	

Fuente: ENAHO

Elaboración: ENAHO

Gráfico 16: Ficha de Características de la Vivienda y del Hogar de ENAHO

100. CARACTERÍSTICAS DE LA VIVIENDA Y DEL HOGAR							
DATOS DE LA VIVIENDA							
101. TIPO DE VIVIENDA:							
Casa independiente.....	1						
Departamento en edificio.....	2						
Vivienda en quinta.....	3						
Vivienda en casa de vecindad (callejón, solar o corralón).....	4						
Chozo o cabaña.....	5						
Vivienda improvisada.....	6						
Local no destinado para habitación humana.....	7						
Otro.....	8						
(Especifique)							
102A. EL MATERIAL PREDOMINANTE EN LAS PAREDES EXTERIORES ES:							
¿Ladrillo o bloque de cemento?.....	1						
¿Piedra o sillar con cal o cemento?.....	2						
¿Adobe?.....	3						
¿Tapia?.....	4						
¿Quincha (caña con barro)?.....	5						
¿Piedra con barro?.....	6						
¿Madera?.....	7						
¿Estera?.....	8						
¿Otro material?.....	9						
(Especifique)							
103. EL MATERIAL PREDOMINANTE EN LOS PISOS ES:							
¿Parquet o madera pulida?.....	1						
¿Láminas asfálticas, vinílicos o similares?.....	2						
¿Losetas, terrazos o similares?.....	3						
¿Madera (entablados)?.....	4						
¿Cemento?.....	5						
¿Tierra?.....	6						
¿Otro material?.....	7						
(Especifique)							
103A. EL MATERIAL PREDOMINANTE EN LOS TECHOS ES:							
¿Concreto armado?.....	1						
104B. ¿LA VIVIENDA :							
1. Cuenta con licencia de construcción?.....	<table border="1"> <tr> <td>Si</td> <td>No</td> <td>No sabe</td> </tr> <tr> <td>1</td> <td>2</td> <td>3</td> </tr> </table>	Si	No	No sabe	1	2	3
Si	No	No sabe					
1	2	3					
2. Fue construida con asistencia técnica de un Ing. civil o arquitecto?.....	<table border="1"> <tr> <td>Si</td> <td>No</td> <td>No sabe</td> </tr> <tr> <td>1</td> <td>2</td> <td>3</td> </tr> </table>	Si	No	No sabe	1	2	3
Si	No	No sabe					
1	2	3					
DATOS DEL HOGAR							
105. LA VIVIENDA QUE OCUPA SU HOGAR ES:							
¿Alquilada?.....	1 → <input type="text"/>						
¿Propia, totalmente pagada?.....	2 → PASE A 107B						
¿Propia, por invasión?.....	3						
¿Propia, comprándola a plazos?.....	4 → <input type="text"/>						
¿Cedida por centro de trabajo?.....	5						
¿Cedida por otro hogar o institución?.....	6						
¿Otra forma?.....	7						
(Especifique)							
106. SI UD. ALQUILARA ESTA VIVIENDA, ¿CUÁNTO CREE QUE LE PAGARÍAN DE ALQUILER MENSUAL?							
Si.....	<input type="text"/>						
<p>A. Encuestador: Transcriba el código de 105.</p> <p><input type="text"/> → Igual a 2, 3 ó 4 continúe con 106A.</p> <p>↓</p> <p>↓ Igual a: 5, 6 ó 7 → PASE A 107B</p>							
106A. ¿ESTA VIVIENDA TIENE TÍTULO DE PROPIEDAD?							
106B. ¿El título de la vivienda está Registrado en la SUNARP?							
Si.....	1 → <table border="1"><tr><td>Si</td><td>No</td></tr><tr><td>1</td><td>2</td></tr></table>	Si	No	1	2		
Si	No						
1	2						
No.....	2						
En trámite de titulación.....	3						

Fuente: ENAHO
Elaboración: ENAHO

Gráfico 17: Ficha de Características de los Miembros del Hogar

200. CARACTERÍSTICAS DE LOS MIEMBROS DEL HOGAR														
Informante N°														
(201)	(202)	(203)	NUCLEOS FAMILIARES		(204)	(205)	(206)	(207)	(208)	(208-A)		PARA 12 AÑOS Y MÁS DE EDAD (209)		
N° DE ORD.	¿CUÁL ES EL NOMBRE Y APELLIDOS DE CADA UNA DE LAS PERSONAS QUE VIVEN PERMANENTEMENTE EN ESTE HOGAR Y LAS QUE ESTÁN ALOJADAS AQUÍ? (NO OLVIDE REGISTRAR A LOS MIEMBROS DEL HOGAR AUSENTES Y RECÍEN NACIDOS)	¿CUÁL ES LA RELACION DE PARENTESCO CON EL JEFE(A) DEL HOGAR?	(203-A) N° DE NÚCLEO FAMILIAR	(203-B) RELACION DE PARENTESCO CON EL JEFE(A) DEL NÚCLEO FAMILIAR (Sólo relación Padres e Hijos)	¿ES MIEMBRO DEL HOGAR?	¿SE ENCUENTRA EN EL HOGAR 30 DÍAS O MÁS?	¿ESTA PRESENTE EN EL HOGAR 30 DÍAS O MÁS?	SEXO	¿QUE EDAD TIENE EN AÑOS CUMPLIDOS?	¿EN QUE PROVINCIA Y DISTRITO NACIÓ?		¿CUÁL ES SU ESTADO CIVIL O CONYUGAL?		
	NOMBRE	APELLIDOS	CÓDIGO	CÓDIGO	SI NO	SI NO	SI NO	Hombre	Mujer	Años	Meses	PROVINCIA	DISTRITO	CÓDIGO
1			1		1 2	1 2	1 2	1 2						1
2					1 2	1 2	1 2	1 2						1
3					1 2	1 2	1 2	1 2						1
4					1 2	1 2	1 2	1 2						1
5					1 2	1 2	1 2	1 2						1
6					1 2	1 2	1 2	1 2						1

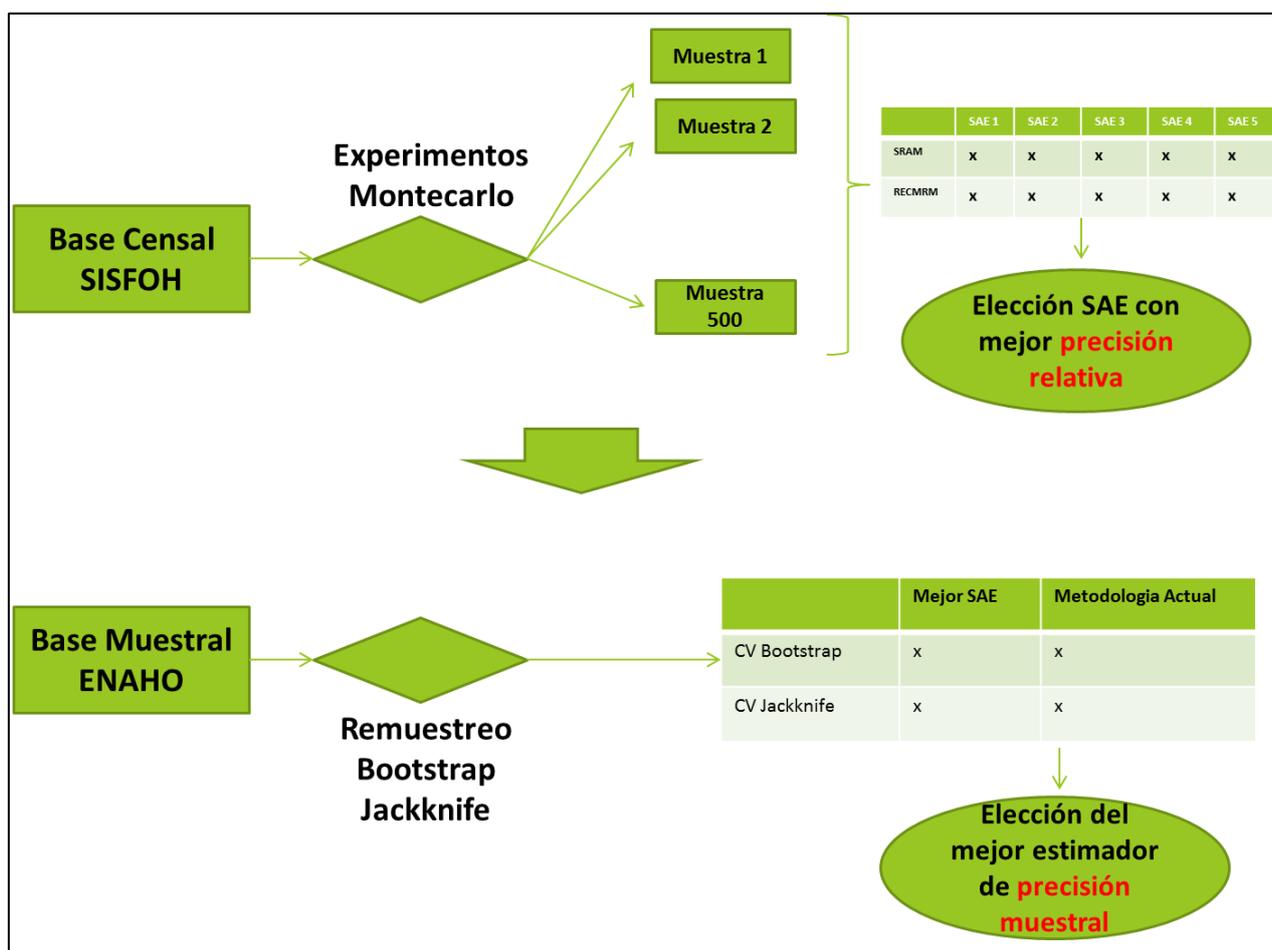
Fuente: ENAHO
Elaboración: ENAHO

3.4 Técnicas de análisis e interpretación de datos

- Se utilizó la simulación Montecarlo para obtener 500 muestras del SISFOH_2012-2013 que sean semejantes al diseño muestral de la ENAHO 2012-2013.
- Se calculó los indicadores de evaluación con las 500 muestras por cada metodología y se eligió el método que posea mejores indicadores de evaluación
- Del método elegido anteriormente se calculó su error estándar pero con la encuesta ENAHO_2012-2013 y así se halló su coeficiente de variación por los métodos de remuestreo bootstrap y jackknife.

A continuación en el gráfico 9 se presenta un resumen del procedimiento

Gráfico 18: Resumen del Procedimiento para el Análisis de Datos



Fuente: Propia

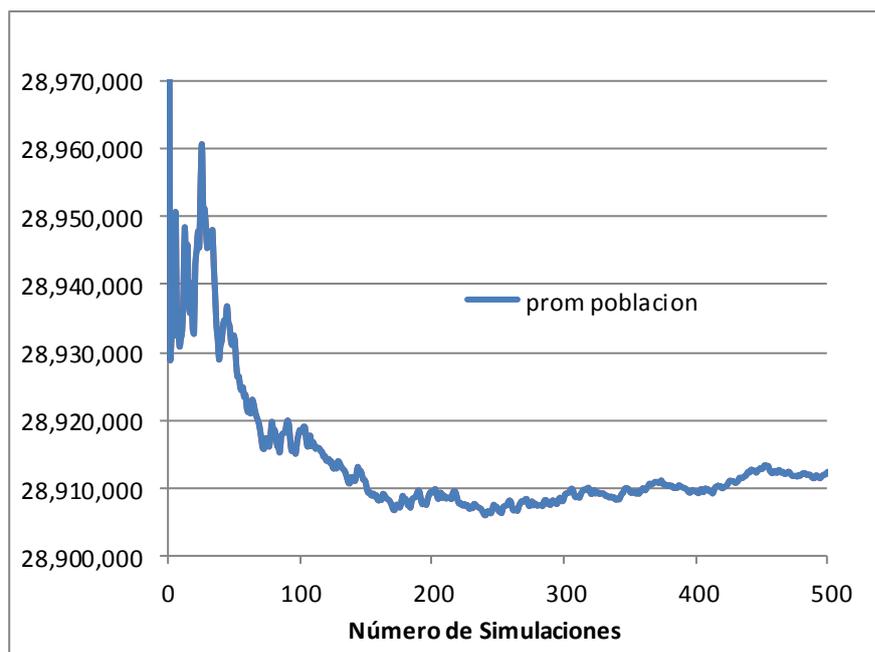
Elaboración: Propia

CAPITULO IV RESULTADOS

Sobre la determinación de la cantidad de simulaciones Montecarlo en las Muestras

Del capítulo 3.2 se observó que la cantidad de personas ponderada con la base ENAHO 2012-2013 fue de 30 464 836 personas a nivel nacional en 48077 registros de vivienda, luego las 500 muestras simuladas desde la base del SISFOH 2012-2013 contenían 47268 registros de vivienda, entonces lo que se analizará a continuación será la convergencia del estimado ponderado de la población total para las simulaciones.

Gráfico 19: Convergencia en Población Total

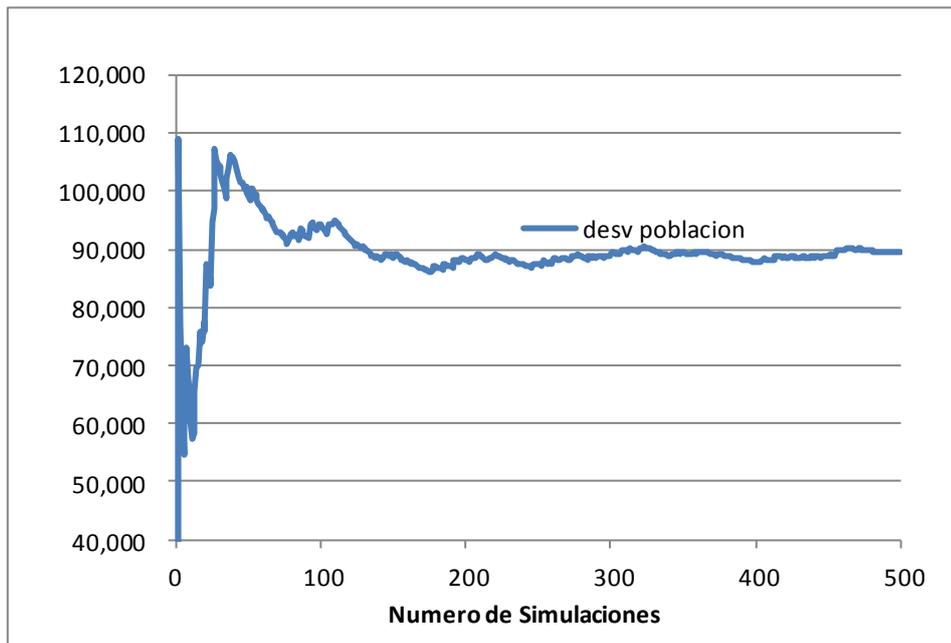


Fuente: Base de Datos del SISFOH 2012-2013 – 500 simulaciones de muestra

Elaboración: Propia

Como se puede observar en el gráfico anterior con 100 simulaciones de muestra aún no se observa una convergencia del total de personas pues la población total ponderada tiende a disminuir. Es a partir de 200 simulaciones en adelante que ya se observa una convergencia del total poblacional, y el valor de convergencia es de aproximadamente 28 910 000 personas.

Gráfico 20: Convergencia en Desviación Poblacional Total



Fuente: Base de Datos del SISFOH 2012-2013 – 500 simulaciones de muestra

Elaboración: Propia

En el gráfico anterior se puede apreciar que a partir de 200 simulaciones el estimado de la desviación poblacional total converge al valor de 90 000

Sobre los estimados de pobreza monetaria con las bases de ENAHO_2012-2013 y SISFOH_2012-2013

En los cuadros y gráficos siguientes se analizará el indicador de la pobreza monetaria para los instrumentos de la ENAHO 2012-2013 y SISFOH 2012-2013.

Cuadro 13: Pobreza Monetaria por departamentos con la base ENAHO 2012-2013

Departamentos	Pobreza Monetaria	Error Típico	Intervalo de Confianza al 95%	
Amazonas	0.47	0.02	0.43	0.51
Áncash	0.24	0.02	0.20	0.28
Apurímac	0.47	0.03	0.42	0.52
Arequipa	0.10	0.01	0.08	0.12
Ayacucho	0.53	0.02	0.48	0.57
Cajamarca	0.54	0.02	0.49	0.58
Callao	0.15	0.01	0.12	0.18
Cusco	0.19	0.02	0.16	0.23
Huancavelica	0.48	0.02	0.44	0.53
Huánuco	0.43	0.02	0.39	0.48
Ica	0.06	0.01	0.05	0.08
Junín	0.20	0.02	0.17	0.24
La Libertad	0.32	0.02	0.28	0.36
Lambayeque	0.22	0.02	0.19	0.26
Lima	0.13	0.01	0.12	0.15
Loreto	0.36	0.02	0.32	0.40
Madre De Dios	0.02	0.01	0.01	0.04
Moquegua	0.09	0.01	0.06	0.11
Pasco	0.44	0.02	0.39	0.49
Piura	0.34	0.02	0.29	0.38
Puno	0.33	0.02	0.29	0.38
San Martín	0.30	0.02	0.26	0.33
Tacna	0.13	0.02	0.10	0.16
Tumbes	0.12	0.01	0.09	0.14
Ucayali	0.12	0.01	0.10	0.15

Fuente: Base de Datos del ENAHO 2012-2013

Elaboración: Propia

El presente cuadro se observa que los departamentos con más pobreza monetaria son los departamentos de Cajamarca y Ayacucho con 0.54 y 0.53 de pobreza monetaria respectivamente, es decir, el 54 % y 53% de la población total de los respectivos departamentos están en la condición de pobres monetarios.

Cuadro 14: Pobreza Monetaria por departamentos con base del SISFOH 2012-2013

Departamentos	Pobreza Monetaria	Error Típico	Intervalo de Confianza al 95%	
Amazonas	0.49	0.01	0.47	0.50
Áncash	0.27	0.01	0.25	0.28
Apurímac	0.49	0.01	0.47	0.50
Arequipa	0.12	0.01	0.11	0.13
Ayacucho	0.51	0.01	0.50	0.53
Cajamarca	0.56	0.01	0.54	0.58
Callao	0.19	0.01	0.18	0.20
Cusco	0.24	0.01	0.22	0.25
Huancavelica	0.48	0.01	0.46	0.51
Huánuco	0.45	0.01	0.43	0.47
Ica	0.07	0.01	0.06	0.08
Junín	0.23	0.01	0.21	0.24
La Libertad	0.32	0.01	0.30	0.33
Lambayeque	0.23	0.01	0.22	0.25
Lima	0.15	0.00	0.14	0.16
Loreto	0.38	0.01	0.36	0.39
Madre De Dios	0.04	0.00	0.03	0.05
Moquegua	0.12	0.01	0.10	0.13
Pasco	0.48	0.01	0.46	0.50
Piura	0.37	0.01	0.35	0.38
Puno	0.38	0.01	0.36	0.39
San Martín	0.31	0.01	0.29	0.33
Tacna	0.16	0.01	0.15	0.17
Tumbes	0.13	0.01	0.11	0.15
Ucayali	0.13	0.01	0.11	0.14

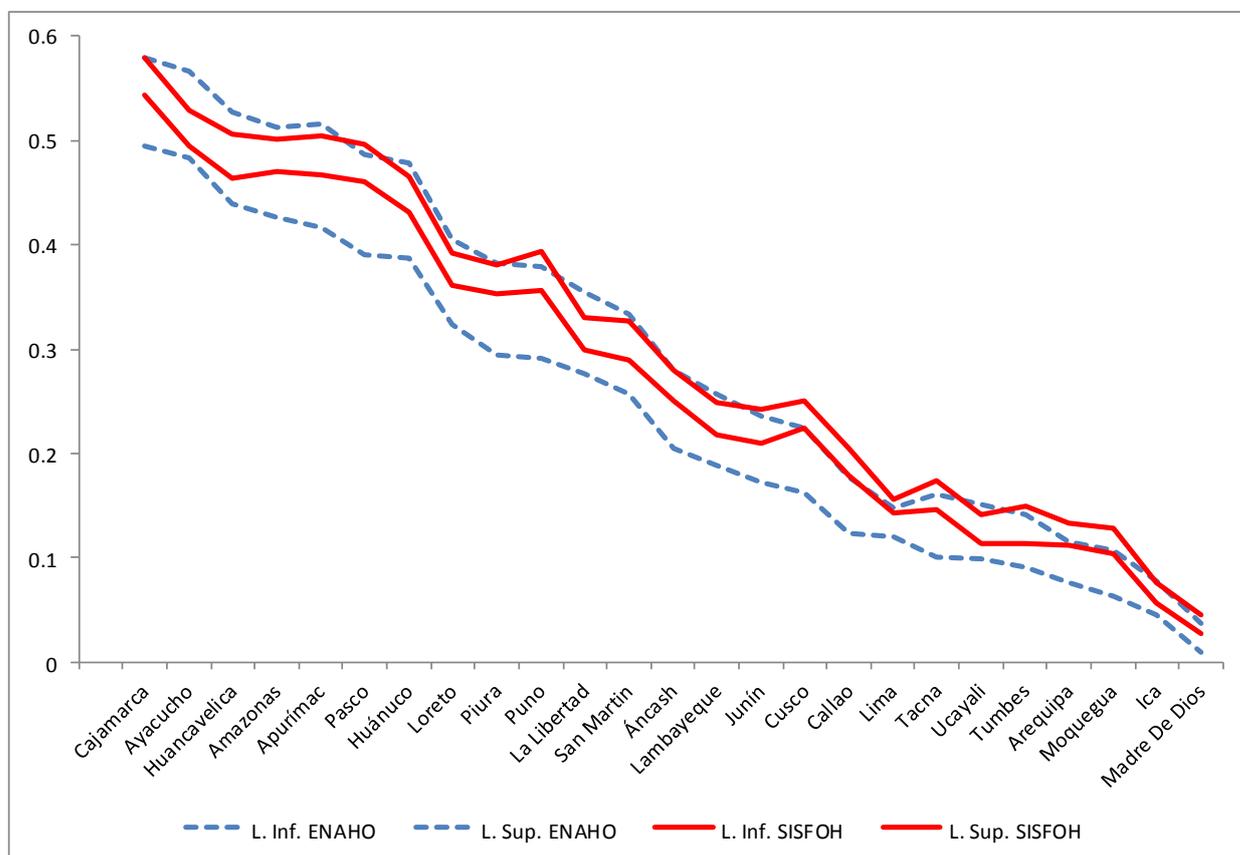
Fuente: Base de Datos del SISFOH 2012-2013

Elaboración: Propia

El presente cuadro se observa que los departamentos con más pobreza monetaria son los departamentos de Cajamarca y Ayacucho con 0.56 y 0.51 de pobreza monetaria respectivamente, es decir, el 56 % y 53% de la población total de los respectivos departamentos están en la condición de pobres monetarios.

decir, el 56 % y 51% de la población total de los respectivos departamentos están en la condición de pobres monetarios.

Gráfico 21: Intervalo de Confianza al 95% de la Pobreza Monetaria de la ENAHO y SISFOH por departamentos



Fuente: Base de Datos de ENAHO 2012-2013 Y SISFOH 2012-2013

Elaboración: Propia

Como se puede apreciar en la mayoría de departamentos el intervalo de confianza al 95% del SISFOH 2012-2013 caen dentro del intervalo de confianza al 95% de la ENAHO 2012-2013, esto es algo de esperarse pues al ser la ENAHO una muestra se obtendrá mayores errores de estimación.

Sobre los resultados de los Indicadores de Evaluación de las metodologías SAE con las 500 muestras simuladas

En los cuadros y gráficos siguientes se analizará el desempeño de los indicadores de evaluación de la pobreza monetaria para 5 metodologías SAE y utilizando las 500 muestras simuladas con el SISFOH 2012-2013.

En el cuadro 15 se presenta los 5 tipos de metodologías en análisis y los estimadores propuestos por cada tipo de metodología. Como se podrá apreciar para las metodologías Post Estratificado y Sintético se presentará 2 estimadores porque se le evaluó con 2 variables distintas (Edad-sexo y estrato socioeconómico). Para el caso del estimador combinado además de ser analizado con dos variables se le adicionó un cambio en el parámetro del modelo (tipo 3 y tipo 4).

En total se analizaron 10 estimadores en las 5 metodologías SAE.

Cuadro 15: Tipos de estimadores SAE analizados

Metodología SAE	Nombre del Estimador	Codigo del Estimador
Estimador Directo	Estimador directo	EST 1
Estimador Post Estratificado	Estimador post estratificado con la variable agrupada edad sexo	EST 2
	Estimador post estratificado con la variable estrato socioeconomico	EST 3
Estimador Sintético	Estimador sintetico para la variable agrupada edad sexo	EST 4
	Estimador sintetico para la variable estrato socioeconomico	EST 5
Estimador Combinado	Estimador combinado para la variable estrato socioeconomico tipo 3	EST 6
	Estimador combinado para la variable estrato socioeconomico tipo 4	EST 7
	Estimador combinado para la variable edad sexo tipo 3	EST 8
	Estimador combinado para la variable edad sexo tipo 4	EST 9
Estimador Basado en Modelo	Estimador basado en modelo logit	EST 10

Fuente: Propia

Elaboración: Propia

Cuadro 16: Indicadores de Evaluación del Estimador EST 1 por departamentos y total

Departamentos	sra	recmr	sram	recmr
Amazonas	4.03	4.47	0.39	0.43
Áncash	11.05	11.26	0.39	0.43
Apurímac	7.33	7.55	0.39	0.43
Arequipa	14.62	15.01	0.39	0.43
Ayacucho	1.71	2.53	0.39	0.43
Cajamarca	7.22	7.34	0.39	0.43
Callao	12.92	13.33	0.39	0.43
Cusco	10.40	10.65	0.39	0.43
Huancavelica	9.66	9.86	0.39	0.43
Huánuco	8.69	8.85	0.39	0.43
Ica	9.10	11.02	0.39	0.43
Junín	2.16	3.86	0.39	0.43
La Libertad	4.52	4.91	0.39	0.43
Lambayeque	5.00	5.95	0.39	0.43
Lima	9.75	9.95	0.39	0.43
Loreto	2.95	3.51	0.39	0.43
Madre De Dios	19.36	20.91	0.39	0.43
Moquegua	13.98	14.39	0.39	0.43
Pasco	1.26	2.15	0.39	0.43
Piura	5.50	5.78	0.39	0.43
Puno	9.08	9.24	0.39	0.43
San Martín	0.18	2.73	0.39	0.43
Tacna	13.66	14.05	0.39	0.43
Tumbes	4.65	6.39	0.39	0.43
Ucayali	6.52	7.64	0.39	0.43

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Ancash, Arequipa, Callao, Cuzco, Madre de Dios, Moquegua y Tacna presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- El departamento de San Martín presenta un sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 17: Indicadores de Evaluación del Estimador EST 2 por departamentos y total

Departamentos	sra	recmr	sram	recmr
Amazonas	4.04	4.49	0.39	0.43
Áncash	11.05	11.26	0.39	0.43
Apurímac	7.34	7.55	0.39	0.43
Arequipa	14.64	15.03	0.39	0.43
Ayacucho	1.70	2.52	0.39	0.43
Cajamarca	7.22	7.34	0.39	0.43
Callao	12.92	13.33	0.39	0.43
Cusco	10.43	10.68	0.39	0.43
Huancavelica	9.66	9.87	0.39	0.43
Huánuco	8.70	8.86	0.39	0.43
Ica	9.10	11.02	0.39	0.43
Junín	2.16	3.86	0.39	0.43
La Libertad	4.53	4.92	0.39	0.43
Lambayeque	5.00	5.95	0.39	0.43
Lima	9.75	9.95	0.39	0.43
Loreto	2.97	3.53	0.39	0.43
Madre De Dios	19.38	20.93	0.39	0.43
Moquegua	13.98	14.39	0.39	0.43
Pasco	1.25	2.14	0.39	0.43
Piura	5.50	5.78	0.39	0.43
Puno	9.08	9.24	0.39	0.43
San Martín	0.18	2.73	0.39	0.43
Tacna	13.67	14.06	0.39	0.43
Tumbes	4.66	6.39	0.39	0.43
Ucayali	6.53	7.64	0.39	0.43

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Ancash, Arequipa, Callao, Cuzco, Madre de Dios, Moquegua y Tacna presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- El departamento de San Martín presenta un sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 18: Indicadores de Evaluación del Estimador EST 3 por departamentos y total

Departamentos	sra	recmr	sram	recmr
Amazonas	4.12	4.56	0.39	0.43
Áncash	11.01	11.23	0.39	0.43
Apurímac	7.38	7.59	0.39	0.43
Arequipa	14.66	15.06	0.39	0.43
Ayacucho	1.66	2.49	0.39	0.43
Cajamarca	7.25	7.36	0.39	0.43
Callao	12.92	13.33	0.39	0.43
Cusco	10.52	10.77	0.39	0.43
Huancavelica	9.68	9.89	0.39	0.43
Huánuco	8.74	8.89	0.39	0.43
Ica	9.10	11.02	0.39	0.43
Junín	2.16	3.86	0.39	0.43
La Libertad	4.55	4.94	0.39	0.43
Lambayeque	5.01	5.95	0.39	0.43
Lima	9.76	9.96	0.39	0.43
Loreto	3.02	3.57	0.39	0.43
Madre De Dios	19.36	20.91	0.39	0.43
Moquegua	14.00	14.40	0.39	0.43
Pasco	1.17	2.10	0.39	0.43
Piura	5.44	5.72	0.39	0.43
Puno	9.09	9.25	0.39	0.43
San Martín	0.19	2.73	0.39	0.43
Tacna	13.70	14.09	0.39	0.43
Tumbes	4.67	6.40	0.39	0.43
Ucayali	6.53	7.65	0.39	0.43

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Ancash, Arequipa, Callao, Cuzco, Madre de Dios, Moquegua y Tacna presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- El departamento de San Martín presentan sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 19: Indicadores de Evaluación del Estimador EST 4 por departamentos y total

Departamentos	sra	recmr	sram	recmrn
Amazonas	1.52	1.53	0.24	0.25
Áncash	4.63	4.63	0.24	0.25
Apurímac	7.13	7.13	0.24	0.25
Arequipa	3.50	3.53	0.24	0.25
Ayacucho	0.17	0.28	0.24	0.25
Cajamarca	0.33	0.36	0.24	0.25
Callao	9.57	9.58	0.24	0.25
Cusco	11.18	11.18	0.24	0.25
Huancavelica	3.67	3.68	0.24	0.25
Huánuco	3.44	3.44	0.24	0.25
Ica	11.11	11.12	0.24	0.25
Junín	0.76	0.85	0.24	0.25
La Libertad	3.65	3.67	0.24	0.25
Lambayeque	0.65	0.74	0.24	0.25
Lima	8.99	8.99	0.24	0.25
Loreto	2.59	2.60	0.24	0.25
Madre De Dios	2.30	2.35	0.24	0.25
Moquegua	8.53	8.53	0.24	0.25
Pasco	0.49	0.53	0.24	0.25
Piura	4.68	4.68	0.24	0.25
Puno	3.27	3.28	0.24	0.25
San Martín	8.35	8.35	0.24	0.25
Tacna	5.58	5.59	0.24	0.25
Tumbes	8.61	8.62	0.24	0.25
Ucayali	7.51	7.52	0.24	0.25

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Cuzco e Ica presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- Los departamentos de Ayacucho, Cajamarca, Junín, Lambayeque y Pasco presentan sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 20: Indicadores de Evaluación del Estimador EST 5 por departamentos y total

Departamentos	sra	recmr	sram	recmr
Amazonas	0.11	0.47	0.22	0.22
Áncash	8.11	8.14	0.22	0.22
Apurímac	2.10	2.17	0.22	0.22
Arequipa	4.64	4.71	0.22	0.22
Ayacucho	3.49	3.53	0.22	0.22
Cajamarca	7.51	7.54	0.22	0.22
Callao	6.48	6.52	0.22	0.22
Cusco	0.79	1.10	0.22	0.22
Huancavelica	3.60	3.63	0.22	0.22
Huánuco	10.00	10.02	0.22	0.22
Ica	2.62	2.69	0.22	0.22
Junín	6.43	6.46	0.22	0.22
La Libertad	5.79	5.84	0.22	0.22
Lambayeque	1.73	1.98	0.22	0.22
Lima	1.39	1.45	0.22	0.22
Loreto	2.33	2.44	0.22	0.22
Madre De Dios	4.08	4.14	0.22	0.22
Moquegua	3.81	4.01	0.22	0.22
Pasco	1.92	2.00	0.22	0.22
Piura	9.23	9.25	0.22	0.22
Puno	0.58	0.92	0.22	0.22
San Martín	0.70	0.94	0.22	0.22
Tacna	8.99	9.05	0.22	0.22
Tumbes	7.56	7.59	0.22	0.22
Ucayali	3.66	3.76	0.22	0.22

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- El departamento de Huánuco presenta diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) en mayor a 10.
- El departamento de Amazonas, Cuzco, Puno y San Martín presenta sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 21: Indicadores de Evaluación del Estimador EST 6 por departamentos y total

Departamentos	sra	recmr	sram	recmrm
Amazonas	2.71	3.04	0.32	0.34
Áncash	10.04	10.16	0.32	0.34
Apurímac	4.22	4.40	0.32	0.34
Arequipa	11.32	11.57	0.32	0.34
Ayacucho	2.27	2.61	0.32	0.34
Cajamarca	7.34	7.39	0.32	0.34
Callao	10.78	11.01	0.32	0.34
Cusco	6.75	6.94	0.32	0.34
Huancavelica	7.66	7.78	0.32	0.34
Huánuco	9.16	9.24	0.32	0.34
Ica	6.94	8.11	0.32	0.34
Junín	3.58	4.21	0.32	0.34
La Libertad	4.97	5.16	0.32	0.34
Lambayeque	3.92	4.55	0.32	0.34
Lima	6.97	7.11	0.32	0.34
Loreto	1.24	1.87	0.32	0.34
Madre De Dios	14.27	15.22	0.32	0.34
Moquegua	10.60	10.89	0.32	0.34
Pasco	0.14	1.24	0.32	0.34
Piura	6.70	6.82	0.32	0.34
Puno	6.26	6.38	0.32	0.34
San Martín	0.36	1.95	0.32	0.34
Tacna	12.13	12.36	0.32	0.34
Tumbes	5.63	6.37	0.32	0.34
Ucayali	3.14	4.15	0.32	0.34

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Ancash, Arequipa, Callao, Madre de Dios, Moquegua y Tacna presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- Los departamentos de Pasco y San Martín presentan sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 22: Indicadores de Evaluación del Estimador EST 7 por departamentos y total

Departamentos	sra	recmr	sram	recmrm
Amazonas	2.01	2.28	0.28	0.30
Áncash	9.56	9.64	0.28	0.30
Apurímac	2.64	2.82	0.28	0.30
Arequipa	9.65	9.83	0.28	0.30
Ayacucho	2.57	2.76	0.28	0.30
Cajamarca	7.38	7.42	0.28	0.30
Callao	9.70	9.86	0.28	0.30
Cusco	4.87	5.04	0.28	0.30
Huancavelica	6.64	6.73	0.28	0.30
Huánuco	9.37	9.42	0.28	0.30
Ica	5.86	6.68	0.28	0.30
Junín	4.29	4.63	0.28	0.30
La Libertad	5.17	5.30	0.28	0.30
Lambayeque	3.37	3.86	0.28	0.30
Lima	5.58	5.68	0.28	0.30
Loreto	0.35	1.22	0.28	0.30
Madre De Dios	11.72	12.39	0.28	0.30
Moquegua	8.90	9.14	0.28	0.30
Pasco	0.37	1.06	0.28	0.30
Piura	7.34	7.41	0.28	0.30
Puno	4.84	4.94	0.28	0.30
San Martín	0.44	1.59	0.28	0.30
Tacna	11.35	11.51	0.28	0.30
Tumbes	6.11	6.53	0.28	0.30
Ucayali	1.44	2.56	0.28	0.30

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Madre de Dios, y Tacna presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- Los departamentos de Loreto, Pasco y San Martín presentan sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 23: Indicadores de Evaluación del Estimador EST 8 por departamentos y total

Departamentos	sra	recmr	sram	recmrm
Amazonas	3.20	3.47	0.29	0.31
Áncash	8.91	9.03	0.29	0.31
Apurímac	2.52	2.80	0.29	0.31
Arequipa	8.59	8.91	0.29	0.31
Ayacucho	1.19	1.75	0.29	0.31
Cajamarca	4.93	5.01	0.29	0.31
Callao	11.81	12.01	0.29	0.31
Cusco	3.23	3.59	0.29	0.31
Huancavelica	7.67	7.79	0.29	0.31
Huánuco	6.95	7.04	0.29	0.31
Ica	9.77	10.63	0.29	0.31
Junín	1.19	2.50	0.29	0.31
La Libertad	1.80	2.24	0.29	0.31
Lambayeque	3.12	3.81	0.29	0.31
Lima	9.50	9.59	0.29	0.31
Loreto	2.84	3.13	0.29	0.31
Madre De Dios	13.69	14.68	0.29	0.31
Moquegua	12.16	12.38	0.29	0.31
Pasco	0.99	1.54	0.29	0.31
Piura	5.22	5.36	0.29	0.31
Puno	4.96	5.10	0.29	0.31
San Martín	2.90	3.44	0.29	0.31
Tacna	10.97	11.19	0.29	0.31
Tumbes	5.98	6.67	0.29	0.31
Ucayali	1.85	3.29	0.29	0.31

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Callao, Madre de Dios, Moquegua y Tacna presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- El departamento de Pasco presenta sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 24: Indicadores de Evaluación del Estimador EST 9 por departamentos y total

Departamentos	sra	recmr	sram	recmrm
Amazonas	2.78	2.97	0.25	0.27
Áncash	7.84	7.92	0.25	0.27
Apurímac	0.11	0.96	0.25	0.27
Arequipa	5.57	5.86	0.25	0.27
Ayacucho	0.94	1.36	0.25	0.27
Cajamarca	3.78	3.84	0.25	0.27
Callao	11.25	11.37	0.25	0.27
Cusco	0.37	1.28	0.25	0.27
Huancavelica	6.67	6.75	0.25	0.27
Huánuco	6.07	6.13	0.25	0.27
Ica	10.10	10.59	0.25	0.27
Junín	0.70	1.84	0.25	0.27
La Libertad	0.44	1.13	0.25	0.27
Lambayeque	2.17	2.74	0.25	0.27
Lima	9.37	9.43	0.25	0.27
Loreto	2.78	2.96	0.25	0.27
Madre De Dios	10.84	11.57	0.25	0.27
Moquegua	11.25	11.39	0.25	0.27
Pasco	0.87	1.26	0.25	0.27
Piura	5.09	5.17	0.25	0.27
Puno	2.91	3.05	0.25	0.27
San Martín	4.27	4.49	0.25	0.27
Tacna	9.62	9.77	0.25	0.27
Tumbes	6.64	7.01	0.25	0.27
Ucayali	0.49	2.15	0.25	0.27

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Callao, Ica, Madre de Dios, y Moquegua presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- Los departamentos de Apurímac, Ayacucho, Cuzco, Junín, La Libertad, Pasco y Ucayali presentan sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 25: Indicadores de Evaluación del Estimador EST 10 por departamentos y total

Departamentos	sra	recmr	sram	recmr
Amazonas	4.31	5.09	0.48	0.74
Áncash	1.96	4.76	0.48	0.74
Apurímac	1.33	2.91	0.48	0.74
Arequipa	40.02	41.84	0.48	0.74
Ayacucho	4.72	5.34	0.48	0.74
Cajamarca	3.38	3.75	0.48	0.74
Callao	3.46	10.24	0.48	0.74
Cusco	13.17	14.07	0.48	0.74
Huancavelica	0.95	3.18	0.48	0.74
Huánuco	4.63	5.49	0.48	0.74
Ica	3.76	23.51	0.48	0.74
Junín	1.76	5.52	0.48	0.74
La Libertad	11.20	11.54	0.48	0.74
Lambayeque	3.35	6.65	0.48	0.74
Lima	31.49	32.23	0.48	0.74
Loreto	1.15	3.27	0.48	0.74
Madre De Dios	5.43	60.93	0.48	0.74
Moquegua	39.26	42.94	0.48	0.74
Pasco	5.45	5.89	0.48	0.74
Piura	8.82	9.24	0.48	0.74
Puno	7.70	8.52	0.48	0.74
San Martín	13.67	14.50	0.48	0.74
Tacna	19.08	22.18	0.48	0.74
Tumbes	1.28	13.91	0.48	0.74
Ucayali	7.45	10.33	0.48	0.74

Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los departamentos de Arequipa, Cuzco, La Libertad, Lima, Moquegua, San Martín y Tacna presentan diferencias con respecto a su valor verdadero de pobreza monetaria pues su indicador de sesgo relativo (SRA) y error cuadrático relativo (RECMR) son en su mayoría mayores a 10.
- El departamento de Huancavelica presenta sesgo relativo menor a 1 lo que significa que los estimados de pobreza monetaria con las bases simuladas están muy próximas al valor verdadero.

Cuadro 26: Análisis descriptivo del Sesgo Relativo por Tipo de Estimador

Tipo de Estimador	Cantidad de Departamentos	
	SRA > 10	SRA < 1
<i>EST 1</i>	7	1
<i>EST 2</i>	7	1
<i>EST 3</i>	7	1
<i>EST 4</i>	2	5
<i>EST 5</i>	1	4
<i>EST 6</i>	6	2
<i>EST 7</i>	2	3
<i>EST 8</i>	4	1
<i>EST 9</i>	4	7
<i>EST 10</i>	7	1

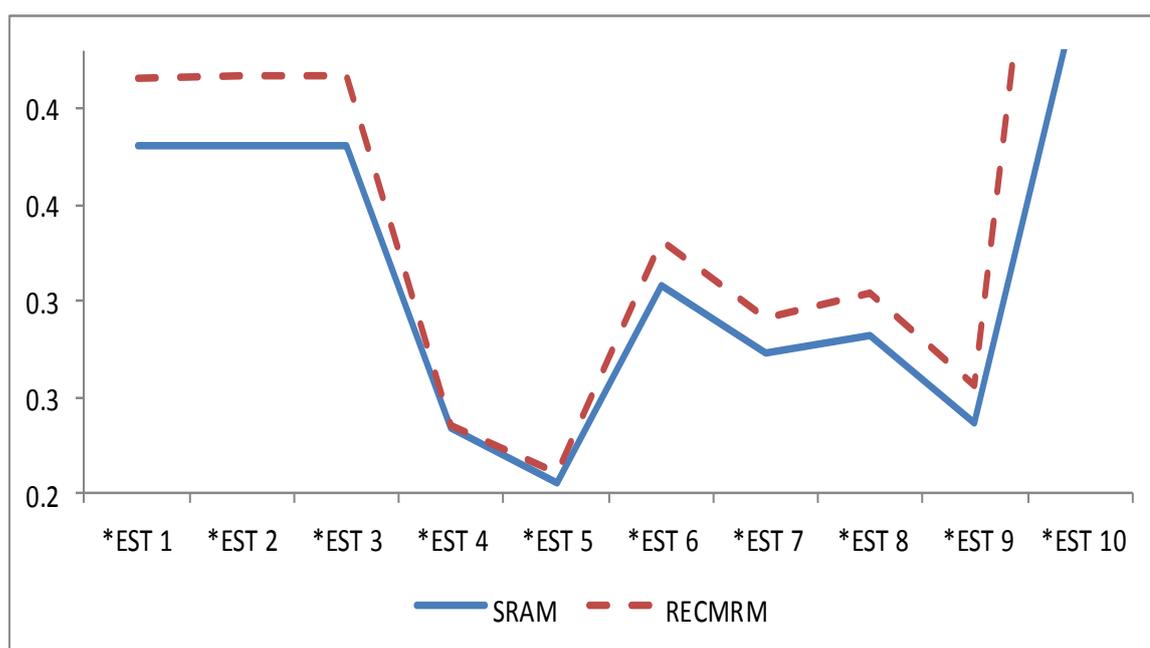
Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Intuitivamente para escoger el mejor estimador sería aquel en donde tenga menos departamentos con sesgo relativo > 10 y aquel estimador que tenga mayor cantidad de departamentos con sesgo relativo < 1.
- En el cuadro anterior se tienen 4 estimadores que podrían cumplir los requisitos mencionados anteriormente y son el “Estimador Sintético con la variable edad-sexo” (EST 4), “Estimador Sintético con la variable estrato socioeconómico” (EST 5), “Estimador Combinado de la variable estrato socioeconómico tipo 4” (EST 7) y “Estimador Combinado de la variable edad-sexo tipo4” (EST 9)

Gráfico 22: Resultados de los indicadores SRAM, RECMRM por tipo de estimador para la Población Total



Fuente: Base de Datos de las 500 simulaciones de muestras del SISFOH 2012-2013

Elaboración: Propia

Observación:

- Los estimadores Directo y Post Estratificado (EST 1, EST 2 y EST 3) poseen similares valores en los indicadores SRAM y RECMRM
- Los estimadores Sintéticos (EST 4 y EST 5) poseen los menores valores en los indicadores SRAM y RECMRM.
- El estimador basado en Modelo Logit (EST 10) presenta los valores más altos de los indicadores SRAM y RECMRM.

Sobre los resultados de los coeficientes de variación con las metodologías de los Estimadores Sintéticos

Como se analizó en el apartado anterior los estimadores Sintéticos poseen características deseables en cuanto a precisión estadística. En los siguientes cuadros y gráficos se comparará los coeficientes de variación de estos dos estimadores sintéticos (EST 4 y EST 5) en la base ENAHO 2012-2013 mediante las técnicas de remuestreo de Jackknife y Bootstrap.

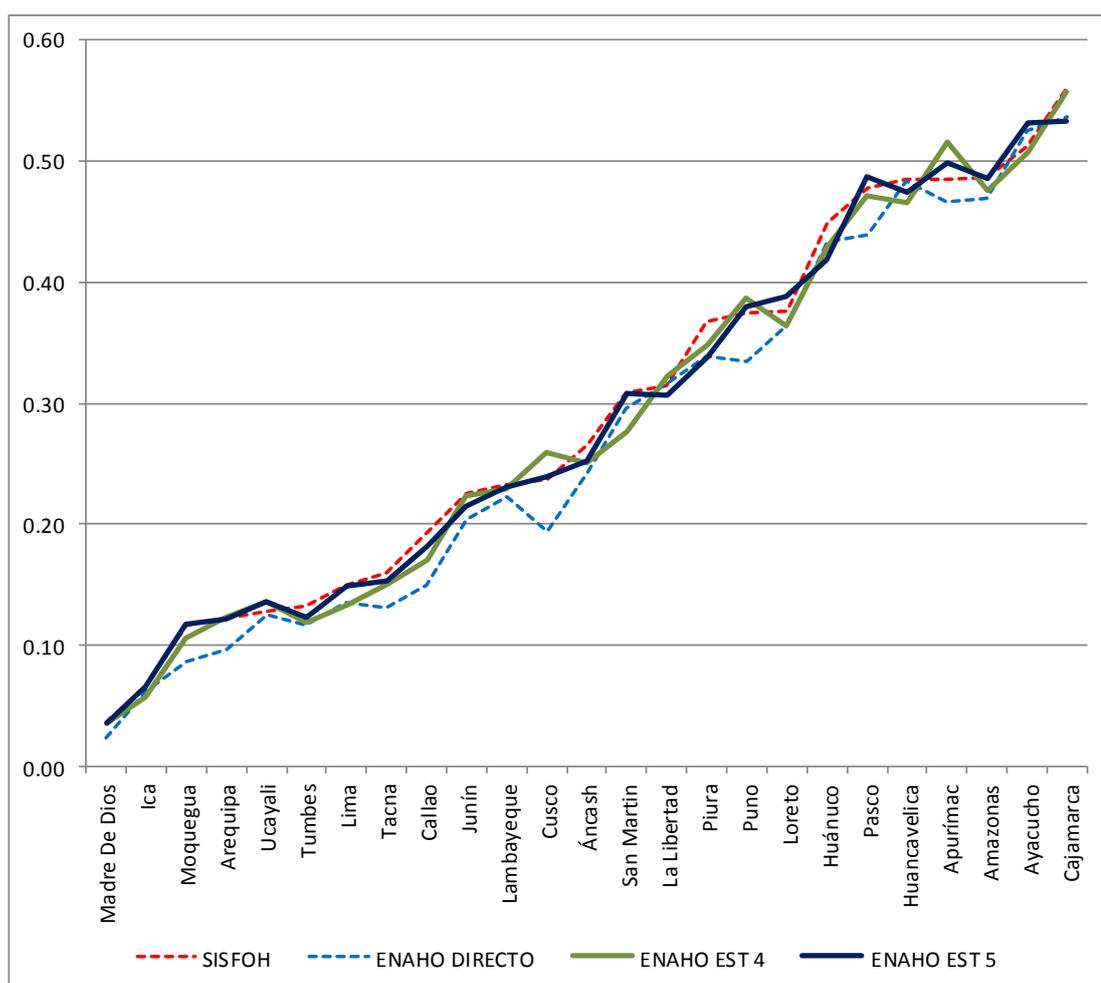
Cuadro 27: Estimados de Pobreza Monetaria con las metodologías Directa y Sintética por Departamentos

DEPARTAMENTOS	SISFOH	ENAHO DIRECTO	ENAHO EST 4	ENAHO EST 5
Amazonas	0.49	0.47	0.48	0.49
Áncash	0.27	0.24	0.25	0.25
Apurímac	0.49	0.47	0.52	0.50
Arequipa	0.12	0.10	0.12	0.12
Ayacucho	0.51	0.53	0.51	0.53
Cajamarca	0.56	0.54	0.56	0.53
Callao	0.19	0.15	0.17	0.18
Cusco	0.24	0.19	0.26	0.24
Huancavelica	0.48	0.48	0.47	0.47
Huánuco	0.45	0.43	0.43	0.42
Ica	0.07	0.06	0.06	0.07
Junín	0.23	0.20	0.22	0.21
La Libertad	0.32	0.32	0.32	0.31
Lambayeque	0.23	0.22	0.23	0.23
Lima	0.15	0.13	0.13	0.15
Loreto	0.38	0.36	0.36	0.39
Madre De Dios	0.04	0.02	0.04	0.04
Moquegua	0.12	0.09	0.11	0.12
Pasco	0.48	0.44	0.47	0.49
Piura	0.37	0.34	0.35	0.34
Puno	0.38	0.33	0.39	0.38
San Martín	0.31	0.30	0.28	0.31
Tacna	0.16	0.13	0.15	0.15
Tumbes	0.13	0.12	0.12	0.12
Ucayali	0.13	0.12	0.14	0.14

Fuente: Base de Datos del SISFOH 2012-2013 y ENAHO 2012-2013

Elaboración: Propia

Gráfico 23: Estimados de Pobreza Monetaria por departamentos



Fuente: Base de Datos del SISFOH 2012-2013 y ENAHO 2012-2013

Elaboración: Propia

Observación:

- El presente gráfico muestra el resultado de las estimaciones del indicador de pobreza monetaria a nivel de departamento de las bases de datos del SISFOH 2012-2013 y ENAHO 2012-2013, bajo las metodologías Directa (EST 1) y Sintética (EST 4 y EST 5).
- Para los departamentos de Cuzco y Puno es de observarse que existe una diferencia de 5 puntos porcentuales entre el SISFOH y ENAHO, por ello que se aprecia cierto alejamiento en dichos departamentos.
- El departamento con más pobreza es el de Cajamarca con aproximadamente 54% de personas pobres en el departamento.
- El departamento de Madre de Dios es el de menos pobreza con apenas 2% de personas pobres

Cuadro 28: Pobreza Monetaria, Errores Típicos y Coeficiente de Variación por departamento

Departamentos	Pobreza Monetaria	Error Típico Jackknife	Coeficiente Variación Jackknife	Error Típico Bootstrap	Coeficiente Variación Bootstrap
Amazonas	0.47	0.007	1.4	0.007	1.4
Áncash	0.24	0.005	2.1	0.005	2.1
Apurímac	0.47	0.008	1.6	0.008	1.6
Arequipa	0.10	0.004	3.8	0.004	3.7
Ayacucho	0.53	0.007	1.3	0.007	1.3
Cajamarca	0.54	0.007	1.3	0.006	1.2
Callao	0.15	0.005	3.1	0.005	3.1
Cusco	0.19	0.005	2.5	0.005	2.5
Huancavelica	0.48	0.007	1.4	0.007	1.5
Huánuco	0.43	0.006	1.5	0.006	1.5
Ica	0.06	0.003	5.2	0.003	5.1
Junín	0.20	0.005	2.6	0.005	2.7
La Libertad	0.32	0.006	1.9	0.006	1.9
Lambayeque	0.22	0.006	2.5	0.005	2.4
Lima	0.13	0.002	1.8	0.003	1.9
Loreto	0.36	0.005	1.4	0.005	1.4
Madre De Dios	0.02	0.003	10.6	0.003	10.5
Moquegua	0.09	0.004	5.0	0.004	5.1
Pasco	0.44	0.008	1.8	0.008	1.8
Piura	0.34	0.006	1.8	0.006	1.9
Puno	0.33	0.007	2.1	0.007	2.0
San Martín	0.30	0.007	2.3	0.007	2.3
Tacna	0.13	0.005	4.2	0.005	4.1
Tumbes	0.12	0.006	4.7	0.005	4.6
Ucayali	0.12	0.004	3.1	0.004	3.0

Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Observación:

- El presente cuadro muestran los errores típicos asociados a la estimación de la pobreza monetaria en la muestra de ENAHO 2012-2013 bajo la metodología clásica o también conocida como Directa y las 2 tipos de remuestro (Jackknife y Bootstrap)
- Se pueden apreciar que los coeficientes de variación de los dos tipos de remuestro son muy semejantes entre sí, lo cual valida más aún el estimar la precisión del indicador de pobreza monetaria.

Cuadro 29: Pobreza Monetaria, Error Típico Jackknife y Coeficiente de Variación del Estimador Sintético EST 4 por departamentos

Departamentos	Pobreza Monetaria	Error Típico Jackknife	Coeficiente de Variación EST 4
Amazonas	0.48	0.001	0.22
Áncash	0.25	0.001	0.24
Apurímac	0.52	0.001	0.19
Arequipa	0.12	0.001	0.43
Ayacucho	0.51	0.001	0.20
Cajamarca	0.56	0.001	0.15
Callao	0.17	0.001	0.31
Cusco	0.26	0.001	0.30
Huancavelica	0.47	0.001	0.20
Huánuco	0.43	0.001	0.19
Ica	0.06	0.000	0.47
Junín	0.22	0.001	0.39
La Libertad	0.32	0.001	0.32
Lambayeque	0.23	0.001	0.34
Lima	0.13	0.000	0.18
Loreto	0.36	0.001	0.27
Madre De Dios	0.04	0.000	0.56
Moquegua	0.11	0.000	0.31
Pasco	0.47	0.001	0.21
Piura	0.35	0.001	0.23
Puno	0.39	0.001	0.25
San Martín	0.28	0.001	0.26
Tacna	0.15	0.001	0.36
Tumbes	0.12	0.000	0.33
Ucayali	0.14	0.001	0.41

Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Observación:

- El presente cuadro muestra la pobreza monetaria, error típico y su coeficiente de variación del estimador Sintético de variables Edad-Sexo por departamento
- El error típico fue hallado por el remuestreo de Jackknife.
- El departamento de Madre de Dios presenta el mayor coeficiente de variación con 0.56
- El departamento de Cajamarca presenta el menor coeficiente de variación con 0.15.

Cuadro 30: Pobreza Monetaria, Error Típico Bootstrap y Coeficiente de Variación del Estimador Sintético EST 4 por departamentos

Departamentos	Pobreza Monetaria	Error Típico Bootstrap	Coeficiente de Variación EST 4
Amazonas	0.48	0.001	0.21
Áncash	0.25	0.001	0.26
Apurímac	0.52	0.001	0.19
Arequipa	0.12	0.000	0.40
Ayacucho	0.51	0.001	0.20
Cajamarca	0.56	0.001	0.15
Callao	0.17	0.000	0.29
Cusco	0.26	0.001	0.29
Huancavelica	0.47	0.001	0.20
Huánuco	0.43	0.001	0.18
Ica	0.06	0.000	0.48
Junín	0.22	0.001	0.38
La Libertad	0.32	0.001	0.32
Lambayeque	0.23	0.001	0.35
Lima	0.13	0.000	0.18
Loreto	0.36	0.001	0.27
Madre De Dios	0.04	0.000	0.56
Moquegua	0.11	0.000	0.30
Pasco	0.47	0.001	0.21
Piura	0.35	0.001	0.23
Puno	0.39	0.001	0.25
San Martín	0.28	0.001	0.26
Tacna	0.15	0.001	0.36
Tumbes	0.12	0.000	0.32
Ucayali	0.14	0.001	0.41

Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Observación:

- El presente cuadro muestra la pobreza monetaria, error típico y su coeficiente de variación del estimador Sintético de variables Edad-Sexo por departamento
- El error típico fue hallado por el remuestreo de Bootstrap.
- El departamento de Madre de Dios presenta el mayor coeficiente de variación con 0.56
- El departamento de Cajamarca presenta el menor coeficiente de variación con 0.15.

Cuadro 31: Pobreza Monetaria, Error Típico Jackknife y Coeficiente de Variación del Estimador Sintético EST 5 por departamentos

Departamentos	Pobreza Monetaria	Error Típico Jackknife	Coeficiente de Variación EST 5
Amazonas	0.49	0.002	0.51
Áncash	0.25	0.002	0.76
Apurímac	0.50	0.003	0.56
Arequipa	0.12	0.001	0.89
Ayacucho	0.53	0.003	0.48
Cajamarca	0.53	0.003	0.61
Callao	0.18	0.001	0.78
Cusco	0.24	0.002	0.77
Huancavelica	0.47	0.002	0.44
Huánuco	0.42	0.002	0.54
Ica	0.07	0.000	0.71
Junín	0.21	0.001	0.68
La Libertad	0.31	0.003	0.87
Lambayeque	0.23	0.002	0.93
Lima	0.15	0.001	0.42
Loreto	0.39	0.002	0.60
Madre De Dios	0.04	0.000	0.77
Moquegua	0.12	0.002	1.38
Pasco	0.49	0.003	0.59
Piura	0.34	0.002	0.57
Puno	0.38	0.003	0.69
San Martín	0.31	0.002	0.70
Tacna	0.15	0.002	1.15
Tumbes	0.12	0.001	0.76
Ucayali	0.14	0.001	0.84

Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Observación:

- El presente cuadro muestra la pobreza monetaria, error típico y su coeficiente de variación del estimador Sintético de variable estrato socioeconómico por departamento
- El error típico fue hallado por el remuestreo de Jackknife.
- El departamento de Moquegua presenta el mayor coeficiente de variación con 1.38
- El departamento de Lima presenta el menor coeficiente de variación con 0.42.

Cuadro 32: Pobreza Monetaria, Error Típico Bootstrap y Coeficiente de Variación del Estimador Sintético EST 5 por departamentos

Departamentos	Pobreza Monetaria	Error Típico Bootstrap	Coeficiente de Variación EST 5
Amazonas	0.49	0.002	0.50
Áncash	0.25	0.002	0.75
Apurímac	0.50	0.003	0.54
Arequipa	0.12	0.001	0.93
Ayacucho	0.53	0.003	0.50
Cajamarca	0.53	0.003	0.60
Callao	0.18	0.001	0.77
Cusco	0.24	0.002	0.75
Huancavelica	0.47	0.002	0.44
Huánuco	0.42	0.002	0.54
Ica	0.07	0.000	0.73
Junín	0.21	0.001	0.70
La Libertad	0.31	0.003	0.89
Lambayeque	0.23	0.002	0.97
Lima	0.15	0.001	0.43
Loreto	0.39	0.002	0.59
Madre De Dios	0.04	0.000	0.81
Moquegua	0.12	0.002	1.36
Pasco	0.49	0.003	0.57
Piura	0.34	0.002	0.57
Puno	0.38	0.003	0.69
San Martín	0.31	0.002	0.71
Tacna	0.15	0.002	1.21
Tumbes	0.12	0.001	0.78
Ucayali	0.14	0.001	0.83

Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Observación:

- El presente cuadro muestra la pobreza monetaria, error típico y su coeficiente de variación del estimador Sintético de variable estrato socioeconómico por departamento
- El error típico fue hallado por el remuestreo de Bootstrap.
- El departamento de Moquegua presenta el mayor coeficiente de variación con 1.36
- El departamento de Lima presenta el menor coeficiente de variación con 0.43.

Cuadro 33: Coeficientes de Variación de los Estimadores Directos y Sintéticos por departamentos

DEPARTAMENTOS	CV DIR JACKKNIFE	CV DIR BOOTSTRAP	CV EST 4 JACKKNIFE	CV EST 4 BOOTSTRAP	CV EST 5 JACKKNIFE	CV EST 5 BOOTSTRAP
Amazonas	1.46	1.45	0.22	0.21	0.51	0.50
Áncash	2.18	2.10	0.24	0.26	0.76	0.75
Apurímac	1.61	1.63	0.19	0.19	0.56	0.54
Arequipa	3.89	3.80	0.43	0.40	0.89	0.93
Ayacucho	1.34	1.33	0.20	0.20	0.48	0.50
Cajamarca	1.26	1.21	0.15	0.15	0.61	0.60
Callao	3.16	3.13	0.31	0.29	0.78	0.77
Cusco	2.59	2.51	0.30	0.29	0.77	0.75
Huancavelica	1.44	1.46	0.20	0.20	0.44	0.44
Huánuco	1.47	1.48	0.19	0.18	0.54	0.54
Ica	5.42	5.27	0.47	0.48	0.71	0.73
Junín	2.67	2.75	0.39	0.38	0.68	0.70
La Libertad	1.89	1.87	0.32	0.32	0.87	0.89
Lambayeque	2.55	2.49	0.34	0.35	0.93	0.97
Lima	1.88	1.90	0.18	0.18	0.42	0.43
Loreto	1.47	1.46	0.27	0.27	0.60	0.59
Madre De Dios	11.44	11.38	0.56	0.56	0.77	0.81
Moquegua	5.24	5.39	0.31	0.30	1.38	1.36
Pasco	1.86	1.79	0.21	0.21	0.59	0.57
Piura	1.84	1.92	0.23	0.23	0.57	0.57
Puno	2.08	2.04	0.25	0.25	0.69	0.69
San Martín	2.37	2.37	0.26	0.26	0.70	0.71
Tacna	4.16	4.13	0.36	0.36	1.15	1.21
Tumbes	4.96	4.82	0.33	0.32	0.76	0.78
Ucayali	3.19	3.14	0.41	0.41	0.84	0.83

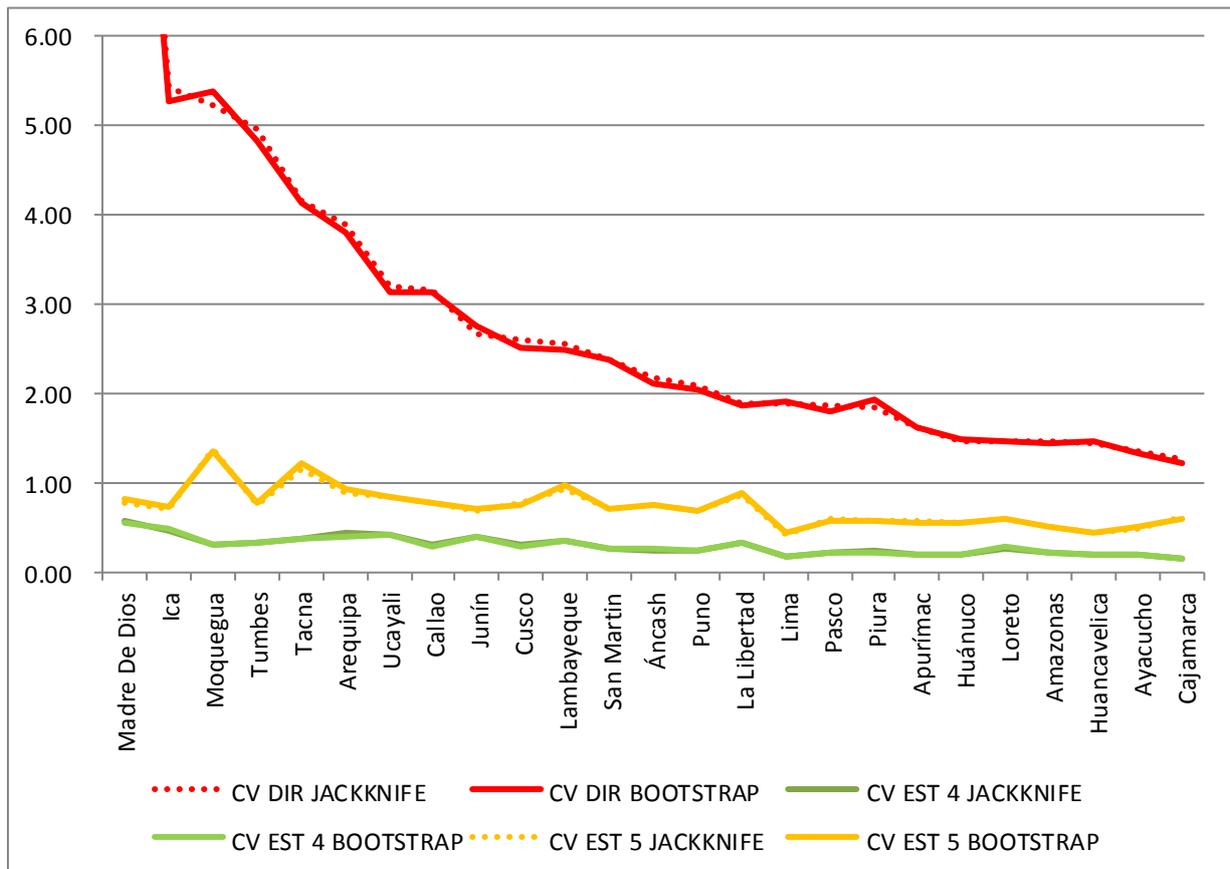
Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Observación:

- El presente cuadro muestra los coeficientes de variación de los estimadores Directos (EST 1) y Sintéticos (EST 4 y EST 5) mediante remuestreo de Jackknife y Bootstrap.
- El coeficiente de variación de la metodología clásica, la que actualmente utilizan las oficinas de estadística para estimar la pobreza monetaria es la que presenta en el cuadro anterior los mayores valores de coeficiente de variación.
- Dentro de los estimadores Sintéticos el que mejor posee características deseables de precisión es el Estimador Sintético de variable Edad-Sexo pues posee menores valores de coeficiente de variación.

Gráfico 24: Coeficientes de Variación de los estimadores Directos y Sintéticos por departamentos



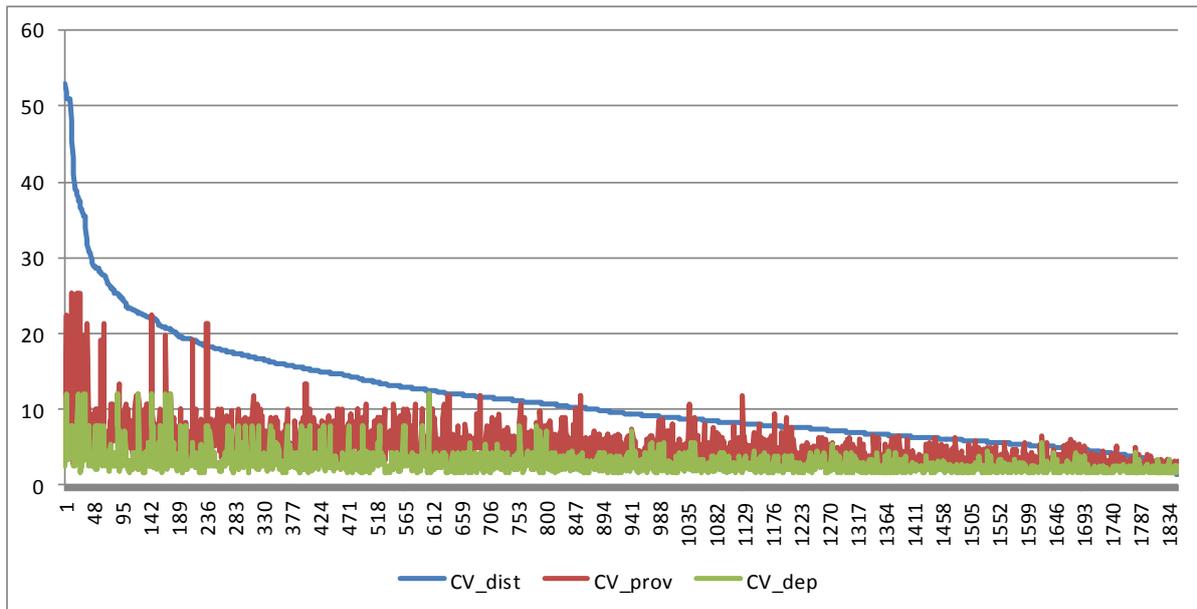
Fuente: Base de Datos de ENAHO 2012-2013

Elaboración: Propia

Observación:

- En el presente gráfico se puede apreciar que el estimador de pobreza monetaria con mejor precisión es el Estimador Sintético de variables Edad-Sexo por tener los menores valores de coeficiente de variación.

Gráfico 25: Coeficientes de Variación de Distritos, Provincias y Departamentos



Fuente: Mapa de Pobreza Monetaria 2013/INEI 2015

Elaboración: Propia

Observación:

- En el presente gráfico muestra que al calcular estimadores de pobreza monetaria para áreas más desagregadas (áreas menores) el coeficiente de variación tiende a aumentar debido a que la cantidad de muestra ya no es suficiente para dar estimaciones precisas.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

Como se apreció en el análisis de convergencia del total poblacional para las simulaciones se puede inferir que con un mínimo de 200 muestras simuladas de la base del SISFOH 2012-2013 se podría trabajar adecuadamente.

Los estimadores “Estimador Sintético de la variable edad-sexo” (EST 4), “Estimador Sintético de la variable estrato socioeconómico” (EST 5) y “Estimador Combinado de la variable estrato socioeconómico tipo 4” presentan un menor número de departamentos con sesgo relativo (SRA) mayor a 10.

Los estimadores “Estimador Sintético de la variable edad-sexo” (EST 4), “Estimador Sintético de la variable estrato socioeconómico” (EST 5) y “Estimador Combinado de la variable edad-sexo tipo 4” (EST 9) presentan un mayor número de departamentos con sesgo relativo (SRA) menor a 1.

Los estimadores Sintéticos (EST 4 y EST 5) se escogieron como los mejores estimadores para la pobreza monetaria debido a que presentaban menores valores respecto a la precisión relativa que se midieron con los indicadores de promedio de sesgo relativo (SRAM) y promedio de error cuadrático relativo (RECMRM)

Debido a que los errores típicos que se calculan con Jackknife y Bootstrap son muy semejantes entre sí se podría tomar indistintamente cualquiera solo que sería recomendable elegir el bootstrap porque el tiempo de procesamiento es menor que el Jackknife

El estimador Sintético de variable Edad-Sexo(EST 4) ha demostrado que es el estimador más preciso en la muestra comparado con el estimador Sintético de variables Estrato Socioeconómico (EST 5) y aún más preciso que del estimador tradicional (EST 1) o vigente y esto debido a la buena información auxiliar (información auxiliar del SISFOH)

A nivel de departamento la presente investigación ha tenido mejoras sustanciales, entonces se recomendaría hacer un posterior estudio a un nivel de Provincias.

REFERENCIAS BIBLIOGRÁFICAS

1. Roberto Hernández Sampieri R., Fernández Collado C. & Pilar Baptista L. (2014). Metodología de la Investigación. McGraw-Hill Education, 600
2. Molina, I. & Rao (2010). Small Area Estimation of Poverty Indicators. Canadian Journal of Statistics, 38, 369–385
3. J. N. K. Rao. (2003). Small Area Estimation. Wiley Series in Survey Methodology, 344
4. Ugarte M. D., Goicoa T., Militino A. F., Sagaseta-López M. (2009). Estimating unemployment in very small áreas. Statistics and Operations Research Transactions, 49-70
5. Palmer Arrache C., Eslava Gómez G. e Méndez Ramírez I. (2001). Métodos de Remuestreo para el Cálculo de Varianzas en Muestreos Complejos. Aplicación a la ENAL 96. Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, 42-121
6. Bravo Alonso J. I. (2010). Perú: Demanda de planificación familiar, en áreas pequeñas. Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, EAP. de Estadística, 63
7. Ministerio de Desarrollo Social de Chile. (2013). Procedimiento de Cálculo de la Tasa de Pobreza a nivel Comunal mediante la aplicación de Metodología de Estimación para Areas Pequeñas (SAE). Serie Documentos Metodológicos, No1, 46
8. López Paños R. (2000). Estimaciones para áreas pequeñas. Estadística Española, 291-338
9. Bay G., CEPAL/CELADE. (2000). Estimaciones Indirectas de Indicadores Demográficos para Areas Menores. Situación en América Latina. División de Población, 20
10. Kaushik Basu. (2015). Los diferentes frentes de batalla sobre las líneas de la pobreza. Voces del Bando Mundial.
<http://blogs.worldbank.org/voices/es/los-frentes-de-batall-sobre-las-l-neas-de-la-pobreza-0>
11. INEI. Enfoques teóricos sobre el Concepto y la Medición de la Pobreza. Proyectos Web INEI
<http://proyectos.inei.gob.pe/web/biblioineipub/bancopub/Est/Lib0384/cap2.htm>
12. Centro de Investigación y Desarrollo (CIDE). (2002). Metodologías para estimar indicadores sociodemográficos en áreas menores. Instituto Nacional de Estadística e Informática, 148
13. Rufino Moya C., Gregorio Saravia A. (2014). Probabilidad e Inferencia Estadística. Editorial San Marcos, 778
14. Rolando Beltrán A. (2008). Complejidad de modelos: Sesgo y Varianza — Notas de clase
<http://dis.unal.edu.co/~fgonza/courses/2008-1/ml/sesgovarianza-rbeltran.pdf>
15. Mancho Corcuera J. (2002). Técnicas de Estimación en Areas Pequeñas. Eustat, 41

16. Medina F. Los Errores de Muestreo en las Encuestas Complejas — CEPAL, 317-347
17. INEI / Microdatos. Ficha técnica de la ENAHO 2013. Instituto Nacional de Estadística e Informática
<http://inei.inei.gob.pe/microdatos/>
18. 21, P. (2014, Junio). *Seccion Politica Peru 21*. Retrieved from Seccion Politica Peru 21:
<http://peru21.pe/politica/ollanta-humala-tenemos-que-defender-proceso-descentralizacion-2186281>
19. 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*, 36.
20. Benavent, R., & Morales, D. (2016). Multivariate Fay-Herriot models for small area estimation. *Computational Statistical & Data Analysis*, 390.
21. Fay, R. E., & 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* , 277.
22. Ghosh, M., & Rao, J. (1994). Small Area Estimation: An Appraisal. *Statistical Science*, 76.
23. INEI. (2015, Abril). *Cifras de Pobreza INEI*. Retrieved from Cifras de Pobreza INEI:
https://www.inei.gob.pe/media/cifras_de_pobreza/informetecnico_pobreza2014.pdf
24. INEI. (n.d.). *Microdatos INEI*. Retrieved from Microdatos INEI: <http://inei.inei.gob.pe/microdatos/>
25. INEI, C. C. (2015, Setiembre). *Mapa de Pobreza*. Retrieved from Mapa de Pobreza:
https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1261/Libro.pdf
26. Instituto Nacional de Estadística e Informática. (2015). *Evolución de la pobreza monetaria 2009-214*. Lima: INEI.
27. Instituto Nacional de Estadística e Informática,. (2015). *Mapa de Pobreza Provincial y Distrital 2013*. Lima: INEI.
28. Marchetti, S., Tzavidis, N., & Pratesi, M. (2012). Non-parametric bootstrap mean squared error estimation for M -quantile estimators of small area averages, quantiles and poverty indicators. *Computational Statistics & Data Analysis*, 2902.
29. Marhuenda, Y., Molina, I., & Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistical & Data Alalysis*, 325.
30. ODM, C. (2005, Agosto). *Objetivos de Desarrollo del Milenio*. Retrieved from Objetivos de Desarrollo del Milenio: <http://www.cepal.org/es/publicaciones/2797-objetivos-desarrollo-milenio-mirada-america-latina-caribe>
31. Tarozzi, A. (2011). Can census data alone signal heterogeneity in the estimation of poverty maps? *Journal of Development Economics*, 185.

32. Viet, C. N. (2011). Poverty projection using a small area estimation method: Evidence from Vietnam. *Journal of Comparative Economics*, 382.
33. Mora R. Propiedades de Muestras Grandes y Simulación. Departamento of Economica, Universidad Carlos III de Madrid, 32
http://www.eco.uc3m.es/~ricmora/MICCUA/materials/S01T11_Spanish_handout.pdf
34. Militino, A. F. (2014). Introducción a la estimación en áreas pequeñas. Departamento de Estadística e I.O. – Universidad Pública de Navarra, Pamplona, España, 108
35. Fay, R. E. and 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.
http://www.jstor.org/stable/2286322?seq=1#page_scan_tab_contents
36. Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An Error-Components Model for Prediction of Country Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, 83, 28-36.
<http://ccp.ucr.ac.cr/noticias/sae/pdf/battese.pdf>
37. Ghosh, M. and Rao, J.N.K.,(1994). Small Area Estimation: An Appraisal, *Statistical Science*, 9, 55-93
http://www.jstor.org/stable/2246284?seq=1#page_scan_tab_contents

GLOSARIO DE TERMINOS

Pobreza Monetaria: Indicador económico de bienestar, mide el porcentaje de personas en la condición de pobres económicos respecto a un total de personas en un dominio o área.

SAE: Siglas de Small Area Estimation. Son técnicas estadísticas de estimación de parámetros en áreas en los cuales no tiene representatividad la muestra.

INEI: Siglas del Instituto Nacional de Estadística e Informática. Es el órgano rector de los Sistemas Nacionales de Estadística e Informática en el Perú. Norma, planea, dirige, coordina, evalúa y supervisa las actividades estadísticas e informáticas oficiales del país.

ENAH: Siglas de la Encuesta Nacional de Hogares. Instrumento muestral de hogares a nivel nacional que utiliza el INEI para estimar la pobreza monetaria cada año, así como también otros indicadores sociales y económicos.

SISFOH: Siglas del Empadronamiento Distrital de Población y Vivienda. Instrumento que contiene el padrón general de hogares del Perú y que es utilizado para la focalización de hogares en condición de pobreza y pobreza extrema.

MONTECARLO: Técnica estadística numérica, usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud.

SRA: Siglas del Sesgo Relativo en valor Absoluto. Es un indicador de evaluación que mide el alejamiento o sesgo de un valor estimado respecto a su valor verdadero para cada departamento

SRAM: Siglas de la Media del Sesgo Relativo en valor Absoluto. Es un indicador de evaluación que mide el alejamiento o sesgo de un valor estimado respecto a su valor verdadero a nivel nacional

RECMR: Siglas de la Raíz cuadrada del Error Cuadrático Medio Relativo. Es un indicador de evaluación que mide la dispersión o varianza de un valor estimado respecto a su valor verdadero para cada departamento

RECMMR: Siglas de la Media de la Raíz cuadrada del Error Cuadrático Medio Relativo. Es un indicador de evaluación que mide la dispersión o varianza de un valor estimado respecto a su valor verdadero a nivel nacional

ESTIMADOR DIRECTO: Método de estimación SAE en donde el cálculo del estimador solo toma en cuenta el factor de ponderación para una determinada área

ESTIMADOR POSTESTRATIFICADO: Método de estimación SAE en donde el cálculo del estimador toma en cuenta los distintos niveles de una variable auxiliar y el factor de ponderación para una determinada área

ESTIMADOR SINTETICO: Método de estimación SAE en donde el cálculo del estimador toma en cuenta los distintos niveles de una variable auxiliar de un área mayor para inferir en un área menor

ESTIMADOR COMPUESTO: Método de estimación SAE en donde el cálculo del estimador toma en cuenta un estimador directo o postestratificado y un estimador sintético

ESTIMADOR LOGIT: Método de estimación SAE en donde el cálculo del estimador toma en cuenta un modelo logit para predecir valores no muestreados

COEFICIENTE DE VARIACION: es una medida de dispersión que describe la cantidad de variabilidad en relación con la media

REMUESTREO: Es una variedad de métodos que permiten estimar la precisión de muestras estadísticas (medianas, variancias, percentiles) mediante el uso de subconjuntos de datos disponibles (jackknife) o tomando datos en forma aleatoria de un conjunto de datos (bootstrap)