Esquema de muestreo

Chihuahua, Colima, Nayarit y Zacatecas

M. Anzarut, F. González, I. Meza, T. Ortiz

### Método

El objetivo del conteo rápido es estimar la proporción de votos en favor de cada candidato. Para cumplir esto, se selecciona una muestra aleatoria de casillas del total y, con la muestra, se calculan intervalos del 95% de confianza para la proporción de votos de cada candidato. A continuación explicamos el criterio bajo el cuál se colectará dicha muestra aleatoria para Chihuahua, Nayarit, Zacatecas y Colima.

El diseño muestral será estratificado en donde, al interior de cada estrato, se seleccionarán casillas mediante un muestreo aleatorio simple sin reemplazo. La asignación de muestra será proporcional al tamaño de cada estrato. Para definir los estratos a emplear:

1. Se crean distintas posibles estratificaciones a partir de cruces de algunas variables.
2. Para cada estado, para cada una de las estratificaciones y para distintos tamaños de muestra, se simulan 200 muestras utilizando como marco la tabla de cómputos distritales del 2018.
3. Con estas muestras se estiman los intervalos de 95% de confianza para la proporción de votos.
4. Se evalúan ciertos criterios con el fin de elegir la estratificación más conveniente.

Los puntos importantes por definir entonces son: ¿Qué estratificaciones se evaluarán? ¿Qué tamaños de muestra se evaluarán? ¿Cómo se estimarán las proporciones de votos? ¿Con qué criterios decidiremos la estratificación óptima?

### Estratifiaciones evaluadas

Utilizamos cuatro variables para construir estratificaciones:

* Distrito federal dentro de cada estado.
* Índice de marginacion a nivel seccion (1 o 2 subgrupos): Esta variable se crea realizando un análisis de componentes principales con los datos del censo de población y vivienda del 2010 [1] y utilizando la primera componente, que se interpreta como un índice de marginación.
* Tamaño de lista nominal (1 o 2 subgrupos).
* Índice de votos en 2015 a nivel sección. (1 o 2 subgrupos): Esta variable se crea realizando un análisis de componentes principales con los datos de votación del 2015 y utilizando la primera componente.

En cada estado, cruzamos el número de distritos federales con una de las siguientes configuraciones:

|  |  |  |  |
| --- | --- | --- | --- |
| estratificacion\_num | indice\_marginacion | lista\_nominal | indice\_votos |
| 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 2 |
| 3 | 1 | 2 | 1 |
| 4 | 1 | 2 | 2 |
| 5 | 2 | 1 | 1 |
| 6 | 2 | 1 | 2 |
| 7 | 2 | 2 | 1 |
| 8 | 2 | 2 | 2 |

Colapsamos algunos estratos del diseño para conseguir un mínimo de 60 casillas por estrato, usando las variables de la tabla anterior de derecha a izquierda y sin colapsar distritos federales. El corte de las variables en subgrupos se hace de forma que cada subgrupo tenga el número más similar de casillas posible.

### Tamaños de muestra evaluados

Mediante el muestreo probabilístico es posible definir estrategias de selección, tamaños de muestra y estimadores, para asegurar que la diferencia entre el estimador y la verdadera proporción no difiera en valor absoluto en más de una cierta cantidad con un 95% de confianza. A la cantidad se le conoce como precisión. Las precisiones indican entonces el tamaño de los intervalos de confianza que presentaremos.

El tamaño de muestra a elegir está directamente relacionado con el nivel de confianza que queremos obtener, las precisiones deseadas y el tamaño de la población total. Buscamos el tamaño mínimo de muestra que nos permita obtener intervalos del 95% de confianza con buenas precisiones. En cuanto a las poblaciones totales, el estado de Chihuahua, en las elecciones del 2018, tenía 5286 casillas, Colima tenía 950, Nayarit 1668 y Zacatecas 2508. Considerando ésto, para el estado de Chihuahua evaluaremos muestras del 4%, 6%, 8% y 10%, mientras que para los estados de Colima, Nayarit y Zacatecas, que tienen un número menor de casillas, evaluaremos muestras del 10%, 15%, 20%, 25% y 30%. Como las fracciones de muestreo consideradas para Chihuahua son distintas presentaremos su análisis por separado.

### Método de estimación

Sea la proporción de votos a favor del candidato , para estimar utilizaremos el estimador de razón combinado , definido como:

donde, para cada estrato , es el tamaño de la muestra, es el número total de casillas, es el número total de estratos, es el número total de candidatos (incluyendo nulos y no registrados) y es el total de votos a favor del candidato en la casilla .

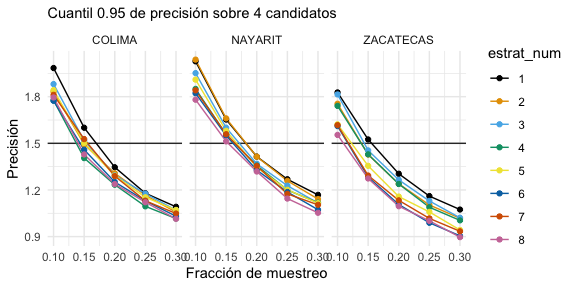
El error estándar de este estimador se calcula utilizando bootrsap no paramétrico (ver, por ejemplo, [2]).

### Criterios de decisión

Entre más fina es una estratificación, mejor es la precisión. Sucede lo mismo con el tamaño de muestra, entre mayor es la muestra mejor es la precisión. Sin embargo, utilizar estratificaciones demasiado finas causa que hayan estratos con pocas casillas, que pueden no ser bien representados a la hora de muestrear. Por otro lado, el considerar muestras demasiado grandes implica un mayor porcentaje de capacitadores-asistentes electorales (CAE) con más de una casilla a reportar, esto complica el trabajo de los CAEs y puede resultar en que no reporten alguna de las casillas. Por esta razón, el diseño muestral evalúa en conjunto estos dos criterios, con el objetivo de encontrar un balance que maximice la precisión sin poner en riesgo la obtención de la muestra.

### Resultados para Colima, Nayarit y Zacatecas

A continuación tenemos una gráfica que muestra el cuantil 0.95 de las precisiones sobre los 4 candidatos mayores para cada estado, para distintos tamaños de muestreo y para cada estratifiación evaluada. La línea negra indica que el cuantil 0.95 de las precisiones es igual a 1.5, esto quiere decir que, bajo esa línea, el 95% de las muestras arrojaron intervalos con precisiones menores a 1.5. Un criterio conservador de selección es buscar fracciones de muestreo y estratificaciones que estén por debajo de la línea negra.



Notamos que para estos 3 estados la estratificación 6 tiene buena precisión, muy similar a la mejor precisión obtenida con la estratifiación 8 que es la más fina, por lo tanto se elige la estratifiación 6 para los 3 estados.

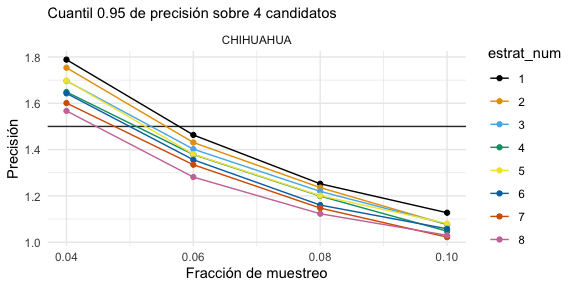
En cuanto a fracciones de muetreo tenemos que para Colima y Nayarit la fracción de 20% nos asegura estar bajo la línea negra, mientras que para Zacatecas basta tomar 15% con la estratificación 6. En la siguiente tabla mostramos el promedio de porcentaje de CAEs con 1, 2 y 3 casillas para la estratificación 6 y algunas fracciones de muestreo que parece razonable analizar:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estado | frac\_muestreo | 1 | 2 | 3 |
| COL | 0.15 | 76 | 21 | 3 |
| COL | 0.2 | 68 | 26 | 5 |
| COL | 0.25 | 61 | 31 | 7 |
| NAY | 0.15 | 73 | 23 | 4 |
| NAY | 0.2 | 64 | 29 | 6 |
| NAY | 0.25 | 56 | 32 | 10 |
| ZAC | 0.15 | 78 | 20 | 2 |
| ZAC | 0.2 | 71 | 25 | 4 |
| ZAC | 0.25 | 63 | 30 | 6 |

El número de CAEs con 1, 2 y 3 casillas parece razonable para fracciones de muestreo de 20% para Colima y Nayarit y 15% para Zacatecas, por lo que confirmamos ésta elección.

## Resultados para Chihuahua

En Chihuahua observamos un patrón similar en cuánto a precisión:



En éste caso, como las fracciones de muestreo consideradas son mucho menores, el porcentaje de casos con 1, 2 y 3 casillas es razonable para cualquiera de éstas fracciones.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estado | frac\_muestreo | 1 | 2 | 3 |
| CHIH | 0.04 | 94 | 6 | 0 |
| CHIH | 0.06 | 90 | 9 | 0 |
| CHIH | 0.08 | 87 | 12 | 1 |
| CHIH | 0.1 | 84 | 15 | 1 |

Elegimos utilizar la estratificación número 3 con una fracción de muestreo de 6%.

## Diseño final

Para el estado de Chihuahua se elige la estratificación 3 que cruza los distritos federales con 2 niveles de lista nominal. Para los estados de Colima, Nayarit y Zacatecas se elige la estratificación 6 que cruza los distritos federales con 2 niveles de índice de marginación y 2 niveles de índice de votos. A continuación aparece el diseño muestral final.

|  |  |  |  |
| --- | --- | --- | --- |
| Estado | Fracción\_muestreo | Número\_estratos | Tamaño\_muestra |
| CHIHUAHUA | 0.06 | 18 | 317 |
| COLIMA | 0.20 | 6 | 190 |
| NAYARIT | 0.20 | 10 | 334 |
| ZACATECAS | 0.15 | 14 | 376 |

### Referencias

[1] <http://gaia.inegi.org.mx/geoelectoral>

[2] <https://tereom.github.io/est_computacional/05-Bootsrap.html>