Maestría en Generación de Información Estadística

Teoría y Técnicas de Muestreo

Augusto E. Hoszowski

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

El Muestreo Probabilístico

Jerzy Neyman

Jerzy Neyman (1934): Muestreo probabilístico

On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection

$\Rightarrow$

Estratificación

Asignación no proporcional de la muestra:

La mejor muestra no es siempre la que respeta las proporcionalidades

El Muestreo Probabilístico

Jerzy Neyman

Jerzy Neyman (1934): Muestreo probabilístico
On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection
$\Rightarrow$

- Estratificación
- Asignación no proporcional de la muestra:
  La mejor muestra no es siempre la que respeta las proporcionalidades

El Muestreo Probabilístico

Jerzy Neyman

Jerzy Neyman (1934): Muestreo probabilístico
On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection

$\Rightarrow$

- Estratificación
- Asignación no proporcional de la muestra:
- La mejor muestra no es siempre la que respeta las proporcionalidades

El Muestreo Probabilístico

Jerzy Neyman

Jerzy Neyman (1934): Muestreo probabilístico
On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection
$\Rightarrow$

- Estratificación
- Asignación no proporcional de la muestra:
- La mejor muestra no es siempre la que respeta las proporcionalidades

Tabla de Contenidos

Definición de Muestreo Probabilístico

Muestra probabilística

Una muestra (o diseño muestral) es *probabilística* si todo elemento del universo U tiene una probabilidad positiva y *conocida* de ser seleccionado

Métodos no probabilísticos

Opinión pública

Muestreo por cuotas
Encuestas telefónicas
Encuestas web
Bola de nieve

Medicina

Selección por conveniencia

Métodos no probabilísticos

Opinión pública

Muestreo por cuotas
Encuestas telefónicas
Encuestas web
Bola de nieve

Medicina

Selección por conveniencia

Métodos no probabilísticos

Opinión pública

Muestreo por cuotas

Encuestas telefónicas

Encuestas web

Bola de nieve

Medicina

Selección por conveniencia

Métodos no probabilísticos

Opinión pública

Muestreo por cuotas

Encuestas telefónicas

Encuestas web

Bola de nieve

Medicina

Selección por conveniencia

Métodos no probabilísticos

Opinión pública

Muestreo por cuotas
Encuestas telefónicas
Encuestas web
Bola de nieve

Medicina

Selección por conveniencia

Métodos no probabilísticos

Opinión pública

Muestreo por cuotas
Encuestas telefónicas
Encuestas web
Bola de nieve

Medicina

Selección por conveniencia

Métodos no probabilísticos

Porqué el muestreo por cuotas no es un método probabilístico?

Consecuencias del Muestreo Probabilístico

Dos ventajas importantes

- Ausencia de *sesgos de selección*
- Posibilidad de estimar medidas de precisión: CV, intervalos de confianza, etc.

siempre que podamos medir a la unidad seleccionada la/s variable/s de interés

Consecuencias del Muestreo Probabilístico

Dos ventajas importantes

- Ausencia de *sesgos de selección*
- Posibilidad de estimar medidas de precisión: CV, intervalos de confianza, etc.

siempre que podamos medir a la unidad seleccionada la/s variable/s de interés

Consecuencias del Muestreo Probabilístico

Dos ventajas importantes

- Ausencia de *sesgos de selección*
- Posibilidad de estimar medidas de precisión: CV, intervalos de confianza, etc.

siempre que podamos medir a la unidad seleccionada la/s variable/s de interés

Consecuencias del Muestreo Probabilístico

Dos ventajas importantes

- Ausencia de *sesgos de selección*
- Posibilidad de estimar medidas de precisión: CV, intervalos de confianza, etc.

siempre que podamos medir a la unidad seleccionada la/s variable/s de interés

Consecuencias del Muestreo Probabilístico

Dos ventajas importantes

- Ausencia de *sesgos de selección*
- Posibilidad de estimar medidas de precisión: CV, intervalos de confianza, etc.

siempre que podamos medir a la unidad seleccionada la/s variable/s de interés

Dificultad de seleccionar una muestra probabilística:

Algunos ejemplos

- Personas afectadas de una enfermedad rara
- Personas sin techo
- Temáticas *sensibles*
  - Dificultad para detectar las unidades
  - Detectamos la unidad pero esta no responde
  - *Poblaciones elusivas*

Dificultad de seleccionar una muestra probabilística:

Algunos ejemplos

- Personas afectadas de una enfermedad rara
- Personas sin techo
- Temáticas *sensibles*
  - Dificultad para detectar las unidades
  - Detectamos la unidad pero esta no responde
  - *Poblaciones elusivas*

Dificultad de seleccionar una muestra probabilística:

Algunos ejemplos

- Personas afectadas de una enfermedad rara
- Personas sin techo
- Temáticas *sensibles*
  - Dificultad para detectar las unidades
  - Detectamos la unidad pero esta no responde
  - *Poblaciones elusivas*

Dificultad de seleccionar una muestra probabilística:

Algunos ejemplos

- Personas afectadas de una enfermedad rara
- Personas sin techo
- Temáticas *sensibles*
  - Dificultad para detectar las unidades
  - Detectamos la unidad pero esta no responde
  - *Poblaciones elusivas*

Dificultad de seleccionar una muestra probabilística:

Algunos ejemplos

- Personas afectadas de una enfermedad rara
- Personas sin techo
- Temáticas *sensibles*
  - Dificultad para detectar las unidades
  - Detectamos la unidad pero esta no responde
  - *Poblaciones elusivas*

Dificultad de seleccionar una muestra probabilística:

Algunos ejemplos

- Personas afectadas de una enfermedad rara
- Personas sin techo
- Temáticas *sensibles*
  - Dificultad para detectar las unidades
  - Detectamos la unidad pero esta no responde
  - *Poblaciones elusivas*

Dificultad de seleccionar una muestra probabilística

Poblaciónes raras

Se desea seleccionar una muestra aleatoria de personas con cierta característica poco frecuente, que se da en el **0.01 %** de las personas.
Los analistas necesitan recolectar una muestra de tamaño aproximado n=2000 personas.
Cuántos hogares deberíamos visitar para llegar a ese tamaño de muestra?
Podemos suponer un tamaño medio de hogar de cuatro personas y que todos los hogares y personas responden.

$$n_h = 2000 \cdot (1/0{,}01) \cdot (1/4) = 50,000$$

Sería una muestra probabilística, pero el costo puede ser prohibitivo

Dificultad de seleccionar una muestra probabilística

Poblaciónes raras

Se desea seleccionar una muestra aleatoria de personas con cierta característica poco frecuente, que se da en el **0.01 %** de las personas. Los analistas necesitan recolectar una muestra de tamaño aproximado n=2000 personas.

Cuántos hogares deberíamos visitar para llegar a ese tamaño de muestra? Podemos suponer un tamaño medio de hogar de cuatro personas y que todos los hogares y personas responden.

$$n_h = 2000 \cdot (1/0{,}01) \cdot (1/4) = 50,000$$

Sería una muestra probabilística, pero el costo puede ser prohibitivo

Dificultad de seleccionar una muestra probabilística

Poblaciónes raras

Se desea seleccionar una muestra aleatoria de personas con cierta característica poco frecuente, que se da en el **0.01 %** de las personas. Los analistas necesitan recolectar una muestra de tamaño aproximado n=2000 personas.
Cuántos hogares deberíamos visitar para llegar a ese tamaño de muestra?
Podemos suponer un tamaño medio de hogar de cuatro personas y que todos los hogares y personas responden.

$$n_h = 2000 \cdot (1/0{,}01) \cdot (1/4) = 50,000$$

Sería una muestra probabilística, pero el costo puede ser prohibitivo

Dificultad de seleccionar una muestra probabilística

Poblaciónes raras

Se desea seleccionar una muestra aleatoria de personas con cierta característica poco frecuente, que se da en el **0.01 %** de las personas.
Los analistas necesitan recolectar una muestra de tamaño aproximado n=2000 personas.
Cuántos hogares deberíamos visitar para llegar a ese tamaño de muestra?
Podemos suponer un tamaño medio de hogar de cuatro personas y que todos los hogares y personas responden.

$$n_h = 2000 \cdot (1/0{,}01) \cdot (1/4) = 50,000$$

Sería una muestra probabilística, pero el costo puede ser prohibitivo

Dificultad de seleccionar una muestra probabilística

Temáticas sensibles

Se desea seleccionar una muestra aleatoria de personas con un hábito relativamente frecuente , *pero condenable socialmente*, de un listado de 10000 personas.

Esa conducta se supone se da aproximadamente en el **0.1 %** de las personas. Por estudios previos se sabe que el **50 %** de las personas con esa hábito no responderán la encuesta o falsearán la respuesta. Los analistas necesitan recolectar una muestra de tamaño aproximado n=200 personas.

Se selecciona una muestra aleatoria de 4600 personas, de las cuales 2200 responden la encuesta, y de los cuales 160 reconocen esa conducta. Se trata de una muestra aleatoria de personas con ese hábito??

Dificultad de seleccionar una muestra probabilística

Temáticas sensibles

Se desea seleccionar una muestra aleatoria de personas con un hábito relativamente frecuente , *pero condenable socialmente*, de un listado de 10000 personas.

Esa conducta se supone se da aproximadamente en el **0.1 %** de las personas. Por estudios previos se sabe que el **50 %** de las personas con esa hábito no responderán la encuesta o falsearán la respuesta. Los analistas necesitan recolectar una muestra de tamaño aproximado n=200 personas.

Se selecciona una muestra aleatoria de 4600 personas, de las cuales 2200 responden la encuesta, y de los cuales 160 reconocen esa conducta. Se trata de una muestra aleatoria de personas con ese hábito??

Dificultad de seleccionar una muestra probabilística

Temáticas sensibles

Se desea seleccionar una muestra aleatoria de personas con un hábito relativamente frecuente , *pero condenable socialmente*, de un listado de 10000 personas.

Esa conducta se supone se da aproximadamente en el **0.1 %** de las personas. Por estudios previos se sabe que el **50 %** de las personas con esa hábito no responderán la encuesta o falsearán la respuesta. Los analistas necesitan recolectar una muestra de tamaño aproximado n=200 personas.

Se selecciona una muestra aleatoria de 4600 personas, de las cuales 2200 responden la encuesta, y de los cuales 160 reconocen esa conducta. Se trata de una muestra aleatoria de personas con ese hábito??

Dificultad de seleccionar una muestra probabilística

Poblaciónes elusivas

Las dos ventajas del muestreo probabilístico:

- Ausencia de sesgo de selección
- Posibilidad de calcular errores de muestreo, intervalos de confianza, etc

desaparecen con la imposibilidad de medir lo que nos interesa a la unidad seleccionada (no respuesta, error de medición, etc)

Dificultad de seleccionar una muestra probabilística

Poblaciónes elusivas

Las dos ventajas del muestreo probabilístico:

- Ausencia de sesgo de selección
- Posibilidad de calcular errores de muestreo, intervalos de confianza, etc

desaparecen con la imposibilidad de medir lo que nos interesa a la unidad seleccionada (no respuesta, error de medición, etc)

Inferencia a partir de muestras no probabilíticas

Ejemplos

- Muestras por conveniencia en medicina y epidemiología
- Muestras electorales: muestra probabilística de *puntos muestra* y muestreo por cuotas de sexo y tramo de edad en cada punto muestra
- Arqueología

Inferencia a partir de muestras no probabilíticas

Ejemplos

- Muestras por conveniencia en medicina y epidemiología
- Muestras electorales: muestra probabilística de *puntos muestra* y muestreo por cuotas de sexo y tramo de edad en cada punto muestra
- Arqueología

Inferencia a partir de muestras no probabilíticas

Ejemplos

- Muestras por conveniencia en medicina y epidemiología
- Muestras electorales: muestra probabilística de *puntos muestra* y muestreo por cuotas de sexo y tramo de edad en cada punto muestra
- Arqueología

Del universo bajo estudio a la estimación

Esquema usual

Universo bajo estudio

Variable/s bajo estudio - Parámetro a estimar (total, media, razón...)

Marco(s) de muestreo
Unidad de muestreo
Unidad de análisis

Diseño muestral / Muestra
Estimador

Del universo bajo estudio a la estimación

Esquema usual

Universo bajo estudio

Variable/s bajo estudio - Parámetro a estimar (total, media, razón...)

Marco(s) de muestreo
Unidad de muestreo
Unidad de análisis

Diseño muestral / Muestra
Estimador

Del universo bajo estudio a la estimación

Esquema usual

Universo bajo estudio

Variable/s bajo estudio - Parámetro a estimar (total, media, razón...)

Marco(s) de muestreo

Unidad de muestreo

Unidad de análisis

Diseño muestral / Muestra

Estimador

Del universo bajo estudio a la estimación

Esquema usual

Universo bajo estudio

Variable/s bajo estudio - Parámetro a estimar (total, media, razón...)

Marco(s) de muestreo
Unidad de muestreo
Unidad de análisis

Diseño muestral / Muestra
Estimador

Del universo bajo estudio a la estimación

Esquema usual

Universo bajo estudio

Variable/s bajo estudio - Parámetro a estimar (total, media, razón...)

Marco(s) de muestreo
Unidad de muestreo
Unidad de análisis

Diseño muestral / Muestra
Estimador

Del universo bajo estudio a la estimación

Esquema usual

Universo bajo estudio

Variable/s bajo estudio - Parámetro a estimar (total, media, razón...)

Marco(s) de muestreo
Unidad de muestreo
Unidad de análisis

Diseño muestral / Muestra
Estimador

Del universo bajo estudio a la estimación

Esquema usual

Universo bajo estudio

Variable/s bajo estudio - Parámetro a estimar (total, media, razón...)

Marco(s) de muestreo
Unidad de muestreo
Unidad de análisis

Diseño muestral / Muestra
Estimador

Diseño muestral y estimador

Informalmente

Diseño muestral: Cómo seleccionar la muestra (incluyendo tamaño de muestra)

Estimador: Una vez la encuesta relevada, qué operaciones hacemos para estimar los parámetros de interés

Marco de Muestreo

Marco de muestreo

Conjunto de listados, archivos, cartografía, etc. a partir de los cuales podemos seleccionar una muestra probabilística

Los marcos de muestreo no tiene porqué contener un listado de todas las unidades finales de muestreo

Marco de Muestreo

Ejemplo

La Encuesta Permanente de Hogares (INDEC)
Muestra probabilística de aproximadamente 17000 hogares efectivos en cada trimestre

La muestra de viviendas se selecciona en dos etapas:

- Primera etapa: Selección de una muestra aleatoria de radios censales. Marco de selección de esta muestra: Listado de (2000) radios censales generados por el CEN2010. **Marco Primario**
- En cada radio seleccionado se construye el listado de viviendas particulares del radio. **Marco Secundario**. A partir del marco secundario se selecciona la muestra aleatoria de viviendas

Marco de Muestreo

Ejemplo

La Encuesta Permanente de Hogares (INDEC)
Muestra probabilística de aproximadamente 17000 hogares efectivos en cada trimestre

La muestra de viviendas se selecciona en dos etapas:

- Primera etapa: Selección de una muestra aleatoria de radios censales. Marco de selección de esta muestra: Listado de (2000) radios censales generados por el CEN2010. **Marco Primario**
- En cada radio seleccionado se construye el listado de viviendas particulares del radio. **Marco Secundario**. A partir del marco secundario se selecciona la muestra aleatoria de viviendas

Diseño muestral

- Qué subconjuntos de U son las muestras posibles?
- Con qué probabilidad seleccionar cada muestra?

$\Rightarrow$ P(s)

- Con qué mecanismo seleccionar la muestra?

Condición: Todo elemento del universo está en alguna muestra posible

Diseño muestral

- Qué subconjuntos de U son las muestras posibles?
- Con qué probabilidad seleccionar cada muestra?

$\Rightarrow$ P(s)

- Con qué mecanismo seleccionar la muestra?

Condición: Todo elemento del universo está en alguna muestra posible

Diseño muestral

- Qué subconjuntos de U son las muestras posibles?
- Con qué probabilidad seleccionar cada muestra?

$\Rightarrow$ P(s)

- Con qué mecanismo seleccionar la muestra?

Condición: Todo elemento del universo está en alguna muestra posible

Diseño muestral

- Qué subconjuntos de U son las muestras posibles?
- Con qué probabilidad seleccionar cada muestra?

$\Rightarrow$ P(s)

- Con qué mecanismo seleccionar la muestra?

Condición: Todo elemento del universo está en alguna muestra posible

Diseño muestral

- Qué subconjuntos de U son las muestras posibles?
- Con qué probabilidad seleccionar cada muestra?

$\Rightarrow$ P(s)

- Con qué mecanismo seleccionar la muestra?

Condición: Todo elemento del universo está en alguna muestra posible

Diseño muestral

- Qué subconjuntos de U son las muestras posibles?
- Con qué probabilidad seleccionar cada muestra?

$\Rightarrow$ P(s)

- Con qué mecanismo seleccionar la muestra?

Condición: Todo elemento del universo está en alguna muestra posible

Diseño muestral

- Qué subconjuntos de U son las muestras posibles?
- Con qué probabilidad seleccionar cada muestra?

$\Rightarrow$ P(s)

- Con qué mecanismo seleccionar la muestra?

Condición: Todo elemento del universo está en alguna muestra posible

Probabilidades de Selección

Ejemplo

En la práctica no nos interesa la probabilidad P(s)

Sino la probabilidad de que un elemento i sea seleccionado

Probabilidad de selección de primer orden $\pi_i = \sum_{s \ni i} P(s)$

Probabilidad de selección de segundo orden $\pi_{ij} = \sum_{s \ni i,j} P(s)$

Factor de expansión $F_i = \frac{1}{\pi_i}$

Probabilidades de Selección

Ejemplo

En la práctica no nos interesa la probabilidad P(s)

Sino la probabilidad de que un elemento *i* sea seleccionado

Probabilidad de selección de primer orden $\pi_i = \sum_{s \ni i} P(s)$

Probabilidad de selección de segundo orden $\pi_{ij} = \sum_{s \ni i,j} P(s)$

Factor de expansión $F_i = \frac{1}{\pi_i}$

Probabilidades de Selección

Ejemplo

En la práctica no nos interesa la probabilidad P(s)

Sino la probabilidad de que un elemento i sea seleccionado

Probabilidad de selección de primer orden $\pi_i = \sum_{s \ni i} P(s)$

Probabilidad de selección de segundo orden $\pi_{ij} = \sum_{s \ni i,j} P(s)$

Factor de expansión $F_i = \frac{1}{\pi_i}$

Probabilidades de Selección

Ejemplo

En la práctica no nos interesa la probabilidad P(s)

Sino la probabilidad de que un elemento *i* sea seleccionado

Probabilidad de selección de primer orden $\pi_i = \sum_{s \ni i} P(s)$

Probabilidad de selección de segundo orden $\pi_{ij} = \sum_{s \ni i,j} P(s)$

Factor de expansión $F_i = \frac{1}{\pi_i}$

Probabilidades de Selección

Ejemplo

En la práctica no nos interesa la probabilidad P(s)

Sino la probabilidad de que un elemento *i* sea seleccionado

Probabilidad de selección de primer orden $\pi_i = \sum_{s \ni i} P(s)$

Probabilidad de selección de segundo orden $\pi_{ij} = \sum_{s \ni i,j} P(s)$

Factor de expansión $F_i = \frac{1}{\pi_i}$

Dos expresiones importantes

$$\sum_{i \in U} \pi_i = E(n(s))$$

$$\sum_{i \in s} 1/\pi_i \approx N$$

Dos expresiones importantes

$$\sum_{i \in U} \pi_i = E(n(s))$$

$$\sum_{i \in s} 1/\pi_i \approx N$$

Probabilidades de Selección

Ejemplo

$$\text{U}= \{Juan, Pedro, Maria\}$$

$$\text{S}= \{\{Juan, Pedro\}, \{Juan, Maria\}, \{Pedro, Maria\}\}$$

$$\text{n(s)}=2\ , \quad \text{card(S)}=3$$

Equiprobales: P(s)=1/3 para cada una de las muestras **s**

$$\pi_{Juan} = 2/3 \quad \pi_{Pedro} = 2/3 \quad \pi_{Maria} = 2/3$$

$$\sum_{i \in U} \pi_i = 3 \cdot 2/3 = 2, \qquad F_i = 1/\pi_i = 3/2$$

Probabilidades de Selección
Ejemplo

$$U = \{Juan, Pedro, Maria\}$$

$$S = \{\{Juan, Pedro\}, \{Juan, Maria\}, \{Pedro, Maria\}\}$$

n(s)=2 , card(S)=3

Equiprobales: P(s)=1/3 para cada una de las muestras **s**

$$\pi_{Juan} = 2/3 \quad \pi_{Pedro} = 2/3 \quad \pi_{Maria} = 2/3$$

$$\sum_{i \in U} \pi_i = 3 \cdot 2/3 = 2, \qquad F_i = 1/\pi_i = 3/2$$

Probabilidades de Selección

Ejemplo

$$U = \{Juan, Pedro, Maria\}$$

$$S = \{\{Juan, Pedro\}, \{Juan, Maria\}, \{Pedro, Maria\}\}$$

$$n(s)=2 \ , \quad card(S)=3$$

Equiprobales: P(s)=1/3 para cada una de las muestras **s**

$$\pi_{Juan} = 2/3 \quad \pi_{Pedro} = 2/3 \quad \pi_{Maria} = 2/3$$

$$\sum_{i \in U} \pi_i = 3 \cdot 2/3 = 2, \qquad F_i = 1/\pi_i = 3/2$$

Probabilidades de Selección

Ejemplo

$$U = \{Juan, Pedro, Maria\}$$

$$S = \{\{Juan, Pedro\}, \{Juan, Maria\}, \{Pedro, Maria\}\}$$

$$n(s)=2 \ , \quad card(S)=3$$

Equiprobales: P(s)=1/3 para cada una de las muestras **s**

$$\pi_{Juan} = 2/3 \quad \pi_{Pedro} = 2/3 \quad \pi_{Maria} = 2/3$$

$$\sum_{i \in U} \pi_i = 3 \cdot 2/3 = 2, \qquad F_i = 1/\pi_i = 3/2$$

Probabilidades de Selección

Ejemplo

$$U = \{Juan, Pedro, Maria\}$$

$$S = \{\{Juan, Pedro\}, \{Juan, Maria\}, \{Pedro, Maria\}\}$$

$$n(s)=2 \;, \quad card(S)=3$$

Equiprobales: P(s)=1/3 para cada una de las muestras **s**

$$\pi_{Juan} = 2/3 \quad \pi_{Pedro} = 2/3 \quad \pi_{Maria} = 2/3$$

$$\sum_{i \in U} \pi_i = 3 \cdot 2/3 = 2, \qquad F_i = 1/\pi_i = 3/2$$

Probabilidades de Selección

Ejemplo

$$U = \{Juan, Pedro, Maria\}$$

$$S = \{\{Juan, Pedro\}, \{Juan, Maria\}, \{Pedro, Maria\}\}$$

$$n(s)=2 \;, \quad card(S)=3$$

Equiprobales: P(s)=1/3 para cada una de las muestras **s**

$$\pi_{Juan} = 2/3 \quad \pi_{Pedro} = 2/3 \quad \pi_{Maria} = 2/3$$

$$\sum_{i \in U} \pi_i = 3 \cdot 2/3 = 2, \qquad F_i = 1/\pi_i = 3/2$$

Probabilidades de Selección

Ejercicio

De un curso con siete alumnos se desea seleccionar una MAS(3) para una encuesta.
U= $\{Juan, Pedro, Maria, Tito, Lila, Carlos, Carla\}$
Hallar:

1. N
2. n
3. card(S) (cantidad de muestras posibles)
4. P(s) (probabilidad de selección de cada muestra)
5. π_i (probabilidad de selección de cada alumnos)
6. F_i (factor de expansión de cada alumno)

Probabilidades de Selección

Ejercicio

De un curso con siete alumnos se desea seleccionar una MAS(3) para una encuesta.
U= $\{Juan, Pedro, Maria, Tito, Lila, Carlos, Carla\}$
Hallar:

1. N
2. n
3. card(S) (cantidad de muestras posibles)
4. P(s) (probabilidad de selección de cada muestra)
5. π_i (probabilidad de selección de cada alumnos)
6. F_i (factor de expansión de cada alumno)

Probabilidades de Selección

Ejercicio

De un curso con siete alumnos se desea seleccionar una MAS(3) para una encuesta.
U= $\{Juan, Pedro, Maria, Tito, Lila, Carlos, Carla\}$
Hallar:

1. N
2. n
3. card(S) (cantidad de muestras posibles)
4. P(s) (probabilidad de selección de cada muestra)
5. π_i (probabilidad de selección de cada alumnos)
6. F_i (factor de expansión de cada alumno)

Probabilidades de Selección

Ejercicio

De un curso con siete alumnos se desea seleccionar una MAS(3) para una encuesta.

U= $\{Juan, Pedro, Maria, Tito, Lila, Carlos, Carla\}$

Hallar:

1. N
2. n
3. card(S) (cantidad de muestras posibles)
4. P(s) (probabilidad de selección de cada muestra)
5. π_i (probabilidad de selección de cada alumnos)
6. F_i (factor de expansión de cada alumno)

Probabilidades de Selección

Ejercicio

De un curso con siete alumnos se desea seleccionar una MAS(3) para una encuesta.
U= $\{Juan, Pedro, Maria, Tito, Lila, Carlos, Carla\}$
Hallar:

1. N
2. n
3. card(S) (cantidad de muestras posibles)
4. P(s) (probabilidad de selección de cada muestra)
5. π_i (probabilidad de selección de cada alumnos)
6. F_i (factor de expansión de cada alumno)

Probabilidades de Selección

Ejercicio

De un curso con siete alumnos se desea seleccionar una MAS(3) para una encuesta.
U= $\{Juan, Pedro, Maria, Tito, Lila, Carlos, Carla\}$
Hallar:

1. N
2. n
3. card(S) (cantidad de muestras posibles)
4. P(s) (probabilidad de selección de cada muestra)
5. π_i (probabilidad de selección de cada alumnos)
6. F_i (factor de expansión de cada alumno)

Probabilidades de Selección

Ejercicio

De un curso con siete alumnos se desea seleccionar una MAS(3) para una encuesta.
U= $\{Juan, Pedro, Maria, Tito, Lila, Carlos, Carla\}$
Hallar:

1. N
2. n
3. card(S) (cantidad de muestras posibles)
4. P(s) (probabilidad de selección de cada muestra)
5. π_i (probabilidad de selección de cada alumnos)
6. F_i (factor de expansión de cada alumno)

Tabla de Contenidos

Muestreo Aleatorio Simple

MAS(n, N)

U: Universo finito

S: Todos los subconjuntos de U de tamaño n

Todas las muestras son equiprobables $\Rightarrow$

$$P = 1/\binom{N}{n}$$

Muestreo Aleatorio Simple

MAS(n, N)

U: Universo finito

S: Todos los subconjuntos de U de tamaño n

Todas las muestras son equiprobables $\Rightarrow$

$$P = 1/\binom{N}{n}$$

Muestreo Aleatorio Simple

MAS(n, N)

U: Universo finito

S: Todos los subconjuntos de U de tamaño n

Todas las muestras son equiprobables $\Rightarrow$

$$P = 1/\binom{N}{n}$$

Muestreo Aleatorio Simple

MAS(n, N)

U: Universo finito

S: Todos los subconjuntos de U de tamaño n

Todas las muestras son equiprobables $\Rightarrow$

$$P = 1/\binom{N}{n}$$

Muestreo Aleatorio Simple

¿Cómo seleccionar una MAS?

- Ordenar 'aleatoriamente' las N unidades
- Seleccionar las n primeras (o las n últimas)
- Hay formas de seleccionar una MAS que no requieren reordenar el archivo

Muestreo Aleatorio Simple

¿Cómo seleccionar una MAS?

- Ordenar 'aleatoriamente' las N unidades
- Seleccionar las n primeras (o las n últimas)
- Hay formas de seleccionar una MAS que no requieren reordenar el archivo

Muestreo Aleatorio Simple

¿Cómo seleccionar una MAS?

- Ordenar 'aleatoriamente' las N unidades
- Seleccionar las n primeras (o las n últimas)
- Hay formas de seleccionar una MAS que no requieren reordenar el archivo

Muestreo Aleatorio Simple

¿Cómo seleccionar una MAS?

- Ordenar 'aleatoriamente' las N unidades
- Seleccionar las n primeras (o las n últimas)
- Hay formas de seleccionar una MAS que no requieren reordenar el archivo

Muestreo Aleatorio Simple

¿Cómo seleccionar una MAS?

- Ordenar 'aleatoriamente' las N unidades
- Seleccionar las n primeras (o las n últimas)
- Hay formas de seleccionar una MAS que no requieren reordenar el archivo

Muestreo Aleatorio Simple

¿Cómo seleccionar una MAS?

- Ordenar 'aleatoriamente' las N unidades
- Seleccionar las n primeras (o las n últimas)
- Hay formas de seleccionar una MAS que no requieren reordenar el archivo

Muestreo Aleatorio Simple

¿Cuántas muestras posibles?

$$\#S = \binom{N}{n}$$

n=2 , N=10 $\#S = 252$

n=20 , N=40

$\#S = 137{,}846{,}528{,}820$

Muestreo Aleatorio Simple

¿Cuántas muestras posibles?

$$\#S = \binom{N}{n}$$

n=2 , N=10 $\#S = 252$

n=20 , N=40

$\#S = 137{,}846{,}528{,}820$

Muestreo Aleatorio Simple

¿Cuántas muestras posibles?

N=10,000 , n=2,000

16555557523550355846089298375274833769686307809901076395012262452792783698032278066240824 99
5318806222772111210005426016020418065598071742873644401690919319335377095372278810640478652
0413339850951599929567643032803416164290936680088121145665954509987077953596641237451927908
5366245926365914714564881420608121809337614087081699727977511397993529081097631668957722811
0919596856791192334318746659600262757013932175504380326709133080441488983122983274425603811
7150720178689066894068507531026417815624234453195871008113238128934831837842040515600131726
0960391232798761539165046472416930838295530819010752780423265026993240120148179690854435505
2385528434122170804525355871678981192929859080385594746155471317881539915068852904830622278
6951038548880400191620565711291586700534540755526276938422405001345270278335726581375322976
0146113329991262165505009516699852893226357290535415654659407445246637262058188665134449520
4818520869743805424667419921175000623063780639488267205333549383140708983099413505886737083
3787098758113596190447219426121568324685764151601296948654893782399960327514764114467176417
1250601334540197087007822824805719350208982047634711216849131907359084143018261401250109369
1016194213027790687455272134662680020109302668903599687603532918015047819158239383782473199
4055511844267891121846403164857127885959745644323971338513739214928092232132691519007718752
7194667508917483274048937834514362518058947363924336172894596464292041241297602733962350332
2048092117538605933105935440926734806737558151600385206036037857107552265095615779105884699
3826792047806030332676423336065499519953076910418838626376480202828151673161942289092221049
2839024106999519123661634690999173102393364546370624825997336062993299235897148756965095480
2966835872346542760275822542764463354994480201097335259997004191897152445021872734562272174
4933664742499521140235707102217164259438766026322532351208348119475549696983427008567651685
9213559660367800804157236880443250995626931244887587281027299477537522287857862009983229788
0143251160834154923406732428021436134694019425135786782053546689135601921990424885927739965
7389914429390105240751239760865282709465029549690591863591028864648910033430399

(281 dígitos)

Probabilidades de Selección y Factores de expansión en el MAS

$$\pi_i = n/N$$

$$F_i = N/n$$

Tabla de Contenidos

Variables objetivo y medidas resumen

Variables *objetivo*

$$i \in U\,,\, i \;\longrightarrow Y_i = (Y_{i1}, Y_{i2}, ..., Y_{ik})$$

Deseamos estimar alguna *medida resumen* de las Y_i:

- Total $Y = \sum_{i=1}^{N} Y_i$
- Media $\overline{Y} = \sum_{i=1}^{N} Y_i / N$
- Cociente de dos totales $R = \frac{Y}{X} = \frac{\overline{Y}}{\overline{X}}$
- $S^2 = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N-1}$ (en R con $var(X)$)
- $S_{xy} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y}) \cdot (X_i - \bar{X})}{N-1}$ (en R con cov(X, Y))

Variables objetivo y medidas resumen

Variables *objetivo*

$$i \in U \,,\, i \longrightarrow Y_i = (Y_{i1}, Y_{i2}, ..., Y_{ik})$$

Deseamos estimar alguna *medida resumen* de las Y_i:

- Total $Y = \sum_{i=1}^{N} Y_i$
- Media $\overline{Y} = \sum_{i=1}^{N} Y_i/N$
- Cociente de dos totales $R = \frac{Y}{X} = \frac{\overline{Y}}{\overline{X}}$
- $S^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}$ (en R con $var(X)$)
- $S_{xy} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})\cdot(X_i - \bar{X})}{N-1}$ (en R con cov(X, Y))

Variables objetivo y medidas resumen

Variables *objetivo*

$$i \in U \, , i \longrightarrow Y_i = (Y_{i1}, Y_{i2}, ..., Y_{ik})$$

Deseamos estimar alguna *medida resumen* de las Y_i:

- Total $Y = \sum_{i=1}^{N} Y_i$
- Media $\overline{Y} = \sum_{i=1}^{N} Y_i / N$
- Cociente de dos totales $R = \frac{Y}{X} = \frac{\overline{Y}}{\overline{X}}$
- $S^2 = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N-1}$ (en R con $var(X)$)
- $S_{xy} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y}) \cdot (X_i - \bar{X})}{N-1}$ (en R con cov(X, Y))

Variables objetivo y medidas resumen

Variables *objetivo*

$$i \in U \, , i \longrightarrow Y_i = (Y_{i1}, Y_{i2}, ..., Y_{ik})$$

Deseamos estimar alguna *medida resumen* de las Y_i:

- Total $Y = \sum_{i=1}^{N} Y_i$
- Media $\overline{Y} = \sum_{i=1}^{N} Y_i/N$
- Cociente de dos totales $R = \frac{Y}{X} = \frac{\overline{Y}}{\overline{X}}$
- $S^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}$ (en R con $var\,(X)$)
- $S_{xy} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})\cdot(X_i - \bar{X})}{N-1}$ (en R con cov(X, Y))

Variables objetivo y medidas resumen

Variables *objetivo*

$$i \in U\,, i \longrightarrow Y_i = (Y_{i1}, Y_{i2}, ..., Y_{ik})$$

Deseamos estimar alguna *medida resumen* de las Y_i:

- Total $Y = \sum_{i=1}^{N} Y_i$
- Media $\overline{Y} = \sum_{i=1}^{N} Y_i/N$
- Cociente de dos totales $R = \frac{Y}{X} = \frac{\overline{Y}}{\overline{X}}$
- $S^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}$ (en R con *var* (X))
- $S_{xy} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})\cdot(X_i - \bar{X})}{N-1}$ (en R con cov(X, Y))

Variables objetivo y medidas resumen

Variables *objetivo*

$$i \in U\,, i \longrightarrow Y_i = (Y_{i1}, Y_{i2}, ..., Y_{ik})$$

Deseamos estimar alguna *medida resumen* de las Y_i:

- Total $Y = \sum_{i=1}^{N} Y_i$
- Media $\overline{Y} = \sum_{i=1}^{N} Y_i/N$
- Cociente de dos totales $R = \frac{Y}{X} = \frac{\overline{Y}}{\overline{X}}$
- $S^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}$ (en R con $var\,(X)$)
- $S_{xy} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})\cdot(X_i - \bar{X})}{N-1}$ (en R con cov(X, Y))

Estimadores muestrales

Parámetro θ. A estimar mediante los valores Y_i observados en $i \in s$

Estimadores muestrales

Variable cuyo valor depende de los valores $\{Y_i\}_{i \in s}$ de la muestra

Los estimadores muestrales son variables aleatorias

Sus valores dependen de la muestra seleccionada

Estimadores muestrales

Parámetro θ. A estimar mediante los valores Y_i observados en $i \in s$

Estimadores muestrales

Variable cuyo valor depende de los valores $\{Y_i\}_{i \in s}$ de la muestra

Los estimadores muestrales son variables aleatorias

Sus valores dependen de la muestra seleccionada

Estimadores muestrales

Parámetro θ. A estimar mediante los valores Y_i observados en $i \in s$

Estimadores muestrales

Variable cuyo valor depende de los valores $\{Y_i\}_{i \in s}$ de la muestra

Los estimadores muestrales son variables aleatorias

Sus valores dependen de la muestra seleccionada

Estimadores muestrales

Estimacion de la media poblacional

Ejemplo

Supongamos una variable bajo estudio Y_i

Seleccionamos s, una MAS(n,N)
A partir de s calculamos la media muestral:

$$\overline{y} = \frac{1}{n} \cdot \sum_{i \in s} Y_i$$

Para cada muestra s tendremos una estimación $\overline{y}$

Estimadores muestrales

Estimacion de la media poblacional

Ejemplo

Supongamos una variable bajo estudio Y_i

Seleccionamos s, una MAS(n,N)
A partir de s calculamos la media muestral:

$$\overline{y} = \frac{1}{n} \cdot \sum_{i \in s} Y_i$$

Para cada muestra s tendremos una estimación $\overline{y}$

Estimadores muestrales

Estimacion de la media poblacional

Ejemplo

Supongamos una variable bajo estudio Y_i

Seleccionamos s, una MAS(n,N)
A partir de s calculamos la media muestral:

$$\overline{y} = \frac{1}{n} \cdot \sum_{i \in s} Y_i$$

Para cada muestra s tendremos una estimación $\overline{y}$

Estimador de Horvitz-Thompson

En el Muestreo Aleatorio Simple

En el Muestreo Aleatorio Simple

$$\widehat{\bar{t}}_{\pi y} = \frac{1}{N}\sum_{i \in s}\frac{Y_i}{\pi_i} = \frac{1}{N}\sum_{i \in s}\frac{Y_i}{n/N} = \bar{y}$$

$$\widehat{t}_{\pi y} = N \cdot \bar{y}$$

Estimador de Horvitz-Thompson

En el Muestreo Aleatorio Simple

En el Muestreo Aleatorio Simple

$$\widehat{\overline{t}}_{\pi y} = \frac{1}{N}\sum_{i \in s}\frac{Y_i}{\pi_i} = \frac{1}{N}\sum_{i \in s}\frac{Y_i}{n/N} = \bar{y}$$

$$\widehat{t}_{\pi y} = N \cdot \bar{y}$$

Estimador de Horvitz-Thompson

En el Muestreo Aleatorio Simple

En el Muestreo Aleatorio Simple

$$\widehat{\bar{t}}_{\pi y} = \frac{1}{N}\sum_{i\in s}\frac{Y_i}{\pi_i} = \frac{1}{N}\sum_{i\in s}\frac{Y_i}{n/N} = \bar{y}$$

$$\widehat{t}_{\pi y} = N \cdot \bar{y}$$

Estimador de Horvitz-Thompson

En el Muestreo Aleatorio Simple

En el Muestreo Aleatorio Simple

$$\widehat{\bar{t}}_{\pi y} = \frac{1}{N}\sum_{i\in s}\frac{Y_i}{\pi_i} = \frac{1}{N}\sum_{i\in s}\frac{Y_i}{n/N} = \bar{y}$$

$$\widehat{t}_{\pi y} = N \cdot \bar{y}$$

Estimador de Horvitz-Thompson

En el Muestreo Aleatorio Simple

En el Muestreo Aleatorio Simple

$$\widehat{\bar{t}}_{\pi y} = \frac{1}{N} \sum_{i \in s} \frac{Y_i}{\pi_i} = \frac{1}{N} \sum_{i \in s} \frac{Y_i}{n/N} = \bar{y}$$

$$\widehat{t}_{\pi y} = N \cdot \bar{y}$$

Algunas propiedades de los estimadores

Propiedades deseables

- Que el *promedio* (ponderado) de los valores posibles del estimador coincidan con el parámetro a estimar: *Estimador insesgado*
- Que haya poca dispersión entre los distintos valores que toma el estimador: *Poca varianza*
- Que los distintos valores del estimador estén cerca del valor a estimar: *Poco error medio cuadrático*

$\Rightarrow$ *Estimador acurado*

Algunas propiedades de los estimadores

Propiedades deseables

- Que el *promedio* (ponderado) de los valores posibles del estimador coincidan con el parámetro a estimar: *Estimador insesgado*
- Que haya poca dispersión entre los distintos valores que toma el estimador: *Poca varianza*
- Que los distintos valores del estimador estén cerca del valor a estimar: *Poco error medio cuadrático*

$\Rightarrow$ *Estimador acurado*

Algunas propiedades de los estimadores

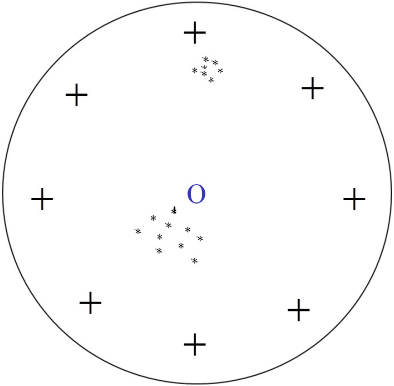
Propiedades deseables

- Que el *promedio* (ponderado) de los valores posibles del estimador coincidan con el parámetro a estimar: *Estimador insesgado*
- Que haya poca dispersión entre los distintos valores que toma el estimador: *Poca varianza*
- Que los distintos valores del estimador estén cerca del valor a estimar: *Poco error medio cuadrático*

$\Rightarrow$ *Estimador acurado*

Representación gráfica de un estimador

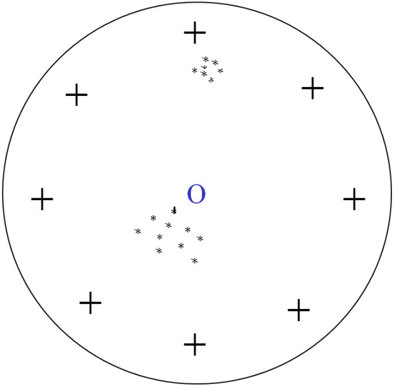
Dispersión, precisión y acuracidad



Varianza Desvío Standard
Sesgo
Error Medio Cuadrático

Representación gráfica de un estimador

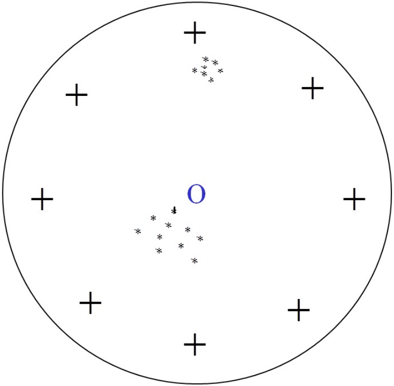
Dispersión, precisión y acuracidad



Varianza Desvío Standard
Sesgo
Error Medio Cuadrático

Representación gráfica de un estimador

Dispersión, precisión y acuracidad



Varianza Desvío Standard
Sesgo
Error Medio Cuadrático

Esperanza, Varianza, Sesgo y EMC de un estimador

Esperanza y Sesgo

$$E(\widehat{\theta}) = \sum_{s} \widehat{\theta}(s) \cdot P(s)$$

$$Sesgo(\widehat{\theta}) = E(\widehat{\theta}) - \theta$$

Esperanza, Varianza, Sesgo y EMC de un estimador

Varianza y Desvío Standard

$$V(\widehat{\theta}) = \sum_{s} (\widehat{\theta}(s) - E(\theta))^2 \cdot P(s)$$

$$DS(\widehat{\theta}) = \sqrt{DS(\widehat{\theta})}$$

Error Medio Cuadrático

$$EMC(\widehat{\theta}) = \sum_{s} (\widehat{\theta}(s) - \theta)^2 \cdot P(s) = Var + Sesgo^2$$

Esperanza, Varianza, Sesgo y EMC de un estimador

Varianza y Desvío Standard

$$V(\widehat{\theta}) = \sum_s (\widehat{\theta}(s) - E(\theta))^2 \cdot P(s)$$

$$DS(\widehat{\theta}) = \sqrt{DS(\widehat{\theta})}$$

Error Medio Cuadrático

$$EMC(\widehat{\theta}) = \sum_s (\widehat{\theta}(s) - \theta)^2 \cdot P(s) = Var + Sesgo^2$$

Estimadores insesgado

Estimador insesgado

Un estimador es insesgado si su valor *esperanza* coincide con el parámetro a estimar *para cualquier vector* $Y_i = Y_{i1}, ..., Y_{iN}$:

$$E(\widehat{\theta}) = \theta$$

El *sesgo* de un estimador será:

$$Sesgo(\widehat{\theta}) = \theta - E(\widehat{\theta})$$

Estimadores insesgados

Ejemplos importantes

En el Muestreo Aleatorio Simple

- $\bar{y}$ es un estimador insegado de la media poblacional $\bar{Y}$
- $N \cdot \bar{y}$ es un estimador insegado del total poblacional $\sum_{i=1}^{N} Y_i$
- s^2 es un estimador insegado de S^2
- s_{xy} es un estimador insegado de S_{xy}

Estimadores insesgados

Ejemplos importantes

En el Muestreo Aleatorio Simple

- $\bar{y}$ es un estimador insegado de la media poblacional $\bar{Y}$
- $N \cdot \bar{y}$ es un estimador insegado del total poblacional $\sum_{i=1}^{N} Y_i$
- s^2 es un estimador insegado de S^2
- s_{xy} es un estimador insegado de S_{xy}

Estimadores insesgados

Ejemplos importantes

En el Muestreo Aleatorio Simple

- $\bar{y}$ es un estimador insegado de la media poblacional $\bar{Y}$
- $N \cdot \bar{y}$ es un estimador insegado del total poblacional $\sum_{i=1}^{N} Y_i$
- s^2 es un estimador insegado de S^2
- s_{xy} es un estimador insegado de S_{xy}

Estimadores insesgados

Ejemplos importantes

En el Muestreo Aleatorio Simple

- $\bar{y}$ es un estimador insegado de la media poblacional $\bar{Y}$
- $N \cdot \bar{y}$ es un estimador insegado del total poblacional $\sum_{i=1}^{N} Y_i$
- s^2 es un estimador insegado de S^2
- s_{xy} es un estimador insegado de S_{xy}

Estimadores insesgados

Ejemplos importantes

En el Muestreo Aleatorio Simple

- $\bar{y}$ es un estimador insegado de la media poblacional $\bar{Y}$
- $N \cdot \bar{y}$ es un estimador insegado del total poblacional $\sum_{i=1}^{N} Y_i$
- s^2 es un estimador insegado de S^2
- s_{xy} es un estimador insegado de S_{xy}

Un estimador importante

Estimador de Horvitz-Thompson

Estimador de Horvitz-Thompson

$$\widehat{t}_{\pi y} = \sum_{i \in s} \frac{Y_i}{\pi_i}$$

(estimador insesgado del total de Y)

Análogamente

$$\widehat{\bar{t}}_{\pi y} = \frac{1}{N} \cdot \sum_{i \in s} \frac{Y_i}{\pi_i}$$

(estimador insesgado de la media de Y)

Un estimador importante

Estimador de Horvitz-Thompson

Estimador de Horvitz-Thompson

$$\widehat{t}_{\pi y} = \sum_{i \in s} \frac{Y_i}{\pi_i}$$

(estimador insesgado del total de Y)

Análogamente

$$\widehat{\overline{t}}_{\pi y} = \frac{1}{N} \cdot \sum_{i \in s} \frac{Y_i}{\pi_i}$$

(estimador insesgado de la media de Y)

Muestreo Aleatorio Simple

Varianza del estimador de la media muestral

Resultado importante: En el MAS se tiene:

$$Var(\bar{y}) = (1 - n/N) \cdot \frac{S^2}{n}$$

con

$$S^2 = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N - 1}$$

Estimador insesgado de $Var(\bar{y})$ en el MAS

$$\hat{Var}(\bar{y}) = (1 - n/N) \cdot \frac{s^2}{n}$$

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Ejemplo de cálculo de varianza de un estimador

U$= \{Juan, Pedro, Maria\}$, Y_i: ingreso de i

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

Deseamos estimar la media de los ingresos, $\bar{Y}$ mediante una Muestra Aleatoria Simple de tamaño n=2
Estimando $\bar{Y}$ con $\bar{y}$

Total de muestras posibles=3
P(s)=1/3 (equiprobables)
Hallar la *varianza* de $\bar{y}$

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Ejemplo de cálculo de varianza de un estimador

U$= \{Juan, Pedro, Maria\}$, Y_i: ingreso de i

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

Deseamos estimar la media de los ingresos, $\bar{Y}$ mediante una Muestra Aleatoria Simple de tamaño n=2
Estimando $\bar{Y}$ con $\bar{y}$

Total de muestras posibles=3
P(s)=1/3 (equiprobables)
Hallar la *varianza* de $\bar{y}$

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Ejemplo de cálculo de varianza de un estimador

U$= \{Juan, Pedro, Maria\}$, Y_i: ingreso de i

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

Deseamos estimar la media de los ingresos, $\bar{Y}$ mediante una Muestra Aleatoria Simple de tamaño n=2

Estimando $\bar{Y}$ con $\bar{y}$

Total de muestras posibles=3
P(s)=1/3 (equiprobables)
Hallar la *varianza* de $\bar{y}$

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Ejemplo de cálculo de varianza de un estimador

U$= \{Juan, Pedro, Maria\}$, Y_i: ingreso de i

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

Deseamos estimar la media de los ingresos, $\bar{Y}$ mediante una Muestra Aleatoria Simple de tamaño n=2

Estimando $\bar{Y}$ con $\bar{y}$

Total de muestras posibles=3

P(s)=1/3 (equiprobables)

Hallar la *varianza* de $\bar{y}$

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Ejemplo de cálculo de varianza de un estimador

U= $\{Juan, Pedro, Maria\}$, Y_i: ingreso de i

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

Deseamos estimar la media de los ingresos, $\bar{Y}$ mediante una Muestra Aleatoria Simple de tamaño n=2
Estimando $\bar{Y}$ con $\bar{y}$

Total de muestras posibles=3
P(s)=1/3 (equiprobables)
Hallar la *varianza* de $\bar{y}$

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Ejemplo de cálculo de varianza de un estimador

U$=\{Juan, Pedro, Maria\}$, Y_i: ingreso de i

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

Deseamos estimar la media de los ingresos, $\bar{Y}$ mediante una Muestra Aleatoria Simple de tamaño n=2
Estimando $\bar{Y}$ con $\bar{y}$

Total de muestras posibles=3
P(s)=1/3 (equiprobables)
Hallar la *varianza* de $\bar{y}$

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

$s_1 = \{Juan, Pedro\} \rightarrow \bar{y}(s_1) = 16{,}5$
$s_2 = \{Juan, Maria\} \rightarrow \bar{y}(s_2) = 13{,}5$
$s_3 = \{Pedro, Maria\} \rightarrow \bar{y}(s_2) = 21{,}0$
$E(\bar{y}) = 17$
Es insesgado!

$Var(\bar{y}) = \frac{1}{3} \cdot (16{,}5 - 17)^2 + (13{,}5 - 17)^2 + (21{,}0 - 17)^2 = 9{,}5$

$S^2 = 57 \Rightarrow$

$(1 - 2/3) \cdot 57 \cdot \frac{1}{2} = 9{,}5$
La fórmula de la varianza es correcta!

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

$s_1 = \{Juan, Pedro\} \rightarrow \bar{y}(s_1) = 16{,}5$
$s_2 = \{Juan, Maria\} \rightarrow \bar{y}(s_2) = 13{,}5$
$s_3 = \{Pedro, Maria\} \rightarrow \bar{y}(s_2) = 21{,}0$
$E(\bar{y}) = 17$
Es insesgado!

$Var(\bar{y}) = \frac{1}{3} \cdot (16{,}5 - 17)^2 + (13{,}5 - 17)^2 + (21{,}0 - 17)^2 = 9{,}5$

$S^2 = 57 \Rightarrow$

$(1 - 2/3) \cdot 57 \cdot \frac{1}{2} = 9{,}5$
La fórmula de la varianza es correcta!

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

$s_1 = \{Juan, Pedro\} \rightarrow \bar{y}(s_1) = 16{,}5$
$s_2 = \{Juan, Maria\} \rightarrow \bar{y}(s_2) = 13{,}5$
$s_3 = \{Pedro, Maria\} \rightarrow \bar{y}(s_2) = 21{,}0$
$E(\bar{y}) = 17$
Es insesgado!

$Var(\bar{y}) = \frac{1}{3} \cdot (16{,}5 - 17)^2 + (13{,}5 - 17)^2 + (21{,}0 - 17)^2 = 9{,}5$

$S^2 = 57 \Rightarrow$

$(1 - 2/3) \cdot 57 \cdot \frac{1}{2} = 9{,}5$
La fórmula de la varianza es correcta!

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

$s_1 = \{Juan, Pedro\} \rightarrow \bar{y}(s_1) = 16{,}5$
$s_2 = \{Juan, Maria\} \rightarrow \bar{y}(s_2) = 13{,}5$
$s_3 = \{Pedro, Maria\} \rightarrow \bar{y}(s_2) = 21{,}0$
$E(\bar{y}) = 17$
Es insesgado!

$Var(\bar{y}) = \frac{1}{3} \cdot (16{,}5 - 17)^2 + (13{,}5 - 17)^2 + (21{,}0 - 17)^2 = 9{,}5$

$S^2 = 57 \Rightarrow$

$(1 - 2/3) \cdot 57 \cdot \frac{1}{2} = 9{,}5$
La fórmula de la varianza es correcta!

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

$s_1 = \{Juan, Pedro\} \rightarrow \bar{y}(s_1) = 16{,}5$
$s_2 = \{Juan, Maria\} \rightarrow \bar{y}(s_2) = 13{,}5$
$s_3 = \{Pedro, Maria\} \rightarrow \bar{y}(s_2) = 21{,}0$
$E(\bar{y}) = 17$
Es insesgado!

$Var(\bar{y}) = \frac{1}{3} \cdot (16{,}5 - 17)^2 + (13{,}5 - 17)^2 + (21{,}0 - 17)^2 = 9{,}5$

$S^2 = 57 \Rightarrow$

$(1 - 2/3) \cdot 57 \cdot \frac{1}{2} = 9{,}5$
La fórmula de la varianza es correcta!

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

$s_1 = \{Juan, Pedro\} \rightarrow \bar{y}(s_1) = 16{,}5$
$s_2 = \{Juan, Maria\} \rightarrow \bar{y}(s_2) = 13{,}5$
$s_3 = \{Pedro, Maria\} \rightarrow \bar{y}(s_2) = 21{,}0$
$E(\bar{y}) = 17$
Es insesgado!

$Var(\bar{y}) = \frac{1}{3} \cdot (16{,}5 - 17)^2 + (13{,}5 - 17)^2 + (21{,}0 - 17)^2 = 9{,}5$

$S^2 = 57 \Rightarrow$

$(1 - 2/3) \cdot 57 \cdot \frac{1}{2} = 9{,}5$
La fórmula de la varianza es correcta!

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

$s_1 = \{Juan, Pedro\} \rightarrow \bar{y}(s_1) = 16{,}5$
$s_2 = \{Juan, Maria\} \rightarrow \bar{y}(s_2) = 13{,}5$
$s_3 = \{Pedro, Maria\} \rightarrow \bar{y}(s_2) = 21{,}0$
$E(\bar{y}) = 17$
Es insesgado!

$Var(\bar{y}) = \frac{1}{3} \cdot (16{,}5 - 17)^2 + (13{,}5 - 17)^2 + (21{,}0 - 17)^2 = 9{,}5$

$S^2 = 57 \Rightarrow$

$(1 - 2/3) \cdot 57 \cdot \frac{1}{2} = 9{,}5$
La fórmula de la varianza es correcta!

Muestreo Aleatorio Simple

Ejemplo de Varianza de la media muestral

Nombre	Y=Ingreso
Juan	9
Pedro	24
Maria	18

$\bar{Y} = (9 + 24 + 18)/3 = 17{,}0$

$s_1 = \{Juan, Pedro\} \rightarrow \bar{y}(s_1) = 16{,}5$
$s_2 = \{Juan, Maria\} \rightarrow \bar{y}(s_2) = 13{,}5$
$s_3 = \{Pedro, Maria\} \rightarrow \bar{y}(s_2) = 21{,}0$
$E(\bar{y}) = 17$
Es insesgado!

$Var(\bar{y}) = \frac{1}{3} \cdot (16{,}5 - 17)^2 + (13{,}5 - 17)^2 + (21{,}0 - 17)^2 = 9{,}5$

$S^2 = 57 \Rightarrow$

$(1 - 2/3) \cdot 57 \cdot \frac{1}{2} = 9{,}5$
La fórmula de la varianza es correcta!

Tabla de Contenidos

Muestreo Aleatorio Simple

En las encuestas una de las estimaciones más usuales son las de proporciones:

Qué proporción de unidades están dentro de cierta categoría C o qué cantidad de unidades están en C

Al estimar proporciones o totales aparecen variables que toman los valores 0 y 1: variables dicotómicas

$$Y_i = \begin{cases} 1, & \text{si } i \in C \\ 0, & i \notin C \end{cases}$$

Muestreo Aleatorio Simple

Caso importante: variables dicotómicas

Es usual escribir $P = \bar{Y}$, $Q = 1 - P$, $p = \bar{y}$

Y se cumple

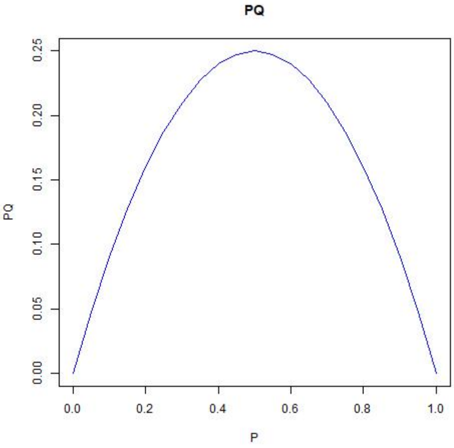
$$S^2 = P \cdot Q \cdot \frac{N}{N-1} \approx P \cdot Q$$

$$\hat{V}(p) = (1 - n/N) \cdot p \cdot q \cdot \frac{1}{n-1}$$

$$CV(p) \approx \sqrt{\frac{Q}{n \cdot P}}$$

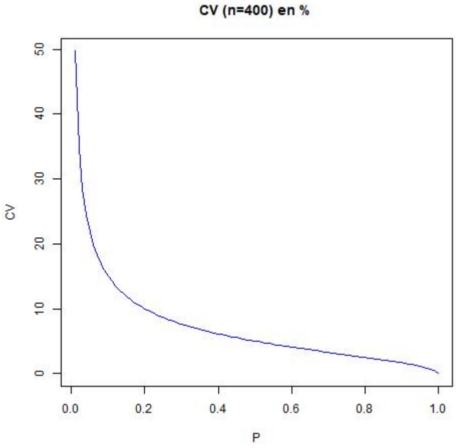
(omitiendo el factor de corrección por población finita)

Variables dicotómicas
S^2 en función de P



Variables dicotómicas

CV en función de P



Muestreo Aleatorio Simple

Caso importante: variables dicotómicas

De lo anterior se deduce que:

El caso más desfavorable repecto al DS es cuando P=0.5

Si $0 \leq a \leq P \leq b \leq$, el caso más desfavorable para el DS es el valor más cercano a 0.5 entre a y b.

Si $P \to 0, CV(p) \to \infty$.

En las encuestas una proporción con un CV alto en general es equivalente a una proporción pequeña.

Muestreo Aleatorio Simple

Caso importante: variables dicotómicas

De lo anterior se deduce que:

El caso más desfavorable repecto al DS es cuando P=0.5

Si $0 \leq a \leq P \leq b \leq$, el caso más desfavorable para el DS es el valor más cercano a 0.5 entre a y b.

Si $P \rightarrow 0, CV(p) \rightarrow \infty$.

En las encuestas una proporción con un CV alto en general es equivalente a una proporción pequeña.

Muestreo Aleatorio Simple

Caso importante: variables dicotómicas

De lo anterior se deduce que:

El caso más desfavorable repecto al DS es cuando P=0.5

Si $0 \leq a \leq P \leq b \leq$, el caso más desfavorable para el DS es el valor más cercano a 0.5 entre a y b.

Si $P \to 0, CV(p) \to \infty$.

En las encuestas una proporción con un CV alto en general es equivalente a una proporción pequeña.

Muestreo Aleatorio Simple

Caso importante: variables dicotómicas

De lo anterior se deduce que:

El caso más desfavorable repecto al DS es cuando P=0.5

Si $0 \leq a \leq P \leq b \leq$, el caso más desfavorable para el DS es el valor más cercano a 0.5 entre a y b.

Si $P \to 0, CV(p) \to \infty$.

En las encuestas una proporción con un CV alto en general es equivalente a una proporción pequeña.

Para qué estimar la varianza de un estimador?

- Coeficiente de Variación: $100 \cdot DS(\hat{\theta})/\theta$ $(\theta \neq 0)$
- Intervalos de confianza
- Efecto Diseño: $V(\hat{\theta})$ / $V_{MAS}(\hat{\theta})$

Para qué estimar la varianza de un estimador?

- Coeficiente de Variación: $100 \cdot DS(\hat{\theta})/\theta$ $(\theta \neq 0)$
- Intervalos de confianza
- Efecto Diseño: $V(\hat{\theta})$ / $V_{MAS}(\hat{\theta})$

Para qué estimar la varianza de un estimador?

- Coeficiente de Variación: $100 \cdot DS(\hat{\theta})/\theta$ $(\theta \neq 0)$
- Intervalos de confianza
- Efecto Diseño: $V(\hat{\theta})$ / $V_{MAS}(\hat{\theta})$

Para qué estimar la varianza de un estimador?

- Coeficiente de Variación: $100 \cdot DS(\hat{\theta})/\theta$ $(\theta \neq 0)$
- Intervalos de confianza
- Efecto Diseño: $V(\hat{\theta})$ / $V_{MAS}(\hat{\theta})$

Coeficiente de Variación de un estimador

$$CV = \ 100 \cdot DS(\hat{\theta})/\theta \ \ (\theta \neq 0)$$

El CV nos dice, aproximadamente, qué proporción del valor a estimar es el error medio del estimador

Ejemplo:

Si θ=120 y $DS(\theta) = 12 \Rightarrow$

$$CV = 100 \cdot 12/120 = 10\,\%$$

Coeficiente de Variación de un estimador

$$CV = \ 100 \cdot DS(\hat{\theta})/\theta \ \ (\theta \neq 0)$$

El CV nos dice, aproximadamente, qué proporción del valor a estimar es el error medio del estimador

Ejemplo:
Si θ=120 y $DS(\theta) = 12 \Rightarrow$

$$CV = 100 \cdot 12/120 = 10\,\%$$

Coeficiente de Variación de un estimador

$$CV = \ 100 \cdot DS(\hat{\theta})/\theta \ \ (\theta \neq 0)$$

El CV nos dice, aproximadamente, qué proporción del valor a estimar es el error medio del estimador

Ejemplo:
Si θ=120 y $DS(\theta) = 12 \Rightarrow$

$$CV = 100 \cdot 12/120 = 10\,\%$$

Coeficiente de Variación de un estimador

Un CV alto no indica que el valor estimado no sea útil
O que no pueda calcularse un intervalo de confianza

Coeficiente de Variación de un estimador

Ejercicio
Un candidato A desea saber si ganará o no un ballotage. Supongamos N (cantidad de electores) igual a 20,000,000. Se selecciona una MAS(800), observándose en la muestra que 415 electores votarán por A.
Se pide:

- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 95 % para la proporción de votantes a A
- El resultado es útil para el condidato?

Coeficiente de Variación de un estimador

Ejercicio

Un candidato A desea saber si ganará o no un ballotage. Supongamos N (cantidad de electores) igual a 20,000,000. Se selecciona una MAS(800), observándose en la muestra que 415 electores votarán por A.

Se pide:

- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 95 % para la proporción de votantes a A
- El resultado es útil para el condidato?

Coeficiente de Variación de un estimador

Ejercicio

Un candidato A desea saber si ganará o no un ballotage. Supongamos N (cantidad de electores) igual a 20,000,000. Se selecciona una MAS(800), observándose en la muestra que 415 electores votarán por A.

Se pide:

- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 95 % para la proporción de votantes a A
- El resultado es útil para el condidato?

Coeficiente de Variación de un estimador

Ejercicio
Un candidato A desea saber si ganará o no un ballotage. Supongamos N (cantidad de electores) igual a 20,000,000. Se selecciona una MAS(800), observándose en la muestra que 415 electores votarán por A.
Se pide:

- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 95 % para la proporción de votantes a A
- El resultado es útil para el condidato?

Coeficiente de Variación de un estimador

Ejercicio

Un programa de salud se pone en marcha si cierta afección supera el 2 % de la población (N=10,000,000 personas). La proporción poblacional P (*desconocida*) de afectados es P=0.012.

En el universo definimos una variable Y_i=1 si **i** afectado, 0 si no.

Se planifica estimar P mediante una muestra aleatoria simple de tamaño n=1000. Se selecciona la muestra y se hallan 10 afectados.

Se pide:

- Hallar CV(Y) y CV(p)
- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 99 % para la proporción de afectados y para el porcentaje de afectados
- Fué útil la información proporcionada por la muestra?

Coeficiente de Variación de un estimador

Ejercicio
Un programa de salud se pone en marcha si cierta afección supera el 2 % de la población (N=10,000,000 personas). La proporción poblacional P (*desconocida*) de afectados es P=0.012.
En el universo definimos una variable Y_i=1 si **i** afectado, 0 si no.
Se planifica estimar P mediante una muestra aleatoria simple de tamaño n=1000. Se selecciona la muestra y se hallan 10 afectados.
Se pide:

- Hallar CV(Y) y CV(p)
- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 99 % para la proporción de afectados y para el porcentaje de afectados
- Fué útil la información proporcionada por la muestra?

Coeficiente de Variación de un estimador

Ejercicio

Un programa de salud se pone en marcha si cierta afección supera el 2 % de la población (N=10,000,000 personas). La proporción poblacional P (*desconocida*) de afectados es P=0.012.

En el universo definimos una variable Y_i=1 si **i** afectado, 0 si no.

Se planifica estimar P mediante una muestra aleatoria simple de tamaño n=1000. Se selecciona la muestra y se hallan 10 afectados.

Se pide:

- Hallar CV(Y) y CV(p)
- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 99 % para la proporción de afectados y para el porcentaje de afectados
- Fué útil la información proporcionada por la muestra?

Coeficiente de Variación de un estimador

Ejercicio
Un programa de salud se pone en marcha si cierta afección supera el 2 % de la población (N=10,000,000 personas). La proporción poblacional P (*desconocida*) de afectados es P=0.012.
En el universo definimos una variable Y_i=1 si **i** afectado, 0 si no.
Se planifica estimar P mediante una muestra aleatoria simple de tamaño n=1000. Se selecciona la muestra y se hallan 10 afectados.
Se pide:

- Hallar CV(Y) y CV(p)
- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 99 % para la proporción de afectados y para el porcentaje de afectados
- Fué útil la información proporcionada por la muestra?

Coeficiente de Variación de un estimador

Ejercicio

Un programa de salud se pone en marcha si cierta afección supera el 2 % de la población (N=10,000,000 personas). La proporción poblacional P (*desconocida*) de afectados es P=0.012.

En el universo definimos una variable Y_i=1 si **i** afectado, 0 si no.

Se planifica estimar P mediante una muestra aleatoria simple de tamaño n=1000. Se selecciona la muestra y se hallan 10 afectados.

Se pide:

- Hallar CV(Y) y CV(p)
- Estimar CV(p) a partir del resultado de la muestra
- Se cumple la regla npq?
- Dar un intervalo de confianza del 99 % para la proporción de afectados y para el porcentaje de afectados
- Fué útil la información proporcionada por la muestra?

Coeficiente de Variación de un estimador

Estimación de proporciones en el MAS

En el MAS, si la variable objetivo es dicotómica, $Y_i = 1/0$, sabemos que

$$Var(p) \approx (1 - n/N) \cdot \frac{P \cdot Q}{n}$$

De aquí se deduce que

$$CV(p) \approx \sqrt{(1 - n/N)} \cdot \sqrt{\frac{Q}{P \cdot n}}$$

$$\Downarrow$$

Para un tamaño de muestra n fijo, si P pequeño CV(p) grande

Efecto diseño

Efecto Diseño:

$$\frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta})}$$

En qué proporción aumenta (o disminuye) la varianza por no haber utilizado un MAS

Cuanta menos (o más) muestra hubiéramos necesitado para alcanzar la misma precisión con un MAS

Ej. Si deff=3, con la tercera
una muestra tres veces menor hubiéramos obtenido igual precisión

Efecto diseño

Efecto Diseño:

$$\frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta})}$$

En qué proporción aumenta (o disminuye) la varianza por no haber utilizado un MAS

Cuanta menos (o más) muestra hubiéramos necesitado para alcanzar la misma precisión con un MAS

Ej. Si deff=3, con la tercera
una muestra tres veces menor hubiéramos obtenido igual precisión

Efecto diseño

Efecto Diseño:

$$\frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta})}$$

En qué proporción aumenta (o disminuye) la varianza por no haber utilizado un MAS

Cuanta menos (o más) muestra hubiéramos necesitado para alcanzar la misma precisión con un MAS

Ej. Si deff=3, con la tercera
una muestra tres veces menor hubiéramos obtenido igual precisión

Estimación de una razón en el MAS

A veces se desea estimar la razón de dos variables X,Y: $R = Y/X$

Ej Y=Ingreso, X=Tamaño del Hogar

Un estimador usual en el MAS es $r = \bar{y}/\bar{x}$. No es en el MAS un estimador insesgado de R, pero el sesgo es en general despreciable
Un resultado de muestreo dice que la varianza *aproximada* de $r = \bar{y}/\bar{x}$ es

$$Var(\bar{y}/\bar{x}) \approx \frac{1}{n \cdot \bar{X}^2} \cdot (1 - n/N) \cdot \sum_{i=1}^{N} (Y_i - RX_i)^2/(N-1)$$

que se estima mediante

$$\hat{V}(\bar{y}/\bar{x}) = \frac{1}{n \cdot \bar{x}^2} \cdot (1 - n/N) \cdot \sum_{i=1}^{n} (y_i - rx_i)^2/(n-1)$$

Estimación de una razón en el MAS

A veces se desea estimar la razón de dos variables X,Y: $R = Y/X$

Ej Y=Ingreso, X=Tamaño del Hogar

Un estimador usual en el MAS es $r = \bar{y}/\bar{x}$. No es en el MAS un estimador insesgado de R, pero el sesgo es en general despreciable

Un resultado de muestreo dice que la varianza *aproximada* de $r = \bar{y}/\bar{x}$ es

$$Var(\bar{y}/\bar{x}) \approx \frac{1}{n \cdot \bar{X}^2} \cdot (1 - n/N) \cdot \sum_{i=1}^{N} (Y_i - RX_i)^2/(N-1)$$

que se estima mediante

$$\hat{V}(\bar{y}/\bar{x}) = \frac{1}{n \cdot \bar{x}^2} \cdot (1 - n/N) \cdot \sum_{i=1}^{n} (y_i - rx_i)^2/(n-1)$$

Estimación de una razón en el MAS

A veces se desea estimar la razón de dos variables X,Y: $R = Y/X$

Ej Y=Ingreso, X=Tamaño del Hogar

Un estimador usual en el MAS es $r = \bar{y}/\bar{x}$. No es en el MAS un estimador insesgado de R, pero el sesgo es en general despreciable
Un resultado de muestreo dice que la varianza *aproximada* de $r = \bar{y}/\bar{x}$ es

$$Var(\bar{y}/\bar{x}) \approx \frac{1}{n \cdot \bar{X}^2} \cdot (1 - n/N) \cdot \sum_{i=1}^{N} (Y_i - RX_i)^2/(N-1)$$

que se estima mediante

$$\hat{V}(\bar{y}/\bar{x}) = \frac{1}{n \cdot \bar{x}^2} \cdot (1 - n/N) \cdot \sum_{i=1}^{n} (y_i - rx_i)^2/(n-1)$$

Estimación de una razón en el MAS

Una forma equivalente de esa aproximación de $\hat{V}(\bar{y}/\bar{x})$ es:

$$Var(\bar{y}/\bar{x}) \approx \frac{(1-f)}{n \cdot \bar{X}^2} \cdot (S_y^2 + R^2 \cdot S_x^2 - 2 \cdot R \cdot \rho \cdot S_{xy})$$

Estimación de una razón en el MAS

Ejemplo

En el siguiente universo de N=4 unidades se desea estimar la razón R=Y/X mediante una Muestra Aleatoria Simple de tamaño n=2, estimando R con $r = \bar{y}/\bar{x}$

i	Y	X
1	8	4
2	12	7
3	6	2
4	4	5

Hallar: R, E(r), Sesgo(r), SesgRel(r), Var(r), EMC(r), CV(r)

Rta: Sesgo(r)=0.0411 , Var(r)=0.1412, CV(r)=22.004 %

Muestreo Aleatorio Simple

Tamaño de muestra en la estimación de medias

La fórmula básica es $Var(\bar{y}) = (1 - n/N) \cdot S^2/n$

Si queremos que la semiamplitud del IC, con un nivel de confianza α, sea d y suponemos n/N despreciable, entonces

$$n = \frac{t_\alpha^2 \cdot S^2}{d^2}$$

Muestreo Aleatorio Simple

Tamaño de muestra en la estimación de proporciones

La fórmula básica es $Var(\bar{y}) \approx (1 - n/N) \cdot PQ/n$

Si queremos que el CV del estimador sea cv y suponemos n/N despreciable, entonces

$$n = \frac{Q}{cv^2}$$

Tabla de Contenidos

Muestreo Sistemático con intervalo entero

Seleccionar uno de cada I elementos

Comenzando en una unidad *aa* seleccionada aleatoriamente ($1 \leq \text{aa} \leq \text{I}$)
en forma equiprobable

Sist(I) I entero, $0 < \text{I} \leq \text{N}$

Seleccionar uno de cada I elementos

I: Intervalo de Selección

Muestreo Sistemático con intervalo entero

Seleccionar uno de cada I elementos
Comenzando en una unidad *aa* seleccionada aleatoriamente ($1 \leq$ aa $\leq$ I)
en forma equiprobable

Sist(I) I entero, $0 <$ I $\leq$ N

Seleccionar uno de cada I elementos

I: Intervalo de Selección

Muestreo Sistemático con intervalo entero

Seleccionar uno de cada I elementos
Comenzando en una unidad *aa* seleccionada aleatoriamente $(1 \leq \text{aa} \leq \text{I})$
en forma equiprobable
Sist(I) I entero, $0 < \text{I} \leq \text{N}$

Seleccionar uno de cada I elementos

I: Intervalo de Selección

Muestreo Sistemático con intervalo entero

Ejemplo

N=10, Sist(3)
Tres muestras posibles:

$$s_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

$$s_2 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

$$s_3 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Las tres muestras equiprobables

¿Cómo seleccionar una muestra sistemática Sist(I)?

- Generar un número entero AA 'aleatorio' $0 < AA \leq$ I
- Seleccionar los elementos AA, AA+I, AA+2I...

Tamaño de muestra en un diseño Sist(I)

- #s=[N/I] o [N/I]+1
- Solo tiene tamaño fijo si I | N

¿Cómo seleccionar una muestra sistemática Sist(I)?

- Generar un número entero AA 'aleatorio' $0 < AA \leq$ I
- Seleccionar los elementos AA, AA+I, AA+2I...

Tamaño de muestra en un diseño Sist(I)

- #s=[N/I] o [N/I]+1
- Solo tiene tamaño fijo si I | N

¿Cómo seleccionar una muestra sistemática Sist(I)?

- Generar un número entero AA 'aleatorio' $0 < AA \leq$ I
- Seleccionar los elementos AA, AA+I, AA+2I...

Tamaño de muestra en un diseño Sist(I)

- #s=[N/I] o [N/I]+1
- Solo tiene tamaño fijo si I | N

¿Cómo seleccionar una muestra sistemática Sist(I)?

- Generar un número entero AA 'aleatorio' $0 < AA \leq$ I
- Seleccionar los elementos AA, AA+I, AA+2I...

Tamaño de muestra en un diseño Sist(I)

- #s=[N/I] o [N/I]+1
- Solo tiene tamaño fijo si I | N

¿Cómo seleccionar una muestra sistemática Sist(I)?

- Generar un número entero AA 'aleatorio' $0 < AA \leq$ I
- Seleccionar los elementos AA, AA+I, AA+2I...

Tamaño de muestra en un diseño Sist(I)

- #s=[N/I] o [N/I]+1
- Solo tiene tamaño fijo si I | N

¿Cómo seleccionar una muestra sistemática Sist(I)?

- Generar un número entero AA 'aleatorio' $0 < AA \leq$ I
- Seleccionar los elementos AA, AA+I, AA+2I...

Tamaño de muestra en un diseño Sist(I)

- #s=[N/I] o [N/I]+1
- Solo tiene tamaño fijo si I | N

Muestreo sistemático

- Selección de viviendas dentro de un radio censal
- Selección de viviendas en una manzana
- Selección de personas a partir de un registro
- Control de calidad

Muestreo sistemático

- Selección de viviendas dentro de un radio censal
- Selección de viviendas en una manzana
- Selección de personas a partir de un registro
- Control de calidad

Muestreo sistemático

- Selección de viviendas dentro de un radio censal
- Selección de viviendas en una manzana
- Selección de personas a partir de un registro
- Control de calidad

Muestreo sistemático

- Selección de viviendas dentro de un radio censal
- Selección de viviendas en una manzana
- Selección de personas a partir de un registro
- Control de calidad

Probabilidades de Selección en un diseño sistemático

Sist(I)

$$\pi_i = 1/I$$

$$F_i = I$$

Muestreo Sistemático de Intervalo Entero I=3

Ejemplo de Esperanza, Varianza, Sesgo y EMC de la media muestral

Supongamos la siguiente población de 7 unidades.

Vivienda	Y=Ingreso
A	10
B	15
C	20
D	25
E	42
F	50
G	55

Deseamos estimar la media de Y mediante un muestreo sistemático, con intervalo de selección I=3.
Compararemos dos estimadores:

1. $\bar{Y}$ con $\bar{y}$
2. $I \cdot \sum y_i / N$

- Hallar $\bar{Y}$
- Listar las tres muestras posibles y su correspondientes valores estimados
- Calcular Varianza, Sesgo, EMC, CV y deff de los dos estimadores. Cuál es preferible?

Muestreo Sistemático de Intervalo Entero I=3

Ejemplo de Esperanza, Varianza, Sesgo y EMC de la media muestral

Supongamos la siguiente población de 7 unidades.

Vivienda	Y=Ingreso
A	10
B	15
C	20
D	25
E	42
F	50
G	55

Deseamos estimar la media de Y mediante un muestreo sistemático, con intervalo de selección I=3.
Compararemos dos estimadores:

1. $\bar{Y}$ con $\bar{y}$
2. $I \cdot \sum y_i / N$

- Hallar $\bar{Y}$
- Listar las tres muestras posibles y su correspondientes valores estimados
- Calcular Varianza, Sesgo, EMC, CV y deff de los dos estimadores. Cuál es preferible?

Muestreo Sistemático de Intervalo Entero I=3

Ejemplo de Esperanza, Varianza, Sesgo y EMC de la media muestral

Supongamos la siguiente población de 7 unidades.

Vivienda	Y=Ingreso
A	10
B	15
C	20
D	25
E	42
F	50
G	55

Deseamos estimar la media de Y mediante un muestreo sistemático, con intervalo de selección I=3.
Compararemos dos estimadores:

1. $\bar{Y}$ con $\bar{y}$
2. $I \cdot \sum y_i / N$

- Hallar $\bar{Y}$
- Listar las tres muestras posibles y su correspondientes valores estimados
- Calcular Varianza, Sesgo, EMC, CV y deff de los dos estimadores. Cuál es preferible?

Muestreo Sistemático de Intervalo Entero I=3

Ejemplo de Esperanza, Varianza, Sesgo y EMC de la media muestral

Supongamos la siguiente población de 7 unidades.

Vivienda	Y=Ingreso
A	10
B	15
C	20
D	25
E	42
F	50
G	55

Deseamos estimar la media de Y mediante un muestreo sistemático, con intervalo de selección I=3.
Compararemos dos estimadores:

1. $\bar{Y}$ con $\bar{y}$
2. $I \cdot \sum y_i / N$

- Hallar $\bar{Y}$
- Listar las tres muestras posibles y su correspondientes valores estimados
- Calcular Varianza, Sesgo, EMC, CV y deff de los dos estimadores. Cuál es preferible?

Muestreo Sistemático de Intervalo Entero I=3

Ejemplo de Esperanza, Varianza, Sesgo y EMC de la media muestral

Supongamos la siguiente población de 7 unidades.

Vivienda	Y=Ingreso
A	10
B	15
C	20
D	25
E	42
F	50
G	55

Deseamos estimar la media de Y mediante un muestreo sistemático, con intervalo de selección I=3.
Compararemos dos estimadores:

1. $\bar{Y}$ con $\bar{y}$
2. $I \cdot \sum y_i / N$

- Hallar $\bar{Y}$
- Listar las tres muestras posibles y su correspondientes valores estimados
- Calcular Varianza, Sesgo, EMC, CV y deff de los dos estimadores. Cuál es preferible?

Muestreo Sistemático de Intervalo Entero I=3

Ejemplo de Esperanza, Varianza, Sesgo y EMC de la media muestral

Supongamos la siguiente población de 7 unidades.

Vivienda	Y=Ingreso
A	10
B	15
C	20
D	25
E	42
F	50
G	55

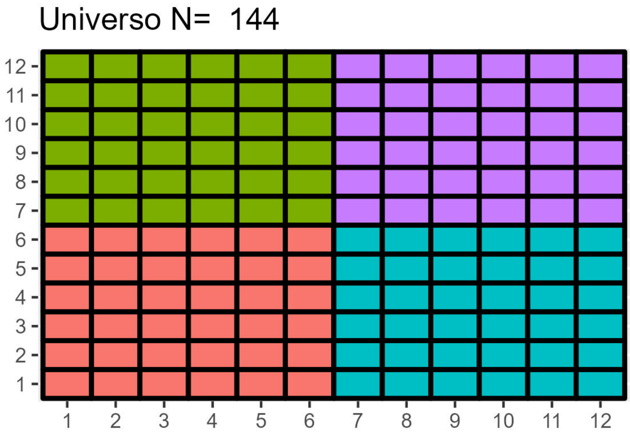
Deseamos estimar la media de Y mediante un muestreo sistemático, con intervalo de selección I=3.
Compararemos dos estimadores:

1. $\bar{Y}$ con $\bar{y}$
2. $I \cdot \sum y_i / N$

- Hallar $\bar{Y}$
- Listar las tres muestras posibles y su correspondientes valores estimados
- Calcular Varianza, Sesgo, EMC, CV y deff de los dos estimadores. Cuál es preferible?

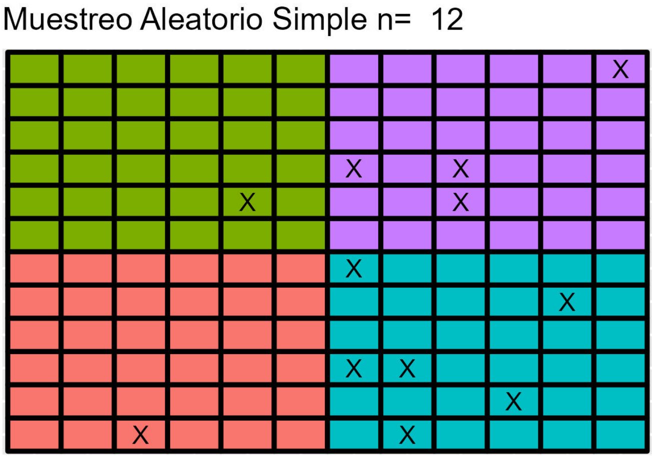
Muestreo sistemático

Aplicación en muestreo espacial



Muestreo sistemático

Aplicación en muestreo espacial



Muestreo sistemático

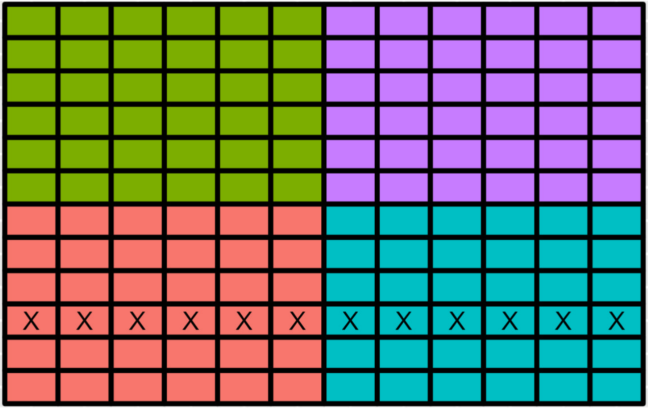
Aplicación en muestreo espacial

12	24	36	48	60	72	84	96	108	120	132	144
11	23	35	47	59	71	83	95	107	119	131	143
10	22	34	46	58	70	82	94	106	118	130	142
9	21	33	45	57	69	81	93	105	117	129	141
8	20	32	44	56	68	80	92	104	116	128	140
7	19	31	43	55	67	79	91	103	115	127	139
6	18	30	42	54	66	78	90	102	114	126	138
5	17	29	41	53	65	77	89	101	113	125	137
4	16	28	40	52	64	76	88	100	112	124	136
3	15	27	39	51	63	75	87	99	111	123	135
2	14	26	38	50	62	74	86	98	110	122	134
1	13	25	37	49	61	73	85	97	109	121	133

Muestreo sistemático

Aplicación en muestreo espacial

Muestreo Estratificado l= 12 aa= 3



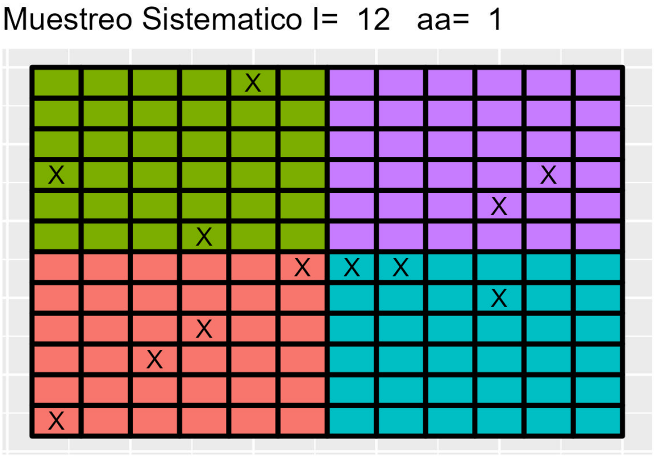
Muestreo sistemático

Aplicación en muestreo espacial

67	79	90	100	109	117	124	130	135	139	142	144
56	68	80	91	101	110	118	125	131	136	140	143
46	57	69	81	92	102	111	119	126	132	137	141
37	47	58	70	82	93	103	112	120	127	133	138
29	38	48	59	71	83	94	104	113	121	128	134
22	30	39	49	60	72	84	95	105	114	122	129
16	23	31	40	50	61	73	85	96	106	115	123
11	17	24	32	41	51	62	74	86	97	107	116
7	12	18	25	33	42	52	63	75	87	98	108
4	8	13	19	26	34	43	53	64	76	88	99
2	5	9	14	20	27	35	44	54	65	77	89
1	3	6	10	15	21	28	36	45	55	66	78

Muestreo sistemático

Aplicación en muestreo espacial



Selección con probabilidades variables

Selección de localidades, Radios Censales, Empresas...

En general es más eficiente seleccionarlas con probabilidad proporcional a alguna medida de tamaño que en forma equiprobable

Localidades

Población o viviendas particulares

Empresas

Facturación, ventas, empleados..

Colegios

Matrícula

Selección con probabilidades variables

Selección de localidades, Radios Censales, Empresas...

En general es más eficiente seleccionarlas con probabilidad proporcional a alguna medida de tamaño que en forma equiprobable

Localidades

Población o viviendas particulares

Empresas

Facturación, ventas, empleados..

Colegios

Matrícula

Selección con probabilidades variables

Selección de localidades, Radios Censales, Empresas...

En general es más eficiente seleccionarlas con probabilidad proporcional a alguna medida de tamaño que en forma equiprobable

Localidades

Población o viviendas particulares

Empresas

Facturación, ventas, empleados..

Colegios

Matrícula

Selección con probabilidades variables

Selección de localidades, Radios Censales, Empresas...

En general es más eficiente seleccionarlas con probabilidad proporcional a alguna medida de tamaño que en forma equiprobable

Localidades

Población o viviendas particulares

Empresas

Facturación, ventas, empleados..

Colegios

Matrícula

Selección con probabilidades proporcionales a una medida de tamaño

En algunos diseño la probabilidad de selección es proporcional a una variable auxiliar $X_i > 0$:

$$\pi_i \approx X_i$$

Selección con probabilidades proporcional a una medida de tamaño

Resultado importante

Si n(s) constante y $X_i \approx Y_i$ entonces

$$Var(\hat{t}_{\pi y}) = 0$$

si las unidades son seleccionadas con probabilidad proporcional a X_i $(X_i > 0)$

Un diseño muy aplicado: Sistemático de Madow

Madow(n, $X_i, X_i > 0$) X_i variable *auxiliar*, $X_i > 0$

- $\lambda_i = n \cdot \frac{X_i}{\sum X_k}$ (suponemos $\lambda_i < 1$)
- Calculamos $V_0 = 0, V_{i+1} = V_i + \lambda_{i+1}, i > 1$
- Generamos $e, v.a.U(0, 1)$
- Si $[V_{i-1} - e, V_i - e]$ contiene un entero, $i > 0 \Rightarrow i \in s$

Diseño con tamaño fijo n
Probabilidad de selección proporcional a X_i
Selección secuencial

Un diseño muy aplicado: Sistemático de Madow

Madow(n, $X_i, X_i > 0$) X_i variable *auxiliar*, $X_i > 0$

- $\lambda_i = n \cdot \frac{X_i}{\sum X_k}$ (suponemos $\lambda_i < 1$)
- Calculamos $V_0 = 0, V_{i+1} = V_i + \lambda_{i+1}, i > 1$
- Generamos $e, v.a.U(0,1)$
- Si $[V_{i-1} - e, V_i - e]$ contiene un entero, $i > 0 \Rightarrow i \in s$

Diseño con tamaño fijo n
Probabilidad de selección proporcional a X_i
Selección secuencial

Un diseño muy aplicado: Sistemático de Madow

Madow(n, $X_i, X_i > 0$) X_i variable *auxiliar*, $X_i > 0$

- $\lambda_i = n \cdot \frac{X_i}{\sum X_k}$ (suponemos $\lambda_i < 1$)
- Calculamos $V_0 = 0, V_{i+1} = V_i + \lambda_{i+1}, i > 1$
- Generamos $e, v.a.U(0,1)$
- Si $[V_{i-1} - e, V_i - e]$ contiene un entero, $i > 0 \Rightarrow i \in s$

Diseño con tamaño fijo n
Probabilidad de selección proporcional a X_i
Selección secuencial

Un diseño muy aplicado: Sistemático de Madow

Madow(n, $X_i, X_i > 0)$ X_i variable *auxiliar*, $X_i > 0$

- $\lambda_i = n \cdot \frac{X_i}{\sum X_k}$ (suponemos $\lambda_i < 1$)
- Calculamos $V_0 = 0, V_{i+1} = V_i + \lambda_{i+1}, i > 1$
- Generamos $e, v.a. U(0, 1)$
- Si $[V_{i-1} - e, V_i - e]$ contiene un entero, $i > 0 \Rightarrow i \in s$

Diseño con tamaño fijo n
Probabilidad de selección proporcional a X_i
Selección secuencial

Un diseño muy aplicado: Sistemático de Madow

Madow(n, $X_i, X_i > 0$) X_i variable *auxiliar*, $X_i > 0$

- $\lambda_i = n \cdot \frac{X_i}{\sum X_k}$ (suponemos $\lambda_i < 1$)
- Calculamos $V_0 = 0, V_{i+1} = V_i + \lambda_{i+1}, i > 1$
- Generamos $e, v.a.U(0,1)$
- Si $[V_{i-1} - e, V_i - e]$ contiene un entero, $i > 0 \Rightarrow i \in s$

Diseño con tamaño fijo n
Probabilidad de selección proporcional a X_i
Selección secuencial

Un diseño muy aplicado: Sistemático de Madow

Madow(n, $X_i, X_i > 0$) X_i variable *auxiliar*, $X_i > 0$

- $\lambda_i = n \cdot \frac{X_i}{\sum X_k}$ (suponemos $\lambda_i < 1$)
- Calculamos $V_0 = 0, V_{i+1} = V_i + \lambda_{i+1}, i > 1$
- Generamos $e, v.a.U(0,1)$
- Si $[V_{i-1} - e, V_i - e]$ contiene un entero, $i > 0 \Rightarrow i \in s$

Diseño con tamaño fijo n

Probabilidad de selección proporcional a X_i

Selección secuencial

Un diseño muy aplicado: Sistemático de Madow

Madow(n, $X_i, X_i > 0$) X_i variable *auxiliar*, $X_i > 0$

- $\lambda_i = n \cdot \frac{X_i}{\sum X_k}$ (suponemos $\lambda_i < 1$)
- Calculamos $V_0 = 0, V_{i+1} = V_i + \lambda_{i+1}, i > 1$
- Generamos $e, v.a. U(0, 1)$
- Si $[V_{i-1} - e, V_i - e]$ contiene un entero, $i > 0 \Rightarrow i \in s$

Diseño con tamaño fijo n
Probabilidad de selección proporcional a X_i
Selección secuencial

Un diseño muy aplicado: Sistemático de Madow

Madow(n, $X_i, X_i > 0$) X_i variable *auxiliar*, $X_i > 0$

- $\lambda_i = n \cdot \frac{X_i}{\sum X_k}$ (suponemos $\lambda_i < 1$)
- Calculamos $V_0 = 0, V_{i+1} = V_i + \lambda_{i+1}, i > 1$
- Generamos $e, v.a. U(0,1)$
- Si $[V_{i-1} - e, V_i - e]$ contiene un entero, $i > 0 \Rightarrow i \in s$

Diseño con tamaño fijo n
Probabilidad de selección proporcional a X_i
Selección secuencial

Sistemático de Madow

- Hay en general probabilidades de segundo orden nulas $\pi_{ij} = 0$ (inconveniente teórico)
- $\Rightarrow$ no hay un estimador insesgado de la varianza de $\hat{t}_{\pi_y}$
- $\Rightarrow$ No es sencillo calcular la varianza de $\hat{t}_{\pi_y}$

Ventaja

- Al ser una selección *sistemática*, podemos darle al archivo un orden conveniente previo a la selección
- $\Rightarrow$ estratificación implícita

Sistemático de Madow

- Hay en general probabilidades de segundo orden nulas $\pi_{ij} = 0$ (inconveniente teórico)
- $\Rightarrow$ no hay un estimador insesgado de la varianza de $\hat{t}_{\pi_y}$
- $\Rightarrow$ No es sencillo calcular la varianza de $\hat{t}_{\pi_y}$

Ventaja

- Al ser una selección *sistemática*, podemos darle al archivo un orden conveniente previo a la selección
- $\Rightarrow$ estratificación implícita

Sistemático de Madow

- Hay en general probabilidades de segundo orden nulas $\pi_{ij} = 0$ (inconveniente teórico)
- $\Rightarrow$ no hay un estimador insesgado de la varianza de $\hat{t}_{\pi_y}$
- $\Rightarrow$ No es sencillo calcular la varianza de $\hat{t}_{\pi_y}$

Ventaja

- Al ser una selección *sistemática*, podemos darle al archivo un orden conveniente previo a la selección
- $\Rightarrow$ estratificación implícita

Sistemático de Madow

- Hay en general probabilidades de segundo orden nulas $\pi_{ij} = 0$ (inconveniente teórico)
- $\Rightarrow$ no hay un estimador insesgado de la varianza de $\hat{t}_{\pi_y}$
- $\Rightarrow$ No es sencillo calcular la varianza de $\hat{t}_{\pi_y}$

Ventaja

- Al ser una selección *sistemática*, podemos darle al archivo un orden conveniente previo a la selección
- $\Rightarrow$ estratificación implícita

Sistemático de Madow

- Hay en general probabilidades de segundo orden nulas $\pi_{ij} = 0$ (inconveniente teórico)
- $\Rightarrow$ no hay un estimador insesgado de la varianza de $\hat{t}_{\pi_y}$
- $\Rightarrow$ No es sencillo calcular la varianza de $\hat{t}_{\pi_y}$

Ventaja

- Al ser una selección *sistemática*, podemos darle al archivo un orden conveniente previo a la selección
- $\Rightarrow$ estratificación implícita

Sistemático de Madow

- Hay en general probabilidades de segundo orden nulas $\pi_{ij} = 0$ (inconveniente teórico)
- $\Rightarrow$ no hay un estimador insesgado de la varianza de $\hat{t}_{\pi_y}$
- $\Rightarrow$ No es sencillo calcular la varianza de $\hat{t}_{\pi_y}$

Ventaja

- Al ser una selección *sistemática*, podemos darle al archivo un orden conveniente previo a la selección
- $\Rightarrow$ estratificación implícita

Ejemplo

Supongamos el siguiente universo de 6 empresas. Se desea estimar el total de Y mediante una muestra aleatoria de tamaño 3, seleccionada mediante Madow, con probabilidad proporcional a X, que es el valor de Y en un período anterior

Empresa	X	Y	pi
A	2	1.5	0.171
B	5	4	0.429
C	4	5	0.343
D	6	7	0.514
E	10	12	0.857
F	8	6	0.686
Total	35	35.5	3.000

Verificar las probabilidades de selección de las seis unidades
Seleccionar con R 24 muestras, hallar las 24 estimaciones del total de Y
Estimar la varianza del estimador con las 24 estimaciones
Estimar el deff

Ejemplo de selección con Madow

Probabilidades desactualizadas

Empresa	Y(t-1)	Y(t)	pi
A	1	20	0.095
B	2	8	0.190
D	8	4	0.762
E	10	2	0.952
Total	21	34	2.000

$s_1 = \{D, E\} \rightarrow \hat{y}(s_1) = 7.35$
$s_2 = \{A, D\} \rightarrow \hat{y}(s_2) = 215.25$

Verificar que s_1 y s_2 son dos muestras posibles
Hallar todas las muestras posibles

Ejemplo de selección con Madow

Probabilidades desactualizadas

Empresa	Y(t-1)	Y(t)	pi
A	1	20	0.095
B	2	8	0.190
D	8	4	0.762
E	10	2	0.952
Total	21	34	2.000

$s_1 = \{D, E\} \to \hat{y}(s_1) = 7.35$
$s_2 = \{A, D\} \to \hat{y}(s_2) = 215.25$

Verificar que s_1 y s_2 son dos muestras posibles
Hallar todas las muestras posibles

Ejemplo de selección con Madow

Probabilidades desactualizadas

Empresa	Y(t-1)	Y(t)	pi
A	1	20	0.095
B	2	8	0.190
D	8	4	0.762
E	10	2	0.952
Total	21	34	2.000

$s_1 = \{D, E\} \rightarrow \hat{y}(s_1) = 7.35$
$s_2 = \{A, D\} \rightarrow \hat{y}(s_2) = 215.25$

Verificar que s_1 y s_2 son dos muestras posibles
Hallar todas las muestras posibles

Selección Sistemática de Madow

Ejercicio

Del siguiente universo de 8 empresas se desea estimar el total de la variable Y mediante una muestra de tamaño n=3, seleccionadas mediante Madow, con probabilidad de selección proporcional a X (una medición anterior de Y), estimando el total mediante el estimador de H-T.

Empresa	X	Y
A	2.1	1.7
B	5.0	4.5
C	4.2	5.0
D	6.5	7.0
E	10.0	12.0
F	8.0	8.2
G	12.0	12.3
H	6.2	5.8

Seleccionar con R 12 muestras y graficar las doce estimaciones mediante un box-plot
Cuáles son las muestras posibles?.
Qué probabilidad de selección tiene cada una de ellas?
Estimar CV($\hat{t}_{\pi y}$)

Selección Sistemática de Madow

Ejercicio

Del siguiente universo de 8 empresas se desea estimar el total de la variable Y mediante una muestra de tamaño n=3, seleccionadas mediante Madow, con probabilidad de selección proporcional a X (una medición anterior de Y), estimando el total mediante el estimador de H-T.

Empresa	X	Y
A	2.1	1.7
B	5.0	4.5
C	4.2	5.0
D	6.5	7.0
E	10.0	12.0
F	8.0	8.2
G	12.0	12.3
H	6.2	5.8

Seleccionar con R 12 muestras y graficar las doce estimaciones mediante un box-plot
Cuáles son las muestras posibles?.
Qué probabilidad de selección tiene cada una de ellas?
Estimar CV($\hat{t}_{\pi y}$)

Selección Sistemática de Madow

Ejercicio

Del siguiente universo de 8 empresas se desea estimar el total de la variable Y mediante una muestra de tamaño n=3, seleccionadas mediante Madow, con probabilidad de selección proporcional a X (una medición anterior de Y), estimando el total mediante el estimador de H-T.

Empresa	X	Y
A	2.1	1.7
B	5.0	4.5
C	4.2	5.0
D	6.5	7.0
E	10.0	12.0
F	8.0	8.2
G	12.0	12.3
H	6.2	5.8

Seleccionar con R 12 muestras y graficar las doce estimaciones mediante un box-plot
Cuáles son las muestras posibles?.
Qué probabilidad de selección tiene cada una de ellas?
Estimar CV($\hat{t}_{\pi y}$)

Efecto diseño

Efecto Diseño - deff

$$Deff = \frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta^*})}$$

Donde el MAS tiene un tamaño de muestra similar al del diseño complejo

- El deff depende tanto del diseño como del estimador
- En las encuestas a hogares, donde en general hay conglomeración, es habitual que $deff > 1$
- En las encuestas a empresas en general $deff < 1$
- En las encuestas a alumnos en colegios en general $deff > 1$

Efecto diseño

Efecto Diseño - deff

$$Deff = \frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta^*})}$$

Donde el MAS tiene un tamaño de muestra similar al del diseño complejo

- El deff depende tanto del diseño como del estimador
- En las encuestas a hogares, donde en general hay conglomeración, es habitual que $deff > 1$
- En las encuestas a empresas en general $deff < 1$
- En las encuestas a alumnos en colegios en general $deff > 1$

Efecto diseño

Efecto Diseño - deff

$$Deff = \frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta^*})}$$

Donde el MAS tiene un tamaño de muestra similar al del diseño complejo

- El deff depende tanto del diseño como del estimador
- En las encuestas a hogares, donde en general hay conglomeración, es habitual que $deff > 1$
- En las encuestas a empresas en general $deff < 1$
- En las encuestas a alumnos en colegios en general $deff > 1$

Efecto diseño

Efecto Diseño - deff

$$Deff = \frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta^*})}$$

Donde el MAS tiene un tamaño de muestra similar al del diseño complejo

- El deff depende tanto del diseño como del estimador
- En las encuestas a hogares, donde en general hay conglomeración, es habitual que $deff > 1$
- En las encuestas a empresas en general $deff < 1$
- En las encuestas a alumnos en colegios en general $deff > 1$

Efecto diseño

Efecto Diseño - deff

$$Deff = \frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta^*})}$$

Donde el MAS tiene un tamaño de muestra similar al del diseño complejo

- El deff depende tanto del diseño como del estimador
- En las encuestas a hogares, donde en general hay conglomeración, es habitual que *deff* > 1
- En las encuestas a empresas en general *deff* < 1
- En las encuestas a alumnos en colegios en general *deff* > 1

Efecto diseño

Efecto Diseño - deff

$$Deff = \frac{V(\hat{\theta})}{V_{MAS}(\hat{\theta^*})}$$

Donde el MAS tiene un tamaño de muestra similar al del diseño complejo

- El deff depende tanto del diseño como del estimador
- En las encuestas a hogares, donde en general hay conglomeración, es habitual que $deff > 1$
- En las encuestas a empresas en general $deff < 1$
- En las encuestas a alumnos en colegios en general $deff > 1$

Efecto diseño

Ejemplo: Un deff=2 indica que con una MAS de la mitad de tamaño, se hubiera logrado la misma precisión

El deff no indica si el diseño muestral es malo o bueno

En una encuesta para algunos estimadores puede ser $deff > 1$ y para otros $deff < 1$

Efecto diseño

Ejemplo: Un deff=2 indica que con una MAS de la mitad de tamaño, se hubiera logrado la misma precisión

El deff no indica si el diseño muestral es malo o bueno

En una encuesta para algunos estimadores puede ser *deff* > 1 y para otros *deff* < 1

Efecto diseño

Ejemplo: Un deff=2 indica que con una MAS de la mitad de tamaño, se hubiera logrado la misma precisión

El deff no indica si el diseño muestral es malo o bueno

En una encuesta para algunos estimadores puede ser $deff > 1$ y para otros $deff < 1$

Tabla de Contenidos

Un paquete para estimar errores de muestreo en R: *survey*
Debemos primero declarar el diseño: decirle a *survey* las características del diseño muestral

Por ejemplo, si el diseño es un MAS escribiremos

$$diseno < -svydesign(id = \sim 1, weights = \sim pondera, fpc = \sim fpc, data = df)$$

donde fpc es una columna de df con la cantidad de unidades de muestreo de la población

En una selección mediante Madow escribiremos

$$diseno < -svydesign(id =\sim 1, weights =\sim pondera, data = df)$$

omitiendo fpc
En un diseño sistemático con intervalo entero

$$diseno < -svydesign(id =\sim 1, weights =\sim pondera, fpc =\sim fpc, data = df)$$

Lo que en general sobrestimará la varianza...

Y en un diseño por conglomerado, con UPE como identificador de las unidades de primera etapa, estratificado en la primer etapa con la variable *estrato*

$diseno <- svydesign(id =\sim UPE, strata =\sim estrato, weights =\sim pondera, fpc =\sim f$