

# Maestría en Generación de Información Estadística

## Teoría y Técnicas de Muestreo

Augusto E. Hoszowski & Nahuel Fernandez

UNTREF

Octubre 2024

# Tabla de Contenidos

# Bibliografía

- Complex Surveys: A Guide to Analysis Using R. T.Lumley (2011). Wiley
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Muestreo: Diseño y Análisis, Sharon Lohr (2000). Thomson
- Practical Tools for Designing and Weighting Survey Samples (2<sup>o</sup>Ed), R Valliant et.al. Springer

# Bibliografía

- Complex Surveys: A Guide to Analysis Using R. T.Lumley (2011). Wiley
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Muestreo: Diseño y Análisis, Sharon Lohr (2000). Thomson
- Practical Tools for Designing and Weighting Survey Samples (2<sup>o</sup>Ed), R Valliant et.al. Springer

# Bibliografía

- Complex Surveys: A Guide to Analysis Using R. T.Lumley (2011). Wiley
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Muestreo: Diseño y Análisis, Sharon Lohr (2000). Thomson
- Practical Tools for Designing and Weighting Survey Samples (2<sup>o</sup>Ed), R Valliant et.al. Springer

# Bibliografía

- Complex Surveys: A Guide to Analysis Using R. T.Lumley (2011). Wiley
- Model Assisted Survey Sampling, C Sarndal, B Swenson, J Wretman (1992). Springer
- Muestreo: Diseño y Análisis, Sharon Lohr (2000). Thomson
- Practical Tools for Designing and Weighting Survey Samples (2<sup>o</sup>Ed), R Valliant et.al. Springer

# Tabla de Contenidos

# Muestreo por Conglomerado

- Es una forma de extraer una muestra aleatoria
- Las unidades de la población se agrupan en subuniversos denominados *conglomerados*.
- Se selecciona una muestra aleatoria de conglomerados



# Muestreo por Conglomerado

- Es una forma de extraer una muestra aleatoria
- Las unidades de la población se agrupan en subuniversos denominados *conglomerados*.
- Se selecciona una muestra aleatoria de conglomerados

# Muestreo por Conglomerado

- Es una forma de extraer una muestra aleatoria
- Las unidades de la población se agrupan en subuniversos denominados *conglomerados*.
- Se selecciona una muestra aleatoria de conglomerados

# Muestreo por Conglomerado

## Muestreo por conglomerado *monoetápico*

- Selección de una muestra de conglomerados
- Encuestamos a todas las unidades de los conglomerados seleccionados

Ejemplo: Seleccionamos una muestra de colegios (conglomerado de alumnos de sexto grado) y encuestamos en los colegios seleccionados a todos los alumnos de sexto grado.

# Muestreo por Conglomerado

## Muestreo por conglomerado *monoetápico*

- Selección de una muestra de conglomerados
- Encuestamos a todas las unidades de los conglomerados seleccionados

Ejemplo: Seleccionamos una muestra de colegios (conglomerado de alumnos de sexto grado) y encuestamos en los colegios seleccionados a todos los alumnos de sexto grado.

# Muestreo por Conglomerado

## Muestreo por conglomerado *monoetápico*

- Selección de una muestra de conglomerados
- Encuestamos a todas las unidades de los conglomerados seleccionados

Ejemplo: Seleccionamos una muestra de colegios (conglomerado de alumnos de sexto grado) y encuestamos en los colegios seleccionados a todos los alumnos de sexto grado.

# Muestreo por Conglomerado

## Muestreo por conglomerado polietápico

- Selección de una muestra de conglomerados
- Encuestamos a una muestra aleatoria de las unidades de los conglomerados seleccionados

Ejemplo: Seleccionamos una muestra de radios censales (conglomerado de viviendas particulares) y seleccionamos luego en cada radio seleccionado una muestra aleatoria de viviendas particulares. Encuestamos luego a todos los hogares de estas viviendas.

# Muestreo por Conglomerado

## Muestreo por conglomerado polietápico

- Selección de una muestra de conglomerados
- Encuestamos a una muestra aleatoria de las unidades de los conglomerados seleccionados

Ejemplo: Seleccionamos una muestra de radios censales (conglomerado de viviendas particulares) y seleccionamos luego en cada radio seleccionado una muestra aleatoria de viviendas particulares. Encuestamos luego a todos los hogares de estas viviendas.

# Muestreo por Conglomerado

## Muestreo por conglomerado polietápico

- Selección de una muestra de conglomerados
- Encuestamos a una muestra aleatoria de las unidades de los conglomerados seleccionados

Ejemplo: Seleccionamos una muestra de radios censales (conglomerado de viviendas particulares) y seleccionamos luego en cada radio seleccionado una muestra aleatoria de viviendas particulares. Encuestamos luego a todos los hogares de estas viviendas.



# Muestreo por Conglomerado

¿Cuándo se selecciona una muestra por conglomerado?

- Ausencia de un listado con las unidades finales de muestreo.
- Disminuir el costo del trabajo de campo

# Muestreo por Conglomerado

¿Cuándo se selecciona una muestra por conglomerado?

- Ausencia de un listado con las unidades finales de muestreo.
- Disminuir el costo del trabajo de campo

# Muestreo por Conglomerado

## Efectos de la conglomeración

En las encuestas a hogares, en general aumento de la varianza respecto a un MAS (a igual  $n$ ) ?  $deff > 1$

- Según la variable bajo estudio
- Tamaño del conglomerado
- Etc.

A mayor homogeneidad de los conglomerados, mayor aumento de la varianza (respecto a un MAS de tamaño similar)

# Muestreo por Conglomerado

## Efectos de la conglomeración

En las encuestas a hogares, en general aumento de la varianza respecto a un MAS (a igual  $n$ ) ?  $deff > 1$

- Según la variable bajo estudio
  - Tamaño del conglomerado
  - Etc.

A mayor homogeneidad de los conglomerados, mayor aumento de la varianza (respecto a un MAS de tamaño similar)

# Muestreo por Conglomerado

## Efectos de la conglomeración

En las encuestas a hogares, en general aumento de la varianza respecto a un MAS (a igual  $n$ ) ?  $deff > 1$

- Según la variable bajo estudio
- Tamaño del conglomerado
- Etc.

A mayor homogeneidad de los conglomerados, mayor aumento de la varianza (respecto a un MAS de tamaño similar)

# Muestreo por Conglomerado

## Efectos de la conglomeración

En las encuestas a hogares, en general aumento de la varianza respecto a un MAS (a igual  $n$ ) ?  $deff > 1$

- Según la variable bajo estudio
- Tamaño del conglomerado
- Etc.

A mayor homogeneidad de los conglomerados, mayor aumento de la varianza (respecto a un MAS de tamaño similar)

# Muestreo por Conglomerado

## Efectos de la conglomeración

En las encuestas a hogares, en general aumento de la varianza respecto a un MAS (a igual  $n$ ) ?  $deff > 1$

- Según la variable bajo estudio
- Tamaño del conglomerado
- Etc.

A mayor homogeneidad de los conglomerados, mayor aumento de la varianza (respecto a un MAS de tamaño similar)

# Muestreo por Conglomerado

## Efectos de la conglomeración

En las encuestas a hogares, en general aumento de la varianza respecto a un MAS (a igual  $n$ ) ?  $deff > 1$

- Según la variable bajo estudio
- Tamaño del conglomerado
- Etc.

A mayor homogeneidad de los conglomerados, mayor aumento de la varianza (respecto a un MAS de tamaño similar)



# Ejemplos de Conglomerados

- Radios censales: conglomerados de viviendas.
- Escuelas: Conglomerados de alumnos
- Aulas: Conglomerados de alumnos
- Hogares: Conglomerados de personas
- Viviendas: Conglomerados de personas.

# Ejemplos de Conglomerados

- Radios censales: conglomerados de viviendas.
- Escuelas: Conglomerados de alumnos
- Aulas: Conglomerados de alumnos
- Hogares: Conglomerados de personas
- Viviendas: Conglomerados de personas.

# Ejemplos de Conglomerados

- Radios censales: conglomerados de viviendas.
- Escuelas: Conglomerados de alumnos
- Aulas: Conglomerados de alumnos
- Hogares: Conglomerados de personas
- Viviendas: Conglomerados de personas.

# Ejemplos de Conglomerados

- Radios censales: conglomerados de viviendas.
- Escuelas: Conglomerados de alumnos
- Aulas: Conglomerados de alumnos
- Hogares: Conglomerados de personas
- Viviendas: Conglomerados de personas.

# Ejemplos de Conglomerados

- Radios censales: conglomerados de viviendas.
- Escuelas: Conglomerados de alumnos
- Aulas: Conglomerados de alumnos
- Hogares: Conglomerados de personas
- Viviendas: Conglomerados de personas.

# Muestreo por Conglomerado monoetápico

## Conglomerados de igual tamaño

Supongamos  $N$  conglomerados de igual tamaño  $M$  y una MAS de  $n$  conglomerados. Si deseamos estimar la media de una variable  $Y_{ij}$ , medida en la unidad  $j$  del conglomerado  $i$ , entonces un estimador insesgado de  $\bar{Y}$  es

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n \bar{y}_i$$

donde  $\bar{y}_i$  es la media en el conglomerado  $i$  - ésimo de la muestra (al ser todos los conglomerados de igual tamaño). De aquí se deduce que

$$Var(\bar{y}) = \frac{(1 - n/N)}{n} \cdot \frac{\sum_{i=1}^N (\bar{Y}_i - \bar{Y})^2}{N - 1}$$

# Muestreo por Conglomerado monoetápico

## Conglomerados de igual tamaño

En el caso anterior, muestreo por conglomerado monoetápico con conglomerados de igual tamaño  $M$ , la varianza de  $\bar{y}$  tiene otra forma, donde aparece no el  $S^2$  a nivel de conglomerado (colegios), sino el  $S^2$  a nivel de las unidades finales (alumnos):

$$\text{Var}(\bar{y}) = \frac{(1 - n/N)}{n} \cdot \frac{NM - 1}{M^2(N - 1)} \cdot S^2 \cdot [1 + (M - 1) \cdot \rho]$$

donde  $\rho$  es el *coeficiente de correlación intraclase*,

$$-\frac{1}{M-1} \leq \rho \leq 1$$

Si  $\rho = 1$ , máxima homogeneidad de los conglomerados

Si  $\rho$  negativo, heterogeneidad de los conglomerados

# Muestreo por Conglomerado monoetápico

Conglomerados de igual tamaño. Una fórmula útil

Supongamos  $N$  conglomerados de igual tamaño  $M$  y una MAS de  $n$  conglomerados,  $N$  grande entonces

$$\text{Var}(\bar{y}) \cong (1 - n/N) \cdot \frac{1}{n \cdot M} \cdot S^2 \cdot [1 + (M - 1) \cdot \rho]$$



# Muestreo por Conglomerado Monoetápico

## Conglomerados de tamaños desiguales

En este caso disponemos de dos estimadores del total de una variable  $Y_i$ :  
 Estimador insesgado (puede tener alta varianza si los tamaños son muy disímiles)

$$\hat{t}_{\pi y} = \sum_{i \in s} F_i \cdot y_i$$

donde  $F_i$  es el factor de expansión del conglomerado  $i$ , y  $Y_i$  es el total en el conglomerado  $i$ .

Estimador sesgado (pero en general con menor EMC)

$$\hat{t}_{\pi y} = N \cdot \sum_{i \in s} F_i \cdot Y_i / \sum_{i \in s} F_i$$

En el caso de conglomerados de tamaño muy desiguales conviene en general estratificar por tamaño

# Muestreo por Conglomerado Monoetápico

MAS de conglomerados - Estimación de medidas y totales

**survey** (u otro soft) estiman los totales

$$\hat{t}_y = \sum_{i \in s} F_i \cdot Y_i$$

pero la media de Y mediante

$$\hat{\bar{t}}_y = \frac{\sum_{i \in s} F_i \cdot Y_i}{\sum_{i \in s} F_i}$$

Por eso veremos que el CV y deff del estimador del total es en general muy distinto al CV y deff del estimador de la media

# Muestreo por Conglomerado Monoetápico

MAS de conglomerados - Estimación de medidas y totales

**survey** (u otro soft) estiman los totales

$$\hat{t}_y = \sum_{i \in s} F_i \cdot Y_i$$

pero la media de Y mediante

$$\hat{\bar{t}}_y = \frac{\sum_{i \in s} F_i \cdot Y_i}{\sum_{i \in s} F_i}$$

Por eso veremos que el CV y deff del estimador del total es en general muy distinto al CV y deff del estimador de la media

# Conglomeración y medidas de homogeneidad

En algunas disciplinas es importante evaluar la homogeneidad de los conglomerados

En muestreo por conglomerado aparece el coeficiente de correlación intraclase  $\rho$ , que puede tomar valores negativos

Pero hay otros coeficientes. Algunos de ellos los estima el paquete **fishmethods**:

Lohr rho

Adjusted r-square

ANOVA rho

que asumen en general valores similares

# Diferencia entre Estratificación y Muestreo por Conglomerado

Veamos gráficamente la diferencia entre la estratificación y el muestreo por conglomerado

(En general, pero no siempre)

Estratificación  $\Rightarrow$  disminución de la varianza

Conglomeración  $\Rightarrow$  aumento de la varianza



# Diferencia entre Estratificación y Muestreo por Conglomerado

Veamos gráficamente la diferencia entre la estratificación y el muestreo por conglomerado  
(En general, pero no siempre)

Estratificación  $\Rightarrow$  disminución de la varianza



Conglomeración  $\Rightarrow$  aumento de la varianza



# Diferencia entre Estratificación y Muestreo por Conglomerado

Veamos gráficamente la diferencia entre la estratificación y el muestreo por conglomerado

(En general, pero no siempre)

Estratificación  $\Rightarrow$  disminución de la varianza

Conglomeración  $\Rightarrow$  aumento de la varianza



# Diferencia entre Estratificación y Muestreo por Conglomerado

Veamos gráficamente la diferencia entre la estratificación y el muestreo por conglomerado

(En general, pero no siempre)

Estratificación  $\Rightarrow$  disminución de la varianza

Conglomeración  $\Rightarrow$  aumento de la varianza





# Diferencia entre Estratificación y Muestreo por Conglomerado

Veamos gráficamente la diferencia entre la estratificación y el muestreo por conglomerado

(En general, pero no siempre)

Estratificación  $\Rightarrow$  disminución de la varianza



Conglomeración  $\Rightarrow$  aumento de la varianza

