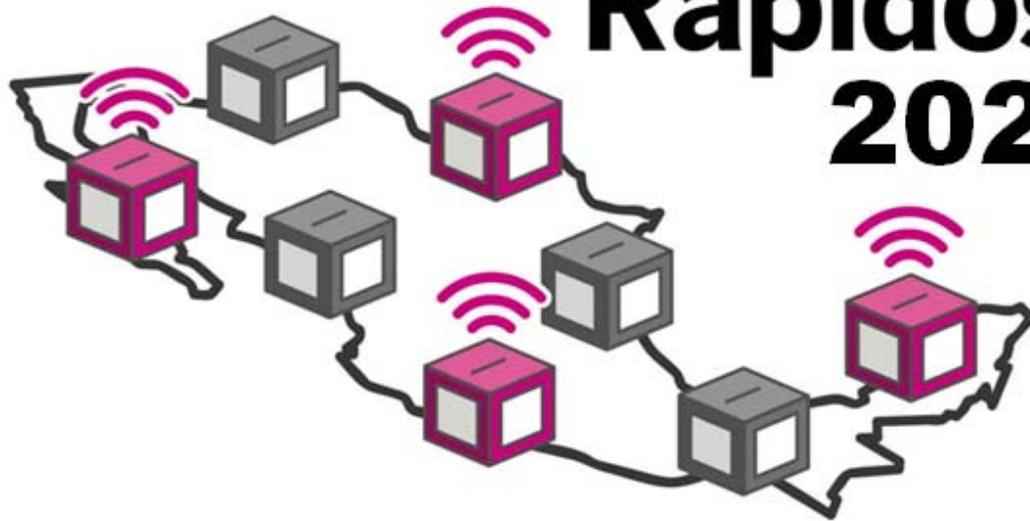


Conteos Rápidos 2024



Criterios científicos, logísticos y operativos para la
realización de los Conteos Rápidos y protocolo
para la selección de las muestras
Procesos Electorales Federales y Locales 2023-2024

CONTENIDO

Glosario y acrónimos	5
1. Antecedentes	7
2. Criterios científicos.....	9
2.1. Elección Presidencial	10
2.1.1. Estratificación y tamaño de muestra	10
2.1.2. Procedimiento de estimación.....	10
2.2. Elección de Senadurías	11
2.2.1. Estratificación y tamaño de muestra	11
2.2.2. Procedimiento de estimación.....	12
2.3. Elección de Diputaciones	12
2.3.1. Estratificación y tamaño de muestra	12
2.3.2. Procedimiento de estimación.....	13
2.4. Elecciones de Gobernatura y jefatura de Gobierno	14
2.4.1. Estratificación y tamaño de muestra	15
2.4.2. Procedimiento de estimación.....	15
2.5. Integración de intervalos.....	16
2.6. Estrategia para la obtención de las muestras	16
2.6.1. Diseños muestrales.....	17
2.6.2. Algoritmo de obtención de muestras	19
2.6.3. Ejemplo ilustrativo de obtención de muestras	20
3. Selección y resguardo de las muestras	26
3.1. Consideraciones generales para la selección de las muestras	26
3.2. Protocolo de selección y resguardo de las muestras	26

3.2.1. Instalación del software para la selección de las muestras.....	26
3.2.2. Selección de las muestras	27
3.2.3. Resguardo de las muestras	27
4. Criterios logísticos y operativos	29
4.1. Operativo de Campo del Conteo Rápido Institucional (OCCRI).....	29
4.1.1. Objetivo	29
4.1.2. Principios	29
4.1.3. Esquema general de funcionamiento	30
4.2. Etapas de la planeación	30
4.2.1. Recursos requeridos	30
4.3. Actividades de preparación.....	34
4.3.1. Capacitación.....	34
4.3.2. Prácticas de captura del SICR	37
4.3.3. Pruebas del SICR	37
4.3.4. Prácticas de marcación a INEtel.....	37
4.3.5. Prácticas de captura y transmisión de la APP MÓVIL	38
4.3.6. Simulacros.....	38
4.3.7. Actividades previas al OCCRI	39
4.4. Ejecución del Operativo de Campo del Conteo Rápido Institucional .	42
4.4.1. Diseño y estructura del FRR.....	42
4.4.2. Recopilación de la información	46
4.4.3. Instrucciones de llenado.....	47
4.4.4. Reporte y captura	47
4.4.5. Flujo de la comunicación	49
4.4.6. Transmisión.....	50
4.5. Plan de contingencia	50

4.6.	Esquema de seguimiento	52
4.6.1.	Ámbito distrital	52
4.6.2.	Ámbito local.....	53
4.6.3.	Ámbito central	54
4.7.	Actividades posteriores al OCCRI.....	55
5.	Medidas de Seguridad en la sede del COTECORA.....	56
6.	Transmisión y difusión del reporte de estimación de los resultados emitidos por el COTECORA.....	57
7.	Anexos. Diseño Muestral y Métodos de Estimación	58
7.1.	Elección Presidencial	58
7.1.1.	Estratificación y tamaño de muestra	58
7.1.2.	Procedimiento de estimación.....	66
7.2.	Elección de Senadores	81
7.2.1.	Estratificación y tamaño de muestra	81
7.2.2.	Procedimiento de estimación.....	98
7.3.	Elección de Diputados.....	111
7.3.1.	Estratificación y tamaño de muestra	111
7.3.2.	Procedimiento de estimación.....	120
7.4.	Elecciones de Gobernatura y Jefatura de Gobierno	131
7.4.1.	Estratificación y tamaño de muestra	131
7.4.2.	Procedimiento de estimación.....	212

Glosario y acrónimos

ACTA	Acta de escrutinio y cómputo de casilla.
APP	Aplicación móvil del Conteo Rápido.
ARE	Área de Responsabilidad Electoral.
AT	Asesor(as) Técnico(as) del Comité Técnico Asesor de los Conteos Rápidos.
CAE	Capacitador(a) Asistente Electoral.
CASILLA	Lugar al que acude la ciudadanía para emitir su voto el día en que celebre la Jornada Electoral.
CAU	Centro de Atención a Usuarios.
CG	Consejo General del Instituto Nacional Electoral.
CR	Conteo Rápido.
COTECORA	Comité Técnico Asesor de los Conteos Rápidos.
CUADERNILLO	Cuadernillo para hacer las operaciones de escrutinio y cómputo para casilla.
DECEYEC	Dirección Ejecutiva de Capacitación Electoral y Educación Cívica.
DEOE	Dirección Ejecutiva de Organización Electoral.
DERFE	Dirección Ejecutiva del Registro Federal de Electores.
DOR	Dirección de Operación Regional de la DEOE.
DPS	Dirección de Planeación y Seguimiento de la DEOE.
ECAE	Estrategia de Capacitación y Asistencia Electoral.
FRR	Formato para recopilación de resultados de la votación.
INE	Instituto Nacional Electoral.
INETEL	Centro de Atención Ciudadana INETEL.
JDE	Junta Distrital Electoral.
JE	Jornada Electoral.
JLE	Junta Local Ejecutiva.

LGIPE	Ley General de Instituciones y Procedimientos Electorales.
MDC	Mesa Directiva de Casilla.
OCCRI	Operativo de Campo del Conteo Rápido Institucional.
OPL	Organismo Público Local.
POL	Programa de Operación Logística.
PEF	Proceso Electoral Federal.
PEL	Proceso Electoral Local.
RE	Reglamento de Elecciones del Instituto Nacional Electoral.
RESRI	Responsable de Sala de Recepción de Información.
SE	Supervisor(a) Electoral.
SICR	Sistema de Información para los Conteos Rápidos.
SIJE	Sistema de Información sobre el desarrollo de la Jornada Electoral.
SIL	Servicio de Internet Local.
UTSI	Unidad Técnica de Servicios de Informática.
VED	Vocal Ejecutivo(a) Distrital.
VEL	Vocal Ejecutivo(a) Local.
VOED	Vocal de Organización Electoral de Junta Distrital Ejecutiva.
VOEL	Vocal de Organización Electoral de Junta Local Ejecutiva.
ZORE	Zona de Responsabilidad Electoral.

1. Antecedentes

El 8 de septiembre de 2023 el Consejo General (CG) del Instituto Nacional Electoral (INE), mediante **Resolución INE/CG530/2023** aprobó ejercer la facultad de asunción parcial para implementar el Conteo Rápido para las elecciones de las personas titulares del Poder Ejecutivo Estatal en Chiapas, Ciudad de México, Guanajuato, Jalisco, Morelos, Puebla, Tabasco, Veracruz y Yucatán.

El 26 de octubre de 2023 el CG del INE, mediante **Acuerdo INE/CG588/2023** determinó realizar los Conteos Rápidos para las elecciones ordinarias de presidencia de los Estados Unidos Mexicanos, Senadurías y Diputaciones Federales; así como para las Gubernaturas de Chiapas, Guanajuato, Jalisco, Morelos, Puebla, Tabasco, Veracruz, Yucatán y la Jefatura de Gobierno de la Ciudad de México, en el marco de los Procesos Electorales Federales y Locales 2023-2024.

En el mismo **Acuerdo INE/CG588/2023** el CG del INE determinó la creación e integración del Comité Técnico Asesor de los Conteos Rápidos 2023-2024 (COTECORA), a fin de pronosticar las tendencias de la votación el día de la jornada electoral para las elecciones federales y las de los titulares del poder ejecutivo que se celebrarán el próximo 2 de junio.

Con base en este acuerdo, el comité quedó integrado por once Asesores Técnicos:

- C. Michelle Anzarut Chacalo.
- C. María Fernanda Gil Leyva Villa.
- C. Manuel Mendoza Ramírez.
- C. Lizbeth Naranjo Albarrán.
- C. Luis Enrique Nieto Barajas.
- C. Gabriel Núñez Antonio.
- C. María Teresa Ortiz Mancera.
- C. Martha Elva Ramírez Guzmán.
- C. Carlos Erwin Rodríguez Hernández-Vela.
- C. Patricia Isabel Romero Mares.
- C. Raúl Rueda Díaz del Campo.

Quienes cumplen con los requisitos estipulados en el RE, son especialistas en el área de Estadística y cuentan con amplia experiencia en diseño muestral. Ocho de ellos han participado en la realización de CR para elecciones federales y

locales que han llevado a cabo el INE o algunos Organismos Públicos Locales, y tres se incorporan por primera vez al Comité.

Para la realización del CR, el comité contará con la colaboración de personal de las direcciones ejecutivas de DERFE y DEOE y de la UTSI.

2. Criterios científicos

En cumplimiento del artículo 367 del RE, el COTECORA presenta los criterios científicos que serán utilizados para realizar las estimaciones de las tendencias de las elecciones de Presidencia de la República, Congreso de la Unión y de los titulares del Poder Ejecutivo de las nueve entidades federativas con PEL.

Los criterios científicos del CR son los procedimientos estadísticos, diseñados con la finalidad de estimar con oportunidad las tendencias de los resultados de una elección, a partir de una muestra probabilística de resultados de actas de escrutinio y cómputo de las CASILLAS electorales, o en su caso, de los CUADERNILLOS para hacer las operaciones de escrutinio y cómputo de CASILLA.

Estos procedimientos, sustentados en la probabilidad y estadística, se usarán para estimar el porcentaje de votos a favor de cada una de las candidaturas a la Presidencia de la República y de las Gubernaturas y Jefatura de Gobierno en entidades con elección del Titular de Poder Ejecutivo, así como para estimar el número de Diputados y de Senadores que obtendrá cada fuerza política en la Cámara correspondiente. Adicionalmente, con los CR también se estimará el porcentaje de participación ciudadana en cada una de estas elecciones.

Los resultados de los CR se presentarán la noche de la Jornada Electoral, con base en el diseño muestral que aquí se define para cada elección.

Las estimaciones deberán realizarse bajo las siguientes consideraciones:

- Los resultados de dichas estimaciones se darán mediante intervalos de confianza/probabilidad.
- Las estimaciones se basarán en los resultados de los CUADERNILLOS para hacer las operaciones de escrutinio y cómputo en CASILLA, obtenidos de las muestras de CASILLAS.
- Las muestras y los modelos de estimación se diseñarán de tal manera que alcancen el 95% de confianza (o probabilidad) y una precisión que genere certidumbre estadística.
- El reporte de resultados que se presentará y será difundido la noche de la JE, especificará las condiciones bajo las cuales se obtuvieron las estimaciones.
- El diseño muestral será estratificado, en cada estrato se seleccionarán CASILLAS mediante un muestreo aleatorio simple sin reemplazo, con lo cual se atenderá lo estipulado en el artículo 373 del RE, el cual determina que la muestra debe abarcar la mayor dispersión geográfica posible.

2.1. Elección Presidencial

2.1.1. Estratificación y tamaño de muestra

En general, para decidir cuál es la mejor estratificación se deben considerar factores como la precisión, o error de estimación, que se pretende obtener, el tamaño de muestra y el nivel de confianza deseado en las estimaciones.

La definición de los estratos para el Conteo Rápido Presidencial se basa en el análisis de los resultados de las elecciones presidenciales de 2012 y 2018. A partir de diferentes escenarios de tamaño de muestra y considerando como criterio de estratificación los distritos federales, se determina un tamaño de muestra de **7,500 casillas** para el cual se calculó una longitud máxima de los intervalos de estimación de 0.47 puntos porcentuales con un nivel de confianza de 95%, considerando los cómputos de la elección de 2018 (véase anexo).

Debido a los diferentes husos horarios que existen en el país y a la dificultad que esto implica para recibir información de manera oportuna, se realizó un análisis de llegada de la muestra en la última elección presidencial, de los resultados obtenidos se determinó que se tomará una sobre muestra en algunos de los estados que tienen una hora de retraso en relación con la hora del centro; en Baja California Sur será de 27.3% y en Sinaloa de 39.0%.

CALCULADO	SOBRE MUESTRA	CRITERIO DE ESTRATIFICACIÓN
7,500	BCS: 27.3% Sinaloa: 39.0%	Distritos federales electorales

2.1.2. Procedimiento de estimación

El COTECORA realizará estimaciones con diferentes enfoques estadísticos: clásico o bayesiano. Dentro de cada uno de estos enfoques, las definiciones específicas de cada uno de los elementos que los integran dan lugar a diferentes procedimientos metodológicos, por lo que se esperan resultados diferentes pero cercanos.

Las estimaciones estarán a cargo de los siguientes equipos:

Equipo 1. Dra. Lizbeth Naranjo Albarrán, Mtra. Patricia Isabel Romero Mares y Dr. Raúl Rueda Díaz del Campo. Cada AT realizará una estimación y entregarán una estimación integrada de sus intervalos.

Equipo 2. Dra. Michelle Anzarut Chacalo, Dr. Manuel Mendoza Ramírez y Mtra. María Teresa Ortiz Mancera.

Equipo 3. Dra. María Fernanda Gil Leyva Villa, Dr. Gabriel Núñez Antonio y Dra. Martha Elva Ramírez Guzmán.

Cada uno de estos equipos producirán intervalos de estimación para el resultado de la elección presidencial que serán combinados con el procedimiento que el comité evalúe como el más apropiado a partir de los ejercicios de simulación que realice.

Los intervalos de estimación serán de al menos el 95% de confianza.

2.2. Elección de Senadurías

2.2.1. Estratificación y tamaño de muestra

Para el Conteo Rápido de la Elección Federal de Senadores, se seguirá un diseño de muestreo estratificado, en donde:

- Los estratos serán las entidades federativas de la República Mexicana por lo que se tendrán 32 estratos.
- El tamaño de muestra total será de 1,620 casillas a nivel nacional, en donde se seleccionarán:

- 50 casillas en cada estado, es decir, un total de 1,600 casillas.
- Una sobre muestra de 10 casillas en dos estados, por diferencia de huso horario y con base en la experiencia de llegada en CR previos.

La sobre muestra total será de 20 casillas, en donde los estados con sobre muestra son: Sonora (10 casillas) y Baja California (10 casillas).

- Con la muestra efectiva, el margen de error en las estimaciones será de:
 - 1% al estimar la participación en la Elección Federal.
 - 0.5% al estimar la Votación Válida Emitida por partido.
 - Entre 6 y 7 asientos mal asignados para el partido con el mayor número de asientos mal asignados al estimar la conformación de la Cámara de Senadores.

Esta estrategia de selección y márgenes de error esperados se obtuvieron tomando como referencia los Cómputos Distritales para las elecciones de senadores de los años 2012 y 2018.

2.2.2. Procedimiento de estimación

En la siguiente tabla se muestra el enfoque estadístico y método de estimación con el que cada uno de los asesores determinará el intervalo, ya sea para la primera, segunda o tercera estimación, de acuerdo con la sección 6.2.2.

ASESOR	ENFOQUE ESTADÍSTICO	TIPO DE ESTIMADOR/MODELO DE ESTIMACIÓN	MÉTODO DE CÁLCULO DE ESTIMACIÓN
Dr. Luis Enrique Nieto Barajas	Bayesiano	Normal	distribución posterior
Dr. Carlos Erwin Rodríguez Hernández-Vela	Bayesiano	Normal	predictiva para el promedio poblacional de votos
Equipo 1	Bayesiano/Clásico	Estimador clásico de un total/Normal	Bootstrap/predictiva del total de votos

2.3. Elección de Diputaciones

2.3.1. Estratificación y tamaño de muestra

Para el Conteo Rápido de la Elección Federal de Diputados, se seguirá un diseño de muestreo estratificado, en donde:

- Los estratos serán los distritos federales, por lo que se tendrán 300 estratos.
- El tamaño de muestra total será de 4,620 casillas a nivel nacional, en donde se seleccionarán:
 - 15 casillas en cada distrito, 4,500 casillas en total.
 - 5 casillas de sobremuestra en algunos distritos debido a diferencias de huso horario y a que en el Conteo Rápido de 2021 se observó una muy baja recepción de casillas en tres estados. La sobre muestra total será de 120 casillas. Los estados con sobre muestra son: Baja California (9 distritos-45 casillas), Guerrero (8 distritos-40 casillas) y Sonora (7 distritos-35 casillas).

- Con la muestra efectiva el margen de error en las estimaciones será de:
 - Entre 6 y 9 asientos mal asignados para el partido con el mayor número de asientos mal asignados al estimar la conformación de la Cámara de Diputados.
 - 0.5% al estimar la participación en la Elección Federal.
 - 0.35% al estimar la Votación Válida Emitida por partido.

Todos estos márgenes de error se calcularon con un 95% de confianza. Estos resultados se obtuvieron tomando como referencia los Cómputos Distritales para las elecciones de Diputaciones en los años 2012, 2015, 2018 y 2021.

- La presión sobre los CAE será mínima. Con información de las casillas instaladas para la elección de diputados de 2021 y seleccionando 5,000 muestras estratificadas por distrito federal (15 casillas por distrito federal para una muestra total de 4,500 casillas), se observó lo siguiente:
 - Del total de CAE que apoyarán para la organización de la elección, sólo 10% participará mandando muestra para el CR de diputados.
 - De este 10%, poco menos del 96% tendrá que reportar sólo una casilla.

2.3.2. Procedimiento de estimación

Para el CR con el que se estimará la conformación de la Cámara de Diputados también se tendrán tres estimaciones, como se indica en el siguiente cuadro.

ASESOR	ENFOQUE ESTADÍSTICO	TIPO DE ESTIMADOR/MODELO DE ESTIMACIÓN	MÉTODO DE CÁLCULO DE ESTIMACIÓN
Dr. Carlos Erwin Rodríguez Hernández-Vela	Bayesiano	Normal	predictiva para el promedio poblacional de votos
Dr. Luis Enrique Nieto Barajas	Bayesiano	Normal	distribución posterior
Mtra. María Teresa Ortiz Mancera (equipo 2)	Clásico	Estimador de razón combinado	Bootstrap

2.4. Elecciones de Gubernatura y jefatura de Gobierno

Para las estimaciones de los conteos rápidos de las elecciones de gubernatura y jefatura de gobierno se acordó que cada equipo de trabajo realizará tres estimaciones para cada entidad federativa, como a continuación se describe:

ENTIDAD	ASESOR TÉCNICO RESPONSABLE	2ª/3ª ESTIMACIÓN
Chiapas	Dra. Lizbeth Naranjo Albarrán	Mtra. Patricia Isabel Romero Mares Dr. Raúl Rueda Díaz del Campo
Ciudad de México	Dr. Raúl Rueda Díaz del Campo	Dra. Lizbeth Naranjo Albarrán Mtra. Patricia Isabel Romero Mares
Guanajuato	Dra. Martha Elva Ramírez Guzmán	Dra. María Fernanda Gil Leyva Villa Dr. Gabriel Núñez Antonio
Jalisco	Mtra. Patricia Isabel Romero Mares	Dra. Lizbeth Naranjo Albarrán Dr. Raúl Rueda Díaz del Campo
Morelos	Dr. Manuel Mendoza Ramírez	Dra. Michelle Anzarut Chacalo Mtra. María Teresa Ortiz Mancera
Puebla	Dr. Gabriel Núñez Antonio	Dra. María Fernanda Gil Leyva Villa Dra. Martha Elva Ramírez Guzmán
Tabasco	Dra. Michelle Anzarut Chacalo	Dr. Manuel Mendoza Ramírez Mtra. María Teresa Ortiz Mancera
Veracruz	Mtra. María Teresa Ortiz Mancera	Dra. Michelle Anzarut Chacalo Dr. Manuel Mendoza Ramírez
Yucatán	Dra. María Fernanda Gil Leyva Villa	Dr. Gabriel Núñez Antonio Dra. Martha Elva Ramírez Guzmán

2.4.1. Estratificación y tamaño de muestra

Con base en los ejercicios realizados para cada entidad (véase Anexo) se determinó el criterio de estratificación y tamaño de muestra a emplearse en el Conteo Rápido de las entidades con elección de Gobernatura/Jefatura de Gobierno, como se muestra en la siguiente tabla.

ESTADO	TAMAÑO DE MUESTRA	CRITERIO DE ESTRATIFICACIÓN
Chiapas	600	Distritos federales electorales
Ciudad de México	600	Distritos federales electorales
Guanajuato	500	Distritos federales electorales
Jalisco	500	Distritos federales electorales
Morelos	200	Distritos federales electorales
Puebla	450	Distritos federales electorales
Tabasco	200	Distritos federales electorales
Veracruz	350	Distritos federales electorales
Yucatán	350	Distritos federales electorales

2.4.2. Procedimiento de estimación

En la siguiente tabla se muestra el enfoque estadístico y método de estimación con el que cada uno de los asesores determinará el intervalo, ya sea para la primera, segunda o tercera estimación, de acuerdo con la sección 6.4.2.

ASESOR	ENFOQUE ESTADÍSTICO	TIPO DE ESTIMADOR/MODELO DE ESTIMACIÓN	MÉTODO DE CÁLCULO DE ESTIMACIÓN
Dra. Michelle Anzarut Chacalo Mtra. María Teresa Ortiz Mancera	Bayesiano y Clásico	Regresión binomial negativa, estructura jerárquica Estimador de razón combinado	Distribución predictiva de la proporción de votos Bootstrap
Dra. María Fernanda Gil Leyva Villa	Bayesiano	Estimador de razón combinado	Bootstrap
Dr. Manuel Mendoza Ramírez	Bayesiano	Normal	Cálculo de la distribución posterior

ASESOR	ENFOQUE ESTADÍSTICO	TIPO DE ESTIMADOR/MODELO DE ESTIMACIÓN	MÉTODO DE CÁLCULO DE ESTIMACIÓN
Dra. Lizbeth Naranjo Albarrán	Bayesiano	Normal ponderado por lista nominal	Distribución posterior
Dr. Gabriel Núñez Antonio	Clásico	Estimador de Razón Combinado	Bootstrap no-paramétrico
Dra. Martha Elva Ramírez Guzmán	Clásico	Estimador de Razón combinado	Remuestreo Bootstrap Montecarlo
Mtra. Patricia Isabel Romero Mares	Clásico	Estimador de razón combinado	Métodos asintóticos
Dr. Raúl Rueda Díaz del Campo	Bayesiano	Modelo normal	Distribución predictiva del total de votos

2.5. Integración de intervalos

Cada uno de los integrantes del COTECORA calculará un intervalo al 95% de confianza o de probabilidad, según corresponda, con los procedimientos antes descritos. Para emitir estimaciones únicas se construirá un intervalo consolidado cuyo procedimiento será definido una vez que se realicen algunas simulaciones y sean discutidas por el COTECORA.

2.6. Estrategia para la obtención de las muestras

La simultaneidad de los conteos rápidos de Presidencia de la República con los de Diputaciones, Senadurías y los nueve conteos rápidos locales requiere establecer un esquema que facilite el trabajo operativo para recabar los datos el día de la Jornada Electoral, para lo cual se estableció la estrategia que a continuación se describe.

Los conteos rápidos de 2024 comprenden 12 procesos electorales:

- Cámara de senadores
- Cámara de diputados
- Presidencia
- 9 elecciones del poder ejecutivo local

Cada elección requiere su propio diseño muestral. Si seleccionáramos las distintas muestras de manera individual, para los estados (y jefatura de gobierno) donde tendremos elección, es posible que una casilla sea seleccionada hasta en 4

muestras. Este escenario complica el trabajo de campo, por lo que el COTECORA desarrolló estrategias que excluyen este escenario particular, y en general que facilitan el trabajo operativo impactando en la menor medida posible las decisiones de diseño de cada conteo rápido.

Tras varias discusiones con personal del INE, el COTECORA concluyó que operativamente los escenarios favorables son:

1. Un mismo CAE reporte los resultados de presidencia y elección local
2. Una misma casilla no sea seleccionada en las muestras de elecciones 1) local-presidencial y 2) alguna de las cámaras.

Para abordar el punto uno, el COTECORA acordó armonizar las muestras de presidencia y elección local, de tal manera que en los estados con elección local, las casillas que formarán parte de la muestra presidencial serán un subconjunto de las casillas seleccionadas para la elección local o viceversa, dependiendo que diseño tenga mayor tamaño de muestra en cada estado. En los estados sin elección local, se seleccionará una muestra presidencial de acuerdo al diseño acordado.

Para atender al segundo punto, se acordó que la selección de la muestra se realice por etapas, asegurando que no haya traslape alguno entre las casillas seleccionadas en las muestras local-presidencial y las cámaras.

En todos los casos, los diseños muestrales seleccionados y la estrategia de selección propuesta, no impactarán los métodos de estimación ni las precisiones objetivo propuestas para cada conteo rápido individual.

2.6.1. Diseños muestrales

La evaluación de diseños de muestra para cada elección se realizó de manera independiente, considerando las necesidades de precisión de cada proceso electoral y la presión a los CAE asignados.

En todos los casos, los diseños de muestra evaluados fueron diseños estratificados, y para asegurar que las casillas entre las muestras local-presidencial y de las cámaras no se traslapen [referencia a demostración], los diseños seleccionados en los conteos locales se restringieron a estratificaciones definidas por Distritos Federales o refinación de Distritos Federales (e.g. Distrito Federal x Tipo de Sección).

Vale la pena notar que el COTECORA convino que restringir los diseños a estratos derivados de Distritos Federales dan precisión y representación aceptable para los objetivos de cada conteo individual.

Es así que los diseños muestrales seleccionados son como sigue:

Tabla 1 Tamaños de muestra por conteo rápido

ELECCIÓN	NÚMERO DE CASILLAS	ESTRATIFICACIÓN
Presidencial ^{a/}	7,500	Distritos Federales Electorales
Senadores	1,620	Estados
Diputados	4,620	Distritos Federales Electorales
Chiapas	600	Distritos Federales Electorales
Ciudad de México	600	Distritos Federales Electorales
Guanajuato	500	Distritos Federales Electorales
Jalisco	500	Distritos Federales Electorales
Morelos	200	Distritos Federales Electorales
Puebla	450	Distritos Federales Electorales
Tabasco	200	Distritos Federales Electorales
Veracruz	350	Distritos Federales Electorales
Yucatán	350	Distritos Federales Electorales

a/ Adicional a las 7,500 casillas se seleccionará una sobre muestra en los estados de Baja California Sur y Sinaloa.

Para repartir la muestra de casillas entre los estratos descritos en la tabla anterior, se seguirá el siguiente procedimiento:

- Para la Cámara de Diputados se asignarán 15 casillas por Distrito Federal electoral, obteniendo 4,500 casillas. Además, 5 casillas de sobremuestra en los 9 distritos de Baja California (45 casillas), en los 8 distritos de Guerrero (40 casillas) y en los 7 distritos de sonora (35 casillas) para una sobremuestra total de 120 casillas y por lo tanto de una muestra total de 4,620 casillas.
- Para la Cámara de Senadores se asignarán 50 casillas por Estado, obteniendo 1,600 casillas. Además, 10 casillas de sobremuestra en Sonora y Baja California para una sobremuestra total de 20 casillas y por lo tanto de una muestra total de 1,620 casillas.
- En el caso de la Elección presidencial, así como todas las elecciones locales, el tamaño de muestra se repartirá de manera proporcional de acuerdo al número total de casillas instaladas en cada estrato. Si n es el tamaño de muestra, N_h el número de casillas instaladas en el estrato h y N el número total de casillas instaladas para la elección particular, entonces el tamaño de la muestra en el estrato h estará dado por

$$n_h^* = n \left(\frac{N_h}{N} \right).$$

El redondeo es crucial porque los cálculos pueden resultar en números fraccionarios, y no se pueden tener fracciones de casillas. Para asegurar que el tamaño de muestra asignado a cada estrato, tras el redondeo, sume el tamaño de muestra global deseado, manteniendo la proporcionalidad y representatividad de cada estrato respecto a la población total, se sigue el siguiente procedimiento:

1. Para cada estrato, se redondea n_h^* hacia abajo al entero más cercano para obtener n_h . Este paso garantiza la asignación de un número entero de casillas por estrato, evitando fracciones de casillas.
2. Se calcula el residuo (e_h) para cada estrato como la diferencia entre el tamaño de muestra proporcional inicial (n_h^*) y el tamaño de muestra redondeado hacia abajo (n_h). Esto es, $e_h = n_h^* - n_h$.
3. Se determina el número total de casillas que faltan para alcanzar el tamaño de muestra global deseado (n) tras el redondeo inicial. Esto se calcula como $e = n - \sum_{h=1}^L n_h$.
4. Se ordenan los residuos (e_h) de manera descendente y se considerando el mismo orden para los tamaños de muestra n_h .
5. Se asigna secuencialmente las observaciones faltantes a los n_h de los estratos, comenzando por el estrato con el residuo más alto: Simplemente se incrementa en +1 cada n_h de acuerdo al orden del paso anterior, hasta agotar el faltante e .

El procedimiento anterior asegura que los tamaños de muestra por estrato n_h , para $h=1, 2, \dots, L$ sean enteros y además que

$$n = \sum_{h=1}^L n_h.$$

2.6.2. Algoritmo de obtención de muestras

El proceso de obtención de la muestra será por etapas de acuerdo a los siguientes pasos:

1. Utilizando el marco nacional de casillas 2024 (*marco-nal-0*) se selecciona la muestra de senadores, con estratos (estados) y tamaños definidos en la Tabla 1.
2. De *marco-nal-0* se eliminan las casillas seleccionadas en el paso 1 y se crea un marco reducido: *marco-nal-1*.

3. Usando *marco-nal-1* se selecciona la muestra de diputados, con estratos (distrito federal) y tamaños definidos en Tabla 1.
4. De *marco-nal-1* eliminan las casillas seleccionadas en el paso 3. y se crea un marco reducido: *marco-nal-2*.
5. Usando *marco-nal-2* se selecciona una muestra con estratos (distritos federales o refinaciones) y tamaños definidos por el máximo entre los tamaños de muestra solicitados para nacional y local en Tabla 1, obtenemos así *muestra-loc-pres-max*. De esta primera muestra se seleccionan las muestras locales con los tamaños correspondientes en Tabla 1.
6. De manera análoga a las muestras locales, usamos *muestra-loc-pres-max* para seleccionar la muestra correspondiente a presidencial, con los tamaños correspondientes.

La estrategia anterior, asegura que las muestras seleccionadas corresponden al diseño muestral propuesto para cada elección, es decir, en cada conteo son muestras estratificadas, y la probabilidad de selección de una casilla en un estrato h es n_h/N_h , donde n_h es el tamaño de muestra en el estrato y N_h es el número total de casillas en el mismo.

2.6.3. Ejemplo ilustrativo de obtención de muestras

Usamos un marco muestral con 156,900 casillas a nivel nacional (correspondiente a 2018 para ilustrar).

Considerando la estratificación por Distrito Federal, observamos que la mayor parte de los estratos tienen entre 400 y 700 casillas.

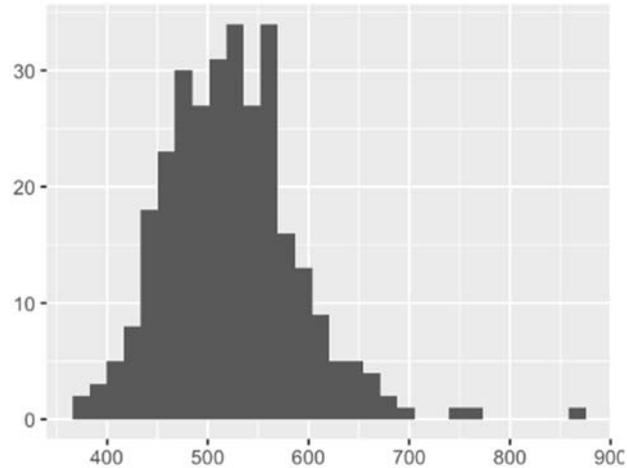
```
library(tidyverse)
library(quickcountmx)
library(readxl)

marco <- read_excel("../datos/MARCO_MJESTRAL_ARE-CASILLAS_110618_FINAL_2018.xlsx")

marco_nal_0 <- marco |>
  mutate(ESTRATO = paste(ID_ESTADO, ID_DISTRITO, sep = "-"),
         consecutivo_casilla = row_number())

marco_nal_0 |>
  count(ESTRATO) |>
  ggplot(aes(x = n)) +
  geom_histogram(bins = 30) +
  labs(x = "", y = "", subtitle = "Número de casillas por estrato marco original")
```

Número de casillas por estrato marco original



1. **Muestra senadores.** Seleccionamos una muestra de tamaño 1,620 asignando 50 casillas en cada entidad federativa, con sobre muestra de 10 casillas para Baja California y Sonora.

```
set.seed(8656)

tab_senadores <- marco_nal_0 |>
  count(ID_ESTADO, NOMBRE_ESTADO, name = "N_h") |>
  mutate(n_h_senadores = ifelse(ID_ESTADO %in% c(26, 2), 60, 50))

muestra_senadores <- select_sample_str(marco_nal_0,
  allocation = tab_senadores,
  sample_size = n_h_senadores, stratum = ID_ESTADO, is_frac = FALSE)

marco_nal_1 <- anti_join(marco_nal_0, muestra_senadores, by = "consecutivo_casilla")

#tamaño de la muestra seleccionada
nrow(muestra_senadores)
```

[1] 1620

```
#tamaño del marco reducido
nrow(marco_nal_1)
```

[1] 155279

2. **Muestra diputados.** Seleccionamos una muestra de tamaño 4,620 asignando 15 casillas por distrito federal y sobre muestra de 5 casillas en los distritos federales de Baja California, Guerrero y Sonora.

```
tab_diputados <- marco_nal_0 |>
  count(ID_ESTADO, NOMBRE_ESTADO, ESTRATO, name = "N_h") |>
  mutate(n_h_diputados = ifelse(ID_ESTADO %in% c(26, 2, 12), 20, 15))

muestra_diputados <- select_sample_str(marco_nal_1,
  allocation = tab_diputados,
  sample_size = n_h_diputados, stratum = ESTRATO, is_frac = FALSE)

marco_nal_2 <- anti_join(marco_nal_1, muestra_diputados,
  by = "consecutivo_casilla")

#tamaño de la muestra seleccionada
nrow(muestra_diputados)
```

```
[1] 4620
```

```
#tamaño del marco reducido
nrow(marco_nal_2)
```

```
[1] 150659
```

3. **Muestra local-presidencial.** Usando los tamaños de Tabla 1 revisamos para cada estado con elección local, el número de casillas siguiendo el diseño de muestra presidencial ($n_{presidencial}$) y el análogo para el diseño local (n_{local}). Para este ejemplo utilizaremos las 7,500 casillas, es decir, ignoramos la sobre muestra.

```

frac_presidencial <- 7500 / nrow(marco_nal_0)

# tamaños de muestra por estrato seleccion nacional
estratos_n_pres <- marco_nal_0 |>
  count(ID_ESTADO, ESTRATO, name = "N_h") |>
  mutate(n_h_nal = round(frac_presidencial * N_h))

# tamaños de muestra por estado diseños locales
tab_muestra <- tribble(~ID_ESTADO, ~edo, ~n_edo,
  7, "Chiapas", 600,
  9, "Ciudad de México", 300,
  11, "Guanajuato", 500,
  14, "Jalisco", 500,
  17, "Morelos", 200,
  21, "Puebla", 450,
  27, "Tabasco", 200,
  30, "Veracruz", 350,
  31, "Yucatán", 350
)

# tamaños de muestra por estrato de las dos elecciones
tab_local_pres <- left_join(estratos_n_pres, tab_muestra, by = "ID_ESTADO") |>
  group_by(ID_ESTADO, edo) |>
  mutate(N_edo = sum(N_h),
    frac_edo = n_edo / N_edo,
    n_h_edo = round(N_h * frac_edo),
    n_h_edo = ifelse(is.na(n_h_edo), 0, n_h_edo),
    n_h = pmax(n_h_nal, n_h_edo, na.rm = TRUE)) |>
  ungroup()

tab_local_pres |>
  group_by(ID_ESTADO, edo) |>
  summarise(n_nal = sum(n_h_nal), n_edo = first(n_edo), .groups = "drop") |>
  na.omit() |>
  arrange(n_nal) |>
  select(Estado = edo, n_presidencial = n_nal, n_local = n_edo) |>
  knitr::kable(format = "pipe")

```

ESTADO	N_PRESIDENCIAL	N_LOCAL
Morelos	115	200
Yucatán	128	350
Tabasco	140	200
Chiapas	301	600
Guanajuato	357	500

ESTADO	N PRESIDENCIAL	N LOCAL
Puebla	361	450
Jalisco	472	500
Veracruz	508	350
Ciudad de México	621	600

Por tanto en el caso de Veracruz y Ciudad de México la muestra local será un subconjunto de la nacional, y lo opuesto para el resto.

3.1. Seleccionamos una muestra del tamaño máximo entre local y presidencial para cada estado (muestra-local-pres-max) usando el marco reducido que obtuvimos en el paso 2.

```
muestra_loc_pres_aux <- select_sample_str(marco_nal_2,
  allocation = tab_local_pres,
  sample_size = n_h, stratum = ESTRATO, is_frac = FALSE)

#tamaño de la muestra conjunta local-presidencial
nrow(muestra_loc_pres_aux)
```

[1] 8421

3.2. Muestra de las elecciones locales.

- Para Veracruz y Ciudad de México: tomamos una muestra estratificada, con asignación proporcional, de tamaño 350 y 600, respectivamente, de muestra-local-pres-max.
- Para el resto de los estados con elección local, seleccionamos todas las casillas obtenidas en 3.1.

```
muestra_local <- select_sample_str(muestra_loc_pres_aux,
  allocation = tab_local_pres,
  sample_size = n_h_edo, stratum = ESTRATO, is_frac = FALSE)

#tamaño de la muestra local
nrow(muestra_local)
```

[1] 3444

3.3. Muestra presidencial. Seleccionamos una muestra de muestra-loc-pres-max (3.1) de acuerdo a los tamaños propuestos, usando estratificación por Distrito Federal y asignación proporcional, como consecuencia, en el caso de los estados sin elección local y en Veracruz y Ciudad de México,

la muestra corresponderá a todas las casillas en muestra-loc-pres-max seleccionada en el inciso 3.1.

```
muestra_presidencial <- select_sample_str(muestra_loc_pres_aux,  
  allocation = tab_local_pres,  
  sample_size = n_h_nal, stratum = ESTRATO, is_frac = FALSE)  
  
casillas_restantes <- marco_nal_2 |>  
  anti_join(muestra_loc_pres_aux, by = "consecutivo_casilla")  
  
#tamaño de la muestra presidencial  
nrow(muestra_presidencial)
```

```
[1] 7501
```

```
#numero de casillas no seleccionadas  
nrow(casillas_restantes)
```

```
[1] 142238
```

Obtenemos así:

- 4 muestras, donde la presidencial y las locales tienen una estructura anidada.
- No hay casillas comunes entre las muestras de diputados, senadores y local-presidencial
- Del total de 156,900 casillas en el marco considerado se utilizan 14,660:
 - 1,620 para la muestra de senadores
 - 4,620 para la muestra de diputados
 - 8,420 para la muestra local-presidencial

3. Selección y resguardo de las muestras

3.1. Consideraciones generales para la selección de las muestras

Las muestras con las que se estimará la votación a favor de las personas candidatas a Presidencia de la República, las Gubernaturas de cada una de las entidades con elección y Jefatura de Gobierno; así como de los partidos políticos y candidaturas independientes para la elección de las cámaras se obtendrán en un acto público el 31 de mayo de 2024, a las 09:00 horas. En el acto estará presente una persona fedataria que atestiguará el desarrollo del protocolo, desde la instalación del software requerido, hasta la obtención y resguardo de las muestras.

Para seleccionar las muestras se hará uso de un equipo de cómputo habilitado con software estadístico. El marco muestral será el listado de las CASILLAS aprobadas para las elecciones del 2 de junio de 2024 y la selección se realizará de acuerdo con el diseño muestral establecido por el COTECORA.

3.2. Protocolo de selección y resguardo de las muestras

Para la selección y resguardo de las muestras se realizarán las siguientes actividades:

3.2.1. Instalación del software para la selección de las muestras

1. Personal de la DERFE realizará ante la persona fedataria la validación del equipo de cómputo donde se instalará el programa para la obtención de las muestras.
2. El COTECORA entregará el software necesario para la selección de las muestras junto con sus códigos de integridad (programa e insumos para la selección de las muestras y la base de datos de CASILLAS).
3. Personal de la DERFE obtendrá los códigos de integridad de los archivos entregados por el COTECORA y los comparará con los entregados en el punto 2. La persona fedataria validará que se trata del mismo código de integridad.
4. A la vista de las y los asistentes se instalará el software estadístico.

5. Se solicitará a un AT del COTECORA iniciar la etapa de ejecución de la selección de las muestras.

3.2.2. Selección de las muestras

1. Para seleccionar las muestras se requiere de un número aleatorio denominado semilla, el cual se utilizará para generar las 9 muestras, una para cada entidad federativa -en el caso de las elecciones para Titular del Poder Ejecutivo local- y para generar las 3 muestras de los CR de las elecciones federales.
2. La semilla se construirá con tres números de seis dígitos.
3. Para construir y capturar los tres números se requiere la participación de seis personas elegidas entre las asistentes. Tres, anotarán un número en un formato diseñado para tal efecto y lo ingresarán en el programa. Para asegurar el correcto ingreso, el programa solicitará la confirmación de los números, los cuales serán nuevamente ingresados por las tres personas restantes.
4. A la vista de la persona fedataria, los números se ingresarán en el programa de selección de las muestras. Los demás asistentes no conocerán estos números.
5. Una vez ingresados los números para construir la semilla, se ejecutará el programa para seleccionar las muestras. Estas últimas quedarán grabadas en el disco duro.
6. Se generará un código de integridad de las muestras. El código de integridad será impreso y se entregará a la persona fedataria y a las y los asistentes que lo soliciten.

3.2.3. Resguardo de las muestras

1. Las muestras serán grabadas en dos dispositivos/medios de almacenamiento removibles.
2. Un dispositivo removible será entregado al Secretario Técnico del COTECORA para su distribución a los VOED, con el fin de que se preparen los trabajos de acopio de las hojas del CUADERNILLO para hacer las operaciones de escrutinio y cómputo.
3. El otro dispositivo con las muestras, los códigos de integridad y los formatos en los que se anotarán los números para construir la semilla serán guardados en un sobre.

4. Asimismo, el equipo de cómputo en el cual fueron generadas las muestras se resguardará en un sobre.
5. Ambos sobres serán sellados y rubricados por la persona fedataria.
6. La persona fedataria entregará al Secretario Ejecutivo del INE, o bien, a la persona que éste último designe, los sobres a los que se refiere el numeral anterior para su resguardo.
7. El Secretario Ejecutivo del INE y el Secretario Técnico del COTECORA, en sus respectivos ámbitos de competencia, serán responsables del resguardo y la secrecía de las muestras, hasta que se publiquen los reportes de las estimaciones de los resultados de cada elección.

Se fortalecerá la secrecía y la cadena de custodia, desde la generación de la muestra, su transmisión a los distritos electorales y, posteriormente, al COTECORA, así como con la copia de resguardo por parte de la Secretaría Ejecutiva.

4. Criterios logísticos y operativos

4.1. Operativo de Campo del Conteo Rápido Institucional (OCCRI)

La DEOE implementará el OCCRI, en coordinación con las JLE y JDE, cuyo propósito y principios son:

4.1.1. Objetivo

Proveer al COTECORA, de manera confiable y oportuna, la información de los resultados de las votaciones asentados en la HOJA DE OPERACIONES de las CASILLAS que conforman la muestra para los CRI, a fin de que realice las estimaciones estadísticas del porcentaje de votación de cada uno de los contendientes de las elecciones y del nivel de participación ciudadana.

4.1.2. Principios

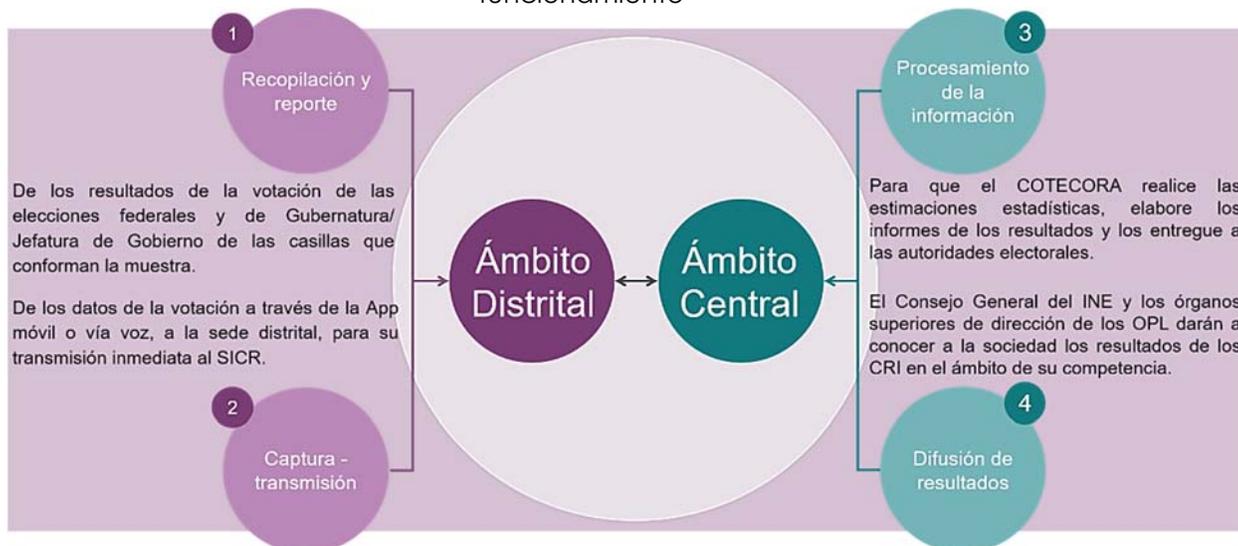
Para el cumplimiento de este propósito, se definieron tres principios de implementación, los cuales tienen un carácter de observancia para el personal involucrado en el OCCRI:



4.1.3. Esquema general de funcionamiento

Se requiere de la planeación, ejecución y seguimiento de una serie de actividades coordinadas que tienen su base en el siguiente procedimiento general:

Imagen 1 Proceso Electoral Concurrente 2023-2024 OCCRI 2024: esquema general de funcionamiento



4.2. Etapas de la planeación

Para que el OCCRI se realice de forma apropiada y exitosa es fundamental determinar su planeación, la cual consistirá en definir los diversos recursos que se requerirán para su ejecución y los procedimientos que deberán realizarse antes, durante y después de la JE.

4.2.1. Recursos requeridos

Con la finalidad de prever las necesidades de los recursos para la adecuada realización del OCCRI, en la noche de la JE, en este apartado se describen los cargos del personal involucrado, recursos materiales y financieros necesarios en las JLE y JDE.

4.2.1.1. Recursos humanos

Vocalía Ejecutiva Local

- ◆ Notificará oportunamente a las VOEL y VED los acuerdos, circulares, oficios, reportes y comunicados relacionados con el OCCRI.
- ◆ Vigilará, en coordinación con el VOEL, el desarrollo del OCCRI, a efecto de garantizar el adecuado flujo del reporte de la información de las CASILLAS de la muestra.

Vocalía de Organización Electoral Local

- ◆ Notificará oportunamente a las VOED los acuerdos, circulares, oficios, reportes y comunicados relacionados con el OCCRI.
- ◆ Vigilará que se apliquen los procedimientos del OCCRI, a efecto de garantizar el adecuado flujo del reporte de la información de las CASILLAS de la muestra.
- ◆ Dará seguimiento puntual del avance en el reporte de las CASILLAS de la muestra.
- ◆ Coordinará el procedimiento que se deba aplicar en caso de la implementación del plan de contingencia.

Vocalía Ejecutiva Distrital

- ◆ Supervisará las actividades que se realicen en el OCCRI, a efecto de garantizar el adecuado flujo del reporte de la información de las CASILLAS de la muestra.
- ◆ Dará seguimiento al procedimiento que se deba aplicar en caso de implementar el plan de contingencia.

Vocalía de Organización Electoral Distrital

- ◆ Será la responsable directa de la ejecución del OCCRI en su ámbito de competencia.
- ◆ Brindará la capacitación correspondiente a las y los CAE, SE, capturistas y a las y los coordinadores distritales que participarán en el OCCRI.
- ◆ Coordinará la implementación de los procedimientos establecidos para el OCCRI.
- ◆ Entregará con antelación los FRR para el OCCRI a las y los CAE y SE.

(continúa)

Vocalía de
Organización
Electoral Distrital

- ◆ Notificará oportunamente a las y los CAE y SE a su cargo respecto de la o las CASILLAS de la muestra que deberán reportar y de los horarios en los que se deberán presentar en el lugar donde se instalaron, diferenciando el horario de la CASILLA con reporte de la elección de Presidencia y/o Gobernatura con el de senadurías y/o diputaciones federales.
- ◆ Analizará anticipadamente las condiciones y situaciones particulares de su Distrito Electoral, la ubicación de las CASILLAS de la muestra y número de elecciones a reportar, a fin de determinar aquellos casos en que las y los SE, otro/a CAE o CAEL deberá apoyar con el reporte de la información de alguna CASILLA de la muestra.
- ◆ Dará seguimiento al reporte de las CASILLAS de la muestra y mantendrá comunicación constante con las y los SE para garantizar el flujo de la información.
- ◆ Instruirá al personal en campo (CAE o SE) respecto a la instancia a la que deberá comunicarse para garantizar el reporte de la información de las CASILLAS de la muestra, en caso de la saturación en las líneas telefónicas o de presentarse situaciones de emergencia que obstaculicen el reporte de la información a la sede distrital.
- ◆ Informará, en su caso, a las instancias correspondientes respecto a las problemáticas y escenarios que se presenten de forma imprevista y que impidan el reporte de la información de las CASILLAS de la muestra a la sede distrital, a fin de determinar la implementación del plan de contingencia.
- ◆ Aplicará, de ser necesario, el plan de contingencia y brindará las indicaciones correspondientes a las y los CAE o SE.
- ◆ Proporcionará, en su caso, la información que requiera el COTECORA para la verificación de datos atípicos.
- ◆ Resguardará los FRR utilizados durante el OCCRI para proporcionarlos a la DPS cuando se les requieran.

Coordinador(a)
Distrital

- ◆ Reportará al CAU y dará seguimiento a las problemáticas que se presenten en el SICR y la APP MÓVIL.
- ◆ Verificará, en su caso, la personalidad de las y los CAE o SE que no cuenten con su clave de autenticación, la cual proporcionará a la o al capturista para que la ingrese al SICR y se proceda a realizar el reporte y la captura de la información.
- ◆ Apoyará a la VOED en el seguimiento del reporte de las CASILLAS de la muestra a través de los reportes que se generen en el SICR.
- ◆ Dará seguimiento, en su caso, al procedimiento que se aplique durante alguna contingencia e informará a la VOED lo conducente.

Capturista

- ◆ Participará activamente en la capacitación para conocer o reforzar los conocimientos, respecto a los procedimientos que se implementarán y las funciones que desempeñará durante la ejecución del OCCRI.
- ◆ La tarde de la JE, ingresará al SICR a las 17:45 horas (horario local) para verificar su acceso; en caso de no poder acceder, reportará la situación a la o el Coordinador Distrital.
- ◆ Recibirá las llamadas telefónicas de las y los CAE y capturará la información que le reporten en el SICR, conforme al protocolo de dictado establecido.
- ◆ Una vez que haya capturado, validado y guardado la información sobre los resultados de una casilla de la muestra en el SICR, proporcionará al personal que la reportó el código de confirmación.
- ◆ Finalizará la llamada telefónica una vez que la o el CAE le haya confirmado que registró el código de confirmación en el FRR.

Supervisor(a) Electoral

- ◆ Apoyará, a solicitud de la VOED, en el acopio y reporte de la información de la o las CASILLAS de la muestra cuando el o la CAE deba reportar más de una CASILLA.
- ◆ Mantendrá comunicación constante con las y los CAE bajo su responsabilidad para conocer el avance del reporte de las CASILLAS de la muestra y atenderá, en su caso, cualquier eventualidad que impida el reporte oportuno de la información.
- ◆ Informará oportunamente a la VOED respecto al avance en el reporte de la información de las CASILLAS de la muestra perteneciente a su ZORE.
- ◆ Establecerá comunicación con las y los CAE a su cargo que no hayan reportado la información de la o las CASILLAS de la muestra, a fin de conocer la situación que impide realizar el reporte y, en su caso, propondrá soluciones.
- ◆ Resguardará los FRR de las y los CAE a su cargo para su posterior entrega a la VOED.

Capacitador(a) Asistente Electoral

- ◆ En su caso, brindará orientación y apoyo a las y los funcionarios de la MDC, para agilizar la ejecución del escrutinio y cómputo de la o las CASILLAS de la muestra asignadas.
- ◆ Recopilará, en el FRR, los resultados de las votaciones de la o las elecciones federales y/o de Gubernatura que se registraron en la HOJA DE OPERACIONES de la o las CASILLAS de la muestra asignadas, conforme a las indicaciones referidas en este formato.
- ◆ Reportará inmediatamente la información que anotó en el FRR, a través de la APP MÓVIL o vía voz a la sede distrital, conforme a los procedimientos establecidos.
- ◆ Mantendrá, en su caso, comunicación permanente con su SE para reportar las causas que impidan el reporte de la información de la o las CASILLAS de la muestra asignada, a fin de recibir indicaciones que le permitan realizar oportunamente el reporte.
- ◆ Resguardará el o los FRR utilizados en el OCCRI y los entregará a la o el SE.
- ◆ Realizará las actividades que les sean encomendadas por la VOED y la o el SE para la implementación y el cumplimiento del objetivo del OCCRI.
- ◆ Responderá las encuestas respecto al funcionamiento de la APP MÓVIL y a los procedimientos del OCCRI, a través de su dispositivo móvil, con la finalidad de proporcionar información que permita identificar áreas de oportunidad.

Capacitador(a) Asistente Electoral Local

- ◆ Participará en la capacitación para conocer los procedimientos que se implementarán y las funciones que deberá desempeñar durante la ejecución del OCCRI.
- ◆ En caso de que la VOED lo indique, tomará la fotografía de los resultados de la votación registrados en la HOJA DE OPERACIONES de la o las elecciones federales y/o locales de la o las CASILLAS de la muestra asignadas y la enviará a la o el CAE correspondiente².

2. Con fundamento en el Artículo 379, Numeral 4 del RE.

4.2.1.2. Recursos materiales

Para la implementación del OCCRI se utilizará la infraestructura instalada en la SARI de cada una de las JDE, la cual se clasifica en dos categorías:

Materiales para el reporte y recepción de la información

- ◆ Las y los CAE y SE: utilizarán principalmente dispositivos móviles, SIL, telefonía pública rural y teléfonos satelitales, estos últimos, en aquellas CASILLAS ubicadas en lugares alejados que no cuentan con servicio de telefonía convencional.
- ◆ Las y los capturistas dispondrán de diademas telefónicas, sistema multilíneas, equipos de cómputo, líneas telefónicas.

En el caso de la transmisión de la información en las que se utilice telefonía satelital y rural, se realizarán las siguientes acciones para el reporte y la captura de la información:

- a) Una vez que el COTECORA seleccione las casillas electorales de la muestra, la DERFE enviará a las 300 VOED la información de esas casillas; por su parte, las VOED harán del conocimiento a las y los CAE que les hayan tocado reportar casillas correspondientes a su ARE; es decir, el único personal que conocerá la identificación de las casillas de la muestra, previo a la Jornada Electoral, serán las VOED y las y los CAE con casillas asignadas de la muestra.
- b) Para el reporte vía voz y para la captura vía APP de la información para cada una de las casillas electorales de la muestra se generará una *clave de autenticación*, la cual estará impresa en el FRR; dicha clave es única para cada casilla y elección a reportar, la cual será la llave para poder realizar la captura de los votos en el SICR y de la APP. Esta información sólo la conocerán las y los CAE que vayan a reportar casillas para el CRI, adicionalmente, deberán proporcionar, al momento del reporte, la identificación de la casilla (entidad, distrito, sección y tipo de casilla), datos que deberán coincidir con la información cargada en el SICR.

4.2.1.3. Recursos financieros

Se asignarán recursos económicos para que las y los CAE cuenten con un medio de comunicación (moneda fraccionaria para el pago del SIL y telefonía pública rural) que les permita transmitir la información de las CASILLAS de la muestra asignada.

4.3. Actividades de preparación

4.3.1. Capacitación

Con la finalidad de cumplir con el objetivo del OCCRI, la DEOE impartirá capacitaciones al personal involucrado, a fin de dotar los conocimientos y habilidades para que realicen todas las actividades de manera precisa, oportuna y con apego a los procedimientos establecidos.

En este sentido, se tiene previsto realizar lo siguiente:

1. **Cursos de capacitación.** Dar a conocer el esquema de funcionamiento del OCCRI, para que implementen adecuadamente los procedimientos y cumplan cabalmente con las actividades asignadas.

2. **Realización de prácticas de captura, prueba de funcionamiento del SICR y la APP MÓVIL, práctica de marcación a INEtel y simulacros.** El objetivo es probar el funcionamiento de los medios informáticos, así como la implementación de los procedimientos establecidos para el OCCRI, a fin de detectar posibles sucesos que podrían afectar el flujo de la información de las CASILLAS de la muestra, lo que permitirá realizar los ajustes necesarios.

4.3.1.1. Cursos de capacitación

En esta etapa de preparación, la VOED tendrá las siguientes responsabilidades:

- ❖ Participar activamente en el curso del OCCRI 2024, que se impartirá a través del Centro Virtual INE.
- ❖ Capacitar a las y los CAE, SE, capturistas, coordinadores distritales y, en su caso, a las y los CAEL para brindarles los conocimientos y habilidades que les permitan desarrollar sus actividades de manera eficiente.
- ❖ Diagnosticar el desempeño de las y los capturistas, en cuanto a la velocidad y exactitud en la captura de la información para establecer, en su caso, acciones correctivas en aras de optimizar el tiempo de captura y mejorar la calidad de los datos registrados.

Curso para las vocalías

OBJETIVO:	DAR A CONOCER EL ESQUEMA DE FUNCIONAMIENTO DEL OCCRI, A FIN DE QUE SE IMPLEMENTEN ADECUADAMENTE LOS PROCEDIMIENTOS Y LAS FUNCIONES QUE REALIZARÁN DURANTE LAS PRÁCTICAS, PRUEBAS, SIMULACROS Y AL CIERRE DE LA JE PARA LA RECOPIACIÓN, REPORTE, CAPTURA Y TRANSMISIÓN DE LA INFORMACIÓN DE LAS CASILLAS DE LA MUESTRA A LA SEDE DEL COTECORA, A TRAVÉS DEL SICR.
Periodo:	Del 4 al 22 de marzo de 2024.
Horario:	Las 24 horas del día.
Modalidad:	En línea, a través del Centro Virtual INE: https://ine.territorio.la/
Dirigido a:	Las VEL, VED, VOEL y VOED.
Responsable:	La DPS de la DEOE.

Curso para el personal operativo

OBJETIVO:	DAR A CONOCER EL ESQUEMA DE FUNCIONAMIENTO DEL OCCRI Y LAS FUNCIONES QUE REALIZARÁN EN LAS PRÁCTICAS, PRUEBAS, SIMULACROS Y AL CIERRE DE LA JE, A FIN DE QUE IMPLEMENTEN CORRECTAMENTE LOS PROCEDIMIENTOS ESTABLECIDOS.
Periodo:	Del 25 de marzo al 12 de abril de 2024 para el personal de captura y en el caso de los CAE y SE durante la segunda etapa de capacitación.
Horario:	De acuerdo con la programación que determine la VOED.
Modalidad:	Presencial.
Dirigido a:	Las y los CAE, SE, capturistas y coordinadores distritales ¹ .
Responsable:	La VOED, en coordinación con las VOEL.



Importante: la programación de los cursos la realizará la VOED, con el visto bueno de la VOEL, tomado en cuenta el periodo establecido.

Se recomienda considerar la participación de las y los integrantes del Consejo Distrital y del personal de las JDE en los cursos que imparta la VOED, con el propósito de que conozcan el esquema de funcionamiento del OCCRI, dando prioridad a lo relacionado con el seguimiento al reporte de las CASILLAS de la muestra.

Es fundamental que la totalidad de las y los CAE y SE participen en todas las etapas de preparación con la finalidad de que adquieran o fortalezcan sus conocimientos, desarrollen las habilidades y estén suficientemente preparados para realizar el reporte de la información de las CASILLAS de la muestra durante la implementación del OCCRI.

En este mismo sentido, es importante que la totalidad del personal que utilizará el SICR, se familiarice con el uso y funcionamiento del mismo, al cual, para el adecuado manejo de la información únicamente tendrá acceso a nivel central con un rol de consulta, el personal de la Dirección de Planeación y Seguimiento de la DEOE, para dar seguimiento a la recepción de reportes, el personal de la UTSI que brindará soporte al sistema y la DERFE definirá la instancia que realizará la carga de la muestra en el sistema; a nivel de junta local, tendrán acceso las vocalías ejecutivas y de organización electoral con un rol de consulta y en caso de contingencia se podrán habilitar capturistas de Junta Local; a nivel distrital, tendrán acceso las vocalías ejecutivas, de organización electoral así como el

¹ Es necesario prever una capacitación para las y los CAEL, a fin de indicarles la participación que tendrán en apoyo al CRI.

coordinador(a) distrital con un rol de consulta para dar seguimiento al reporte de información desde campo y a las y los capturistas con un rol de captura, quienes recibirán los reportes vía voz y registrarán la información en el sistema.

4.3.2. Prácticas de captura del SICR

OBJETIVO:	EVALUAR A LAS Y LOS CAPTURISTAS EN LA PRECISIÓN, VELOCIDAD Y CAPACIDAD DE CAPTURA DE LA INFORMACIÓN EN EL SICR, PARA ESTABLECER, EN SU CASO, LAS MEDIDAS CORRECTIVAS QUE CONTRIBUYAN A GARANTIZAR LA CALIDAD DE LOS DATOS REGISTRADOS.
Periodo:	Del 8 al 12 de abril de 2024.
Horario:	Libre.
Modalidad:	Presencial.
Dirigido a:	Las y los capturistas, coordinadores distritales y personal de apoyo de las JLE y JDE.
Responsable:	Las VED, en coordinación con las VOED y VOEL.

4.3.3. Pruebas del SICR

OBJETIVO:	Verificar el adecuado funcionamiento del SICR y familiarizarse con la interfaz.
Periodo:	Del 15 al 22 de abril de 2024.
Horario:	15 y 16 de abril de 2024 de las 14:00 a 18:00 horas. Del 17 al 22 de abril de 2024 de las 10:00 a las 14:00 horas (horario del centro del país).
Modalidad:	Presencial.
Dirigido a:	Las y los capturistas, coordinadores distritales y personal de apoyo de las JLE y JDE.
Responsable:	La VED, en coordinación con la VOED y VOEL.

4.3.4. Prácticas de marcación a INEtel

OBJETIVO:	Familiarizar a las y los cae y se con la secuencia de marcación para que realicen el reporte de la información de la casilla asignada a INEtel, conforme a los procedimientos establecidos.
Periodo:	Del 10 al 12 y del 15 al 19 de abril de 2024, conforme a la programación establecida por la DPS.
Horario:	De las 09:00 a las 17:00 horas (horario del centro del país).
Modalidad:	A distancia.
Dirigido a:	Las y los CAE, SE y asistentes de INEtel.
Responsable:	La VED, en coordinación con la VOED y VOEL.

4.3.5. Prácticas de captura y transmisión de la APP MÓVIL

OBJETIVO:	PROBAR EL FUNCIONAMIENTO DE LA APP MÓVIL Y FAMILIARIZARSE CON LOS COMPONENTES QUE INTEGRA CADA MÓDULO, PARA IDENTIFICAR POSIBLES FALLAS, ASÍ COMO EVALUAR A LAS Y LOS CAE EN LA PRECISIÓN Y VELOCIDAD DE CAPTURA DE LA INFORMACIÓN, A FIN DE ESTABLECER, EN SU CASO, MEDIDAS CORRECTIVAS QUE CONTRIBUYAN A GARANTIZAR LA CALIDAD DE LOS DATOS REGISTRADOS.
Periodo:	Entre el 30 de abril y 1 de mayo de 2024 ² .
Horario:	Libre.
Modalidad:	<ul style="list-style-type: none"> ❖ A distancia para los y las CAE. ❖ Presencial para las y los capturistas, Coordinador(a) Distrital y personal de apoyo de las JLE y JDE.
Dirigido a:	Las y los CAE, capturistas ³ , coordinadores distritales y personal de apoyo de las JLE y JDE.
Responsable:	Las VED, en coordinación con las VOED y VOEL.

4.3.6. Simulacros

OBJETIVO:	Probar el funcionamiento de los medios informáticos y de comunicación, así como la implementación de los procedimientos
------------------	---

² La fecha se estableció considerando la conclusión en el desarrollo de la APP MÓVIL señalada por la UTSI a través del oficio INE/UTSI/1119/2024.

³ Las y los capturistas participarán en estas pruebas, monitoreando el avance en el SICR y descargando los reportes que le sean solicitados por parte de la VOED.

	establecidos para el OCCRI, a fin de detectar posibles sucesos que pudieran afectar la recopilación, el reporte, la captura y transmisión de la información de las casillas de la muestra para la estimación de los resultados de los CRI, lo que permitirá realizar los ajustes necesarios para garantizar que la información se envíe de manera oportuna al COTECORA, a través del SICR.
Fechas:	<ul style="list-style-type: none"> ❖ Primero 25 de abril 2024. ❖ Segundo 5 de mayo 2024. ❖ Tercero 19 de mayo 2024.
Horario:	De las 14:00 a las 18:00 horas (horario del centro del país).
Modalidad:	Presencial.
Dirigido a:	Las y los CAE, SE, capturistas, coordinadores distritales y personal de apoyo de las JLE y JDE.
Responsable:	La VED, en coordinación con la VOED y VOEL.



Importante: la totalidad de las y los CAE, SE, capturistas, coordinadores distritales y personal de apoyo de las JLE y JDE deberán participar en los tres simulacros para asegurar que conozcan y apliquen correctamente los procedimientos del OCCRI.

4.3.7. Actividades previas al OCCRI

4.3.7.1. Entrega del material

La VOED entregará a la totalidad de las y los CAE y SE, durante las reuniones de coordinación previas al OCCRI, los documentos que se indican a continuación:

- ❖ **Carta dirigida a la Presidencia de la MDC** para informarle que su CASILLA podría ser seleccionada para el ejercicio del CRI, con el propósito de que permita a la o el CAE, SE o CAEL realice la recopilación de los resultados de la votación asentados en la HOJA DE OPERACIONES. La carta se entregará al o la Presidente de la MDC cuando se le proporcione la documentación y los materiales electorales o al inicio de la JE.
- ❖ **FRR para la JE**, uno por cada CASILLA, se deberá imprimir en dos tantos, una impresión para la o el CAE y otra para la o el SE.

La o el CAE o SE al momento de recibir los FRR deberá verificar que los datos de la identificación de las CASILLAS coincidan con las asignadas a su ARE, así como identificar la clave de autenticación con la que se realizará el

reporte. Por su parte la o el SE deberá verificar que le sea entregado un FRR por cada CASILLA que habrá de instalarse en su ZORE.

- ❖ **Un tríptico que se imprimirá y entregará a cada CAE y SE**, donde se describen las actividades específicas que deben realizar al cierre de la o las CASILLAS de la muestra que le fueron asignadas, para la recopilación y el reporte de la información.
- ❖ **Un tríptico que se imprimirá y entregará a la Presidencia de cada MDC**, donde se describen las actividades en las que participarán las y los funcionarios de las MDC, al cierre de la CASILLA, para la recopilación y el reporte de la información. Se entregará cuando se le proporcione la documentación y los materiales electorales o al inicio de la JE.



Importante: estos documentos estarán disponibles a partir del 20 de mayo de 2024. En el caso de los FRR se descargarán del SICR, mientras que la carta y el tríptico que se entregará a la Presidencia de la MDC, así como el tríptico para cada SE y CAE se descargarán del sitio colaborativo de la DPS.

4.3.7.2. Envío y verificación de la recepción del listado de las casillas de la muestra

La estimación y selección de las muestras de las CASILLAS, por parte del COTECORA, que se utilizarán para los CRI de las tres elecciones federales y, en su caso, de la elección de Gobernatura/Jefatura de Gobierno de la Ciudad de México, se llevará a cabo el 31 de mayo de 2024.

Una vez que se obtengan las muestras, la UTSI en coordinación con la DERFE, remitirá exclusivamente a las y los titulares o encargados de la VOED, vía correo electrónico, el listado de las CASILLAS correspondiente a sus Distrito Electoral Federal.

Por su parte, el personal de la DEOE se comunicará con las y los titulares o encargados de las VOED, el mismo día que se envíe la información, para verificar que la hayan recibido correctamente. Para ello, se aplicará un breve cuestionario, a fin de corroborar la recepción del correo electrónico y que la relación de las CASILLAS de la muestra corresponda a su Distrito Electoral Federal.

4.3.7.3. Notificación al personal de campo

Es importante que la VOED o, en su caso, la o el Coordinador Distrital establezca comunicación con las y los CAE y SE que les asignarán CASILLAS de la muestra para el reporte al CRI.

La VOED tendrá la libertad de definir la estrategia que considere adecuada para garantizar que la totalidad de las y los CAE y SE tengan conocimiento de la o las CASILLAS pertenecientes a su ARE o ZORE que fueron seleccionadas para el reporte al CRI, especificándoles la o las CASILLAS y el o los tipos de elecciones que deberán transmitir; así como el o los horarios en que se deberán presentar en la sede donde se instalaron, considerando un horario para el reporte de la elección de Presidencia y/o Gubernatura y otro para el de senadurías y/o diputaciones federales.

En este sentido, se recomienda que a las y los CAE y SE que se trasladen a zonas incomunicadas se les notifique respecto a la o las CASILLAS que deberán reportar antes de quedar incomunicados y al resto de las y los CAE y SE, sin problemas de comunicación, se les notifique durante el transcurso de la mañana del día de la JE, preferentemente antes de las 11:00 horas.

En caso de requerir el apoyo de las o los CAEL, la VOED se comunicará con dicho personal para indicarle la o las CASILLAS y el o los tipos de elección que deberá reportar a la o el CAE, conforme al procedimiento establecido en el **Anexo 1. Flujo del reporte del CAEL**.



Importante: la VOED comunicará a las y los CAE y SE, durante las reuniones previas a la JE, que en caso de no recibir la llamada o que no aparezca información en la APP MÓVIL de las CASILLAS pertenecientes a su ARE o ZORE, no deberán reportar la información para el CRI.

4.3.7.4. Definición de estrategias

Una vez que la VOED reciba la información de las CASILLAS de la muestra del CRI, solo la podrá usar para lo señalado el apartado *Notificación al personal de campo*, así como para fines de planeación y definición de estrategias operativas con el propósito de:

- ❖ Detectar eventualidades en la cobertura de la señal telefónica o del SIL en las ARE donde se ubicarán las CASILLAS de la muestra.

- ❖ Identificar a las y los CAE que tengan que reportar la información de más de una CASILLA de la muestra en diferente domicilio.
- ❖ Como recomendación para apoyar a las y los CAE que se encuentren en esa situación, la VOED podrá considerar las siguientes opciones de apoyo:
 - Solicitar el apoyo del o la SE a cargo de esa ZORE.
 - Requerir el apoyo de algún CAE que no tenga CASILLAS a reportar al CRI.
 - Considerar el apoyo de algún CAEL, conforme a lo establecido.

Con base en la identificación de las eventualidades que pudieran impedir el flujo de la información a la sede distrital, a través de la APP MÓVIL o vía VOZ, la VOED deberá establecer alternativas de solución y estrategias de apoyo y de comunicación oportuna y expedita para las y los CAE y SE que pudieran enfrentar alguna problemática durante el OCCRI, a fin de transmitir oportunamente la información al COTECORA, cuidando en todo momento la confidencialidad de la muestra.

4.4. Ejecución del Operativo de Campo del Conteo Rápido Institucional

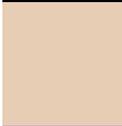
El 2 de junio de 2024, al cierre de la CASILLA iniciará el OCCRI. Previamente el o la CAE o SE entregará a la Presidencia de MDC, la carta de aviso en la que se le informará que su CASILLA podría ser seleccionada para el CRI, así como el tríptico que describe las actividades del CRI en las que participarán las y los funcionarios de la MDC.

4.4.1. Diseño y estructura del FRR

El FRR es el instrumento que utilizará la o el CAE para la recopilación de los resultados de la votación de la CASILLA. Para el OCCRI del PEC 2023-2024, se diseñaron cuatro modelos de acuerdo con el tipo de elección, diferenciándolos en cuanto a la tonalidad del color y logotipo representativo, con la finalidad de que el personal en campo identifique el tipo de elección que debe reportar. Los colores y logotipos que se utilizarán en los FRR se muestran en el cuadro siguiente.

Cuadro 1 Proceso Electoral concurrente 2023-2024 OCCRI: colores y logotipos que se utilizarán en el diseño de los FRR, por tipo de elección

TIPO DE ELECCIÓN	COLOR	LOGOTIPO
❖ Presidencia de la República		

TIPO DE ELECCIÓN	COLOR	LOGOTIPO
❖ Senaduría		 SENADURÍA
❖ Diputación Federal		 DIPUTACIÓN FEDERAL
❖ Gobernatura/Jefatura de Gobierno de la Ciudad de México		Del OPL correspondiente. *
		

(*): A las entidades federativas de Chiapas, Ciudad de México, Guanajuato, Jalisco, Morelos, Puebla, Tabasco, Veracruz y Yucatán.

En lo que respecta a su estructura, se conforma de tres secciones, que se describen en la siguiente imagen:

Imagen 2 Proceso Electoral Concurrente 2023-2024
OCCRI 2024: Formato para la Recopilación de los Resultados de la votación⁴

Sección I
Encabezado e identificación de la CASILLA



CONTEO RÁPIDO 2024
ELECCIÓN DE PRESIDENCIA DE LA REPÚBLICA
AGUASCALIENTES

Nombre completo: _____ CAE SE Distrito Electoral: _____ ARE: _____
Marca con "a"

Identificación de la casilla: _____ Sección electoral _____ Casilla _____ Clave de Autenticación: _____

Resultados de la votación				Si te equivocás al transcribir los datos, utiliza esta columna para corregir
Partido político o coalición	Dictado	Votos (con número)		
	A			
	B			
	C			
	D			
	E			
	F			
	G			
CI	H			
1 2 3	I			
1 2	J			
1 3	K			
2 3	L			
4 5 6	M			
4 5	N			
4 6	O			
5 6	P			
CANDIDATOS/A NO REGISTRADOS/A	Q			
VOTOS NULOS	R			

INSTRUCCIONES:

- Verifica que la identificación de la CASILLA corresponda a la indicada por la VOED.
- Para llenar el FRR, utiliza únicamente tinta azul o negra, evita mezclar diversos colores de tinta.
- Transcribe los resultados de la votación de la Hoja para hacer las operaciones de escrutinio y cómputo de casilla de la elección que tengas que reportar.
- Anota un dígito por recuadro. Si no ocupas espacios a la izquierda, relíenalos con "0", no deberás dejar espacios en blanco. En caso de no registrar datos deberás poner ceros por ejemplo, 000, 001 o 012.
- Asegúrate que los datos anotados en este formato coincidan con los de la Hoja para hacer las operaciones de escrutinio y cómputo de casilla.
- Reporta INMEDIATAMENTE a través de la APP MÓVIL o, en su caso, a los números siguientes:

Teléfono de Junta Distrital Ejecutiva
800 433 2000
Teléfono de INEtel

Teléfono de la Junta Local Distrital (en caso de contingencia)
Anota la hora de término del llenado, utiliza formato de 24 horas (hh:mm):
[] : []

Firma de la/el CAE o SE
Anota el código de confirmación que emita la APP MÓVIL o que te dicte la/el capturista:
[] [] [] [] [] [] [] [] [] []

Guarda el formato para su entrega a la/el Supervisor Electoral.

Sección III
Instrucciones y confirmación.

Sección I. Encabezado e identificación de la CASILLA

Está conformada por campos de información prellenada que servirán para que la o el CAE identifique el tipo de elección y la CASILLA que deberá reportar al CRI.

Sección II. Resultados de la votación

En esta sección, la o el CAE anotará los resultados de la votación de la elección y de la CASILLA que deberá reportar al CRI; se divide en tres columnas principales:

⁴ Esta imagen es un ejemplo de un modelo de FRR que se utilizará el día de la JE, el cual incluirá los logotipos de cada partido político, coaliciones y candidaturas independientes que contendrán en cada elección, de acuerdo con el tipo de elección y datos de identificación.

1. Cintillo con la leyenda del simulacro correspondiente o de la JE.
2. Resultados de la votación, columna que se divide en tres apartados:
 - ❖ **Partido político o coalición**, contiene los logotipos de los partidos políticos, coaliciones, candidaturas comunes y candidaturas independientes, candidaturas no registradas y votos nulos.
 - ❖ **Dictado**, contiene las letras del abecedario, con la finalidad de apoyar visualmente y agilizar el reporte de resultados mediante el protocolo de dictado.
 - ❖ **Votos (con número)**, se divide en tres recuadros, donde se registrarán los resultados de la votación, un número por cada recuadro.
3. Columna con la leyenda *Si te equivocas al transcribir los datos, utiliza esta columna para corregir*, se utilizará en caso de haber registrado erróneamente los datos en la columna previa.

Sección III. Instrucciones y confirmación

Se conforma por los siguientes campos de información:

- ❖ **Instrucciones**, describe las acciones que el o la CAE deberá realizar para verificar la identificación de la CASILLA y el llenado de la Sección II. *Resultados de la votación*, marcará con una X cada una de ellas, conforme las vaya realizando.
- ❖ Tres espacios para anotar los números telefónicos, para el reporte vía voz, de:
 - La Junta Distrital Ejecutiva.
 - INEtel.
 - La Junta Local Ejecutiva.
- ❖ **Registro de la hora** en que se concluyó la recopilación de la información, se anotará en formato de 24 horas (hh:mm).
- ❖ Nueve recuadros para anotar el **código de confirmación**, dato que es obligatorio para corroborar que la información fue reportada.
- ❖ Instrucción para el resguardo y entrega del FRR.

Para la recopilación de la información, se implementarán tres etapas principales, de acuerdo con la imagen siguiente.

Imagen 3 Proceso Electoral Concurrente 2023-2024
OCCRI 2024: etapas para la implementación



4.4.2. Recopilación de la información

Actividades que realizarán las y los CAE, para recopilar los resultados de la votación:

1. Se presentarán a las 18:00 horas (horario local) en la sede de la CASILLA que tenga que reportar para el caso de la elección de Presidencia y/o Gubernatura y para la elección de senadurías y/o diputaciones federales en el horario que le indique la VOED o el o la SE.
2. De ser necesario, orientarán a las y los integrantes de la MDC, respecto al desarrollo del escrutinio y cómputo, a fin de agilizar la obtención de los resultados de la votación de las elecciones que se hayan llevado a cabo en la CASILLA.
3. Una vez que la Secretaría de la MDC concluya con el llenado de la "Columna 1" del apartado "Resultados de la votación" de la HOJA DE OPERACIONES, la o el CAE procederá inmediatamente a transcribir los datos de la votación en el FRR.
4. Si se requiere, la o el CAEL brindará apoyo en la recopilación de la información, tomando una fotografía de la HOJA DE OPERACIONES y la enviará inmediatamente a la o el CAE correspondiente.

El apoyo de la o el CAEL lo determinará la VOED, esto derivado de la cantidad de CASILLAS y de elecciones que tenga que reportar de manera

simultánea alguna o algún CAE durante el OCCRI, conforme al **Anexo 1. Flujo del reporte del CAEL**.

4.4.3. Instrucciones de llenado

La o el CAE deberá atender las indicaciones que se describen a continuación:

1. Usará de preferencia bolígrafo de tinta azul o negra.
2. Escribirá su nombre completo en el espacio asignado, empezando por el o los nombres, seguidos de primer y segundo apellidos.
3. Leerá cuidadosamente las instrucciones y las marcará conforme vaya realizando cada actividad.
4. Anotará los resultados de la votación en la columna "Votos (con número)", en el mismo orden que se encuentran en la HOJA DE OPERACIONES, anotará un dígito por recuadro; en caso de no ocupar los recuadros a la izquierda deberá anotar ceros para evitar dejar espacios vacíos, por ejemplo: **009** o **050**.
5. En caso de cometer alguna equivocación, cancelará el dato erróneo con un tache o cruz abarcando los tres cuadros y en la columna "*Si te equivocas al transcribir los datos, utiliza esta columna para corregir*", en la misma fila, anotará la cifra correcta, conforme al punto 4.
6. Anotará el horario en el que concluyó la recopilación de la información, en formato de 24 horas en horario local.
7. Cuando haya concluido y verificado el correcto llenado del FRR, deberá firmarlo en el espacio indicado.

4.4.4. Reporte y captura

1. La o el CAE, al concluir la firma del FRR, **reportará de inmediato** la información considerando el medio de comunicación asignado y la modalidad del reporte:
 - a. Vía APP MÓVIL
 - b. Vía voz



Importante: en elecciones concurrentes, cuando la o el CAE tenga que reportar más de una elección de la misma CASILLA, dará prioridad al reporte de la información de la elección federal y posteriormente a la local o, en su caso, lo hará conforme se vaya concluyendo el escrutinio y cómputo de cada elección.

Vía APP MÓVIL

La o el CAE dará prioridad a este reporte, a fin de agilizar el flujo de la información; conforme a lo siguiente:

1. Ingresará con su usuario y contraseña que le fueron asignados para el SIJE.
2. Seleccionará el tipo de elección que reportará y la o las CASILLAS de la muestra asignadas, correspondientes a su ARE. Si no aparece información, no reportará al CRI.
3. Seleccionará la CASILLA a reportar, registrará la clave de autenticación.
4. Capturará los resultados de la votación en dos ocasiones para su validación.
5. Registrará en el FRR el código de confirmación emitida por la APP MÓVIL. En caso contrario, capturará nuevamente las letras con inconsistencias, conforme lo solicite la APP MÓVIL, hasta que la información sea correcta.
6. Verificará el registro y envío de la información en el módulo de *Consulta*, donde visualizará la o las CASILLAS reportadas.
7. En caso de que aparezca la leyenda "*Hay capturas pendientes de envío*", pulsará en el icono "*Actualizar*", para que se realice nuevamente el envío de la información.
8. De no recibir la confirmación del envío, después de dos minutos de haberla capturado, realizará el reporte inmediatamente vía voz, conforme al subapartado *Vía voz*.

Vía voz

Cuando la o el CAE establezca comunicación vía telefónica, a la sede distrital con la o el capturista, realizará el siguiente procedimiento conforme al **Anexo 2. Protocolo de dictado:**

1. **Datos de identificación.** Dictará, por cifras de dos dígitos, los datos de identificación de la CASILLA y la clave de autenticación.
2. **Reporte y captura de los resultados de la votación.** A la indicación de la o el capturista, dictará en dos momentos los resultados de la votación, el primero será dígito por dígito y el segundo a cifras completas (omitiendo los ceros a la izquierda y que no forman parte integral de la cantidad, en ambos casos).
3. **Confirmación de los datos capturados.** Si la información es consistente, el SICR emitirá el código de confirmación, será dictado, en cifras de dos dígitos, por la o el capturista a la o el CAE para su registro en el FRR.

En caso contrario, el sistema, solicitará las cifras de las letras que no coincidieron, hasta que la información sea consistente y se emita el código de confirmación.

4. **Conclusión de la llamada telefónica.** Una vez que la o el CAE le haya confirmado a la o el capturista que registró el código de confirmación en el FRR, la llamada telefónica concluirá cuando lo indique la o el capturista.

Si la o el CAE tuviera asignadas más CASILLAS de la muestra, deberá trasladarse de inmediato a la CASILLA que se encuentre más próxima a la recién reportada y realizar el mismo procedimiento.

4.4.5. Flujo de la comunicación

En caso de que tenga que hacer uso de telefonía pública rural o satelital, deberá trasladarse de inmediato al lugar donde le brinden el servicio o tenga señal para transmitir la información. En el siguiente diagrama se muestra el procedimiento que deberá realizar el personal de campo (CAE o SE) para hacer el reporte de resultados a las instancias correspondientes:

Imagen 4 Proceso Electoral Concurrente 2023-2024
OCCRI 2024: flujo de comunicación para el reporte de la información



Para establecer contacto con INEtel, deberá considerarse el **Anexo 3. Protocolo de marcación a INEtel.**

4.4.6. Transmisión

La información de la base de datos del SICR será enviada automáticamente al COTECORA, en remesas de información generadas periódicamente, para su procesamiento y realización de las estimaciones estadísticas correspondientes.

4.5. Plan de contingencia

Se implementará ante escenarios imprevistos que obstaculicen el reporte de la información de las CASILLAS de la muestra a la sede distrital o vía APP MÓVIL, entre los que se pudieran presentar:

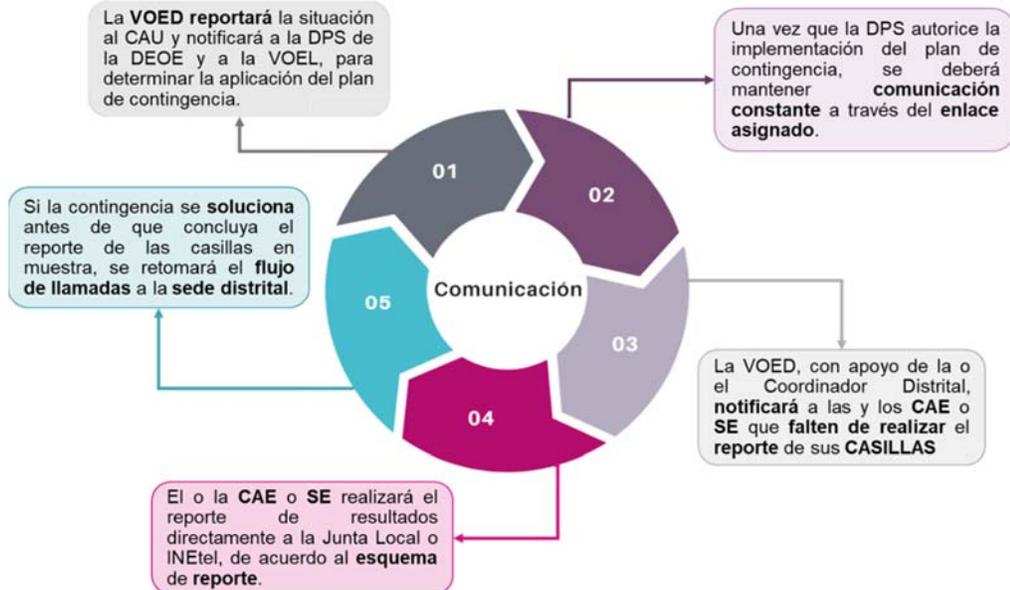
Imagen 5 Proceso Electoral Concurrente 2023-2024
OCCRI 2024: escenarios imprevistos



En estos supuestos o cualquiera otro que obstaculice el reporte de información, se implementará el plan de contingencia para garantizar la continuidad del reporte de las CASILLAS de la muestra; para ello, la JLE correspondiente e INEtel fungirán como primera y segunda instancias de apoyo, respectivamente.

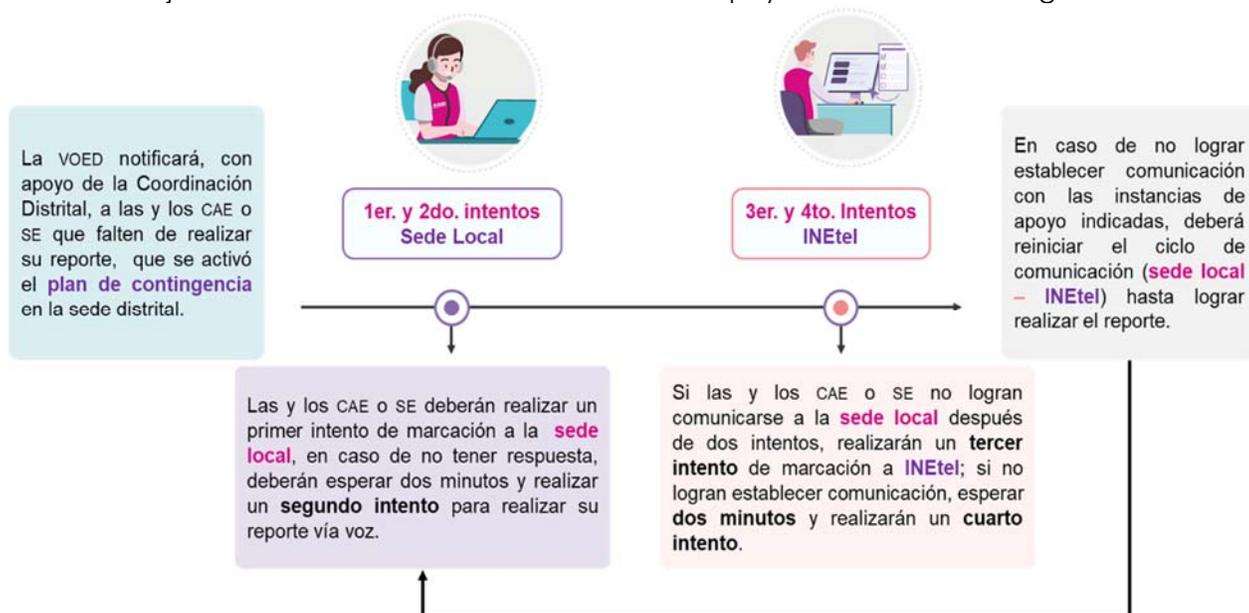
Para implementar el plan de contingencia, se deberá considerar el siguiente procedimiento:

Imagen 6 Proceso Electoral Concurrente 2023-2024
OCCRI 2024: flujo para la activación del plan de contingencia



En cuanto al flujo de comunicación para reportar vía voz, durante la activación del plan de contingencia en la sede distrital, se deberá contemplar el esquema siguiente:

Imagen 7 Proceso Electoral Concurrente 2023-2024
OCCRI: flujo de comunicación con las instancias de apoyo en caso de contingencia



4.6. Esquema de seguimiento

Para asegurar que el reporte de la información de las CASILLAS de la muestra se lleve de manera eficaz y oportuna, se implementarán los esquemas de seguimiento en el ámbito distrital, local y central, los cuales se describen a continuación:

4.6.1. Ámbito distrital

La VOED, con el apoyo de la o el Coordinador Distrital, será la responsable de coordinar las acciones para el adecuado seguimiento al reporte de las CASILLAS de la muestra, considerando lo siguiente:

1. Definir un procedimiento para establecer comunicación constante y expedita con las y los SE y CAE, donde considere grupos de chat en medios de comunicación alternos a los institucionales.

2. Generar, a través del SICR, los reportes de las CASILLAS de la muestra que no hayan sido reportadas.
3. A partir de las 20:00 horas (horario local) y en intervalos de 10 minutos, dará seguimiento al reporte de las CASILLAS de la muestra, a fin de identificar aquellas que se encuentren pendientes de reportar al SICR.
4. En los casos donde identifique que no hay avance en el reporte, se comunicará con las o los SE correspondientes, quienes a su vez contactarán a las y los CAE a su cargo para conocer los contratiempos que estarían impidiendo el reporte de la información. De ser necesario, implementará las acciones pertinentes para garantizar el flujo constante de la transmisión de la información.
5. En el supuesto que tenga conocimiento que se presentó alguna situación que obstaculice el reporte de la información de las CASILLAS de la muestra, la cual tardaría en solucionarse, deberá valorar si se activa el plan de contingencia.
6. Atenderá, en su caso, las peticiones que requiera la DPS respecto a proporcionar la información o imagen del FRR de alguna de las CASILLAS de la muestra en la que el COTECORA haya identificado datos atípicos de los resultados de la votación.

4.6.2. Ámbito local

La VOEL será la responsable del seguimiento al reporte de las CASILLAS de la muestra; por lo que tendrá que realizar lo siguiente:

1. Generará, a través del SICR, los informes de las CASILLAS de la muestra que no hayan sido reportadas.
2. A partir de las 20:00 horas (horario local), y en intervalos de cada 10 minutos, dará seguimiento al reporte de las CASILLAS de la muestra, a fin de identificar a los distritos electorales que no muestren avances en el reporte y a las CASILLAS que se encuentren pendientes por reportar.
3. En caso de identificar que no hay avance en el reporte en alguno de los distritos electorales, se comunicará con la o el titular de la VOED o, en su caso, con la o el Coordinador Distrital correspondiente, a fin de conocer los contratiempos que impiden el reporte. Si fuera el caso, indicará las acciones que considere pertinentes para dar continuidad al flujo de la transmisión de la información.

4. Atenderá, en su caso, las peticiones que requiera la DPS respecto a proporcionar la información o imagen del FRR de alguna de las CASILLAS de la muestra en la que el COTECORA haya identificado datos atípicos de los resultados de la votación.
5. En el supuesto que se implemente el plan de contingencia, dará seguimiento puntual al reporte de las CASILLAS faltantes, a través de los informes que se generen en el SICR.

4.6.3. Ámbito central

El personal de la DPS llevará a cabo el seguimiento para verificar el avance de las CASILLAS de la muestra reportadas y no reportadas; para ello, atenderá lo siguiente:

1. Integrará un directorio que contenga los datos de los medios de comunicación mediante los cuales se contactará a las y los titulares de las VOEL y VOED o, en su defecto, al personal que designen para apoyar el seguimiento o proporcionar información que sea requerida en materia del CRI.
2. Verificará constantemente el avance en el reporte de las CASILLAS de la muestra, a través de los informes que se generen en el SICR, para identificar a los distritos electorales que no muestren avances en el reporte y a las CASILLAS que estén pendientes de reportar. En su caso, informará al superior jerárquico a fin de recibir las indicaciones correspondientes.
3. Se comunicará con la o el titular de la VOEL o con la VOED correspondiente, cuando se identifique un retraso en el reporte de la información de las CASILLAS de la muestra, a efecto de conocer la o las problemáticas que estarían impidiendo el flujo de la transmisión e informará al superior jerárquico lo conducente.
4. Atenderá, en su caso, las peticiones que formule el COTECORA respecto a verificar o proporcionar la información de alguna de las CASILLAS de la muestra en la que se hayan reportado datos atípicos de los resultados de la votación.
5. La persona titular de la DPS compartirá, cuando lo requiera la o el titular de la DEOE, la información del avance del reporte de las CASILLAS de la muestra y de las problemáticas que impidan la transmisión de la información al COTECORA.

4.7. Actividades posteriores al OCCRI

Una vez concluido el OCCRI, se deberán realizar las actividades que se describen a continuación:

1. La o el CAE resguardará el o los FRR que utilizaron para el reporte al CRI y los entregarán a la o el SE o a la VOED. Previamente, verificará que estén debidamente llenados, firmados y que incluya su respectivo código de confirmación.
2. La o el SE que haya recibido los FRR, deberá verificar que estén debidamente requisitados y los entregará a la VOED.
3. La VOED deberá elaborar una relación de los FRR, los resguardará y entregará digitalizados a la DPS, cuando le sean requeridos; esto para disponer de información que servirá para realizar el análisis correspondiente, que permita identificar áreas de oportunidad para fortalecer la capacitación y la operación logística para los próximos procesos electorales o ejercicios de participación ciudadana donde se determine la realización de los CRI.
4. Las y los CAE y SE que participaron en el OCCRI deberán responder la "Encuesta de percepción del OCCRI 2024", la cual estará disponible los días 3 y 4 de junio de 2024, a través de la APP MÓVIL en el módulo "Encuestas".
5. La VOED capturará, los días 3 y 4 de junio de 2024, en la **App DPS_Seguimiento_JE_CRI_2024, que estará disponible en el concentrado de aplicaciones y tableros de la DEOE**, la información requerida de las CASILLAS que no fueron reportadas o que tuvieron dificultades durante el reporte, esto para conocer las causas que afectaron la transmisión de la información.
6. La DPS elaborará un informe de resultados del OCCRI 2024, donde se describirán las actividades realizadas antes, durante y posteriores, así como los logros obtenidos, lo que permitirá, en su caso, identificar áreas de oportunidad para mejorar los procedimientos y el funcionamiento de los instrumentos informáticos.

5. Medidas de Seguridad en la sede del COTECORA

El día 2 de junio de 2024, el recinto acondicionado (búnker) para llevar a cabo la estimación de las tendencias de los resultados de la votación de la Presidencia de los Estados Unidos Mexicanos, Senadurías y Diputaciones Federales, así como para las Gubernaturas de Chiapas, Guanajuato, Jalisco, Morelos, Puebla, Tabasco, Veracruz y Yucatán y la Jefatura de Gobierno de la Ciudad de México, desde el inicio de sus operaciones y hasta que se haya entregado el último Reporte de las estimaciones de los resultados de votación, estará sujeto a lo siguiente:

- Sólo tendrá acceso al recinto personal de apoyo adscrito a la DERFE, UTSI e integrantes del Comité Técnico y sus colaboradores.

La DERFE, quince días antes de la jornada electoral, presentará para su aprobación a la CRFE los nombres de las personas que podrán acceder al búnker y cuando menos integrará los nombres completos y área de adscripción en el Instituto.

- Las personas que tengan acceso al recinto deberá registrar su ingreso en una bitácora donde se señale la fecha y hora de entrada y salida.
- El acceso a Internet en cualquier dispositivo está prohibido dentro del recinto. Sólo podrá usarse la red local dispuesta por el INE exclusivamente para las computadoras en las que se procesará la información relativa a los respectivos Conteos Rápidos.
- El ingreso al recinto con teléfonos celulares o cualquier dispositivo de comunicación está terminantemente prohibido. La DERFE deberá solicitar a todas las personas con acceso al recinto la entrega de dichos dispositivos, los cuales deberán quedar registrados y mantenerse resguardados por personal del Instituto.
- Las personas que no formen parte del Comité Técnico pero estén presentes en el búnker, tendrán un papel operativo y solo podrán participar como observadoras en las deliberaciones del Comité.

6. Transmisión y difusión del reporte de estimación de los resultados emitidos por el COTECORA

Hasta que los reportes de resultados con la estimación de la votación no sean difundidos por la Presidencia del CG o de los Órganos Superiores de Dirección de los OPL, la información se considerará confidencial y reservada.

La DERFE, quince días previos al día de la jornada electoral, presentará para su aprobación a la CRFE el procedimiento de envío del reporte al CG u Órgano Superior de Dirección correspondiente que, cuando menos, deberá contener:

- Formato del reporte y resultados.
- Nombre de la o las personas servidoras públicas que remitirán y recibirán el reporte para entregarlo a la Presidencia del órgano correspondiente previo a su difusión.
- Vías para remitir el reporte.
- Medidas de seguridad, logística y control para salvaguardar la secrecía del reporte hasta su difusión.
- En el caso de los OPL, una guía de mensajes para dar a conocer los resultados del Conteo Rápido.

7. Anexos. Diseño Muestral y Métodos de Estimación

7.1. Elección Presidencial

7.1.1. Estratificación y tamaño de muestra

7.1.1.1. Dr. Manuel Mendoza Ramírez

Objetivos.

El objetivo del Conteo Rápido 2024, para la elección Presidencial, es obtener estimaciones del porcentaje de votos emitidos a favor de los candidatos al cargo de Presidente de la República, así como del porcentaje de ciudadanos que acudan a votar. El Conteo Rápido es un ejercicio estadístico y para llevarlo a cabo se desarrollarán diversos procedimientos técnicos que tienen sustento en la Teoría de la Inferencia Estadística.

Antecedentes.

La Inferencia Estadística se ha desarrollado progresivamente durante más de 100 años y a partir de su estructura conceptual, el rigor de sus resultados teóricos y su desempeño en la práctica, se ha establecido como una disciplina científica muy reconocida y como un estándar aceptado internacionalmente para el estudio de los fenómenos que presentan variabilidad. En el caso específico de este conteo rápido se recurrirá, por una parte, a la Teoría de Muestreo para Poblaciones Finitas con el propósito de establecer, en primer lugar, un diseño eficiente de la muestra que proveerá la información para la estimación. Por otra parte, de entre los Métodos Estadísticos de Estimación, se seleccionarán los mecanismos que se habrán de utilizar, con los datos que ofrezca la muestra, para producir las estimaciones.

Diseño Muestral.

Consideraciones generales.

Para efectos de este conteo rápido se considera como población bajo estudio a la totalidad de las casillas que se instalarán en el país, el día de la jornada electoral. En esa población es de interés estimar, como objetivo central, el porcentaje de votos emitidos a favor de cada candidata o candidato al cargo de Presidente. Además, interesa también el número total de votos que se

recibirán para la elección de Presidente y consecuentemente, el porcentaje de ciudadanos, en la lista nominal, que acudirán a ejercer ese derecho.

Estas estimaciones se producirán a partir de la información recolectada, al cierre de la votación, en una muestra de casillas que se habrá de definir mediante un mecanismo de selección probabilística la cual, con la finalidad de atender a lo dispuesto en el Artículo 373 del Reglamento de Elecciones del Instituto Nacional Electoral, debe abarcar la mayor dispersión geográfica posible. Para cumplir con ese propósito, se utilizará un diseño muestral estratificado. Específicamente, el país se dividirá en regiones que servirán como estratos y de cada uno de los estratos se seleccionará una muestra de casillas completamente al azar, sin reemplazo. El tamaño de la muestra en cada estrato será proporcional al tamaño del propio estrato.

Este diseño, además de responder al requerimiento del reglamento, es muy común en la práctica del muestreo. Produce estimaciones más precisas que las correspondientes a una muestra que se selecciona con un muestreo aleatorio simple, sin restricciones geográficas o de otro tipo. Para definir un diseño estratificado es necesario establecer, en primer lugar, los criterios con que se dividirá el país en estratos y, una vez con tales criterios, el siguiente paso es seleccionar el tamaño de total de la muestra.

La determinación de los criterios de estratificación se lleva a efecto tomando en cuenta que el diseño produce mejores resultados en la medida en que las unidades de muestreo dentro de cada estrato son más homogéneas entre sí y, al mismo tiempo, diferentes de las que pertenecen a otros estratos. La homogeneidad en esta conceptualización, se refiere al comportamiento de la variable, objeto de estudio; en este caso, la votación observada.

Por otra parte, el tamaño de la muestra se fija de manera que la estimación alcance niveles aceptables de confiabilidad y precisión. En el caso del Conteo Rápido para la elección de Presidente de la República en 2024, el nivel de confiabilidad, como también lo establece el Reglamento de Elecciones, se ha fijado en 0.95. De esta forma, para cada propuesta de estratificación, se examina la precisión que producen los distintos tamaños de muestra, considerando este nivel de confiabilidad.

Criterio de estratificación.

Para determinar el criterio de estratificación en 2024, el Comité Técnico de los Conteos Rápidos acordó, en su sesión del 20 de marzo, tomar en cuenta la experiencia que ofrecen los resultados de los conteos rápidos que se han llevado a cabo para la elección presidencial en 2006, 2012 y 2018. En estos casos, la estratificación se diseñó a partir de la división del país en sus 300 distritos

electorales federales. En 2006 y 2012, los distritos se dividieron, además, cuando su composición lo hizo posible, en sus partes urbana y no urbana. La conveniencia de esta estratificación se validó con los estudios de simulación que establecieron los niveles de precisión y confiabilidad apropiados. Para 2018, estos estudios verificaron que la división de los estratos en esas componentes ya no aportaba una mejora sustancial en la calidad de las inferencias y simplemente se utilizaron 300 estratos, uno por cada distrito electoral federal.

A partir de esta información y de los resultados obtenidos en los conteos rápidos correspondientes, para el ejercicio de 2024, el Comité acordó utilizar una estratificación con 300 estratos, uno por cada distrito electoral federal.

Determinación del tamaño de muestra.

De la misma forma en que los tres conteos rápidos previos para la elección presidencial han empleado la misma estratificación, con los distritos electorales federales como estratos, también han empleado el mismo tamaño de muestra básico: 7,500 casillas. Este tamaño se denomina básico porque es objeto de un incremento que tiene por objeto contar con una sobre muestra en las entidades federativas más occidentales del país donde el huso horario sea diferente al centro del país.

En este caso, el Comité acordó que, si bien 7,500 era una opción razonable, era conveniente conducir estudios de simulación para evaluar otros tamaños de muestra alternativos. Así, una vez establecida la estratificación, el procedimiento para seleccionar el tamaño de muestra óptimo hizo uso de los resultados de las elecciones Presidenciales de 2012 y 2018. A partir de las bases de datos que contienen los resultados de las respectivas votaciones en cada una de las casillas instaladas, y con el conocimiento de los resultados finales, se procedió, para distintos tamaños de muestra, a evaluar la precisión de las estimaciones, sujeta a un nivel de confiabilidad de 0.95. El mecanismo de evaluación de los distintos tamaños de muestra se desarrolla como se describe a continuación para cada base de datos (2012 y 2018).

Dado un tamaño total de muestra n , que se distribuye entre los estratos en forma proporcional al tamaño de cada estrato, se procede de la siguiente manera.

1. Se selecciona una muestra con ese tamaño, de acuerdo al esquema de estratificación.
2. Con la información de esa muestra, se utiliza el procedimiento descrito en Mendoza y Nieto (2016) para calcular la distribución final de la proporción de votos a favor de cada uno de los candidatos a gobernador.

3. Para cada candidato, por ejemplo el i -ésimo candidato, se calcula la estimación puntual de la proporción de votos a su favor \hat{p}_i y, con el valor que finalmente obtuvo en esa elección, p_i , se determina el correspondiente error de estimación $|p_i - \hat{p}_i|$.
4. También para el i -ésimo candidato, se calcula el intervalo de estimación de p_i con un nivel de confiabilidad del 95% y se registra la longitud de ese intervalo.
5. Considerando los resultados de todos los candidatos, con esa muestra se reporta M , el máximo de los errores de estimación y L , el valor máximo de las longitudes de los intervalos de estimación.

Los pasos 1 a 5 se repiten un número K de veces (en este caso, $K = 1000$) a fin de estudiar el comportamiento de los indicadores M y L , además de monitorear la cobertura de los intervalos de estimación.

Con los resultados de este estudio es posible estudiar el efecto de los distintos tamaños de muestra para seleccionar uno que conduzca a un nivel de precisión que genere la certidumbre requerida. Los tamaños de muestra considerados van de 6,000 a 8,000 casillas, con incrementos de 500 unidades.

Estos tamaños equivalían, en 2012, a un porcentaje entre 4.2% y 5.6%, de las casillas totales. En el caso de 2018, equivalían a un porcentaje entre 3.8% y 5.1% de las casillas totales. Finalmente, tomando en cuenta el total de casillas preliminar para 2024, equivalen a un porcentaje que va de 3.5% a 4.7%.

Tanto en la elección Presidencial de 2012 como en la de 2018 se presentaron 4 candidatos así que el ejercicio determina el máximo de los cuatro errores de estimación y el valor máximo de las cuatro longitudes de los intervalos de estimación (de confiabilidad 95%) para cada muestra en el estudio y se analiza el comportamiento de estos indicadores en las 1000 muestras respectivas.

Figura 1 Distribución del error máximo de estimación. Datos 2012.

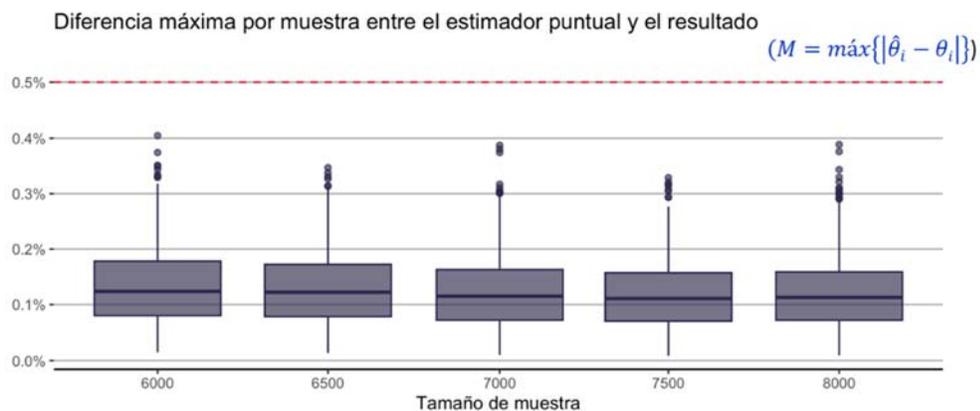
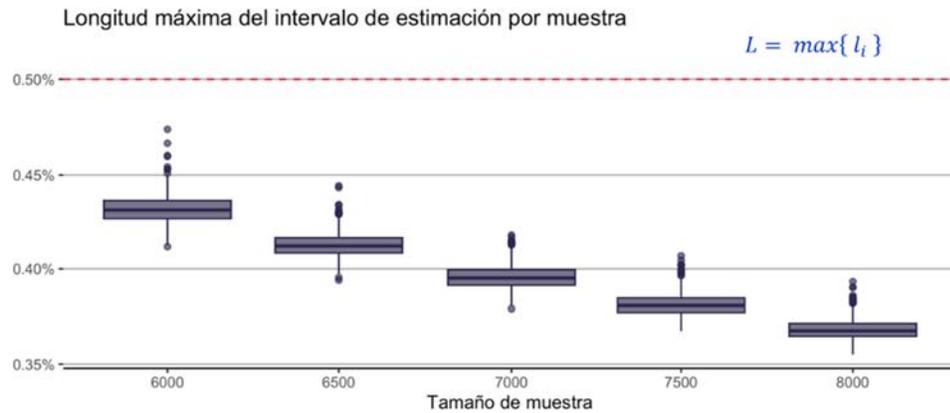


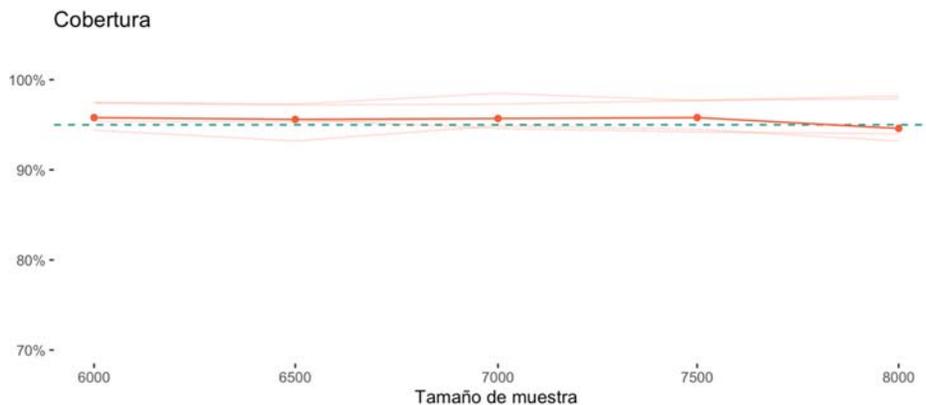
Figura 2 Distribución de la longitud máxima del intervalo de estimación. Datos 2012.



En la Figura 1 se puede observar que en el rango de tamaños de muestra considerados no se presentan diferencias especialmente relevantes en lo que se refiere a la distribución del error máximo de estimación.

Por otra parte, de la Figura 2, se tiene que en lo que se refiere a la longitud máxima del intervalo de estimación se aprecia una disminución a medida que aumenta el tamaño de la muestra. Particularmente, es a partir de 7,500 casillas que prácticamente la totalidad de las muestras reportan una longitud máxima por debajo de 0.4%.

Figura 3 Cobertura de los intervalos de estimación. Datos 2012.



En la Figura 3 se exhibe la cobertura que ofrecen los intervalos de estimación de los porcentajes poblacionales de votación y, como puede observarse, para los cuatro candidatos esta oscila alrededor de la cobertura nominal de 95%.

El mismo análisis se aplicó a la información proveniente de la elección de 2018. En el caso del error máximo de estimación ahora se observa los errores son

ligeramente mayores comparados con 2012 y que la distribución presenta pequeñas variaciones a la baja cuando el tamaño de la muestra aumenta. Este efecto se aprecia con mayor claridad con los tamaños 7,500 y 8,000.

En lo que se refiere a la longitud máxima de los intervalos de estimación con los datos de 2018, se tienen, en todos los casos, longitudes máximas más grandes. Por ejemplo, mientras en 2012, con un tamaño de 7,500 casillas, la longitud máxima no rebasaba el 0.4%, en 2018 con el mismo número de casillas, la mayoría de las muestras reportan longitudes máximas superiores a 0.45%.

Figura 4 Distribución del error máximo de estimación. Datos 2018.
Diferencia máxima por muestra entre el estimador puntual y el resultado

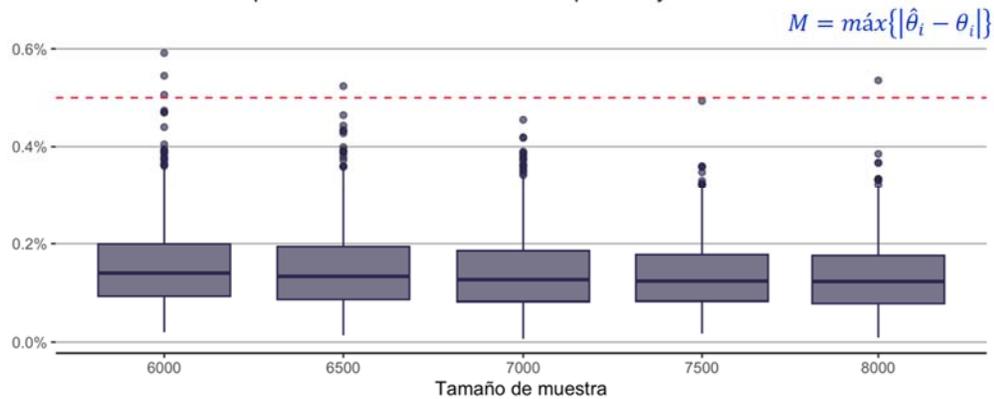
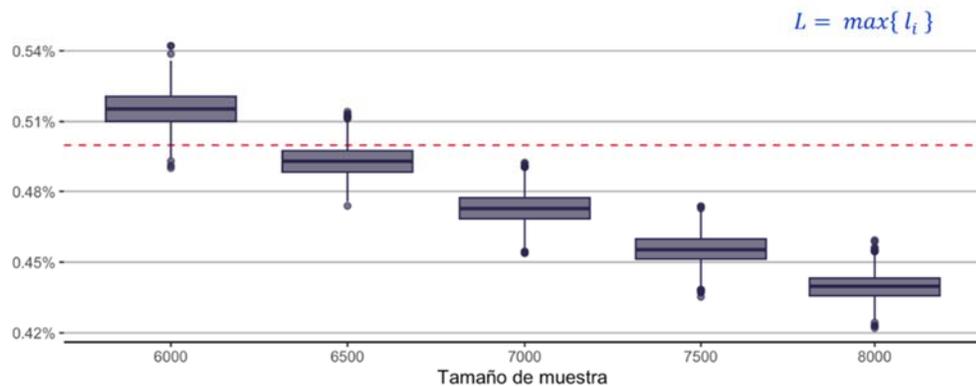


Figura 5 Distribución de la longitud máxima del intervalo de estimación. Datos 2018.

Longitud máxima del intervalo de estimación por muestra



En cuanto a la cobertura, en 2018 este indicador para los cuatro candidatos se encuentra alrededor del valor nominal 95% si bien en este caso presenta una mayor variabilidad.

Figura 6 Cobertura de los intervalos de estimación. Datos 2018.



En resumen,

- a) El comportamiento del error máximo de estimación es relativamente estable tanto en 2012 como en 2018 cuando se consideran tamaños de muestra entre 6,000 y 9,000 casillas.
- b) En 2018, se aprecia una ligera disminución en la distribución de ese indicador, cuando el tamaño de la muestra aumenta, especialmente a partir de 7,500 casillas.
- c) En términos de la longitud máxima de los intervalos, existe una tendencia a la disminución en la medida en que el tamaño de la muestra se incrementa y eso ocurre tanto en 2012 como en 2018.
- d) Las magnitudes de estas longitudes son apreciablemente mayores con los datos de la elección de 2018 en comparación con lo ocurrido en 2012.

Consideraciones complementarias.

Antes de formular conclusiones, es necesario recordar que las estimaciones en este estudio se produjeron con sólo uno de los varios métodos de estimación que se utilizarán en el conteo rápido de 2024.

Con la finalidad de ilustrar, así sea parcialmente, lo que se puede esperar en 2024, se presentan los resultados del conteo rápido que se dieron a conocer en 2018 para la elección Presidencial.

Tabla 2 Intervalos de estimación finales. Conteo Rápido Elección Presidencial 2018.

Candidato	Lim inf	Lim sup	Longitud
RAC	22.1	22.8	0.7
JAMK	15.7	16.3	0.6
AMLO	53.0	53.8	0.8
JHRC	5.3	5.5	0.2

La característica más notable de estos intervalos es que las longitudes superan claramente los niveles que en este estudio se reportan para la longitud máxima. Existen, al menos, dos posibles explicaciones para este fenómeno.

Por una parte, en este documento, como ya se señaló, se reportan las longitudes de los intervalos de estimación fruto de sólo uno de los métodos que habrán de utilizarse en el conteo rápido 2024. La noche de la jornada electoral lo que se dará a conocer, como ocurrió en 2018, son los intervalos que se obtienen de combinar los resultados de tres métodos de estimación distintos y este mecanismo de consolidación puede producir intervalos de mayor longitud.

El otro aspecto que posiblemente tenga un mayor impacto en las longitudes es el hecho de que en este ejercicio se utilizó la muestra completa para producir las estimaciones mientras que en 2018, los intervalos que parecen en la Tabla 1 se produjeron con la información disponible en la última remesa que se recibió y que contenía sólo el 67.5% de la muestra originalmente planeada, lo que equivale aproximadamente a 5,000 casillas.

Conclusiones.

En primer lugar, es importante insistir en que los resultados en este estudio suponen que la muestra, sea cual sea su tamaño, se recibe completa y las estimaciones que reporta proceden de sólo uno de los métodos que habrán de utilizarse en el conteo rápido de 2024.

Tomando en cuenta esas condiciones, los resultados apuntan en el sentido de que un tamaño de muestra de 7,500 casillas es razonable y que, en principio, otras posibilidades de tamaños menores también podrían ser razonables.

Sin embargo, insistiendo en que la noche de la elección los intervalos por reportarse serán combinados, posiblemente con una longitud mayor y ante la expectativa de que la muestra no se reciba completa, la propuesta es asumir una posición conservadora para este ejercicio que es, con la importancia que todos los conteos de esa noche tienen, el más relevante y de mayor impacto. Así se

propone mantener el tamaño de 7,500 que se ha empleado en conteos rápidos previos.

Finalmente, el tamaño de la muestra final se determinará como la suma de estas 7,500 casillas más una cantidad relativamente menor de casillas adicionales que, con naturaleza de sobre muestra, se añadirán en los estados de Baja California Sur y Sinaloa, entidades que por el huso horario que les aplica, iniciarán y concluirán la jornada electoral con una hora de retraso respecto al centro del país.

Referencias.

Mendoza, M. y Nieto-Barajas, L.E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. *Electoral Studies* 43, 124-132.

7.1.2. Procedimiento de estimación

7.1.2.1. Equipo 1

Dra. Lizbeth Lizbeth Naranjo Albarrán, Dra. Ruth Selene Fuentes García y Act. Miguel Ángel Chong Rodríguez

Método de Estimación.

Se utiliza el método normal ponderado por la lista nominal, propuesto por Mendoza y Nieto-Barajas (2016).

Sea θ_{ij} la proporción de personas en la lista nominal del estrato i cuya preferencia es por el candidato j , donde $i = 1, \dots, N$ y $j = 1, \dots, J$, N es el número de estratos, J es el número total de candidatos, incluyendo los candidatos no registrados, los nulos, y las abstenciones. Sea n_i la lista nominal de votantes en el estrato i , para $i = 1, \dots, N$, con $n = \sum_{i=1}^N n_i$ el tamaño de la lista nominal total.

La proporción de votos, relativa a lista nominal, a favor del candidato j está dada por:

$$\theta_j = \sum_{i=1}^N \frac{n_i}{n} \theta_{ij}$$

La proporción efectiva de votos a favor del candidato j se puede calcular como:

$$\lambda_j = \frac{\theta_j}{\sum_{l=1}^{J-1} \theta_l}$$

Sea X_{ij}^k el número de personas a favor del candidato j , en la casilla k del estrato i , donde $k = 1, \dots, K_i$, y K_i es el número de casillas del i -ésimo estrato. El modelo normal ponderado por lista nominal supone que X_{ij}^k sigue una distribución normal, en específico:

$$X_{ij}^k | \theta_{ij}, \tau_{ij} \sim N \left(n_i^k \theta_{ij}, \frac{\tau_{ij}}{n_i^k} \right)$$

donde n_i^k es el tamaño de la lista nominal de la casilla k en el estrato i , y τ_{ij} es un parámetro de precisión para $k = 1, \dots, K_i$, $i = 1, \dots, N$ y $j = 1, \dots, J$.

Las distribuciones iniciales para los parámetros θ_{ij} y τ_{ij} se consideran independientes y no informativas, es decir:

$$p(\theta_{ij}, \tau_{ij}) = U(\theta_{ij} | 0, 1) \text{Ga}(\tau_{ij} | 0.5, 0.05)$$

Usando el Teorema de Bayes se calcula la distribución posterior de θ_{ij} y τ_{ij} (Berdardo y Smith, 2000). Los estimadores puntuales y por intervalo para θ_{ij} se aproximarán mediante simulaciones de la distribución posterior por métodos de Monte Carlo (Chen et al., 2000).

Referencias

- Bernardo, J.M. y Smith, A.F.M. (2000). Bayesian Theory. Wiley: Chichester.
- Chen, M.-H., Shao, Q.-M. e Ibrahim, J.G. (2000). Monte Carlo methods in Bayesian computation. Springer, New York.
- Mendoza, M. y Nieto-Barajas, L.E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. Electoral Studies 43, 124-132.

Mtra. Patricia Isabel Romero Mares y Lic. Luis Enrique Reyes Romero.

El diseño muestral será estratificado con L estratos, que son los Distritos Federales, y una selección de casillas con muestreo aleatorio simple en cada uno de ellos y distribución proporcional de la muestra a los estratos.

Suponga un estado fijo o el país completo. La estimación de la proporción de ciudadanos/as que voten por el candidato p en el estado (país) se calculará utilizando el estimador de razón combinado para un esquema de muestreo estratificado con selección de una muestra aleatoria simple y sin reemplazo en cada estrato, y con asignación proporcional de la muestra a los L estratos:

$$\hat{R}_p = \frac{\hat{Y}_p}{\hat{X}} = \frac{\sum_{h=1}^L \hat{Y}_{hp}}{\sum_{h=1}^L \hat{X}_h} = \frac{\sum_{h=1}^L N_h * \bar{y}_{hp}}{\sum_{h=1}^L N_h * \bar{x}_h}$$

Donde:

\hat{R}_p : estimador de la proporción de votos a favor del candidato p , $p=1, \dots, \text{número de candidatos}$

\hat{Y}_p : estimador del total de votos a favor del candidato p , $p=1, \dots, \text{número de candidatos}$

\hat{X} : estimador del total de votos emitidos en el estado (o nacional)

\hat{Y}_{hp} : estimador del total de votos a favor del candidato p en el estrato h

\hat{X}_h : estimador del total de votos emitidos en el estrato h

\bar{y}_{hp} : estimador del promedio de votos por casilla a favor del candidato p , en el estrato h

\bar{x}_h : estimador del promedio de votos emitidos por casilla en el estrato h

N_h : total de casillas en el estrato h

El estimador de la varianza del estimador de la proporción de votos a favor del candidato p es el siguiente:

$$\hat{V}(\hat{R}_p) = \frac{1}{\hat{X}^2} * \sum_{h=1}^L N_h^2 * \left(1 - \frac{n_h}{N_h}\right) * \frac{s_h^2}{n_h}$$

Donde:

$$s_h^2 = \frac{\sum_{i=1}^{n_h} [(y_{hi}^p - \bar{y}_h^p) - \hat{R}_p(x_{hi} - \bar{x}_h)]^2}{n_h - 1}$$

y_{hi}^p : total de votos a favor del candidato p en la casilla i del estrato h

\bar{y}_h^p : promedio de votos a favor del candidato p del estrato h

x_{hi} : total de votos emitidos en la casilla i del estrato h

\bar{x}_h : promedio de votos emitidos por casilla en el estrato h

n_h : número de casillas en muestra del estrato h

La precisión observada para la estimación de la proporción de votos al candidato p se calcula como:

$$precisión = z_{(1-\frac{\alpha}{2})} \sqrt{\hat{V}(\hat{R}_p)}$$

Considerando una confianza del 95% y el cuantil de la Normal estándar $z=1.96$ correspondiente a esta confianza, el intervalo de confianza para la proporción de votos a favor del candidato p se obtiene de la siguiente manera:

$$(\hat{R}_p - precisión, \hat{R}_p + precisión)$$

De igual manera, la participación P , se estimará como un estimador de razón combinado, con la siguiente expresión:

$$\hat{P} = \frac{\hat{X}}{\hat{LN}}$$

Con estimador de su varianza e intervalo de confianza, equivalentes a lo presentado arriba.

Referencias:

- Kish, L. (1970). "Survey Sampling". John Wiley and Sons. NY, USA.
- Raj, Des. (1968). "Sampling Theory". McGRAW-Hill. NY, USA.
- Särndal, Carl Erik, et al. (1992). "Model Assisted Survey Sampling". Springer-Verlag. NY, USA.

Dr. Raúl Rueda Díaz del Campo y Mtro. Javier Santibáñez Cortés

El objetivo de cualquier Conteo Rápido es estimar la proporción de votos de cada candidato, y la participación ciudadana, basándose en una muestra de casillas. El diseño utilizado consiste en dividir a la población objetivo en k estratos y seleccionar una muestra de tamaño m_i en cada uno de ellos con un muestreo aleatorio simple.

En este ejercicio se estima el total de votos para cada candidato a nivel estatal (o nacional) y con estos valores es inmediato estimar las proporciones deseadas.

El total poblacional está dado por

$$T_p = \sum_{i=1}^k \sum_{j=1}^{M_i} X_{ij}$$

en donde $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijp})$ son los votos emitidos en la casilla j del estrato i p mediciones

M_i , es el número total de casillas en el i -ésimo estrato y k es el número de estratos.

No es difícil demostrar que este total puede expresarse como

$$T_p = \sum_{i=1}^k (m_i \bar{x}_i + (M_i - m_i) \bar{\mathbf{X}}_i^*)$$

con \bar{x}_i la media muestral observada en el estrato i , m_i el tamaño de muestra observado y $\bar{\mathbf{X}}_i^*$ la media **no** observada en el mismo estrato. Así que para estimar T_p basta estimar $\bar{\mathbf{X}}_i^*$ en todos los estratos.

Suponiendo normalidad en las observaciones y una distribución inicial poco informativa, es posible demostrar que la distribución **predictiva** de $\bar{\mathbf{X}}_i^*$ es una distribución Student de dimensión p con $m_i + 0.5$ grados de libertad, parámetro de localización \bar{x}_i y parámetro de escala $\frac{(m_i - \nu + 1)(M_i - m_i)}{M_i} \mathbf{s}_i^{-1}$, donde \mathbf{s}_i es la matriz de varianzas y covarianzas muestral y ν es el parámetro (conocido) de la inicial.

Se genera una muestra de esta distribución y eso permite obtener una muestra de la distribución predictiva de T_p condicional a la información observada. Finalmente, se obtienen muestras de la distribución predictiva de las proporciones deseadas y, usando las listas nominales, de la participación ciudadana. Los

intervalos buscados se construyen con los cuantiles correspondientes de la muestra generada.

7.1.2.2. Equipo 2

Dra. Michelle Anzarut Chacalo y Dr. Luis Felipe González Pérez

Resumen

El conteo rápido consiste en generar estimaciones de los resultados de una contienda electoral, utilizando los votos registrados en una muestra. Su principal desafío radica en que la estimación se efectúa con muestras incompletas, donde la ausencia de datos no es aleatoria. El método de estimación que emplearemos es una regresión binomial negativa con estructura jerárquica, permitiendo que algunos coeficientes varíen según el estrato. Adicionalmente, implementaremos un procedimiento de ajuste por sesgo al estimar con muestras incompletas. El modelo proporciona intervalos de probabilidad con una cobertura aproximada del 95%, incluso frente a ciertos patrones de muestras sesgadas observados en elecciones previas.

Antecedentes

La idea del modelo que aplicaremos provino del modelo de regresión multinivel y postestratificación (ver, por ejemplo, [1]). En este modelo de regresión multinivel y postestratificación, (1) se segmenta a la población en celdas según combinaciones de distintos atributos, (2) se utilizan las muestras para obtener simulaciones posteriores de los parámetros, y (3) se calcula la variable de respuesta para cada celda, ponderando las estimaciones para realizar una estimación a nivel poblacional.

Llevaremos a cabo los pasos (1) y (2), pero en el paso (3), adoptaremos un enfoque basado en modelos (ver [2]), simulando la distribución posterior del número total de votos para cada candidato y para cada una de las casillas electorales del país. Así, emplearemos el modelo para predecir los valores no muestreados de la población. Es importante destacar que esto proporciona una estimación inmediata a nivel poblacional.

Para ejecutar el paso (1), ajustamos una regresión multinivel para cada casilla en función de ciertas covariables. Es decir, para cada candidato, modelamos el número de votos en la casilla i , y_i , con una distribución de probabilidad $p(y|\theta_1, \theta_2)$, donde, para una parte de los parámetros, digamos θ_1 , ajustamos una regresión multinivel.

La ventaja de esta estructura de modelos multinivel radica en que, ante la ausencia de respuesta, atrae los parámetros de un grupo hacia la media del mismo. Como consecuencia, logramos un manejo uniforme de los datos faltantes en las muestras y una mejora en las propiedades de cobertura de intervalos cuando los datos de la muestra presentan sesgos.

La distribución p es binomial negativa. Esta distribución se adapta adecuadamente a los datos y su robustez contribuye a la del modelo, permitiéndole ajustarse eficazmente tanto a candidatos con una alta como con una baja cantidad de votos. Además, ofrece una capa adicional de protección ante posibles errores en la base de datos.

Modelo

A continuación, definimos el modelo que utilizaremos, que se obtuvo después de varias iteraciones del flujo de trabajo bayesiano en [4].

Denotamos con BN a la distribución binomial negativa con una parametrización directamente usando un parámetro de media, $\mu_i \in \mathbb{R}^+$, y un parámetro $\phi_i \in \mathbb{R}^+$. Para $y \in \mathbb{N}$,

$$\text{BN}(y|\mu_i, \phi_i) \propto \left(\frac{\mu_i}{\mu_i + \phi_i} \right)^y.$$

Nuestro primer objetivo es predecir el número total de votos a favor de cada candidato y el número total de votos nulos en cada casilla. Teniendo esas predicciones podemos deducir cualquier cantidad que sea de nuestro interés, por ejemplo, la votación total emitida, la participación, intervalos de confianza para la proporción de votos obtenida por cada candidato o la conformación de la Cámara de Diputados.

Por simplicidad de notación consideraremos a los votos nulos como un candidato extra. Sea p el número total de candidatos, modelamos el número de votos obtenidos para el candidato $j \in \{1, \dots, p\}$ y para la casilla i como

$$y_{i,j} \sim \text{NB} \left(\mu_{i,j} := n_i w_i \theta_{i,j}, \phi_{i,j} := \frac{n_i w_i \theta_{i,j}}{\kappa_{s(i),j}} \right).$$

Aquí $\theta_{i,j}$ representa la proporción de personas que votaron por j del total de votos en la casilla i , w_i representa la proporción de personas en la lista nominal que votaron en la casilla i y $\kappa_{s(i),j} > 0$ controla la sobredispersión en relación a la media $\mu_{i,j}$.

El vector $\theta_i = (\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,p})$ se define como

$$\theta_i = \text{softmax}(\beta_{s(i)}^{0,1} + \beta_1^T x_i, \dots, \beta_{s(i)}^{0,p} + \beta_p^T x_i),$$

donde x_i es la i -ésima fila de X , una matriz de covariables escaladas, y

$$\text{softmax}(z_1, \dots, z_p)_j = \frac{\exp(z_j)}{\sum_{i=1}^p \exp(z_i)}.$$

Los coeficientes β_1, \dots, β_p tienen una distribución inicial gaussiana centrada en cero y con una desviación estándar de 0.25. Sin embargo, para $k = 1, \dots, p$,

$$\beta_{s(i)}^{0,k} \sim \mathcal{N}(\beta_{0,k}, \sigma_k),$$

añadiendo una estructura jerárquica dependiendo del estrato. Los detalles de las distribuciones iniciales pueden consultarse en [8] en donde, adicionalmente, se prueba la consistencia de la elección de todas las distribuciones iniciales.

Estimación

La metodología general de la estimación se describe a continuación. Inicialmente, para cada casilla que no forma parte de la muestra, generamos simulaciones de votos basadas en el modelo. Posteriormente, juntamos los valores observados en las casillas con los valores simulados, obteniendo así los votos totales de todas las casillas. A partir de estos votos, es directo calcular los resúmenes que sean necesarios para cada elección.

La estimación se hace con el software Stan, llamado desde R utilizando el paquete CmdStanR [6]. El código en Stan es legible, lo que hace que el trabajo sea más transparente. Stan también permite una serie de diagnósticos, lo que ayudó a obtener una inferencia confiable más rápida. La implementación se puede consultar y reproducir en su totalidad en el paquete de R quickcountmx [7].

Para asegurarnos de abarcar una diversidad de posibles resultados, realizamos una calibración bayesiana con los datos de las elecciones nacionales de 2018. La calibración bayesiana se enfoca en verificar que el modelo produzca inferencias adecuadamente calibradas desde un enfoque frecuentista; es decir, busca asegurar que el modelo entregue intervalos de credibilidad posterior con una cobertura cercana al 95% en muestreos repetidos. La calibración evidencia que, ante una proporción significativa de casillas no reportadas, el modelo por sí solo no es suficiente para ajustar los sesgos presentes en las muestras censuradas.

El problema del sesgo

El sistema de información comienza a operar a las 6 p.m. y, cada 5 minutos, recibe nueva información de la muestra. De esta manera, el sistema genera una secuencia de muestras parciales acumulativas, que son analizadas mediante métodos de estimación para monitorear la tendencia de los resultados. Los resultados del conteo rápido se publican utilizando una de estas muestras parciales, debido a que un porcentaje significativo de la muestra completa tarda en ser reportado.

La ausencia de datos en las muestras parciales no es aleatoria; factores como si la casilla se encuentra en una zona urbana o rural, las condiciones climáticas o la lista nominal, influyen en el momento en que una casilla llega a la muestra.

Realizamos un análisis exhaustivo de sesgo en elecciones anteriores y nos enfrentamos al desafío de mitigar este sesgo en todas las etapas de nuestro flujo de trabajo. Para llevar a cabo este análisis, necesitábamos un procedimiento para generar muestras parciales que se asemeje al verdadero proceso de generación de datos. Así, ajustamos modelos de supervivencia para los tiempos de llegada de las casillas en las elecciones de 2018. Los detalles se pueden consultar en [8].

Usamos el modelo de supervivencia para obtener una medida del sesgo en diferentes horas o con diferentes porcentajes de muestra obtenidos.

El ajuste que haremos al modelo dependerá del porcentaje de muestra observado, al que denotamos con p_{obs} . En primer lugar, estimamos la distribución de probabilidad del número total de votos de cada candidato en cualquier casilla con el modelo descrito. Después, a la hora de simular, en lugar de utilizar el vector θ_i , usamos una perturbación, dada por

$$\theta'_i = \text{softmax}(\theta_{i,1} + \eta_{i,1}, \dots, \theta_{i,p} + \eta_{i,p}),$$

donde η_1, \dots, η_p son variables aleatorias independientes e idénticamente distribuidas $N(0, 0.1(1 - p_{obs}))$.

Notemos que el tamaño de los intervalos de probabilidad producidos crece a medida que crece $\eta_{i,k}$ y, conforme nos acercamos a la muestra completa, $\eta_{i,k}$ se acerca a 0 para cualquier candidato k . El modelo, con el ajuste por sesgo, provee coberturas cercanas al 95% tanto con muestras completas como con muestras censuradas.

Método de respaldo

Como método de respaldo, implementaremos el estimador de razón combinado.

Sea p_j la proporción de votos a favor del candidato j , para estimar p_j utilizaremos \hat{p}_j , definido como:

$$\hat{p}_j = \frac{\sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hij}}{\sum_{j=1}^p \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hij}},$$

donde, para cada estrato h , n_h es el tamaño de la muestra, N_h es el número total de casillas, L es el número total de estratos, p es el número total de candidatos (incluyendo nulos y no registrados) y x_{hij} es el total de votos a favor del candidato j en la casilla i .

El error estándar de este estimador se calcula utilizando bootstrap no paramétrico (ver, por ejemplo, [9]).

Referencias

- [1] Park, D.K., Gelman, A., Bafumi, J.: Bayesian multilevel estimation with poststratification: state-level estimates from national polls. *Political Analysis* 12(4), 375–385 (2004)
- [2] Little, R.J.: Calibrated Bayes, an alternative inferential paradigm for official statistics. *Journal of official statistics* 28(3), 309 (2012)
- [3] Anzarut, M., González, L.F., Ortiz, M.T.: A heavy-tailed multilevel mixture model for the quick count in the mexican elections of 2018. In: *National Statistics Forum (FNE) and Latin-American Congress of Statistical Societies (CLATSE)*, pp. 1–13. Springer (2018)
- [4] Schad, D., Betancourt, M., Vasishth, S.: Toward a principled bayesian workflow in cognitive science. *Psychological Methods* 26 (2020). DOI 10.1037/met0000275
- [5] Barnard, John, Robert McCulloch, and Xiao-Li Meng. 2000.: Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage. *Statistica Sinica*, 1281–1311.
- [6] Stan Development Team: *CmdStanR: the R interface to CmdStan* (2021). URL <https://mc-stan.org/r-packages/>
- [7] Anzarut, M., González, L.F., Ortiz, M.T.: quickcountmx: Fuctions for the mexican quick-counts 2022 (2022). URL <https://github.com/cotecora-team-2/quickcountmx>

[8] Anzarut, M., González, L.F., Ortiz, M.T.: Bayesian Hierarchical Multinomial Modeling of the 2021 Mexican Election Outcomes with Censored Samples. In: National Statistics Forum (FNE). Springer (2022)

[9] Ortiz, M.T. Bootstrap no paramétrico, Estadística Computacional Capítulo 5, https://tereom.github.io/est_computacional/05-Bootstrap.html

Dr. Manuel Mendoza Ramírez y Dr. Felipe Javier Medina Aguayo

Introducción.

La estadística Bayesiana es una alternativa a la estadística frecuentista para la producción de inferencias sobre los parámetros desconocidos de un modelo. Aborda los problemas de inferencia como problemas de decisión en ambiente de incertidumbre, y recurre a una teoría de la decisión con fundamento axiomático para proveer las soluciones. (Bernardo y Smith, 2000).

Tanto bajo el enfoque frecuentista como bajo el Bayesiano, el fenómeno de interés es sujeto de observación a través de mecanismos que garantizan que los datos que se recolectan forman muestras aleatorias. En consecuencia, el comportamiento, la variabilidad, de estas muestras se describe a través de modelos de probabilidad.

En este contexto, las características del fenómeno que interesa conocer se identifican con el valor de uno o más parámetros del modelo y es en esos términos que los procesos de inferencia se concentran en aumentar y precisar el conocimiento sobre los parámetros.

Desde la perspectiva Bayesiana, la incertidumbre sobre los parámetros también se describe a través de modelos de probabilidad. En este caso no se describe variabilidad alguna sino el estado de conocimiento que, en un momento determinado, tiene el científico sobre el fenómeno bajo estudio. Con diferentes modelos se puede reflejar desde un conocimiento relativamente preciso hasta un desconocimiento casi completo sobre el valor del parámetro. De esta forma, bajo este paradigma, tanto las variables aleatorias observables (la muestra) como los parámetros fijos (pero desconocidos) son descritos mediante modelos de probabilidad.

La cuantificación de la incertidumbre sobre un parámetro puede producirse antes de contar con una muestra. El conocimiento teórico, histórico o incluso empírico, previo a la recolección de la muestra da lugar a esta valoración que se refleja en un modelo o distribución inicial o a priori $f(\theta)$.

Si es posible obtener la información muestral $\underline{x} = (X_1, X_2, \dots, X_n)$, proveniente del modelo $f(x|\theta)$, el estado de conocimiento inicial es susceptible de actualizarse para combinar la información inicial con la que provee la muestra.

Esta actualización se lleva a cabo mediante el mecanismo que establece el teorema de Bayes. El resultado es el modelo final o posterior $f(\theta|\underline{x})$, que combina las dos fuentes de información a través de la relación

$$f(\theta|\underline{x}) = f(\underline{x}|\theta)f(\theta)/f(\underline{x}).$$

A partir del modelo final se producen las inferencias de interés sobre el parámetro θ , de manera que es indispensable establecer las características de $f(\theta|\underline{x})$. Si bien la fórmula de Bayes es una expresión matemática simple, en la práctica, la caracterización analítica de la distribución final se puede complicar debido, por ejemplo, a las dificultades en el cálculo de la constante de normalización $f(\underline{x})$. Gracias a los avances computacionales recientes y en especial a los algoritmos de simulación Monte Carlo vía cadenas de Markov (MCMC), es posible conocer las características de cualquier distribución final mediante métodos de simulación sin necesidad de calcular la constante de normalización (Chen et al. 2000).

Definiciones específicas.

El objetivo del conteo rápido es producir estimaciones confiables de la proporción efectiva de votos a favor de cada uno de los candidatos. Esta proporción se calcula como el cociente entre el número de votos a favor de un candidato y el número total de votos emitidos por todos los candidatos incluyendo los votos nulos y los candidatos no registrados.

Respecto a la información muestral con la que se producen los resultados del conteo rápido, se adopta el supuesto de que corresponde con un diseño estratificado de casillas. En esos términos, sea N el número de estratos, cada uno con una lista nominal de tamaño n_i para $i = 1, \dots, N$. Además, se define θ_{ij} como la proporción de personas en la lista nominal del estrato i cuya preferencia es por el candidato j , donde $j = 1, \dots, J$ con J el número total de candidatos incluyendo los candidatos no registrados, los nulos y el abstencionismo. En este caso, la proporción de personas en la lista nominal de todo el estado cuya preferencia es por el candidato j está dada por

$$\theta_j = \sum_{i=1}^N \frac{n_i}{n} \theta_{ij}$$

donde n es el tamaño de la lista nominal del estado, si se trata de una elección local o la lista nominal del país si se trata de una elección federal. Finalmente, la proporción efectiva de votos a favor del candidato j se puede calcular como

$$\lambda_j = \frac{\theta_j}{\sum_{l=1}^{J-1} \theta_l}$$

El Modelo y el Mecanismo de Inferencia

Tanto el modelo que se utiliza como el mecanismo para producir las inferencias se presentan en detalle en Mendoza y Nieto-Barajas (2016). En términos generales, se pueden describir como sigue. Del total de casillas en el estado (o país, según sea el caso) se seleccionan, mediante un mecanismo aleatorio, c de ellas (el tamaño total de la muestra), de manera que c_i de ellas provienen del i -ésimo estrato $i = 1, \dots, N$. De esta forma, c_i es el tamaño de la muestra en el estrato i .

Sea ahora, X_{ij}^k el número de personas a favor del candidato j , en la casilla k del estrato i . En esta formulación, $k = 1, \dots, K_i$ con K_i el número total de casillas en el estrato i .

El modelo que se emplea asume que

$$X_{ij}^k | \theta_{ij}, \tau_{ij} \sim N \left(n_i^k \theta_{ij}, \frac{\tau_{ij}}{n_i^k} \right)$$

donde n_i^k es el tamaño de la lista nominal de la casilla k en el estrato i , y τ_{ij} es un parámetro de precisión para $k = 1, \dots, c_i$, $i = 1, \dots, N$ y $j = 1, \dots, J$.

El modelo se complementa con la elección de las distribuciones iniciales para los parámetros θ_{ij} y τ_{ij} . En este caso, se utilizan iniciales que reflejan poca información a priori (distribuciones iniciales no informativas). El objetivo es obtener la distribución final para cada uno de los parámetros λ_j , con $j = 1, \dots, J - 1$ de manera que las inferencias que interesan se producen a partir de ese modelo de probabilidad. Como en otros casos, en este modelo esa distribución final es analíticamente compleja así que se recurre a métodos de simulación para obtener los resultados.

Referencias

- Bernardo, J.M. y Smith, A.F.M. (2000). Bayesian Theory. Wiley: Chichester.
- Chen, M.-H., Shao, Q.-M. e Ibrahim, J.G. (2000). Monte Carlo methods in Bayesian computation. Springer, New York.
- Mendoza, M. y Nieto-Barajas, L.E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. Electoral Studies 43, 124-132.

7.1.2.3. Equipo 3

Dr. Gabriel Núñez Antonio y Mtro. Emiliano Geneyro Squarzon

Procedimiento de estimación

El reconocimiento explícito de la incertidumbre es crucial en las ciencias estadísticas. En muchas situaciones simples, la incertidumbre de un estimador puede visualizarse a través de cálculos analíticos basados en algún modelo probabilístico, adecuado. Sin embargo, en problemas más complejos esta tarea puede ser complicada, y los resultados llegan a ser potencialmente incorrectos si los supuestos básicos no se cumplen o si se realizan aproximaciones con la finalidad de simplificar los cálculos. Los avances computacionales, permiten hoy en día, el empleo de métodos denominados de remuestreo que ayudan a soslayar los problemas anteriores, a la vez que permiten obtener estimaciones de errores estándar, intervalos de confianza, y otras medidas de incertidumbre.

La idea principal de los métodos bootstrap es el remuestreo a partir de los datos originales, ya sea en forma directa o vía un modelo ajustado; con la finalidad de obtener muestras replicadas a partir de las cuales se pueda evaluar la variabilidad de las cantidades de interés. Los métodos bootstrap también se pueden aplicar en problemas simples para verificar las características de las medidas de incertidumbre, para relajar supuestos, o para dar rápidas soluciones aproximadas. Un ejemplo de lo anterior es el remuestreo aleatorio para estimar la distribución permutacional de alguna estadística de prueba no paramétrica.

Es verdad que en muchas aplicaciones se puede confiar ampliamente en un modelo paramétrico particular y en el correspondiente análisis clásico basado en dicho modelo. Aun así, puede ser de utilidad investigar qué tanto se puede inferir sin asumir los supuestos de un modelo paramétrico particular. Esto es la esencia de la *robustez* del análisis estadístico realizado. El *bootstrap no-paramétrico* permite hacer esto. De acuerdo con lo anterior los métodos bootstrap se pueden aplicar tanto en los casos en los que se cuente con un modelo probabilístico bien definido para los datos, como en los casos en los que no se disponga de dicho modelo.

A pesar de su alcance y utilidad, el remuestreo debe ser aplicado en forma cuidadosa. A menos que las ideas básicas queden completamente claras, es muy fácil producir una solución al problema equivocado, o una mala solución al problema correcto. Los métodos de remuestreo como el bootstrap pretenden, de hecho lo hacen, evitar el empleo de procedimientos basados en supuestos distribucionales cuestionables.

La propuesta de estimación en el Conteo Rápido para la Presidencia de la República, en las elecciones del 2 de junio de 2024, consideran el diseño de

muestral definido en este documento; con selección de casillas aleatoria sin reemplazo dentro de cada estrato y con tamaño de muestra proporcional al número de casillas en cada estrato.

Tipo de estimadores puntuales que se emplearán

Sea h el estrato a considerar, $h=1, \dots, k$, con N_h elementos en cada estrato. En cada estrato se toma una muestra aleatoria simple de tamaño n_h . Para estimar el porcentaje de votos de cada candidato p_j , $j=1, \dots, b$, los estimadores puntuales propuestos para la proporción de votos están dados por:

$$\hat{p}_j = \frac{\hat{T}_j}{\hat{T}} \quad \forall j = 1, \dots, b,$$

donde \hat{T}_j es un estimador estratificado del total de votos para cada candidato j , $j=1, \dots, b$, y \hat{T} es un estimador estratificado del total de votos obtenidos. Específicamente,

$$\hat{T}_j = \sum_h \frac{N_h}{n_h} Y_{j,h}, \quad \hat{T} = \sum_h \frac{N_h}{n_h} Y_h,$$

donde $Y_{j,h}$ es el total de votos del candidato j en el estrato h y Y_h es el total de votos en el estrato h .

Procedimiento para construir intervalos de confianza

Para obtener intervalos de confianza, con una precisión adecuada, para la proporción de votos para cada candidato se pueden seguir diferentes estrategias. Por ejemplo se puede usar una aproximación normal asintótica. Si se considera que la aproximación normal asintótica puede no ser adecuada, otra opción es construir vía simulación la distribución muestral del estimador que se esté utilizando y obtener los cuantiles de orden 2.5% y 97.5% o a partir de muestras bootstrap obtener un estimador de la varianza del estimador correspondiente y con ellos construir el intervalo correspondiente.

7.2. Elección de Senadores

7.2.1. Estratificación y tamaño de muestra

7.2.1.1. Dr. Luis Enrique Nieto Barajas⁵

Introducción

Para el diseño muestral de la elección federal del 2024 nos basaremos en los resultados de la elección federal del 2018, ya que esta es la elección más reciente.

Las elecciones al Senado de México de 2018 son parte de las elecciones que se celebraron en México el 1 de julio de 2018 en las que se renovaron todos los cargos federales de elección popular: presidente de la República, diputados federales y senadores; así como diversos cargos de carácter local: gobernadores, congresos locales y alcaldes.

La cámara de senadores se forma por 128 escaños. Dos son obtenidos por mayoría relativa en cada una de las 32 entidades del país, 64 en total; se otorga un escaño más a la primera minoría, 32 en total. Finalmente se asignan 32 escaños por la vía de representación proporcional.

El número de escaños obtenidos oficialmente por cada uno de los partidos políticos en la conformación del senado se presenta en la Figura 7.

Determinación del tamaño de muestra y error de estimación

Como ya se mencionó, el diseño muestral estará basado en un muestreo estratificado por entidad federativa. Para la determinación del tamaño de muestra y de los errores de estimación alcanzados con cada tamaño, nos basaremos en un estudio de simulación.

Se generaron 1,000 muestras aleatorias estratificadas de casillas de tamaños 320, 640, 1280, 1920, 2560, 3200, y 3,840 que divididas entre las 32 entidades federativas equivalen a 10, 20, 40, 60, 80, 100 y 120 casillas por estrato, respectivamente.

Para cada muestra se estimarán los parámetros $\theta_{i,j}$ de manera bayesiana y se obtendrán 1,000 realizaciones de la distribución posterior. Para cada una de estas realizaciones calculamos el número de escaños NE_j . Suponiendo una función de

⁵ Luis E. Nieto B., Carlos S. Pérez P. @ COTECORA

pérdida valor absoluto, estimamos NE_j mediante la mediana posterior a la que denotaremos por \widehat{NE}_j .

Definimos el error de estimación como el número de escaños mal asignados por partido

$NEMA_j = |\widehat{NE}_j - NE_j|$, donde NE_j es el valor oficial de escaños obtenidos por el partido j descrito en la Figura 1. Resumimos el error de estimación en la conformación de toda la cámara mediante dos medidas: el promedio de escaños mal asignados $PEMA = \frac{1}{J-2} \sum_{j=1}^{J-2} NEMA_j$; y el máximo de escaños mal asignados $MEMA = \max_{\{j=1, \dots, J-2\}} NEMA_j$.

Los resultados del ejercicio de simulación se muestran a continuación. La Figura 8 contiene histogramas para los 1,000 valores de los errores de estimación por partido $NEMA_j$.

Notamos que el partido con mayor variabilidad es el PRI con errores de hasta 6 escaños, seguido de PAN y MORENA con errores de hasta 4 y 2 escaños respectivamente. Los demás partidos presentan muy poca variabilidad con errores alrededor de cero. Recordemos que el indicador $NEMA_j$ se basa en el estimador puntual \widehat{NE}_j ; sin embargo, en la práctica no se reportará el estimador puntual sino un intervalo de credibilidad al 95%. Comparando la distribución muestral de los errores por partido en los distintos tamaños de muestra, se aprecia un decremento en la variabilidad de los errores para el tamaño de muestra mayores, no obstante, para los tamaños de muestra mayores a 1920 no se aprecian diferencias considerables entre las distribuciones.

Al promediar todos los errores de estimación de todos los partidos, la Figura 9 muestra las distribuciones muestrales (histogramas de los 1,000 valores) de los errores promedios de la cámara $PEMA$. Se observa un cambio en la dispersión de las distribuciones del error promedio al aumentar el tamaño de muestra. La distribución muestral con 640 casillas es más dispersa que la que se obtiene con 3,840 casillas que está muy concentrada alrededor del 0.

Para una mejor comparación entre las distribuciones, el Cuadro 1 muestra los cuantiles del número promedio de escaños mal asignados $PEMA$ para toda la cámara de senadores con distintos tamaños de muestra. Las medianas $q_{0.5}$ son de 0.8 para un tamaño de muestra de 320, de 0.5 para un tamaño de muestra de 1280, de 0.4 para un tamaño de muestra de 1920 en adelante.

Por otro lado, la Figura 10 muestra las distribuciones muestrales del máximo número de escaños mal asignados $MEMA$ con distintos tamaños de muestra. Al aumentar el tamaño de muestra toda la distribución se mueve hacia la izquierda, lo que indica un menor error. Para 320 casillas la moda se sitúa en 6 escaños, para los tamaños de muestra de 640 y 1,280 la moda es de 5 y 4, respectivamente y

para 1,920 casillas en adelante la moda es de 3 escaños. Los cuantiles de estas distribuciones de muestreo se incluyen en el Cuadro 2. Prácticamente para tamaños de muestra de 1,920 en adelante no hay diferencias.

Relación entre casillas por ARE y tamaño de muestra

En el ejercicio electoral de 2018 se reporta un total de áreas de responsabilidad electoral (ARE) de 38,752 con un total de casillas de 157,069. En este caso, cada casilla es identificada a través de los campos Estado, Distrito Federal y Número de ARE de forma anidada. No obstante, las ARE contienen casillas que responden a una división operativa diferente a la de los distritos, por lo que para una sección puede haber más de una ARE.

Para cada una de las 1,000 muestras simuladas se calcula la distribución de casillas por ARE y se promedian. Denotamos por *PCARE* al número promedio de casillas en muestra que están asignadas a un ARE. Este indicador, para los distintos tamaños de muestra, se reporta en la Figura 11. Observamos que, para todos los tamaños de muestra, al menos 93% de los ARE en muestra tienen una sola casilla, lo que hace muy factible el levantamiento de la información en su totalidad.

Consideraciones Finales

Tomando en cuenta los resultados aquí presentados, los errores de estimación en la conformación de la cámara resumidos en los indicadores *PEMA* y *MEMA*, sus distribuciones de muestreo varían muy poco con muestras de tamaño mayores o iguales a 1920.

Por otro lado, a sugerencia de la Dirección Ejecutiva de Organización Electoral (DEOE), es recomendable anidar las muestras de los conteos rápidos de las elecciones de diputados con la de senadores. En el diseño muestral del conteo rápido para la elección de diputados, se propuso un muestreo aleatorio estratificado de casillas con los estratos definidos por los distritos electorales federales. El tamaño de muestra propuesto por distrito es de 20 casillas.

Para compatibilizar ambos diseños, para las elecciones de diputados y senadores, sugerimos usar el nivel más bajo de estratificación, que en este caso serían los distritos electorales federales y tomar un tamaño de muestra que al agregarse por estado sea de aproximadamente de 60 casillas.

Dado que cada estado tiene un número distinto de distritos, algunos tienen 2 y otros tienen 24 distritos. Entonces los tamaños de muestra propuestos se incluyen en el Cuadro 3. La muestra total resultante es de $n = 1,969$ casillas, que es muy

similar al tamaño de muestra sugerido inicialmente. Los máximos tamaños de error con un 95% de credibilidad serían $q_{0.95} = 0.7$ para *PEMA* y de $q_{0.95} = 4$ para *MEMA*.

Una segunda opción de anidamiento de las muestras para los conteos de diputados y senadores es disminuir el tamaño de muestra a 15 casillas por distrito para el conteo de senadores y repartir de manera proporcional las casillas en los distritos para el conteo de senadores. Esto nos lleva a sugerir tomar 4 casillas por distrito para el conteo de senadores. los tamaños de muestra totales, en este caso, serían de 4,500 casillas para el conteo de diputados y de 1,200 casillas para el conteo de senadores.

Dados los tamaños de muestra tan pequeños que estamos sugiriendo, y por conveniencia del proceso de estimación en la elección de senadores, una tercera opción para las muestras de ambos conteos sería considerar muestras independientes, de 15 casillas por distrito para el conteo de diputados y 40 casillas por estado para el conteo de senadores. En este caso los tamaños de muestra totales serían de 4,500 casillas para el conteo de diputados y de 1,280 casillas para el conteo de senadores.

Después de discutir estas 3 opciones de tamaños de muestra con los demás miembros del COTECORA y con la gente de DEOE, se decidió que las muestras para las dos elecciones de diputados y senadores fueran como en la tercera opción, con una pequeña sobremuestra en los estados de Baja California y Sonora de 5 casillas más por distrito para diputados y de 10 casillas más por estado para senadores. Los tamaños de muestra finales serían de $n = 4,580$ casillas para la elección de diputados y de $n = 1,300$ casillas para la elección de senadores, en donde ambas muestras se tomarían de forma independiente.

Consideramos que este tamaño de muestra es muy manejable por los capacitadores y asistentes electorales (CAE), lo que se esperaba que la muestra final recolectada para hacer la estimación sea de casi el 100%.

Cuadro 1 Cuantiles de la distribución de muestreo de *PEMA*

TAMAÑO DE MUESTRA	$q_{0.025}$	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	$q_{0.975}$
320	0.3	0.4	0.6	0.8	1	1.2	1.3
640	0.2	0.3	0.5	0.6	0.8	1	1.1
1280	0.1	0.2	0.4	0.5	0.7	0.9	1
1920	0.1	0.2	0.3	0.4	0.6	0.8	0.9
3200	0	0.1	0.2	0.4	0.5	0.7	0.8
3840	0	0.1	0.2	0.4	0.5	0.7	0.7
2560	0.1	0.1	0.3	0.4	0.6	0.7	0.8

Cuadro 2 Cuantiles de la distribución de muestreo de MEMA

TAMAÑO DE MUESTRA	q _{0.025}	q _{0.05}	q _{0.25}	q _{0.5}	q _{0.75}	q _{0.95}	q _{0.975}
320	4	4	5	6	7	8	8
640	3	4	4	5	5	6	6
1280	3	3	3	4	4	5	5
1920	2	3	3	3	4	5	5
2560	2	2	3	3	4	4	5
3840	2	2	2	3	3	4	4
3200	2	2	3	3	3	4	4

Cuadro 3 Tamaños de muestra por estado para la elección de senadores

ESTADO	DISTRITOS	CASILLAS X DISTRITO	TOTAL CASILLAS
AGS	3	20	60
BC	9	7	63
BCS	2	20	40
CAM	2	20	40
COA	8	8	64
COL	2	20	40
CHIS	13	5	65
CHIH	9	7	63
CDMX	22	3	66
DUR	4	15	60
GTO	15	4	60
GUE	8	8	64
HDO	7	9	63
JAL	20	3	60
MEX	40	2	80
MICH	11	6	66
MOR	5	12	60
NAY	3	20	60
NL	14	5	70
OAX	10	6	60
PUE	16	4	64

ESTADO	DISTRITOS	CASILLAS X DISTRITO	TOTAL CASILLAS
QUE	6	11	66
QROO	4	15	60
SLP	7	9	63
SIN	7	9	63
SON	7	9	63
TAB	6	10	60
TAM	8	8	64
TLAX	3	20	60
VER	19	4	76
YUC	6	11	66
ZAC	4	15	60

Figura 7 Desglose de Resultados Elección 2018

Partido	Senadores Mayoría relativa	Senadores Primera minoría	Senadores Rep. Proporcional	Total
 Movimiento Regeneración Nacional	38	4	13	55
 Partido Acción Nacional	7	10	6	23
 Partido Revolucionario Institucional	1	6	6	13
 Partido de la Revolución Democrática	1	5	2	8
 Partido Encuentro Social	7	1	0	8
 Partido Verde Ecologista de México	1	4	2	7
 Movimiento Ciudadano	4	1	2	7
 Partido del Trabajo	5	0	1	6
 Nueva Alianza	0	1	0	1
Total	64	32	32	128

Figura 8 Distribución de muestreo de $NEMA_j$ para los 10 partidos del 2018, por tamaño de muestra.

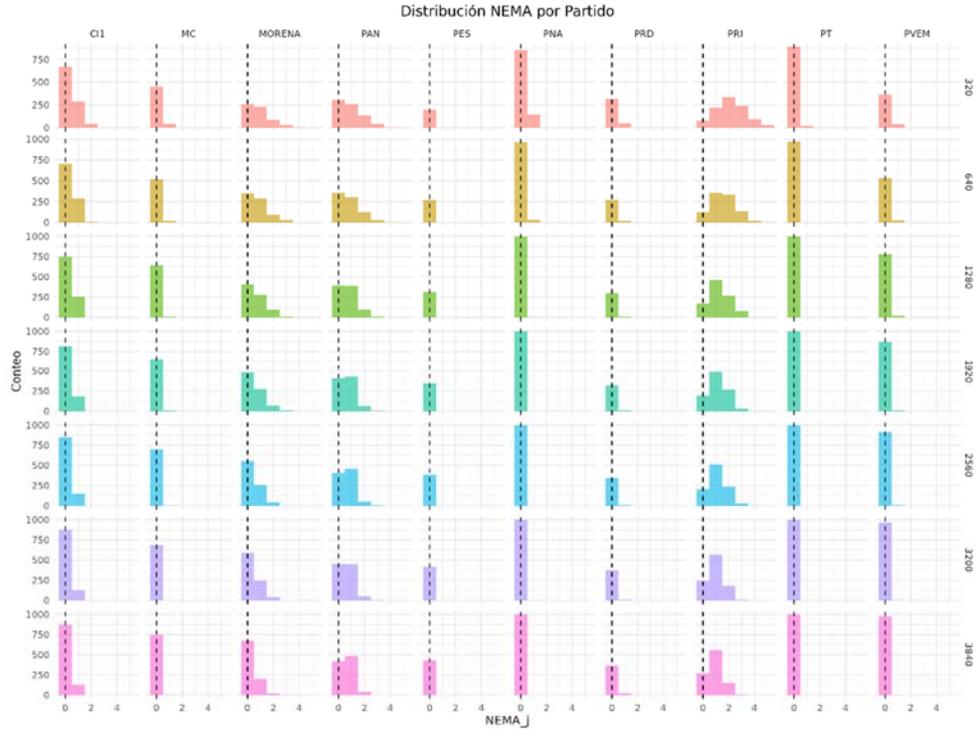


Figura 9 Distribución de muestreo de $PEMA$ por tamaño de muestra.

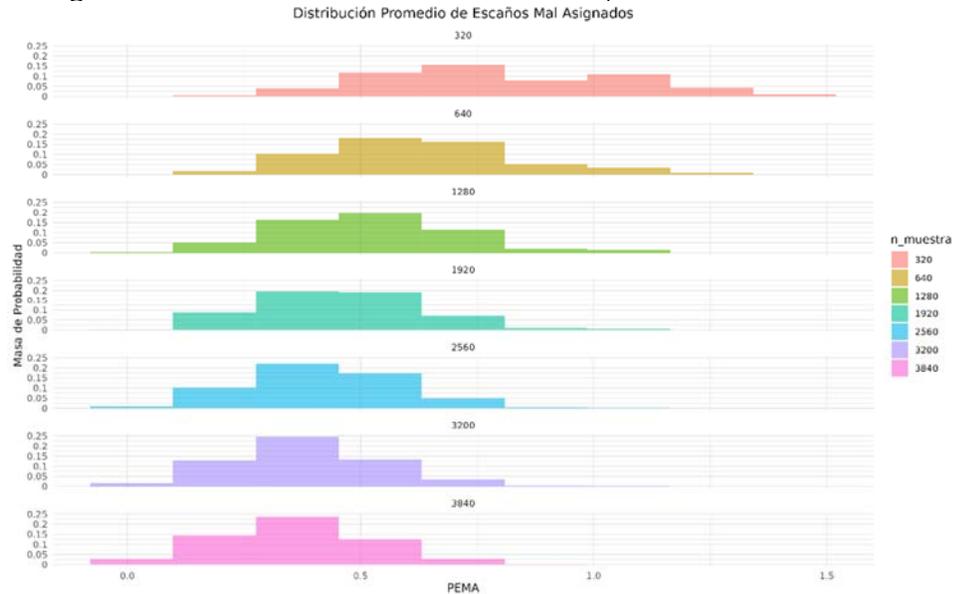


Figura 10 Distribución de muestreo de MEMA por tamaño de muestra.
Distribución Máximo de Escaños Mal Asignados

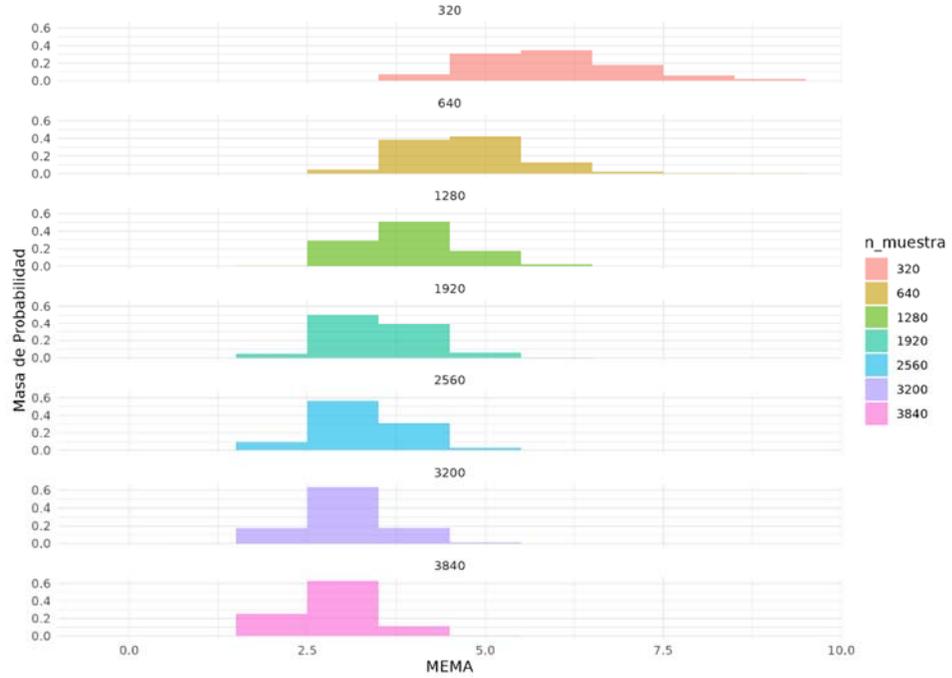


Figura 11 Distribución del número de casillas por ARE para cada



7.2.1.2. Dr. Carlos Erwin Rodríguez Hernández-Vela

Resumen

Para el Conteo Rápido (CR) en la Elección Federal de Senadores, se seguirá un diseño de muestreo estratificado, en donde:

- Los estratos serán los estados de la República Mexicana por lo que se tendrán 32 estratos.
- El tamaño de muestra total será de 1,620 asillas a nivel nacional, en donde se seleccionarán:
 - 50 casillas en cada estado, 1,600 casillas en total.
 - Una sobre-muestra de 10 casillas por diferencia de huso horario. La sobre-muestra total será de 20 casillas, en donde los estados con sobre-muestra son; Sonora (10 casillas) y Baja California (10 casillas).
- Con la muestra efectiva el margen de error en las estimaciones será de:
 - Entre 3 y 7 asientos mal asignados para el partido con el mayor número de asientos mal asignados al estimar la conformación de la cámara de senadores.
 - 0.76% al estimar la participación en la Elección Federal.
 - 0.83% al es estimar la Votación Válida Emitida por partido.

Esta estrategia de selección y márgenes de error esperados se obtuvieron tomando como referencia los Cómputos Distritales para las elecciones de senadores en los años 2012 y 2018.

- La presión sobre los CAE será mínima. Con información de las casillas instaladas para la elección de senadores del 2018 y seleccionando 5,000 muestras estratificadas por estado (50 casillas por estado para una muestra total de 1,600 casillas), se observó lo siguiente:
 - Del total de CAE que apoyarán para la organización de la elección, sólo el 4% participará mandando muestra para el CR de senadores.
 - De este 4%, poco más del 97% tendrá que reportar sólo una casilla.

Objetivos

En el Conteo Rápido (CR) para la Elección Federal de Senadores del 2 de junio de 2024, se establecieron los tres objetivos siguientes:

1. Estimar la conformación de la cámara de senadores, i.e. de los 128 senadores Federales, el número de asientos que obtiene cada partido político, así como el número de curules obtenidos por candidatos independientes:
 - a. 64 senadores por mayoría relativa. El partido (o coalición) que alcanza la mayor votación en un determinado estado obtiene dos escaños en la cámara de senadores. Lo anterior se replica en los 32 estados en los que se divide el país.
 - b. 32 senadores por primera minoría.
 - c. 32 senadores por representación proporcional, para calcular esta parte se necesita
 - i. La votación total emitida (VTE), la votación válida emitida (VVE) y la votación nacional emitida (VNE)⁶.
 - ii. Las reglas para distribuir asientos por representación proporcional como marca la LEGIPE.
2. Estimar el porcentaje de votos obtenido por cada partido tomando como referencia la VVE.
3. Estimar el porcentaje de participación nacional en la elección de senadores.

Todas las estimaciones se presentarán en forma de intervalo de confianza del 95%.

Consideraciones preliminares

De los tres objetivos establecidos para el CR Federal de Senadores, la estimación de la conformación de la cámara de senadores es el de mayor complejidad, además

- Contiene a los otros dos objetivos de manera natural.
- Estimar la conformación de la cámara de senadores no es un problema estándar y la teoría del muestreo no se puede aplicar directamente.

⁶ En el LEGIPE se definen tres tipos de votación: VTE = suma de todos los votos, la VVE = VTE – votos candidatos no registrados – votos nulos y VNE = VVE – votación de los partidos que no hayan rebasado el 3% de la VVE – votación candidatos independientes.

Por lo tanto, el diseño de muestreo del CR de senadores estará enfocado en la estimación de la conformación de la cámara de senadores.

Margen de Error para la Estimación de la Conformación de la Cámara de Senadores

Tal como en el caso de la cámara de diputados, para el caso de la estimación de la conformación de la cámara de senadores se definirá el error en la estimación como el número máximo de asientos mal asignados. Formalmente lo definimos como

$$\max_{j \in \{1,2,\dots,k\}} |NA_j - \widehat{NA}_j|. \quad (1)$$

En donde

k : Número de partidos registrados ante el INE a nivel nacional y en su caso, también se cuentan los candidatos independientes.

NA_j : Número de asientos en la cámara de senadores asignados al partido j , una vez concluido el cómputo distrital de votos (final).

\widehat{NA}_j : Número de asientos en la cámara de senadores asignados al partido j , obtenida mediante la estimación del conteo rápido

Para definir el tamaño de muestra en la Elección Federal, primero se establecerá un máximo error permisible en la asignación de asientos, denotado como **d** , y a continuación se buscará determinar el tamaño de muestra necesario para garantizar que

$$P(\max_{j \in \{1,2,\dots,k\}} |NE_j - \widehat{NE}_j| \leq d) = 0.95, \quad (2)$$

No es posible encontrar expresiones analíticas que permitan despejar el tamaño de muestra necesario para alcanzar un margen de error **d** este caso. Sin embargo, mediante ejercicios de simulación es posible aproximar la distribución para (1) usando varios tamaños de muestra y así poder aproximar el cuantil que satisfaga (2).

Simulación estocástica para determinar el margen de error con varios tamaños de muestra

Para definir el diseño de muestreo para el Conteo Rápido 2024 para estimar la conformación de la Cámara de Senadores, se usaron las bases de datos de referencia con los resultados de los cómputos distritales en las elecciones de senadores de 2012 y 2018.

En los experimentos se siguió un diseño de muestreo estratificado por estado en donde, en cada iteración se siguieron los siguientes pasos;

- Selección por MASSR de n_h casillas al interior de cada estado, considerando una de las bases de datos de referencia.
- Se usó el estimador común del total para aproximar el total de votos en favor de cada fuerza política.
- Considerando las coaliciones y los votos totales en cada estado se obtuvieron los senadores por los principios de mayoría relativa, así como por primera minoría.
- Se usó la teoría del muestreo aleatorio estratificado para combinar los resultados de los 32 estados y estimar los distintos tipos de votaciones (votación total emitida, votación válida emitida y votación nacional emitida).
- Utilizando el estimador de razón combinado se realizó la estimación de la participación, así como del porcentaje de la VVE que alcanzó cada partido político.
- Se estimó la conformación de la cámara de senadores.
 - En este punto ya se cuenta con estimaciones para la VTE, VVE y VNE.
 - Se siguieron las reglas marcadas en la LEGIPE y en la Constitución para distribuir asientos por representación proporcional. Ver Apéndice A.

Lo anterior constituye una realización de la simulación, se realizaron 5,000 iteraciones y en cada iteración se calculó el margen de error mostrado en la expresión (2).

Se probaron los tamaños de muestra al interior de cada estrato de

$$n_h = 5, 10, 15, 20, 35, 40, 50, 65, 80 \text{ y } 100 \text{ casillas.}$$

Dado que son 32 estratos y en cada estrato se usará el mismo tamaño de muestra, lo anterior implica que los tamaños de muestra total a considerar son de

$$n = 160, 320, 480, 640, 1,120, 1,280, 1,600, 2,080, 2,560 \text{ y } 3,200 \text{ casillas.}$$

En la Tabla 4 se presentan los resultados de las simulaciones, considerando todas las votaciones de referencia y los distintos tamaños de muestra.

Tabla 3 Cuantil 0.95 del máximo número de asientos mal asignados

n_h	2012	2018
5	8	6
10	8	5
15	8	4
35	7	4
40	7	3
50	7	3
65	6	3
80	6	3
100	6	3

A partir de la muestra de 50 casillas por estado ($32 \times 50 = 1,600$ a nivel nacional), no se observan mejoras significativas, en el máximo número de escaños mal asignados, por lo que para el conteo rápido del 2024 se usará este tamaño de muestra para estimar la conformación de la cámara de senadores.

Estimación de porcentajes

Con las mismas 5,000 iteraciones se monitorearon tanto los porcentajes de la VVE en favor de cada fuerza política, así como la participación en la elección. En el caso de porcentajes, el margen de error está dado mediante la siguiente ecuación.

$$P(|p - \hat{p}| \leq \epsilon) = 0.95, \quad (3)$$

En donde p es el porcentaje poblacional calculado con la base de datos de la votación completa, \hat{p} es el estimador puntual y ϵ es el margen de error. Esta estimación es estándar, por lo que la expresión (3) puede encontrarse en los libros de texto de muestreo probabilístico, por ejemplo, ver Cochran (1977) y Lohr (2021).

Desarrollando el valor absoluto se tiene que

$$P(-\epsilon \leq p - \hat{p} \leq \epsilon) = 0.95,$$

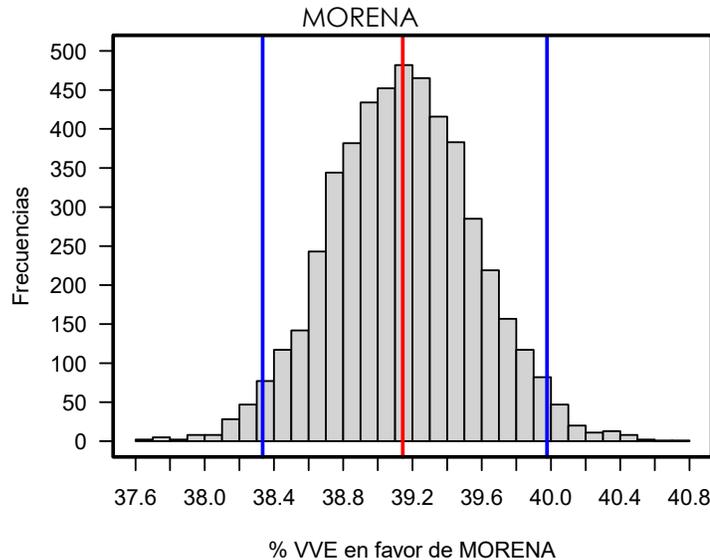
ya que p es una constante y la distribución de muestreo de \hat{p} es aproximadamente simétrica, podemos aproximar el margen de error mediante

$$\epsilon \approx \max\{p - q_{0.025}, q_{0.975} - p\},$$

en donde $q_{0.025}$ y $q_{0.975}$ son los cuantiles 0.025 y 0.975 de la distribución de muestreo de \hat{p} .

En la Figura 13 se presenta la distribución de muestreo para el porcentaje de la VVE, en favor del partido MORENA, en la elección de senadores del 2018. Las estimaciones puntuales se obtuvieron con 50 casillas por estado.

Figura 12 Distribución de muestreo para el porcentaje de la VVE en favor del partido



En la gráfica se observa el porcentaje poblacional de la VVE en favor de MORENA (en rojo), así como los cuantiles 0.025 y 0.975 de las 5,000 estimaciones puntuales (en azul). Por lo tanto, el margen de error aproximado en la estimación será menor al 0.83 por ciento. Es importante mencionar que las estimaciones para el porcentaje de la VVE en favor de MORENA fueron las que exhibieron mayor variabilidad por lo que son las que necesitarán mayor tamaño de muestra para alcanzar el mismo margen de error que las estimaciones para los porcentajes en favor de los demás partidos.

Haciendo el mismo ejercicio para la participación en la elección, con el mismo tamaño de muestra de 50 casillas por estado y considerando la elección de senadores del 2018 se obtuvo que, con este tamaño de muestra, el margen de error para estimar la participación sería menor al 0.76 por ciento.

Presión sobre los CAE

Los CAE desempeñan un papel crucial en la organización de las elecciones en México. Contratados y capacitados por el Instituto Nacional Electoral (INE), los CAE tienen la responsabilidad de reclutar y capacitar a los funcionarios de casilla. Estos últimos son fundamentales en la administración del proceso de votación en las casillas durante el día de la elección. Sus tareas abarcan desde verificar la

identidad de los votantes hasta entregar las papeletas electorales, supervisar el proceso de votación, realizar el conteo de votos al cierre de las urnas y garantizar que todo el proceso se lleve a cabo de manera transparente y ordenada. La labor coordinada entre los CAE y los funcionarios de casilla es esencial para asegurar la integridad y eficiencia del proceso electoral.

Durante la jornada electoral, los CAE se dedican a brindar apoyo activo a los funcionarios de casilla ante cualquier eventualidad que pueda surgir durante el proceso. En promedio, cada CAE es responsable de respaldar a cuatro casillas electorales, adaptando esta cifra según la densidad poblacional: menor en áreas rurales y mayor en entornos urbanos.

Además de su función principal de respaldo a los funcionarios de casilla, los CAE asumen otras responsabilidades cruciales, tales como la entrega y recolección del material electoral, la coordinación con las autoridades electorales, la transmisión de información relevante sobre el desarrollo de las elecciones, y, en algunos casos, la colaboración en la transmisión de datos de votación de casillas seleccionadas para el conteo rápido. Esta versatilidad en sus labores refleja la importancia y amplitud del rol de los CAE en el eficiente desarrollo de las elecciones.

Dado que la participación de los CAE en el conteo rápido es solo una de las diversas actividades que llevan a cabo durante la jornada electoral y no necesariamente la más prioritaria, es imperativo planificar cuidadosamente el tamaño y diseño de la muestra. El objetivo es minimizar la cantidad de CAE que deban enviar información para el conteo rápido, asegurando que no se vean abrumados y puedan cumplir eficazmente su responsabilidad principal de respaldar a los funcionarios de casilla.

Utilizando la base de datos de casillas instaladas para la elección de senadores del 2018 se seleccionaron 5,000 muestras estratificadas por estado: considerando 50 casillas por estrato para un tamaño de muestra total de 1,600 casillas. Esta base de datos contaba con 156,974 casillas aprobadas para instalarse en la elección, así como 38,721 CAE que apoyarían a los funcionarios de casillas en ese proceso electoral.

Para cada una de las 5,000 muestras estratificadas se monitoreo

- El número de CAE que participarían en el Conteo Rápido. En promedio fue de 1,558 CAE; aproximadamente el 4% del total de CAE que apoyarían en la elección.
- De los CAE que participarían en el Conteo Rápido, la distribución de casillas que tendrían que reportar. Esto se muestra en la siguiente tabla

CASILLAS A REPORTAR	1	2	3	4
% de CAE	97.33	2.61	0.05	0

Por lo que prácticamente 97% de los aproximadamente 1,558 CAE que participarían en el Conteo Rápido para la estimación de la conformación de la Cámara reportarían sólo una casilla.

Referencias

Cochran, W. G. (1977). *Sampling Techniques*, 3rd, John Wiley.

Lohr, S.L. (2021). *Sampling: Design and Analysis* (3rd ed.). Chapman and Hall/CRC.

CPEUM (2023). *Constitución política de los estados unidos mexicanos, Compendio de Legislación Nacional Electoral Tomo I*, INE.

LEGIPE (2023). *Ley General de Instituciones y Procedimientos Electorales, Compendio de Legislación Nacional Electoral Tomo II*, INE.

Apéndice A. Conformación de la Cámara de Senadores

De acuerdo con el artículo 56 de la CPEUM (2023) La Cámara de Senadores está integrada por 128 miembros, elegidos mediante tres métodos distintos: 64 mediante el principio de mayoría relativa (MR), 32 a través de la primera minoría (PM), y los 32 restantes por el método de representación proporcional (RP). De esta forma, la conformación de la Cámara se define por el total de asientos, de los 128 disponibles, que cada partido político logra asegurar. Este órgano legislativo se renueva cada seis años con elecciones simultáneas en los 32 estados de la República Mexicana. Después de estas elecciones, se establece la composición de la Cámara siguiendo el siguiente procedimiento:

1. A partir de la votación a nivel casilla se obtienen los sub-totales de votos recibidos en cada estado, aquí se tendrán votos en favor de cada partido, así como por las combinaciones válidas si es que existen coaliciones, candidatos independientes, votos nulos y candidatos no registrados.
2. En caso de existir coaliciones en un determinado estado, los votos por las combinaciones válidas para votar se reparten por igual entre el número de partidos que aparezcan en la combinación. Por lo tanto, después de este paso tendremos una base de datos de votos para los 32 estados, en donde cada renglón tendrá el número de votos en favor de cada fuerza

política en el estado (partidos y candidatos independientes), así como votos nulos y votos por candidatos no registrados.

3. En cada estado se realiza una elección de senadores, así por el principio de MR se designan dos senadores de la siguiente manera:
 - a. Si existen coaliciones, entonces los votos por cada partido de la coalición se suman a la coalición. Si la coalición obtiene el mayor número de votos en el estado, entonces se asignan los dos senadores a la coalición ganadora y se consideran los convenios de coalición para determinar qué partidos (o partido) de la coalición se queda con los dos asientos. Los acuerdos de coalición identifican una primera y segunda fórmula, la primera tiene prioridad.
 - b. Si el mayor número de votos en el estado es obtenido por un partido que va en solitario, entonces los dos asientos se le asignan a este partido.

Es importante mencionar que, si en un estado gana una coalición, entonces ninguno de los partidos de la coalición puede entrar a la competencia por el asiento por PM. De manera equivalente, si gana un partido en solitario, entonces este partido ya no puede participar en la competencia de PM.

4. Un senador se asigna a la primera minoría (PM), esta es la coalición o partido que haya alcanzado el segundo lugar en número de votos obtenidos en el estado.
 - a. Si la primera minoría es obtenida por una coalición, entonces se asigna el senador a la primera fórmula establecida en los acuerdos de la coalición.
 - b. Si la primera minoría es obtenida por un partido que va en solitario, entonces se le asigna un asiento a este partido.
5. Para asignar los 32 senadores por representación proporcional (RP), se sigue el siguiente procedimiento (artículo 21 de la LEGIPE (2023)):
 - a. Se suman los sub-totales a nivel estatal para obtener totales a nivel nacional. Estos totales representan la votación total emitida (VTE) en favor de las fuerzas políticas (partidos y candidatos independientes), votos nulos y candidatos no registrados.
 - b. Se obtiene la votación válida emitida (VVE), esta es simplemente la VTE eliminando los votos en favor de los candidatos no registrados y los votos nulos.

- c. Se calculan porcentajes de votos para cada partido político, candidatos independientes, considerando la suma total de votos de la VVE.
- d. De la VVE se quitan los votos para 1) los candidatos independientes y 2) los partidos que no superen el 3% que se calculó en el punto anterior. Esta votación recibe el nombre de votación nacional emitida (VNE). Es importante observar que este agregado de votos ya sólo tiene votos en favor de partidos políticos.
- e. Los 32 diputados se reparten de forma proporcional considerando la VNE y sólo a los partidos que la forman.
- f. El reparto anterior genera decimales, por lo que simplemente se resuelven los problemas de redondeo mediante la técnica del resto mayor para tener enteros.

7.2.2. Procedimiento de estimación

7.2.2.1. Dr. Luis Enrique Nieto Barajas y Mtro. Carlos Samuel Pérez Pérez

Metodología

El problema de estimación de la conformación de la cámara de senadores, desde un punto de vista estadístico, requiere de hacer estimaciones a dos niveles: primero, a nivel de cada uno de los 32 estados de la república, para los senadores elegidos por mayoría relativa; y en segundo lugar, a nivel nacional para la determinación de los 32 senadores por representación proporcional. Para el primer caso, se necesita tener representación muestral a nivel estado, y para el segundo, se necesita tener representación a nivel nacional. Es por esto que el diseño muestral más apropiado para el problema en cuestión es un muestreo aleatorio estratificado por estado.

El reglamento de elecciones en sus artículos 416 y 417 establece los principios para la determinación de los senadores electos por mayoría relativa, y el artículo 425 establece los principios para la asignación de senadores por representación proporcional.

Aunque los criterios se basan en la votación total y los votos por partido, es posible representar los criterios en términos de las proporciones de votos. Para este efecto introducimos la siguiente notación.

Sea θ_{ij} la proporción de votos, relativa al listado nominal, a favor del candidato/partido j en el estrato (estado) i , con $j = 1, \dots, J$ e $i = 1, \dots, N$. J es el número total de candidatos (que incluye a los candidatos por partido político y

los independientes) más dos categorías extras: los votos nulos y no registrados y las abstenciones. $N = 32$ es el número de estratos (estados). Sea n_i el listado nominal de votantes en el distrito i para $i = 1, \dots, N$, y $n = \sum_{i=1}^N n_i$ el tamaño de la lista nominal nacional. Entonces $\theta_j = \sum_{i=1}^N \frac{n_i}{n} \theta_{i,j}$ es la proporción de votos, relativa al listado nominal, a favor del candidato j a nivel nacional.

Para la determinación de los senadores ganadores por mayoría relativa definimos

$$mr_{i,j} = \begin{cases} 1, & \text{si } \theta_{i,j} > \theta_{i,l} \text{ } l \neq j, l = 1, \dots, J - 2 \\ 0, & \text{en otro caso} \end{cases}$$

una indicadora que nos dice si el candidato j ganó o no en el estado i , previa aplicación de los convenios de coalición. Este proceso se repite al considerar la primera y segunda fórmulas.

Finalmente, definimos $mr_j = \sum_{i=1}^N mr_{i,j}$ como el número de senadurías ganadas por el partido j por el principio de mayoría relativa.

Para la determinación de senadores ganadores por primera minoría definimos

$$pm_{i,j} = \begin{cases} 1, & \text{si } \theta_{i,j} \text{ es el segundo mayor de } \theta_{i,l} \text{ } l \neq j, l = 1, \dots, J - 2 \\ 0, & \text{en otro caso} \end{cases}$$

una indicadora que nos dice si el candidato j quedó o no en el segundo lugar del estado i , previa aplicación de los convenios de coalición usando la primera fórmula. Finalmente, definimos $pm_j = \sum_{i=1}^N pm_{i,j}$ como el número de senadurías ganadas por el partido j por el principio de primera minoría.

Para la determinación de senadores ganadores por primera minoría definimos primero $\lambda_j = \frac{\theta_j}{\sum_{l=1}^{J-2} \theta_l}$ la proporción de votos válidos (se excluyen las abstenciones,

los nulos y los candidatos no registrados), y $\eta_j = \frac{\theta_j}{\sum_{l=1}^{J-3} \theta_{ll}(\lambda_j > 0.03)}$ como la proporción después de excluir a los partidos que no alcanzan el 3% en la votación válida, los independientes y los nulos y no registrados.

El algoritmo consiste en calcular primero $rp_j = \lfloor 32\eta_j \rfloor$ los curules por representación proporcional cruda para el partido j . Posteriormente se calculan los decimales remanentes $32\eta_j - rp_j$ y se ordenan de mayor a menor y se les asigna una senaduría más a cada uno hasta completar 32.

Finalmente, el número de escaños por partido político NE_j resulta de sumar los escaños por mayoría relativa mr_j , primera minoría pm_j y los de representación proporcional rp_j . Es decir, $NE_j = mr_j + rp_j + pm_j$.

Para la estimación de los parámetros antes descritos, y en particular para estimar θ_{ij} , usaremos un enfoque bayesiano. En específico usaremos el modelo propuesto por Mendoza y Nieto-Barajas (2016) ⁷.

Dicho modelo supone que X_{ij}^k , denota los votos a favor del partido/candidato j en el estrato i en la casilla k . Esta variable observable sigue un comportamiento descrito por la distribución siguiente

$$X_{ij}^k | \theta_{ij}, \tau_{ij} \sim N \left(n_i^k \theta_{ij}, \frac{\tau_{ij}}{n_i^k} \right)$$

con $k = 1, \dots, c_i$, con c_i casillas en muestra en el distrito i , $i = 1, \dots, N$ y $j = 1, \dots, J$.

Donde n_i^k es la lista nominal para la casilla k en el distrito i y τ_{ij} es la precisión que se supone constante dentro del estrato e independiente de θ_{ij} , más aún X_{ij}^k independiente de $X_{ij}^{k'}$, para $j \neq j'$.

La verosimilitud para cada estrato $i = 1, \dots, N$ y cada candidato $j = 1, \dots, J$ está dada por

$$L(\theta_{ij}, \tau_{ij} | x_{ij}) \propto \tau_{ij}^{\frac{c_i}{2}} \exp \left\{ -\frac{\tau_{ij}}{2} \sum_{k=1}^{c_i} \frac{1}{n_i^k} (x_{ij}^k - n_i^k \theta_{ij})^2 \right\}$$

La distribución inicial para los parámetros desconocidos del modelo, $\theta_{i,j}$ y $\tau_{i,j}$ se considera independiente y no informativa, es decir

$$p(\theta_{ij}, \tau_{ij}) = U(\theta_{ij} | 0,1) \text{Ga}(\tau_{ij} | 0.5, 0.05)$$

Así, usando el Teorema de Bayes, la distribución posterior resulta ser proporcional al producto de una normal truncada para θ_{ij} condicional en τ_{ij} y una distribución gamma para τ_{ij} . Los estimadores puntuales y por intervalo para θ_{ij} se aproximarán mediante simulaciones de la distribución posterior.

7.2.2.2. Dr. Carlos Erwin Rodríguez Hernández-Vela y Lic. José Alberto Florencio Chávez

Resumen

⁷ Mendoza, M. y Nieto-Barajas L.E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. Electoral Studies 43, 124–132.

Para el Conteo Rápido (CR) en la Elección Federal de Senadores, se seguirá un diseño de muestreo estratificado, en donde:

- Los estratos serán los estados de la República Mexicana por lo que se tendrán 32 estratos.
- El tamaño de muestra total será de 1,620 asillas a nivel nacional, en donde se seleccionarán:
 - 50 casillas en cada estado, 1,600 casillas en total.
 - Una sobre-muestra de 10 casillas por diferencia de huso horario. La sobre-muestra total será de 20 casillas, en donde los estados con sobre-muestra son; Sonora (10 casillas) y Baja California (10 casillas).
- Con la muestra efectiva el margen de error en las estimaciones será de:
 - Entre 3 y 7 asientos mal asignados para el partido con el mayor número de asientos mal asignados al estimar la conformación de la cámara de senadores.
 - 0.76% al estimar la participación en la Elección Federal.
 - 0.83% al es estimar la Votación Válida Emitida por partido.

Esta estrategia de selección y márgenes de error esperados se obtuvieron tomando como referencia los Cómputos Distritales para las elecciones de senadores en los años 2012 y 2018.

- La presión sobre los CAE será mínima. Con información de las casillas instaladas para la elección de senadores del 2018 y seleccionando 5,000 muestras estratificadas por estado (50 casillas por estado para una muestra total de 1,600 casillas), se observó lo siguiente:
 - Del total de CAE que apoyarán para la organización de la elección, sólo el 4% participará mandando muestra para el CR de senadores.
 - De este 4%, poco más del 97% tendrá que reportar sólo una casilla.

Objetivos

En el Conteo Rápido (CR) para la Elección Federal de Senadores del 2 de junio de 2024, se establecieron los tres objetivos siguientes:

4. Estimar la conformación de la cámara de senadores, i.e. de los 128 senadores Federales, el número de asientos que obtiene cada partido

político, así como el número de curules obtenidos por candidatos independientes:

- a. 64 senadores por mayoría relativa. El partido (o coalición) que alcanza la mayor votación en un determinado estado obtiene dos escaños en la cámara de senadores. Lo anterior se replica en los 32 estados en los que se divide el país.
 - b. 32 senadores por primera minoría.
 - c. 32 senadores por representación proporcional, para calcular esta parte se necesita
 - i. La votación total emitida (VTE), la votación válida emitida (VVE) y la votación nacional emitida (VNE)⁸.
 - ii. Las reglas para distribuir asientos por representación proporcional como marca la LEGIPE.
5. Estimar el porcentaje de votos obtenido por cada partido tomando como referencia la VVE.
 6. Estimar el porcentaje de participación nacional en la elección de senadores.

Todas las estimaciones se presentarán en forma de intervalo de confianza del 95%.

Consideraciones preliminares

De los tres objetivos establecidos para el CR Federal de Senadores, la estimación de la conformación de la cámara de senadores es el de mayor complejidad, además

- Contiene a los otros dos objetivos de manera natural.
- Estimar la conformación de la cámara de senadores no es un problema estándar y la teoría del muestreo no se puede aplicar directamente.

Por lo tanto, el diseño de muestreo del CR de senadores estará enfocado en la estimación de la conformación de la cámara de senadores.

⁸ En el LEGIPE se definen tres tipos de votación: VTE = suma de todos los votos, la VVE = VTE – votos candidatos no registrados – votos nulos y VNE = VVE – votación de los partidos que no hayan rebasado el 3% de la VVE – votación candidatos independientes.

Margen de Error para la Estimación de la Conformación de la Cámara de Senadores

Tal como en el caso de la cámara de diputados, para el caso de la estimación de la conformación de la cámara de senadores se definirá el error en la estimación como el número máximo de asientos mal asignados. Formalmente lo definiremos como

$$\max_{j \in \{1,2,\dots,k\}} |NA_j - \widehat{NA}_j|. \quad (1)$$

En donde

k: Número de partidos registrados ante el INE a nivel nacional y en su caso, también se cuentan los candidatos independientes.

NA_j: Número de asientos en la cámara de senadores asignados al partido j, una vez concluido el cómputo distrital de votos (final).

\widehat{NA}_j : Número de asientos en la cámara de senadores asignados al partido j, obtenida mediante la estimación del conteo rápido

Para definir el tamaño de muestra en la Elección Federal, primero se establecerá un máximo error permisible en la asignación de asientos, denotado como **d**, y a continuación se buscará determinar el tamaño de muestra necesario para garantizar que

$$P(\max_{j \in \{1,2,\dots,k\}} |NE_j - \widehat{NE}_j| \leq d) = 0.95, \quad (2)$$

No es posible encontrar expresiones analíticas que permitan despejar el tamaño de muestra necesario para alcanzar un margen de error **d** este caso. Sin embargo, mediante ejercicios de simulación es posible aproximar la distribución para (1) usando varios tamaños de muestra y así poder aproximar el cuantil que satisfaga (2).

Simulación estocástica para determinar el margen de error con varios tamaños de muestra

Para definir el diseño de muestreo para el Conteo Rápido 2024 para estimar la conformación de la Cámara de Senadores, se usaron las bases de datos de referencia con los resultados de los cómputos distritales en las elecciones de senadores de 2012 y 2018.

En los experimentos se siguió un diseño de muestreo estratificado por estado en donde, en cada iteración se siguieron los siguientes pasos;

- Selección por MASSR de n_h casillas al interior de cada estado, considerando una de las bases de datos de referencia.
- Se uso el estimador común del total para aproximar el total de votos en favor de cada fuerza política.
- Considerando las coaliciones y los votos totales en cada estado se obtuvieron los senadores por los principios de mayoría relativa, así como por primera minoría.
- Se usó la teoría del muestreo aleatorio estratificado para combinar los resultados de los 32 estados y estimar los distintos tipos de votaciones (votación total emitida, votación válida emitida y votación nacional emitida).
- Utilizando el estimador de razón combinado se realizó la estimación de la participación, así como del porcentaje de la VVE que alcanzó cada partido político.
- Se estimó la conformación de la cámara de senadores.
 - En este punto ya se cuenta con estimaciones para la VTE, VVE y VNE.
 - Se siguieron las reglas marcadas en la LEGIPE y en la Constitución para distribuir asientos por representación proporcional. Ver Apéndice B.

Lo anterior constituye una realización de la simulación, se realizaron 5,000 iteraciones y en cada iteración se calculó el margen de error mostrado en la expresión (2).

Se probaron los tamaños de muestra al interior de cada estrato de

$$n_h = 5, 10, 15, 20, 35, 40, 50, 65, 80 \text{ y } 100 \text{ casillas.}$$

Dado que son 32 estratos y en cada estrato se usará el mismo tamaño de muestra, lo anterior implica que los tamaños de muestra total a considerar son de

$$n = 160, 320, 480, 640, 1,120, 1,280, 1,600, 2,080, 2,560 \text{ y } 3,200 \text{ casillas.}$$

En la Tabla 4 se presentan los resultados de las simulaciones, considerando todas las votaciones de referencia y los distintos tamaños de muestra.

Tabla 4 Cuantil 0.95 del máximo número de asientos mal asignados

n_h	2012	2018
5	8	6
10	8	5

15	8	4
35	7	4
40	7	3
50	7	3
65	6	3
80	6	3
100	6	3

A partir de la muestra de 50 casillas por estado ($32 \times 50 = 1,600$ a nivel nacional), no se observan mejoras significativas, en el máximo número de escaños mal asignados, por lo que para el conteo rápido del 2024 se usará este tamaño de muestra para estimar la conformación de la cámara de senadores.

Estimación de porcentajes

Con las mismas 5,000 iteraciones se monitorearon tanto los porcentajes de la VVE en favor de cada fuerza política, así como la participación en la elección. En el caso de porcentajes, el margen de error está dado mediante la siguiente ecuación.

$$P(|p - \hat{p}| \leq \epsilon) = 0.95, \quad (3)$$

En donde p es el porcentaje poblacional calculado con la base de datos de la votación completa, \hat{p} es la estimador puntual y ϵ es el margen de error. Esta estimación es estándar, por lo que la expresión (3) puede encontrarse en los libros de texto de muestreo probabilístico, por ejemplo, ver Cochran (1977) y Lohr (2021).

Desarrollando el valor absoluto se tiene que

$$P(-\epsilon \leq p - \hat{p} \leq \epsilon) = 0.95,$$

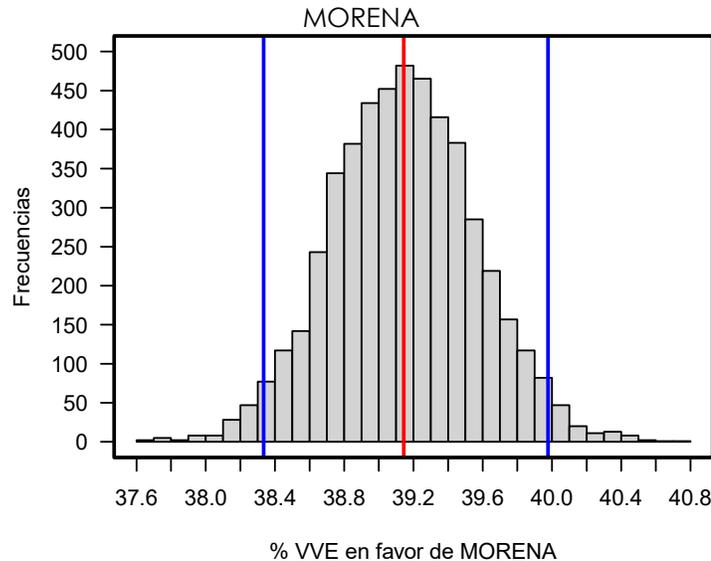
ya que p es una constante y la distribución de muestreo de \hat{p} es aproximadamente simétrica, podemos aproximar el margen de error mediante

$$\epsilon \approx \max\{p - q_{0.025}, q_{0.975} - p\},$$

en donde $q_{0.025}$ y $q_{0.975}$ son los cuantiles 0.025 y 0.975 de la distribución de muestreo de \hat{p} .

En la Figura 13 se presenta la distribución de muestreo para el porcentaje de la VVE, en favor del partido MORENA, en la elección de senadores del 2018. Las estimaciones puntuales se obtuvieron con 50 casillas por estado.

Figura 13 Distribución de muestreo para el porcentaje de la VVE en favor del partido



En la gráfica se observa el porcentaje poblacional de la VVE en favor de MORENA (en rojo), así como los cuantiles 0.025 y 0.975 de las 5,000 estimaciones puntuales (en azul). Por lo tanto, el margen de error aproximado en la estimación será menor al 0.83 por ciento. Es importante mencionar que las estimaciones para el porcentaje de la VVE en favor de MORENA fueron las que exhibieron mayor variabilidad por lo que son las que necesitarán mayor tamaño de muestra para alcanzar el mismo margen de error que las estimaciones para los porcentajes en favor de los demás partidos.

Haciendo el mismo ejercicio para la participación en la elección, con el mismo tamaño de muestra de 50 casillas por estado y considerando la elección de senadores del 2018 se obtuvo que, con este tamaño de muestra, el margen de error para estimar la participación sería menor al 0.76 por ciento.

Presión sobre los CAE

Los CAE desempeñan un papel crucial en la organización de las elecciones en México. Contratados y capacitados por el Instituto Nacional Electoral (INE), los CAE tienen la responsabilidad de reclutar y capacitar a los funcionarios de casilla. Estos últimos son fundamentales en la administración del proceso de votación en las casillas durante el día de la elección. Sus tareas abarcan desde verificar la identidad de los votantes hasta entregar las papeletas electorales, supervisar el proceso de votación, realizar el conteo de votos al cierre de las urnas y garantizar que todo el proceso se lleve a cabo de manera transparente y ordenada. La labor coordinada entre los CAE y los funcionarios de casilla es esencial para asegurar la integridad y eficiencia del proceso electoral.

Durante la jornada electoral, los CAE se dedican a brindar apoyo activo a los funcionarios de casilla ante cualquier eventualidad que pueda surgir durante el proceso. En promedio, cada CAE es responsable de respaldar a cuatro casillas electorales, adaptando esta cifra según la densidad poblacional: menor en áreas rurales y mayor en entornos urbanos.

Además de su función principal de respaldo a los funcionarios de casilla, los CAE asumen otras responsabilidades cruciales, tales como la entrega y recolección del material electoral, la coordinación con las autoridades electorales, la transmisión de información relevante sobre el desarrollo de las elecciones, y, en algunos casos, la colaboración en la transmisión de datos de votación de casillas seleccionadas para el conteo rápido. Esta versatilidad en sus labores refleja la importancia y amplitud del rol de los CAE en el eficiente desarrollo de las elecciones.

Dado que la participación de los CAE en el conteo rápido es solo una de las diversas actividades que llevan a cabo durante la jornada electoral y no necesariamente la más prioritaria, es imperativo planificar cuidadosamente el tamaño y diseño de la muestra. El objetivo es minimizar la cantidad de CAE que deban enviar información para el conteo rápido, asegurando que no se vean abrumados y puedan cumplir eficazmente su responsabilidad principal de respaldar a los funcionarios de casilla.

Utilizando la base de datos de casillas instaladas para la elección de senadores del 2018 se seleccionaron 5,000 muestras estratificadas por estado: considerando 50 casillas por estrato para un tamaño de muestra total de 1,600 casillas. Esta base de datos contaba con 156,974 casillas aprobadas para instalarse en la elección, así como 38,721 CAE que apoyarían a los funcionarios de casillas en ese proceso electoral.

Para cada una de las 5,000 muestras estratificadas se monitoreo

- El número de CAE que participarían en el Conteo Rápido. En promedio fue de 1,558 CAE; aproximadamente el 4% del total de CAE que apoyarían en la elección.
- De los CAE que participarían en el Conteo Rápido, la distribución de casillas que tendrían que reportar. Esto se muestra en la siguiente tabla

CASILLAS A REPORTAR	1	2	3	4
% de CAE	97.33	2.61	0.05	0

Por lo que prácticamente 97% de los aproximadamente 1,558 CAE que participarían en el Conteo Rápido para la estimación de la conformación de la Cámara reportarían sólo una casilla.

Referencias

Cochran, W. G. (1977). *Sampling Techniques*, 3rd, John Wiley.

Lohr, S.L. (2021). *Sampling: Design and Analysis* (3rd ed.). Chapman and Hall/CRC.

CPEUM (2023). *Constitución política de los estados unidos mexicanos, Compendio de Legislación Nacional Electoral Tomo I*, INE.

LEGIPE (2023). *Ley General de Instituciones y Procedimientos Electorales, Compendio de Legislación Nacional Electoral Tomo II*, INE.

Apendice B. Conformación de la Cámara de Senadores

De acuerdo con el artículo 56 de la CPEUM (2023) La Cámara de Senadores está integrada por 128 miembros, elegidos mediante tres métodos distintos: 64 mediante el principio de mayoría relativa (MR), 32 a través de la primera minoría (PM), y los 32 restantes por el método de representación proporcional (RP). De esta forma, la conformación de la Cámara se define por el total de asientos, de los 128 disponibles, que cada partido político logra asegurar. Este órgano legislativo se renueva cada seis años con elecciones simultáneas en los 32 estados de la República Mexicana. Después de estas elecciones, se establece la composición de la Cámara siguiendo el siguiente procedimiento:

A partir de la votación a nivel casilla se obtienen los sub-totales de votos recibidos en cada estado, aquí se tendrán votos en favor de cada partido, así como por las combinaciones válidas si es que existen coaliciones, candidatos independientes, votos nulos y candidatos no registrados.

En caso de existir coaliciones en un determinado estado, los votos por las combinaciones válidas para votar se reparten por igual entre el número de partidos que aparezcan en la combinación. Por lo tanto, después de este paso tendremos una base de datos de votos para los 32 estados, en donde cada renglón tendrá el número de votos en favor de cada fuerza política en el estado (partidos y candidatos independientes), así como votos nulos y votos por candidatos no registrados.

En cada estado se realiza una elección de senadores, así por el principio de MR se designan dos senadores de la siguiente manera:

Si existen coaliciones, entonces los votos por cada partido de la coalición se suman a la coalición. Si la coalición obtiene el mayor número de votos en el estado, entonces se asignan los dos senadores a la coalición ganadora y se consideran los convenios de coalición para determinar qué partidos (o partido) de la coalición se queda con los dos asientos. Los acuerdos de coalición identifican una primera y segunda fórmula, la primera tiene prioridad.

Si el mayor número de votos en el estado es obtenido por un partido que va en solitario, entonces los dos asientos se le asignan a este partido.

Es importante mencionar que, si en un estado gana una coalición, entonces ninguno de los partidos de la coalición puede entrar a la competencia por el asiento por PM. De manera equivalente, si gana un partido en solitario, entonces este partido ya no puede participar en la competencia de PM.

Un senador se asigna a la primera minoría (PM), esta es la coalición o partido que haya alcanzado el segundo lugar en número de votos obtenidos en el estado.

Si la primera minoría es obtenida por una coalición, entonces se asigna el senador a la primera fórmula establecida en los acuerdos de la coalición.

Si la primera minoría es obtenida por un partido que va en solitario, entonces se le asigna un asiento a este partido.

Para asignar los 32 senadores por representación proporcional (RP), se sigue el siguiente procedimiento (artículo 21 de la LEGIPE (2023)):

Primero, se suman los sub-totales a nivel estatal para obtener totales a nivel nacional. Estos totales representan la votación total emitida (VTE) en favor de las fuerzas políticas (partidos y candidatos independientes), votos nulos y candidatos no registrados.

Se obtiene la votación válida emitida (VVE), esta es simplemente la VTE eliminando los votos en favor de los candidatos no registrados y los votos nulos.

Se calculan porcentajes de votos para cada partido político, candidatos independientes, considerando la suma total de votos de la VVE.

De la VVE se quitan los votos para 1) los candidatos independientes y 2) los partidos que no superen el 3% que se calculó en el punto anterior. Esta votación recibe el nombre de votación nacional emitida (VNE). Es importante observar que este agregado de votos ya sólo tiene votos en favor de partidos políticos.

Los 32 diputados se reparten de forma proporcional considerando la VNE y sólo a los partidos que la forman.

El reparto anterior genera decimales, por lo que simplemente se resuelven los problemas de redondeo mediante la técnica del resto mayor para tener enteros.

7.2.2.3. Equipo 1

Dra. Lizbeth Lizbeth Naranjo Albarrán, Mtra. Patricia Isabel Romero Mares y
Dr. Raúl Rueda Díaz del Campo

La Cámara de Senadores está integrada por 128 miembros, 96 son los senadores por Mayoría Relativa y 32 los de Representación Proporcional.

Para estimar el número de senadores por Mayoría relativa, se registra el número de senadores que gana cada partido, considerando que en cada estado del país se reparten 3 senadores por este concepto, dos para el partido o coalición que tenga la primera mayoría y un senador asignado al partido o coalición que tenga la primera minoría. En caso en que sean coaliciones las que obtengan el mayor número de votos en el estado, se consideran los acuerdos de coalición que previamente construyeron los partidos que la integran, para asignar los o el senador al partido correspondiente a estos acuerdos.

Entonces, dada la muestra de casillas para una entidad federativa, se tiene que estimar el total de votos en dicha entidad para cada partido y coalición existente. Con esta estimación se asignan los senadores a los partidos o coaliciones como se explicó anteriormente.

La forma de estimar el total de votos a cada partido o coalición se lleva a cabo según el enfoque estadístico utilizado por cada uno de los miembros del equipo 1, en forma clásica o bayesiana.

Posteriormente, se reparten los votos de las coaliciones entre los partidos involucrados y se suman las votaciones a nivel nacional, con las cuales, y haciendo uso de las reglas para asignar los senadores por representación proporcional, se asignan estos senadores. Sumando el número de senadores por MR con los asignados por RP en cada partido se obtiene una configuración de la cámara de senadores.

El procedimiento completo se repite un número grande de veces, digamos B , lo que produce B configuraciones de la cámara de senadores. De aquí se calculan los cuantiles que representan el intervalo de 95% de confianza para la configuración de la cámara.

7.3. Elección de Diputados

7.3.1. Estratificación y tamaño de muestra

7.3.1.1. Dr. Carlos Erwin Rodríguez Hernández-Vela y Lic. José Alberto Florencio Chávez

Resumen

Para el Conteo Rápido (CR) en la Elección Federal de Diputados, se seguirá un diseño de muestreo estratificado, en donde:

- Los estratos serán los distritos federales, por lo que se tendrán 300 estratos.
- El tamaño de muestra total será de 4,620 asillas a nivel nacional, en donde se seleccionarán:
 - 15 casillas en cada distrito, 4,500 casillas en total.
 - 5 casillas de sobre-muestra debido a diferencias de huso horario y a que en el Conteo Rápido de 2021 se observó una muy baja recepción de casillas en estos estados. La sobre-muestra total será de 120 casillas. Los estados con sobre-muestra son; Sonora (7 distritos – 35 casillas), Baja California (9 distritos – 45 casillas) y Guerrero (8 distritos-40 casillas).
- Con la muestra efectiva el margen de error en las estimaciones será de:
 - Entre 6 y 9 asientos mal asignados para el partido con el mayor número de asientos mal asignados al estimar la conformación de la cámara de diputados.
 - 0.5% al estimar la participación en la Elección Federal.
 - 0.35% al es estimar la Votación Válida Emitida por partido.

Todos estos márgenes de error se observarán con un 95% de confianza. Estos resultados se obtuvieron tomando como referencia los Cómputos Distritales para las elecciones de diputados en los años 2012, 2015, 2018 y 2021.

- La presión sobre los CAE será mínima. Con información de las casillas instaladas para la elección de diputados del 2021 y seleccionando 5,000 muestras estratificadas por distrito federal (15 casillas por distrito federal para una muestra total de 4,500 casillas), se observó lo siguiente:

- Del total de CAE que apoyarán para la organización de la elección, sólo el 10% participará mandando muestra para el CR de diputados.
- De este 10%, poco menos del 96% tendrá que reportar sólo una casilla.

Objetivos

En el Conteo Rápido (CR) para la Elección Federal de Diputados del 2 de junio de 2024, se establecieron los tres objetivos siguientes:

1. Estimar la conformación de la cámara de diputados, i.e. de los 500 diputados que Federales, el número de asientos que obtiene cada partido político, así como el número de curules obtenidos por candidatos independientes:
 - a. 300 diputados por mayoría relativa. El partido que alcanza la mayor votación en un determinado distrito obtiene un escaño en la cámara de diputados. Lo anterior se replica en los 300 distritos electorales en los que se divide el país. Esta vía es la única en la que un candidato independiente puede obtener un escaño en la cámara.
 - b. 200 diputados por representación proporcional. Para calcular esta parte se necesita
 - i. La conformación de los 300 diputados por mayoría relativa.
 - ii. La votación total emitida (VTE), la votación válida emitida (VVE) y la votación nacional emitida (VNE)⁹.
 - iii. Las reglas para distribuir asientos por representación proporcional como marcan tanto el LEGIPE como la Constitución Política de los Estados Unidos Mexicanos.
2. Estimar el porcentaje de votos obtenido por cada partido tomando como referencia la VVE. Esta votación es particularmente importante pues es la que toma como referencia la ley para determinar si un partido pierde o

⁹ En el LEGIPE se definen tres tipos de votación: VTE = suma de todos los votos, la VVE = VTE – votos candidatos no registrados – votos nulos y VNE = VVE – votación de los partidos que no hayan rebasado el 3% de la VVE – votación candidatos independientes.

mantiene el registro. Si un partido iguala o supera el 3% de la VVE mantiene su registro, en otro caso lo pierde.

3. Estimar el porcentaje de participación nacional en la elección.

Todas las estimaciones se presentarán en forma de intervalo de confianza del 95%.

Consideraciones preliminares

De los tres objetivos establecidos para el CR Federal de Diputados, la estimación de la conformación de la cámara de diputados es el de mayor complejidad, además

- Contiene a los otros dos objetivos de manera natural.
- Estimar la conformación de la cámara de diputados no es un problema estándar y la teoría del muestreo no se puede aplicar directamente.

Por lo tanto, el diseño de muestreo del CR de diputados estará enfocado en la estimación de la conformación de la cámara de diputados.

Margen de Error para la Estimación de la Conformación de la Cámara de Diputados

Lo que se busca en esta sección es introducir un margen de error para la estimación de la conformación de la cámara. Primero es necesario definir una medida de error. Con esta meta se presenta la Tabla 5, que contiene lo sucedido en el Conteo Rápido de 2003.

Tabla 5 Conformación de la cámara de diputados en la Elección Federal de 2003:
Estimaciones puntuales de los tres equipos

NO	PARTIDOS	RESULTADO FINAL (12 DE JULIO, 2003)	ESTIMACIÓN PUNTUAL DE LA CONFORMACIÓN DE LA CÁMARA DE DIPUTADOS 2003 (6 DE JULIO, 2003)		
			IGNACIO Y HORTENSIA	RAÚL Y PATY	FEDERICO Y KARIM
1	PAN	153	155	152	155
2	PRI	226	223	226	224
3	PRD	95	96	96	95
4	PVEM	15	15	15	15
5	PT	6	6	6	6

NO	PARTIDOS	RESULTADO FINAL (12 DE JULIO, 2003)	ESTIMACIÓN PUNTUAL DE LA CONFORMACIÓN DE LA CÁMARA DE DIPUTADOS 2003 (6 DE JULIO, 2003)		
			IGNACIO Y HORTENSIA	RAÚL Y PATY	FEDERICO Y KARIM
6	Convergencia	5	5	5	5

Para el caso de la estimación de la conformación de la cámara definiremos el error en la estimación como el número máximo de asientos mal asignados. Formalmente lo definimos como

$$\max_{j \in \{1,2,\dots,k\}} |NA_j - \widehat{NA}_j|. \quad (1)$$

En donde

k Número de partidos registrados ante el INE a nivel nacional y en su caso, también se cuentan los candidatos independientes.

NA_j Número de asientos en la cámara de diputados asignados al partido j , una vez concluido el cómputo distrital de votos (final).

\widehat{NA}_j Número de asientos en la cámara de diputados asignados al partido j , obtenida mediante la estimación del conteo rápido.

El máximo número de asientos mal asignados es una de muchas posibilidades para la medida de error. Es posible definir otros criterios de forma similar. Sin embargo, el criterio (1) es fácil de comprender y en los ejercicios realizados se observaron buenos resultados. En la Tabla 6 se muestra 1) el número de asientos mal asignados para cada fuerza política y el máximo número de asientos mal asignados de cada uno de los tres equipos en el Conteo Rápido de 2003 (Véase también la Tabla 5).

Tabla 6 Error en la estimación de la conformación de la cámara de diputados en la Elección Federal de 2003

PARTIDOS	ASIENTOS MAL ASIGNADOS	ASIENTOS MAL ASIGNADOS	ASIENTOS MAL ASIGNADOS
	IGNACIO Y HORTENSIA	RAÚL Y PATY	FEDERICO Y KARIM
PAN	2	1	2
PRI	1	2	0
PRD	1	1	0
PVEM	2	2	2
PT	0	0	0

PARTIDOS	ASIENTOS MAL ASIGNADOS IGNACIO Y HORTENSIA	ASIENTOS MAL ASIGNADOS RAÚL Y PATY	ASIENTOS MAL ASIGNADOS FEDERICO Y KARIM
Convergencia	0	0	0
Máximo número de asientos mal asignados	2	2	2

Para definir el tamaño de muestra en la Elección Federal, primero se establecerá un máximo error permisible en la asignación de asientos, denotado como d , y a continuación se buscará determinar el tamaño de muestra necesario para garantizar que

$$P(\max_{j \in \{1,2,\dots,k\}} |NE_j - \widehat{NE}_j| \leq d) = 0.95, \quad (2)$$

No es posible encontrar expresiones analíticas que permitan despejar el tamaño de muestra necesario para alcanzar un margen de error d este caso. Sin embargo, mediante ejercicios de simulación es posible aproximar la distribución para (1), usando varios tamaños de muestra, y así poder aproximar el cuantil que satisfaga (2).

Simulación estocástica

Para definir el diseño de muestreo para el Conteo Rápido 2024, se usaron las bases de datos de referencia con los resultados de los cómputos distritales en las elecciones de diputados 2012, 2015, 2018 y 2021.

En todos los experimentos se siguió un diseño de muestreo estratificado por distrito federal electoral en donde, en cada iteración de un ejercicio de simulación estocástica se siguieron los siguientes pasos;

- Selección por Muestreo Aleatorio Simple Sin Reemplazo (MASSR) de n_h casillas al interior de cada distrito, considerando una de las bases de referencia.
- Se uso el estimador común del total para aproximar el total de votos en favor de cada fuerza política.
- Considerando las coaliciones y los votos totales en cada distrito se obtuvieron los diputados por el principio de mayoría relativa.
- Se usó la teoría del muestreo aleatorio estratificado para combinar los resultados de los 300 distritos y estimar los distintos tipos de votaciones

(votación total emitida, votación válida emitida y votación nacional emitida).

- Utilizando el estimador de razón combinado se realizó la estimación de la participación, así como del porcentaje de la VVE que alcanzó cada partido político.
- Se estimó la conformación de la cámara de diputados.
 - En este punto ya se cuenta con estimaciones para los diputados por MR, así como para la VTE y VVE.
 - Se siguieron las reglas marcadas en la LEGIPE y en la Constitución para distribuir asientos por representación proporcional.

Lo anterior constituye una realización de la simulación, se realizaron 5,000 iteraciones y en cada iteración se calculó el margen de error mostrado en la expresión (2).

En los ejercicios de simulación se probaron los tamaños de muestra al interior de cada estrato de $n_h = 5, 10, 15, 20, 25$ y 30 , dado que son 300 estratos y en cada estrato se usará el mismo tamaño de muestra, lo anterior implica que los tamaños de muestra a considerar son de

$$n = 1,500, 3,000, 4,500, 6,000, 7,500 \text{ y } 9,000 \text{ casillas.}$$

En la Tabla 7 se presentan los resultados de las simulaciones, considerando todas las votaciones de referencia y los distintos tamaños de muestra por estrato.

Tabla 7 Cuantil 0.95 del máximo número de escaños mal asignados

n_h	2012	2015	2018	2021
5	9	10	9	11
10	7	9	7	9
15	6	9	6	8
20	6	9	6	7
25	5	8	5	7
30	5	8	5	6

A partir de la muestra de 15 casillas por distrito federal electoral ($300 \times 15 = 4,500$ a nivel nacional), no se observan mejoras significativas, en el máximo número de escaños mal asignados, por lo que para el conteo rápido del 2024 se usará este tamaño de muestra para estimar la conformación de la cámara de diputados.

Estimación de porcentajes

Con las mismas 5,000 iteraciones se monitorearon tanto los porcentajes de la VVE en favor de cada fuerza política, así como la participación en la elección. En el caso de porcentajes, el margen de error está dado mediante la siguiente ecuación.

$$P(|p - \hat{p}| \leq \epsilon) = 0.95, \quad (3)$$

En donde p es el porcentaje poblacional calculado con la base de datos de la votación completa, \hat{p} es el estimador puntual y ϵ es el margen de error. Esta estimación es estándar, por lo que la expresión (3) puede encontrarse en los libros de texto de muestreo probabilístico, por ejemplo, ver Cochran (1977) y Lohr (2021).

Desarrollando el valor absoluto se tiene que

$$P(-\epsilon \leq p - \hat{p} \leq \epsilon) = 0.95,$$

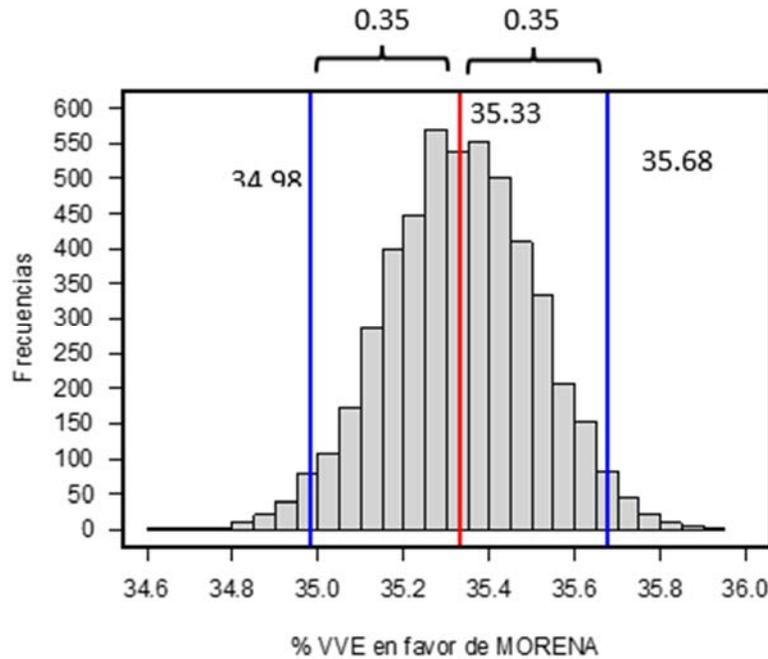
ya que p es una constante y la distribución de muestreo de \hat{p} es aproximadamente simétrica, podemos aproximar el margen de error mediante

$$\epsilon \approx \max\{p - q_{0.025}, q_{0.975} - p\},$$

en donde $q_{0.025}$ y $q_{0.975}$ son los cuantiles 0.025 y 0.975 de la distribución de muestreo de \hat{p} .

En la Figura 14 se presenta la distribución de muestreo para el porcentaje de la VVE, en favor del partido MORENA, en la elección de diputados del 2021. Las estimaciones puntuales se obtuvieron con 15 casillas por distrito federal electoral.

Figura 14 Distribución de muestreo para el porcentaje de la VVE en favor del partido MORENA



En la gráfica se observa el porcentaje poblacional de la VVE en favor de MORENA (en rojo), así como los cuantiles 0.025 y 0.975 de las 5,000 estimaciones puntuales (en azul). Por lo tanto, el margen de error aproximado en la estimación será menor al 0.35 por ciento. Es importante mencionar que las estimaciones para el porcentaje de la VVE en favor de MORENA fueron las que exhibieron mayor variabilidad por lo que son las que necesitarán mayor tamaño de muestra para alcanzar el mismo margen de error que las estimaciones para los porcentajes en favor de los demás partidos.

Haciendo el mismo ejercicio para la participación en la elección, con el mismo tamaño de muestra de 15 casillas por distrito federal y considerando la elección de diputados del 2021 se obtuvo que, con este tamaño de muestra, el margen de error para estimar la participación sería menor al 0.5 por ciento.

Presión sobre los CAE

Los CAE desempeñan un papel crucial en la organización de las elecciones en México. Contratados y capacitados por el Instituto Nacional Electoral (INE), los CAE tienen la responsabilidad de reclutar y capacitar a los funcionarios de casilla. Estos últimos son fundamentales en la administración del proceso de votación en las casillas durante el día de la elección. Sus tareas abarcan desde verificar la identidad de los votantes hasta entregar las papeletas electorales, supervisar el

proceso de votación, realizar el conteo de votos al cierre de las urnas y garantizar que todo el proceso se lleve a cabo de manera transparente y ordenada. La labor coordinada entre los CAE y los funcionarios de casilla es esencial para asegurar la integridad y eficiencia del proceso electoral.

Durante la jornada electoral, los CAE se dedican a brindar apoyo activo a los funcionarios de casilla ante cualquier eventualidad que pueda surgir durante el proceso. En promedio, cada CAE es responsable de respaldar a cuatro casillas electorales, adaptando esta cifra según la densidad poblacional: menor en áreas rurales y mayor en entornos urbanos.

Además de su función principal de respaldo a los funcionarios de casilla, los CAE asumen otras responsabilidades cruciales, tales como la entrega y recolección del material electoral, la coordinación con las autoridades electorales, la transmisión de información relevante sobre el desarrollo de las elecciones, y, en algunos casos, la colaboración en la transmisión de datos de votación de casillas seleccionadas para el conteo rápido. Esta versatilidad en sus labores refleja la importancia y amplitud del rol de los CAE en el eficiente desarrollo de las elecciones.

Dado que la participación de los CAE en el conteo rápido es solo una de las diversas actividades que llevan a cabo durante la jornada electoral y no necesariamente la más prioritaria, es imperativo planificar cuidadosamente el tamaño y diseño de la muestra. El objetivo es minimizar la cantidad de CAE que deban enviar información para el conteo rápido, asegurando que no se vean abrumados y puedan cumplir eficazmente su responsabilidad principal de respaldar a los funcionarios de casilla.

Utilizando la base de datos de casillas instaladas para la elección de diputados del 2021 se seleccionaron 5,000 muestras estratificadas por distrito federal electoral: considerando 15 casillas por estrato para un tamaño de muestra total de 4,500 casillas. Esta base de datos contaba con 162,809 casillas aprobadas para instalarse en la elección, así como 41,060 CAE que apoyarían a los funcionarios de casillas en ese proceso electoral.

Para cada una de las 5,000 muestras estratificadas se monitoreo

- El número de CAE que participarían en el Conteo Rápido. En promedio fue de 4,319 CAE; aproximadamente el 10% del total de CAE que apoyarían en la elección.
- De los CAE que participarían en el Conteo Rápido, la distribución de casillas que tendrían que reportar. Esto se muestra en la siguiente tabla.

CASILLAS A REPORTAR	1	2	3	4
% de CAE	95.89	4.03	0.08	0

Por lo que prácticamente 96% de los aproximadamente 4,319 CAE que participarían en el Conteo Rápido para la estimación de la conformación de la Cámara reportarían sólo una casilla.

Referencias

Cochran, W. G. (1977). *Sampling Techniques*, 3rd, John Wiley.

Lohr, S.L. (2021). *Sampling: Design and Analysis* (3rd ed.). Chapman and Hall/CRC.

CPEUM (2023). *Constitución política de los estados unidos mexicanos, Compendio de Legislación Nacional Electoral Tomo I*, INE.

LEGIPE (2023). *Ley General de Instituciones y Procedimientos Electorales, Compendio de Legislación Nacional Electoral Tomo II*, INE.

7.3.2.Procedimiento de estimación

7.3.2.1.Dr. Carlos Erwin Rodríguez Hernández-Vela y Lic. José Alberto Florencio Chávez

Estimación

El propósito principal de este documento es describir de manera general la estrategia para estimar la conformación de las Cámaras de Diputados y Senadores. Dichas conformaciones se determinan mediante los totales de votos, los cuales se recopilan a nivel de distrito federal para los Diputados, y a nivel estatal para los Senadores. Por lo tanto, resulta imprescindible realizar estimaciones de estos totales tanto a nivel de distrito federal como estatal, dependiendo de la cámara en cuestión. Utilizando estos totales, junto con las reglas establecidas para la asignación de asientos, podemos determinar la conformación de ambas cámaras. A posteriori, el análisis se centrará en la estimación de porcentajes de votación; tarea que se simplifica significativamente una vez se han calculado los sub-totales.

Tomando en cuenta lo discutido previamente, el enfoque de muestreo adoptado para la Cámara de Diputados utiliza los distritos federales como estratos, mientras que el diseño correspondiente a la Cámara de Senadores emplea los estados

como base. Este apartado detalla la metodología empleada para estimar inicialmente los totales por estrato. Posteriormente, estos totales se agregan a un nivel nacional, y, en la fase final, se procede a calcular los porcentajes correspondientes.

Denotando a y_{ij} como el número de votos en la casilla i -ésima en favor de la fuerza política j -ésima, entonces el total de votos a favor de la fuerza política j -ésima en la población está dado por

$$t_j = \sum_{i=1}^N y_{ij} = N \bar{y}_{Nj}, \quad (1)$$

en donde N representa el número total de casillas instaladas para la elección. Por lo tanto, las proporciones que se buscará estimar están dadas por

$$p_j = \frac{t_j}{\sum_{j=1}^k t_j}, \quad (2)$$

para $j = 1, 2, \dots, k$ (se asumirán k fuerzas políticas). Es importante notar, como lo indica la expresión (1), que es posible calcular las proporciones requeridas como función de los promedios poblacionales de votos en favor de cada candidato.

Como se señaló anteriormente, la selección de casillas, para ambas cámaras, se realizará mediante muestreos aleatorios estratificados. Por lo que la inferencia será en una primera instancia sobre los promedios de votos poblacionales para cada partido en cada estrato. Es fácil ver que se puede descomponer el promedio poblacional de votos en favor del candidato j -ésimo en el estrato h -ésimo en una parte observada y otra parte no observada, i.e.

$$\bar{y}_{N_h j} = \left(\frac{N_h - n_h}{N_h} \right) \bar{y}_{N_h - n_h j} + \left(\frac{n_h}{N_h} \right) \bar{y}_{n_h j},$$

la parte observada es la que se recupera vía la muestra aleatoria de n_h casillas, de un total de N_h casillas instaladas en el estrato h . En este caso es el promedio muestral de votos $\bar{y}_{n_h j}$ en favor del candidato j -ésimo en el estrato h y tendremos muestra en cada uno de los estratos: $y_{i,j,h}$ indicará el número de votos en la casilla i , en favor del candidato j en el estrato h . El promedio no observado está denotado como $\bar{y}_{N_h - n_h j}$.

Asumiendo que la muestra de votos en las casillas en cada estrato para el candidato j -ésimo se pueden describir mediante una distribución $N(y|\mu_{hj}, \sigma_{hj}^2)$. Y utilizando un prior NIX (Normal-Ji cuadrada inversa)

$$p(\mu_{hj}, \sigma_{hj}^2) \propto N(\mu_{hj}|\mu_0, \sigma_0^2/k_0) \chi^{-2}(\sigma_{hj}^2|v_0, \sigma_0^2), \quad (3)$$

esto indica que se usarán los mismos valores para todos los candidatos y en todos los estratos. Por lo que se estaría dando la misma información a priori acerca de las votaciones en favor de cada fuerza política.

La interpretación de los parámetros de la prior NIX es la siguiente:

- μ_0 representa la mejor estimación o creencia inicial sobre el valor verdadero de $\mu_{h,j}$ (la media del modelo normal).
- k_0 es un parámetro de precisión o concentración que controla cuánto se confía en la creencia inicial sobre μ_0 . Un valor mayor de k_0 indica una mayor certeza en la creencia inicial de que μ_0 esté cerca de μ_h . Por el contrario, un valor menor de k_0 indica una mayor incertidumbre.
- σ_0^2 es la mejor estimación o creencia inicial para σ_{hj}^2 (la varianza del modelo normal).
- ν_0 es el número de grados de libertad para la distribución ji-cuadrada inversa, que controla la forma de la distribución prior de la varianza σ_{hj}^2 . Cuanto mayor sea ν_0 , más se concentra la distribución prior alrededor de σ_0^2 , lo que indica una mayor confianza en la creencia inicial sobre la varianza. Un valor más bajo de ν_0 implica una mayor incertidumbre o dispersión en torno a σ_0^2 .

Es posible calcular la distribución predictiva para el promedio poblacional de votos en favor del candidato k -ésimo en el estrato h , la cual está dada por

$$\bar{y}_{N_{hj}} | \mathbf{y}_{1,j,h}, \dots, \mathbf{y}_{n_h,j,h} \sim t_{\nu_{n_h}}(\mu_{N_{j,h}}, \sigma_{N_{j,h}}^2), \text{ para } j = 1, 2, \dots, k$$

una distribución t de Student con ν_{n_h} grados de libertad, en donde

$$\mu_{N_{j,h}} = \left(\frac{n_h}{N_h}\right) \bar{y}_{n_{nj}} + \left(\frac{N_h - n_h}{N_h}\right) \mu_{n_{hj}},$$

$$\sigma_{N_{j,h}}^2 = \frac{(N_h - n_h)(N_h - n_h - k_{n_h})}{N_h^2} \frac{\sigma_{n_{hj}}^2}{k_{n_h}}$$

y

$$\mu_{n_{hj}} = \frac{k_0 \mu_0 + n_h \bar{y}_{n_{nj}}}{k_0 + n_h},$$

$$\sigma_{n_{hj}}^2 = \frac{1}{\nu_0 + n_h} \left(\nu_0 \sigma_0^2 + n_h s_{n_{nj}}^2 + \frac{n_h k_0}{k_0 + n_h} (\mu_0 - \bar{y}_{n_{nj}})^2 \right),$$

$$k_{n_h} = k_0 + n_h,$$

$$v_{n_h} = v_0 + n_h,$$

con $\bar{y}_{n_{nj}}$ el promedio de votos y $s_{n_{nj}}^2$ la varianza muestral de los votos en favor del candidato j -ésimo en el estrato h .

Estimación por intervalo de la conformación

Simulando $\bar{y}_{N_{hj}}$, para $h=1,2,\dots, L$ es posible estimar el total de votos en favor de cada una de las fuerzas políticas simplemente multiplicando por el número de casillas instaladas en cada estrato, como se indica en la ecuación (1), y usando las reglas para distribuir asientos de la Cámara de Senadores (o la Cámara de Diputados) se obtendría una conformación. Esto se repite $m = 5,000$ veces obteniendo el mismo número de conformaciones, entonces se obtienen los cuantiles del número de asientos para cada fuerza política simplemente calculando los cuantiles muestrales.

Estimación de proporciones

Los totales en los estratos se suman para estimar el total de votos en favor de cada fuerza política a nivel nacional. Entonces, sumando nuevamente se obtiene una estimación para el total de votos emitidos en la elección. Así, es trivial estimar las proporciones (2). Ya que estos pasos se repiten 5,000 veces, obteniendo 5,000 proporciones para las k fuerzas políticas. Para obtener la estimación por intervalo se calculan los cuantiles correspondientes.

Estimación bajo muestras incompletas

Resulta crucial establecer una estrategia efectiva para las estimaciones preliminares, teniendo en cuenta que la información completa de todas las casillas según los distintos estratos no estará disponible hasta aproximadamente las 10 de la noche, o quizás más tarde. No obstante, se espera que las primeras remesas de datos de votación comiencen a llegar entre las 6:30 pm y las 7 pm. Para abordar esta situación, se optó por el uso del prior NIX, permitiendo así realizar estimaciones incluso antes de tener datos completos por estrato. Es vital ajustar los parámetros $(\mu_0, k_0, \sigma_0^2, v_0)$ adecuadamente para evitar discrepancias en las estimaciones preliminares. Para asegurar la consistencia de los resultados, se llevaron a cabo análisis y experimentos destinados a alinear las estimaciones obtenidas mediante el prior NIX con aquellas derivadas del modelo de referencia

$$p(\mu_{hj}, \sigma_{hj}^2) \propto \sigma_{hj}^{-2}.$$

Esta consideración es especialmente relevante cuando se dispone de muestras en todos los estratos, dado que el modelo estadístico normal, utilizando la prioridad de referencia, presupone la existencia de datos de muestra en cada uno de los estratos. Bajo esta premisa, se procedió a determinar los siguientes valores:

$$(\mu_0, k_0, \sigma_0^2, v_0) = (100, 1, 1000, 1).$$

Entonces, a priori estamos diciendo que

- $\mu_0 = 100$, indica que, de acuerdo con nuestro conocimiento previo en cada casilla, de todos los estratos, el promedio de votos en favor de cada fuerza política será de 100.
- $k_0 = 1$, indica un alto grado de incertidumbre en que el promedio de votos en favor de cada fuerza política sea de $\mu_0 = 100$.
- $\sigma_0^2 = 1,000$, es nuestra mejor estimación previa acerca de σ_{hj}^2 (la varianza del modelo normal).
- $v_0 = 1$, indica un alto grado de incertidumbre en que la varianza de los votos en favor de cada fuerza política, en todos los estratos, sea de $\sigma_0^2 = 1,000$.

Referencias

Lohr, S.L. (2021). Sampling: Design and Analysis (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9780429298899>

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2013). Bayesian Data Analysis (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>

7.3.2.2. Dr. Luis Enrique Nieto Barajas y Mtro. Carlos Samuel Pérez Pérez

Metodología

El problema de estimación de la conformación de la cámara de senadores, desde un punto de vista estadístico, requiere de hacer estimaciones a dos niveles: primero, a nivel de cada uno de los 32 estados de la república, para los senadores elegidos por mayoría relativa; y en segundo lugar, a nivel nacional para la determinación de los 32 senadores por representación proporcional. Para el primer caso, se necesita tener representación muestral a nivel estado, y para el segundo, se necesita tener representación a nivel nacional. Es por esto que el

diseño muestral más apropiado para el problema en cuestión es un muestreo aleatorio estratificado por estado.

El reglamento de elecciones en sus artículos 416 y 417 establece los principios para la determinación de los senadores electos por mayoría relativa, y el artículo 425 establece los principios para la asignación de senadores por representación proporcional.

Aunque los criterios se basan en la votación total y los votos por partido, es posible representar los criterios en términos de las proporciones de votos. Para este efecto introducimos la siguiente notación.

Sea θ_{ij} la proporción de votos, relativa al listado nominal, a favor del candidato/partido j en el estrato (estado) i , con $j = 1, \dots, J$ e $i = 1, \dots, N$. J es el número total de candidatos (que incluye a los candidatos por partido político y los independientes) más dos categorías extras: los votos nulos y no registrados y las abstenciones. $N = 32$ es el número de estratos (estados). Sea n_i el listado nominal de votantes en el distrito i para $i = 1, \dots, N$, y $n = \sum_{i=1}^N n_i$ el tamaño de la lista nominal nacional. Entonces $\theta_j = \sum_{i=1}^N \frac{n_i}{n} \theta_{i,j}$ es la proporción de votos, relativa al listado nominal, a favor del candidato j a nivel nacional.

Para la determinación de los senadores ganadores por mayoría relativa definimos

$$mr_{i,j} = \begin{cases} 1, & \text{si } \theta_{i,j} > \theta_{i,l} \text{ } l \neq j, l = 1, \dots, J - 2 \\ 0, & \text{en otro caso} \end{cases}$$

una indicadora que nos dice si el candidato j ganó o no en el estado i , previa aplicación de los convenios de coalición. Este proceso se repite al considerar la primera y segunda fórmulas.

Finalmente, definimos $mr_j = \sum_{i=1}^N mr_{i,j}$ como el número de senadurías ganadas por el partido j por el principio de mayoría relativa.

Para la determinación de senadores ganadores por primera minoría definimos

$$pm_{i,j} = \begin{cases} 1, & \text{si } \theta_{i,j} \text{ es el segundo mayor de } \theta_{i,l} \text{ } l \neq j, l = 1, \dots, J - 2 \\ 0, & \text{en otro caso} \end{cases}$$

una indicadora que nos dice si el candidato j quedó o no en el segundo lugar del estado i , previa aplicación de los convenios de coalición usando la primera fórmula. Finalmente, definimos $pm_j = \sum_{i=1}^N pm_{i,j}$ como el número de senadurías ganadas por el partido j por el principio de primera minoría.

Para la determinación de senadores ganadores por primera minoría definimos primero $\lambda_j = \frac{\theta_j}{\sum_{l=1}^{J-2} \theta_l}$ la proporción de votos válidos (se excluyen las abstenciones,

los nulos y los candidatos no registrados), y $\eta_j = \frac{\theta_j}{\sum_{l=1}^{J-3} \theta_l I(\lambda_j > 0.03)}$ como la proporción después de excluir a los partidos que no alcanzan el 3% en la votación válida, los independientes y los nulos y no registrados.

El algoritmo consiste en calcular primero $rp_j = \lfloor 32\eta_j \rfloor$ los curules por representación proporcional cruda para el partido j . Posteriormente se calculan los decimales remanentes $32\eta_j - rp_j$ y se ordenan de mayor a menor y se les asigna una senaduría más a cada uno hasta completar 32.

Finalmente, el número de escaños por partido político NE_j resulta de sumar los escaños por mayoría relativa mr_j , primera minoría pm_j y los de representación proporcional rp_j . Es decir, $NE_j = mr_j + rp_j + pm_j$.

Para la estimación de los parámetros antes descritos, y en particular para estimar θ_{ij} , usaremos un enfoque bayesiano. En específico usaremos el modelo propuesto por Mendoza y Nieto-Barajas (2016) ¹⁰.

Dicho modelo supone que X_{ij}^k , denota los votos a favor del partido/candidato j en el estrato i en la casilla k . Esta variable observable sigue un comportamiento descrito por la distribución siguiente

$$X_{ij}^k | \theta_{ij}, \tau_{ij} \sim N\left(n_i^k \theta_{ij}, \frac{\tau_{ij}}{n_i^k}\right)$$

con $k = 1, \dots, c_i$, con c_i casillas en muestra en el distrito i , $i = 1, \dots, N$ y $j = 1, \dots, J$.

Donde n_i^k es la lista nominal para la casilla k en el distrito i y τ_{ij} es la precisión que se supone constante dentro del estrato e independiente de θ_{ij} , más aún X_{ij}^k independiente de $X_{ij}^{k'}$, para $j \neq j'$.

La verosimilitud para cada estrato $i = 1, \dots, N$ y cada candidato $j = 1, \dots, J$ está dada por

$$L(\theta_{ij}, \tau_{ij} | x_{ij}) \propto \tau_{ij}^{\frac{c_i}{2}} \exp\left\{-\frac{\tau_{ij}}{2} \sum_{k=1}^{c_i} \frac{1}{n_i^k} (x_{ij}^k - n_i^k \theta_{ij})^2\right\}$$

La distribución inicial para los parámetros desconocidos del modelo, $\theta_{i,j}$ y $\tau_{i,j}$ se considera independiente y no informativa, es decir

¹⁰ Mendoza, M. y Nieto-Barajas L.E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. Electoral Studies 43, 124–132.

$$p(\theta_{ij}, \tau_{ij}) = U(\theta_{ij}|0,1)Ga(\tau_{ij}|0.5, 0.05)$$

Así, usando el Teorema de Bayes, la distribución posterior resulta ser proporcional al producto de una normal truncada para θ_{ij} condicional en τ_{ij} y una distribución gamma para τ_{ij} . Los estimadores puntuales y por intervalo para θ_{ij} se aproximarán mediante simulaciones de la distribución posterior.

7.3.2.3. Equipo 2

Mtra. María Teresa Ortiz Mancera y Lic. Ana María Madrigal Gómez

Resumen

El conteo rápido consiste en generar estimaciones de los resultados de una contienda electoral, utilizando los votos registrados en una muestra. Su principal desafío radica en que la estimación se efectúa con muestras incompletas, donde la ausencia de datos no es aleatoria. El método de estimación que emplearemos es una regresión binomial negativa con estructura jerárquica, permitiendo que algunos coeficientes varíen según el estrato. Adicionalmente, implementaremos un procedimiento de ajuste por sesgo al estimar con muestras incompletas. El modelo proporciona intervalos de probabilidad con una cobertura aproximada del 95%, incluso frente a ciertos patrones de muestras sesgadas observados en elecciones previas.

Antecedentes

La idea del modelo que aplicaremos provino del modelo de regresión multinivel y postestratificación (ver, por ejemplo, [1]). En este modelo de regresión multinivel y postestratificación, (1) se segmenta a la población en celdas según combinaciones de distintos atributos, (2) se utilizan las muestras para obtener simulaciones posteriores de los parámetros, y (3) se calcula la variable de respuesta para cada celda, ponderando las estimaciones para realizar una estimación a nivel poblacional.

Llevaremos a cabo los pasos (1) y (2), pero en el paso (3), adoptaremos un enfoque basado en modelos (ver [2]), simulando la distribución posterior del número total de votos para cada candidato y para cada una de las casillas electorales del país. Así, emplearemos el modelo para predecir los valores no muestreados de la población. Es importante destacar que esto proporciona una estimación inmediata a nivel poblacional.

Para ejecutar el paso (1), ajustamos una regresión multinivel para cada casilla en función de ciertas covariables. Es decir, para cada candidato, modelamos el

número de votos en la casilla i , y_i , con una distribución de probabilidad $p(y|\theta_1, \theta_2)$, donde, para una parte de los parámetros, digamos θ_1 , ajustamos una regresión multinivel.

La ventaja de esta estructura de modelos multinivel radica en que, ante la ausencia de respuesta, atrae los parámetros de un grupo hacia la media del mismo. Como consecuencia, logramos un manejo uniforme de los datos faltantes en las muestras y una mejora en las propiedades de cobertura de intervalos cuando los datos de la muestra presentan sesgos.

La distribución p es binomial negativa. Esta distribución se adapta adecuadamente a los datos y su robustez contribuye a la del modelo, permitiéndole ajustarse eficazmente tanto a candidatos con una alta como con una baja cantidad de votos. Además, ofrece una capa adicional de protección ante posibles errores en la base de datos.

Modelo

A continuación, definimos el modelo que utilizaremos, que se obtuvo después de varias iteraciones del flujo de trabajo bayesiano en [4].

Denotamos con BN a la distribución binomial negativa con una parametrización directamente usando un parámetro de media, $\mu_i \in \mathbb{R}^+$, y un parámetro $\phi_i \in \mathbb{R}^+$. Para $y \in \mathbb{N}$,

$$\text{BN}(y|\mu_i, \phi_i) \propto \left(\frac{\mu_i}{\mu_i + \phi_i} \right)^y.$$

Nuestro primer objetivo es predecir el número total de votos a favor de cada candidato y el número total de votos nulos en cada casilla. Teniendo esas predicciones podemos deducir cualquier cantidad que sea de nuestro interés, por ejemplo, la votación total emitida, la participación, intervalos de confianza para la proporción de votos obtenida por cada candidato o la conformación de la Cámara de Diputados.

Por simplicidad de notación consideraremos a los votos nulos como un candidato extra. Sea p el número total de candidatos, modelamos el número de votos obtenidos para el candidato $j \in \{1, \dots, p\}$ y para la casilla i como

$$y_{i,j} \sim \text{NB} \left(\mu_{i,j} := n_i w_i \theta_{i,j}, \phi_{i,j} := \frac{n_i w_i \theta_{i,j}}{\kappa_{s(i),j}} \right).$$

Aquí $\theta_{i,j}$ representa la proporción de personas que votaron por j del total de votos en la casilla i , w_i representa la proporción de personas en la lista nominal que votaron en la casilla i y $\kappa_{s(i),j} > 0$ controla la sobredispersión en relación a la media $\mu_{i,j}$.

El vector $\theta_i = (\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,p})$ se define como

$$\theta_i = \text{softmax}(\beta_{s(i)}^{0,1} + \beta_1^T x_i, \dots, \beta_{s(i)}^{0,p} + \beta_p^T x_i),$$

donde x_i es la i -ésima fila de X , una matriz de covariables escaladas, y

$$\text{softmax}(z_1, \dots, z_p)_j = \frac{\exp(z_j)}{\sum_{i=1}^p \exp(z_i)}.$$

Los coeficientes β_1, \dots, β_p tienen una distribución inicial gaussiana centrada en cero y con una desviación estándar de 0.25. Sin embargo, para $k = 1, \dots, p$,

$$\beta_{s(i)}^{0,k} \sim N(\beta_{0,k}, \sigma_k),$$

añadiendo una estructura jerárquica dependiendo del estrato. Los detalles de las distribuciones iniciales pueden consultarse en [8] en donde, adicionalmente, se prueba la consistencia de la elección de todas las distribuciones iniciales.

Estimación

La metodología general de la estimación se describe a continuación. Inicialmente, para cada casilla que no forma parte de la muestra, generamos simulaciones de votos basadas en el modelo. Posteriormente, juntamos los valores observados en las casillas con los valores simulados, obteniendo así los votos totales de todas las casillas. A partir de estos votos, es directo calcular los resúmenes que sean necesarios para cada elección.

La estimación se hace con el software Stan, llamado desde R utilizando el paquete CmdStanR [6]. El código en Stan es legible, lo que hace que el trabajo sea más transparente. Stan también permite una serie de diagnósticos, lo que ayudó a obtener una inferencia confiable más rápida. La implementación se puede consultar y reproducir en su totalidad en el paquete de R quickcountmx [7].

Para asegurarnos de abarcar una diversidad de posibles resultados, realizamos una calibración bayesiana con los datos de las elecciones nacionales de 2018. La calibración bayesiana se enfoca en verificar que el modelo produzca inferencias adecuadamente calibradas desde un enfoque frecuentista; es decir, busca asegurar que el modelo entregue intervalos de credibilidad posterior con una cobertura cercana al 95% en muestreos repetidos. La calibración evidencia que, ante una proporción significativa de casillas no reportadas, el modelo por sí solo no es suficiente para ajustar los sesgos presentes en las muestras censuradas.

El problema del sesgo

El sistema de información comienza a operar a las 6 p.m. y, cada 5 minutos, recibe nueva información de la muestra. De esta manera, el sistema genera una secuencia de muestras parciales acumulativas, que son analizadas mediante métodos de estimación para monitorear la tendencia de los resultados. Los resultados del conteo rápido se publican utilizando una de estas muestras parciales, debido a que un porcentaje significativo de la muestra completa tarda en ser reportado.

La ausencia de datos en las muestras parciales no es aleatoria; factores como si la casilla se encuentra en una zona urbana o rural, las condiciones climáticas o la lista nominal, influyen en el momento en que una casilla llega a la muestra.

Realizamos un análisis exhaustivo de sesgo en elecciones anteriores y nos enfrentamos al desafío de mitigar este sesgo en todas las etapas de nuestro flujo de trabajo. Para llevar a cabo este análisis, necesitábamos un procedimiento para generar muestras parciales que se asemeje al verdadero proceso de generación de datos. Así, ajustamos modelos de supervivencia para los tiempos de llegada de las casillas en las elecciones de 2018. Los detalles se pueden consultar en [8].

Usamos el modelo de supervivencia para obtener una medida del sesgo en diferentes horas o con diferentes porcentajes de muestra obtenidos.

El ajuste que haremos al modelo dependerá del porcentaje de muestra observado, al que denotamos con p_{obs} . En primer lugar, estimamos la distribución de probabilidad del número total de votos de cada candidato en cualquier casilla con el modelo descrito. Después, a la hora de simular, en lugar de utilizar el vector θ_i , usamos una perturbación, dada por

$$\theta'_i = \text{softmax}(\theta_{i,1} + \eta_{i,1}, \dots, \theta_{i,p} + \eta_{i,p}),$$

donde η_1, \dots, η_p son variables aleatorias independientes e idénticamente distribuidas $N(0, 0.1(1 - p_{obs}))$.

Notemos que el tamaño de los intervalos de probabilidad producidos crece a medida que crece $\eta_{i,k}$ y, conforme nos acercamos a la muestra completa, $\eta_{i,k}$ se acerca a 0 para cualquier candidato k . El modelo, con el ajuste por sesgo, provee coberturas cercanas al 95% tanto con muestras completas como con muestras censuradas.

7.4. Elecciones de Gubernatura y Jefatura de Gobierno

7.4.1. Estratificación y tamaño de muestra

7.4.1.1. Chiapas

Dra. Lizbeth Naranjo Albarrán, Dra. Ruth Selene Fuentes García y
Act. Miguel Ángel Chong Rodríguez

Diseño de muestra para la elección de la Gubernatura en el estado de Chiapas

El diseño muestral elegido es muestreo estratificado aleatorio, donde los estratos son los Distritos Federales, la selección de la muestra en cada estrato se hará con muestreo aleatorio simple (sin reemplazo), y se distribuirá el tamaño de muestra estatal ($n=600$) a los estratos en forma proporcional al número de casillas.

Esta propuesta se basa en el análisis que se presenta a continuación, donde se evaluaron distintas estratificaciones y tamaños de muestra para las elecciones de gubernatura de Chiapas 2024. Se calculó el tamaño de muestra para diferentes precisiones, y se realizó un estudio de simulación para calcular la precisión de los intervalos de estimación para distintos tamaños de muestra. Se obtuvo también un aproximado del porcentaje de casillas por CAE.

Método

Usando un diseño de muestreo estratificado aleatorio, donde en cada estrato se seleccionan casillas con un muestreo aleatorio simple (sin reemplazo), con distribución proporcional de la muestra a los estratos; para estimar la proporción de votos para un candidato mediante un estimador de razón combinada.

Para cada una de las estratificaciones consideradas, se obtiene lo siguiente:

1. Determinación del tamaño de muestra.
 - a. Se calculan para distintos valores de precisión, usando un nivel de confianza del 95%.
2. Cálculo de la precisión con un estudio de simulación.
 - a. Para distintos tamaños de muestra, se simulan 500 muestras, utilizando como marco muestral los resultados de las elecciones de gubernatura de Chiapas 2018.
 - b. Se estiman los intervalos del 95% de confianza.
 - c. Se estiman las precisiones (márgenes de error de los intervalos de confianza).

3. Se evalúa la carga para los CAE.
4. Se revisa el tiempo de llegada de la muestra.

Estratificaciones evaluadas

Con la información de la elección de gobernador de Chiapas en el año 2018, y la definición de estratos de 2023, se estudiaron las siguientes estratificaciones:

1. Sin estratos (1 estrato).
2. Distritos Federales (13 estratos).
3. Distritos Locales (24 estratos).
4. Distritos Federales por Tipo de Sección (25 estratos): Urbano vs. No Urbano.

El candidato con mayor varianza fue el C3, el cual se tomará como referencia.

Resultados del tamaño de muestra

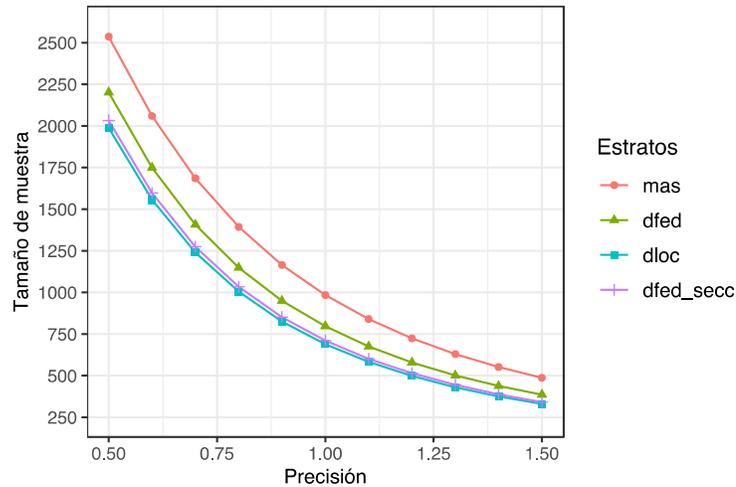
Utilizando diferentes tamaños de precisión, 0.5, 0.6, ..., 1.4, 1.5, se calcula el tamaño de muestra n .

Tabla 8 Tamaños de muestra n para el candidato con mayor varianza C3, para diferentes tamaños de precisión en las diferentes estratificaciones.

PRECISIÓN	M.A.S.	DTO.FED.	DTO.LOC.	DTO.FED. X TIPO SECC.
0.5	2536	2202	1988	2032
0.6	2060	1749	1558	1597
0.7	1685	1407	1240	1274
0.8	1393	1148	1004	1033
0.9	1165	950	826	851
1	984	797	689	711
1.1	840	676	583	601
1.2	724	580	498	515
1.3	630	502	431	445
1.4	552	438	375	388
1.5	487	386	330	341

Figura 15 Tamaños de muestra n para el candidato con mayor varianza C3, para diferentes tamaños de precisión en las diferentes estratificaciones.

Chiapas 2023, tamaño de muestra



Resultados de la precisión con simulación

Utilizando diferentes tamaños de muestra n , 300, 350, ..., 850, 900, se estima la precisión. En cada simulación se registra la precisión del candidato con mayor varianza, el cual se toma como referencia. Se grafica la media, y los cuantiles 0.05 y 0.95.

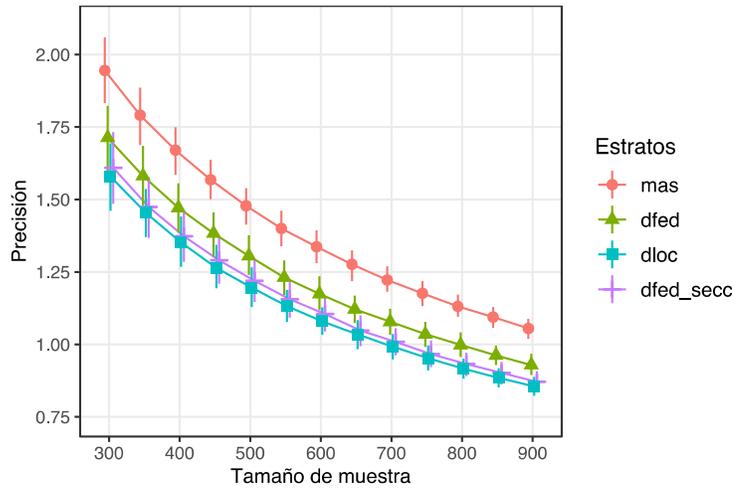
Tabla 9 Tamaños de precisión δ para el candidato con mayor varianza C3, para diferentes tamaños de muestra en las diferentes estratificaciones.

n	M.A.S.	DTO.FED.	DTO.LOC.	DTO.FED. X TIPO SECC.
300	1.95	1.71	1.58	1.61
350	1.79	1.58	1.45	1.47
400	1.67	1.47	1.35	1.37
450	1.57	1.38	1.26	1.29
500	1.48	1.31	1.20	1.22
550	1.40	1.23	1.13	1.16
600	1.34	1.17	1.08	1.10
650	1.28	1.12	1.03	1.05
700	1.22	1.08	0.99	1.01
750	1.18	1.04	0.95	0.97
800	1.13	1.00	0.92	0.93
850	1.09	0.96	0.88	0.90

n	M.A.S.	DTO.FED.	DTO.LOC.	DTO.FED. X TIPO SECC.
900	1.06	0.93	0.85	0.87

Figura 16 Tamaños de precisión δ para el candidato con mayor varianza C3, para diferentes tamaños de muestra en las diferentes estratificaciones.

Chiapas 2023, precisión

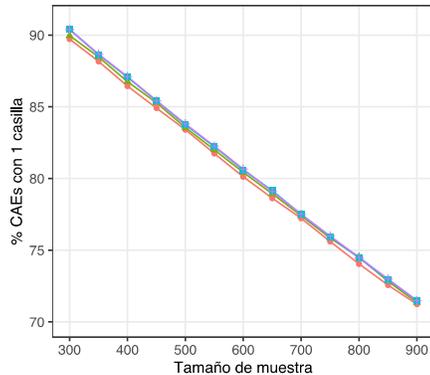


Resultados de la proporción de CAE

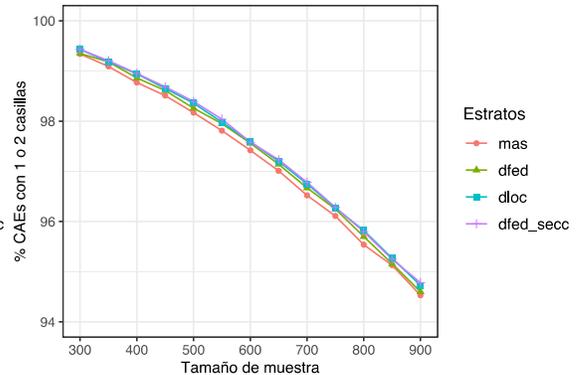
Utilizando diferentes tamaños de muestra n, se calcula el porcentaje de CAE que serán responsables de 1 casilla, y de 2 o más casillas.

Figura 17 Porcentaje de CAE con 1 sola casilla, y con 1 y 2 casillas, para diferentes tamaños de muestra en las diferentes estratificaciones.

Chiapas 2023, % CAEs con 1 casilla



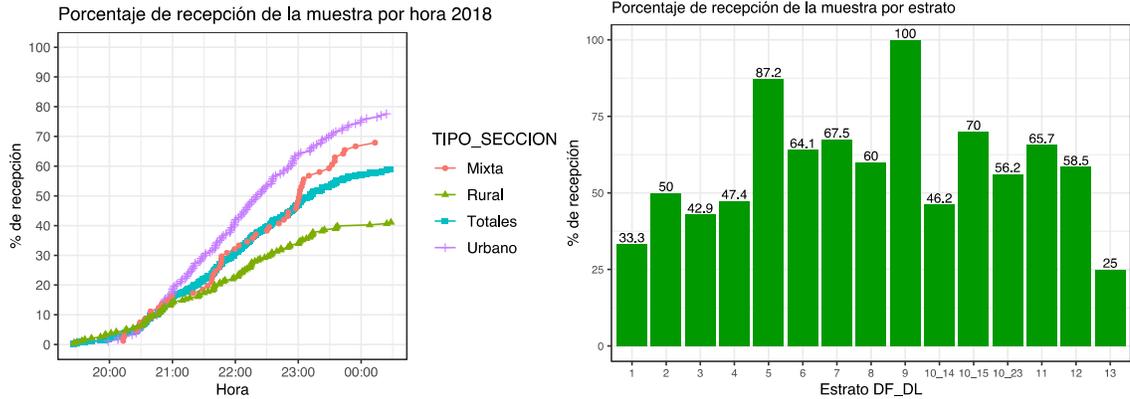
Chiapas 2023, % CAEs con 1 o 2 casillas



Resultados del tiempo de llegadas de la muestra

Revisamos los tiempos de llegada de la muestra.

Figura 18 Porcentaje de recepción de la muestra por hora y por estrato.



Determinación del tamaño de muestra y de la precisión

La determinación del tamaño de muestra, n , se realizó con la siguiente expresión:

$$n = \frac{Z_{(1-\alpha/2)}^2 N \times \sum_{h=1}^L N_h S_h^2}{X^2 \delta^2 + Z_{(1-\alpha/2)}^2 \times \sum_{h=1}^L N_h S_h^2}$$

El cálculo de la precisión, δ se realizó con la siguiente expresión:

$$\delta = \sqrt{\hat{V}(\hat{R}^p)} \times Z_{(1-\alpha/2)} \times 100$$

El estimador de razón combinado \hat{R}^p para la proporción de votos para el candidato p es:

$$\hat{R}^p = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} Y_{hi}^p w_h}{\sum_{h=1}^L \sum_{i=1}^{n_h} X_{hi} w_h}$$

Donde:

n : tamaño de muestra;

δ : precisión;

$Z_{(1-\alpha/2)}$: cuantil $1 - \alpha/2$ de una distribución Normal estándar;

N : tamaño de la población, número de casillas;

L : número de estratos;

N_h : número de casillas en el estrato h ;

w_h : factor de expansión para el estrato h , $w_h = N_h/N$;

n_h : número de casillas en muestra del estrato h ;

Y_{hi}^p : total de votos en casilla i del estrato h para el candidato p ;

\bar{Y}_h^p : promedio de votos por casilla en el estrato h para el candidato p ,

$$\bar{Y}_h^p = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi}^p$$

X_{hi} : total de votos emitidos en la casilla i del estrato h ;

\bar{X}_h : promedio de votos emitidos en el estrato h ,

$$\bar{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} X_{hi}$$

X : total de votos emitidos,

$$X = \sum_{h=1}^L \sum_{i=1}^{n_h} X_{hi}$$

S_h^2 : varianza del estrato h ,

$$S_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} [(Y_{hi}^p - \bar{Y}_h^p) - \hat{R}^p \times (X_{hi} - \bar{X}_h)]^2$$

$\hat{V}(\hat{R}^p)$: estimador de la varianza de \hat{R}^p ,

$$\hat{V}(\hat{R}^p) = \frac{1}{(\sum_{h=1}^L \sum_{i=1}^{n_h} X_{hi} w_h)^2} \times \sum_{h=1}^L N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2$$

Considerando una confianza del 95% y el cuantil de la Normal estándar $Z_{0.975} = 1.96$, el intervalo para la proporción de votos del candidato p se obtiene de la siguiente manera:

$$\left(\hat{R}^p - Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{R}^p)}, \hat{R}^p + Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{R}^p)} \right)$$

Bibliografía

Cochran, W. G. (1977). Sampling techniques. Wiley.

Lohr, S. L. (2011). Sampling: Design and Analysis. Brooks/Cole, 2nd edition.

Tabla 10 Tamaño de muestra Diseño M.A.S. de Casillas

PRECISIÓN	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
0.5	1057	1860	2536	2312	237
0.6	781	1445	2060	1850	167
0.7	597	1144	1685	1496	124
0.8	470	922	1393	1226	95
0.9	378	756	1165	1018	76
1	310	629	984	855	61
1.1	259	531	840	727	51
1.2	219	453	724	625	43
1.3	188	391	630	542	37
1.4	163	341	552	474	32
1.5	143	299	487	417	27

Tabla 11 Tamaño de muestra Diseño Estratificado (13 Distritos Federales) distribución proporcional.

PRECISIÓN	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
0.5	948	1670	2202	1928	210
0.6	696	1282	1749	1505	148
0.7	530	1006	1407	1195	109
0.8	415	806	1148	965	84
0.9	334	658	950	793	67
1	274	545	797	661	54
1.1	228	459	676	558	45
1.2	193	391	580	477	38
1.3	165	337	502	412	32
1.4	143	293	438	359	28
1.5	125	257	386	315	24

Tabla 12 Tamaño de muestra Diseño Estratificado (24 Distritos Locales) distribución proporcional.

PRECISIÓN	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
0.5	872	1550	1988	1767	202

PRECISIÓN	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
0.6	637	1181	1558	1365	142
0.7	483	922	1240	1076	105
0.8	378	736	1004	864	81
0.9	303	599	826	707	64
1	248	495	689	587	52
1.1	207	416	583	495	43
1.2	175	354	498	422	36
1.3	150	305	431	364	31
1.4	130	265	375	317	27
1.5	113	232	330	278	23

Tabla 13 Tamaño de muestra Diseño Estratificado (25 Distritos Federales x Tipo Sección, Urbano y No Urbano) distribución proporcional.

PRECISIÓN	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
0.5	924	1622	2032	1788	202
0.6	677	1242	1597	1383	142
0.7	515	972	1274	1091	105
0.8	403	777	1033	877	81
0.9	324	634	851	718	64
1	265	525	711	597	52
1.1	221	441	601	503	43
1.2	187	376	515	429	36
1.3	160	324	445	370	31
1.4	139	281	388	322	27
1.5	121	247	341	283	23

Tabla 14 Tamaño de precisión Diseño M.A.S. de Casillas.

N	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
300	1.02	1.51	1.95	1.78	0.43
350	0.94	1.37	1.79	1.65	0.40
400	0.87	1.28	1.67	1.53	0.37
450	0.82	1.20	1.57	1.44	0.35
500	0.77	1.13	1.48	1.36	0.33
550	0.73	1.07	1.40	1.29	0.31

N	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
600	0.70	1.02	1.34	1.23	0.30
650	0.67	0.98	1.28	1.17	0.28
700	0.64	0.94	1.22	1.12	0.27
750	0.61	0.91	1.18	1.08	0.27
800	0.59	0.87	1.13	1.04	0.25
850	0.57	0.84	1.09	1.00	0.25
900	0.55	0.81	1.06	0.97	0.24

Tabla 15 Tamaño de precisión Diseño Estratificado (13 Distritos Federales) distribución proporcional.

N	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
300	0.94	1.37	1.71	1.54	0.40
350	0.87	1.26	1.58	1.42	0.38
400	0.82	1.18	1.47	1.31	0.35
450	0.77	1.11	1.38	1.24	0.33
500	0.72	1.05	1.31	1.17	0.31
550	0.68	0.99	1.23	1.11	0.30
600	0.65	0.94	1.17	1.05	0.28
650	0.62	0.90	1.12	1.01	0.27
700	0.60	0.87	1.08	0.96	0.26
750	0.57	0.83	1.04	0.93	0.25
800	0.55	0.80	1.00	0.90	0.24
850	0.53	0.78	0.96	0.86	0.23
900	0.52	0.75	0.93	0.83	0.22

Tabla 16 Tamaño de precisión Diseño Estratificado (24 Distritos Locales) distribución proporcional

N	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
300	0.91	1.30	1.58	1.44	0.40
350	0.83	1.20	1.45	1.33	0.36
400	0.77	1.11	1.35	1.23	0.34
450	0.73	1.05	1.26	1.15	0.32
500	0.69	0.99	1.20	1.10	0.30
550	0.65	0.93	1.13	1.04	0.29

N	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
600	0.62	0.89	1.08	0.98	0.27
650	0.59	0.86	1.03	0.94	0.26
700	0.57	0.83	0.99	0.91	0.25
750	0.55	0.79	0.95	0.87	0.24
800	0.52	0.76	0.92	0.84	0.23
850	0.51	0.73	0.88	0.81	0.23
900	0.49	0.71	0.85	0.78	0.22

Tabla 17 Tamaño de precisión Diseño Estratificado (25 Distritos Federales x Tipo Sección, Urbano y No Urbano) distribución proporcional.

N	PAN_PRD_MC	PRI_NA	PT_MORENA_ES	PVEM_PCU_MVC	CAND_IND1
300	0.94	1.34	1.61	1.45	0.40
350	0.86	1.25	1.47	1.33	0.36
400	0.80	1.17	1.37	1.24	0.34
450	0.75	1.08	1.29	1.16	0.32
500	0.70	1.02	1.22	1.10	0.30
550	0.67	0.97	1.16	1.05	0.29
600	0.64	0.93	1.10	1.00	0.28
650	0.61	0.89	1.05	0.95	0.26
700	0.59	0.85	1.01	0.91	0.25
750	0.56	0.81	0.97	0.87	0.24
800	0.55	0.78	0.93	0.84	0.23
850	0.52	0.76	0.90	0.82	0.22
900	0.51	0.74	0.87	0.79	0.22

7.4.1.2.Ciudad de México

Dr. Raúl Rueda Díaz del Campo y Lic. Javier Santibáñez Cortés.

Con la información recibida en el archivo CDMX 2018 conARE conGeoref2023.txt

- Lista nominal 7,626,537
- Número de casillas 12,971
- Número de distritos federales (DF2018) 24; (DF2023) 22

- Número de distritos locales (DL2018) 33
- Hubo redistribución en el periodo de 2018 a 2023
- Número de ARE 3,120 a los que les corresponden entre 2 y 6 casillas

Elección de gobernador CDMX 2018

Siete candidatos:

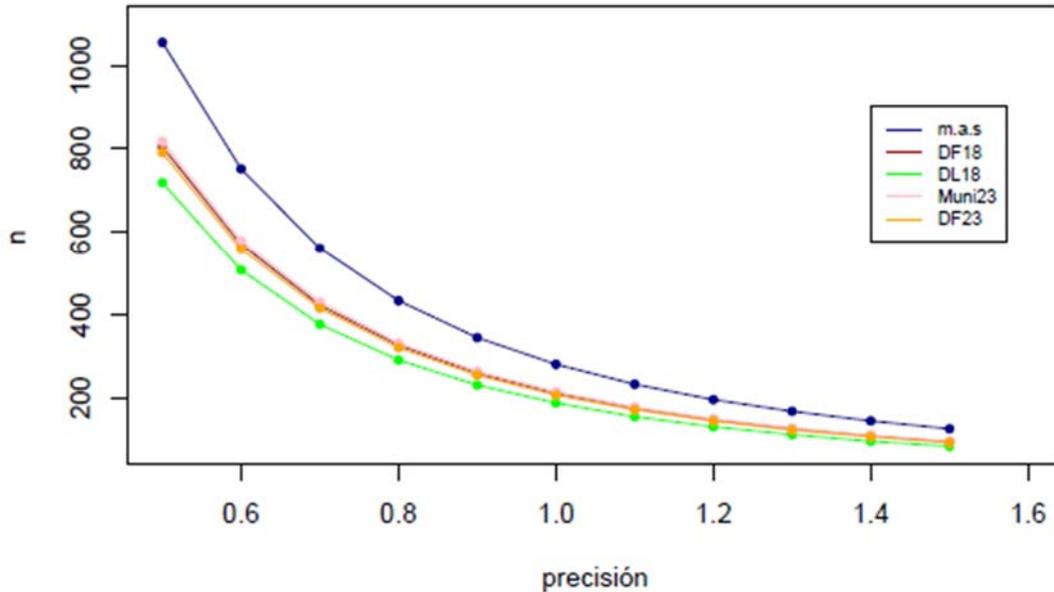
- C1. Coalición:
 - Partido Acción Nacional (PAN)
 - Partido de la Revolución Democrática (PRD)
 - Movimiento Ciudadano (MC)
- C2. Partido Revolucionario Institucional (PRI)
- C3. Partido Verde Ecologista de México (PVEM)
- C4. Coalición:
 - Movimiento Regeneración Nacional (MORENA)
 - Partido del Trabajo (PT)
 - Partido Encuentro Social (PES)
- C5. Partido Nueva Alianza (PNAL)
- C6. PH
- C7. Candidatos Independiente

Con la información de la elección de gobernador de CDMX en el año 2018, se estudiaron las siguientes estratificaciones (en todos los casos se selecciona la muestra con m.a.s.):

- sin estratificar (solo como referencia)
- DF 2018 (24 estratos)
- DF 2023 (22 estratos)
- DL 2018 (33 estratos)
- Municipios 2023 (16 estratos)

Se calculó el tamaño de muestra necesario con diferentes precisiones para cada una de las estratificaciones descritas y un nivel de **confianza del 95%**.

CDMX, tamaño de muestra



Con una **precisión** aproximada de **0.60%**, y considerando solo los candidatos con mayor varianza, el tamaño de muestra para cada estratificación es:

Estratificación	C1	C4
DF18	339	570
DL18	290	508
DF23	316	596
Muni23	291	579

Diseño de muestra

Muestreo estratificado aleatorio, donde los estratos son los **distritos federales**, la selección de la muestra en cada estrato es con muestreo aleatorio simple (sin reemplazo) y se distribuirá el tamaño de muestra estatal (**n=600**) a los estratos en forma proporcional al número de casillas que tiene cada distrito federal.

Análisis para la estratificación por Distrito Federal

- Información de N = 12;971 casillas electorales de la elección para gobernador 2018.
- Se seleccionaron 1,000 muestras con el diseño especificado.
 - 24 Distritos Federales, n = 600
- Se calculó el promedio de los estimadores de razón combinado y de la precisión.
- Se calculó la cobertura de los intervalos del 95% de confianza.
- Se calculó el porcentaje de ARE que les corresponde entre 1 y 5 casillas en muestra.

Análisis para la estratificación por Distrito Federal

	$\bar{R}\%$	Precisión%	Cobertura
C1	31.0	0.44	94.5
C2	12.8	0.39	93.9
C3	3.8	0.11	95.1
C4	47.1	0.58	95.1
C5	0.7	0.05	94.9
C6	.95	0.04	95.8
C7	1.2	0.06	94.2

Análisis para la estratificación por Distrito Federal n=600

	Casillas				
	1	2	3	4	5
ARES	92.73%	7.01%	0.25%	0.0%	0.0%

7.4.1.3. Guanajuato

Dra. Marta Elva Ramírez Guzman y Lic. Gabriela López Reyes

Objetivo del Conteo Rápido

El objetivo del Conteo Rápido para 2024, es la estimación por intervalo del porcentaje de votación de cada una de las candidaturas de las gubernaturas, así como la estimación por intervalo del porcentaje de participación ciudadana. Una manera para estimar dicho porcentaje es estimar una proporción o razón (R), del número de votantes a favor de un cierto candidato entre el total de votación, para luego multiplicarlo por 100 y de esta manera obtener el porcentaje de interés. Para estimar esta proporción, existen diferentes diseños muestrales. A continuación se describen los más comunes.

Tipos de diseño muestral

Existen varios diseño de muestreo para estimar R , algunos más populares son: el diseño completamente al azar (MAS) y el diseño estratificado. Cuando la población a muestrear es homogénea, es suficiente con aplicar un diseño completamente al azar, cuando esto no sucede, es importante realizar una estratificación de la población con objeto de identificar subpoblaciones cuyos miembros que las componen se caracterizan por presentar homogeneidad respecto a la variable de interés. Para el caso que nos ocupa, la estratificación de la población podría ser compuesta por Distritos Electorales Federales (DF), Distritos Electorales Locales (DL), Tipos de Sección (TS) y las combinaciones de estos. Cabe destacar que estas distritaciones electorales, tienen el objetivo de garantizar una representación política equilibrada de las personas habitantes en cada distrito electoral.

Con estas estratificaciones, se podrá seleccionar el tamaño de muestra, tal que su error de precisión fuera del 1% cuando mucho y con una confianza del 95% de probabilidad, restricciones acordadas por COTECORA y representantes del INE.

Otra restricción a considerar es la carga de trabajo para el Capacitador Asistente Electoral (CAE), ya que con objeto de minimizar su carga de trabajo se ha instituido por el INE y COTECORA, que la asignación de una sola casilla por CAE para el trabajo de Conteo Rápido sea de al menos del 80%. Por lo tanto, estas tres restricciones deben ser consideradas en la propuesta del esquema de muestreo para la Gubernatura de Guanajuato.

Es importante mencionar que entre mayor es el grado de estratificación (intersección de variables de clasificación), menor será el número de casillas en el estrato, lo cual implica el riesgo de que dicho estrato no quede representado por la falta de envío de la información por los Capacitadores Asistentes

Electorales (CAE) o Supervisor Electoral (SE's). Por lo anterior, es fundamental definir estratos en donde exista un mínimo de casillas electorales. De acuerdo al documento intitulado: Criterios científicos, logísticos y operativos para la realización de los Conteos Rápidos y protocolo para la selección de las muestras COTECORA 2017-2018 de 2018 (mayo de 2018) y los respectivos reportes de los miembros del COTECORA para dicha elección, la cantidad mínima de estratos fue de 15, 30 o 50 casillas por estrato, en donde la primera cantidad fue para estados con poca población como Yucatán, la segunda para Guanajuato y la tercera para Veracruz.

Tamaño de muestra

Para seleccionar el tamaño de muestra óptimo, se procederá a identificar qué tamaño de muestra cumple la restricción de producir un error de estimación, que cumpla con: $d \leq 1\%$ con una probabilidad del 95% de confianza. El tamaño de muestra que cumpla con esta restricción bajo diferentes esquemas de estratificación, será un tamaño de muestra adecuado. Este tamaño de muestra depende de la variabilidad del número de votantes en el estrato h ($h = 1, 2, \dots, L$). Dicha variabilidad se estima como: $S_{hp}^2 = Y_{hp} - R_{hp}X_h$, en donde Y_{hp} es el total de votos a favor del candidato p en el estrato h y X_h es el total de votos en el estrato h y R_{hp} es la proporción de votos a favor del candidato p en el estrato h en una muestra piloto (datos distritales de 2018).

Bases de datos

El archivo denominado: Gubernatura2018GuanajuatoCasilla_conDto2023, fue la base de datos utilizada para estimar una muestra piloto, la cual se utilizó para estimar la variabilidad del número de votantes en el estrato h (S_h^2) y para candidato. Posteriormente, esta variabilidad fue incluirla en la estimación del tamaño de muestra con y sin estratificación para el Conteo Rápido de 2024. Esta base de datos consistió de 7,479 casillas, sin embargo, al remover los registros de casillas sin identificación de DF, DL y TC la base final con la que se trabajó fue de 7,171 registros.

Resultados

Identificación de una estratificación adecuada

Con objeto de identificar la mejor estratificación que fuera representativa del estado de Guanajuato, en función de la base de datos de 2018, el Cuadro 4, muestra el total de estratos, el número mínimo, el máximo y el promedio de casillas

para las estratificaciones: TS, DF, DL, intersección de DF y TS ($DF \cap TS$), intersección de DL y TS ($DL \cap TS$) e intersección de DF y DL ($DF \cap DL$). Programa: G_Estratificacion_BoxPlots.R.

Cuadro 4 Cómputos distritales elección de gobernador 2012: Estadísticas básicas para cada estratificación y las casillas instaladas en cada estrato.

ESTATIFICACIÓN	NÚMERO DE ESTRATOS	MAXIMO DE CASILLAS X ESTRATO	MÍNIMO DE CASILLAS X ESTRATO	PROMEDIO DE CASILLAS X ESTRATO
MAS	1	7,171	7,171	7,171
TS	3	4,445	690	2,390
DF	15	619	385	478
DL	22	457	197	326
$DF \cap TS$	45	442	5	159
$DL \cap TS$	63	457	4	114
$DF \cap DL$	41	444	4	175

Como podemos observar, la estratificación definida por DL es la que mantiene las mejores condiciones, debido a que el número mínimo de casillas (197) sobrepasa el mínimo de casillas recomendado por el COTECORA (en promedio 30 casillas), para la elección gubernamental del 2018, así también, respecto al número de estratos, podemos observar que se encuentra en un punto intermedio, esto es, no tiene tan pocos (entre 1 y 3 estratos), como las estratificaciones correspondientes a MAS y TA, ni tantos como las respectivas intersecciones: $DF \cap TS$, $DL \cap TS$ y $DF \cap DL$. Sin embargo, dado que un acuerdo de COTECORA fue armonizar el tamaño de muestra de gubernaturas que estarán sujetas a votación, igual que la presidencial, se optó por anidar las casillas que constituirán la muestra para presidente dentro de las casillas que constituirán la muestra para la gubernatura. Por lo tanto, toda la energía se enfocó en estimar el tamaño de muestra en función de los estratos de acuerdo a DF. Sin embargo, para fines comparativos también se trabajó con las intersecciones de DF y TS, así como DF y DL, para definir el tamaño de muestra.

Precisiones

A continuación (Cuadro 5), se muestran las precisiones estimadas mediante simulación bootstrap con $B=1,000$ remuestreos, para los tres partidos punteros en la elección del 2018, en función del tamaño de la muestra: $n=200, 300, 400, 500$ y 600 . El denominador considerado para estimar la proporción de ciudadanos que votaron por un partido fue el total de votos, incluyendo candidatos no registrados y nulos. En el mismo cuadro, se muestran las precisiones derivadas para la

participación ciudadana, para ésta el denominador considerado fue el número de personas registrados en la Lista Nominal (LN). Estos tamaños se definieron por ser los que generaban porcentajes de precisión alrededor del 1%, valor definido por COTECORA e INE.

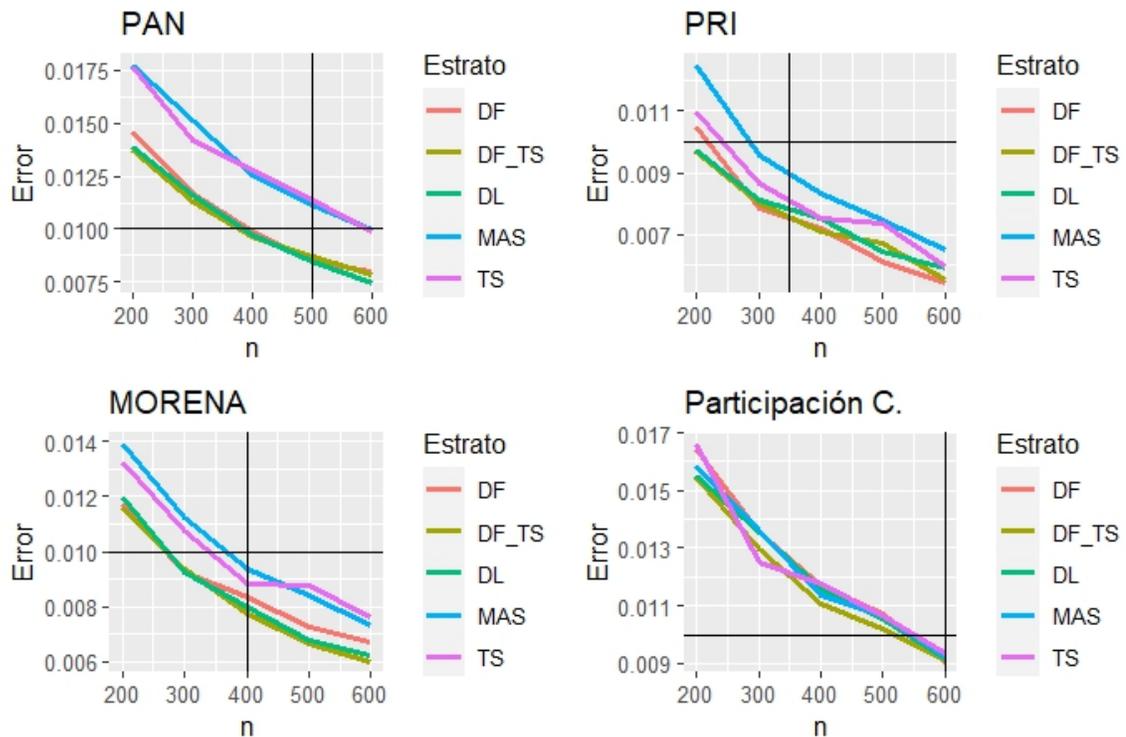
Cuadro 5 Precisiones en porcentaje de la estimación para los tres partidos punteros de 2018 con estratificación por: DF, DL, TS, DF_TS y MAS.

N	PAN	PRI	MORENA	PARTICIPACIÓN
DF				
200	0.01456	0.01048	0.01175	0.01645
300	0.01172	0.00784	0.00926	0.01356
400	0.0099	0.0072	0.00832	0.01173
500	0.00849	0.00615	0.00724	0.01075
600	0.00796	0.00545	0.00674	0.00907
DL				
200	0.01392	0.00973	0.01194	0.01551
300	0.01159	0.0081	0.00926	0.01355
400	0.00973	0.00752	0.008	0.01158
500	0.0085	0.00646	0.00676	0.01054
600	0.00747	0.00592	0.00622	0.00918
TS				
200	0.01768	0.01095	0.01324	0.01663
300	0.01422	0.00863	0.01076	0.01251
400	0.01277	0.00753	0.00885	0.01177
500	0.01134	0.00734	0.00878	0.01068
600	0.00986	0.00595	0.00764	0.00935
DF_TS				
200	0.01373	0.00969	0.01159	0.01543
300	0.01132	0.00801	0.00934	0.01297
400	0.00968	0.00711	0.00771	0.01108
500	0.00867	0.00673	0.00665	0.0102
600	0.00783	0.00556	0.00601	0.00913
MAS				
200	0.01777	0.01249	0.01392	0.01584
300	0.01514	0.00957	0.01123	0.01358
400	0.01251	0.00831	0.00939	0.01138
500	0.01115	0.00745	0.00838	0.01065

N	PAN	PRI	MORENA	PARTICIPACIÓN
600	0.00995	0.00651	0.00734	0.00923

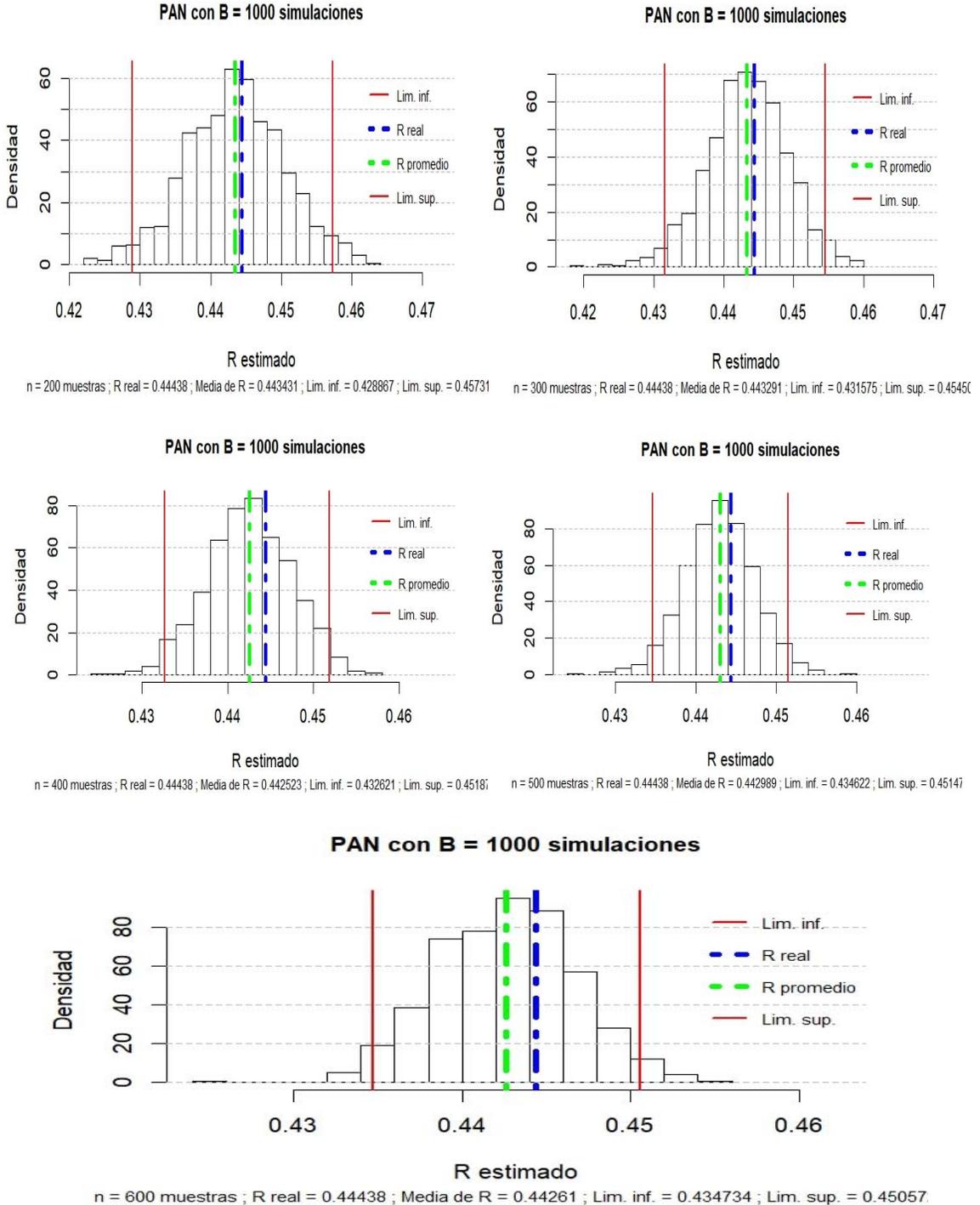
El PAN, que es el partido con mayor variabilidad de votación, muestra que a partir de $n=500$, la precisión se mantiene por debajo de 1%, considerando DF como estratos, en consecuencia, los partidos PRI y MORENA presentan mejores precisiones con $n=500$. Respecto a la participación ciudadana, esta gráfica muestra que con un tamaño de muestra entre 500 y 600 alcanza en promedio la precisión mencionada (Gráfica 1).

Gráfica 1 Errores de precisión y tamaño de muestra asociados a los partidos punteros de 2018.

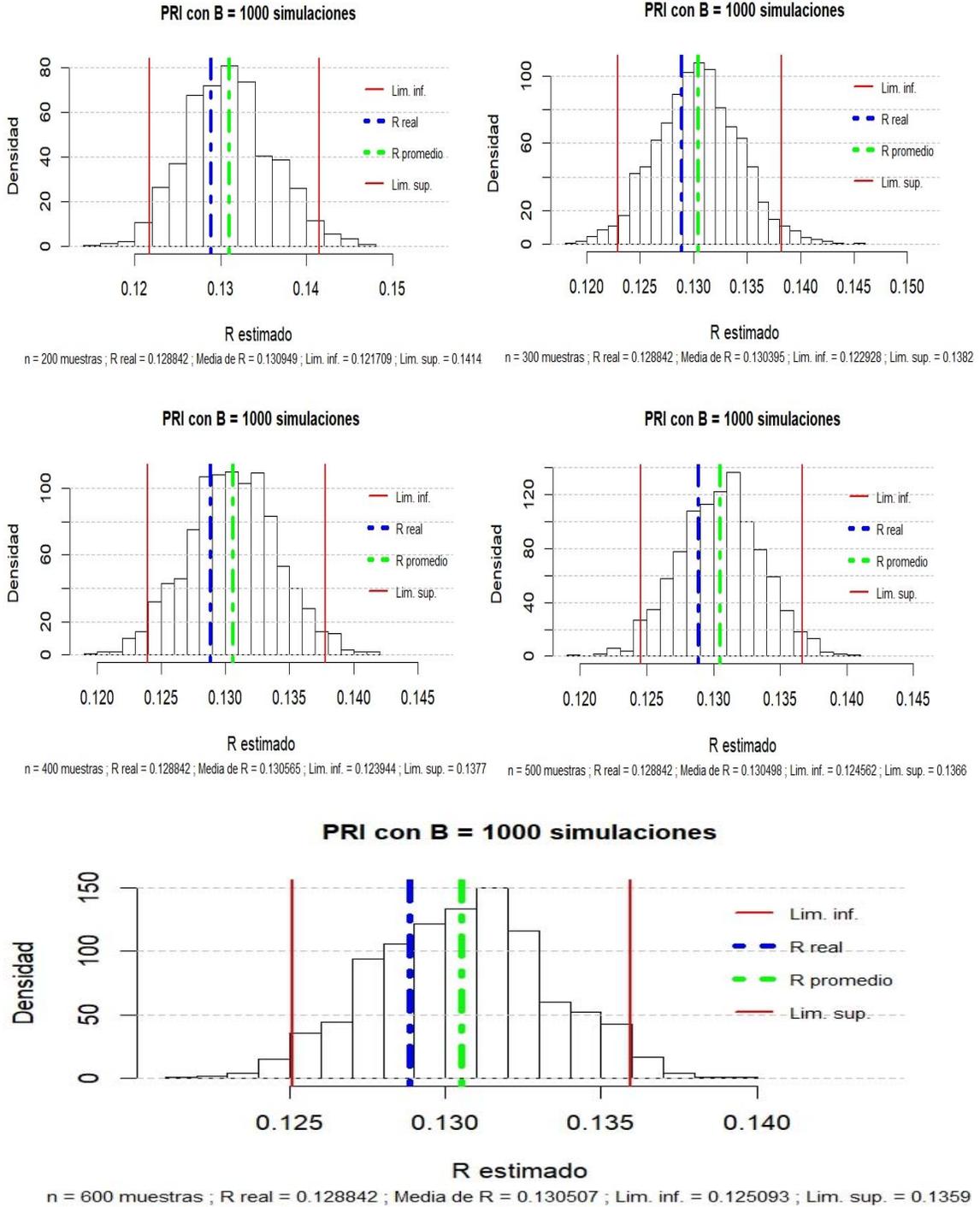


Las Gráficas 2, 3 y 4 muestran los histogramas de las simulaciones bootstrap para los partidos de PAN, PRI y MORENA, respectivamente, para tamaños de muestra de 200, 300, 400, 500 y 600, respectivamente. El número de simulaciones para cada muestra fue de 1,000. En estas gráficas se aprecia que el valor real (línea azul) de la proporción de votantes (R_p) "cae dentro" del intervalo del 95% de confianza (color rojo) del estimador de razón combinado (\hat{R}_p), que se muestra en color verde. Podemos apreciar que la diferencia entre R_p y \hat{R}_p es del orden de milésimas. Programa: G_GraficaPrecisionesPorPartido.

Gráfica 2 Estimación bootstrap de la fracción de votación para el PAN, con DF como estratos, con cuantiles de probabilidad 0.025 y 0.975 (color rojo) con B=1,000 simulaciones.

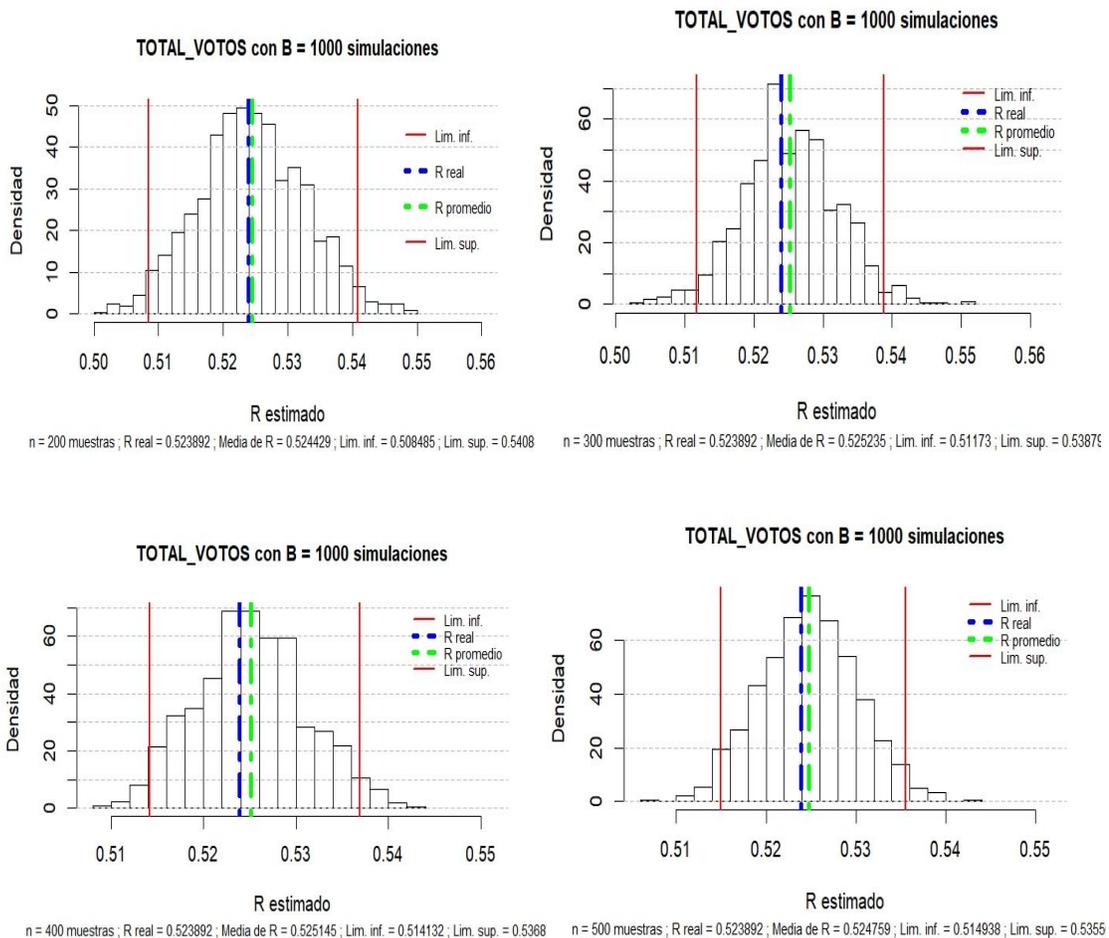


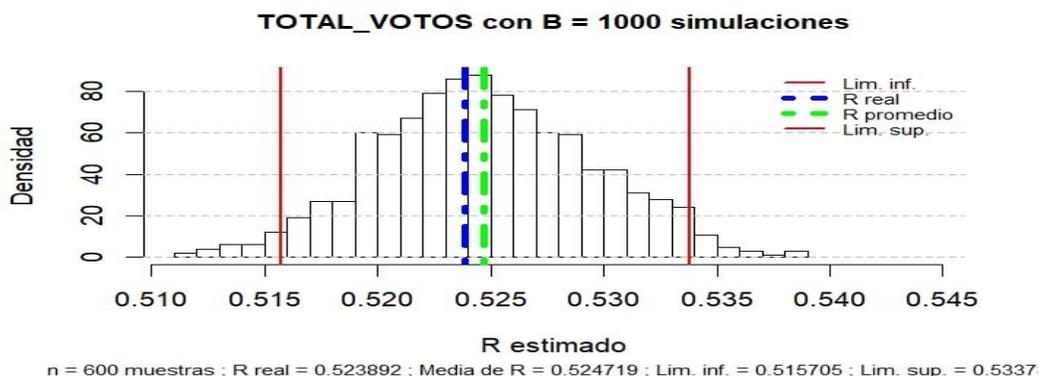
Gráfica 3 Estimación bootstrap de la fracción de votación para el PRI, con DF como estratos, con cuantiles de probabilidad 0.025 y 0.975 (color rojo) con B=1,000 simulaciones.



La Gráfica 5 muestra los histogramas de las simulaciones bootstrap para el porcentaje de Participación Ciudadana, para tamaños de muestra de 200, 300, 400, 500 y 600, respectivamente. El número de simulaciones para cada muestra fue de 1,000. En esta gráfica se aprecia que el valor real (línea azul) de la proporción de votantes (R_p) “cae dentro” del intervalo del 95% de confianza (color rojo) del estimador de razón combinado (\hat{R}_p), que se muestra en color verde. Podemos apreciar que la diferencia entre R_p y \hat{R}_p es del orden de milésimas. Programa: G_GraficaPrecisionesPorPartido.

Gráfica 5 Estimación bootstrap de la fracción de votación para Participación Ciudadana, con DF como estratos, con cuantiles de probabilidad 0.025 y 0.975 (color rojo) con B=1,000 simulaciones.





Si bien, con $n=500$ se espera una buena precisión ($d < 1\%$) para los candidatos punteros, así como para la participación ciudadana. Sin embargo, si por diversas razones llegara solo el 80% (400) de la muestra, aún así se tendría una precisión por debajo del 1% en promedio, para los partidos punteros y 1.17% para la estimación de la Participación Ciudadana. Si llegara sólo el 60% ($n=300$) de la muestra se esperarían precisiones de 1.17, 0.78 y 0.92%, para PAN, PRI y MORENA, respectivamente y 1.35% para proporción de la participación ciudadana (Cuadro 5).

Tamaño de muestra por tipo de estratificación

Los Cuadro 6, 7, 8, 9 y 10 muestran el tamaño de muestra por tipo de estratificación.

Cuadro 6 Tamaño de muestra sin considerar estratos (MAS).

ESTRATO	NI	200	300	400	500	600
1	7171	200	300	400	500	600

En particular el tamaño de muestra para la estratificación con DF, oscila entre 27 y 43 casillas, lo cual implica un cierto equilibrio entre los estratos (Cuadro 7).

Cuadro 7 Tamaño de muestra considerando los DF como estratos.

ESTRATO	NI	200	300	400	500	600
1	463	13	19	26	32	39
2	499	14	21	28	35	42
3	471	13	20	26	33	39
4	439	12	18	24	31	37
5	385	11	16	21	27	32

ESTRATO	NI	200	300	400	500	600
6	389	11	17	22	27	33
7	414	12	17	23	29	35
8	513	14	21	29	35	43
9	452	13	19	25	32	38
10	619	17	26	35	43	51
11	446	12	19	25	31	37
12	519	14	22	29	36	43
13	596	17	25	33	42	50
14	552	15	23	31	38	46
15	414	12	17	23	29	35

Cuadro 8 Tamaño de muestra considerando los DL como estratos.

ESTRATO	NI	200	300	400	500	600
1	374	10	16	21	25	31
2	321	9	13	18	22	27
3	227	6	10	13	16	19
4	205	6	10	11	14	18
5	197	5	8	11	14	16
6	345	10	14	19	24	29
7	457	13	19	25	32	38
8	317	9	13	18	22	27
9	309	9	13	17	22	26
10	339	9	14	19	24	28
11	332	9	14	19	23	28
12	311	9	13	17	22	26
13	367	10	15	20	26	31
14	342	10	14	19	24	29
15	363	10	15	20	25	30
16	245	7	10	14	17	20
17	310	9	13	17	22	26
18	376	10	16	21	26	31
19	340	9	14	19	24	28
20	390	11	16	22	27	33

ESTRATO	NI	200	300	400	500	600
21	260	8	11	15	18	22
22	444	12	19	25	31	37

Cuadro 9 Tamaño de muestra considerando los TS como estratos.

ESTRATO	NI	200	300	400	500	600
M	690	19	29	38	48	58
R	2036	57	85	114	142	170
U	4445	124	186	248	310	372

Cuadro 10 Tamaño de muestra considerando DF \cap TS como estratos, con estratificaciones colapsadas (Estrato 25: con menos de 50 casillas y Estrato 26: con casillas entre 51 y 99). Sin colapsar dio origen a 57 estratos, con cinco casillas como mínimo.

	ESTRATO	NI	200	300	400	500	600
1	1_R	254	7	11	14	18	21
2	1_U	146	4	6	8	10	12
3	10_R	171	5	7	10	12	14
4	10_U	405	11	17	23	28	34
5	11_U	389	11	16	22	27	33
6	12_U	442	12	18	25	31	37
7	13_R	310	9	13	17	22	26
8	13_U	212	6	9	12	15	18
9	14_R	255	7	11	14	18	21
10	14_U	223	6	9	12	16	19
11	15_U	337	9	14	19	23	28
12	2_R	238	7	10	13	17	20
13	2_U	201	6	8	11	14	17
14	3_U	436	12	18	24	30	36
15	4_R	199	6	8	11	14	17
16	4_U	195	5	8	11	14	16
17	5_U	374	10	16	21	26	31
18	6_U	326	9	14	18	23	27
19	7_R	186	5	8	10	13	16
20	7_U	187	5	8	10	13	16

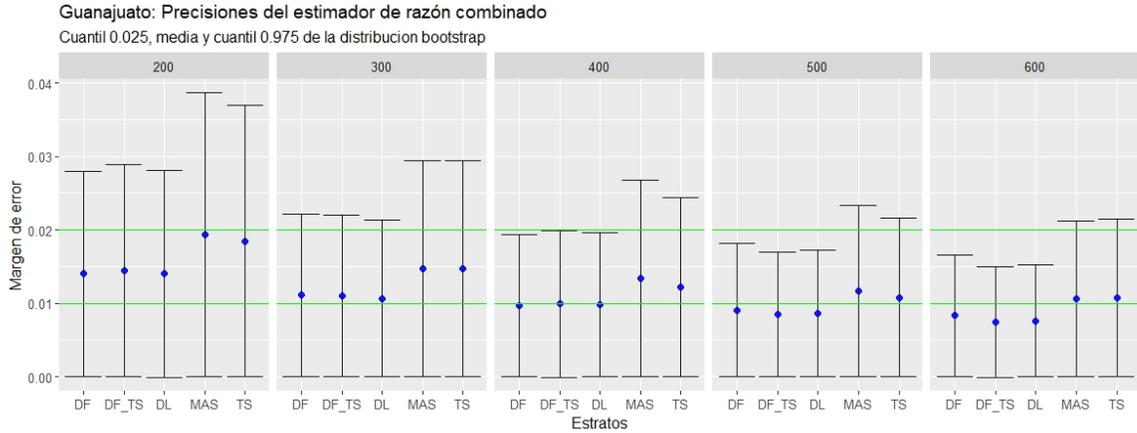
	ESTRATO	NI	200	300	400	500	600
21	8_R	135	4	6	8	9	11
22	8_U	316	9	13	18	22	26
23	9_R	123	3	5	7	9	10
24	9_U	256	7	11	14	18	21
25	Combinacion_1	449	13	19	25	31	38
26	Combinacion_2	406	11	17	23	28	34

Comparación entre diferentes tipos de estratificación respecto a sus correspondientes errores de estimación

Antes de realizar la comparación entre los distintos tipos de estratificación empleados, es conveniente mencionar que al elaborar la combinación de estratificación DF con TS ($DF \cap T$), se obtuvieron 57 combinaciones de estratificación, con tamaño un tamaño mínimo de estrato (N_h) de 5 casillas. Con objeto de minimizar la falta de representación, se colapsaron estratos, esto dio origen a 26 estratos (Cuadro 10).

Con el propósito de comparar la precisión del estimador de razón combinado de las diferentes estratificaciones evaluadas, se procedió a graficar los cuantiles 0.025, 0.0975 y media de 1,000 simulaciones considerando los tamaños de muestra 200, 300, 400, 500 y 600 (Gráfica 6). Como podemos observar, sólo las estratificaciones: DF, DF_TS y DL, presenta un promedio de error de estimación por debajo del 1% a partir de $n=500$. Al considerar el cuantil $q_{0.975}$, podemos observar que dicho límite, se encuentra por debajo del 2%, esto implica que un tamaño de muestra de $n = 500$, puede ser considerado como conveniente. Programas: GS1_LecturaDatos_con_estratos_DF_TS_26_Estratos.R y GS5_grafica_precisiones_cuantiles.R.

Gráfica 6 Cuantiles 0.025 y 0.975 de los errores de estimación asociados a las estratificaciones: DF, DF_TS, DL, MAS y TS.

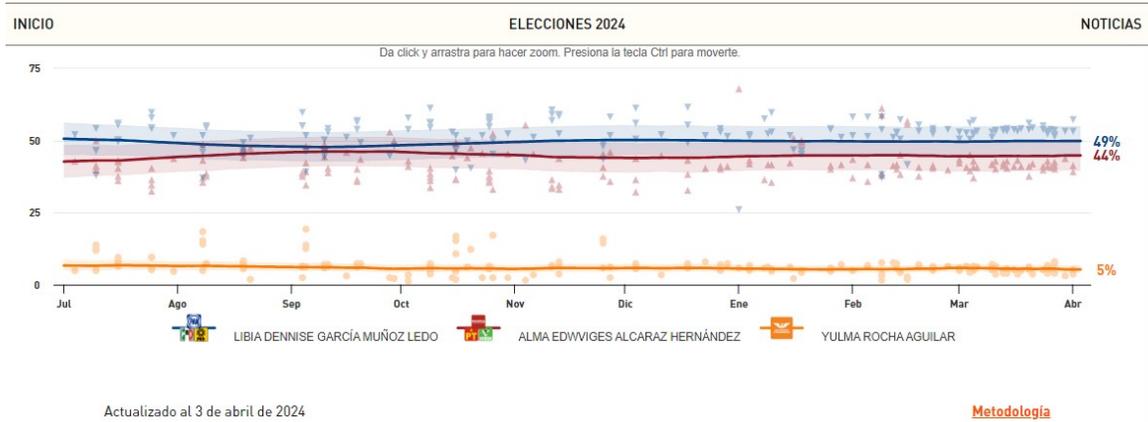


Información de Polls.mx

El Modelo Polls.mx es una oficina de agregación de encuestas, que estima la votación para presidente, gubernaturas y alcaldías de CDMX para este año, mediante un modelo bayesiano dinámico el cual promedia las estimaciones de varias empresas encuestadoras (<https://polls.mx/presidencia/>). Para el caso de Guanajuato presenta los porcentajes de votación potencial para las coaliciones: 1) Fuerza y Corazón por México, la cual integra a los partidos: PAN, PRI y PRD, 2) Seguimos Haciendo Historia y para el partido Movimiento Ciudadano, la cual integra a los partidos: MORENA-PT-PVEM-PANAL-FXP y 3) Partido MC.

La Gráfica 7 muestra que existe una diferencia de 5 puntos porcentuales entre Fuerza y Corazón por México y Seguimos Haciendo Historia, de tal suerte que si consideramos que el error de estimación promedio es del 1%, en teoría se podría identificar un candidato a la gubernatura, de seguir la misma tendencia hasta el 2 de junio del año en curso.

Gráfica 7 Tendencias de las alianzas: Fuerza y Corazón por México (línea azul), Seguimos Haciendo Historia (línea roja) y partido Movimiento Ciudadano (línea anaranjada).
Fecha 3 de abril de 2024.



Porcentaje de CAE con una casilla por tamaño de muestra y tipo de estratificación

Los Cuadro 11 a 16 8 a 12 muestran el porcentaje de CAE con una casilla de acuerdo al tamaño de muestra por cada tipo de estratificación. La simulación se realizó con B=1000 simulaciones bootstrap. Como podemos observar, sólo los estratos DF y DL cumplen exitosamente con la restricción de que a un CAE le corresponda una sola casilla en al menos 80%. En tanto que la estratificación por TS sólo cumple para los estratos mixto, para los tamaños de muestra de n=200 a 600, en tanto que en el estrato Rural sólo se cumpliría para el tamaño de muestra n=200 (Cuadro 15). Programa: G1_LecturaDatos_con_estratos_CAES_MERG.R

Cuadro 11 Porcentaje de CAE con una casilla en función del tamaño de muestra considerando estratificación por DF (15 estratos).

ESTRATO	200	300	400	500	600
1	95.98	93.80	91.36	89.61	87.12
2	95.30	93.01	90.66	88.42	86.38
3	95.65	93.38	91.32	88.85	87.09
4	96.25	93.76	91.58	89.68	87.92
5	96.64	94.84	92.85	90.91	89.46
6	96.28	94.74	92.41	91.13	89.18
7	96.00	93.99	92.54	90.54	88.25
8	95.54	93.12	90.04	88.25	86.00
9	95.76	93.96	91.31	89.46	87.42
10	94.43	91.42	88.40	86.18	83.25

ESTRATO	200	300	400	500	600
11	95.94	93.21	91.40	89.52	87.83
12	95.38	92.61	90.49	87.93	85.98
13	94.07	91.78	89.07	86.31	84.04
14	95.05	92.59	89.55	87.62	84.98
15	96.13	94.59	92.40	90.26	88.20

Cuadro 12 Porcentaje de CAE con una casilla en función del tamaño de muestra considerando estratificación por DF_TS (26 estratos). (Estrato 25: con menos de 50 casillas y Estrato 26: con casillas entre 51 y 99).

	ESTRATO	200	300	400	500	600
1	1_R	34.26	24.16	19.46	15.91	13.78
2	1_U	51.3	38.53	31.25	26.03	22.38
3	10_R	44.08	34.77	26.15	22.43	19.70
4	10_U	23.93	16.69	12.67	10.57	8.79
5	11_U	24.07	17.38	13.16	10.87	9.03
6	12_U	22.52	15.77	11.74	9.56	8.08
7	13_R	28.27	20.79	16.61	13.12	11.31
8	13_U	38.45	28.6	22.33	18.51	15.81
9	14_R	34.89	24.02	19.64	15.79	13.75
10	14_U	38.42	28.54	22.51	17.41	15.13
11	15_U	28.42	19.69	15.13	12.69	10.52
12	2_R	34.53	25.99	20.93	16.52	14.48
13	2_U	38.98	31.00	23.97	19.88	16.62
14	3_U	22.58	15.88	12.19	9.84	8.33
15	4_R	38.52	31.27	24.07	19.56	16.78
16	4_U	44.38	31.10	24.09	19.81	17.38
17	5_U	26.21	17.44	13.68	11.27	9.59
18	6_U	28.11	19.44	15.76	12.64	10.85
19	7_R	43.16	31.10	26.01	20.90	17.46
20	7_U	43.92	31.08	26.06	20.75	17.49
21	8_R	51.17	38.97	31.48	28.21	23.8
22	8_U	28.12	20.96	15.88	13.15	11.33
23	9_R	61.67	43.74	34.69	28.27	25.87
24	9_U	34.57	24.03	19.65	15.81	13.84

	ESTRATO	200	300	400	500	600
25	Combinacion_1	20.95	15.15	11.74	9.65	7.91
26	Combinacion_2	24.05	16.56	12.63	10.52	8.80

Cuadro 13 Porcentaje de CAE con una casilla en función del tamaño de muestra considerando estratificación DL (22 estratos).

ESTRATO	200	300	400	500	600
1	96.55	94.95	92.83	91.14	89.54
2	97.59	95.93	93.77	92.71	90.8
3	98.23	96.79	95.72	94.62	93.44
4	97.53	96.98	96.26	95.4	94.30
5	98.64	97.33	96.44	95.71	94.58
6	96.74	94.74	93.59	91.74	90.24
7	95.65	93.72	91.59	89.16	87.45
8	97.23	95.59	94.27	92.55	91.19
9	96.98	95.79	94.37	92.70	91.53
10	97.34	95.33	93.57	92.19	90.85
11	96.90	95.46	93.81	92.52	90.64
12	97.17	95.54	94.25	92.45	91.43
13	96.67	95.01	93.25	91.38	89.62
14	96.50	95.51	93.70	92.12	90.62
15	97.02	95.20	93.47	91.75	90.13
16	97.81	96.84	95.29	94.11	93.33
17	96.97	96.05	94.28	92.50	91.34
18	96.76	94.65	93.03	91.33	89.39
19	96.84	95.26	93.73	91.88	90.39
20	96.28	94.64	92.61	91.11	89.25
21	97.84	96.25	95.28	93.93	92.82
22	95.91	93.51	91.46	89.64	87.90

Cuadro 14 Porcentaje de CAE con una casilla en función del tamaño de muestra considerando estratificación MAS (1 estrato).

ESTRATO	200	300	400	500	600
11	52.59	40.9	32.84	27.25	23.16

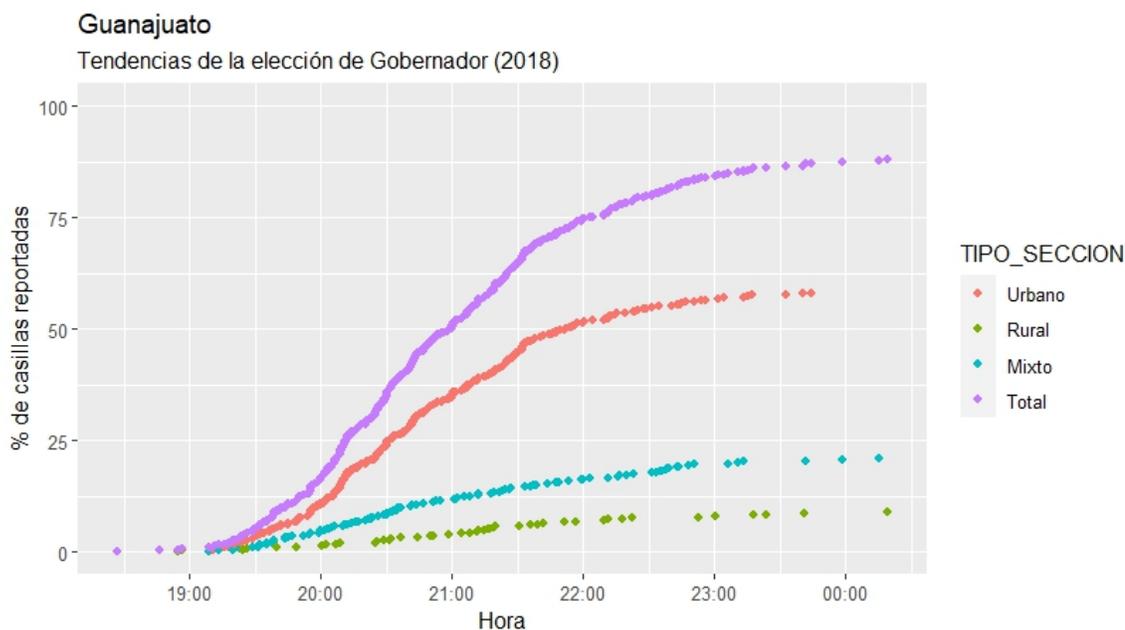
Cuadro 15 Porcentaje de CAE con una casilla en función del tamaño de muestra considerando estratificación TS (3 estratos).

ESTRATO	200	300	400	500	600
M	93.52	90.33	87.58	84.33	81.63
R	81.87	74.67	68.05	62.43	57.40
U	66.00	54.78	46.52	39.95	34.79

Tendencias de la elección a través del tiempo en 2018

Las tendencias del Conteo Rápido de la elección para gobernador para Guanajuato en 2018, empezaron a registrarse a las 18:26 para terminar a las 23:58 horas. De las n=500 casillas solicitadas para el Conteo Rápido, sólo el 89.4% (447 casillas) llegaron, de las cuales el 66.22, 10.07 y 23.71%, correspondieron a Urbano, Rural y Mixto, respectivamente (Gráfica 8). Podemos observar que las tendencias de la votación del medio rural son las más tempranas en estabilizarse, seguido de la mixta y la más tardía es la tendencia que proviene de la zona urbana (G_TendenciasDeEleccionFinal_4.R).

Gráfica 8 Tendencias del porcentaje de casillas a través del tiempo, que fueron enviadas a la cede del INE en 2018.



Conclusión

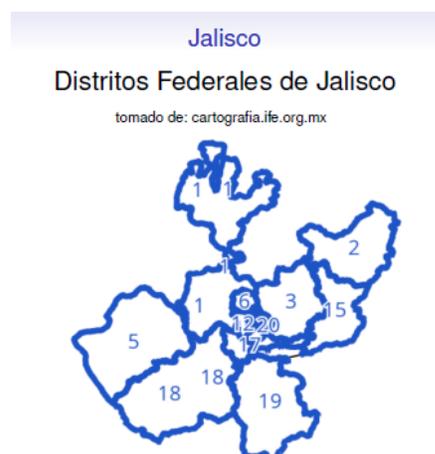
Se concluye que el diseño muestral con un tamaño de muestra de $n=500$, en donde los Distritos Federales (DF) son considerados como estratos, es el tamaño de muestra indicado, debido a que esta estratificación cumple con las restricciones de que el error promedio de estimación se encuentre por debajo del 1%, con una confianza del 95% de probabilidad y además un CAE le corresponda una casilla con una probabilidad de al menos el 80%.

7.4.1.4. Jalisco

Mtra. Patricia Isabel Romero Mares y Lic. Luis Enrique Reyes Romero.

Con la información recibida en el archivo Jalisco_2028_conARE_conGeoref2023.txt

- Lista nominal 5,903,281
- Número de casillas 9,794
- Número de distritos federales (DF) 20
- Número de distritos locales (DL) 20
- Hubo redistribución en el periodo de 2018 a 2023
- Los DF2018 son iguales a los DL2018
- Los DF2023 son prácticamente iguales a los DL2023 excepto en 61 casillas
- Número de ARE 2,363 a los que les corresponden entre 1 y 7 casillas



Elección de gobernador Jalisco 2018

Siete candidatos:

- C1. Partido Acción Nacional (PAN)
- C2. Partido Revolucionario Institucional (PRI)
- C3. Partido de la Revolución Democrática (PRD)
- C4. Partido Verde Ecologista de México (PVEM)
- C5. Movimiento Ciudadano (MC)
- C6. Partido Nueva Alianza (PNAL)
- C7. **Coalición**
 - Movimiento Regeneración Nacional (MORENA)
 - Partido del Trabajo (PT)
 - Partido Encuentro Social (PES)

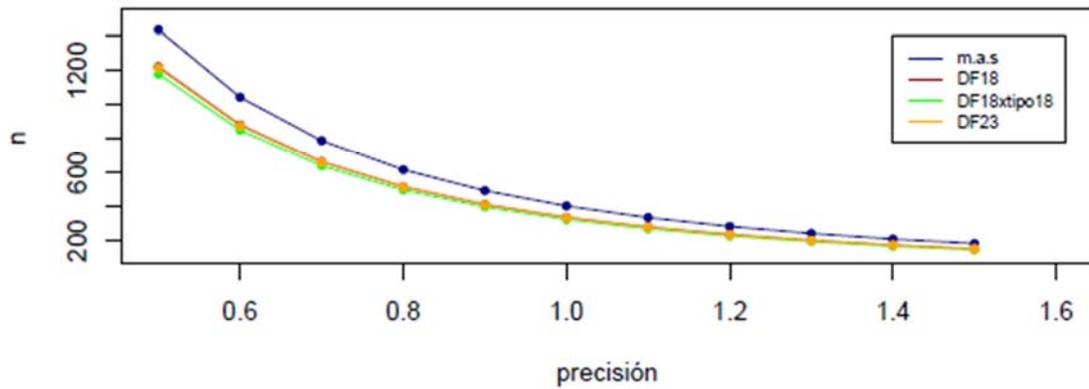
Diseño de muestra

Con la información de la elección de gobernador de Jalisco en el año 2018, se estudiaron las siguientes estratificaciones (en todos los casos se selecciona la muestra con m.a.s.):

- sin estratificar (solo como referencia)
- DF 2018 (20 estratos)
- DF 2023 (20 estratos)
- cruce de DF2018 por tipo de sección 2018 (urbana, no urbana) (35 estratos)

Se calculó el tamaño de muestra necesario con diferentes precisiones para cada una de las estratificaciones descritas y un nivel de **confianza del 95%**.

Jalisco, tamaño de muestra



Con una **precisión** aproximada de **0.81%**, y considerando solo los candidatos con mayor varianza, el tamaño de muestra para cada estratificación es:

Estratificación	C5	C7
DF18	505	487
DF18xTS18	490	420
DF23	500	482

Muestreo estratificado aleatorio, donde los estratos son los **distritos federales**, la selección de la muestra en cada estrato es con muestreo aleatorio simple (sin reemplazo) y se distribuirá el tamaño de muestra estatal ($n=500$) a los estratos en forma proporcional al número de casillas que tiene cada distrito federal, $n_h \in [19,31]$.

**Tomado de polls.mx, actualizado al 29 de febrero 2024



Análisis para la estratificación por Distrito Federal

- Información de N = 9807 casillas electorales de la elección para gobernador 2018.
- Se seleccionaron 1,000 muestras con el diseño especificado.
 - 20 Distritos Federales, n = 500
- Se calculó el promedio de los estimadores de razón combinado y de la precisión.
- Se calculó la cobertura de los intervalos del 95% de confianza.
- Se calculó el porcentaje de ARE que les corresponde entre 1 y 5 casillas en muestra.

Análisis para la estratificación por Distrito Federal

	R%	Precisión%	Cobertura
C1	10.64	0.58	94.2
C2	16.63	0.55	95.4
C3	1.02	0.14	93.0
C4	2.79	0.29	93.8
C5	39.06	0.81	94.4
C6	1.98	0.20	93.3
C7	24.64	0.76	95.4

Análisis para la estratificación por Distrito Federal n=500

	Casillas				
	1	2	3	4	5
ARES	91.79%	7.88%	0.33%	0.0%	0.0%

7.4.1.5. Morelos

Dr. Manuel Mendoza Ramírez y Dr. Felipe Javier Medina Aguayo

Objetivos

El objetivo del Conteo Rápido Morelos 2024, para la elección de Gobernador, es obtener estimaciones del porcentaje de votos emitidos a favor de los candidatos a Gobernador del estado, así como del porcentaje de ciudadanos que acudan

a votar. El Conteo Rápido es un ejercicio estadístico y para llevar a cabo se desarrollarán diversos procedimientos técnicos que tienen sustento en la Teoría de la Inferencia Estadística.

Antecedentes.

La Inferencia Estadística se ha desarrollado progresivamente durante más de 100 años y a partir de su estructura conceptual, el rigor de sus resultados teóricos y su desempeño en la práctica, se ha establecido como una disciplina científica muy reconocida y como un estándar aceptado internacionalmente para el estudio de los fenómenos que presentan variabilidad. En el caso específico de este conteo rápido se recurrirá, por una parte, a la Teoría de Muestreo para Poblaciones Finitas, con el propósito de establecer, en primer lugar, un diseño eficiente de la muestra que proveerá la información para la estimación. Por otra parte, de entre los Métodos Bayesianos de Estimación Paramétrica, se seleccionará el mecanismo que habrá de utilizar con los datos que ofrezca la muestra para producir las estimaciones.

Diseño Muestral.

Consideraciones generales.

Para efectos de este conteo rápido se considera como población bajo estudio a la totalidad de las casillas que se instalarán en la entidad el día de la jornada electoral. En esa población es de interés estimar, como objetivo central, el porcentaje de votos emitidos a favor de cada candidato a Gobernador. Además, interesa también el número total de votos que se recibirán para la elección de Gobernador y consecuentemente, el porcentaje de ciudadanos, en la lista nominal, que acudirán a ejercer ese derecho.

Estas estimaciones se producirán a partir de la información recolectada, al cierre de la votación, en una muestra de casillas que se habrá de definir mediante un mecanismo de selección probabilística la cual, con la finalidad de atender a lo dispuesto en el Artículo 373 del Reglamento de Elecciones del Instituto Nacional Electoral, debe abarcar la mayor dispersión geográfica posible. Para cumplir con ese propósito, se utilizará un diseño muestral estratificado geográficamente. De esta forma, la entidad se dividirá en regiones que servirán como estratos y de cada uno de los estratos se seleccionará una muestra de casillas completamente al azar, sin reemplazo. El tamaño de la muestra en cada estrato será proporcional al tamaño del propio estrato.

Este diseño, además de responder al requerimiento del reglamento, es muy común en la práctica del muestreo. Produce estimaciones más precisas que las

correspondientes a una muestra que se selecciona con un muestreo aleatorio simple, sin restricción geográfica o de otro tipo. Para definir un diseño estratificado es necesario establecer, en primer lugar, los criterios con que se dividirá la entidad en estratos y, una vez con tales criterios, el siguiente paso es seleccionar el tamaño de total de la muestra.

La determinación de los criterios de estratificación se lleva a efecto tomando en cuenta que el diseño produce mejores resultados en la medida en que las unidades de muestreo dentro de cada estrato son más homogéneas entre sí y, al mismo tiempo, diferentes de las que pertenecen a otros estratos. La homogeneidad en esta conceptualización, se refiere al comportamiento de la variable, objeto de estudio; en este caso, la votación observada.

Por otra parte, el tamaño de la muestra se fija de manera que la estimación alcance niveles aceptables de confiabilidad y precisión. En el caso del Conteo Rápido para la elección de Gobernador en Morelos en 2024, el nivel de confiabilidad, como también lo establece el Reglamento de Elecciones, se ha fijado en 0.95. De esta forma, para cada propuesta de estratificación, es necesario examinar la precisión que producen los distintos tamaños de muestra, considerando este nivel de confiabilidad.

Criterios de estratificación.

Para un conteo rápido estatal una posibilidad, relativamente natural, es considerar como posible estratificación la división de la entidad en sus distritos electorales. Ahora bien, en México, existen dos tipos de distritos electorales: los distritos locales y los distritos federales. Los primeros son la base para los procesos electorales que sólo incumben a la entidad como las elecciones de Gobernador y del Congreso Estatal. Los segundos, juegan el mismo papel, pero para los procesos federales, en particular para la elección de Presidente de la República y de la Cámara de Diputados Federal.

Tanto la división de la entidad en distritos locales como la correspondiente a los federales son objeto de actualizaciones periódicas por parte de la autoridad electoral y puede ocurrir que en alguna ocasión coincidan, aunque esta es una situación poco frecuente. También puede suceder que los distritos locales resulten en una subdivisión de los distritos federales. Al iniciar el proceso electoral federal 2023-2024 el estado de Morelos está dividido, como resultado de la más reciente redistribución, en 5 distritos electorales federales y 12 distritos electorales locales y estos últimos no se obtienen como una subdivisión de los distritos federales.

Figura 19 Los 12 Distritos locales del Estado de Morelos en 2024.



Figura 20 Los 5 Distritos federales del Estado de Morelos en 2024.



Para los fines de este estudio, pueden considerarse, como punto de partida, dos estratificaciones alternativas: i) la que divide a la entidad en sus 12 distritos electorales locales y ii) la que la divide en sus 5 distritos electorales federales. De hecho, también se consideró una estratificación intermedia con 7 estratos, resultado del cruce de las dos estratificaciones iniciales a la que fue necesario imponer restricciones para evitar estratos excesivamente pequeños.

Cuando el número de casillas por estrato disminuye, esta condición aumenta la probabilidad de que el día de la jornada electoral, si la muestra llega incompleta, no se cuente con información de todo un estrato, como ya ha ocurrido en procesos electorales previos. Por esta razón es conveniente establecer una restricción sobre el tamaño mínimo de los estratos resultantes en el diseño.

En otros ejercicios similares, se han incorporado otros factores para refinar aún más la eventual estratificación. Un ejemplo es la división de los distritos en sus componentes urbana y no urbana. En este caso, adicionalmente a la idea de

evitar estratos pequeños y dado que el conteo rápido para el estado de Morelos se llevará a cabo como parte de un sistema de conteos rápidos que incluirán las elecciones de gobernador o jefe de gobierno en otras 8 entidades, la elección presidencial y las elecciones de las cámaras de diputados y de senadores, el Comité ha decidido privilegiar los diseños más simples posibles, en tanto provean los niveles de precisión apropiados, con el propósito de facilitar el trabajo de campo en la recolección de todas las muestras necesarias. De esta manera, en el caso de Morelos, se consideraron tres propuestas de estratificación a partir de estos criterios: Distrito Electoral Federal (DF), Distrito Electoral Local (DL) y Cruce de Distrito Electoral Federal con Distrito Electoral Local (DFxDL).

Evaluación de las alternativas de estratificación.

Una vez establecidas las tres estratificaciones posibles, el procedimiento para seleccionar la alternativa óptima hizo uso de los resultados de la elección local para Gobernador de 2018. A partir de la base de datos que contiene los resultados de la votación en cada una de las casillas instaladas en esa jornada electoral, y con el conocimiento de los resultados finales, se procedió, para cada estratificación propuesta y considerando distintos tamaños de muestra, a evaluar la precisión de las estimaciones sujeta a un nivel de confiabilidad de 0.95.

Vale la pena señalar que los distritos vigentes en 2018 fueron objeto de un proceso de redistribución con posteridad a esa elección y aún cuando el número de distritos en 2024 (tanto locales como federales) se mantiene inalterado, la partición geográfica que representan es sustancialmente distinta. Por esta razón, y gracias al apoyo del personal del INE que estableció, para cada casilla instalada en 2018, el estrato de 2024 al que pertenecería, fue posible obtener resultados en los dos escenarios: con los distritos de 2018 y con los de 2024.

En términos generales, el mecanismo de evaluación de las propuestas de diseño se desarrolla como sigue: Para una estratificación fija y un tamaño total de muestra n , que se distribuye entre los estratos en forma proporcional al tamaño de cada estrato, se procede de la siguiente manera.

1. Se selecciona una muestra con ese tamaño, de acuerdo al esquema de estratificación.
2. Con la información de esa muestra, se utiliza el procedimiento descrito en Mendoza y Nieto (2016) para calcular la distribución final de la proporción de votos a favor de cada uno de los candidatos a gobernador.
3. Para cada candidato, por ejemplo el i -ésimo candidato, se calcula la estimación puntual de la proporción de votos a su favor \hat{p}_i y, con el valor que finalmente obtuvo en esa elección, p_i , se determina el correspondiente error de estimación $|p_i - \hat{p}_i|$.

4. También para el i -ésimo candidato, se calcula el intervalo de estimación de p_i con un nivel de confiabilidad del 95% y se registra la longitud de ese intervalo.
5. Considerando los resultados de todos los candidatos, con esa muestra se reporta M , el máximo de los errores de estimación y L , el valor máximo de las longitudes de los intervalos de estimación.

Los pasos 1 a 5 se repiten un número K de veces (en este caso, $K = 1000$) a fin de estudiar el comportamiento de los indicadores M y L , además de monitorear la cobertura de los intervalos de estimación.

Con los resultados de este estudio es posible comparar, considerando un tamaño de muestra fijo, el desempeño de las distintas estratificaciones para elegir la alternativa óptima. Una vez con esta decisión, se puede estudiar el efecto de distintos tamaños de muestra para seleccionar uno que conduzca a un nivel de precisión que genere la certidumbre requerida.

Otro elemento que se considera de la mayor importancia para decidir el tamaño de la muestra, en un ejercicio de conteo rápido, es la carga de trabajo que representa la recolección de los datos en campo, el día de la jornada electoral. Quienes se ocupan de esta tarea son los Capacitadores Asistentes Electorales (CAE). Estos Asistentes tienen asignada una variedad de responsabilidades a lo largo del proceso electoral y, específicamente, el día de la jornada electoral, entre otras tareas, tienen a su cargo un grupo de casillas que han de monitorear para comprobar que han sido instaladas, que operan normalmente y concluyen sus funciones de acuerdo a lo previsto.

El trabajo de presentarse, hacia el final de la jornada, en una casilla seleccionada en la muestra, para recabar la información que resulte del proceso de escrutinio a fin de reportarla al Comité, les demanda un tiempo que puede interferir momentáneamente con otras actividades que también son relevantes. Para contribuir a que ese tiempo no sea excesivo, el tamaño de la muestra se elige de manera que la proporción de los CAE que deban reportar la información de más de una casilla sea la menor posible.

De esta forma, en este estudio se analiza, primero, el efecto que distintos tamaños de muestra tienen en la carga de trabajo de los CAE para, después, con un rango de tamaños de muestra de referencia, explorar el desempeño de las distintas estratificaciones.

Resultados de la elección para Gobernador en 2018

En la elección de 2018 contendieron ocho candidatos por el cargo de Gobernador del estado de Morelos. (CB) Cuauhtémoc Blanco, por la coalición

Juntos Haremos Historia, (VC) Víctor Caballero por la coalición Por Morelos al Frente, (RG) Rodrigo Gayosso por la coalición Juntos por Morelos, (JM) Jorge Meade por el Partido Revolucionario Institucional, (NL) Nadia Lara por el Partido Verde Ecologista, (AV) Alejandro Vera por el Partido Nueva Alianza, (MR) Mario Rojas por el Partido Humanista de Morelos y finalmente, (FD) Fidel Demédecis, Candidato Independiente.

Estructura del estudio de estratificación

Como ya se señaló, se consideraron los siguientes criterios para la establecer la estratificación.

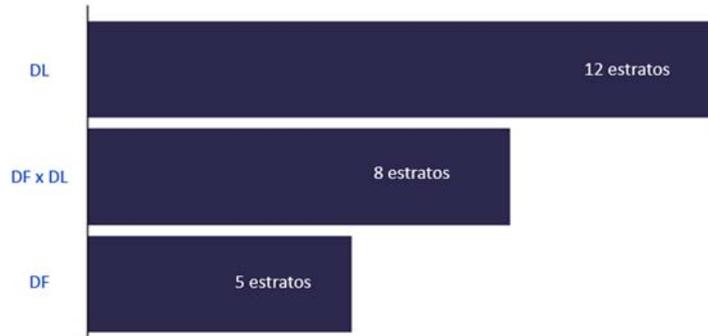
- Distrito Electoral Federal (DF).
- Distrito Electoral Local (DL).
- Cruce Distrito Federal con Local (DFxDL)

En 2018 se instalaron 2,421 casillas en la entidad. En promedio, 484 casillas por Distrito Electoral Federal. En este estudio se consideraron fracciones de muestreo entre 4% y 9%, que representan entre 97 y 218 casillas en total. Estas fracciones equivalen a entre 20 y 44 casillas en muestra, en promedio, por Distrito Electoral Federal.

Figura 21 Número de estratos en las tres estratificaciones (2018).



Figura 22 Número de estratos en las tres estratificaciones (2024).



Análisis de la carga de trabajo para los CAE.

Como ya se indicó, en una primera fase, se analizó la carga de trabajo que diferentes tamaños de muestra implicarían para los CAE. Para llevar a efecto este análisis, se consideraron las tres estratificaciones propuestas, en cada uno de los escenarios de distritación (2018 y 2024).

Se simularon 1000 muestras para cada fracción de muestreo (desde 4% hasta 9%, con intervalos de un punto porcentual). En cada muestra se calcula el porcentaje de CAE con sólo una casilla en muestra. Se reporta también el porcentaje de CAE con dos y con más de dos casillas en muestra.

Figura 23 Porcentaje de CAE con una casilla (2018).

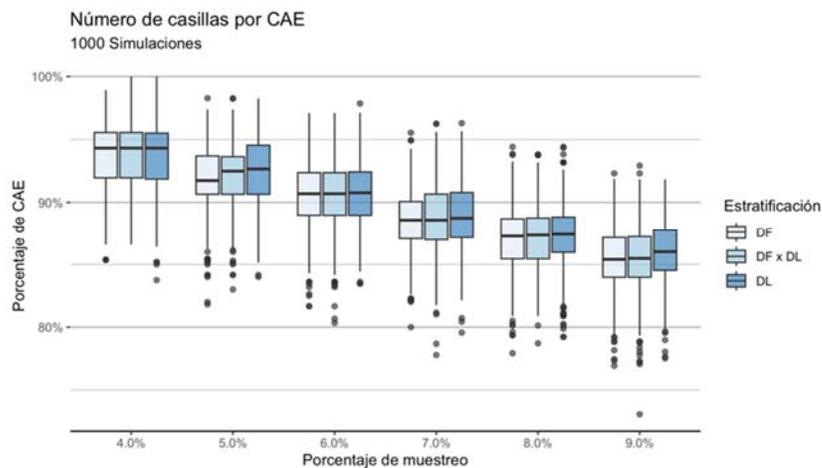
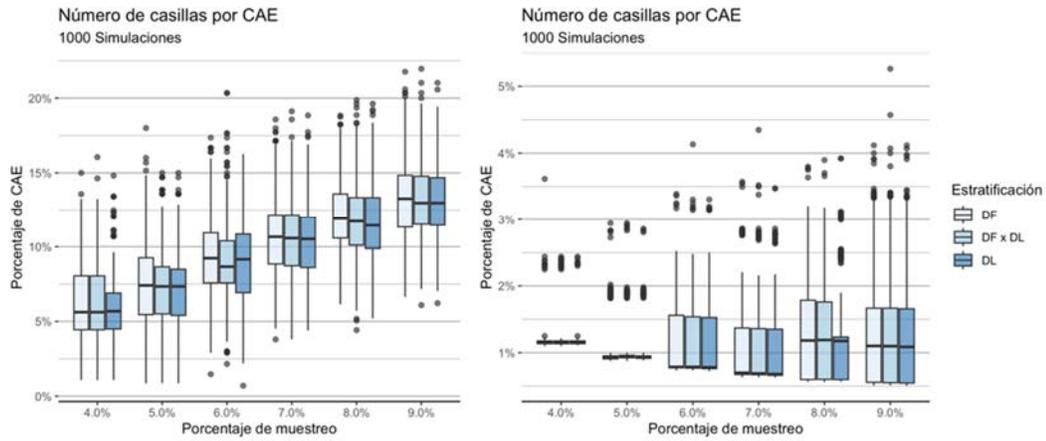


Figura 24 Porcentaje de CAE con dos casillas (izq.) y 3 o más casillas (der.) (2018).



En las Figuras 22 y 23 se observa que, en el tema de la carga de trabajo para los CAE, sin importar cuál sea la fracción de muestreo, no hay una diferencia sustantiva entre las tres estratificaciones cuando se consideran los distritos de 2018. En general, se puede asegurar que con una fracción de muestreo que no supere el 8%, la mayoría de las muestras reportan más de 80% de los CAE con sólo una casilla por reportar. De la misma forma, la mayoría de las muestras presentan menos del 13% de los CAE con dos casillas por reportar y menos del 3% de los CAE con tres o más casillas por reportar.

Figura 25 Porcentaje de CAE con una casilla (2024).

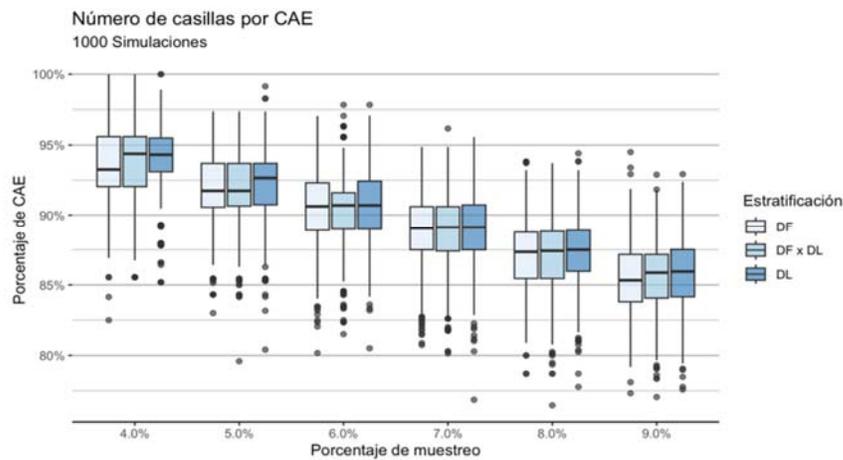
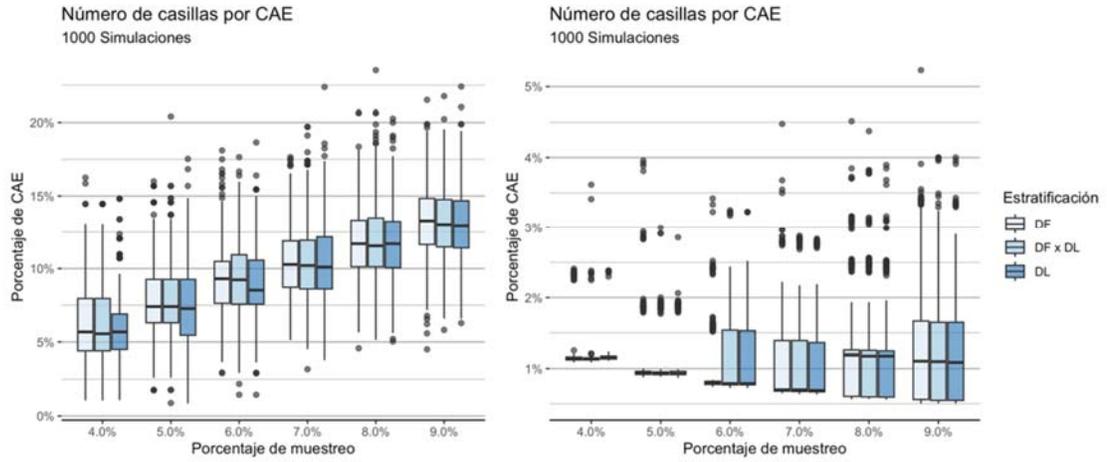


Figura 26 Porcentaje de CAE con dos casillas (izq.) y 3 o más casillas (der.) (2024).



Cuando se analiza la selección de las muestras, utilizando los distritos de 2024, los resultados son prácticamente los mismos. De esta forma, se puede concluir que, para efectos de la carga de trabajo en campo, una fracción de muestreo de hasta 8% es razonable para los CAE.

Análisis de las posibles estratificaciones.

Tomando como partida los resultados de la carga de trabajo en campo, las tres estratificaciones de interés se analizaron, tanto con los distritos de 2018 como con los distritos de 2024, considerando una fracción de muestreo de 7%, como referencia, y el procedimiento de evaluación es el descrito en la subsección 3.3 de este documento.

Figura 27 Error máximo de estimación. Distritos 2018.

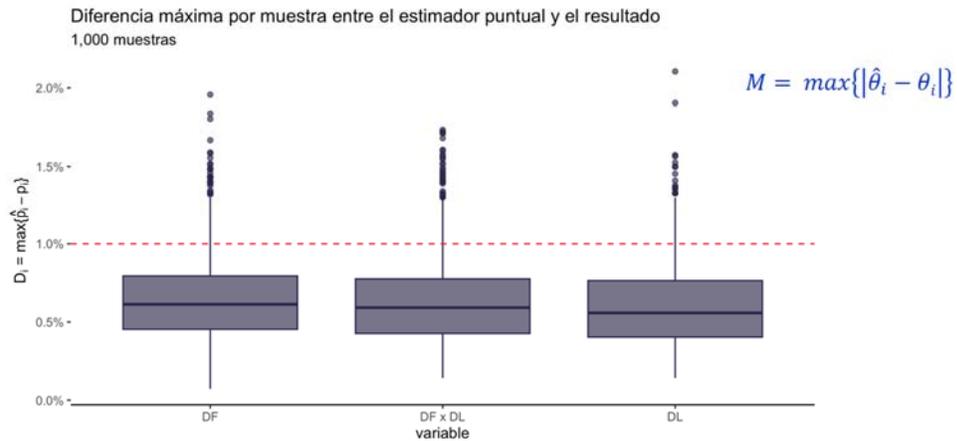
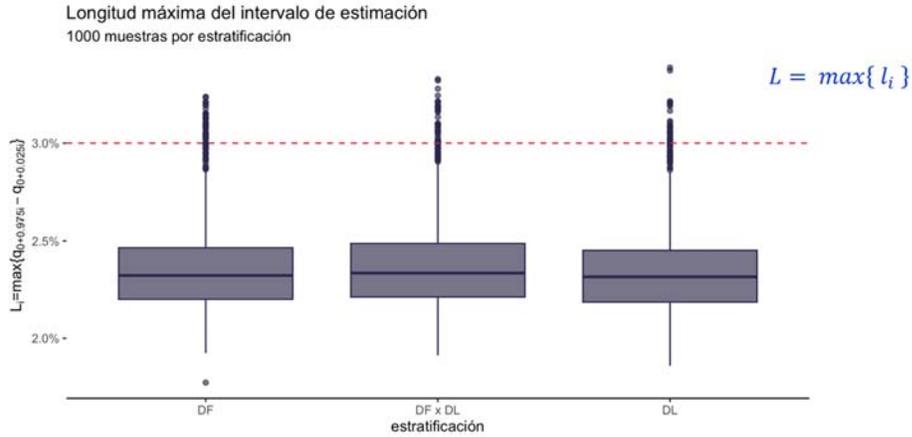


Figura 28 Longitud máxima del intervalo de estimación. Distritos 2018.



Como puede observarse en las Figuras 26 y 27, cuando se consideran los distritos de 2018, no se aprecian diferencias relevantes en los resultados que se obtienen con las tres estratificaciones. En todos los casos se tiene un error máximo de estimación que, en la mayoría de las muestras, es claramente inferior al 1% y una longitud máxima del intervalo de estimación inferior a 2.5%, también en la mayoría de las muestras.

Figura 29 Error máximo de estimación. Distritos 2024.

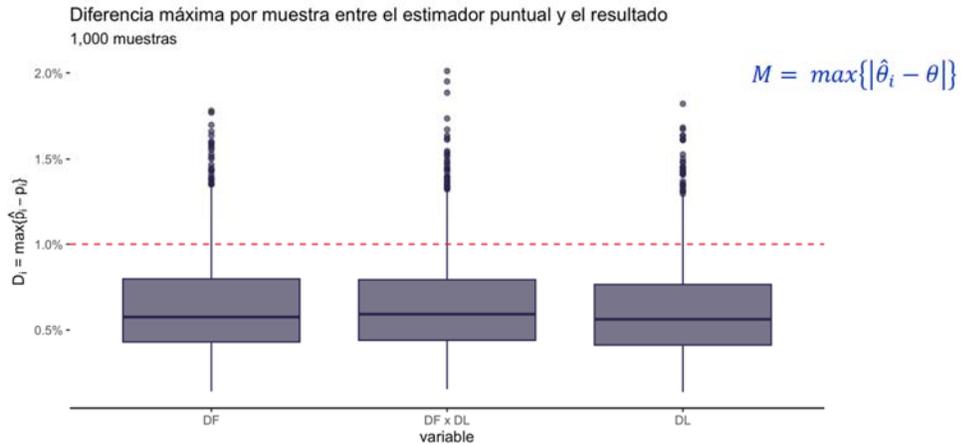
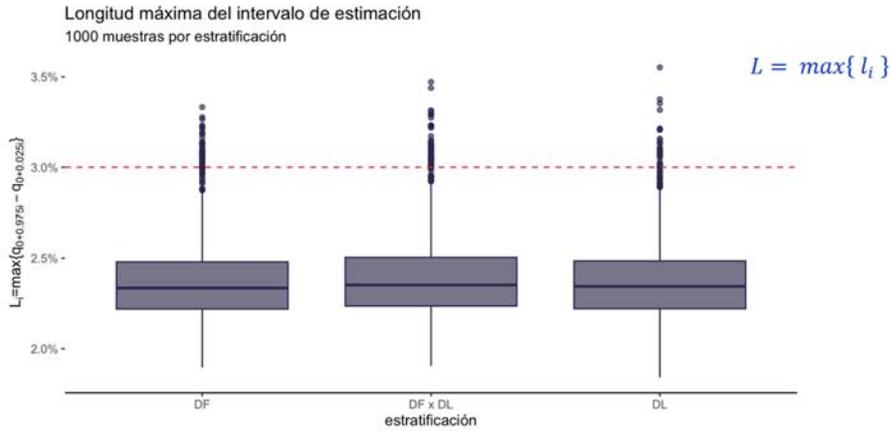


Figura 30 Longitud máxima del intervalo de estimación. Distritos 2024.

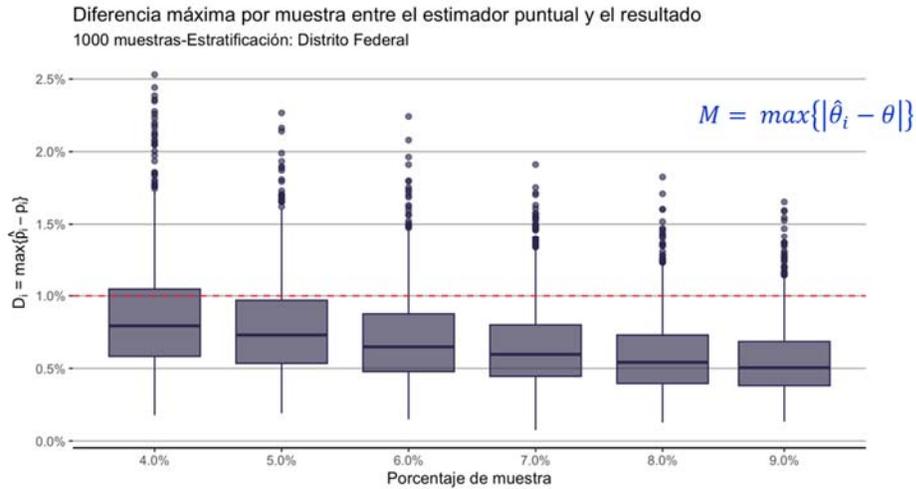


Con el apoyo de las Figuras 28 y 29, se concluye que este patrón se replica casi exactamente si ahora se utilizan los distritos de 2024. De esta forma, tomando en cuenta que el diseño se simplifica con la estructura más sencilla y en el afán de que este diseño sea compatible y susceptible de armonizarse con los demás que forman el sistema de conteos rápidos 2024, se concluye que para el conteo rápido de la elección de gobernador en el estado de Morelos es razonable utilizar la estratificación correspondiente a los cinco distritos electorales federales vigentes es 2024.

Determinación del tamaño de muestra.

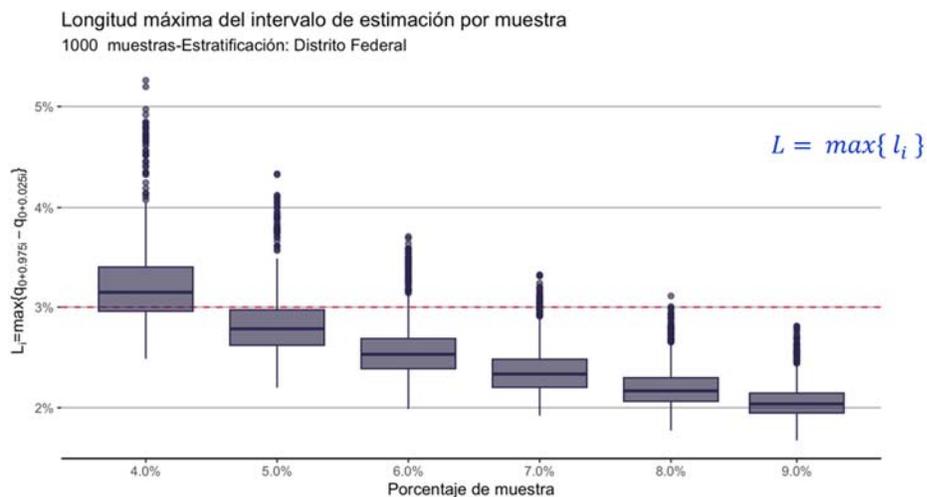
Para fijar en definitiva el tamaño de muestra apropiado y tomando en cuenta los resultados de la sección previa, se considera solamente la estratificación con los distritos electorales federales (DF). Si bien durante el proceso electoral en curso estarán vigentes los distritos federales 2024, con fines informativos también se analizaron los resultados que se obtendrían con los distritos federales 2018.

Figura 31 Error de estimación máximo como función del tamaño de la muestra. Distritos Electorales Federales 2018.



El análisis explora el desempeño de las inferencias con muestras completas que corresponden a distintas fracciones de muestreo. Con 1000 muestras de cada tamaño se evalúan los indicadores M , el máximo de los errores de estimación y L , el valor máximo de las longitudes de los intervalos de estimación. Los resultados se pueden observar en a Figuras 30 y 31.

Figura 32. Longitud máxima del intervalo de estimación como función del tamaño de la muestra. Distritos Electorales Federales 2018.



En estas gráficas se puede observar que el error de estimación máximo disminuye a medida que aumenta el tamaño de la muestra, y que con una fracción de muestreo de 7% o mayor, se tiene que, no solamente la mayoría de las muestras producen un error de estimación máximo inferior a 1% sino que todas las muestras

analizadas reportan un error máximo de estimación inferior a 2%. Por lo que se refiere a la longitud máxima de los intervalos de estimación, esta también disminuye a medida que el tamaño de la muestra aumenta y se puede observar que, si la fracción de muestreo es 8% o mayor, prácticamente se puede asegurar que la longitud máxima será menor a 3 puntos porcentuales.

Figura 33 Error de estimación máximo como función del tamaño de la muestra. Distritos Electorales Federales 2024.

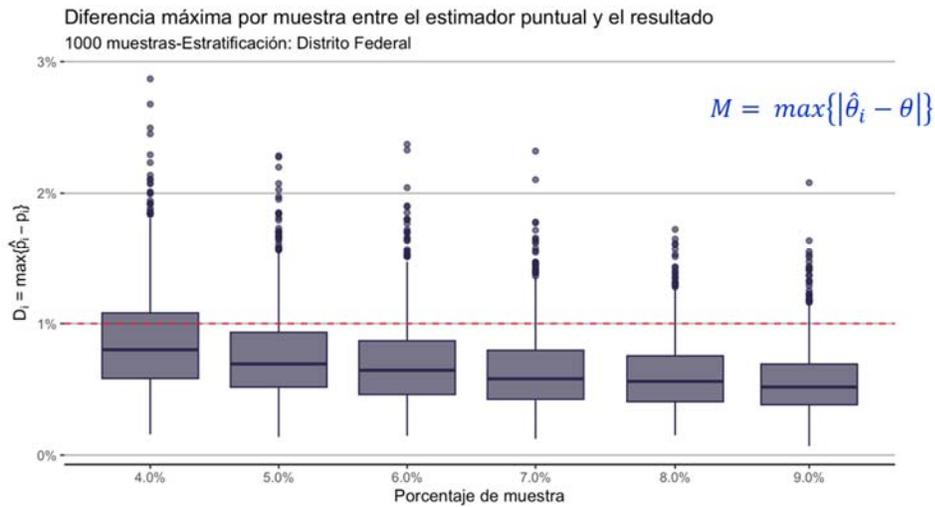
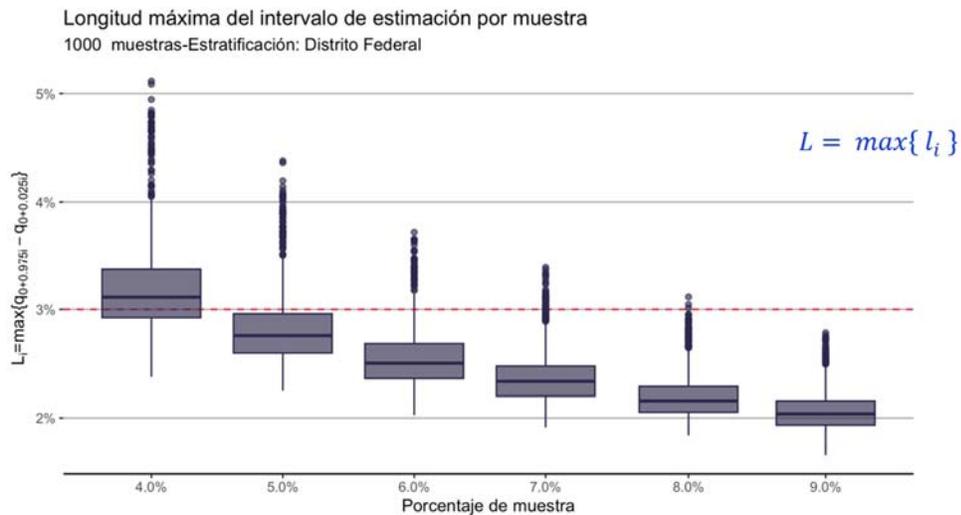


Figura 34 Longitud máxima del intervalo de estimación como función del tamaño de la muestra. Distritos Electorales Federales 2024.



Como puede comprobarse en las Figuras 32 y 33, el comportamiento del error máximo de estimación y la longitud máxima de los intervalos de estimación con

los distritos de 2024 es prácticamente el mismo que se observó con los distritos de 2018.

Como resumen, se puede concluir que, a partir de los resultados de este estudio, es razonable utilizar la estratificación que emplea como estratos a los cinco distritos electorales vigentes en 2024 con un tamaño de muestra equivalente a una fracción de muestreo entre 7% y 8%

Más específicamente, y tomando en cuenta que, de acuerdo a las previsiones disponibles para el número de casillas que se instalarán en el estado de Morelos para el proceso electoral 2024, el 7% de las casillas equivaldrá a un poco más de 180 casillas mientras que el 8% resultará en aproximadamente 210 casillas, se propone utilizar una muestra de 200 casillas para este ejercicio.

Consideraciones Complementarias.

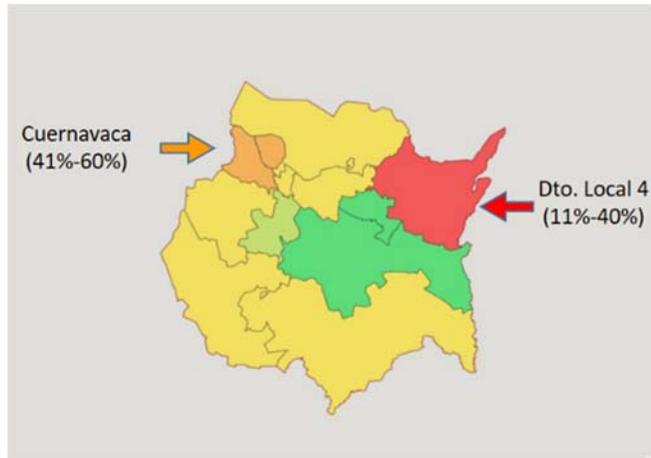
Como ya se ha indicado, a partir de los resultados obtenidos se propone utilizar la estratificación que utiliza como estratos a los 5 Distritos Electorales Federales vigentes en 2024, con una muestra de 200 casillas distribuidas proporcionalmente conforme al tamaño de los estratos.

Es importante destacar que esta propuesta supone que las estimaciones se llevan a cabo con la muestra completa. En otros términos, el análisis describe los niveles de calidad estadística que pueden alcanzarse el día de la jornada electoral si la muestra que se recibe para producir los resultados contiene el 100% de las casillas seleccionadas en la muestra o un porcentaje muy cercano a esta cifra.

Con el propósito de explorar cuan razonable es la expectativa de recibir la muestra completa y, en todo caso, sugerir alguna medida para evitar retrasos y faltantes en la muestra recolectada se analizó el arribo de la muestra que se empleó en 2108 para el mismo tipo de ejercicio, en la misma entidad, el estado de Morelos.

En 2018, el Comité solicitó una muestra también de 200 casillas y, a las 22:40, el informe del conteo se produjo con la información de 160 casillas lo que equivale al 80% de la muestra prevista. En esa ocasión se consideraron como estratos los 12 Distritos Electorales Locales. Entre los hallazgos más interesantes es que el patrón de arribo no fue homogéneo entre los distintos estratos.

Figura 35 Cobertura de la muestra recibida en la elección de 2018.



A partir del sistema de información geográfica que se operó en 2018, se puede verificar, en la Figura 35, que sólo de dos distritos se obtuvo la correspondiente muestra prácticamente completa (en verde). Sin embargo, es más importante observar que los estratos (distritos) con menor porcentaje de muestra recibida son los que en ese momento correspondían al distrito electoral local 4 (en rojo) y los distritos locales en el entorno de la ciudad de Cuernavaca (en marrón). El caso particularmente extremo es el del entonces distrito local 4 con un nivel de muestra recibida entre 11 y 40 por ciento.

Ese distrito local se ha transformado después de la nueva distritación y, además, es conveniente recordar que en 2024 los estratos corresponderán con los distritos electorales federales y no locales. En todo caso es conveniente señalar esta región del estado para investigar si es posible intervenir de alguna forma en este fenómeno de retraso tan específico en términos geográficos.

Figura 36 Distrito Electoral Federal con menor cobertura en 2018.



Como puede observarse en la Figura 36, la zona con menor cantidad de muestra recogida en 2018, corresponde en la actualidad a la parte más oriental del distrito

electoral federal 5. Esta información se ha compartido con la Dirección Ejecutiva de Organización Electoral a fin de que puedan explorar tanto las causas de este retraso como las medidas que puedan mitigar el problema.

Otro aspecto potencialmente relevante es la naturaleza de las secciones electorales que conduce a la clasificación de las casillas en urbanas, rurales o mixtas. En diversas ocasiones se ha manifestado preocupación por las casillas rurales que, dada su condición, podrían encontrar más dificultad en la comunicación y, por esa causa, presentar retrasos. Con el objetivo de explorar esa idea se consideraron las 2,421 casillas que se instalaron en total en el estado de Morelos en 2018 y que se distribuyeron como sigue:

Tabla 18 Distribución de las casillas por tipo en la entidad.

Tipo	Casillas	Porcentaje
Urbanas	1815	75%
Rurales	174	7%
Mixtas	432	18%

De esas 2, 421 casillas, como ya se señaló, se tomó una muestra de 200 casillas para la estimación del conteo rápido y el informe de resultados se produjo con las 160 casillas que arribaron hasta las 22:40. Estas 160 casillas se distribuyeron de acuerdo a la siguiente Tabla.

Tabla 19 Distribución de las casillas por tipo en la muestra.

Tipo	Casillas	Porcentaje
Urbanas	126	79%
Rurales	11	7%
Mixtas	23	14%

Es inmediato comprobar que la proporción de las casillas rurales en la muestra es la misma que en la población completa. Esto sugiere que la condición de rural no es factor para el no arribo de la muestra. Un análisis similar apunta en el sentido de que son las casillas de tipo mixto las que llegaron en una proporción inferior a la esperada.

Figura 37 Patrón de arribo de todas las casillas en la muestra 2018.



En la Figura 37 se exhibe el patrón de arribo de todas las casillas en la muestra. La gráfica no sugiere anomalía alguna. La tasa de llegada parece constante, sin interrupciones. Sin embargo, la situación es diferente si se analizan por separado las casillas urbanas y las mixtas.

Figura 38 Patrón de arribo de todas las casillas Urbanas y Mixtas en la muestra 2018.



En la Figura 38 se observa la diferencia entre los patrones de arribo de los dos tipos de casillas que refuerza la presunción de que el retraso en la llegada de las casillas en la muestra se manifiesta en particular en el caso de las casillas de tipo mixto.

En resumen, el análisis de esta sección identifica la zona oriental del actual distrito electoral federal número 5 como la región en donde se produjo el mayor retraso en el arribo de las casillas de la muestra de 2018. También exhibe evidencia en el sentido de que se retraso se manifiesta especialmente en el caso de las casillas de tipo mixto. Esta información ha sido transmitida al Instituto Nacional Electoral con el fin de que sea considerada para evaluar medidas que puedan contribuir a mitigar el retraso en la llegada de la muestra del conteo rápido de 2024.

Conclusiones.

A partir de los resultados en este estudio, se propone que el diseño la muestra para realizar el conteo rápido de la elección de Gobernador en el estado de Morelos, sea estratificado utilizando los 5 distritos electorales federales vigentes a la fecha como estratos. Además, se propone un tamaño de muestra igual a 200 casillas que se distribuirán entre los estratos en forma proporcional al tamaño del estrato.

Referencias.

- Mendoza, M. y Nieto-Barajas, L.E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. *Electoral Studies* 43, 124-132.

7.4.1.6.Puebla

Dr. Gabriel Núñez Antonio y Mtro. Emiliano Geneyro Squarzon.

El objetivo general de los ejercicios de Conteo Rápido (CR) es contribuir a dar certidumbre, confianza y transparencia en un proceso electoral. Lo anterior, ofreciendo una estimación de las tendencias de la votación el mismo día de la jornada electoral. Se debe enfatizar que el éxito de los Conteo Rápidos depende del trabajo conjunto de varios actores: Los miembros del Comité Técnico Asesor del CR (COTECORA), las diferentes áreas del INE involucradas, los miembros de las Juntas Locales, los miembros de los Organismos Públicos Locales Electorales (OPL), etc.

Desde el punto de vista estadístico un Conteo Rápido es un procedimiento que, con base en la información de una muestra aleatoria de casillas de escrutinio y cómputo, permite estimar con cierta precisión y confianza, las tendencias de los resultados la misma noche de la elección.

Uno de los aspectos técnicos responsabilidad del COTECORA es el denominado Diseño Muestral, el cual se puede definir como el conjunto de estrategias y procedimientos definidos para seleccionar una muestra de una población objetivo de estudio, la cual debe cumplir con ciertas características estadísticas adecuadas. El Diseño muestral también considera procedimientos que minimicen los errores de sesgo y cobertura, asociados a los procedimientos de estimación que se emplean. Así, desde un punto de vista moderno, se puede decir que un Diseño muestral queda definido por la determinación de un tamaño de muestra, una estrategia de selección de la muestra y un método de estimación. Lo anterior, con la finalidad de lograr niveles aceptables de precisión en las estimaciones ofrecidas, con un alto nivel de confianza estadística.

A continuación, se presentan algunos análisis y consideraciones técnicas que permitieron definir el Diseño muestral que se empleará para el Conteo Rápido en la elección del 2 de junio de 2024 a la Gubernatura del Estado de Puebla.

Estrategia de Selección

Considerando el Artículo 373 del Reglamento de Elecciones (donde se establece que la muestra debe abarcar la mayor dispersión geográfica posible) y tomando en cuenta la experiencia técnica adquirida en anteriores ejercicios de Conteo Rápido, se concluye que la estrategia de selección de las casillas de escrutinio y cómputo sea un muestreo aleatorio estratificado. Un aspecto importante en un muestreo estratificado es la definición o construcción de los estratos, los cuales son sub-grupos de la población que cumplen con dos características: no se traslapan y la unión de todos los estratos conforma a toda la población objetivo.

Para el caso del Conteo Rápido del Estado de Puebla, la población está conformada por el total de casillas aprobadas para el día de la Jornada Electoral. Para definir los estratos se consideraron diferentes estratificaciones (forma de definir estratos) entre las cuales se encuentran:

- Estratificación por Distritos Federales.
- Estratificación por Distritos Locales.
- Estratificación considerando las intersecciones de Distritos Federales y Distritos Locales.
- Estratificación considerando las intersecciones por Distritos Federales y tipo de sección.

Para tomar una decisión final de cuál es la mejor estratificación a emplear, se debe considerar el impacto de varios factores como:

1. La precisión esperada (límite máximo deseable para el error de estimación) que se pretende obtener en las estimaciones reportadas. En términos prácticos es la mitad de la longitud esperada de los intervalos de estimación que se desea obtener.
2. Porcentaje de CAE que participarán reportando 1 sola casilla para el Conteo Rápido.
3. El tamaño de la muestra que se va a seleccionar.
4. El tamaño de muestra asignado en cada uno de los estratos.

5. Un nivel de confianza deseado en las estimaciones, el cual se ha fijado en 95%.

Se debe mencionar que, para la definición del tamaño de muestra, adicionalmente se deben considerar factores como:

- El Proceso de llegada y la calidad de la información muestral (características de la muestra recibida y la hora de llegada).
- El Tamaño de la no-respuesta en ejercicios de Conteo Rápido anteriores.
- La composición de la muestra recibida. Aunque se menciona poco, este aspecto, la composición de la muestra recibida, es igualmente relevante que la cantidad (porcentaje) recibida. Por este motivo, más adelante se incluye un apartado específico sobre este aspecto.

Por otro lado, también se debe mencionar que en la definición del Diseño muestral se ha considerado tener Diseños muestrales donde un alto porcentaje de los CAE (al menos 80%) apoyando al Conteo Rápido, solo reporte 1 casilla. Lo anterior, ha sido una consideración que los miembros del COTECORA han asumido en la definición técnica del Diseño muestral en los últimos Conteos Rápidos. Lo anterior, con el fin de contribuir a que la presión de los CAE el día de la jornada electoral no sea excesiva y, con esto propiciar condiciones favorables (desde el punto de vista técnico-estadístico) para lograr la recepción de un mayor porcentaje de muestra el día de la jornada electoral.

A continuación, se presentan algunos conceptos probabilísticos para ir complementando algunos de los puntos antes mencionados.

Algunas ideas básicas del muestreo probabilístico para el Conteo Rápido

Uno de los objetivos técnicos del Conteo Rápido es estimar la proporción p de votos a favor de cada uno de los candidatos a la Gobernatura¹¹ del Estado de Puebla. Para lo anterior, se selecciona una muestra aleatoria (bajo algún esquema de muestreo o estrategia de selección) de n casillas de un total de N y con la información muestral disponible se calcula un estimador \hat{p} . Mediante muestreo probabilístico es posible analizar el impacto de diferentes estrategias de selección, tamaños de muestra n y estimadores, para asegurar (teóricamente) que:

¹¹ Esto se hará para cada contendiente, así como, para la estimación de la votación en favor de candidatos no registrados y votación nula; sin embargo, para efectos de la explicación nos concentraremos en la estimación para un solo contendiente.

$$|p - \hat{p}| \leq d, \quad (1)$$

con un 95% de confianza. A la cantidad d comúnmente se le conoce como precisión o límite máximo aceptable para el error de estimación. La expresión (1) se puede escribir de forma equivalente en términos de un intervalo de confianza al $(1-\alpha)100\%$, como

$$\hat{p} - d \leq p \leq \hat{p} + d \quad (2)$$

Esto significa, por ejemplo, que si $(1-\alpha)=0.95$, $d=0.02$, y se extraen 1,000 muestras distintas e independientes una de otra (cada una siguiendo la misma estrategia de selección, usando el mismo tamaño de muestra y el mismo estimador) y, con cada muestra se hace una estimación, entonces se tendrían 1000 estimaciones puntuales $\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5, \dots, \hat{p}_{1000}$, de las cuales se esperaría que el 95% de estas estimaciones cumpliera con la condición (1). Es decir, que su distancia máxima (error de estimación) respecto al verdadero valor será de 2 puntos porcentuales (2%) con una confianza del 95%.

Precisión en las estimaciones

El procedimiento anterior permite analizar diferentes diseños muestrales para obtener las precisiones bajo varias estrategias de selección, diferentes tamaños de muestra y diferentes estimadores. En consecuencia, mediante estudios de simulación podemos decidir cuál es el diseño de muestreo más adecuado para llevar a cabo el procedimiento correspondiente del Conteo Rápido en la elección para la Gobernatura del Estado de Puebla.

A continuación, se presentan y analizan las precisiones para diferente estratificación y diferentes tamaños de muestra. En todos los análisis realizados se tomó como referencia los cómputos distritales de la elección a la Gobernatura del Estado de Puebla en 2018, y los cómputos distritales de la elección extraordinaria a la Gobernatura del Estado de Puebla en 2019. Lo anterior, considerando la distritación 2023. Las estratificaciones consideradas en la definición del diseño muestral fueron:

1. Distritos Federales \Rightarrow 16 estratos
2. Distritos Locales \Rightarrow 26 estratos
3. Distritos Federales x Distritos locales \Rightarrow 48 estratos
4. Distritos Federales x Tipo sección \Rightarrow 30 estratos
5. Distritos Federales x Distritos locales \Rightarrow 36 estratos
6. Con algunos estratos colapsados o unidos

Para el análisis de las precisiones, se consideró el estimador de razón

$$\hat{p}_i = \frac{\sum_{h=1}^L N_h \bar{y}_{hi}}{\sum_i \sum_{h=1}^L N_h \bar{y}_{hi}} \quad (3)$$

En donde el sub-índice i representa al candidato i -ésimo (incluyendo también el número de votos por candidatos no registrados y nulos), h es el número de estrato, L el número de estratos totales, N_h es el número de casillas consideradas en el estrato h y \bar{y}_{hi} es el número promedio de votos por casilla en muestra, para el candidato i en el estrato h . Las estimaciones se realizaron para los candidatos, candidatos no registrado y votos nulos). Sin embargo, en aras de una mejor exposición sólo se muestra el análisis relativo al candidato que registró una mayor varianza, y por tanto requería un tamaño de muestra mayor, para alcanzar al menos la misma precisión en todos los candidatos.

La Tabla 20 muestra los resultados promedio, basados en un estudio de 5000 simulaciones, de las precisiones alcanzadas para las diferentes estratificaciones y diferentes tamaños de muestra.

Tabla 20 Análisis de precisiones para el CR a la Gobernatura de Puebla

N	DTO. FED (16 ESTRATOS)	DTO. LOC (26 ESTRATOS)	DTO. FED X DTO. LOC (48 ESTRATOS)	DTO. FED X TIPO DE SEC. (30 ESTRATOS)	DTO. FED X DTOS. LOC COLAPSOS (36 ESTRATOS)
250	1.27	1.21	1.17	1.19	1.19
300	1.16	1.13	1.10	1.10	1.10
350	1.08	1.04	1.00	1.02	1.02
400	1.01	0.98	0.95	0.96	0.96
450	0.96	0.93	0.90	0.91	0.91
500	0.91	0.88	0.86	0.87	0.87
550	0.87	0.84	0.82	0.83	0.83
600	0.83	0.81	0.78	0.79	0.80

Se puede observar que, en general las diversas estrategias de estratificación arrojan precisiones similares, excepto la estratificación por distrito federal que arroja los errores de estimación más grandes. También, se puede ver que la estratificación que arroja las mejores precisiones es la del cruce del distrito federal con distrito local. Sin embargo, esta última estratificación representa tener 48 estratos. Por la experiencia en otros ejercicios de CR, esta cantidad de estratos puede llevar a tener un tamaño de muestra pequeño en los estratos. Por lo

anterior, se analizó el caso de colapsar algunos de esos estratos, lo cual produjo los resultados de la última columna de la Tabla 20.

NOTAS:

1. Se debe señalar que los análisis anteriores se hicieron previamente al acuerdo reciente de los miembros del COTECORA sobre la armonización de todas las muestras involucradas en los CR para esta jornada 2024
2. Para cumplir con la armonización de las diferentes muestras que se van a usar en los diferentes Conteos Rápidos (Presidencial-Gubernaturas, Diputados y Senadores) se realizaron nuevos estudios sobre la pertinencia de emplear una estratificación definida solo por los Distritos Federales. La conclusión a la que se llegó es que, una estratificación por Distrito Federal, aunque produce una menor precisión en las estimaciones, presenta un nivel de cobertura aceptable. Por todo lo anterior y con el fin de contribuir a una correcta armonización de las muestras a utilizar, se decidió usar la estratificación por Distrito Federal.

Presión sobre los CAE apoyando el CR

Por otro lado, dada la preocupación en la carga que han tenido los CAE en los últimos ejercicios de Conteo Rápido, se analizó también la proporción de CAE a los que les tocaría reportar 1, 2 y 3 o más casillas en el Conteo Rápido a la Gobernatura del Estado de Puebla. Para este análisis se consideró la información de la elección de 2018, usando la distritación vigente. Los resultados se muestran en la Tabla 21.

Analizando los resultados derivados de la Tabla 21, se puede ver que si se consideran tamaños de muestra de 450 o 500 aún se tienen porcentajes altos (al alrededor del 86.1%-87.6) de CAE a los cuales solo les tocaría reportar 1 casilla, considerando la estratificación por Distrito Federal.

Tabla 21 Porcentajes de CAE con 1, 2, y 3 o más casillas

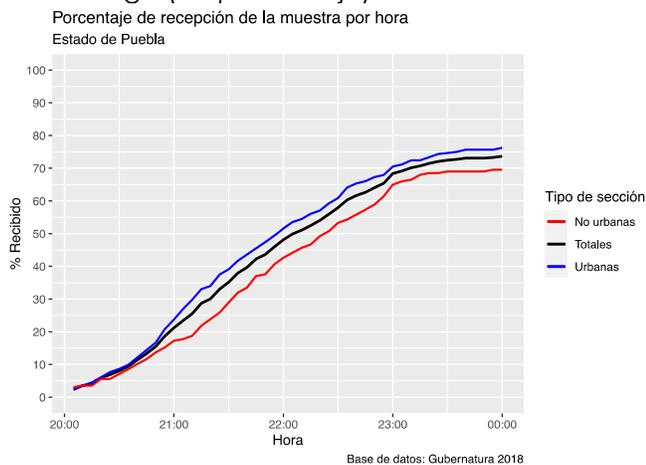
N	CASILLAS X CAE	DTO. FED (16 ESTRATOS)	DTO. LOC (26 ESTRATOS)	DTO. FED.	DTO. FED.	DTO. FED
				X	X	X
				DTO. LOCAL (48 ESTRATOS)	DTO. LOCAL (COLAPSOS) (36 ESTRATOS)	TIPO SECCIÓN (30 ESTRATOS)
400	1	89.2%	91.2%	91.2%	91.2%	90.1%
	2	10.1%	8.4%	8.4%	8.4%	9.4%
	3 o más	0.7%	0.4%	0.4%	0.4%	0.5%
450	1	87.6%	89.9%	90.1%	90%	88.9%

N	CASILLAS X CAE	DTO. FED (16 ESTRATOS)	DTO. LOC (26 ESTRATOS)	DTO. FED.	DTO. FED.	DTO. FED.
				X	X	X
				DTO. LOCAL (48 ESTRATOS)	DTO. LOCAL (COLAPSOS) (36 ESTRATOS)	TIPO SECCIÓN (30 ESTRATOS)
	2	11.4%	9.5%	9.4%	9.5%	10.5%
	3 o más	1%	0.6%	0.5%	0.5%	0.6%
	1	86.1%	88.7%	89%	88.7%	87.8%
500	2	12.6%	10.6%	10.3%	10.6%	11.4%
	3 o más	1.3%	0.7%	0.7%	0.7%	0.8%

El Proceso de llegada de la información muestral

Un aspecto importante en la determinación del tamaño de muestra es la calidad de la información recibida (y en general, también la estrategia de selección). En este sentido, se analizó el proceso de llegada (tiempo) de la información. Para tener una perspectiva de este aspecto, se analizó la hora de llegada de la muestra relativa al Estado de Puebla en el CR de la elección de 2018. La Gráfica 9 muestra este proceso de llegada, diferenciando el proceso para casillas en secciones urbanas y no urbanas, así como, para el caso general.

Gráfica 9 Proceso de llega (en porcentaje) de las casillas solicitadas en muestra



En 2018, el número de casillas en muestra en el estado de Puebla fue de 509. De estas casillas a las 10:00 pm (hora del centro) del día de la jornada electoral, solo se habían recibido el 48.1% de la muestra solicitada. A las 11:00 pm se tenía el 68.4% y a las 23:45 se tenía 73.1% de la muestra solicitada. Este análisis es relevante ya que muestra la preocupación sobre el flujo de llegada de la información para el COTECORA. Si esta tasa de llegada de la información se presenta en la jornada

electoral del 2 de junio de 2024, se podría dar una situación donde las estimaciones no tengan la precisión y cobertura adecuadas, antes de las 11:30 pm.

Por otro lado, reconociendo que la Jornada Electoral a la Gobernatura del Estado de Puebla en 2018 tuvo algunas vicisitudes, se analizó el proceso de llegada de la información muestral en la Elección extraordinaria de 2019. En este proceso se solicitó un tamaño de muestra de 522 casillas. La información de la primera casilla que recibió el COTECORA fue a las 18:30 y a las 20:05 (hora de salida del Conteo Rápido) ya se había recibido el 82.2% (429 casillas). Es decir, comparado con el proceso de llegada de la información muestral de 2018, el proceso de llegada en 2019 fue totalmente diferente. Por ese y otros motivos, nos dimos la tarea de analizar el proceso de llegada de la información muestral de algunos otros estados en donde se realizaron Conteos Rápidos en 2021. A continuación, se presentan estos resultados en la Tabla 22.

Tabla 22 Muestra solicitada vs. Muestra recibida, CR 2019

	GUERRERO	NUEVO LEÓN	SINALOA	TLAXCALA	BAJA CALIFORNIA SUR
% de CAE con 1 casilla	78%	79%	76%	73.4	70.4
% Muestra recibida	72.8%	65.9%	65.1%	72.2	72.12
Hora de Salida	23:30	23:10	23:00	23:40	23:00 hr. local

Hay varias conclusiones que se pueden derivar del análisis de la Tabla 22. Una de estas conclusiones es que, en general, prácticamente no se recibe al menos un tamaño de muestra igual al porcentaje de CAE a los que les toca reportar solo 1 casilla. Otra conclusión es, que en procesos locales en elecciones concurrentes (no es el caso de 2021, pero se puede inferir) en promedio no se recibe más del 70% de la muestra solicitada.

Con los avances tecnológicos que ha tenido el INE y el compromiso de todas sus áreas, para que en este proceso 2023-2024 se reciba un mayor porcentaje de muestra para los Conteos Rápidos locales, nosotros consideramos un porcentaje de muestra a recibir del 80%.

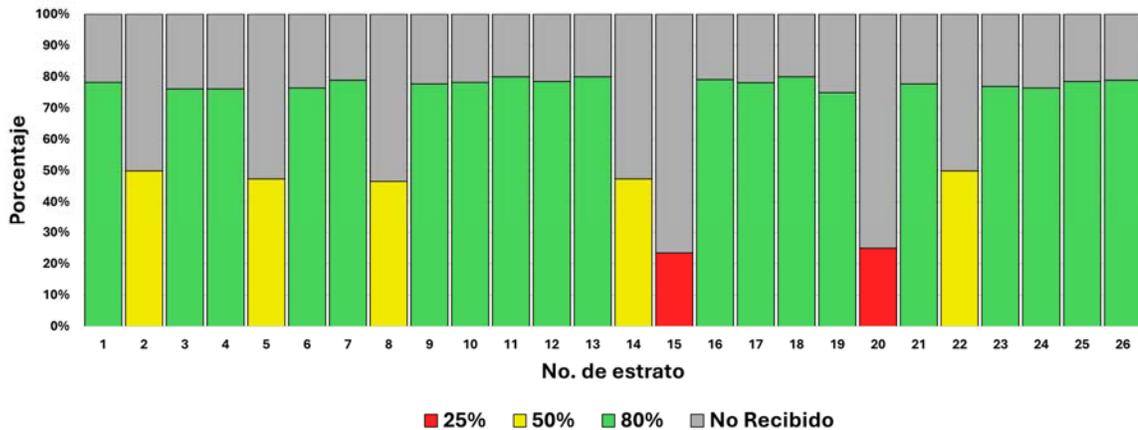
Calidad (COMPOSICIÓN) de la información muestral recibida

Algunas de las preguntas relevantes en los ejercicios de Conteo Rápido asociados a una Jornada Electoral son ¿Con qué porcentaje de la muestra se

pueden dar resultados?, ¿A qué hora se podría salir? Las preguntas anteriores están muy correlacionadas, pero, un aspecto que no hay que descuidar y que es más importante que el porcentaje de la muestra recibida es la Composición de la muestra. No es lo mismo recibir una muestra del 80% donde para algunos estratos no se tenga información, que contar con una muestra del 80% de casillas donde todos los estratos tienen una cantidad de casillas recibidas más homogénea.

Por la gran importancia de la Composición de la información recibida, nosotros hemos realizado desde ejercicios anteriores de Conteo Rápido, un análisis del impacto de la Composición de la muestra recibida en aspectos como la precisión esperada de las estimaciones, así como, de la cobertura esperada (probabilidad de que el intervalo estimado contenga al verdadero valor). Solo a manera de ejemplo, presentamos los resultados de una simulación para el Conteo Rápido del Estado de Puebla, con una muestra completa de 500 casillas versus los resultados de una muestra incompleta al 70%, pero donde el 10% de los Estratos recibe solo el 25% de la muestra, el 20% de los estratos recibe solo el 50% de los estratos y el 70% de los estratos recibe solo el 80% de la muestra.

Gráfica 10 Muestra solicitada vs. Muestra recibida



MUESTRA COMPLETA	MUESTRA RECIBIDA (70%)
500 casillas	350 casillas
Precisión esperada: 0.88	Precisión esperada: 1.15
Cobertura: 95%	Cobertura: 93%

Como una conclusión general se puede decir que se pierde más precisión y cobertura en una muestra sesgada al 70% comparado con una muestra que se recibe de manera homogénea del 70% de la muestra original que equivaldría a una muestra de 350 casillas (ver Tabla 20).

Determinación del tamaño de muestra

Derivado del análisis presentado en todas las secciones anteriores, el día de la jornada electoral estaríamos esperando recibir alrededor de las 11:30 pm al menos el 80% de la muestra solicitada. Adicionalmente, derivado del análisis de la presión de los CAE que participaran el día de la jornada electoral, podemos decir que si elegimos un tamaño de muestra de 500 casillas al menos el 86.1% de ellos tendrían que reportar solo 1 casilla. Sin embargo, dado que esta Jornada Electoral del 2 de junio de 2024 es la más compleja en la historia y con el afán de contribuir a una mejor dinámica de los correspondientes procesos de Conteo Rápido, decidimos disminuir en un 10% el tamaño de muestra.

Derivado de todos los análisis realizados, el tamaño de muestra solicitado para el Conteo Rápido a la Gubernatura del Estado de Puebla, que nos permitiría tener estimaciones con precisiones adecuadas y con una carga muy moderada a los CAE, es de $n=450$ casillas. Lo anterior, nos llevaría a esperar un tamaño efectivo de muestra recibida de al menos 360 casillas para poder producir resultados definitivos en el caso de recibir una muestra homogénea.

Estratificación final

Al corte del 25 de marzo de 2024 se contaba con 8341 casillas asignadas al Estado de Puebla. Así, la estratificación final por Distrito Federal tomando en cuenta todas las condiciones anteriores, se muestran en la Tabla 23. Se debe mencionar que si hay un ajuste en el número de casillas los valores de N_h y n_h presentarán modificaciones y esto hay que tenerlo muy presente, sobre todo para el sistema elaborado por el área de Cartografía de la DERFE.

Tabla 23 Definición de la Estratificación final para el CR a la Gubernatura del Estado de Puebla.

ESTRATO	DISTRITOS FEDERALES	N_h	n_h
1	1	539	29
2	2	528	28
3	3	526	28
4	4	504	27
5	5	492	27
6	6	529	29
7	7	444	24
8	8	461	25

ESTRATO	DISTRITOS FEDERALES	Nh	nh
9	9	629	34
10	10	503	27
11	11	529	29
12	12	581	31
13	13	487	26
14	14	584	32
15	15	480	26
16	16	525	28
Totales		8341	450

Total de casillas en cada estrato (Nh) y tamaños de muestra en cada estrato (nh). Las Nh y nh son calculadas con base al corte de casillas al 25 de marzo de 2024.

Con la estratificación anterior y usando nuevamente el estimador de razón definido en (3) se espera obtener (teóricamente) una precisión de aproximadamente 1.0 punto porcentual en las estimaciones, con un porcentaje de al menos 87% de CAE con una sola casilla asignada.

Algunas consideraciones importantes

1. Se debe señalar que la precisión en las estimaciones se maneja tanto en la etapa de planeación del ejercicio como en la presentación de los resultados finales del mismo. En la etapa de planeación se usa para determinar un tamaño de muestra adecuado teóricamente, para alcanzar el nivel de precisión y confianza deseadas en las estimaciones. Sin embargo, debido a que se fija este tamaño de muestra tomando como referencia los resultados de elecciones previas y a que se tendrá un porcentaje de no respuesta (real) que no conocemos hasta el día de la jornada electoral, una vez concluida la elección y con toda la información disponible se determinará la precisión realmente obtenida.
2. Se debe señalar que hasta el momento (marzo de 2024) se cuenta con la proyección de contratación y designación de 34 CAE por Distrito Federal para apoyar el CR para la elección a la Gobernatura en el Estado de Puebla (información proporcionada por la DEOE en el Anexo de Capacidad Máxima de CAE para CR 2023-2024). Así, se tiene la capacidad de 544 CAE para apoyar el CR para la elección a la Gobernatura en el Estado de Puebla. De esta estimación de CAE solo se emplearían aproximadamente 396 CAE (asociados a una muestra de 450 casillas), de los cuales al menos el 87.6% (345) solo tendrá que reportar 1 sola casilla. Por lo anterior, es pertinente señalar que el tamaño de muestra

de 450 casillas, no solo se encuentra dentro de los límites de la capacidad operativa de campo, sino también representa un gran apoyo a la disminución de carga del conjunto total de CAE disponibles.

3. Cabe que mencionar, por un lado, que, si se recibe un tamaño de muestra menor a 450 casillas, el día de la Jornada Electoral, todavía se pueden tener estimaciones confiables si la distribución de la muestra recibida es homogénea. Sin embargo, las precisiones y las coberturas de las estimaciones por intervalos serán menores. Por otro lado, como lo hemos señalado enfáticamente, si la composición de la muestra recibida presenta un sesgo importante y además se recibe una muestra menor al 80% (de las 450 casillas originalmente solicitadas), difícilmente se podrán producir resultados que ofrezcan estimaciones adecuadas.
4. Finalmente, hay que señalar que los diferentes procedimientos de estimación que implementarán los tres miembros del COTECORA, la Dra. María Fernanda Gil Leyva Villa, la Dra. Martha Elva Ramírez Guzmán y el Dr. Gabriel Núñez Antonio, para la realización del CR del Estado de Puebla, estarán basados en el Diseño muestral definido en este apartado.

Integración de las estimaciones

Cada uno de los tres miembros del grupo (la Dra. María Fernanda Gil Leyva Villa, la Dra. Martha Elva Ramírez Guzmán y el Dr. Gabriel Núñez Antonio) del COTECORA, con la misma información de las casillas recibidas el día de la jornada electoral, realizarán una estimación para la votación de cada uno de los candidatos a la Gubernatura del Estado de Puebla, así como, para el porcentaje de participación. Por lo anterior, para integrar los resultados anteriores se propone usar la unión de al menos dos de las tres estimaciones obtenidas. Sin embargo, la manera definitiva de integrar las estimaciones dependerá de los resultados que se obtengan en futuros ejercicios de simulación y en los simulacros.

7.4.1.7.Tabasco

Dra. Michelle Anzarut Chacalo y Dr. Luis Felipe Gonzalez Perez

Método

El objetivo del conteo rápido es estimar la proporción de votos en favor de cada candidato la misma noche de la elección. Para cumplir este objetivo, se selecciona una muestra aleatoria de casillas del total y, con la muestra recibida, se calculan intervalos del 95% de confianza para la proporción de votos obtenidos por cada candidato, así como para la participación ciudadana. A

continuación explicamos como se seleccionó el diseño de dicha muestra aleatoria para Tabasco.

Evaluamos distintos diseños muestrales: muestreo aleatorio simple y tres diseños de muestra estratificados, en donde al interior de cada estrato se seleccionan las casillas mediante muestreo aleatorio simple sin reemplazo, con la asignación de la muestra proporcional al tamaño de cada estrato. Adicionalmente, para cada estratificación propuesta variamos el tamaño de muestra en un rango de 100 a 500 casillas.

Cada uno de los diseños muestrales se evaluó siguiendo los siguientes pasos:

1. Se simularon 100 muestras utilizando los resultados de las elecciones estatales del 2018 con las variables de distritación correspondientes a 2023.
2. Para cada muestra se estiman intervalos de 95% de confianza para la proporción de votos en favor de cada candidato.
3. Se calculan márgenes de error y se evalúa la carga para los CAE

Finalmente, comparamos los resultados obtenidos en los distintos diseños para elegir la estratificación más conveniente con un criterio que considera tanto un margen de error aceptable como una carga a los CAE razonable.

Estratificaciones evaluadas

Para los diseños estratificados consideramos estratos definidos por:

- Distrito Federal.
- Distrito Local.
- El cruce de Distrito Federal con tipo de sección (urbana o no urbana).

Veamos cuántos estratos hay para cada una de estas estratificaciones:

ESTRATIFICACIÓN	NÚMERO DE ESTRATOS
Muestreo aleatorio simple	1
Distrito Federal	6
Distrito Federal*Tipo Sección	12
Distrito Local	22

Tamaños de muestra evaluados

El tamaño de muestra a elegir está relacionado con el nivel de confianza que queremos obtener, el tamaño de intervalo deseado y, en menor medida, el número de casillas en el marco. Buscamos el tamaño mínimo de muestra que nos permita obtener intervalos de longitudes razonables con un nivel del 95% de confianza.

Para el estado de Tabasco, el marco con el que se realizó este ejercicio cuenta con información de 2,912 casillas. Analizaremos los resultados con muestras de 100, 150, 200, 250, 300, 350, 400, 450 y 500 casillas.

Método de estimación

Sea p_j la proporción de votos a favor del candidato j , estimaremos p_j utilizando el estimador de razón combinado \hat{p}_j , definido como:

$$\hat{p}_j = \frac{\sum_{h=1}^L \frac{N_h}{n_h} (\sum_{i=1}^{n_h} x_{hij})}{\sum_{j=1}^J \sum_{h=1}^L \frac{N_h}{n_h} (\sum_{i=1}^{n_h} x_{hij})},$$

donde, para cada estrato h , n_h es el tamaño de la muestra, N_h es el número total de casillas, L es el número total de estratos, J es el número total de candidatos (incluyendo votos nulos y candidatos no registrados) y x_{hij} es el total de votos a favor del candidato j en la casilla i .

El error estándar de este estimador se calcula utilizando bootstrap no paramétrico (ver, por ejemplo, [1]).

Criterios de decisión

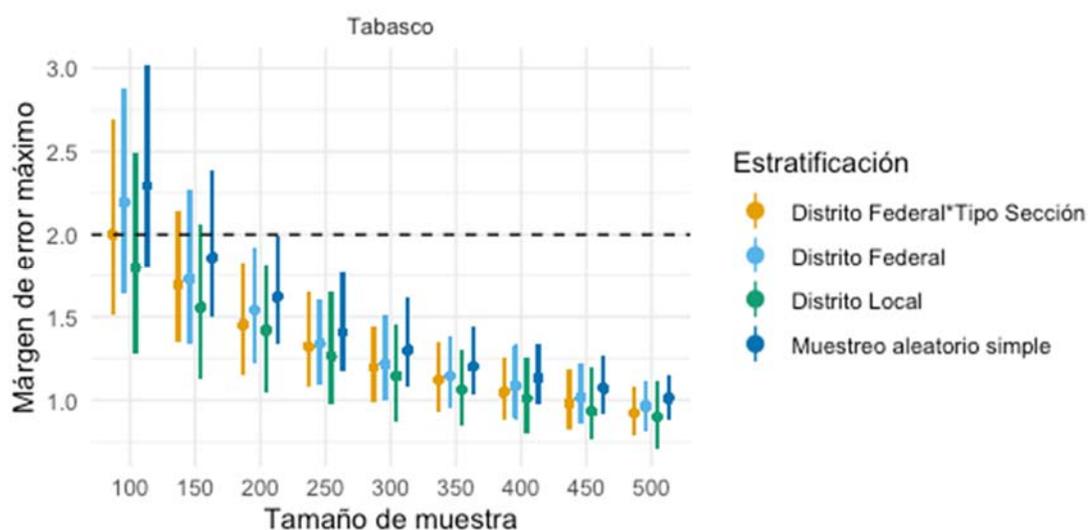
Buscamos que la longitud de los intervalos sea razonable, por lo que como medida de decisión consideraremos la mitad del intervalo más grande sobre todos los candidatos, esto se conoce como margen de error máximo.

Entre mayor es el tamaño de la muestra, menores son los márgenes de error. Por otro lado, muestras más grandes implican un mayor porcentaje de capacitadores-asistentes electorales (CAE) con más de una casilla a reportar, esto complica el trabajo operativo y puede resultar en que no se reporten algunas de las casillas. Por esta razón, el diseño muestral evalúa en conjunto estos dos criterios, con el objetivo de encontrar un balance que optimice el tamaño de los intervalos sin poner en riesgo la obtención de la muestra.

En el mismo sentido que el tamaño de muestra, las estratificaciones más finas, entre las evaluadas, están asociadas a menores márgenes de error. Sin embargo, en las estratificaciones más finas tenemos estratos con muy pocas casillas, que añaden complejidad al proceso de selección de la muestra y al proceso de estimación.

Resultados

En cada simulación registramos el 197argen de error máximo, y graficamos el cuantil 0.05, la media y el cuantil 0.95 de los 197argen de error máximo a lo largo de las simulaciones.

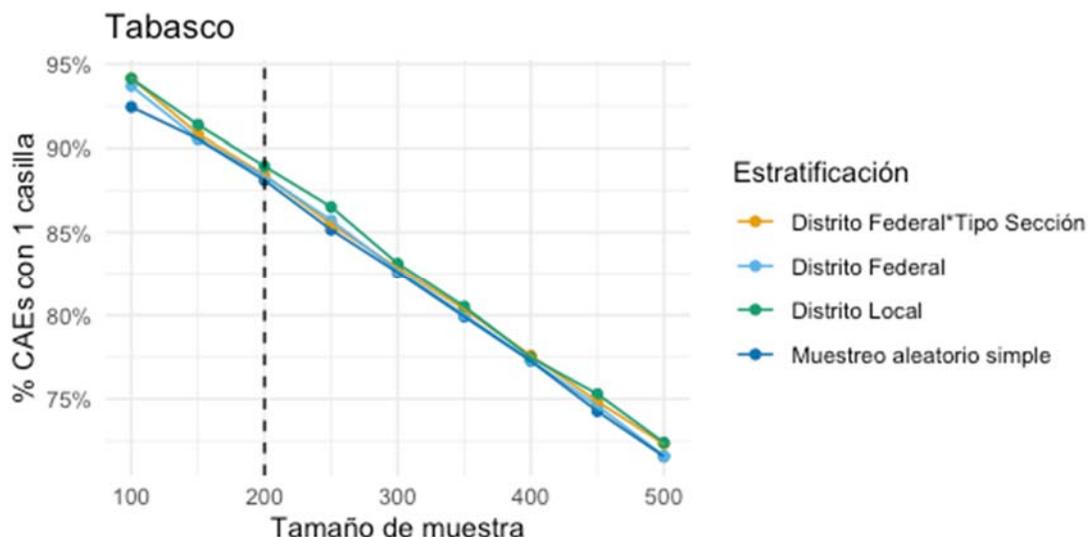


Se busca que el diseño seleccionado tenga un margen de error máximo por debajo de la línea negra que indica 2 puntos porcentuales. Se elige una muestra de 200 casillas (cerca del 7% del total) y la estratificación por Distrito Federal. A pesar de que el día de la elección es posible que no contemos con el total de la muestra al momento de realizar las estimaciones, no se recomienda aumentar el tamaño de muestra para este estado ya que la censura no es aleatoria y al crecer el tamaño de muestra incrementa también el riesgo de que la cobertura real se encuentre por debajo de la cobertura nominal del 95%.

Presión sobre los CAE

El segundo criterio a evaluar es la presión sobre los CAE. En la siguiente gráfica mostramos en promedio cuál es el porcentaje de CAE con 1 casilla para cada diseño considerado.

Para una muestra de 200, la estratificación basada en Distrito Federal presenta aproximadamente un 88% de CAE asignados a una casilla.



Podemos analizar también el porcentaje de CAE con 2 casillas y con 3 o más casillas, para las estratificación y tamaño de muestra seleccionados.

ESTRATIFICACIÓN	TAMAÑO MUESTRA	CAE CON 2 CASILLAS	CAE CON 3 O MÁS CASILLAS
Distrito Federal	200	10%	1%

Consideramos así que 200 casillas es una carga de trabajo aceptable.

Diseño final

A continuación aparece el diseño muestral final. Se elige una muestra de 200 casillas y se propone estratificación de Distrito Federal, mostramos el promedio del margen de error máximo (Promedio MEM) y el porcentaje de CAE con 1 o 2 casillas.

ESTRATIFICACIÓN	TAMAÑO MUESTRA	PROMEDIO MEM	CAE CON 1 CASILLA	CAE CON 2 CASILLAS
Distrito Federal	200	1.5	88%	10%

Referencias

[1] https://tereom.github.io/est_computacional/05-Bootsrap.html

7.4.1.8. Veracruz

Mtra. María Teresa Ortiz Mancera y Lic. Ana María Madrigal Gómez

Método

El objetivo del conteo rápido es estimar la proporción de votos en favor de cada candidato la misma noche de la elección. Para cumplir este objetivo, se selecciona una muestra aleatoria de casillas del total y, con la muestra recibida, se calculan intervalos del 95% de confianza para la proporción de votos obtenidos por cada candidato, así como para la participación ciudadana. A continuación explicamos como se seleccionó el diseño de dicha muestra aleatoria para Veracruz.

Evaluamos distintos diseños muestrales: muestreo aleatorio simple y tres diseños de muestra estratificados, en donde al interior de cada estrato se seleccionan las casillas mediante muestreo aleatorio simple sin reemplazo, con la asignación de la muestra proporcional al tamaño de cada estrato. Adicionalmente, para cada estratificación propuesta variamos el tamaño de muestra en un rango de 100 a 500 casillas.

Cada uno de los diseños muestrales se evaluó siguiendo los siguientes pasos:

1. Se simularon 100 muestras utilizando los resultados de las elecciones estatales del 2018 con las variables de distritación correspondientes a 2023.
2. Para cada muestra se estiman intervalos de 95% de confianza para la proporción de votos en favor de cada candidato.
3. Se calculan márgenes de error y se evalúa la carga para los CAE.

Finalmente, comparamos los resultados obtenidos en los distintos diseños para elegir la estratificación más conveniente con un criterio que considera tanto un margen de error aceptable como una carga a los CAE razonable.

Estratificaciones evaluadas

Para los diseños estratificados consideramos estratos definidos por:

- Distrito Federal.

- Distrito Local.
- El cruce de Distrito Federal con tipo de sección (urbana o no urbana).

Veamos cuántos estratos hay para cada una de estas estratificaciones:

ESTRATIFICACIÓN	NÚMERO DE ESTRATOS
Muestreo aleatorio simple	1
Distrito Federal	19
Distrito Local	30
Distrito Federal*Tipo Sección	37

Tamaños de muestra evaluados

El tamaño de muestra a elegir está relacionado con el nivel de confianza que queremos obtener, el tamaño de intervalo deseado y, en menor medida, el número de casillas en el marco. Buscamos el tamaño mínimo de muestra que nos permita obtener intervalos de longitudes razonables con un nivel del 95% de confianza.

Para el estado de Veracruz, el marco con el que se realizó este ejercicio cuenta con información de 10,595 casillas. Analizaremos los resultados con muestras de 100, 150, 200, 250, 300, 350, 400, 450 y 500 casillas.

Sea p_j la proporción de votos a favor del candidato j , estimaremos p_j utilizando el estimador de razón combinado \hat{p}_j , definido como:

$$\hat{p}_j = \frac{\sum_{h=1}^L \frac{N_h}{n_h} (\sum_{i=1}^{n_h} x_{hij})}{\sum_{j=1}^J \sum_{h=1}^L \frac{N_h}{n_h} (\sum_{i=1}^{n_h} x_{hij})}$$

donde, para cada estrato h , n_h es el tamaño de la muestra, N_h es el número total de casillas, L es el número total de estratos, J es el número total de candidatos (incluyendo votos nulos y candidatos no registrados) y x_{hij} es el total de votos a favor del candidato j en la casilla i .

El error estándar de este estimador se calcula utilizando bootstrap no paramétrico (ver, por ejemplo, [1]).

Criterios de decisión

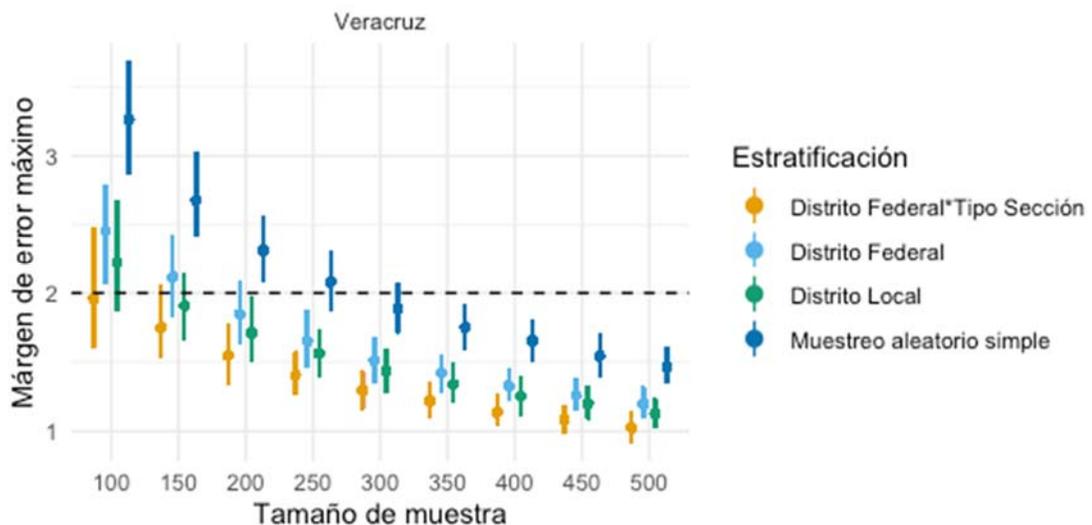
Buscamos que la longitud de los intervalos sea razonable, por lo que como medida de decisión consideraremos la mitad del intervalo más grande sobre todos los candidatos, esto se conoce como margen de error máximo.

Entre mayor es el tamaño de la muestra, menores son los márgenes de error. Por otro lado, muestras más grandes implican un mayor porcentaje de capacitadores-asistentes electorales (CAE) con más de una casilla a reportar, esto complica el trabajo operativo y puede resultar en que no se reporten algunas de las casillas. Por esta razón, el diseño muestral evalúa en conjunto estos dos criterios, con el objetivo de encontrar un balance que optimice el tamaño de los intervalos sin poner en riesgo la obtención de la muestra.

En el mismo sentido que el tamaño de muestra, las estratificaciones más finas, entre las evaluadas, están asociadas a menores márgenes de error. Sin embargo, en las estratificaciones más finas tenemos estratos con muy pocas casillas, que añaden complejidad al proceso de selección de la muestra y al proceso de estimación.

Resultados

En cada simulación registramos el margen de error máximo, y graficamos el cuantil 0.05, la media y el cuantil 0.95 de los márgenes de error máximo a lo largo de las simulaciones.

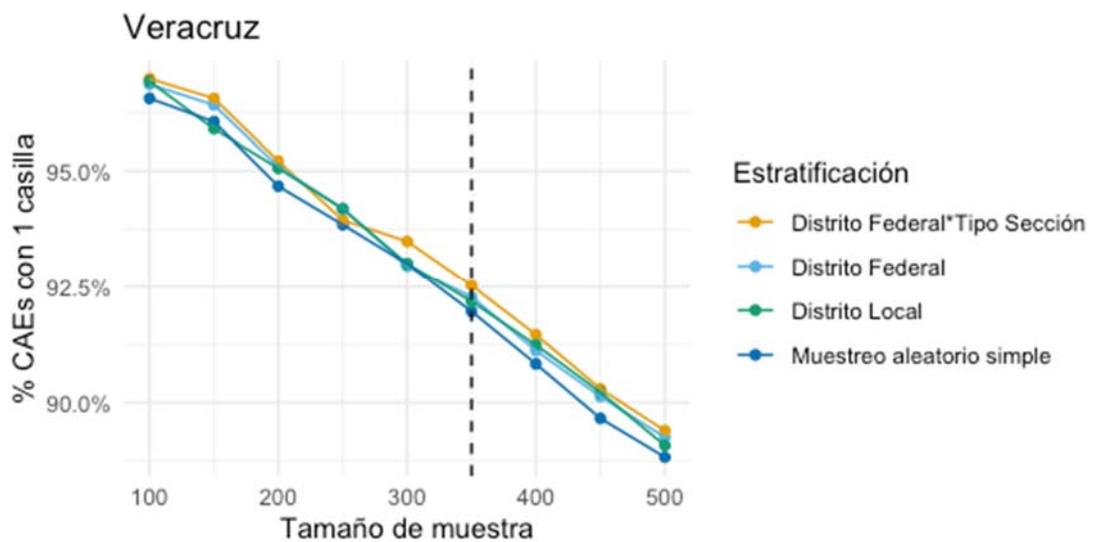


Se busca que el diseño seleccionado tenga un margen de error máximo por debajo de la línea negra que indica 2 puntos porcentuales. Se elige una muestra de 350 casillas (cerca del 3.5% del total) y la estratificación por Distrito Federal.

Presión sobre los CAE

El segundo criterio a evaluar es la presión sobre los CAE. En la siguiente gráfica mostramos en promedio cuál es el porcentaje de CAE con 1 casilla para cada estratificación y los distintos tamaños de muestra.

Para una muestra de 350, la estratificación de Distrito Federal presenta 92% de CAE asignados a una casilla.



Podemos analizar también el porcentaje de CAE con 2 casillas y con 3 o más casillas, para las estratificación y tamaño de muestra seleccionados.

ESTRATIFICACIÓN	TAMAÑO MUESTRA	CAE CON 2 CASILLAS	CAE CON 3 O MÁS CASILLAS
Distrito Federal	350	7%	1%

Consideramos que 350 casillas es una carga de trabajo aceptable.

Diseño final

A continuación aparece el diseño muestral final. Se elige una muestra de 350 casillas y estratificación basada en Distrito Federal, mostramos el promedio del

margen de error máximo (Promedio MEM) y el porcentaje de CAE con 1 o 2 casillas.

ESTRATIFICACIÓN	TAMAÑO MUESTRA	PROMEDIO MEM	CAE CON 1 CASILLA	CAE CON 2 CASILLAS
Distrito Federal	350	1.4	92%	7%

Referencias

[1] https://tereom.github.io/est_computacional/05-Bootsrap.html

7.4.1.9. Yucatán

Dra. María Fernanda Gil Leyva Villa y Dr. Alan Riva Palacio Cohen.

El objetivo general de los ejercicios de Conteo Rápido es a dar certidumbre y confianza a la población en un proceso electoral, ofreciendo una estimación de las tendencias de votación el mismo día de la jornada electoral. A grandes rasgos, este ejercicio consiste en seleccionar una muestra aleatoria de casillas de escrutinio y cómputo cuyos resultados se reportan a los miembros del Comité Técnico Asesor del Conteo Rápido (COTECORA). Con base en esta información, para cada candidato se estima un intervalo que con alta probabilidad contenga a la proporción real votos a favor. Cabe mencionar que el éxito de los Conteo Rápidos depende del trabajo conjunto de varios actores: Los miembros del COTECORA, las diferentes áreas del INE involucradas, los miembros de las Juntas Locales, los miembros de los Organismos Públicos Locales Electorales (OPLEs), entre otros.

Uno de los aspectos clave es el denominado Diseño Muestral, se puede definir como el conjunto de estrategias para determinar el procedimiento bajo el cual se selecciona la muestra aleatoria. El Diseño Muestral debe de tomar en cuenta diversos factores tanto técnicos como operativos buscando garantizar que la muestra cumpla con los criterios establecidos en Reglamento de Elecciones, que sea representativa de la población objetivo y que esta llegue en su completitud a los miembros del COTECORA el día de la jornada electoral. El Diseño muestral también considera procedimientos que minimicen errores de estimación y maximicen la cobertura del método estadístico.

El resto de este documento detalla las consideraciones técnicas que permitieron definir el Diseño muestral que se empleará para el Conteo Rápido en la elección del 2 de junio de 2024 a la Gobernatura del Estado de Yucatán.

Estrategia de Muestreo

Una muestra de una población objetivo es simplemente un subconjunto aleatorio de esta población, la estrategia de muestreo establece como será seleccionado este subconjunto aleatorio. Esta, queda determinada por dos factores: el tamaño de la muestra y como se repartirá en la población objetivo. En el contexto presente, la población objetivo está conformada por todas las casillas aprobadas para el día de la jornada electoral. De forma que, si N es número total de casillas aprobadas, el tamaño de la muestra es simplemente un número m menor que N . En cuanto a la repartición de estas m casillas en la población, se ha optado por un Muestro Aleatorio Simple Estratificado (MASE), considerando el Artículo 373 del Reglamento de Elecciones (donde se establece que la muestra debe abarcar la mayor dispersión geográfica posible) y tomando en cuenta el éxito que ha tenido el MASE en anteriores ejercicios de Conteo Rápido. El MASE consiste en definir estratos (subconjuntos de la población objetivo) que satisfacen que cada casilla pertenece a exactamente un estrato. En otras palabras, los estratos no se traslapan y la unión de todos los estratos es la población objetivo. Una vez definida la estratificación, la muestra de tamaño m se reparte en los estratos de manera proporcional al tamaño de cada estrato. Esto es si hay H estratos y el número de casillas en el estrato h es N_h , se seleccionan $m_h = m * (N_h/N)$ casillas en el estrato h , de forma que $m_1 + \dots + m_h = m$. Finalmente, dentro de cada estrato, las m_h casillas son seleccionadas de acuerdo a un muestreo aleatorio simple, es decir, al azar.

Variables a determinar

Una vez hechas las precisiones anteriores, queda claro que para determinar la estrategia de muestreo basta con definir el tamaño de la muestra y la estratificación.

Tomando en consideración características tanto geográficas como poblacionales propias del estado de Yucatán, se ha optado por evaluar tamaños de muestra que van desde las 50 casillas hasta las 400 casillas y las siguientes estratificaciones:

- Simple: Consiste de un solo estrato que es la población total.
- Por Distrito Federal (FED): Consiste de 6 estratos que son los Distritos Federales (2023).
- Por Distrito Federal y tipo de sección (FEDxSEC): Consiste de 9 estratos, de los cuales 3 estratos coinciden con un Distritos Federal (2023), los otros 6

estratos son los 3 Distritos Federales restantes refinados por el tipo de sección (urbana/no urbana).

- Por Distrito Local (LOC): Consiste de 21 estratos que coinciden con los Distritos Locales (2023).

Para tomar una decisión final sobre el tamaño de muestra y la estratificación a emplear, se realizan extensos estudios de simulación que consisten en generar muestras, por medio de MASE, para cada combinación de tamaño de muestra y estratificación. Con ello se analiza el error de estimación, la cobertura del método estadístico y la carga de trabajo de campo que cada combinación (tamaño de muestra y estratificación) representa. Estos estudios se han realizado sobre datos simulados y para los datos de las Elecciones a la Gubernatura de Yucatán 2018 (YUC18). Los detalles se proporcionan a continuación.

Detalles del estudio de Simulación

Para cada tamaño de muestra ($m = 50, 100, \dots, 400$) y para cada estratificación (Simple, FED, FEDxSEC y LOC) se repite el siguiente experimento que consta de tres pasos:

1. Se genera una muestra de tamaño m mediante el MASE.
2. Usando la muestra generada y mediante Bootstrap Bayesiano se calcula:
 - a. Un estimador puntual, \hat{P}_j , de la proporción de votos al candidato j , P_j . Cabe mencionar que este estimador está centrado en el estimador de razón combinada:

$$\frac{\sum_h m_h \bar{x}_{h,j}}{\sum_j \sum_h m_h \bar{x}_{h,j}}$$

donde $\bar{x}_{h,j}$ es el promedio de votos al candidato j de las casillas en muestra del estrato h , y recordando que m_h es el número de casillas en muestra en el estrato h .

- b. Un intervalo $[a_j, b_j]$ que contengan a la proporción de votos, P_j , con una probabilidad de al menos 0.95.
3. Para la muestra y las estimaciones obtenidas en los pasos anteriores se analizan:
 - a. El error máximo de estimación: $\varepsilon = \max_j |\hat{P}_j - P_j|$.
 - b. La longitud los intervalos: $l_j = b_j - a_j$, para cada candidato.

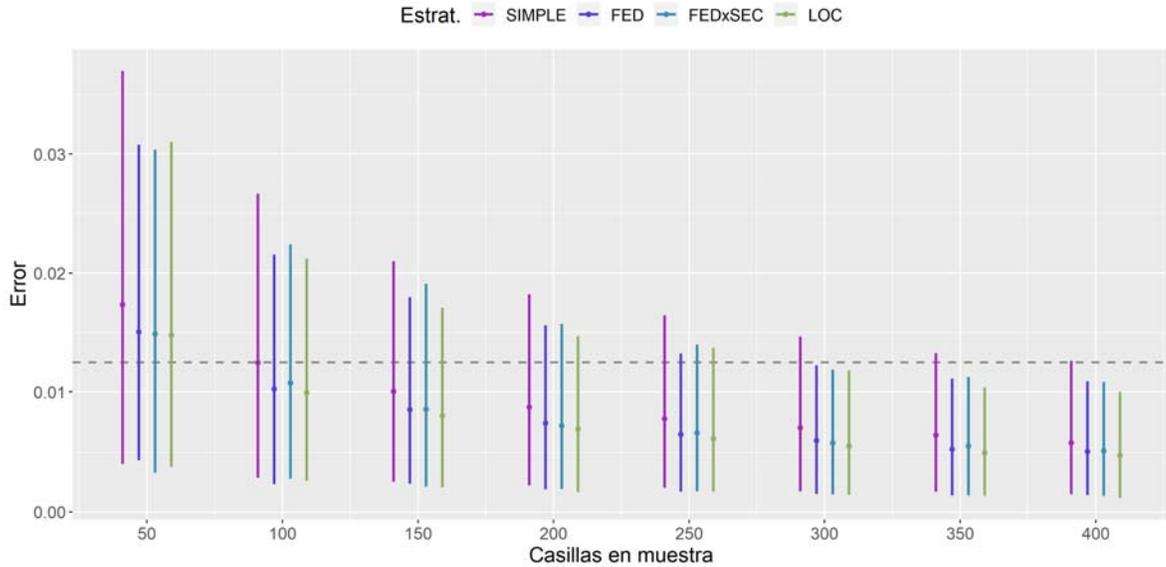
- c. La cobertura de los intervalos: Si el valor real de P_j pertenece a $[a_j, b_j]$, para cada candidato.
- d. La carga a los Capacitadores Asistentes Electorales (CAE): La proporción de CAE a los que les corresponde reportar los resultados de 1, 2 o más casillas para el Conteo Rápido.

El experimento (1-3) se repite 1000 veces para cada combinación de tamaño de muestra y estratificación. A continuación se presentan los resultados obtenidos para la base de datos YUC18.

Análisis del error de estimación

En la Figura 39 se muestran los resultados de los errores de estimación, ε , para la base de datos YUC18. En esta figura, para cada tamaño de muestra y estratificación se exhibe una línea con un punto. El punto representa el promedio de las observaciones de ε . Las partes inferior y superior de la línea representan los cuantiles 0.025 y 0.975, respectivamente, correspondientes a la distribución empírica de las observaciones de ε obtenidas en los 1000 experimentos. Aquí observamos que la estratificación Simple es la que exhibe mayores errores de estimación. Mientras que para las otras tres estratificaciones el cambio en los errores de estimación no es significativa. Así mismo, se aprecia que, a partir de un tamaño de muestra de 300 casillas, y para las estratificaciones FED, FEDxSECC y LOC, el error de estimación es menor que 0.0125 en más del 97.5% de los experimentos (las líneas correspondientes a estas estratificaciones y tamaños de muestra están por debajo de la línea punteada). Esta figura sugiere que un tamaño de muestra de 300 o más casillas y cualquiera de las estratificaciones FED, FEDxSECC y LOC son adecuadas.

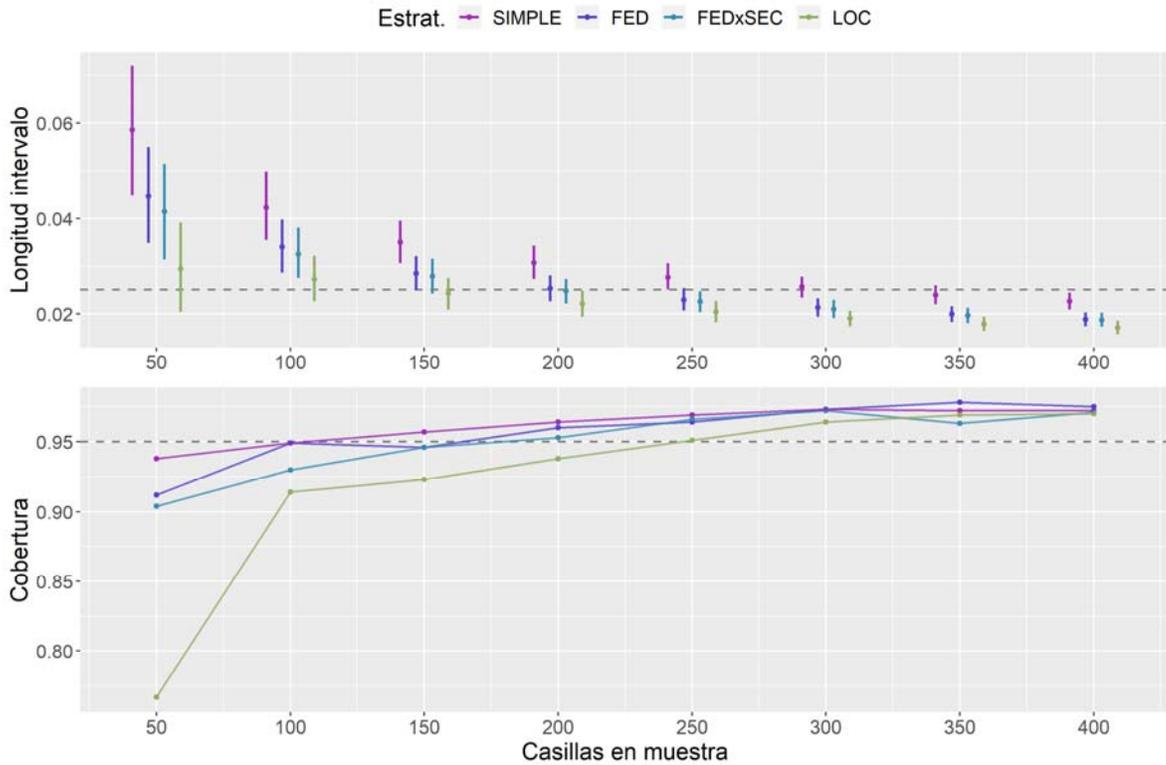
Figura 39 Promedio y cuantiles 0.025 y 0.975 de los errores de estimación, ϵ , por tamaño de muestra y estratificación, para la base de datos YUC18.



Análisis de los intervalos estimados

Antes de ahondar en los resultados de los intervalos estimados cabe mencionar que, a grandes rasgos, las mayores coberturas corresponden a intervalos más grandes. Es decir, si se estiman dos intervalos distintos para la misma cantidad y centrados en el mismo valor, el intervalo más amplio estará dotado de una mejor cobertura. Idealmente, se estimarían intervalos pequeños con amplia cobertura, esto se puede lograr (para cualquier estratificación y método estadístico correctamente aplicado) cuando los tamaños de muestra son suficientemente grandes. Aquí lo que se busca es un balance entre la longitud y la cobertura de los intervalos.

Figura 40 Resultados de los intervalos estimados, $[a_j, b_j]$, para el candidato que presenta los intervalos estimados más grandes, por tamaño de muestra y estratificación, y para la base de datos YUC18.

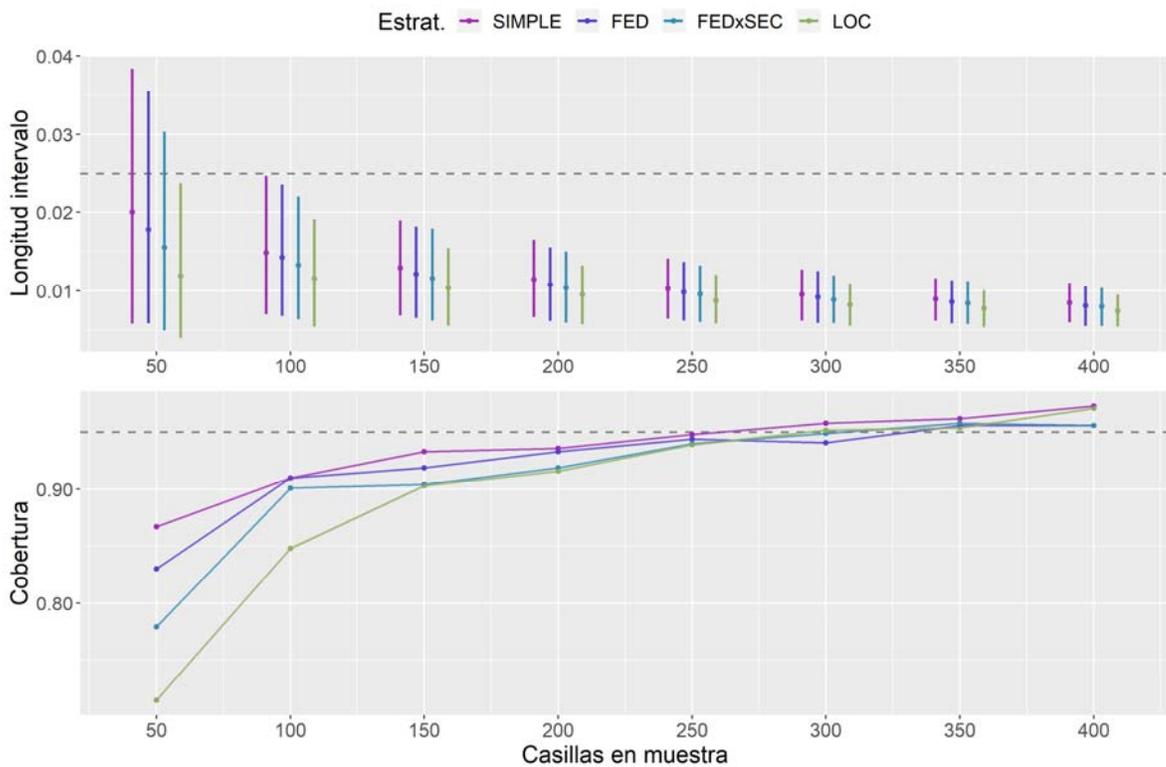


A continuación, se reportan los resultados de dos candidatos. Uno de ellos es el que presenta los intervalos estimados de mayor longitud (Figura 40), al otro le corresponden los intervalos con menor cobertura (Figura 41). La gráfica en la parte superior de las Figuras 39 y 40 exhibe una línea con un punto para cada tamaño de muestra y estratificación. El punto representa el promedio de las observaciones de $l_j = b_j - a_j$. Las partes inferior y superior de la línea representan los cuantiles 0.025 y 0.975, respectivamente, correspondientes a la distribución empírica de las observaciones de l_j obtenidas en los 1000 experimentos. En la gráfica inferior de ambas figuras se exhibe una línea para cada estratificación. Los puntos representan la cobertura de los intervalos para cada estratificación y tamaño de muestra. Esto es, la proporción de experimentos para los que se cumple que el valor real de P_j , pertenece al intervalo estimado $[a_j, b_j]$.

Para el candidato con intervalos más grandes, en la Figura 40, se puede observar que la estratificación Simple es la que arroja intervalos de mayor longitud, mientras que la estratificación LOC estima intervalos más pequeños. A partir de tamaños de muestra de 150 casillas, las estratificaciones FED y FEDxSECC arrojan intervalos de longitud similar para este candidato. También se aprecia que, a partir de las 250 casillas, la longitud de los intervalos estimados es menor que 0.025

más del 97.5% de las veces, para las estratificaciones LOC, FED y FEDxSECC. Otra observación que vale la pena resaltar es que si bien la estratificación LOC es la que exhibe los intervalos más pequeños también es la estratificación que sufre de una menor cobertura.

Figura 41 Resultados de los intervalos estimados, $[a_j, b_j]$, para el candidato que presenta la menor cobertura, por tamaño de muestra y estratificación, y para la base de datos YUC18.



Para el candidato con menor cobertura, en la Figura 41 se aprecia que, a partir de las 300 casillas, la cobertura es adecuada para todas las estratificaciones. Esto debido a que la proporción real de votos al candidato correspondiente, pertenece al intervalo estimado aproximadamente el 95% de las veces. Para tamaños de muestra menores a 200 observamos que la cobertura se ve deteriorada. Al igual que en la Figura 40, aquí notamos que la estratificación LOC tiene la cobertura más pobre para tamaños de muestra pequeños. Esto no es un detalle menor, ya que es importante que el método estadístico este protegido, en la medida de lo posible, ante eventualidades que conlleven a no recibir la muestra completa el día de la jornada electoral. En cuanto a las longitudes de los intervalos estimados para este candidato, observamos que estas son pequeñas para todos los tamaños de muestra mayores a 100 casillas y para todas las estratificaciones.

Este análisis sobre los intervalos estimados sugiere que tamaños de muestra mayores o iguales a 300 casillas son adecuados, así como las estratificaciones FED y FEDxSECC, con una ligera inclinación por la estratificación FED debido a que para tamaños de muestra pequeños esta exhibe una mejor cobertura.

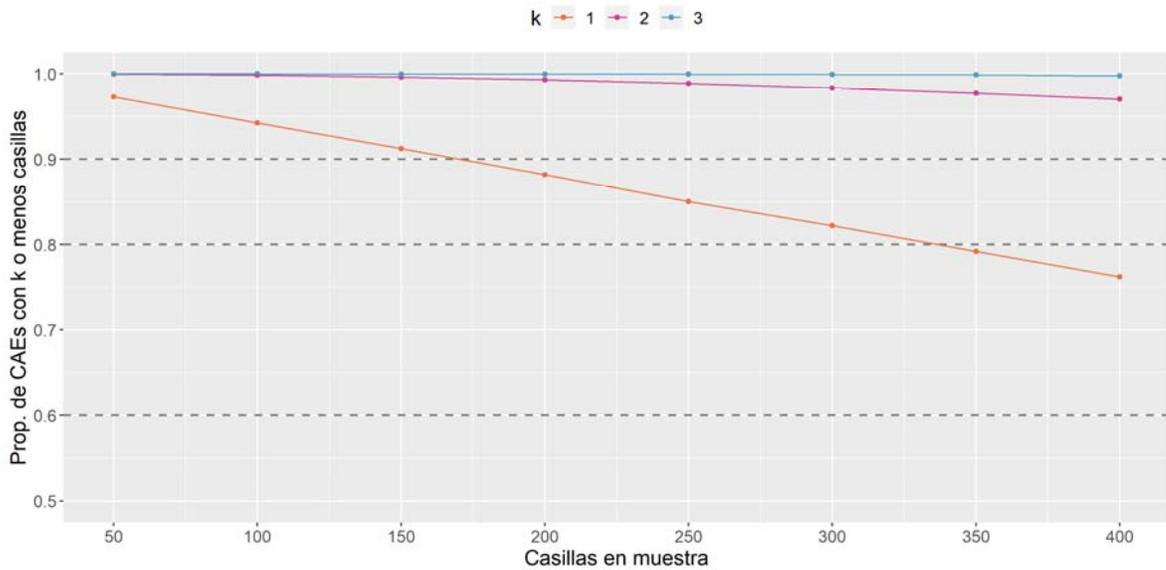
Análisis de la carga de trabajo a los CAE

Tomando en cuenta los resultados de los errores de estimación y de los intervalos estimados, consideramos que tamaños de muestra mayores o iguales a 300 casillas son adecuados. Así mismo, hemos optado por la estratificación FED. Esta elección es tomada considerando únicamente criterios técnico-estadísticos, sin embargo, queda por estudiar el trabajo de campo que representa. Los miembros del COTECORA, conjuntamente con distintas áreas del INE, han acordado que resulta conveniente que se satisfagan los siguientes dos criterios:

1. Al menos el 80% de los CAE deben reportar los resultados de una única casilla en muestra.
2. Al menos el 95% de los CAE deben reportar los resultados de no más de 2 casillas.

Considerando una estratificación por Distrito Federal (FED), en la Figura 42 observamos la carga de trabajo de campo que distintos tamaños de muestra conllevan. La línea naranja describe la proporción de CAE con exactamente 1 casilla en muestra, vemos que para tamaños de muestra menores o iguales a 300 se satisface que esta línea está por arriba del 80%. Además, resaltamos que, para un tamaño de muestra de 350 casillas, la proporción de CAE con una casilla en muestra está apenas por debajo del 80%. La línea morada exhibe la proporción de CAE a los que corresponde reportar los resultados de 1 o 2 casillas, esta línea se encuentra por arriba del 95% para todos los tamaños de muestra. Finalmente, la línea azul corresponde a la proporción de CAE que tiene 3 o menos casillas en muestra, observamos que para muestras de tamaño 350 o menos casillas, prácticamente todos los CAE están en este escenario.

Figura 42 Resultados de la carga a los CAE para la base de datos YUC18, considerando la estratificación FED y para distintos tamaños de muestra.



Consideraciones adicionales y decisión sobre la estrategia de muestreo

Aunado a los resultados de los estudios de simulación, para tomar una decisión final sobre el tamaño de muestra y la estratificación a emplear, es importante tomar en cuenta los siguientes puntos:

Para el año 2024, habrá un número mayor de casillas y de CAE en Yucatán, en comparación al año 2018, mismo que fue considerado para el estudio de simulación realizado. Esto significa que, un mismo tamaño de muestra corresponderá a una menor proporción de casillas en muestra en 2024. Así mismo, la carga de trabajo campo será menor en 2024 que la aquí reportada.

Es posible que el escenario que se presente en 2024 sea más complejo de estimar que el que ocurrió en 2018. En este caso se requeriría un mayor tamaño de muestra para lograr resultados equivalentes a los aquí presentados, bajo la hipótesis de que la muestra llega completa a los miembros del COTECORA el día de la jornada electoral.

Debido a esto se ha llegado a la determinación de elegir un tamaño de muestra de 350 casillas, así como una estratificación por Distrito Federal (FED).

Integración de las estimaciones

Concluimos este documento aclarando que cada uno de los tres miembros del grupo (la Dra. María Fernanda Gil Leyva Villa, la Dra. Martha Elva Ramírez Guzmán y el Dr. Gabriel Núñez Antonio) del COTECORA, con la misma información de las casillas recibidas el día de la jornada electoral, realizarán una estimación para la votación de cada uno de los candidatos a la Gubernatura del Estado de Yucatán, así como, para el porcentaje de participación. Por lo anterior, para integrar los resultados anteriores, se propone usar la unión de al menos dos de las tres estimaciones obtenidas. Sin embargo, la manera definitiva de integrar las estimaciones dependerá de los resultados que se obtengan en futuros ejercicios de simulación y en los simulacros.

7.4.2. Procedimiento de estimación

7.4.2.1. Chiapas

Dra. Lizbeth Naranjo Albarrán, Dra. Ruth Selene Fuentes García y
 Act. Miguel Ángel Chong Rodríguez.

Método de Estimación.

Se utiliza el método normal ponderado por la lista nominal, propuesto por Mendoza y Nieto-Barajas (2016).

Sea θ_{ij} la proporción de personas en la lista nominal del estrato i cuya preferencia es por el candidato j , donde $i = 1, \dots, N$ y $j = 1, \dots, J$, N es el número de estratos, J es el número total de candidatos, incluyendo los candidatos no registrados, los nulos, y las abstenciones. Sea n_i la lista nominal de votantes en el estrato i , para $i = 1, \dots, N$, con $n = \sum_{i=1}^N n_i$ el tamaño de la lista nominal total.

La proporción de votos, relativa a lista nominal, a favor del candidato j está dada por:

$$\theta_j = \sum_{i=1}^N \frac{n_i}{n} \theta_{ij}$$

La proporción efectiva de votos a favor del candidato j se puede calcular como:

$$\lambda_j = \frac{\theta_j}{\sum_{l=1}^{J-1} \theta_l}$$

Sea X_{ij}^k el número de personas a favor del candidato j , en la casilla k del estrato i , donde $k = 1, \dots, K_i$, y K_i es el número de casillas del i -ésimo estrato. El modelo

normal ponderado por lista nominal supone que X_{ij}^k sigue una distribución normal, en específico:

$$X_{ij}^k | \theta_{ij}, \tau_{ij} \sim N\left(n_i^k \theta_{ij}, \frac{\tau_{ij}}{n_i^k}\right)$$

donde n_i^k es el tamaño de la lista nominal de la casilla k en el estrato i , y τ_{ij} es un parámetro de precisión para $k = 1, \dots, K_i$, $i = 1, \dots, N$ y $j = 1, \dots, J$.

Las distribuciones iniciales para los parámetros θ_{ij} y τ_{ij} se consideran independientes y no informativas, es decir:

$$p(\theta_{ij}, \tau_{ij}) = U(\theta_{ij} | 0, 1) \text{Ga}(\tau_{ij} | 0.5, 0.05)$$

Usando el Teorema de Bayes se calcula la distribución posterior de θ_{ij} y τ_{ij} (Berdardo y Smith, 2000). Los estimadores puntuales y por intervalo para θ_{ij} se aproximarán mediante simulaciones de la distribución posterior por métodos de Monte Carlo (Chen et al., 2000).

Referencias

- Bernardo, J.M. y Smith, A.F.M. (2000). Bayesian Theory. Wiley: Chichester.
- Chen, M.-H., Shao, Q.-M. e Ibrahim, J.G. (2000). Monte Carlo methods in Bayesian computation. Springer, New York.
- Mendoza, M. y Nieto-Barajas, L.E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. Electoral Studies 43, 124-132.

7.4.2.2. Ciudad de México

Dr. Raúl Rueda Díaz del Campo y Mtro. Javier Santibáñez Cortés

El objetivo de cualquier Conteo Rápido es estimar la proporción de votos de cada candidato, y la participación ciudadana, basándose en una muestra de casillas. El diseño utilizado consiste en dividir a la población objetivo en k estratos y seleccionar una muestra de tamaño m_i en cada uno de ellos con un muestreo aleatorio simple.

En este ejercicio se estima el total de votos para cada candidato a nivel estatal (o nacional) y con estos valores es inmediato estimar las proporciones deseadas.

El total poblacional está dado por

$$T_P = \sum_{i=1}^k \sum_{j=1}^{M_i} \mathbf{X}_{ij}$$

en donde $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijp})$ son los votos emitidos en la casilla j del estrato i p mediciones

M_i , es el número total de casillas en el i -ésimo estrato y k es el número de estratos.

No es difícil demostrar que este total puede expresarse como

$$T_P = \sum_{i=1}^k (m_i \bar{x}_i + (M_i - m_i) \bar{\mathbf{X}}_i^*)$$

con \bar{x}_i la media muestral observada en el estrato i , m_i el tamaño de muestra observado y $\bar{\mathbf{X}}_i^*$ la media **no** observada en el mismo estrato. Así que para estimar T_P basta estimar $\bar{\mathbf{X}}_i^*$ en todos los estratos.

Suponiendo normalidad en las observaciones y una distribución inicial poco informativa, es posible demostrar que la distribución **predictiva** de $\bar{\mathbf{X}}_i^*$ es una distribución Student de dimensión p con $m_i + 0.5$ grados de libertad, parámetro de localización \bar{x}_i y parámetro de escala $\frac{(m_i - \nu + 1)(M_i - m_i)}{M_i} \mathbf{s}_i^{-1}$, donde \mathbf{s}_i es la matriz de varianzas y covarianzas muestral y ν es el parámetro (conocido) de la inicial.

Se genera una muestra de esta distribución y eso permite obtener una muestra de la distribución predictiva de T_P condicional a la información observada. Finalmente, se obtienen muestras de la distribución predictiva de las proporciones

deseadas y, usando las listas nominales, de la participación ciudadana. Los intervalos buscados se construyen con los cuantiles correspondientes de la muestra generada.

7.4.2.3. Guanajuato

Dra. Marta Elva Ramírez Guzman y Lic. Gabriela López Reyes

Procedimiento de estimación mediante simulación bootstrap con Montecarlo

La simetría o asimetría de la distribución de la proporción (R), depende de la proporción real de la población y del tamaño de la muestra. Una proporción cercana a 0.5 y un tamaño de muestra grande tienden a producir una distribución muestral simétrica, mientras que desviaciones significativas de 0.5 o muestras pequeñas pueden resultar en una distribución muestral asimétrica. Ante la incertidumbre del tipo de distribución de R , una alternativa para estimarla es la simulación bootstrap. El bootstrap de simulación Montecarlo es especialmente útil cuando no se conocen las distribuciones teóricas exactas de los estadísticos de interés o cuando los cálculos analíticos son complicados. Al simular múltiples muestras de una muestra, se obtiene una aproximación de la distribución de la estadística de interés, lo que permite realizar inferencias estadísticas de manera robusta.

El procedimiento de bootstrap es una técnica estadística que se utiliza para estimar la distribución de una estadística de interés a partir de una muestra de datos. La simulación de Montecarlo, por otro lado, es una técnica numérica que se basa en la generación de números aleatorios para estimar resultados en situaciones donde el cálculo analítico puede ser difícil de obtener. La combinación de ambos, conocida como bootstrap de simulación Montecarlo, es una forma poderosa de obtener estimaciones de intervalos de confianza o realizar inferencias sobre parámetros desconocidos.

El procedimiento de bootstrap de simulación Montecarlo generalmente sigue estos pasos:

1. Obtención de la muestra inicial: Se comienza con una muestra de datos observados, la cual puede ser de cualquier tamaño.
2. Remuestreo (muestreo con reemplazo): Se generan múltiples (B) muestras de datos (muestras bootstrap) del mismo tamaño que la muestra original, seleccionando observaciones con reemplazo de la muestra inicial. Esto significa que algunas observaciones pueden aparecer múltiples veces en una muestra bootstrap mientras que otras pueden no aparecer en absoluto.

3. Cálculo de la estadística de interés: Para cada una de las muestras generadas en el paso anterior, se calcula la estadística de interés. En nuestro caso la estadística de interés es \hat{R} .
4. Construcción de la distribución de la estadística (\hat{R}): Se construye la distribución de la estadística de interés a partir de las estadísticas calculadas en las muestras bootstrap.
5. Inferencias y análisis: Esta distribución aproximada se utiliza para realizar inferencias sobre el parámetro de interés, como intervalos de confianza, errores de precisión o pruebas de hipótesis, de manera similar a como se haría con una muestra grande de datos.

Por lo anterior, el procedimiento de bootstrap de simulación Montecarlo, se utilizará para estimar el porcentaje de votantes en favor del p -ésimo candidato (\hat{R}_p), así como el error de estimación (d) e intervalos de confianza del 95%. En donde es deseable que el error de estimación sea tal que: $d \leq 1\%$.

Estimador de razón combinado

Respecto al estimador de R , éste se realizará mediante el estimador de razón combinado:

$$\hat{R}_p = \frac{\hat{Y}_p}{\hat{X}} = \frac{\sum_{h=1}^L \hat{Y}_{hp}}{\sum_{h=1}^L \hat{X}_h} = \frac{\sum_{h=1}^L N_h \bar{y}_{hp}}{\sum_{h=1}^L N_h \bar{x}_h}$$

Donde:

\hat{R}_p : estimador de la proporción de votos a favor del candidato p en la entidad

\hat{Y}_p : estimador del total de votos a favor del candidato p en la entidad

\hat{X} : estimador del total de votos emitidos en la entidad

\hat{Y}_{hp} : estimador del total de votos a favor del candidato p en el estrato h

\hat{X}_h : estimador del total de votos emitidos en el estrato h

\bar{y}_{hp} : estimador del promedio de votos a favor del candidato p , en el estrato h

\bar{x}_h : estimador del promedio de votos en el estrato h

N_h : total de casillas en el estrato h

7.4.2.4. Jalisco

Mtra. Patricia Isabel Romero Mares y Lic.. Enrique Reyes Romero.

El diseño muestral será estratificado con L estratos, que son los Distritos Federales, y una selección de casillas con muestreo aleatorio simple en cada uno de ellos y distribución proporcional de la muestra a los estratos.

Suponga un estado fijo o el país completo. La estimación de la proporción de ciudadanos/as que voten por el candidato p en el estado (país) se calculará utilizando el estimador de razón combinado para un esquema de muestreo estratificado con selección de una muestra aleatoria simple y sin reemplazo en cada estrato, y con asignación proporcional de la muestra a los L estratos:

$$\hat{R}_p = \frac{\hat{Y}_p}{\hat{X}} = \frac{\sum_{h=1}^L \hat{Y}_{hp}}{\sum_{h=1}^L \hat{X}_h} = \frac{\sum_{h=1}^L N_h * \bar{y}_{hp}}{\sum_{h=1}^L N_h * \bar{x}_h}$$

Donde:

\hat{R}_p : estimador de la proporción de votos a favor del candidato p , $p=1, \dots, \text{número de candidatos}$

\hat{Y}_p : estimador del total de votos a favor del candidato p , $p=1, \dots, \text{número de candidatos}$

\hat{X} : estimador del total de votos emitidos en el estado (o nacional)

\hat{Y}_{hp} : estimador del total de votos a favor del candidato p en el estrato h

\hat{X}_h : estimador del total de votos emitidos en el estrato h

\bar{y}_{hp} : estimador del promedio de votos por casilla a favor del candidato p , en el estrato h

\bar{x}_h : estimador del promedio de votos emitidos por casilla en el estrato h

N_h : total de casillas en el estrato h

El estimador de la varianza del estimador de la proporción de votos a favor del candidato p es el siguiente:

$$\hat{V}(\hat{R}_p) = \frac{1}{\hat{X}^2} * \sum_{h=1}^L N_h^2 * \left(1 - \frac{n_h}{N_h}\right) * \frac{s_h^2}{n_h}$$

Donde:

$$s_h^2 = \frac{\sum_{i=1}^{n_h} [(y_{hi}^p - \bar{y}_h^p) - \hat{R}_p(x_{hi} - \bar{x}_h)]^2}{n_h - 1}$$

y_{hi}^p : total de votos a favor del candidato p en la casilla i del estrato h

\bar{y}_h^p : promedio de votos a favor del candidato p del estrato h

x_{hi} : total de votos emitidos en la casilla i del estrato h

\bar{x}_h : promedio de votos emitidos por casilla en el estrato h

n_h : número de casillas en muestra del estrato h

La precisión observada para la estimación de la proporción de votos al candidato p se calcula como:

$$precisión = z_{(1-\frac{\alpha}{2})} \sqrt{\hat{V}(\hat{R}_p)}$$

Considerando una confianza del 95% y el cuantil de la Normal estándar $z=1.96$ correspondiente a esta confianza, el intervalo de confianza para la proporción de votos a favor del candidato p se obtiene de la siguiente manera:

$$(\hat{R}_p - precisión, \hat{R}_p + precisión)$$

De igual manera, la participación P , se estimará como un estimador de razón combinado, con la siguiente expresión:

$$\hat{P} = \frac{\hat{X}}{LN}$$

Con estimador de su varianza e intervalo de confianza, equivalentes a lo presentado arriba.

Referencias:

- Kish, L. (1970). "Survey Sampling". John Wiley and Sons. NY, USA.
- Raj, Des. (1968). "Sampling Theory". McGRAW-Hill. NY, USA.
- Särndal, Carl Erik, et al. (1992). "Model Assisted Survey Sampling". Springer-Verlag. NY, USA.

7.4.2.5. Morelos

Dr. Manuel Mendoza Ramírez y Dr. Felipe Javier Medina Aguayo

Introducción.

La estadística Bayesiana es una alternativa a la estadística frecuentista para la producción de inferencias sobre los parámetros desconocidos de un modelo. Aborda los problemas de inferencia como problemas de decisión en ambiente de incertidumbre, y recurre a una teoría de la decisión con fundamento axiomático para proveer las soluciones. (Bernardo y Smith, 2000).

Tanto bajo el enfoque frecuentista como bajo el Bayesiano, el fenómeno de interés es sujeto de observación a través de mecanismos que garantizan que los datos que se recolectan forman muestras aleatorias. En consecuencia, el comportamiento, la variabilidad, de estas muestras se describe a través de modelos de probabilidad.

En este contexto, las características del fenómeno que interesa conocer se identifican con el valor de uno o más parámetros del modelo y es en esos términos que los procesos de inferencia se concentran en aumentar y precisar el conocimiento sobre los parámetros.

Desde la perspectiva Bayesiana, la incertidumbre sobre los parámetros también se describe a través de modelos de probabilidad. En este caso no se describe variabilidad alguna sino el estado de conocimiento que, en un momento determinado, tiene el científico sobre el fenómeno bajo estudio. Con diferentes modelos se puede reflejar desde un conocimiento relativamente preciso hasta un desconocimiento casi completo sobre el valor del parámetro. De esta forma, bajo este paradigma, tanto las variables aleatorias observables (la muestra) como los parámetros fijos (pero desconocidos) son descritos mediante modelos de probabilidad.

La cuantificación de la incertidumbre sobre un parámetro puede producirse antes de contar con una muestra. El conocimiento teórico, histórico o incluso empírico, previo a la recolección de la muestra da lugar a esta valoración que se refleja en un modelo o distribución inicial o a priori $f(\theta)$.

Si es posible obtener la información muestral $\underline{x} = (X_1, X_2, \dots, X_n)$, proveniente del modelo $f(x|\theta)$, el estado de conocimiento inicial es susceptible de actualizarse para combinar la información inicial con la que provee la muestra.

Esta actualización se lleva a cabo mediante el mecanismo que establece el teorema de Bayes. El resultado es el modelo final o posterior $f(\theta|\underline{x})$, que combina las dos fuentes de información a través de la relación

$$f(\theta|\underline{x}) = f(\underline{x}|\theta)f(\theta)/f(\underline{x}).$$

A partir del modelo final se producen las inferencias de interés sobre el parámetro θ , de manera que es indispensable establecer las características de $f(\theta|\underline{x})$. Si bien la fórmula de Bayes es una expresión matemática simple, en la práctica, la caracterización analítica de la distribución final se puede complicar debido, por

ejemplo, a las dificultades en el cálculo de la constante de normalización $f(\underline{x})$. Gracias a los avances computacionales recientes y en especial a los algoritmos de simulación Monte Carlo vía cadenas de Markov (MCMC), es posible conocer las características de cualquier distribución final mediante métodos de simulación sin necesidad de calcular la constante de normalización (Chen et al. 2000).

Definiciones específicas.

El objetivo del conteo rápido es producir estimaciones confiables de la proporción efectiva de votos a favor de cada uno de los candidatos. Esta proporción se calcula como el cociente entre el número de votos a favor de un candidato y el número total de votos emitidos por todos los candidatos incluyendo los votos nulos y los candidatos no registrados.

Respecto a la información muestral con la que se producen los resultados del conteo rápido, se adopta el supuesto de que corresponde con un diseño estratificado de casillas. En esos términos, sea N el número de estratos, cada uno con una lista nominal de tamaño n_i para $i = 1, \dots, N$. Además, se define θ_{ij} como la proporción de personas en la lista nominal del estrato i cuya preferencia es por el candidato j , donde $j = 1, \dots, J$ con J el número total de candidatos incluyendo los candidatos no registrados, los nulos y el abstencionismo. En este caso, la proporción de personas en la lista nominal de todo el estado cuya preferencia es por el candidato j está dada por

$$\theta_j = \sum_{i=1}^N \frac{n_i}{n} \theta_{ij}$$

donde n es el tamaño de la lista nominal del estado, si se trata de una elección local o la lista nominal del país si se trata de una elección federal. Finalmente, la proporción efectiva de votos a favor del candidato j se puede calcular como

$$\lambda_j = \frac{\theta_j}{\sum_{l=1}^{J-1} \theta_l}$$

El Modelo y el Mecanismo de Inferencia

Tanto el modelo que se utiliza como el mecanismo para producir las inferencias se presentan en detalle en Mendoza y Nieto-Barajas (2016). En términos generales, se pueden describir como sigue. Del total de casillas en el estado (o país, según sea el caso) se seleccionan, mediante un mecanismo aleatorio, c de ellas (el tamaño total de la muestra), de manera que c_i de ellas provienen del i -ésimo estrato $i = 1, \dots, N$. De esta forma, c_i es el tamaño de la muestra en el estrato i .

Sea ahora, X_{ij}^k el número de personas a favor del candidato j , en la casilla k del estrato i . En esta formulación, $k = 1, \dots, K_i$ con K_i el número total de casillas en el estrato i .

El modelo que se emplea asume que

$$X_{ij}^k | \theta_{ij}, \tau_{ij} \sim N \left(n_i^k \theta_{ij}, \frac{\tau_{ij}}{n_i^k} \right)$$

donde n_i^k es el tamaño de la lista nominal de la casilla k en el estrato i , y τ_{ij} es un parámetro de precisión para $k = 1, \dots, c_i$, $i = 1, \dots, N$ y $j = 1, \dots, J$.

El modelo se complementa con la elección de las distribuciones iniciales para los parámetros θ_{ij} y τ_{ij} . En este caso, se utilizan iniciales que reflejan poca información a priori (distribuciones iniciales no informativas). El objetivo es obtener la distribución final para cada uno de los parámetros λ_j , con $j = 1, \dots, J - 1$ de manera que las inferencias que interesan se producen a partir de ese modelo de probabilidad. Como en otros casos, en este modelo esa distribución final es analíticamente compleja así que se recurre a métodos de simulación para obtener los resultados.

Referencias

- Bernardo, J.M. y Smith, A.F.M. (2000). Bayesian Theory. Wiley: Chichester.
- Chen, M.-H., Shao, Q.-M. e Ibrahim, J.G. (2000). Monte Carlo methods in Bayesian computation. Springer, New York.
- Mendoza, M. y Nieto-Barajas, L.E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. Electoral Studies 43, 124-132.

7.4.2.6. Puebla

Dr. Gabriel Núñez Antonio y Mtro. Emiliano Geneyro Squarzon.

Procedimiento de estimación

El reconocimiento explícito de la incertidumbre es crucial en las ciencias estadísticas. En muchas situaciones simples, la incertidumbre de un estimador puede visualizarse a través de cálculos analíticos basados en algún modelo probabilístico, adecuado. Sin embargo, en problemas más complejos esta tarea puede ser complicada, y los resultados llegan a ser potencialmente incorrectos si los supuestos básicos no se cumplen o si se realizan aproximaciones con la

finalidad de simplificar los cálculos. Los avances computacionales, permiten hoy en día, el empleo de métodos denominados de remuestreo que ayudan a soslayar los problemas anteriores, a la vez que permiten obtener estimaciones de errores estándar, intervalos de confianza, y otras medidas de incertidumbre.

La idea principal de los métodos bootstrap es el remuestreo a partir de los datos originales, ya sea en forma directa o vía un modelo ajustado; con la finalidad de obtener muestras replicadas a partir de las cuales se pueda evaluar la variabilidad de las cantidades de interés. Los métodos bootstrap también se pueden aplicar en problemas simples para verificar las características de las medidas de incertidumbre, para relajar supuestos, o para dar rápidas soluciones aproximadas. Un ejemplo de lo anterior es el remuestreo aleatorio para estimar la distribución permutacional de alguna estadística de prueba no paramétrica.

Es verdad que en muchas aplicaciones se puede confiar ampliamente en un modelo paramétrico particular y en el correspondiente análisis clásico basado en dicho modelo. Aun así, puede ser de utilidad investigar qué tanto se puede inferir sin asumir los supuestos de un modelo paramétrico particular. Esto es la esencia de la *robustez* del análisis estadístico realizado. El *bootstrap no-paramétrico* permite hacer esto. De acuerdo con lo anterior los métodos bootstrap se pueden aplicar tanto en los casos en los que se cuente con un modelo probabilístico bien definido para los datos, como en los casos en los que no se disponga de dicho modelo.

A pesar de su alcance y utilidad, el remuestreo debe ser aplicado en forma cuidadosa. A menos que las ideas básicas queden completamente claras, es muy fácil producir una solución al problema equivocado, o una mala solución al problema correcto. Los métodos de remuestreo como el bootstrap pretenden, de hecho lo hacen, evitar el empleo de procedimientos basados en supuestos distribucionales cuestionables.

La propuesta de estimación en el Conteo Rápido para la Gubernatura del estado de Puebla, en las elecciones del 2 de junio de 2024, consideran el diseño de muestral definido en este documento; con selección de casillas aleatoria sin reemplazo dentro de cada estrato y con tamaño de muestra proporcional al número de casillas en cada estrato.

Tipo de estimadores puntuales que se emplearán

Sea h el estrato a considerar, $h=1, \dots, k$, con N_h elementos en cada estrato. En cada estrato se toma una muestra aleatoria simple de tamaño n_h . Para estimar el porcentaje de votos de cada candidato p_j , $j=1, \dots, b$, los estimadores puntuales propuestos para la proporción de votos están dados por:

$$\hat{p}_j = \frac{\hat{T}_j}{\hat{T}} \quad \forall j = 1, \dots, b,$$

donde \hat{T}_j es un estimador estratificado del total de votos para cada candidato j , $j=1, \dots, b$, y \hat{T} es un estimador estratificado del total de votos obtenidos. Específicamente,

$$\hat{T}_j = \sum_h \frac{N_h}{n_h} Y_{j,h}, \quad \hat{T} = \sum_h \frac{N_h}{n_h} Y_h,$$

donde $Y_{j,h}$ es el total de votos del candidato j en el estrato h y Y_h es el total de votos en el estrato h .

Procedimiento para construir intervalos de confianza

Para obtener intervalos de confianza, con una precisión adecuada, para la proporción de votos para cada candidato se pueden seguir diferentes estrategias. Por ejemplo se puede usar una aproximación normal asintótica. Si se considera que la aproximación normal asintótica puede no ser adecuada, otra opción es construir vía simulación la distribución muestral del estimador que se esté utilizando y obtener los cuantiles de orden 2.5% y 97.5% o a partir de muestras bootstrap obtener un estimador de la varianza del estimador correspondiente y con ellos construir el intervalo correspondiente. Estas dos últimas estrategias son las que se explorarían para construir los correspondientes intervalos de confianza, considerando un proceso de remuestreo con al menos 5,000 simulaciones. Lo anterior derivado de los diferentes estudios y análisis realizados con las bases de datos de la elección de gobernador de Puebla en 2018 y 2019.

7.4.2.7.Tabasco

Michelle Anzarut, Felipe González, Ana María Madrigal, Teresa Ortiz

Resumen

El conteo rápido consiste en generar estimaciones de los resultados de una contienda electoral, utilizando los votos registrados en una muestra. Su principal desafío radica en que la estimación se efectúa con muestras incompletas, donde la ausencia de datos no es aleatoria. El método de estimación que emplearemos es una regresión binomial negativa con estructura jerárquica, permitiendo que algunos coeficientes varíen según el estrato. Adicionalmente, implementaremos un procedimiento de ajuste por sesgo al estimar con muestras incompletas. El modelo proporciona intervalos de probabilidad con una cobertura aproximada del 95%, incluso frente a ciertos patrones de muestras sesgadas observados en elecciones previas.

Antecedentes

La idea del modelo que aplicaremos provino del modelo de regresión multinivel y postestratificación (ver, por ejemplo, [1]). En este modelo de regresión multinivel y postestratificación, (1) se segmenta a la población en celdas según combinaciones de distintos atributos, (2) se utilizan las muestras para obtener simulaciones posteriores de los parámetros, y (3) se calcula la variable de respuesta para cada celda, ponderando las estimaciones para realizar una estimación a nivel poblacional.

Llevaremos a cabo los pasos (1) y (2), pero en el paso (3), adoptaremos un enfoque basado en modelos (ver [2]), simulando la distribución posterior del número total de votos para cada candidato y para cada una de las casillas electorales del país. Así, emplearemos el modelo para predecir los valores no muestreados de la población. Es importante destacar que esto proporciona una estimación inmediata a nivel poblacional.

Para ejecutar el paso (1), ajustamos una regresión multinivel para cada casilla en función de ciertas covariables. Es decir, para cada candidato, modelamos el número de votos en la casilla i , y_i , con una distribución de probabilidad $p(y|\theta_1, \theta_2)$, donde, para una parte de los parámetros, digamos θ_1 , ajustamos una regresión multinivel.

La ventaja de esta estructura de modelos multinivel radica en que, ante la ausencia de respuesta, atrae los parámetros de un grupo hacia la media del mismo. Como consecuencia, logramos un manejo uniforme de los datos faltantes en las muestras y una mejora en las propiedades de cobertura de intervalos cuando los datos de la muestra presentan sesgos.

La distribución p es binomial negativa. Esta distribución se adapta adecuadamente a los datos y su robustez contribuye a la del modelo, permitiéndole ajustarse eficazmente tanto a candidatos con una alta como con una baja cantidad de votos. Además, ofrece una capa adicional de protección ante posibles errores en la base de datos.

Modelo

A continuación, definimos el modelo que utilizaremos, que se obtuvo después de varias iteraciones del flujo de trabajo bayesiano en [4].

Denotamos con BN a la distribución binomial negativa con una parametrización directamente usando un parámetro de media, $\mu_i \in \mathbb{R}^+$, y un parámetro $\phi_i \in \mathbb{R}^+$. Para $y \in \mathbb{N}$,

$$\text{BN}(y|\mu_i, \phi_i) \propto \left(\frac{\mu_i}{\mu_i + \phi_i} \right)^y.$$

Nuestro primer objetivo es predecir el número total de votos a favor de cada candidato y el número total de votos nulos en cada casilla. Teniendo esas predicciones podemos deducir cualquier cantidad que sea de nuestro interés, por ejemplo, la votación total emitida, la participación, intervalos de confianza para la proporción de votos obtenida por cada candidato o la conformación de la Cámara de Diputados.

Por simplicidad de notación consideraremos a los votos nulos como un candidato extra. Sea p el número total de candidatos, modelamos el número de votos obtenidos para el candidato $j \in \{1, \dots, p\}$ y para la casilla i como

$$y_{i,j} \sim \text{NB} \left(\mu_{i,j} := n_i w_i \theta_{i,j}, \phi_{i,j} := \frac{n_i w_i \theta_{i,j}}{\kappa_{s(i),j}} \right).$$

Aquí $\theta_{i,j}$ representa la proporción de personas que votaron por j del total de votos en la casilla i , w_i representa la proporción de personas en la lista nominal que votaron en la casilla i y $\kappa_{s(i),j} > 0$ controla la sobredispersión en relación a la media $\mu_{i,j}$.

El vector $\theta_i = (\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,p})$ se define como

$$\theta_i = \text{softmax} \left(\beta_{s(i)}^{0,1} + \beta_1^T x_i, \dots, \beta_{s(i)}^{0,p} + \beta_p^T x_i \right),$$

donde x_i es la i -ésima fila de X , una matriz de covariables escaladas, y

$$\text{softmax}(z_1, \dots, z_p)_j = \frac{\exp(z_j)}{\sum_{i=1}^p \exp(z_i)}$$

Los coeficientes β_1, \dots, β_p tienen una distribución inicial gaussiana centrada en cero y con una desviación estándar de 0.25. Sin embargo, para $k = 1, \dots, p$,

$$\beta_{s(i)}^{0,k} \sim \mathcal{N}(\beta_{0,k}, \sigma_k),$$

añadiendo una estructura jerárquica dependiendo del estrato. Los detalles de las distribuciones iniciales pueden consultarse en [8] en donde, adicionalmente, se prueba la consistencia de la elección de todas las distribuciones iniciales.

Estimación

La metodología general de la estimación se describe a continuación. Inicialmente, para cada casilla que no forma parte de la muestra, generamos simulaciones de votos basadas en el modelo. Posteriormente, juntamos los valores observados en las casillas con los valores simulados, obteniendo así los votos totales de todas las casillas. A partir de estos votos, es directo calcular los resúmenes que sean necesarios para cada elección.

La estimación se hace con el software Stan, llamado desde R utilizando el paquete CmdStanR [6]. El código en Stan es legible, lo que hace que el trabajo sea más transparente. Stan también permite una serie de diagnósticos, lo que ayudó a obtener una inferencia confiable más rápida. La implementación se puede consultar y reproducir en su totalidad en el paquete de R quickcountmx [7].

Para asegurarnos de abarcar una diversidad de posibles resultados, realizamos una calibración bayesiana con los datos de las elecciones nacionales de 2018. La calibración bayesiana se enfoca en verificar que el modelo produzca inferencias adecuadamente calibradas desde un enfoque frecuentista; es decir, busca asegurar que el modelo entregue intervalos de credibilidad posterior con una cobertura cercana al 95% en muestreos repetidos. La calibración evidencia que, ante una proporción significativa de casillas no reportadas, el modelo por sí solo no es suficiente para ajustar los sesgos presentes en las muestras censuradas.

El problema del sesgo

El sistema de información comienza a operar a las 6 p.m. y, cada 5 minutos, recibe nueva información de la muestra. De esta manera, el sistema genera una secuencia de muestras parciales acumulativas, que son analizadas mediante métodos de estimación para monitorear la tendencia de los resultados. Los

resultados del conteo rápido se publican utilizando una de estas muestras parciales, debido a que un porcentaje significativo de la muestra completa tarda en ser reportado.

La ausencia de datos en las muestras parciales no es aleatoria; factores como si la casilla se encuentra en una zona urbana o rural, las condiciones climáticas o la lista nominal, influyen en el momento en que una casilla llega a la muestra.

Realizamos un análisis exhaustivo de sesgo en elecciones anteriores y nos enfrentamos al desafío de mitigar este sesgo en todas las etapas de nuestro flujo de trabajo. Para llevar a cabo este análisis, necesitábamos un procedimiento para generar muestras parciales que se asemeje al verdadero proceso de generación de datos. Así, ajustamos modelos de supervivencia para los tiempos de llegada de las casillas en las elecciones de 2018. Los detalles se pueden consultar en [8].

Usamos el modelo de supervivencia para obtener una medida del sesgo en diferentes horas o con diferentes porcentajes de muestra obtenidos.

El ajuste que haremos al modelo dependerá del porcentaje de muestra observado, al que denotamos con p_{obs} . En primer lugar, estimamos la distribución de probabilidad del número total de votos de cada candidato en cualquier casilla con el modelo descrito. Después, a la hora de simular, en lugar de utilizar el vector θ_i , usamos una perturbación, dada por

$$\theta'_i = \text{softmax}(\theta_{i,1} + \eta_{i,1}, \dots, \theta_{i,p} + \eta_{i,p}),$$

donde η_1, \dots, η_p son variables aleatorias independientes e idénticamente distribuidas $N(0, 0.1(1 - p_{obs}))$.

Notemos que el tamaño de los intervalos de probabilidad producidos crece a medida que crece $\eta_{i,k}$ y, conforme nos acercamos a la muestra completa, $\eta_{i,k}$ se acerca a 0 para cualquier candidato k . El modelo, con el ajuste por sesgo, provee coberturas cercanas al 95% tanto con muestras completas como con muestras censuradas.

Método de respaldo

Como método de respaldo, implementaremos el estimador de razón combinado.

Sea p_j la proporción de votos a favor del candidato j , para estimar p_j utilizaremos \hat{p}_j , definido como:

$$\hat{p}_j = \frac{\sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hij}}{\sum_{j=1}^p \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hij}},$$

donde, para cada estrato h , n_h es el tamaño de la muestra, N_h es el número total de casillas, L es el número total de estratos, p es el número total de candidatos (incluyendo nulos y no registrados) y x_{hij} es el total de votos a favor del candidato j en la casilla i .

El error estándar de este estimador se calcula utilizando bootstrap no paramétrico (ver, por ejemplo, [9]).

Referencias

- [1] Park, D.K., Gelman, A., Bafumi, J.: Bayesian multilevel estimation with poststratification: state-level estimates from national polls. *Political Analysis* 12(4), 375–385 (2004)
- [2] Little, R.J.: Calibrated Bayes, an alternative inferential paradigm for official statistics. *Journal of official statistics* 28(3), 309 (2012)
- [3] Anzarut, M., González, L.F., Ortiz, M.T.: A heavy-tailed multilevel mixture model for the quick count in the mexican elections of 2018. In: National Statistics Forum (FNE) and Latin-American Congress of Statistical Societies (CLATSE), pp. 1–13. Springer (2018)
- [4] Schad, D., Betancourt, M., Vasishth, S.: Toward a principled bayesian workflow in cognitive science. *Psychological Methods* 26 (2020). DOI 10.1037/met0000275
- [5] Barnard, John, Robert McCulloch, and Xiao-Li Meng. 2000.: Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage. *Statistica Sinica*, 1281–1311.
- [6] Stan Development Team: CmdStanR: the R interface to CmdStan (2021). URL <https://mc-stan.org/r-packages/>
- [7] Anzarut, M., González, L.F., Ortiz, M.T.: quickcountmx: Fuctions for the mexican quick-counts 2022 (2022). URL <https://github.com/cotecora-team-2/quickcountmx>
- [8] Anzarut, M., González, L.F., Ortiz, M.T.: Bayesian Hierarchical Multinomial Modeling of the 2021 Mexican Election Outcomes with Censored Samples. In: National Statistics Forum (FNE). Springer (2022)
- [9] Ortiz, M.T. Bootstrap no paramétrico, *Estadística Computacional* Capítulo 5, https://tereom.github.io/est_computacional/05-Bootsrap.html

7.4.2.8. Veracruz

Michelle Anzarut, Felipe González, Ana María Madrigal, Teresa Ortiz

Resumen

El conteo rápido consiste en generar estimaciones de los resultados de una contienda electoral, utilizando los votos registrados en una muestra. Su principal desafío radica en que la estimación se efectúa con muestras incompletas, donde la ausencia de datos no es aleatoria. El método de estimación que emplearemos es una regresión binomial negativa con estructura jerárquica, permitiendo que algunos coeficientes varíen según el estrato. Adicionalmente, implementaremos un procedimiento de ajuste por sesgo al estimar con muestras incompletas. El modelo proporciona intervalos de probabilidad con una cobertura aproximada del 95%, incluso frente a ciertos patrones de muestras sesgadas observados en elecciones previas.

Antecedentes

La idea del modelo que aplicaremos provino del modelo de regresión multinivel y postestratificación (ver, por ejemplo, [1]). En este modelo de regresión multinivel y postestratificación, (1) se segmenta a la población en celdas según combinaciones de distintos atributos, (2) se utilizan las muestras para obtener simulaciones posteriores de los parámetros, y (3) se calcula la variable de respuesta para cada celda, ponderando las estimaciones para realizar una estimación a nivel poblacional.

Llevaremos a cabo los pasos (1) y (2), pero en el paso (3), adoptaremos un enfoque basado en modelos (ver [2]), simulando la distribución posterior del número total de votos para cada candidato y para cada una de las casillas electorales del país. Así, emplearemos el modelo para predecir los valores no muestreados de la población. Es importante destacar que esto proporciona una estimación inmediata a nivel poblacional.

Para ejecutar el paso (1), ajustamos una regresión multinivel para cada casilla en función de ciertas covariables. Es decir, para cada candidato, modelamos el número de votos en la casilla i , y_i , con una distribución de probabilidad $p(y|\theta_1, \theta_2)$, donde, para una parte de los parámetros, digamos θ_1 , ajustamos una regresión multinivel.

La ventaja de esta estructura de modelos multinivel radica en que, ante la ausencia de respuesta, atrae los parámetros de un grupo hacia la media del mismo. Como consecuencia, logramos un manejo uniforme de los datos faltantes en las muestras y una mejora en las propiedades de cobertura de intervalos cuando los datos de la muestra presentan sesgos.

La distribución p es binomial negativa. Esta distribución se adapta adecuadamente a los datos y su robustez contribuye a la del modelo, permitiéndole ajustarse eficazmente tanto a candidatos con una alta como con una baja cantidad de votos. Además, ofrece una capa adicional de protección ante posibles errores en la base de datos.

Modelo

A continuación, definimos el modelo que utilizaremos, que se obtuvo después de varias iteraciones del flujo de trabajo bayesiano en [4].

Denotamos con BN a la distribución binomial negativa con una parametrización directamente usando un parámetro de media, $\mu_i \in \mathbb{R}^+$, y un parámetro $\phi_i \in \mathbb{R}^+$. Para $y \in \mathbb{N}$,

$$\text{BN}(y|\mu_i, \phi_i) \propto \left(\frac{\mu_i}{\mu_i + \phi_i} \right)^y.$$

Nuestro primer objetivo es predecir el número total de votos a favor de cada candidato y el número total de votos nulos en cada casilla. Teniendo esas predicciones podemos deducir cualquier cantidad que sea de nuestro interés, por ejemplo, la votación total emitida, la participación, intervalos de confianza para la proporción de votos obtenida por cada candidato o la conformación de la Cámara de Diputados.

Por simplicidad de notación consideraremos a los votos nulos como un candidato extra. Sea p el número total de candidatos, modelamos el número de votos obtenidos para el candidato $j \in \{1, \dots, p\}$ y para la casilla i como

$$y_{i,j} \sim \text{NB} \left(\mu_{i,j} := n_i w_i \theta_{i,j}, \phi_{i,j} := \frac{n_i w_i \theta_{i,j}}{\kappa_{s(i),j}} \right).$$

Aquí $\theta_{i,j}$ representa la proporción de personas que votaron por j del total de votos en la casilla i , w_i representa la proporción de personas en la lista nominal que votaron en la casilla i y $\kappa_{s(i),j} > 0$ controla la sobredispersión en relación a la media $\mu_{i,j}$.

El vector $\theta_i = (\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,p})$ se define como

$$\theta_i = \text{softmax} \left(\beta_{s(i)}^{0,1} + \beta_1^T x_i, \dots, \beta_{s(i)}^{0,p} + \beta_p^T x_i \right),$$

donde x_i es la i -ésima fila de X , una matriz de covariables escaladas, y

$$\text{softmax}(z_1, \dots, z_p)_j = \frac{\exp(z_j)}{\sum_{i=1}^p \exp(z_i)}$$

Los coeficientes β_1, \dots, β_p tienen una distribución inicial gaussiana centrada en cero y con una desviación estándar de 0.25. Sin embargo, para $k = 1, \dots, p$,

$$\beta_{s(i)}^{0,k} \sim \mathcal{N}(\beta_{0,k}, \sigma_k),$$

añadiendo una estructura jerárquica dependiendo del estrato. Los detalles de las distribuciones iniciales pueden consultarse en [8] en donde, adicionalmente, se prueba la consistencia de la elección de todas las distribuciones iniciales.

Estimación

La metodología general de la estimación se describe a continuación. Inicialmente, para cada casilla que no forma parte de la muestra, generamos simulaciones de votos basadas en el modelo. Posteriormente, juntamos los valores observados en las casillas con los valores simulados, obteniendo así los votos totales de todas las casillas. A partir de estos votos, es directo calcular los resúmenes que sean necesarios para cada elección.

La estimación se hace con el software Stan, llamado desde R utilizando el paquete CmdStanR [6]. El código en Stan es legible, lo que hace que el trabajo sea más transparente. Stan también permite una serie de diagnósticos, lo que ayudó a obtener una inferencia confiable más rápida. La implementación se puede consultar y reproducir en su totalidad en el paquete de R quickcountmx [7].

Para asegurarnos de abarcar una diversidad de posibles resultados, realizamos una calibración bayesiana con los datos de las elecciones nacionales de 2018. La calibración bayesiana se enfoca en verificar que el modelo produzca inferencias adecuadamente calibradas desde un enfoque frecuentista; es decir, busca asegurar que el modelo entregue intervalos de credibilidad posterior con una cobertura cercana al 95% en muestreos repetidos. La calibración evidencia que, ante una proporción significativa de casillas no reportadas, el modelo por sí solo no es suficiente para ajustar los sesgos presentes en las muestras censuradas.

El problema del sesgo

El sistema de información comienza a operar a las 6 p.m. y, cada 5 minutos, recibe nueva información de la muestra. De esta manera, el sistema genera una secuencia de muestras parciales acumulativas, que son analizadas mediante métodos de estimación para monitorear la tendencia de los resultados. Los

resultados del conteo rápido se publican utilizando una de estas muestras parciales, debido a que un porcentaje significativo de la muestra completa tarda en ser reportado.

La ausencia de datos en las muestras parciales no es aleatoria; factores como si la casilla se encuentra en una zona urbana o rural, las condiciones climáticas o la lista nominal, influyen en el momento en que una casilla llega a la muestra.

Realizamos un análisis exhaustivo de sesgo en elecciones anteriores y nos enfrentamos al desafío de mitigar este sesgo en todas las etapas de nuestro flujo de trabajo. Para llevar a cabo este análisis, necesitábamos un procedimiento para generar muestras parciales que se asemeje al verdadero proceso de generación de datos. Así, ajustamos modelos de supervivencia para los tiempos de llegada de las casillas en las elecciones de 2018. Los detalles se pueden consultar en [8].

Usamos el modelo de supervivencia para obtener una medida del sesgo en diferentes horas o con diferentes porcentajes de muestra obtenidos.

El ajuste que haremos al modelo dependerá del porcentaje de muestra observado, al que denotamos con p_{obs} . En primer lugar, estimamos la distribución de probabilidad del número total de votos de cada candidato en cualquier casilla con el modelo descrito. Después, a la hora de simular, en lugar de utilizar el vector θ_i , usamos una perturbación, dada por

$$\theta'_i = \text{softmax}(\theta_{i,1} + \eta_{i,1}, \dots, \theta_{i,p} + \eta_{i,p}),$$

donde η_1, \dots, η_p son variables aleatorias independientes e idénticamente distribuidas $N(0, 0.1(1 - p_{obs}))$.

Notemos que el tamaño de los intervalos de probabilidad producidos crece a medida que crece $\eta_{i,k}$ y, conforme nos acercamos a la muestra completa, $\eta_{i,k}$ se acerca a 0 para cualquier candidato k . El modelo, con el ajuste por sesgo, provee coberturas cercanas al 95% tanto con muestras completas como con muestras censuradas.

Método de respaldo

Como método de respaldo, implementaremos el estimador de razón combinado.

Sea p_j la proporción de votos a favor del candidato j , para estimar p_j utilizaremos \hat{p}_j , definido como:

$$\hat{p}_j = \frac{\sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hij}}{\sum_{j=1}^p \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hij}},$$

donde, para cada estrato h , n_h es el tamaño de la muestra, N_h es el número total de casillas, L es el número total de estratos, p es el número total de candidatos (incluyendo nulos y no registrados) y x_{hij} es el total de votos a favor del candidato j en la casilla i .

El error estándar de este estimador se calcula utilizando bootstrap no paramétrico (ver, por ejemplo, [9]).

Referencias

- [1] Park, D.K., Gelman, A., Bafumi, J.: Bayesian multilevel estimation with poststratification: state-level estimates from national polls. *Political Analysis* 12(4), 375–385 (2004)
- [2] Little, R.J.: Calibrated Bayes, an alternative inferential paradigm for official statistics. *Journal of official statistics* 28(3), 309 (2012)
- [3] Anzarut, M., González, L.F., Ortiz, M.T.: A heavy-tailed multilevel mixture model for the quick count in the mexican elections of 2018. In: National Statistics Forum (FNE) and Latin-American Congress of Statistical Societies (CLATSE), pp. 1–13. Springer (2018)
- [4] Schad, D., Betancourt, M., Vasishth, S.: Toward a principled bayesian workflow in cognitive science. *Psychological Methods* 26 (2020). DOI 10.1037/met0000275
- [5] Barnard, John, Robert McCulloch, and Xiao-Li Meng. 2000.: Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage. *Statistica Sinica*, 1281–1311.
- [6] Stan Development Team: CmdStanR: the R interface to CmdStan (2021). URL <https://mc-stan.org/r-packages/>
- [7] Anzarut, M., González, L.F., Ortiz, M.T.: quickcountmx: Fuctions for the mexican quick-counts 2022 (2022). URL <https://github.com/cotecora-team-2/quickcountmx>
- [8] Anzarut, M., González, L.F., Ortiz, M.T.: Bayesian Hierarchical Multinomial Modeling of the 2021 Mexican Election Outcomes with Censored Samples. In: National Statistics Forum (FNE). Springer (2022)
- [9] Ortiz, M.T. Bootstrap no paramétrico, *Estadística Computacional* Capítulo 5, https://tereom.github.io/est_computacional/05-Bootsrap.html

7.4.2.9. Yucatán

Dra. María Fernanda Gil Leyva Villa y Dr. Alan Riva Palacio Cohen.

Para cada tamaño de muestra ($m = 50, 100, \dots, 400$) y para cada estratificación (Simple, FED, FEDxSEC y LOC) se repite el siguiente experimento que consta de tres pasos:

1. Se genera una muestra de tamaño m mediante el MASE.
2. Usando la muestra generada y mediante Bootstrap Bayesiano se calcula:
 - a. Un estimador puntual, \hat{P}_j , de la proporción de votos al candidato j , P_j . Cabe mencionar que este estimador está centrado en el estimador de razón combinada:

$$\frac{\sum_h m_h \bar{x}_{h,j}}{\sum_j \sum_h m_h \bar{x}_{h,j}}$$

donde $\bar{x}_{h,j}$ es el promedio de votos al candidato j de las casillas en muestra del estrato h , y recordando que m_h es el número de casillas en muestra en el estrato h .

- b. Un intervalo $[a_j, b_j]$ que contengan a la proporción de votos, P_j , con una probabilidad de al menos 0.95.
3. Para la muestra y las estimaciones obtenidas en los pasos anteriores se analizan:
 - a. El error máximo de estimación: $\varepsilon = \max_j |\hat{P}_j - P_j|$.
 - b. La longitud los intervalos: $l_j = b_j - a_j$, para cada candidato.
 - c. La cobertura de los intervalos: Si el valor real de P_j pertenece a $[a_j, b_j]$, para cada candidato.
 - d. La carga a los Capacitadores Asistentes Electorales (CAE): La proporción de CAE a los que les corresponde reportar los resultados de 1, 2 o más casillas para el Conteo Rápido.