Maestría en Generación de Información Estadística

Teoría y Técnicas de Muestreo

Augusto E. Hoszowski , [1]

[1]UNTREF

2024

Tabla de Contenidos

Bibliografía

- Muestreo: Diseño y Análisis, Sharon Lohr (2000) . Thomson
- Teoría de Muestreo, Yves Tillé (en castellano, 2005)
- Sampling: Design and Analysis, Sharon Lohr (2021, 3Ed) . Thomson
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Practical Tools for Designing and Weighting Survey Samples (2ºEd), R Valliant et.al. Springer
- Complex Surveys: A Guide to Analysis Using R. T. Lumley. Wiley

Bibliografía

- Muestreo: Diseño y Análisis, Sharon Lohr (2000) . Thomson
- Teoría de Muestreo, Yves Tillé (en castellano, 2005)
- Sampling: Design and Analysis, Sharon Lohr (2021, 3Ed) . Thomson
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Practical Tools for Designing and Weighting Survey Samples (2ºEd), R Valliant et.al. Springer
- Complex Surveys: A Guide to Analysis Using R. T. Lumley. Wiley

Bibliografía

- Muestreo: Diseño y Análisis, Sharon Lohr (2000) . Thomson
- Teoría de Muestreo, Yves Tillé (en castellano, 2005)
- Sampling: Design and Analysis, Sharon Lohr (2021, 3Ed) . Thomson
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Practical Tools for Designing and Weighting Survey Samples (2ºEd), R Valliant et.al. Springer
- Complex Surveys: A Guide to Analysis Using R. T. Lumley. Wiley

Bibliografía

- Muestreo: Diseño y Análisis, Sharon Lohr (2000) . Thomson
- Teoría de Muestreo, Yves Tillé (en castellano, 2005)
- Sampling: Design and Analysis, Sharon Lohr (2021, 3Ed) . Thomson
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Practical Tools for Designing and Weighting Survey Samples (2ºEd), R Valliant et.al. Springer
- Complex Surveys: A Guide to Analysis Using R. T. Lumley. Wiley

Bibliografía

- Muestreo: Diseño y Análisis, Sharon Lohr (2000) . Thomson
- Teoría de Muestreo, Yves Tillé (en castellano, 2005)
- Sampling: Design and Analysis, Sharon Lohr (2021, 3Ed) . Thomson
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Practical Tools for Designing and Weighting Survey Samples (2ºEd), R Valliant et.al. Springer
- Complex Surveys: A Guide to Analysis Using R. T. Lumley. Wiley

Bibliografía

- Muestreo: Diseño y Análisis, Sharon Lohr (2000) . Thomson
- Teoría de Muestreo, Yves Tillé (en castellano, 2005)
- Sampling: Design and Analysis, Sharon Lohr (2021, 3Ed) . Thomson
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Practical Tools for Designing and Weighting Survey Samples (2ºEd), R Valliant et.al. Springer
- Complex Surveys: A Guide to Analysis Using R. T. Lumley. Wiley

Tabla de Contenidos

Estratificación

Algunos objetivos de la estratificación

- Garantizar un tamaño mínimo de muestra en ciertos dominios
- Reducir la dispersión entre las estimaciones a partir de las diferentes muestras: reducir la varianza de los estimadores
- Garantizar que las muestra esté bien distribuída (no necesariamente en forma proporcional)

Estratificación

Algunos objetivos de la estratificación

- Garantizar un tamaño mínimo de muestra en ciertos dominios
- Reducir la dispersión entre las estimaciones a partir de las diferentes muestras: reducir la varianza de los estimadores
- Garantizar que las muestra esté bien distribuída (no necesariamente en forma proporcional)

Estratificación

Algunos objetivos de la estratificación

- Garantizar un tamaño mínimo de muestra en ciertos dominios
- Reducir la dispersión entre las estimaciones a partir de las diferentes muestras: reducir la varianza de los estimadores
- Garantizar que las muestra esté bien distribuída (no necesariamente en forma proporcional)

Estratificación

Estratificación

$$U = U_1 \cup U_2 ..... U_H$$

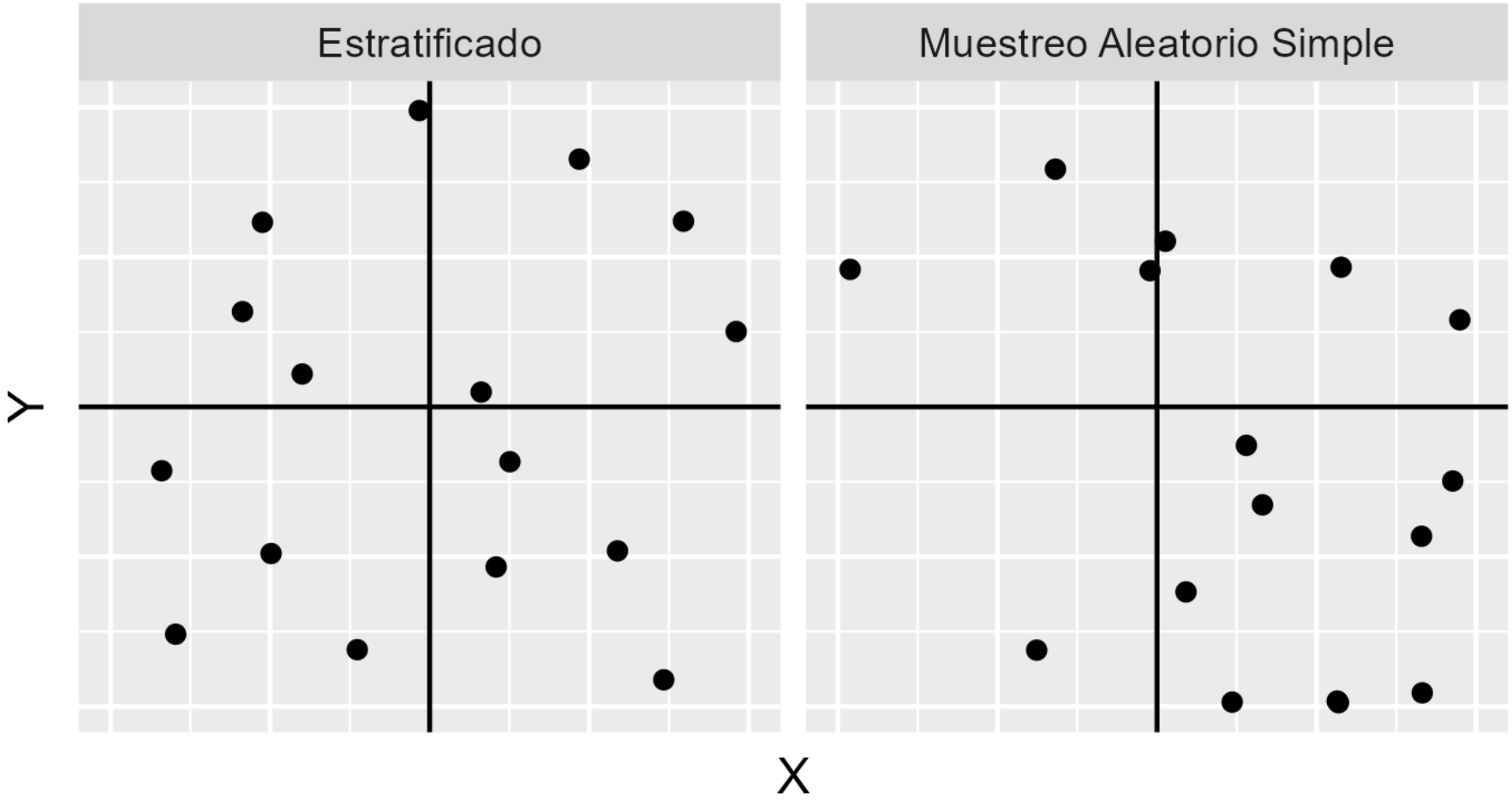
$$U_i \cap U_j = \emptyset \ si \quad i \neq j$$

En cada U_i seleccionamos una muestra aleatoria, en forma independiente

de estrato a estrato

Estratificación

Muestreo Aleatorio Simple y Muestreo Estratificado



Estratificación

En las encuestas es usual que la muestra **no** se distribuya en forma proporcional en los estratos

- Encuesta Permanente de Hogares
- Encuestas a Empresas
- PISA
- Etcétera

Graficamente:

Estratificación

En las encuestas es usual que la muestra **no** se distribuya en forma proporcional en los estratos

- Encuesta Permanente de Hogares
- Encuestas a Empresas
- PISA
- Etcétera

Graficamente:

Estratificación

En las encuestas es usual que la muestra **no** se distribuya en forma proporcional en los estratos

- Encuesta Permanente de Hogares
- Encuestas a Empresas
- PISA
- Etcétera

Graficamente:

Estratificación

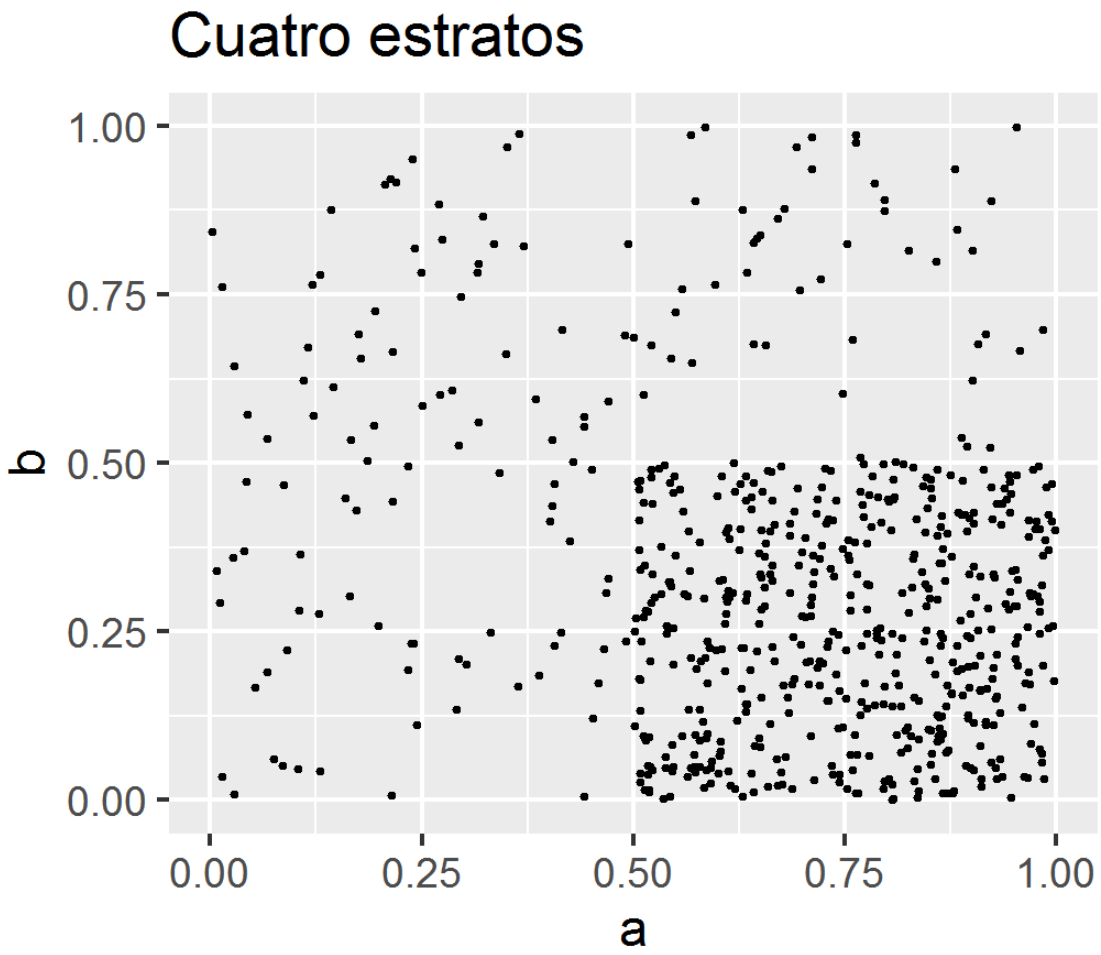
En las encuestas es usual que la muestra **no** se distribuya en forma proporcional en los estratos

- Encuesta Permanente de Hogares
- Encuestas a Empresas
- PISA
- Etcétera

Graficamente:

Estratificación

Muestreo Estratificado



Estratificación

Ejemplos de estratificaciones

- Muestra de empresas: Rama de actividad y tamaño
- Muestra de radios censales: Departamento y nivel educativo del jefe de hogar
- Muestra de alumnos: Carrera, sexo, tramo de edad
- ....

Estratificación

Ejemplos de estratificaciones

- Muestra de empresas: Rama de actividad y tamaño
- Muestra de radios censales: Departamento y nivel educativo del jefe de hogar
- Muestra de alumnos: Carrera, sexo, tramo de edad
- ....

Estratificación

Ejemplos de estratificaciones

- Muestra de empresas: Rama de actividad y tamaño
- Muestra de radios censales: Departamento y nivel educativo del jefe de hogar
- Muestra de alumnos: Carrera, sexo, tramo de edad
- ....

Estratificación

Ejemplos de estratificaciones

- Muestra de empresas: Rama de actividad y tamaño
- Muestra de radios censales: Departamento y nivel educativo del jefe de hogar
- Muestra de alumnos: Carrera, sexo, tramo de edad
- ....

Estratificación

- No necesariamente la estratificación reduce la varianza respecto a una muestra aleatoria simple!
- No necesariamente más estratos es un mejor diseño

Estratificación

- No necesariamente la estratificación reduce la varianza respecto a una muestra aleatoria simple!
- No necesariamente más estratos es un mejor diseño

Estratificación

Notación

- $N = \sum_{h=1}^{H} N_h$
- $n = \sum_{h=1}^{H} n_h$
- $W_h = N_h/N$

Estratificación

Notación

- $N = \sum_{h=1}^{H} N_h$
- $n = \sum_{h=1}^{H} n_h$
- $W_h = N_h / N$

Estratificación

Notación

- $N = \sum_{h=1}^{H} N_h$
- $n = \sum_{h=1}^{H} n_h$
- $W_h = N_h/N$

Estratificación

Estimador estratificado

Si $\hat{\bar{t}}_{hy}$ es un estimador insesgado de $\bar{Y}_h$ entonces

$$\bar{y}_{st} = \sum_h W_h \cdot \hat{\bar{t}}_{hy}$$

será un estimador insesgado de $\bar{Y}$
con varianza

$$V(\bar{y}_{st}) = \sum_h W_h^2 \cdot V(\hat{\bar{t}}_{hy})$$

que será (probablemente) *pequeña* si los estratos son *homogéneos*

Estratificación

Estimador estratificado

Si $\hat{\bar{t}}_{hy}$ es un estimador insesgado de $\bar{Y}_h$ entonces

$$\bar{y}_{st} = \sum_h W_h \cdot \hat{\bar{t}}_{hy}$$

será un estimador insesgado de $\bar{Y}$
con varianza

$$V(\bar{y}_{st}) = \sum_h W_h^2 \cdot V(\hat{\bar{t}}_{hy})$$

que será (probablemente) *pequeña* si los estratos son *homogéneos*

Estratificación

Estimador estratificado

Si $\hat{V}_{hy}$ es un estimador insesgado de $Var(\hat{\bar{Y}}_{hy})$ entonces

$$\hat{V}_y = \sum_h W_h^2 \cdot \hat{V}_{hy}$$

será un estimador insesgado de $Var(\bar{y}_{st})$

Estratificación

MAS en cada estrato

Si en cada estrato seleccionamos una MAS(n_h), entonces el estimador de H-T de la media de una variable Y es

$$\bar{y}_{st} = \sum_h W_h \cdot \bar{y}_h$$

y su varianza es

$$V(\bar{y}_{st}) = \sum_h W_h^2 \cdot (1 - n_h/N_h) \cdot \frac{S_h^2}{n_h}$$

Y un estimador insesgado será (n_h¿1)

$$\hat{V}(\bar{y}_{st}) = \sum_h W_h^2 \cdot (1 - n_h/N_h) \cdot \frac{s_h^2}{n_h}$$

Estratificación

Estimación de varianzas

- En la práctica, si hay muchos estratos, la selección de la muestra se facilita mediante algún software estadístico (SPSS, R-survey...)
- La estimación de varianzas (CV, *deff*, intervalos de confianza...) en general se hace con algún soft
- Es usual probar diferentes alternativas de estratificación (con bases simuladas o de operativos anteriores)
- Luego de estratificar el marco de muestreo y seleccionar la muestra asignaremos a cada unidad su factor de expansión (inversa de la probabilidad de selección). Para poder luego realizar las estimaciones

Estratificación

Estimación de varianzas

- En la práctica, si hay muchos estratos, la selección de la muestra se facilita mediante algún software estadístico (SPSS, R-survey...)
- La estimación de varianzas (CV, *deff*, intervalos de confianza...) en general se hace con algún soft
- Es usual probar diferentes alternativas de estratificación (con bases simuladas o de operativos anteriores)
- Luego de estratificar el marco de muestreo y seleccionar la muestra asignaremos a cada unidad su factor de expansión (inversa de la probabilidad de selección). Para poder luego realizar las estimaciones

Estratificación

Estimación de varianzas

- En la práctica, si hay muchos estratos, la selección de la muestra se facilita mediante algún software estadístico (SPSS, R-survey...)
- La estimación de varianzas (CV, *deff*, intervalos de confianza...) en general se hace con algún soft
- Es usual probar diferentes alternativas de estratificación (con bases simuladas o de operativos anteriores)
- Luego de estratificar el marco de muestreo y seleccionar la muestra asignaremos a cada unidad su factor de expansión (inversa de la probabilidad de selección). Para poder luego realizar las estimaciones

Estratificación

Estimación de varianzas

- En la práctica, si hay muchos estratos, la selección de la muestra se facilita mediante algún software estadístico (SPSS, R-survey...)
- La estimación de varianzas (CV, *deff*, intervalos de confianza...) en general se hace con algún soft
- Es usual probar diferentes alternativas de estratificación (con bases simuladas o de operativos anteriores)
- Luego de estratificar el marco de muestreo y seleccionar la muestra asignaremos a cada unidad su factor de expansión (inversa de la probabilidad de selección). Para poder luego realizar las estimaciones

Estratificación

Asignación de la muestra

Una vez estratificado el *Marco de Muestreo*, debemos distribuir la muestra en los estratos:

Asignación de la muestra

Uniforme
Proporcional
Optima (Neyman) ....

Estratificación

Asignación de la muestra

Asignación Uniforme

$$n_h = \frac{n}{H}$$

Si seleccionamos una MAS en cada estrato, entonces

$$F_{hi} = \frac{1}{\pi_{hi}} = \frac{N_h}{n_h}$$

Estratificación

Asignación de la muestra

Asignación proporcional

$$n_h = n \cdot \frac{N_h}{N}$$

Si seleccionamos una MAS en cada estrato, entonces

$$F_{hi} = \frac{1}{\pi_{hi}} = \frac{N_h}{n \cdot N_h/N} = \frac{N}{n}$$

(diseño autoponderado)

Estratificación

Asignación de Neyman

Supongamos una función de costo sencilla:

$$C = C_0 + \sum_h n_h \cdot c_h$$

La asignación de muestra que minimiza la varianza de $\bar{y}_{st}$ suponiendo una MAS en cada estrato, es

$$\frac{n_h}{n} = \frac{N_h \cdot S_h/\sqrt{c_h}}{\sum_k N_k S_k/\sqrt{c_k}}$$

Estratificación

Asignación de Neyman

Supongamos una función de costo sencilla:

$$C = C_0 + \sum_h n_h \cdot c_h$$

La asignación de muestra que minimiza la varianza de $\bar{y}_{st}$ suponiendo una MAS en cada estrato, es

$$\frac{n_h}{n} = \frac{N_h \cdot S_h/\sqrt{c_h}}{\sum_k N_k S_k/\sqrt{c_k}}$$

Estratificación

Asignación de Neyman

Si queremos suavizar la diferencia entre los S_h o entre los N_h una alternativa es

$$n_h = n \cdot \frac{N_h^{\beta} \cdot S_h^{\alpha} / \sqrt{c_k}}{\sum_k N_k^{\beta} S_k^{\alpha} / \sqrt{c_k}}$$

con $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$

Estratificación

Efecto de la estratificación

En general la estratificación tiene un efecto moderado en la estimación de medias o totales de variables dicotómicas (proporciones)
Mientras que reduce fuertemente la varianza en el caso de encuestas a empresas por ejemplo

Estratificación: Ejercicio I

En cierta universo de N=4 unidades se desea estimar el total de la variable Y, seleccionando una muestra aleatoria estratificada de tamaño n=2, estratificando por la variable Zona. Hallar Varianza , CV y deff del estimador de Horvitz-Thompson del total

Zona	Unidad	Y
A	1	2
A	2	3
B	3	5
B	4	6
B	5	10

Estratificación: Ejercicio II

En cierta localidad se selecciona una muestra aleatoria de hogares para estimar el total y proporción de hogares pobres. Se estratifica un marco de hogares por Zona y en cada Zona se selecciona una MAS de hogares. Suponiendo que la tabla siguiente contiene totales de hogares y los resultados de la encuesta, estimar:
Total de hogares pobres (mediante H-T) y los correspondientes CV y deff
Proporción de hogares pobres (mediante H-T) y los correspondientes CV y deff
Proporción de hogares pobres (mediante estimador de razón)
Proporción de población pobre

Zona	Total Hogares	n	Hogares pobres en la muestra	Población pobre en la muestra
A	10,000	200	50	150
B	5,000	100	20	40
C	15,000	300	25	100

Tabla de Contenidos

Estratificación con una variable continua

- Estratificar un listado de empresas de cierta rama de actividad según facturación
- Estratificar un listado de localidades de cierta región según población
- Estratificar un listado de escuelas según puntaje medio obtenido en una prueba
- Estratificar un listado de hogares según ingreso

1. Cuántos estratos?
2. Qué cortes hacer?

En principio para minimizar la varianza. Qué estimadores?

Estratificación con una variable continua

- Estratificar un listado de empresas de cierta rama de actividad según facturación
- Estratificar un listado de localidades de cierta región según población
- Estratificar un listado de escuelas según puntaje medio obtenido en una prueba
- Estratificar un listado de hogares según ingreso

1. Cuántos estratos?
2. Qué cortes hacer?

En principio para minimizar la varianza. Qué estimadores?

Estratificación con una variable continua

- Estratificar un listado de empresas de cierta rama de actividad según facturación
- Estratificar un listado de localidades de cierta región según población
- Estratificar un listado de escuelas según puntaje medio obtenido en una prueba
- Estratificar un listado de hogares según ingreso

1. Cuántos estratos?
2. Qué cortes hacer?

En principio para minimizar la varianza. Qué estimadores?

Estratificación con una variable continua

- Estratificar un listado de empresas de cierta rama de actividad según facturación
- Estratificar un listado de localidades de cierta región según población
- Estratificar un listado de escuelas según puntaje medio obtenido en una prueba
- Estratificar un listado de hogares según ingreso

1. Cuántos estratos?
2. Qué cortes hacer?

En principio para minimizar la varianza. Qué estimadores?

Estratificación con una variable continua

- Estratificar un listado de empresas de cierta rama de actividad según facturación
- Estratificar un listado de localidades de cierta región según población
- Estratificar un listado de escuelas según puntaje medio obtenido en una prueba
- Estratificar un listado de hogares según ingreso

1. Cuántos estratos?
2. Qué cortes hacer?

En principio para minimizar la varianza. Qué estimadores?

Estratificación con una variable continua

- Estratificar un listado de empresas de cierta rama de actividad según facturación
- Estratificar un listado de localidades de cierta región según población
- Estratificar un listado de escuelas según puntaje medio obtenido en una prueba
- Estratificar un listado de hogares según ingreso

1. Cuántos estratos?
2. Qué cortes hacer?

En principio para minimizar la varianza. Qué estimadores?

Estratificación con una variable continua

Variable estratificadora muy asimétrica

Estrato autorepresentado

Es habitual en las muestras de empresas que las unidades de mayor tamaño (según alguna medida) sean seleccionadas con probabilidad 1:

Estratificación con una variable continua

Variable estratificadora muy asimétrica

Y puede haber un estrato *take none*: unidades con probabilidad cero de ser seleccionadas:

Unidades que aportan muy poco al total general

Complicadas de relevar

Que no serán ubicadas (muerte de empresas)

Estratificación con una variable continua

Variable estratificadora muy asimétrica

Y puede haber un estrato *take none*: unidades con probabilidad cero de ser seleccionadas:

Unidades que aportan muy poco al total general

Complicadas de relevar

Que no serán ubicadas (muerte de empresas)

Estratificación con una variable continua

Variable estratificadora muy asimétrica

Y puede haber un estrato *take none*: unidades con probabilidad cero de ser seleccionadas:

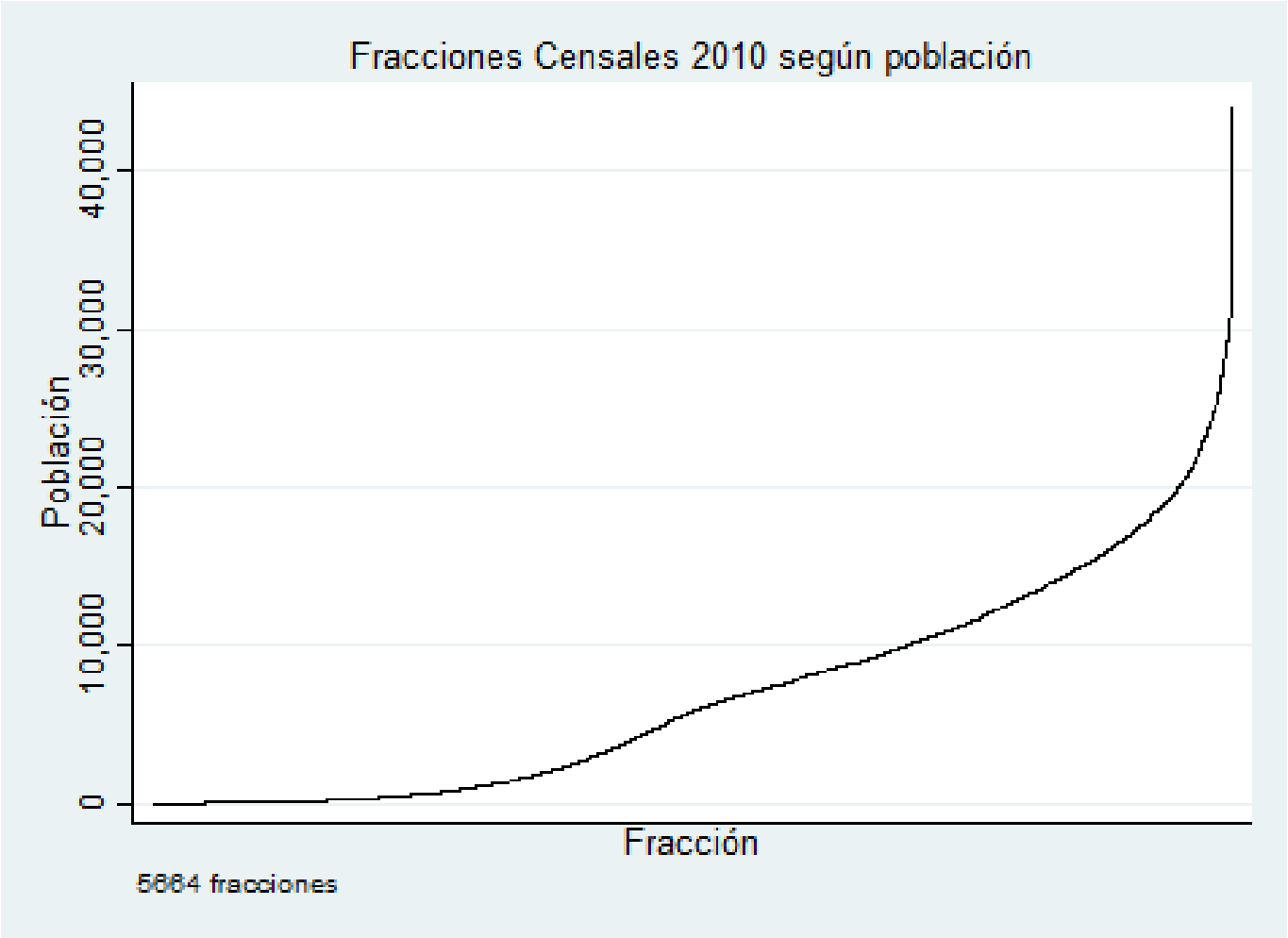
Unidades que aportan muy poco al total general

Complicadas de relevar

Que no serán ubicadas (muerte de empresas)

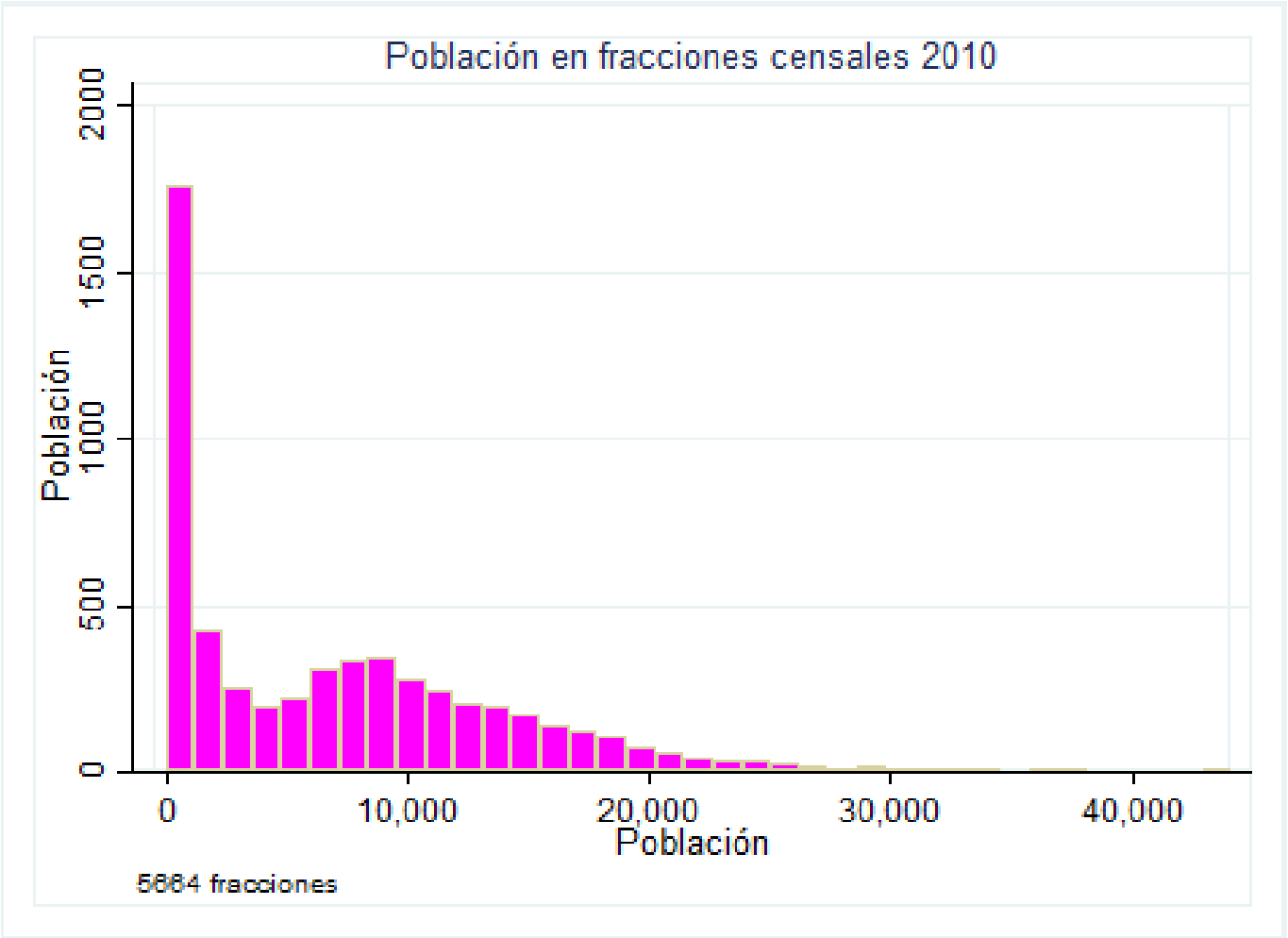
Estratificación con una variable continua

Estratificación universo de fracciones censales 2010



Estratificación con una variable continua

Estratificación universo de fracciones censales 2010



Estratificación con una variable continua

Antecedentes

Michel Hidiroglou(1986). Estratificar en dos estratos, uno de ellos *autorepresentado*, para minimizar la varianza de la estimación de un total, con n prefijado

Michel Hidiroglou - Pierre Lavallée (1988). Estratificar en H estratos, uno *autorepresentado*, para minimizar la varianza de la estimación de un total, con n prefijado.

En estos dos métodos suponemos que la variables estratificadora es la variable bajo estudio.
Y que en cada estrato seleccionamos una MAS.

Estratificación con una variable continua

Antecedentes

Michel Hidiroglou(1986). Estratificar en dos estratos, uno de ellos *autorepresentado*, para minimizar la varianza de la estimación de un total, con n prefijado

Michel Hidiroglou - Pierre Lavallée (1988). Estratificar en H estratos, uno *autorepresentado*, para minimizar la varianza de la estimación de un total, con n prefijado.

En estos dos métodos suponemos que la variables estratificadora es la variable bajo estudio.
Y que en cada estrato seleccionamos una MAS.

Estratificación con una variable continua

Antecedentes

Michel Hidiroglou(1986). Estratificar en dos estratos, uno de ellos *autorepresentado*, para minimizar la varianza de la estimación de un total, con n prefijado

Michel Hidiroglou - Pierre Lavallée (1988). Estratificar en H estratos, uno *autorepresentado*, para minimizar la varianza de la estimación de un total, con n prefijado.

En estos dos métodos suponemos que la variables estratificadora es la variable bajo estudio.
Y que en cada estrato seleccionamos una MAS.

Estratificación con una variable continua

Caso típico

Caso típico

Estratificar un listado de empresas provenientes de un censo anterior
Posibles problemas:

1. Muerte de empresas
2. Cambios en Y_i
3. Nacimiento de empresas

stratification permite simular algunas alternativas de los casos 1. y 2.

Estratificación con una variable continua

Caso típico

Caso típico

Estratificar un listado de empresas provenientes de un censo anterior
Posibles problemas:

1. Muerte de empresas
2. Cambios en Y_i
3. Nacimiento de empresas

stratification permite simular algunas alternativas de los casos 1. y 2.

Estratificación con una variable continua

Caso típico

Caso típico

Estratificar un listado de empresas provenientes de un censo anterior
Posibles problemas:

1. Muerte de empresas
2. Cambios en Y_i
3. Nacimiento de empresas

stratification permite simular algunas alternativas de los casos 1. y 2.

Estratificación con una variable continua

Caso típico

Caso típico

Estratificar un listado de empresas provenientes de un censo anterior
Posibles problemas:

1. Muerte de empresas
2. Cambios en Y_i
3. Nacimiento de empresas

stratification permite simular algunas alternativas de los casos 1. y 2.

Estratificación con una variable continua

Caso típico

Caso típico

Estratificar un listado de empresas provenientes de un censo anterior
Posibles problemas:

1. Muerte de empresas
2. Cambios en Y_i
3. Nacimiento de empresas

stratification permite simular algunas alternativas de los casos 1. y 2.

Estratificación con una variable continua

Caso típico

Caso típico

Estratificar un listado de empresas provenientes de un censo anterior
Posibles problemas:

1. Muerte de empresas
2. Cambios en Y_i
3. Nacimiento de empresas

stratification permite simular algunas alternativas de los casos 1. y 2.

Estratificación con una variable continua

Alternativa sencilla: Estratificación geométrica

Estratificación Geométrica

Los límites de los L estratos se determinan mediante

a $= k_0$ mínimo valor de Y
$k_h = a \cdot r^h$, con $r = (k_L/k_0)^{(}\, 1/L)$

El supuesto es que en cada estrato la distribución de la variables es *aproximadamente* uniforme

Estratificación con **stratification**

Opciones básicas

Comando strata.LH (LH de Lavallée-Hidiroglou)

strata.LH(x → Variable estratificadora, Ls → Nro de estratos a considerar (por defecto 3) ,
n → Tamaño de muestra, CV → CV. Debe indicarse n o CV,
certain → Vector posicion de elementos de inclusión forzosa. Default es NULL ,
takeall → Cantidad de estratos de inclusión forzosa. Default:0 ,
takenone → 1/0 , bias.penality → penalidad por el sesgo de takenone (entre 0 y 1) ,
model → Modelo que relaciona la variable estratificadora y la variable objetivo. Si es la misma se utiliza NONE ,
alloc → Lista que indica la forma de asignacion: q1, q2 y q3 . Son los exponentes de N_h, $\bar{Y}_h$ y S_h^2 respectivamente en la formula de asignacion general de Hidiroglou y Srinath
→ Asignacion de Neyman : (0.5, 0, 0.5) (default)
→ Asignacion proporcional: (0.5, 0.5, 0)
. . .)