Estimación en áreas pequeñas

Indicadores de pobreza y métodos directos

Andrés Gutiérrez
Comisión Económica para América Latina y el Caribe

2019

Referencias

Referencias

(2018) Molina, Isabel. *Estudio de los límites de desagregación de dates en encuestas de hogares para subgrupos de población y áreas geográficas y los requerimientos para superarlos: Fase II*. CEPAL.

Referencias

(2018) Molina, Isabel. *Estudio de los límites de desagregación de dates en encuestas de hogares para subgrupos de población y áreas geográficas y los requerimientos para superarlos: Fase II*. CEPAL.

Referencias

- (2018) Molina, Isabel. *Estudio de los límites de desagregación de dates en encuestas de hogares para subgrupos de población y áreas geográficas y los requerimientos para superarlos: Fase II*. CEPAL.
- (2015) Rao, J.N.K y Isabel Molina. *Small Area Estimation*. Second ed. Wiley Series in Survey Methodology.

Introducción

Introducción

- Una encuesta es realizada con un tamaño muestral establecido.

Introducción

- Una encuesta es realizada con un tamaño muestral establecido.
- Después de una encuesta realizada, a menudo se produce una demanda para estimaciones en áreas más desagregadas.

Introducción

- Una encuesta es realizada con un tamaño muestral establecido.
- Después de una encuesta realizada, a menudo se produce una demanda para estimaciones en áreas más desagregadas.
- Por ejemplo, se realiza un muestreo para estimar niveles de pobreza en departamentos, pero después, el cliente quiere que se realicen estas estimaciones a nivel de municipio.

Introducción

- Cuando eso pasa, se puede aumentar los tamaños muestrales en las áreas en las que sea necesario.

Introducción

- Cuando eso pasa, se puede aumentar los tamaños muestrales en las áreas en las que sea necesario.
- Hay varios métodos para mejorar el diseño muestral.

Introducción

- Cuando eso pasa, se puede aumentar los tamaños muestrales en las áreas en las que sea necesario.
- Hay varios métodos para mejorar el diseño muestral.
- No obstante, esto podría ser caro, y el cliente podría pedir más de lo que es posible.

Introducción

- Las subdivisiones para las cuales se desean estimaciones se llaman “áreas” o “dominios”.

Introducción

- Las subdivisiones para las cuales se desean estimaciones se llaman "áreas" o "dominios".
- "áreas" pueden ser no solo áreas geográficas, sino también grupos socioeconómicos, o un cruce de ambos tipos.

Introducción

- Las subdivisiones para las cuales se desean estimaciones se llaman “áreas” o “dominios”.
- “áreas” pueden ser no solo áreas geográficas, sino también grupos socioeconómicos, o un cruce de ambos tipos.
- A la hora de estimar indicadores en estas áreas, se puede usar un *estimador directo*, lo que usa solamente los datos de la encuesta para esa área.

Introducción

- Las subdivisiones para las cuales se desean estimaciones se llaman "áreas" o "dominios".
- "áreas" pueden ser no solo áreas geográficas, sino también grupos socioeconómicos, o un cruce de ambos tipos.
- A la hora de estimar indicadores en estas áreas, se puede usar un *estimador directo*, lo que usa solamente los datos de la encuesta para esa área.
- Habitualmente son insesgados o prácticamente insesgados con respecto al diseño muestral.

Introducción

- Las subdivisiones para las cuales se desean estimaciones se llaman "áreas" o "dominios".
- "áreas" pueden ser no solo áreas geográficas, sino también grupos socioeconómicos, o un cruce de ambos tipos.
- A la hora de estimar indicadores en estas áreas, se puede usar un *estimador directo*, lo que usa solamente los datos de la encuesta para esa área.
- Habitualmente son insesgados o prácticamente insesgados con respecto al diseño muestral.
- En esta presentación nos enfocaremos en estos estimadores.

Introducción

- Como se ha dicho, en algunas áreas, el tamaño muestral es demasiado pequeño, lo que incrementa errores de muestreo en los estimadores directos para esas áreas.

Introducción

- Como se ha dicho, en algunas áreas, el tamaño muestral es demasiado pequeño, lo que incrementa errores de muestreo en los estimadores directos para esas áreas.
- Cuando esto pasa, estas áreas se llaman *areas pequeñas*.

Introducción

- Como se ha dicho, en algunas áreas, el tamaño muestral es demasiado pequeño, lo que incrementa errores de muestreo en los estimadores directos para esas áreas.
- Cuando esto pasa, estas áreas se llaman *areas pequeñas*.
- Esto no refiere al tamaño poblacional del área, sino áreas para las que no se disponen estimadores directos eficientes debido a tamaños muestrales pequeños.

Indicadores comunes de pobreza y desigualdad

Indicadores comunes de pobreza y desigualdad

- El indicador más común para medir pobreza es *la incidencia o tasa de pobreza*, también se conoce como tasa en riesgo de pobreza.

Indicadores comunes de pobreza y desigualdad

- El indicador más común para medir pobreza es *la incidencia o tasa de pobreza*, también se conoce como tasa en riesgo de pobreza.
- Otro indicador es la *brecha de la pobreza*, que mide la magnitud de pobreza en lugar de frecuencia.

Indicadores comunes de pobreza y desigualdad

- El indicador más común para medir pobreza es *la incidencia o tasa de pobreza*, también se conoce como tasa en riesgo de pobreza.
- Otro indicador es la *brecha de la pobreza*, que mide la magnitud de pobreza en lugar de frecuencia.
- Estos dos son parte de una familia de indicadores más amplia definidos por Foster, Greer y Thorbecke (1984), que llamaremos *indicadores FGT*.

Indicadores comunes de pobreza y desigualdad

- El indicador más común para medir pobreza es *la incidencia o tasa de pobreza*, también se conoce como tasa en riesgo de pobreza.
- Otro indicador es la *brecha de la pobreza*, que mide la magnitud de pobreza en lugar de frecuencia.
- Estos dos son parte de una familia de indicadores más amplia definidos por Foster, Greer y Thorbecke (1984), que llamaremos *indicadores FGT*.
- Ambos indicadores tienen la ventaja de ser aditivos.

Indicadores comunes de pobreza y desigualdad

- Llamemos U a la población objetivo de tamaño N, la cuál se divide en D subpoblaciones de tamaños $N_1, \ldots, N_D$.

Indicadores comunes de pobreza y desigualdad

- Llamemos U a la población objetivo de tamaño N, la cuál se divide en D subpoblaciones de tamaños $N_1, \ldots, N_D$.
- Llamemos E_{di} al poder adquisitivo (e.g.medida de ingresos o gastos) del individuo i en área d.

Indicadores comunes de pobreza y desigualdad

- Llamemos U a la población objetivo de tamaño N, la cuál se divide en D subpoblaciones de tamaños $N_1, \ldots, N_D$.
- Llamemos E_{di} al poder adquisitivo (e.g.medida de ingresos o gastos) del individuo i en área d.
- Llamamos z al umbral predefinido de pobreza, por debajo del cual un individuo se considera en riesgo de pobreza.

Indicadores comunes de pobreza y desigualdad

- Los indicadores FGT para el área d pueden ser definidos por:

$$F_{\alpha d} = \frac{1}{N_d}\sum_{i=1}^{N_d}\left(\frac{z - E_{di}}{z}\right)^{\alpha} I(E_{di} < z), \quad d = 1, \ldots, D, \ \alpha \geq 0$$

donde $I(E_{di} < z)$ es una función indicadora que toma el valor 1 si $E_{di} < z$ y 0 en caso contrario. Note que:

Indicadores comunes de pobreza y desigualdad

- Los indicadores FGT para el área d pueden ser definidos por:

$$F_{\alpha d} = \frac{1}{N_d}\sum_{i=1}^{N_d}\left(\frac{z-E_{di}}{z}\right)^{\alpha} I(E_{di} < z), \quad d = 1, \ldots, D, \ \alpha \geq 0$$

donde $I(E_{di} < z)$ es una función indicadora que toma el valor 1 si $E_{di} < z$ y 0 en caso contrario. Note que:

- Con $\alpha = 0$, obtenemos la *tasa de pobreza*

Indicadores comunes de pobreza y desigualdad

- Los indicadores FGT para el área d pueden ser definidos por:

$$F_{\alpha d} = \frac{1}{N_d}\sum_{i=1}^{N_d}\left(\frac{z - E_{di}}{z}\right)^{\alpha} I(E_{di} < z), \quad d = 1, \ldots, D, \ \alpha \geq 0$$

donde $I(E_{di} < z)$ es una función indicadora que toma el valor 1 si $E_{di} < z$ y 0 en caso contrario. Note que:

- Con $\alpha = 0$, obtenemos la *tasa de pobreza*
- Con $\alpha = 1$, obtenemos la *brecha de pobreza*

Métodos directos para la desagregación de datos de pobreza

Métodos directos

- En esta sección, se describirán estimadores directos para la media de una variable en un área, dada por:

$$\overline{Y}_d = N_d^{-1} \sum_{i=1}^{N_d} Y_{di}$$

donde Y_{di} es el valor de la variable de individuo i en área d.

Métodos directos

- Los indicadores FGT,

$$F_{\alpha,di} = \left(\frac{z - E_{di}}{z}\right)^{\alpha} I(E_{di} < z),$$

también se pueden escribir en la forma de la diapositiva anterior.

Métodos directos

- Los indicadores FGT,

$$F_{\alpha,di} = \left(\frac{z - E_{di}}{z}\right)^{\alpha} I(E_{di} < z),$$

  también se pueden escribir en la forma de la diapositiva anterior.
- Llamemos $F_{\alpha d}$ a la media de $Y_{di} = F_{\alpha,di}$ en el dominio d.

Métodos directos

- Los indicadores FGT,

$$F_{\alpha,di} = \left(\frac{z - E_{di}}{z}\right)^{\alpha} I(E_{di} < z),$$

  también se pueden escribir en la forma de la diapositiva anterior.
- Llamemos $F_{\alpha d}$ a la media de $Y_{di} = F_{\alpha,di}$ en el dominio d.
- Entonces,

$$F_{\alpha d} = N_d^{-1} \sum_{i=1}^{N_d} F_{\alpha,di}$$

Métodos directos

- Los indicadores FGT,

$$F_{\alpha,di} = \left(\frac{z - E_{di}}{z}\right)^{\alpha} I(E_{di} < z),$$

  también se pueden escribir en la forma de la diapositiva anterior.
- Llamemos $F_{\alpha d}$ a la media de $Y_{di} = F_{\alpha,di}$ en el dominio d.
- Entonces,

$$F_{\alpha d} = N_d^{-1} \sum_{i=1}^{N_d} F_{\alpha,di}$$

- Este estimador es *directo* pues solo usa los datos del dominio d en cuestión.

Métodos directos: Estimadores Horvitz-Thompson y Hájek

Métodos directos: Horvitz-Thompson (HT)

- El estimador Horvitz-Thompson es un estimador directo insesgado con respecto al diseño muestral para la media de área d, $\bar{Y}_d$.

Métodos directos: Horvitz-Thompson (HT)

- El estimador Horvitz-Thompson es un estimador directo insesgado con respecto al diseño muestral para la media de área d, $\bar{Y}_d$.
- El estimador HT es conocido por

$$\hat{\bar{Y}}_d = N_d^{-1} \sum_{i \in s_d} w_{di} Y_{di}$$

Métodos directos: Horvitz-Thompson (HT)

- El estimador Horvitz-Thompson es un estimador directo insesgado con respecto al diseño muestral para la media de área d, $\bar{Y}_d$.
- El estimador HT es conocido por

$$\hat{\bar{Y}}_d = N_d^{-1} \sum_{i \in s_d} w_{di} Y_{di}$$

- Para este estimador, es necesario conocer el tamaño poblacional, N_d.

Métodos directos: Horvitz-Thompson (HT)

- El estimador Horvitz-Thompson es un estimador directo insesgado con respecto al diseño muestral para la media de área d, $\bar{Y}_d$.
- El estimador HT es conocido por

$$\hat{\bar{Y}}_d = N_d^{-1} \sum_{i \in s_d} w_{di} Y_{di}$$

- Para este estimador, es necesario conocer el tamaño poblacional, N_d.
- En cambio, para el estimador HT del total, $\hat{Y}_d = \sum_{i \in s_d} w_{di} Y_{di}$, no se necesita el tamaño poblacional.

Métodos directos: Horvitz-Thompson (HT)

- Un estimador para la varianza del estimador HT viene dado por

$$\widehat{\text{var}}_{\pi}(\hat{\hat{Y}}_d) = N_d^{-2}\left\{\sum_{i\in s_d}\frac{Y_{di}^2}{\pi_{di}^2}(1-\pi_{di}) + 2\sum_{i\in s_d}\sum_{\substack{j\in s_d\\ j>i}}\frac{Y_{di}Y_{dj}}{\pi_{di}\pi_{dj}}\left(\frac{\pi_{d,ij}-\pi_{di}\pi_{dj}}{\pi_{d,ij}}\right)\right\}$$

donde $\pi_{d,ij}$ es la probabilidad de inclusión de segundo orden

Métodos directos: Horvitz-Thompson (HT)

- Un estimador para la varianza del estimador HT viene dado por

$$\widehat{\text{var}}_{\pi}(\hat{\bar{Y}}_d) = N_d^{-2}\left\{\sum_{i\in s_d}\frac{Y_{di}^2}{\pi_{di}^2}(1-\pi_{di}) + 2\sum_{i\in s_d}\sum_{\substack{j\in s_d\\ j>i}}\frac{Y_{di}Y_{dj}}{\pi_{di}\pi_{dj}}\left(\frac{\pi_{d,ij}-\pi_{di}\pi_{dj}}{\pi_{d,ij}}\right)\right\}$$

  donde $\pi_{d,ij}$ es la probabilidad de inclusión de segundo orden

- Este estimador es insesgado si $\pi_{di} > 0$ para todo $i = 1, \ldots, N_d$.

Métodos directos: Horvitz-Thompson (HT)

- Un estimador para la varianza del estimador HT viene dado por

$$\widehat{\text{var}}_{\pi}(\hat{\hat{Y}}_d) = N_d^{-2}\left\{\sum_{i\in s_d}\frac{Y_{di}^2}{\pi_{di}^2}(1-\pi_{di}) + 2\sum_{i\in s_d}\sum_{\substack{j\in s_d\\ j>i}}\frac{Y_{di}Y_{dj}}{\pi_{di}\pi_{dj}}\left(\frac{\pi_{d,ij}-\pi_{di}\pi_{dj}}{\pi_{d,ij}}\right)\right\}$$

  donde $\pi_{d,ij}$ es la probabilidad de inclusión de segundo orden

- Este estimador es insesgado si $\pi_{di} > 0$ para todo $i = 1, \ldots, N_d$.
- Si se supone que $\pi_{d,ij} \approx \pi_{di}\pi_{dj}$, el estimador queda definido por:

$$\widehat{\text{var}}_{\pi}(\hat{\hat{Y}}_d) = N_d^{-2}\sum_{i\in s_d} w_{di}(w_{di}-1)Y_{di}^2$$

Métodos directos: Horvitz-Thompson (HT)

- Como se ha mencionado, los indicadores FGT se pueden escribir como una media para individuos en un área,

$$F_{\alpha d} = N_d^{-1} \sum_{i=1}^{N_d} F_{\alpha, di}$$

Métodos directos: Horvitz-Thompson (HT)

- Como se ha mencionado, los indicadores FGT se pueden escribir como una media para individuos en un área,

$$F_{\alpha d} = N_d^{-1} \sum_{i=1}^{N_d} F_{\alpha,di}$$

- Por consiguiente, el estimador HT de $F_{\alpha d}$ es,

$$\hat{F}_{\alpha d} = N_d^{-1} \sum_{i \in s_d} w_{di} F_{\alpha,di}$$

Métodos directos: Horvitz-Thompson (HT)

- Podemos usar el estimador HT, $\hat{Y}_d = \sum_{i \in s_d} w_{di} Y_{di}$, para estimar el total poblacional, es decir,

$$\hat{Y} = \sum_{d=1}^{D} \hat{Y}_d$$

Métodos directos: Horvitz-Thompson (HT)

- Podemos usar el estimador HT, $\hat{Y}_d = \sum_{i \in s_d} w_{di} Y_{di}$, para estimar el total poblacional, es decir,

$$\hat{Y} = \sum_{d=1}^{D} \hat{Y}_d$$

- Esta propiedad se llama *benchmarking*, donde los estimadores para áreas desagregadas suman al estimador para el total.

Métodos directos: Horvitz-Thompson (HT), comentario sobre benchmarking

- Cuando no se cumple la propiedad de benchmarking, $\hat{Y} = \sum_{d=1}^{D} \hat{Y}_d$, es común ajustar de la siguiente manera:

$$\hat{Y}_d^{AEST} = \hat{Y}_d^{EST} \frac{\hat{Y}}{\sum_{d=1}^{D} \hat{Y}_d^{EST}}, \quad d = 1, \ldots, D$$

Métodos directos: Hájek

- Aunque el estimador HT es insesgado, puede tener una varianza bajo el diseño muestral muy grande.

Métodos directos: Hájek

- Aunque el estimador HT es insesgado, puede tener una varianza bajo el diseño muestral muy grande.
- El estimador de Hájek es ligeramente insesgado pero con una varianza menor que la de HT, escrito de la siguiente forma,

$$\hat{\bar{Y}}_d^{HA} = \hat{N}_d^{-1} \sum_{i \in s_d} w_{di} Y_{di}, \text{ donde } \hat{N}_d = \sum_{i \in s_d} w_{di}$$

Métodos directos: Hájek

- Aunque el estimador HT es insesgado, puede tener una varianza bajo el diseño muestral muy grande.
- El estimador de Hájek es ligeramente insesgado pero con una varianza menor que la de HT, escrito de la siguiente forma,

$$\hat{\bar{Y}}_d^{HA} = \hat{N}_d^{-1} \sum_{i \in s_d} w_{di} Y_{di}, \text{ donde } \hat{N}_d = \sum_{i \in s_d} w_{di}$$

- Observe que no se necesita el tamaño poblacional con el estimador de Hájek.

Métodos directos: Hájek

- Un estimador de la varianza de Hájek, $\hat{\bar{Y}}_d^{HA}$, se obtiene con un proceso de linealización de Taylor.

Métodos directos: Hájek

- Un estimador de la varianza de Hájek, $\hat{\bar{Y}}_d^{HA}$, se obtiene con un proceso de linealización de Taylor.
- Si suponemos que $\pi_{d,ij} \approx \pi_{di}\pi_{dj}$ para todo $j \neq i$, y que todo $\pi_{di} > 0$, obtenemos:

$$\widehat{\text{var}}_\pi(\hat{\bar{Y}}_d) = \hat{N}_d^{-2} \sum_{i \in s_d} w_{di}(w_{di} - 1)(Y_{di} - \hat{\bar{Y}}_d^{HA})^2$$

Métodos directos: Hájek

- Como se ha mencionado, variables FGT se pueden escribir como una media para individuos en un área.

Métodos directos: Hájek

- Como se ha mencionado, variables FGT se pueden escribir como una media para individuos en un área.
- Por consiguiente, el estimador de Hájek de $F_{\alpha d}$ es,

$$\hat{F}_{\alpha d}^{HA} = \hat{N}_d^{-1} \sum_{i \in s_d} w_{di} F_{\alpha, di}$$

Resumen de estimadores HT y Hájek

- Indicadores objetivos:

Resumen de estimadores HT y Hájek

- Indicadores objetivos:
  - Parámetros aditivos (que son sumas de ciertas variables para cada individuo del área).

Resumen de estimadores HT y Hájek

- Indicadores objetivos:
  - Parámetros aditivos (que son sumas de ciertas variables para cada individuo del área).
  - Pueden ser funciones de variables de interés, por ejemplo, $F_{\alpha,di} = f(E_{di})$.

Resumen de estimadores HT y Hájek

- Indicadores objetivos:
  - Parámetros aditivos (que son sumas de ciertas variables para cada individuo del área).
  - Pueden ser funciones de variables de interés, por ejemplo, $F_{\alpha,di} = f(E_{di})$.
- Requerimientos de datos:

Resumen de estimadores HT y Hájek

- Indicadores objetivos:
  - Parámetros aditivos (que son sumas de ciertas variables para cada individuo del área).
  - Pueden ser funciones de variables de interés, por ejemplo, $F_{\alpha,di} = f(E_{di})$.
- Requerimientos de datos:
  - Pesos muestrales w_{di} para individuos en grupo d.

Resumen de estimadores HT y Hájek

- Indicadores objetivos:
  - Parámetros aditivos (que son sumas de ciertas variables para cada individuo del área).
  - Pueden ser funciones de variables de interés, por ejemplo, $F_{\alpha,di} = f(E_{di})$.
- Requerimientos de datos:
  - Pesos muestrales w_{di} para individuos en grupo d.
  - Para el estimador HT de la media y estimador de Hájek del total, tamaño poblacional del área N_d.

Resumen de estimadores HT y Hájek

- Ventajas:

Resumen de estimadores HT y Hájek

- Ventajas:
  - El estimador HT es insesgado y el de Hájek es ligeramente insesgado.

Resumen de estimadores HT y Hájek

- Ventajas:
  - El estimador HT es insesgado y el de Hájek es ligeramente insesgado.
  - Ambos son consistentes cuando n_d crece.

Resumen de estimadores HT y Hájek

- Ventajas:
  - El estimador HT es insesgado y el de Hájek es ligeramente insesgado.
  - Ambos son consistentes cuando n_d crece.
  - Son no paramétricos porque no se supone nada de la distribución de Y_{di}.

Resumen de estimadores HT y Hájek

- Desventajas:

Resumen de estimadores HT y Hájek

- Desventajas:
  - Son muy ineficientes para áreas pequeñas.

Resumen de estimadores HT y Hájek

- Desventajas:
    - Son muy ineficientes para áreas pequeñas.
    - No se puede calcular un estimador cuando $n_d = 0$, o cuando el área no es muestreada.

Métodos directos: Estimadores GREG y de calibración

Métodos directos: Estimador GREG

- El estimador generalizado de regresión (*generalized regression*), GREG, utiliza información auxiliar.

Métodos directos: Estimador GREG

- El estimador generalizado de regresión (*generalized regression*), GREG, utiliza información auxiliar.
- Este estimador requiere el total $\mathbf{X}_d = \sum_{i=1}^{N_d} \mathbf{x}_{di}$, o la media $\overline{\mathbf{X}_d} = N_d^{-1} \sum_{i=1}^{N_d} \mathbf{x}_{di}$, para el área d.

Métodos directos: Estimador GREG

- El estimador generalizado de regresión (*generalized regression*), GREG, utiliza información auxiliar.
- Este estimador requiere el total $\mathbf{X}_d = \sum_{i=1}^{N_d} \mathbf{x}_{di}$, o la media $\overline{\mathbf{X}_d} = N_d^{-1} \sum_{i=1}^{N_d} \mathbf{x}_{di}$, para el área d.
- El vector $\mathbf{x}_{di}$ consiste de valores de p variables auxiliares relacionadas con Y_{di}, para el individuo i en el área d.

Métodos directos: Estimador GREG

- Asumamos que existe un modelo de la forma

$$Y_{di} = \mathbf{x}'_{di}\beta_d + \epsilon_{di}, \quad i = 1, \ldots, N_d$$

Métodos directos: Estimador GREG

- Asumamos que existe un modelo de la forma

$$Y_{di} = \mathbf{x}'_{di}\beta_d + \epsilon_{di}, \quad i = 1, \ldots, N_d$$

- Entonces, podemos definir un estimador

$$\hat{\mathbf{B}}_d = \left(\sum_{i \in s_d} w_{di} \mathbf{x}_{di} \mathbf{x}'_{di} / c_{di} \right)^{-1} \sum_{i \in s_d} w_{di} \mathbf{x}_{di} Y_{di} / c_{di}$$

Métodos directos: Estimador GREG

- Asumamos que existe un modelo de la forma

$$Y_{di} = \mathbf{x}'_{di}\beta_d + \epsilon_{di}, \quad i = 1, \ldots, N_d$$

- Entonces, podemos definir un estimador

$$\hat{\mathbf{B}}_d = \left(\sum_{i \in s_d} w_{di}\mathbf{x}_{di}\mathbf{x}'_{di}/c_{di}\right)^{-1} \sum_{i \in s_d} w_{di}\mathbf{x}_{di}Y_{di}/c_{di}$$

- En el modelo, los errores ϵ_{di} son independientes con esperanza igual a 0 y varianza $\sigma^2 c_{di}$, con $c_{di} > 0$ siendo constantes que representan la posible heteroscedasticidad, $i = 1, \ldots, N_d$.

Métodos directos: Estimador GREG

- $\hat{\bar{\mathbf{X}}}_d = N_d^{-1} \sum_{i \in s_d} w_{di} \mathbf{x}_{di}$ es el estimador de HT de $\bar{\mathbf{X}}_d$

Métodos directos: Estimador GREG

- $\hat{\overline{\mathbf{X}}}_d = N_d^{-1} \sum_{i \in s_d} w_{di} \mathbf{x}_{di}$ es el estimador de HT de $\overline{\mathbf{X}}_d$
- Podemos usar la regresión mencionada para estimar $\overline{Y}_d$

Métodos directos: Estimador GREG

- $\hat{\overline{\mathbf{X}}}_d = N_d^{-1} \sum_{i \in s_d} w_{di} \mathbf{x}_{di}$ es el estimador de HT de $\overline{\mathbf{X}}_d$
- Podemos usar la regresión mencionada para estimar $\overline{Y}_d$
- Este estimador está dado por:

$$\hat{\overline{Y}}_d^{GREG} = \hat{\overline{Y}}_d + \left(\overline{\mathbf{X}}_d - \hat{\overline{\mathbf{X}}}_d\right)' \hat{\mathbf{B}}_d$$

Métodos directos: Estimador GREG

- El estimador GREG es más eficiente que el estimador directo $\hat{\bar{Y}}$ si las variables auxiliares están linealmente relacionadas con Y_{di},

Métodos directos: Estimador GREG

- El estimador GREG es más eficiente que el estimador directo $\hat{\bar{Y}}$ si las variables auxiliares están linealmente relacionadas con Y_{di},
- Es difícil encontrar auxiliares $\mathbf{x}_{di}$ relacionadas con $F_{\alpha,di} = I\{(z - E_{di})/z\}^{\alpha} I(E_{di} < z)$, porque es una función compleja.

Métodos directos: Estimador GREG

- Si $\pi_{d,ij} \approx \pi_{di}\pi_{di}$, para $j \neq i$, el estimador de varianza para GREG viene dado por:

$$\widehat{\text{var}}_{\pi}(\hat{\bar{Y}}_d^{GREG}) = N_d^{-2} \sum_{i \in s_d} w_{di}(w_{di} - 1)\tilde{e}_{di}^2$$

donde $\tilde{e}_{di} = Y_{di} - \mathbf{x}_{di}'\hat{\mathbf{B}}_d$.

Métodos directos: Estimador de calibración

- Este método utiliza los pesos h_{di} para estimar el total de una variable de interés usando p variables auxiliares.

Métodos directos: Estimador de calibración

- Este método utiliza los pesos h_{di} para estimar el total de una variable de interés usando p variables auxiliares.
- h_{di} son los pesos más cercanos a los pesos originales, w_{di}, sujeto a

$$\sum_{i \in s_d} h_{di} \mathbf{x}_{di} = \mathbf{X}_d$$

Métodos directos: Estimador de calibración

- Este método utiliza los pesos h_{di} para estimar el total de una variable de interés usando p variables auxiliares.
- h_{di} son los pesos más cercanos a los pesos originales, w_{di}, sujeto a

$$\sum_{i \in s_d} h_{di} \mathbf{x}_{di} = \mathbf{X}_d$$

- Una posibilidad viene dada por

$$h_{di} = w_{di} \left\{ 1 + \mathbf{x}'_{di} \left(\sum_{i \in s_d} w_{di} \mathbf{x}_{di} \mathbf{x}'_{di} / c_{di} \right)^{-1} \left(\mathbf{X}_d - \sum_{i \in s_d} w_{di} \mathbf{x}_{di} / c_{di} \right) \right\}, i \in s_d$$

Métodos directos: Estimador de calibración

- El estimador de calibración de $\bar{Y}_d$ se obtiene igual que el estimador de HT

$$\hat{\bar{Y}}_d^{CAL} = N_d^{-1} \sum_{i \in s_d} h_{di} Y_{di}$$

Métodos directos: Estimador de calibración

- El estimador de calibración de $\bar{Y}_d$ se obtiene igual que el estimador de HT

$$\hat{\bar{Y}}_d^{CAL} = N_d^{-1} \sum_{i \in s_d} h_{di} Y_{di}$$

- Se puede mostrar que, bajo ciertas condiciones de regularidad para $G_{di}(\cdot, \cdot)$, el estimador de calibración es asintóticamente igual al GREG y comparten la misma varianza asintótica.

Resumen de estimadores GREG y de calibración

- Indicadores objetivo: Medias/totales de la variable de interés.

Resumen de estimadores GREG y de calibración

- Indicadores objetivo: Medias/totales de la variable de interés.
- Requerimientos de datos:

Resumen de estimadores GREG y de calibración

- Indicadores objetivo: Medias/totales de la variable de interés.
- Requerimientos de datos:
  - Pesos muestrales w_{di} para individuos de la muestra en el área d.

Resumen de estimadores GREG y de calibración

- Indicadores objetivo: Medias/totales de la variable de interés.
- Requerimientos de datos:
  - Pesos muestrales w_{di} para individuos de la muestra en el área d.
  - Para el estimador de la media, tamaño poblacional del área N_d.

Resumen de estimadores GREG y de calibración

- Indicadores objetivo: Medias/totales de la variable de interés.
- Requerimientos de datos:
  - Pesos muestrales w_{di} para individuos de la muestra en el área d.
  - Para el estimador de la media, tamaño poblacional del área N_d.
  - Observaciones muestrales de las p variables auxiliares.

Resumen de estimadores GREG y de calibración

- Indicadores objetivo: Medias/totales de la variable de interés.
- Requerimientos de datos:
  - Pesos muestrales w_{di} para individuos de la muestra en el área d.
  - Para el estimador de la media, tamaño poblacional del área N_d.
  - Observaciones muestrales de las p variables auxiliares.
  - Totales $\mathbf{X}_d$ o medias $\bar{\mathbf{X}}_d$ poblacionales de las p variables auxiliares.

Resumen de estimadores GREG y de calibración

- Ventajas:

Resumen de estimadores GREG y de calibración

- Ventajas:
  - Son aproximadamente insesgados con respecto al diseño muestral.

Resumen de estimadores GREG y de calibración

- Ventajas:
    - Son aproximadamente insesgados con respecto al diseño muestral.
    - Pueden mejorar a los estimadores directos básicos si el modelo de regresión se verifica.

Resumen de estimadores GREG y de calibración

- Ventajas:
  - Son aproximadamente insesgados con respecto al diseño muestral.
  - Pueden mejorar a los estimadores directos básicos si el modelo de regresión se verifica.
  - No requieren la verificación del modelo considerado para las variables de interés Y_{di}; son no paramétricos.

Resumen de estimadores GREG y de calibración

- Desventajas:

Resumen de estimadores GREG y de calibración

- Desventajas:
  - Pueden ser ineficientes para áreas pequeñas.

Resumen de estimadores GREG y de calibración

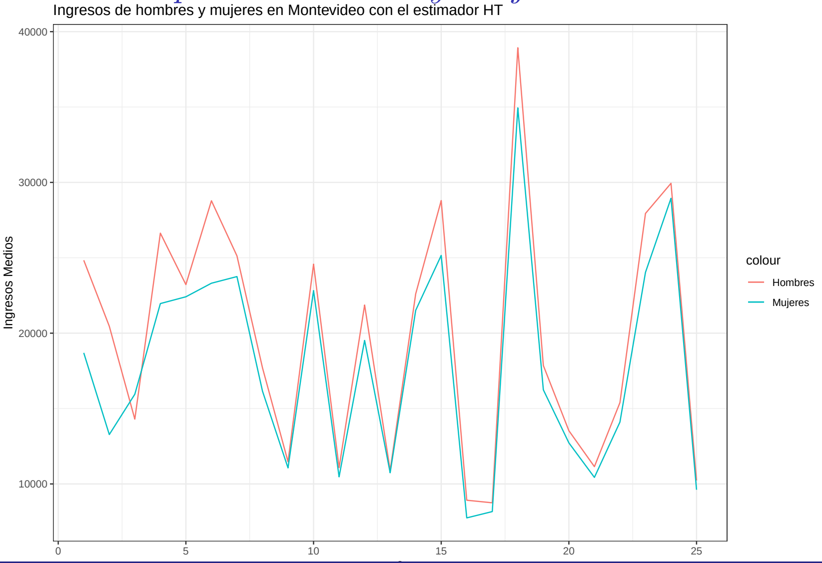
- Desventajas:
  - Pueden ser ineficientes para áreas pequeñas.
  - No se pueden calcular en áreas con un tamaño muestro n_d igual a 0.

Resultados: Estimación de ingreso medio en sectores de Montevideo

Horvitz Thompson: Hombres y Mujeres en Montevideo

sec2	ntotal	Hombres	Mujeres
2	121	20461	13277
1	167	24837	18694
3	186	14299	15951
4	319	26635	21965
6	320	28784	23314
5	495	23223	22414
21	3165	11148	10435
13	3556	10897	10742
18	3950	38932	34943
11	3963	11080	10473
17	4373	8750	8167
10	6302	24576	22823

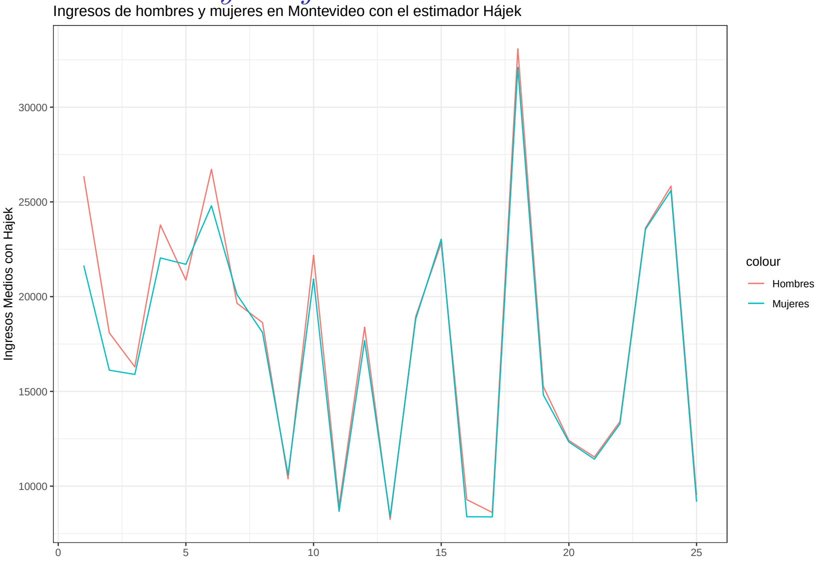
Horvitz Thompson: Hombres y Mujeres en Montevideo



Hájek: Hombres y Mujeres en Montevideo

sec2	ntotal	Hombres	Mujeres
2	121	18088	16120
1	167	26363	21644
3	186	16294	15896
4	319	23786	22044
6	320	26723	24798
5	495	20874	21706
21	3165	11539	11424
13	3556	8248	8384
18	3950	33081	32103
11	3963	8954	8675
17	4373	8612	8377
10	6302	22186	20929

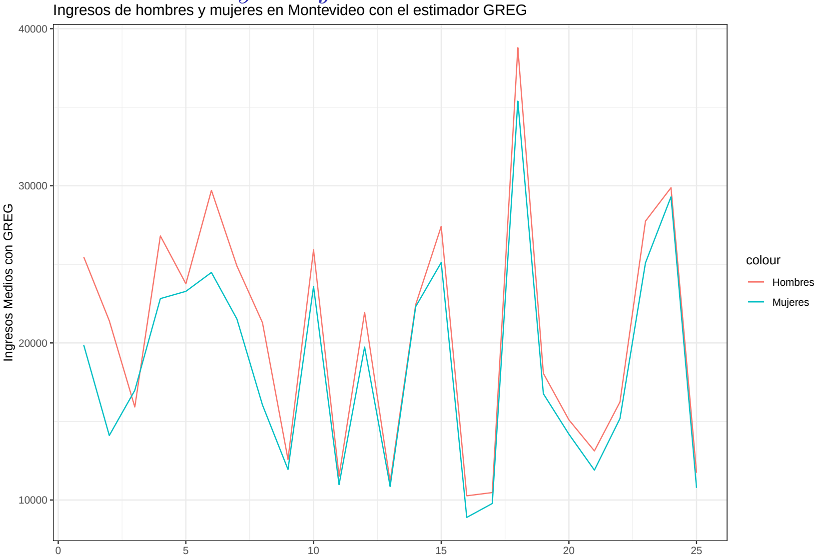
Hájek: Hombres y Mujeres en Montevideo



GREG: Hombres y Mujeres en Montevideo

sec2	ntotal	Hombres	Mujeres
2	121	21410	14107
1	167	25468	19861
3	186	15921	16981
4	319	26809	22819
6	320	29710	24484
5	495	23763	23282
21	3165	13125	11901
13	3556	11156	10862
18	3950	38789	35391
11	3963	11510	10977
17	4373	10473	9777
10	6302	25921	23589

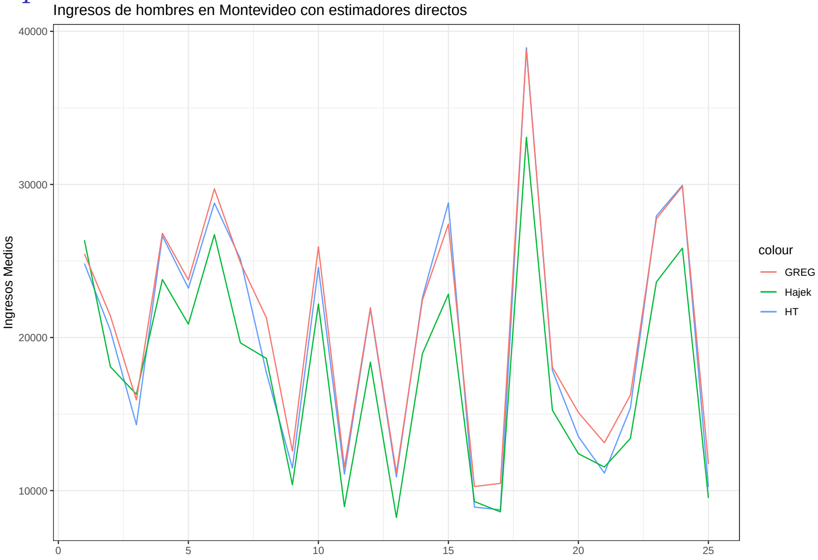
GREG: Hombres y Mujeres en Montevideo



Comparando los estimadores: Hombres

sec2	ntotal	HT	Hajek	GREG
2	121	20461	18088	21410
1	167	24837	26363	25468
3	186	14299	16294	15921
4	319	26635	23786	26809
6	320	28784	26723	29710
5	495	23223	20874	23763
21	3165	11148	11539	13125
13	3556	10897	8248	11156
18	3950	38932	33081	38789
11	3963	11080	8954	11510
17	4373	8750	8612	10473
10	6302	24576	22186	25921

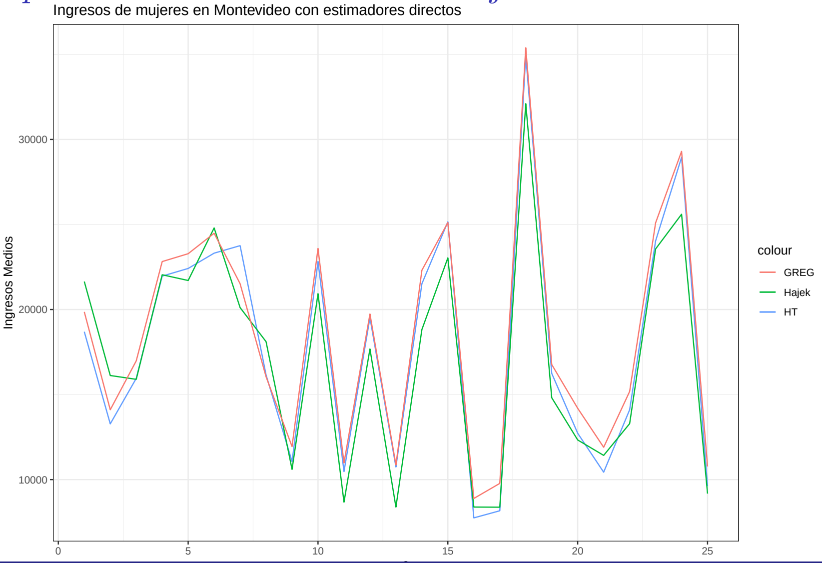
Comparando los estimadores: Hombres



Comparando los estimadores: Mujeres

sec2	ntotal	HT	Hajek	GREG
2	121	13277	16120	14107
1	167	18694	21644	19861
3	186	15951	15896	16981
4	319	21965	22044	22819
6	320	23314	24798	24484
5	495	22414	21706	23282
21	3165	10435	11424	11901
13	3556	10742	8384	10862
18	3950	34943	32103	35391
11	3963	10473	8675	10977
17	4373	8167	8377	9777
10	6302	22823	20929	23589

Comparando los estimadores: Mujeres



¡Gracias!

Email: andres.gutierrez@cepal.org