

Desagregación de Estimaciones en Áreas Pequeñas usando R

Curso Internacional

CEPAL - ISWGHS

16 de octubre de 2022 – 22 de diciembre de 2023

Organización

El curso internacional “Desagregación de Estimaciones en Áreas Pequeñas usando R” es organizado por la Comisión Económica para América Latina y el Caribe (CEPAL) con apoyo del Intersecretariat Working Group on Household Surveys y la División de Estadísticas de las Naciones Unidas en Nueva York.

Antecedentes

Las encuestas de hogares permiten monitorear continuamente los indicadores a lo largo del tiempo para que los gobiernos puedan adoptar políticas públicas que se centren primero en los más vulnerables. Sin embargo, la precisión de las estimaciones de las encuestas de hogares disminuye cuando se trata de hacer inferencias en subgrupos para los cuales no se diseñó la encuesta. Sin embargo, es posible utilizar las encuestas de hogares como información para producir estimaciones basadas en la información auxiliar disponible que proviene de los censos o registros administrativos.

El énfasis en la Agenda 2030 para el Desarrollo Sostenible en no dejar a nadie atrás (*left no one behind*) y alcanzar a los más vulnerables primero presenta a la comunidad estadística muchos desafíos y oportunidades. La desagregación de datos es esencial para comprender si los frutos del desarrollo están beneficiando a todo el espectro de la sociedad, incluidos los más vulnerables y los más atrasados. Será un desafío para los sistemas estadísticos nacionales, incluidos los de los países desarrollados, producir datos desglosados por diversas características de la población y por diferentes niveles de geografía. Nuevas y mejores metodologías estadísticas, técnicas innovadoras y el mayor uso de datos administrativos se encuentran entre las iniciativas esenciales necesarias para responder a los enormes desafíos que presenta la necesidad de datos más detallados. La Agenda 2030 brinda oportunidades para mejorar y mejorar la capacidad estadística de los países para producir información relevante, actualizada y detallada sobre la población.

De esta manera, el desarrollo de capacidades en nuevas técnicas de estimación se vuelve crucial para proporcionar estadísticas sociales para aquellos subgrupos que no están bien representados en una encuesta. La División de Estadísticas de la CEPAL ha aunado sus esfuerzos para abordar el problema de la desagregación de estimaciones y para brindar orientación a los países sobre este tema. Como parte de esa colaboración, se ha desarrollado el presente curso, con el que se busca mejorar la capacidad de los

países para adoptar nuevas metodologías para producir indicadores ODS desagregados, complementando el uso de datos de encuestas de hogares con el uso de técnicas estadísticas adecuadas para su desagregación

El curso ayudará a lograr un entendimiento común y mejorado entre los participantes sobre los problemas técnicos y los desafíos de la desagregación cuando se usan datos de encuestas de hogares para producir estadísticas sociales. También proporcionará valiosos aportes para el desarrollo continuo de directrices y material de capacitación sobre la desagregación de datos para medir y monitorear los ODS. Este curso pretende presentar una guía práctica para la integración de información proveniente del análisis de encuestas de hogares, censos y registros administrativos, mediante la utilización del software estadístico R.

Metodología y aspectos por considerar

Objetivos generales

El curso permitirá discutir entre los técnicos de los sistemas estadísticos nacionales la utilidad de las encuestas de hogares junto con otras fuentes de datos para producir estadísticas sociales para áreas pequeñas (subgrupos de población específicos). Igualmente, se revisarán las metodologías actuales para la combinación de fuentes de datos para lograr niveles más altos de precisión y compartir los desafíos y las mejores prácticas para cumplir con el llamado de no dejar a nadie atrás.

Objetivos específicos

1. Introducir a los asistentes en el paradigma del error total de muestreo como base de análisis de las encuestas de hogares.
2. Repasar las técnicas usuales de ponderación y estimación directa en las encuestas de hogares basadas en el principio de representatividad.
3. Presentar las metodologías de estimación indirecta en el cálculo de estimadores de parámetros desagregados
4. Introducir a los participantes en el paradigma de la inferencia basada en la modelación sobre la población finita.
5. Presentar las metodologías de estimación de *Fay-Herriot* (modelos de área) y *Battese-Harter-Fuller* (modelos de unidad)
6. Presentar la antigua metodología del Banco Mundial (ELL) y del predictor *Empirical Best*.
7. Presentar las metodologías asociadas a parámetros complejos que se desvían del supuesto de la normalidad.
8. Realizar prácticas computacionales en R para todas las metodologías de estimación en áreas pequeñas.
9. Realizar mapas de indicadores desagregados a nivel geográfico.

Dirigido a

Este curso está dirigido al personal técnico de los INE y demás instituciones de los sistemas estadísticos nacionales. Asimismo, funcionarios técnicos de Ministerios Nacionales y áreas de gobiernos regionales, provinciales o distritales. De la misma manera, a los miembros de organizaciones que realizan investigación social, y a miembros de la academia, como profesores o estudiantes de programas afines al tópico del curso.

Requisitos

Este curso introducirá paulatinamente los conceptos fundamentales de la teoría del muestreo y el ajuste estadístico de modelos mixtos. Se espera que el asistente tenga destreza en el manejo básico de algunos conceptos estadísticos, como medidas de tendencia central, medidas de dispersión y conceptos de probabilidad. De igual forma, el estudiante debe tener conocimientos básicos en el manejo de información sistematizada en hojas de cálculo. El conocimiento y previo manejo del software R es deseable.

¿Por qué debo tomar este curso?

- Para entender las diferentes características de la estimación desagregada de los indicadores de los Objetivos de Desarrollo Sostenible.
- Para obtener capacitación en las técnicas modernas de estimación indirecta basadas en encuestas de hogares, censos y registros administrativos.
- Para aprender a seleccionar la mejor metodología a la hora de obtener estimaciones desagregadas y sus correspondientes errores estándar.
- Para poder implementar un sistema de estimación en áreas pequeñas dentro de los Institutos Nacionales de Estadística, Ministerios de Gobierno y demás instituciones estatales en los países.
- Para tener la oportunidad de interactuar directamente con instructores expertos en la estimación de indicadores desagregados y discutir con profesionales que cuentan con bastante experiencia en este tópico.

Materiales

Debido a la diversidad de herramientas computacionales disponibles para la modelación estadística y estimación en encuestas de hogares, se propone trabajar con el software estadístico R, puesto que su acceso es universal, es un software libre y además es multiplataforma (Windows, Mac, Linux). El ILPES de CEPAL proveerá la plataforma virtual de este curso en donde se podrá acceder a los materiales como presentaciones, videos, foros, bases de datos, códigos computacionales, entre otros.

Duración del curso

El Curso tiene un total de 80 horas académicas, las cuales se desarrollarán del 16 de octubre de 2023 al 22 de diciembre de 2023, repartidos en diez módulos. El curso tiene una duración de diez semanas y exige una hora diaria de dedicación para lectura de material, visualización de videos y resolución de actividades propuestas.

Metodología del curso

Las clases serán virtuales a cargo de funcionarios de la División de Estadísticas e instructores internacionales para que los estudiantes puedan replicar los conceptos adquiridos en el curso a través de 10 módulos que repasan los distintos contenidos. Consta de material audiovisual para ejemplificar los aspectos más importantes de las metodologías de estimación en áreas pequeñas y de prácticas computacionales con bases de datos de encuestas de hogares y censos. De manera general en el curso:

- El progreso del estudiante será valorado por medio de tareas, evaluaciones y un proyecto final.
- El estudiante tendrá acceso a los materiales del curso a través de la plataforma virtual Moodle, y los contenidos se liberarán de manera diaria.
- Aunque el estudiante tendrá total libertad en el manejo de su horario de ingreso a la plataforma, se espera que pueda dedicar al menos 30 minutos diarios al estudio del contenido.
- Durante el desarrollo del curso, todos los viernes de cada semana se tendrán sesiones virtuales computacionales sincrónicas que tendrán lugar a través de la plataforma MS Teams desde las 10h00 hasta las 11h30 de Santiago de Chile (GMT -3).

Contenidos temáticos

0. Objetivos de Desarrollo Sostenible y SAE.
1. Estimadores Directos en encuestas de hogares.
2. Estimadores sintéticos y compuestos.
3. Modelos de área.
4. Modelos de unidad.
5. Mejor predictor empírico.
6. Modelos bayesianos y generalizados.
7. Marco de referencia para la aplicación de modelos SAE.
8. Mapeo de indicadores.
9. Repaso de las metodologías SAE.

Evaluación

El estudiante deberá presentar tres tareas a lo largo del curso, responder los cuestionarios de seguimiento semanales y realizar un proyecto final. Tanto las tareas como el proyecto se deberán subir a la plataforma virtual en las fechas acordadas al inicio del curso y tendrán una calificación que variará de cero a cien puntos. La ponderación de la evaluación es como sigue a continuación:

- Tareas (15% cada una)
- Cuestionarios de seguimiento (40% en total)
- Proyecto final (15%)

La asistencia a las sesiones sincrónicas y la participación diaria en el foro virtual son componentes fundamentales en el desarrollo del curso. Es por lo anterior que se requiere que el estudiante participe en al menos el 85% de los foros. La nota mínima aprobatoria del curso será de 70 puntos para los estudiantes que acrediten la participación mínima en las sesiones.

Instructores

El personal docente del curso está integrado por personal de CEPAL, así como por profesores con amplio reconocimiento internacional y consultores altamente calificados. A continuación, se presentan los instructores del curso

- PhD. Isabel Molina. Universidad Complutense de Madrid (España).
- PhD. Pedro Silva. ENCE/IBGE (Brasil).
- PhD. Ángela Luna. Universidad de Southampton (Inglaterra).
- PhD. Natalia Rojas. Universidad de los Emiratos Árabes Unidos (EAU).
- PhD. Cristian Téllez. Universidad Santo Tomás (Colombia).
- MSc. Diego Lemus. DNP (Colombia).
- PhD. Andrés Gutiérrez. UN-CEPAL.
- MSc. Felipe Molina. UN-CEPAL.

Coordinación académica

Para información sobre el contenido del curso contactar a:

Sr. Andrés Gutiérrez

Email: andres.gutierrez@cepal.org

Teléfono: +56 2 2210 2665

Sr. Cristian Téllez

Email: cftellezp@unal.edu.co

Teléfono: +57 300 2451617

Cronograma

Módulo	Título del tema
Módulo 0 - Objetivos de Desarrollo Sostenible y SAE Andrés Gutiérrez 16 al 20 de octubre de 2023	Objetivos, metas e indicadores
	No dejar a nadie atrás
	Coefficiente de variación
	Criterios de calidad
	Áreas pequeñas
Módulo 1 - Estimadores Directos Pedro Silva 23 al 27 de octubre de 2023	Introducción
	Indicadores comunes de pobreza
	Métodos directos para la desagregación de datos de pobreza
	Estimador tipo Hájek
	Estimadores GREG y de calibración

	Aplicación
Módulo 2 - Estimadores sintéticos Diego Lemus 30 de octubre al 3 de noviembre de 2023	Introducción
	Estimador postestratificado sintético
	Estimador de regresión a nivel de área (I)
	Estimador de regresión a nivel de área (II)
	Estimador de regresión a nivel de individuo
	Estimadores compuestos
	Aplicación
Módulo 3 - Modelos de área Cristian Téllez 6 al 10 de noviembre de 2023	Introducción
	Efectos aleatorios
	Error cuadrado medio
	Ventajas y desventajas
	Aplicación
Módulo 4 - Modelos de unidad Diego Lemus 13 al 17 de noviembre de 2023	Introducción
	Modelos con errores anidados
	Modelo de Battese-Harter-Fuller (I)
	Modelo de Battese-Harter-Fuller (II)
	Modelo de Elbers-Lanjouw-Lanjouw
Aplicación	
Módulo 5 - Mejor predictor empírico Ángela Luna 20 al 24 de noviembre de 2023	Introducción
	Modelo con errores anidados
	Cálculo del predictor empírico
	Estimación del error cuadrático medio
	Aplicación
Módulo 6 - Modelos bayesianos y generalizados Andrés Gutiérrez 27 de noviembre al 1ro de diciembre de 2023	Introducción
	Métodos bayesianos (I)
	Métodos bayesianos (II)
	Modelos lineales generalizados (I)
	Modelos lineales generalizados (II)
Aplicación	
Módulo 7 - Marco de referencia Natalia Rojas 4 al 8 de diciembre de 2023	Especificación
	Análisis y adaptación
	Transformaciones
	Evaluación
	Resumen
	Emdi
Módulo 8 - Mapas Felipe Molina 11 al 15 de diciembre de 2023	Datos cartográficos
	Introducción a ggplot2
	Mapas en ggplot2
	Mapas en tmap
	Mapas en leaflet

Módulo 9 - Repaso de las metodologías SAE Isabel Molina 18 al 22 de diciembre de 2023	Introducción
	Estimadores directos
	Estimadores indirectos
	Modelos de área
	Modelos de unidad
	Modelos lineales generalizados
	Extensiones