

# Universidad Nacional de Tres de Febrero

Maestría en Generación de Información Estadística

Teoría y Técnicas de Muestreo

Muestreo polietápico y muestreo en dos fases

Augusto E. Hoszowski

## Ejercicio I

El siguiente universo artificial es el listado de 27 alumnos en siete colegios. A cada alumno se le asocia una variable  $Y$ , cuyo total se desea estimar mediante una muestra bietápica de tamaño total en alumnos 12. Compararemos dos estrategias:

A)

- Primera etapa: Muestreo estratificado de 4 colegios, estratificando los colegios según Tipo,  $n=2$  en cada estrato seleccionados mediante MAS
- Segunda etapa: Muestreo aleatorio simple tres alumnos por colegio (o todos si  $M_i < 3$ )
- 1. Hallar las probabilidades de selección de cada alumno
- 2. Seleccionar una muestra de alumnos con ese diseño
- 3. A partir de la muestra seleccionada, con **survey** estimar (declarando ambas etapas de selección) el total de  $Y$ , deff y CV, declarando *ambas etapas de selección*
- 4. (\*) Seleccionar 2000 muestras independientes y estimar con ellas el CV del estimador del total

B)

- Primera etapa: Muestreo estratificado de 4 colegios, estratificando los colegios según Tipo,  $n=2$  en cada estrato seleccionados. Selección mediante Madow, con probabilidad de selección proporcional a la matrícula. Ordenando por código de colegio.
- Segunda etapa: Muestreo aleatorio simple de tres alumnos por colegio
- 1. Hallar las probabilidades de selección de cada alumno
- 2. Seleccionar una muestra de alumnos con ese diseño
- 3. A partir de la muestra seleccionada, con **survey** (ahora sólo declaramos la primera etapa) estimar el total de  $Y$ , deff y CV
- 4. (\*) Seleccionar 2000 muestras independientes y estimar con ellas el CV del estimador de la media
- 5. (\*) Qué diseño tiene menor CV, dos etapas o dos fases?

Tipo Colegio	Colegio	Alumno	Sexo	Y
A	1	1	M	4
A	1	2	M	2
A	1	3	F	6
A	1	4	M	3
A	2	1	M	10
A	2	2	F	18
A	2	3	M	8
A	3	1	M	12
A	3	2	M	6
A	3	3	M	10
A	3	4	M	7
A	3	5	F	12
B	1	1	M	20
B	1	2	M	12
B	1	3	F	24
B	1	4	M	10
B	2	1	M	26
B	2	2	F	30
B	2	3	F	28
B	3	1	M	21
B	3	2	F	36
B	3	3	F	28
B	4	1	F	38
B	4	2	M	29
B	4	3	M	50
B	4	4	F	50
B	4	5	F	50

## Ejercicio II

Con la tabla anterior, deseamos ahora estimar el total de Y pero con un muestreo en dos fases, aprovechando que se sabe que las mujeres tienen, en promedio, un valor de Y superior a los varones. Suponemos que en el marco que nos entregan no está la variable Sexo, por lo que se decide un diseño en dos fases (no hay recursos para encuestar a todas las personas). El diseño será entonces así (de igual tamaño final que el anterior)

- Primera fase: Muestreo estratificado de 4 colegios, estratificado según Tipo, dos colegios en cada estrato seleccionados mediante MAS
- Segunda fase: Muestreo aleatorio estratificado de 12 (seleccionados a partir de la muestra de colegios), seis alumnos en cada categoría de la variable sexo (o todos los que haya si son menos de seis)

En el muestreo en dos fases es complicado hallar las probabilidades de selección de cada individuo. Se calcula entonces una probabilidad  $\pi^*$  así:

$$\pi^* = \pi_1 \times \pi_2$$

donde  $\pi_1$  es la probabilidad de selección en la primera etapa y  $\pi_2$  es la probabilidad de selección en la segunda etapa, *condicional a la muestra seleccionada en la primera fase*. Y el factor de expansión de cada unidad será la inversa del  $\pi^*$  correspondiente.

1. Seleccionar una muestra
2. Hallar las probabilidades  $\pi^*$  de cada alumno de la muestra
3. Estimar el total de Y
4. (\*) Con **survey** estimar el total de Y, deff y CV, declarando ambas etapas de selección
5. (\*) Seleccionar 2000 muestras independientes y estimar con ellas el CV del estimador de la media