



Criterios científicos, logísticos y operativos para
la realización de los Conteos Rápidos y protocolo
para la selección de las muestras
COTECORA 2017-2018

CONTENIDO

1. Antecedentes.....	4
2. Criterios científicos.....	6
2.1. Elección Presidencial.....	6
2.1.1.....Estratificación y tamaño de muestra	7
2.1.2.....Procedimientos de estimación	8
2.1.3.....Integración de las estimaciones	9
2.2. Elecciones de Gobernatura y Jefatura de Gobierno.....	11
2.2.1.....Estratificación y tamaño de muestra	12
2.2.2.....Procedimientos de estimación	13
2.2.3.....Integración de las estimaciones	14
3. Selección y resguardo de las muestras.....	16
3.1. Consideraciones generales para la selección de las muestras.....	16
3.2. Protocolo de selección y resguardo de las muestras.....	16
3.2.1.....Instalación	16
3.2.2.....Selección de las muestras	17
3.2.3.....Resguardo de las muestras	17
4. Criterios logísticos y operativos.....	19
4.1. Objetivos de los criterios logísticos y operativos.....	19

4.2.	Esquema general de funcionamiento del Conteo Rápido.....	19
4.2.1.....	Etapas de planeación	22
4.2.2.....	Preparación	26
4.2.3.....	Ejecución del operativo de campo	28
4.2.4.....	Esquemas de contingencia	31
4.2.5.....	Esquema de seguimiento	33
4.3.	Medidas de Seguridad en la sede del COTECORA.....	35
5.	Anexos. Diseño Muestral y Métodos de Estimación.....	36
5.1.	Elección Presidencial.....	36
5.1.1.....	Estratificación y tamaño de muestra	36
5.1.2.....	Procedimiento de estimación	41
5.2.	Elecciones de Gobernatura y Jefatura de Gobierno.....	53
5.2.1.....	Estratificación y tamaño de muestra	53
5.2.2.....	Procedimiento de estimación	131

1. Antecedentes

El 8 de septiembre de 2017, en sesión extraordinaria del Consejo General, el Consejero Presidente emitió un pronunciamiento para dar formal inicio al Proceso Electoral Federal 2017-2018.

Asimismo, entre el 1° de septiembre y el 31 de diciembre de 2017, dieron inicio los Procesos Electorales Locales 2017-2018, en las diversas entidades federativas que celebran elecciones locales, en forma concurrente con la federal, el domingo 1° de julio de 2018.

El 30 de octubre de 2017, mediante Resolución **INE/CG503/2017**, el Consejo General determinó asumir el diseño, implementación y operación del Conteo Rápido para el Proceso Electoral Local Ordinario 2017-2018 a desarrollarse en Tabasco, junto con la elección federal ordinaria de la Presidencia de los Estados Unidos Mexicanos.

El 22 de noviembre de 2017, el Consejo General modificó, mediante Acuerdo **INE/CG565/2017**, diversas disposiciones del Reglamento de Elecciones (RE), en términos de lo dispuesto por el artículo 441 del mencionado reglamento.

Esa misma fecha, el Consejo General determinó, mediante Resolución **INE/CG568/2017**, asumir el diseño, implementación y operación del programa de Conteo Rápido para la elección de titulares del Poder Ejecutivo Estatal en las entidades federativas de Chiapas, Guanajuato, Jalisco, Morelos, Puebla, Veracruz y Yucatán, así como para la elección del Jefe de Gobierno de la Ciudad de México durante sus Procesos Electorales Locales ordinarios 2017-2018, junto con la elección federal ordinaria de la Presidencia de los Estados Unidos Mexicanos.

Asimismo, en dicha sesión el Consejo General aprobó, mediante Acuerdo **INE/CG569/2017**, la realización del Conteo Rápido basado en actas de escrutinio y cómputo de casilla para la elección federal ordinaria de la Presidencia de los Estados Unidos Mexicanos, a fin de conocer la estimación de las tendencias de los resultados de la votación el día de la Jornada Electoral del 1° de julio de 2018, así como la creación e integración del Comité Técnico Asesor de los Conteos Rápidos para los Procesos Electorales Federal y Locales 2017-2018 (**COTECORA**).

El 14 de febrero de 2018, la Sala Superior del Tribunal Electoral del Poder Judicial de la Federación emitió la sentencia recaída dentro del expediente **SUP-RAP-749/2017** y sus acumulados, en el sentido de modificar el acuerdo **INE/CG565/2017**, ordenando al Instituto Nacional Electoral (INE), entre otras cuestiones, derogar los párrafos 4, 5, 6, 7 y 8 del artículo 246 del RE, relativo al procedimiento previo de revisión de urnas aprobado por este Consejo General, la posibilidad de levantar actas al término de cada escrutinio y

cómputo, así como a diversas especificaciones a llevarse a cabo en los mismos.

El 19 de febrero de 2018, este Consejo General emitió el Acuerdo **INE/CG111/2018** por el que, en acatamiento a la sentencia dictada por la Sala Superior, se modificó el Acuerdo **INE/CG565/2017**.

El 28 de febrero de 2018, el Consejo General aprobó el Acuerdo **INE/CG122/2018**, por el que se determina que la realización del Conteo Rápido para la elección de titulares de los Ejecutivos Federal y Locales para conocer la estimación de las tendencias de los resultados de la votación el día de la Jornada Electoral del 1° de julio de 2018, se realice con base en los datos obtenidos de las hojas del cuadernillo para hacer las operaciones del escrutinio y cómputo en casilla.

El 22 de marzo de 2018, el Tribunal Electoral del Poder Judicial de la Federación resolvió en el **SUP-RAP-42/2018**, confirmando el acuerdo **INE/CG122/2018**, en que se establece la utilización de los cuadernillos de apoyo u hojas de operaciones para obtener los datos que requiere el Conteo Rápido ya que no representa afectación alguna al principio de certeza.

Derivado de lo anterior y en apego a las disposiciones normativas en la materia, el **COTECORA** desarrolló los criterios científicos y revisó el diseño de los procedimientos operativos y logísticos para el acopio de los datos contenidos en las hojas del cuadernillo para hacer las operaciones del escrutinio y cómputo en casilla, así como el protocolo para la selección de muestras.

2. Criterios científicos

Los criterios científicos del Conteo Rápido son todos los procedimientos que, con base en la probabilidad y en la estadística, se usarán para estimar el porcentaje de votos a favor de los candidatos a la Presidencia, Gubernaturas y Jefatura de Gobierno, así como para estimar el porcentaje de ciudadanos que acudan a votar¹.

Debido a que se realizarán diez Conteos Rápidos, uno federal y nueve locales, las Asesoras y Asesores Técnicos distribuyeron las actividades, de manera que cada uno de ellos sería responsable de uno de los estados, en el sentido de establecer la definición del diseño muestral. En el caso del Conteo Rápido Federal, todos los integrantes del comité acordarían el diseño muestral que emplearán para realizar las estimaciones de las tendencias.

2.1. Elección Presidencial

Derivado de los acuerdos adoptados en la tercera reunión de trabajo, el COTECORA conformó tres equipos de trabajo, quedando de la siguiente manera:

- Dra. Michelle Anzarut, Dr. Manuel Mendoza y Dr. Luis Enrique Nieto.
- Dr. Carlos Hernández, Dr. Gabriel Núñez y Dr. Carlos Erwin Rodríguez.
- Dr. Alberto Alonso, Mtra. Patricia Romero y Dr. Raúl Rueda.

Cada equipo de trabajo realizará una estimación de la elección Presidencial, estas tres estimaciones estarán basadas en la misma información muestral y deberán realizarse bajo las siguientes consideraciones:

- Las estimaciones se basarán en las hojas del cuadernillo para hacer las operaciones del escrutinio y cómputo obtenidos de las muestras de casillas.
- Las muestras se diseñarán con al menos una confianza del 95 por ciento y una precisión tal que genere certidumbre estadística para cumplir con el objetivo.

¹ De acuerdo con el Artículo 356, numeral 1, del RE del INE, el Conteo Rápido se define como *“el procedimiento estadístico diseñado con la finalidad de estimar con oportunidad las tendencias de los resultados finales de una elección, a partir de una muestra probabilística de resultados de actas de escrutinio y cómputo de las casillas electorales, cuyo tamaño y composición se establecen previamente, de acuerdo con un esquema de selección específico de una elección determinada, y cuyas conclusiones se presentan la noche de la Jornada Electoral”*.

- En las entidades federativas con elección de Gobernatura o Jefatura de Gobierno, la muestra que se utilizará para la elección de la Presidencia de la República será un subconjunto de la muestra de la entidad.
- Los resultados de dichas estimaciones se darán mediante intervalos.
- El reporte de resultados que se presentará y será difundido la noche de la Jornada Electoral especificará las condiciones bajo las cuales se obtuvieron las estimaciones y las conclusiones que de ellas puedan derivarse.

Considerando el Artículo 373 del RE donde se establece que la muestra debe abarcar la mayor dispersión geográfica posible, se concluye que el diseño muestral sea estratificado, donde al interior de cada estrato se seleccionarán casillas mediante un muestreo aleatorio simple sin reemplazo.

2.1.1. Estratificación y tamaño de muestra

En general, para decidir cuál es la mejor estratificación se deben considerar factores como la precisión (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 federal se basa en el análisis de los resultados de las elecciones presidenciales de 2006 y 2012. A partir de diferentes escenarios de tamaño de muestra y considerando dos criterios de estratificación: a) distritos federales y b) distritos federales y tipo de sección, se determina que los distritos federales constituyen la estratificación más conveniente para alcanzar una precisión de 0.25 puntos porcentuales con un nivel de confianza de 95% y un tamaño de muestra de 7,500 casillas (véase anexo).

La simultaneidad del conteo rápido federal con los nueve conteos rápidos locales, requiere establecer un diseño que facilite el trabajo operativo para recabar los datos el día de la Jornada Electoral, por lo que la estratificación se establecerá considerando dos grupos de entidades: las que tienen y las que no tienen elección de Gobernatura/Jefatura de Gobierno. En el primer caso, los estratos corresponderán a la estratificación definida para la elección local y en el segundo caso los estratos estarán constituidos por los distritos federales.

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 tomará una sobre muestra en los estados que tienen dos horas de retraso en relación con la hora del centro, de tal manera que el número de casillas en muestra será de 7,787.

TAMAÑO DE MUESTRA

CRITERIO DE ESTRATIFICACIÓN

TOTAL	CALCULADO	POR HUSO HORARIO	
7,787	7,500	287	<u>Entidades con elección de Gubernatura:</u> Estratificación definida para la elección de Gubernatura <u>Entidades sin elección de Gubernatura:</u> Distritos federales

* 113 corresponden a Baja California y 174 a Sonora

2.1.2. Procedimientos de estimación

Los enfoques estadísticos que se usarán para la estimación de los resultados de la elección federal son el clásico y el bayesiano. Cabe mencionar que, 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.

En total se usarán tres procedimientos de estimación que tendrán el mismo objetivo; sin embargo, dado que los distintos métodos que se utilizarán se basan en supuestos particulares bajo cada enfoque, no se esperan resultados idénticos pero sí muy similares. Las estimaciones se integrarán para presentar intervalos consolidados, al menos al 95% de confianza, referidos al porcentaje de votos a favor de cada candidato a la Presidencia.

De manera general, los procedimientos de estimación del Conteo Rápido federal son:

- El Dr. Raúl Rueda, la Mtra. Patricia Isabel Romero y el Dr. Alberto Alonso (equipo 1) realizarán una estimación conjunta con base en el enfoque de estadística clásica, utilizarán un muestreo estratificado con un estimador de razón combinado.
- El Dr. Luis Enrique Nieto, la Dra. Michelle Anzarut y el Dr. Manuel Mendoza (equipo 2) harán una estimación conjunta basada en el enfoque bayesiano.
- El Dr. Gabriel Núñez, el Dr. Carlos Hernández y el Dr. Carlos Erwin Rodríguez (equipo 3) realizarán, cada uno, las estimaciones mediante el enfoque de estadística clásica y tomarán la mediana de los puntos mínimos y máximos para obtener los intervalos de los resultados de la elección Presidencial.

2.1.3. Integración de las estimaciones

Cada uno de los equipos del COTECORA calculará un intervalo de confianza/probabilidad para cada uno de los candidatos a la Presidencia de la República.

Para emitir estimaciones únicas se construirán intervalos consolidados con base en las estimaciones de cada uno de los tres equipos de especialistas.

El enfoque y método de estimación que emplearán los equipos se resumen a continuación.

EQUIPO DE ASESORES	ENFOQUE ESTADÍSTICO	TIPO DE ESTIMADOR/MODELO DE ESTIMACIÓN	FORMA DE ESTIMACIÓN DEL INTERVALO
Equipo 1	Clásico	Estimador de razón combinado	Métodos asintóticos
Equipo 2	Bayesiano	Método estratificado basado en un modelo normal y ponderado con el listado nominal	Con la distribución posterior
Equipo 3	Clásico	Estimador de razón ponderado con el listado nominal	Métodos asintóticos
	Clásico	Estimador de razón combinado para muestreo estratificado	Re-muestreo
	Clásico	Estimador de razón combinado para muestreo estratificado	Re-muestreo <i>Bootstrap</i>

Para establecer el mecanismo más apropiado para integrar el intervalo de estimación, el Comité realizará los estudios basados en consideraciones metodológicas, así como en simulaciones que consideren una variedad de escenarios.

En el anexo de este documento se presentan detalladamente los procedimientos de estimación.

2.2. Elecciones de Gubernatura y Jefatura de Gobierno

Derivado de los acuerdos adoptados en la tercera reunión de trabajo se determinó que cada entidad federativa quedará bajo la responsabilidad de uno de los Asesores Técnicos del COTECORA.

Asimismo, se acordó que cada equipo de trabajo realizará una o dos estimaciones adicionales para las entidades, como a continuación se describe:

ENTIDAD	ASESOR TÉCNICO RESPONSABLE	2ª/3ª ESTIMACIÓN
Chiapas	Dr. Manuel Mendoza Ramírez	Dr. Luis Enrique Nieto
Ciudad de México	Dr. Carlos Hernández Garciadiego	Dr. Carlos Erwin Rodríguez Dr. Gabriel Núñez
Guanajuato	Dra. Michelle Anzarut Chacalo	Dr. Manuel Mendoza
Jalisco	Dr. Raúl Rueda Díaz del Campo	Dr. Alberto Alonso Mtra. Patricia Isabel Romero
Morelos	Dr. Luis Enrique Nieto Barajas	Dra. Michelle Anzarut
Puebla	Mtra. Patricia Isabel Romero Mares	Dr. Raúl Rueda Dr. Alberto Alonso
Tabasco	Dr. Alberto Alonso y Coria	Mtra. Patricia Isabel Romero Dr. Raúl Rueda
Veracruz	Dr. Gabriel Núñez Antonio	Dr. Carlos Erwin Rodríguez Dr. Carlos Hernández
Yucatán	Dr. Carlos Erwin Rodríguez Hernández-Vela	Dr. Gabriel Núñez Dr. Carlos Hernández

Las estimaciones para cada entidad estarán basadas en la misma información muestral y deberán realizarse bajo las siguientes consideraciones:

- Los resultados de dichas estimaciones se darán mediante intervalos de confianza/probabilidad.
- Las estimaciones se basarán en las hojas del cuadernillo para hacer las operaciones del escrutinio y cómputo obtenidas de las muestras de casillas.

- Las muestras se diseñarán para producir estimaciones con al menos una confianza del 95 por ciento y una precisión tal que genere certidumbre estadística para cumplir con el objetivo.
- El reporte de resultados que se presentará y será difundido la noche de la Jornada Electoral especificará las condiciones bajo las cuales se obtuvieron las estimaciones y las conclusiones que de ellas puedan derivarse.
- El porcentaje esperado de CAE's con una casilla en muestra será de alrededor de 80%.

Considerando el Artículo 373 del RE, el cual determina que la muestra debe abarcar la mayor dispersión geográfica posible, se establece que el diseño muestral sea estratificado, donde al interior de cada estrato se seleccionarán casillas mediante un muestreo aleatorio simple sin reemplazo. Cabe mencionar que la estratificación y tamaño de muestra para hacer las estimaciones son las que determine el Asesor Técnico responsable de la entidad.

2.2.1. 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 cada entidad con elección de Gobernatura, como se muestra en la siguiente tabla.

ESTADO	TAMAÑO DE MUESTRA			CRITERIO DE ESTRATIFICACIÓN	PRECISIÓN
	TOTAL	CALCULADO CON VARIANZA Y ESTRATIFICACIÓN	SOBRE MUESTRA		
Chiapas	500	500	0	Distritos federales con excepción del distrito federal 10 que se divide en tres. Para la división se siguió el criterio de las áreas ocupadas por los distritos locales 14, 15 y 23	1.29% IC 95%: [1.24%,1.38%]
Ciudad de México	1,108	700	408	Distritos locales	0.5%
Guanajuato	500	500	0	Distritos locales	0.75%
Jalisco	467	467	0	Distritos federales	0.85%

ESTADO	TAMAÑO DE MUESTRA			CRITERIO DE ESTRATIFICACIÓN	PRECISIÓN
	TOTAL	CALCULADO CON			
		VARIANZA Y ESTRATIFICACIÓN	SOBRE MUESTRA		
Morelos	200	200	0	Distritos locales	1%
Puebla	509	424	85	Distritos locales	1%
Tabasco	450	450	0	Distritos locales colapsados	1%
Veracruz	1,100	1,100	0	Distrito federal por tipo de sección colapsando 12 estratos asociados al primer cuartil.	0.692%
Yucatán	300	205	95	Distritos locales colapsados	0.96%

2.2.2. Procedimientos 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 2.2.

ASESOR	ENFOQUE ESTADÍSTICO	TIPO DE ESTIMADOR/MODELO DE ESTIMACIÓN	MÉTODO DE CÁLCULO DE ESTIMACIÓN
Dr. Manuel Mendoza Ramírez	Bayesiano	Método estratificado basado en un modelo normal y ponderado con el listado nominal	Con la distribución posterior
Dr. Carlos Hernández Garciadiego	Clásico	Estimador de razón ponderado con el listado nominal	Métodos asintóticos
Dra. Michelle Anzarut Chacalo	Bayesiano	Regresión multinivel con postestratificación	Con la distribución posterior
Dr. Raúl Rueda Díaz del Campo	Bayesiano	Modelo normal para los votos por casilla dentro de los estratos, con independencia	Distribución predictiva de la proporción de votos

ASESOR	ENFOQUE ESTADÍSTICO	TIPO DE ESTIMADOR/MODELO DE ESTIMACIÓN	MÉTODO DE CÁLCULO DE ESTIMACIÓN
Dr. Luis Enrique Nieto Barajas	Bayesiano	entre estratos Método estratificado basado en un modelo normal y ponderado con el listado nominal	Con la distribución posterior
Mtra. Patricia Isabel Romero Mares	Clásico	Estimador de razón combinado	Métodos asintóticos
Dr. Alberto Alonso y Coria	Clásico	Estimador de razón separado y combinado para muestreo estratificado	Métodos asintóticos
Dr. Gabriel Núñez Antonio	Clásico	Estimador de razón combinado para muestreo estratificado	Re-muestreo
Dr. Carlos Erwin Rodríguez Hernández-Vela	Clásico	Estimador de razón combinado para muestreo estratificado	Re-muestreo Bootstrap

2.2.3. Integración de las estimaciones

Cada uno de los miembros del COTECORA calculará un intervalo de confianza/probabilidad para cada uno de los candidatos contendientes de la entidad correspondiente e integrarán sus estimaciones como se describe a continuación:

- El Dr. Raúl Rueda, la Mtra. Patricia Isabel Romero y el Dr. Alberto Alonso, cada uno realizará estimaciones para los siguientes estados: Jalisco, Puebla y Tabasco.
- El Dr. Luis Enrique Nieto, la Dra. Michelle Anzarut y el Dr. Manuel Mendoza realizarán dos estimaciones por entidad federativa y reportarán un intervalo consolidado de ambas estimaciones para cada estado: Chiapas, Guanajuato y Morelos.
- El Dr. Gabriel Núñez, el Dr. Carlos Hernández y el Dr. Carlos Erwin Rodríguez, cada uno realizará estimaciones para las tres entidades: Ciudad de México, Veracruz y Yucatán. El procedimiento para reportar

los intervalos será tomando la mediana de los límites inferiores y la mediana de los límites superiores de los tres intervalos.

3. Selección y resguardo de las muestras

3.1. Consideraciones generales para la selección de las muestras

Las muestras con las que se estimará por un lado, la votación a favor de los candidatos a la Presidencia de la República y por el otro, la votación a favor de los candidatos a las Gubernaturas/Jefatura de Gobierno de cada una de las entidades con elección concurrente se obtendrán en un acto público el 30 junio de 2018. En el acto estará presente un Fedatario que dará fe del desarrollo del protocolo desde la instalación del software requerido hasta la obtención y resguardo de las muestras definitivas.

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 1° de julio de 2018 y la selección se realizará de acuerdo con el diseño de muestreo 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

1. Personal de la Dirección Ejecutiva del Registro Federal de Electores (DERFE) realizará ante Fedatario 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 para la selección de las muestras y la base de datos de casillas así como sus correspondientes códigos de integridad).
3. Personal de la DERFE obtendrá los códigos de integridad de los archivos entregados por el COTECORA y los comparará con los entregados. El Fedatario validará que se trata del mismo código de integridad.
4. A la vista de los asistentes se instalará el software estadístico.
5. Se solicitará a un miembro 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 diez muestras.
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 los 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 del Fedatario, 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, un miembro del COTECORA 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 al Fedatario y a los asistentes que lo soliciten.

3.2.3. Resguardo de las muestras

1. Las muestras serán grabadas en dos discos compactos no regrabables.
2. Un disco será entregado al Director Ejecutivo de Organización Electoral 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 disco con las muestras, los códigos de integridad y los formatos en los que se anotaron 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 el Fedatario.

6. El Fedatario entregará al Secretario Ejecutivo del INE o a la persona que éste último designe los sobres a los que se refiere el numeral anterior para su resguardo.
7. El Secretario Ejecutivo y la DEOE, 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.

4. Criterios logísticos y operativos

4.1. Objetivos de los criterios logísticos y operativos

La operación logística del Conteo Rápido considera la definición de los recursos necesarios para planear el operativo de campo, así como de las acciones que se implementarán para asegurar el adecuado flujo de la información de las casillas de las muestras, federal y locales, al COTECORA el día de la Jornada Electoral. En razón de lo anterior, se busca cumplir con los siguientes objetivos:

a) General

Proveer de manera confiable y oportuna la información de los resultados de las votaciones asentados en los cuadernillos para hacer las operaciones de escrutinio y cómputo de las casillas de las muestras correspondientes, con la finalidad de que el Comité elabore las estimaciones estadísticas para conocer las tendencias de las votaciones de la elección federal Presidencial y de las elecciones locales de Gubernatura o Jefatura de Gobierno el próximo 1º de julio de 2018.

b) Específicos

- Determinar los requerimientos para la etapa de planeación de la operación logística del Conteo Rápido.
- Precisar las funciones que desarrollará el personal involucrado en la ejecución de la operación logística del Conteo Rápido.
- Definir los procedimientos para la recopilación, reporte y captura de los datos de la votación emitida en cada una de las casillas de la muestra.
- Definir un esquema de seguimiento en las juntas distritales ejecutivas para asegurar la oportunidad en la transmisión de los datos de las votaciones.

4.2. Esquema general de funcionamiento del Conteo Rápido

Con la finalidad de brindar información oportuna a la población, tras el cierre de las casillas, sobre las tendencias de los resultados de las votaciones de la elección federal Presidencial y de las elecciones locales de Gubernatura y Jefatura de Gobierno, el INE llevará a cabo el próximo 1º de julio de 2018 los

ejercicios de Conteo Rápido, con base en la información asentada en los cuadernillos para hacer las operaciones de escrutinio y cómputo de casilla.

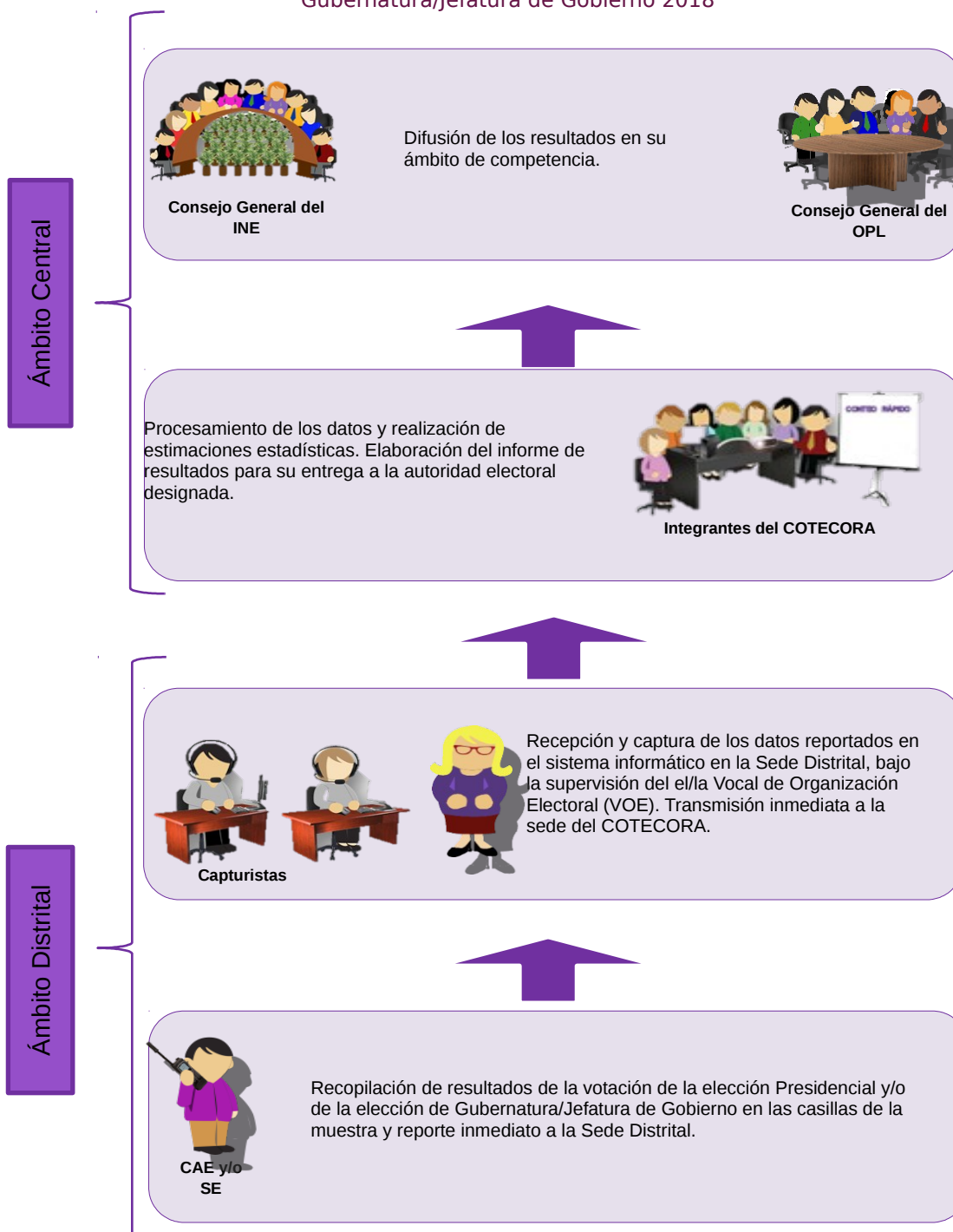
Para lograr este propósito, se requiere planear y ejecutar una serie de actividades coordinadas que tienen su base en el siguiente procedimiento general, mismo que tendrá verificación el día de la Jornada Electoral:

1. El personal en campo, Capacitador Asistente Electoral (CAE) o Supervisor Electoral (SE), se encargará de recabar en el formato diseñado para tal fin, los resultados asentados en los cuadernillos para hacer las operaciones de escrutinio y cómputo de la(s) casilla(s) seleccionada(s) dentro de la muestra, asignada(s) a su Área de Responsabilidad Electoral (ARE).
2. Una vez llenados los formatos, de manera inmediata llamará a través del medio de comunicación que le fue asignado (teléfono celular, satelital o telefonía pública), a la Sede Distrital² correspondiente para reportar la información recopilada.
3. En la Sede Distrital, los/las capturistas recibirán las llamadas e ingresarán directamente en el Sistema de Información del Conteo Rápido 2018 los datos que le sean comunicados, esto para su transferencia inmediata a la sede del COTECORA.
4. Los integrantes del COTECORA procesarán la información proporcionada por el sistema y realizarán las estimaciones estadísticas correspondientes. A partir de ello, elaborarán un informe sobre los resultados obtenidos y lo enviarán a la autoridad electoral que corresponda para su difusión.
5. La autoridad electoral respectiva dará a conocer a la opinión pública, la noche de la jornada electoral, los resultados del Conteo Rápido de la elección correspondiente a su ámbito de competencia.

El procedimiento general antes descrito, desde el acopio de datos en campo por parte de los/las CAE's y SE, hasta la difusión de resultados en la noche de la Jornada Electoral, se representa gráficamente en el siguiente esquema:

² El reporte se realizará al número telefónico del sistema multilíneas para funcionamiento de la Sala del SIJE.

Figura 1. Procedimiento general del Conteo Rápido para las elecciones federal Presidencial y de Gobernatura/Jefatura de Gobierno 2018



Al respecto, es importante hacer notar que la logística se desarrollará básicamente en el ámbito distrital con la participación fundamental de los/las CAE's y SE.

En lo que sigue, se puntualizarán las etapas de planeación en donde se hace referencia a los diversos recursos requeridos para la ejecución de este proyecto, así como los procedimientos que deberán realizarse antes, durante y después de la Jornada Electoral.

4.2.1. Etapas de planeación

4.2.1.1. Definición de recursos requeridos

A continuación, se describen los recursos humanos, y sus principales funciones, así como los recursos materiales y financieros necesarios en cada una de las Juntas Distritales Ejecutivas.

I. Recursos humanos

- Vocal Ejecutivo(a) Distrital (VED).
 - ✓ Con el apoyo de el/la Vocal de Organización Electoral Distrital asegurará la oportuna disponibilidad y funcionamiento de todos los recursos humanos, materiales y financieros necesarios para el Conteo Rápido.
 - ✓ Supervisará la realización de la prueba de captura y simulacros.
 - ✓ Vigilará el desarrollo del operativo de campo del Conteo Rápido durante la Jornada Electoral a efecto de garantizar el adecuado flujo en el reporte de información.
 - ✓ Dará puntual seguimiento a los procedimientos que deban aplicarse en caso de alguna contingencia.
- Vocal de Organización Electoral Distrital (VOED)
 - ✓ Responsable directo de la ejecución del operativo de campo del Conteo Rápido.
 - ✓ Brindará la capacitación correspondiente a los/las CAE's, SE, y los/las capturistas que tomarán parte en el proyecto.
 - ✓ Coordinará la realización de los simulacros.
 - ✓ Dará aviso al personal en campo a quienes les haya sido asignada una o más casillas de la muestra para que, al término del escrutinio y cómputo de los votos el día de la Jornada Electoral, reporten los resultados de la votación correspondiente.

- ✓ Analizará, en su caso, si las y los SE deberán apoyar en el reporte de resultados de alguna casilla de la muestra.
- ✓ Dará seguimiento oportuno al reporte de resultados de las casillas en muestra y, si es el caso, mantendrá comunicación constante con el/la SE para garantizar el flujo de la información.
- ✓ En caso de presentarse saturación en las líneas telefónicas para el reporte de los resultados, deberá definir los mecanismos de información para indicar oportunamente al personal de campo a qué instancia comunicarse y así garantizar el correcto flujo de la información.
- ✓ De ser necesario aplicar el plan de contingencia, será el responsable de brindar las indicaciones pertinentes al personal en campo con el objetivo de cumplir con el flujo de la información.
- Coordinador/a Distrital³
 - ✓ Apoyará a el/la VOED en el cumplimiento de sus funciones.
 - ✓ Auxiliará a los/las capturistas en caso de que se presenten fallas en el sistema informático.
 - ✓ En su caso, se encargará de verificar la personalidad de los/las CAE's o SE que por alguna razón no cuenten con su clave de autenticación para el reporte de datos.
- Capturistas⁴
 - ✓ Recibirán capacitación sobre las funciones que deberán desempeñar en la ejecución del operativo de campo del Conteo Rápido.
 - ✓ Participarán en la prueba de captura del sistema informático y en los simulacros.
 - ✓ Recibirán las llamadas del personal en campo para capturar, en el sistema informático correspondiente, los datos de la votación que les sean comunicados.
 - ✓ Proporcionarán la clave de confirmación del reporte de resultados a los/las CAE's o SE, una vez que la información haya sido corroborada (de acuerdo con las precisiones de dictado indicadas en la capacitación) y capturada en el sistema informático.

3 Corresponde al Coordinador Distrital de la Sala del SIJE.

4 Básicamente son las/las capturistas que atenderán el SIJE.

- Supervisores/as electorales
 - ✓ Recibirán capacitación sobre las funciones que deberán desempeñar en la ejecución del operativo de campo del Conteo Rápido.
 - ✓ Todos ellos participarán en la realización de los tres simulacros.
 - ✓ Verificarán que los/las CAE's bajo su responsabilidad reporten la información de las casillas de la muestra que les corresponda.
 - ✓ A solicitud de el/la VOED, apoyarán en el acopio y reporte de datos de votación de las casillas incluidas en la muestra, esto cuando algún/a CAE tenga que reportar más de una.
 - ✓ Mantendrán comunicación constante con el/la VOED para dar seguimiento oportuno al reporte de resultados de las casillas asignadas a su Zona de Responsabilidad Electoral (ZORE).
- Capacitadores/as Asistentes Electorales
 - ✓ Recibirán capacitación sobre las funciones que deberán desempeñar en la ejecución del operativo de campo del Conteo Rápido.
 - ✓ Todos ellos participarán en la realización de los tres simulacros.
 - ✓ Recopilarán los resultados de las votaciones de las elecciones Presidencial y/o de Gobernatura/Jefatura de Gobierno registrados en los cuadernillos para hacer las operaciones de escrutinio y cómputo de casilla de la muestra, conforme a las indicaciones recibidas por parte de el/la VOED y los reportarán inmediatamente a la Sede Distrital.
 - ✓ Para agilizar la ejecución del escrutinio y cómputo de la casilla de la muestra, podrán brindar orientación y apoyo a los funcionarios de casilla en los procedimientos correspondientes.
 - ✓ En su caso, mantendrá comunicación permanente con su SE para reportar las causas que impidan el reporte de resultados de la votación de la(s) casilla(s) asignada(s).

II. Recursos materiales

El operativo de campo del Conteo Rápido basa su funcionamiento en la capacidad instalada para implementar el *Sistema de Información de la*

Jornada Electoral (SIJE) en cada una de las juntas distritales ejecutivas. Los recursos materiales se agrupan en dos categorías:

1. *Para el reporte y recepción de la información.* Son los medios de comunicación que utilizarán los/las CAE's y SE desde campo, así como aquellos que emplearán los/las capturistas en cada Junta Distrital Ejecutiva:
 - ✓ Los/las CAE's y SE utilizarán principalmente teléfonos celulares y en menor medida teléfonos públicos rurales y teléfonos satelitales, estos últimos principalmente en aquellas casillas ubicadas en lugares alejados que no cuentan con servicios de telefonía convencional.
 - ✓ Los/las capturistas dispondrán de diademas telefónicas vinculadas a un sistema multilíneas para atender las llamadas de las/los CAE's y SE.
 - ✓ Para la captura y transmisión de la información. Se instalarán equipos de cómputo con conexión a la RedINE, para el registro y transmisión inmediata de los datos a la sede del COTECORA a través del sistema informático diseñado para tal fin. Se procurará que no estén muy cercanos a fin de evitar distorsiones de la información que recibirá cada capturista.
2. *Para acondicionamiento.*
 - ✓ Mobiliario. Se contará con sillas, mesas de trabajo y equipos de cómputo -preferentemente de los que se encuentran bajo el esquema de arrendamiento- en cantidad igual al número de líneas telefónicas en el sistema multilíneas.
 - ✓ Plantas de energía eléctrica. Cada Junta Distrital Ejecutiva deberá disponer de este equipo para prevenir posibles eventualidades por falta de suministro de energía eléctrica.

III. Recursos financieros

Se contará con la asignación de recursos para el uso de telefonía celular y satelital -en menor medida-, así como moneda fraccionaria para el pago de llamadas a través de telefonía pública rural. Estos recursos estarán destinados para la realización de simulacros y para su uso el día de la Jornada Electoral.

4.2.2. Preparación

Previo a la Jornada Electoral del 1º de julio de 2018, se llevarán a cabo dos etapas de preparación para la ejecución del operativo de campo: 1) capacitación de los recursos humanos y 2) realización de, por lo menos, una prueba de captura y tres simulacros.

Para ello, la DEOE se encargará de remitir oportunamente a través de el/la Vocal de Organización Electoral (VOE) de la Junta Local Ejecutiva, los lineamientos y documentos de apoyo para que los haga del conocimiento de los/las VOED y coordine la realización de las actividades a que haya lugar.

4.2.2.1. Capacitación

Las actividades que se llevarán a cabo están dirigidas a la preparación y entrenamiento del personal involucrado en la ejecución del operativo de campo del Conteo Rápido.

De manera inicial, durante el mes de abril, se llevará a cabo la capacitación a las/los VOED, con la participación fundamental de los VOE locales y personal de la Dirección de Planeación y Seguimiento (DPS) de la DEOE, encaminada a fortalecer los conocimientos que tendrán que transmitir al personal en campo, ya que esta actividad será responsabilidad de el/la VOED, bajo la estricta supervisión de el/la VED.

A diferencia de los procesos electorales anteriores, en esta ocasión se prevé la participación de personal de las Juntas Locales Ejecutivas, en aquellas entidades federativas en las se implementará de forma exclusiva el Conteo Rápido federal de la elección Presidencial, actuando como instancia de atención de contingencias. Asimismo, se instrumentará un mecanismo de capacitación más supervisado, específico y particular para reforzar el proceso de aprendizaje en todos los niveles de responsabilidad, con mayor énfasis en el caso de las entidades federativas donde adicionalmente a dicho Conteo, se implementará el correspondiente a la elección local de Gobernatura o Jefatura de Gobierno.

Por otra parte, el/la VOED incluirá dentro del Segundo Taller de Capacitación para SE y CAE's (del 2 al 7 de mayo de 2018), una explicación detallada sobre el operativo de campo a este personal, así como a los/las capturistas y a el/la Coordinador/a Distrital. En dicho taller, será conveniente considerar la asistencia de quienes integren el Consejo Distrital, así como del personal de las Juntas Locales, con el objetivo de que se capaciten y conozcan los procedimientos y actividades que se implementarán para el cabal cumplimiento del proyecto.

Para ello, la DEOE a través de la DPS, remitirá la *Guía de procedimientos para el operativo de campo* en la que se detallarán las funciones y actividades que

deberá realizar el personal involucrado en la recopilación, reporte, captura y transmisión de datos de las casillas de la muestra. Asimismo, enviará material audiovisual en el que se integrarán los aspectos principales contenidos en la guía para facilitar el proceso de capacitación.

Con base en la experiencia de los procesos electorales previos en los que se han efectuado ejercicios de Conteo Rápido, **es fundamental que la totalidad de los/las CAE's y SE cuenten con los conocimientos suficientes sobre la operación logística que se ejecutará, con la finalidad de que estén preparados para realizar el reporte de resultados de alguna de las casillas bajo su responsabilidad, en la noche de la Jornada Electoral.**

En esta etapa de preparación, el/la VOED en su carácter de coordinador/a del operativo de campo del Conteo Rápido, tendrá la responsabilidad de:

- Revisar los documentos enviados por la DEOE en los que se describen los procedimientos y actividades que se llevarán a cabo, así como los formatos que se utilizarán para la recopilación, reporte y captura de la información.
- Capacitar a el/la Coordinador/a Distrital, SE, CAE, así como a las/los capturistas para que cumplan a cabalidad con sus respectivas funciones.
- Distribuir a los/las CAE's y SE los formatos con los que se recabará la información de la votación recibida de las casillas que integrarán la muestra. Asimismo, dar a conocer los números telefónicos de la Sede Distrital y, en su caso, de las instancias de apoyo para la atención de contingencias.

4.2.2.2. Realización de prueba de captura y simulacros

Con el fin de recrear el contexto y situaciones que pudieran ocurrir en la Jornada Electoral, previamente se llevarán a cabo, por lo menos, una prueba de captura y tres simulacros de reporte de resultados.

a) Prueba de captura

Este ejercicio persigue varios objetivos relacionados con el sistema informático entre los que destacan: a) verificar el acceso de las/los capturistas; b) familiarizar al personal con el diseño y funcionamiento del sistema; c) detectar los aspectos que pudieran ocasionar errores a la hora de ingresar datos; d) identificar posibles fallas para proceder a subsanarlas. Además, se busca probar las capacidades de capturistas para la recepción e ingreso de datos a efecto de detectar posibles debilidades y corregirlas.

Esta prueba se realizará en la segunda quincena de mayo de 2018 con la participación exclusiva de quienes actuarán como capturistas, bajo la coordinación y supervisión de el/la VOED con apoyo de el/la Coordinador/a Distrital.

b) Simulacros

La realización de estos ejercicios tiene como objetivos: a) probar el funcionamiento de los medios de comunicación asignados a los/las CAE's y SE desde campo, así como la adecuada ejecución de los procedimientos de reporte de datos; b) verificar la correcta captura y transmisión de la información; c) comprobar el funcionamiento del sistema informático. Todo ello con la finalidad de detectar oportunamente cualquier posible falla en esos aspectos y realizar los ajustes necesarios para garantizar el óptimo desarrollo de la operación logística el día de la Jornada Electoral.

En este sentido, se realizarán, al menos, simulacros los días 3, 10 y 17 en el mes de junio. Para su ejecución, será necesario contar previamente con la relación de casillas seleccionadas por el COTECORA, así como datos de votación ficticia que deberá emplearse, para que los/las CAE's y SE participen en el reporte de la información conforme les instruya el/la VOED.

4.2.3. Ejecución del operativo de campo

Se ejecutarán tres fases principales: a) recopilación, b) reporte, y c) captura-transmisión. A continuación, se presentan con mayor detalle cada una de estas fases.

4.2.3.1. Recopilación de la información

Las acciones que ejecutarán las/las CAE's, o en su caso SE, para el acopio de los datos de las casillas de la muestra son las siguientes:

- a) En la Jornada Electoral (1º de julio de 2018), acudirá a las 18:00 horas (hora local) a la casilla seleccionada de su ARE. Ahí permanecerá para brindar el apoyo y orientación necesarios a los funcionarios de Mesa Directiva de Casilla (MDC) para agilizar los procedimientos de escrutinio y cómputo de cada una de las elecciones que tengan lugar.
- b) Una vez que el Secretario/a concluya con el llenado del cuadernillo para hacer las operaciones de escrutinio y cómputo de casilla de la elección que corresponda reportar (Presidencial y/o Gobernatura o Jefatura de Gobierno), procederá de inmediato a transcribir los resultados en el *Formato para recopilación de resultados de la*

votación de dicha elección, asegurándose de que cada uno de los datos transcritos corresponde fielmente con los del documento fuente.

- c) Solicitará al Presidente o Secretario que valide la información asentada en el formato y llamará de inmediato a la Sede Distrital para realizar el reporte. En su caso, los CAE's que hagan uso de telefonía pública rural deberán trasladarse de inmediato al lugar donde ésta se ubique para realizar el reporte.
- d) Si eventualmente, se le hubiera asignado otra casilla para reportar y esta estuviera próxima a la recién reportada, tendrá que trasladarse de inmediato a dicha ubicación para realizar el mismo procedimiento.
- e) En aquellos casos donde a un/una CAE le corresponda realizar el reporte de resultados de la elección federal y de la elección local de Gobernatura o Jefatura de Gobierno en una misma casilla, **deberá llenar el formato correspondiente a cada elección y realizar el reporte de ambos resultados a la Sede Distrital dentro de la misma llamada**. Los formatos estarán debidamente diferenciados en hojas de distinto color.

4.2.3.2. Reporte inmediato a la Sede Distrital

Como se ha indicado, una vez que el personal en campo termine de copiar los datos de los cuadernillos para hacer las operaciones de escrutinio y cómputo en el formato asignado, **de manera inmediata**, llamará a la Sede Distrital a que pertenezca para reportar la información recopilada.

En la Sede Distrital estará disponible un número definido de capturistas equipados con diademas telefónicas conectadas a un sistema multilíneas y equipos informáticos conectados a la RedINE.

Cuando se establezca comunicación vía telefónica con el/la capturista, las/los CAE's deberán reportar los datos recabados conforme a lo siguiente:

- a) Dictará el tipo de elección (Presidencial o Gobernatura/Jefatura de Gobierno), los datos de identificación de la casilla y la clave de autenticación⁵ (conformada por cuatro dígitos), con lo cual se acreditará la personalidad de quien reporta la información. Dicha clave es la llave de acceso al *Sistema de Información del Conteo Rápido 2018* para el ingreso de los datos de votación.

⁵ Las claves de autenticación utilizadas durante los simulacros serán distintas de las que se emplearán durante la Jornada Electoral. En este sentido, para la Jornada Electoral, el/a Coordinador/a Distrital deberá disponer de la relación de las claves asignadas al personal en campo en la que se incorporen datos personales para que, de ser necesario, tenga elementos para identificar la personalidad de éstos en caso de que no cuenten con dicha clave.

- b) A la indicación de el/la capturista, el/la CAE dictará los resultados de la votación que obtuvo cada partido político, coalición, candidatos independientes (en su caso), candidatos no registrados y los votos nulos, para que el/la capturista los ingrese en dicho sistema informático. Las cifras de los resultados de las votaciones de cada casilla serán proporcionadas una por una y en dos ocasiones, conforme al protocolo de dictado que se defina.
- c) De coincidir la doble captura de datos, el sistema guardará la información y mostrará una clave de confirmación que deberá ser dictada por la/el capturista a el/la CAE, para asegurar que cumplió con el reporte de la información. En caso contrario, el/la capturista solicitará a el/la CAE, la confirmación de los datos nuevamente hasta que éstos coincidan.
- d) El/la CAE concluirán la llamada hasta que el/la capturista se lo indique.

En el caso exclusivo de las entidades federativas que atenderán los Conteos Rápidos de la elección federal Presidencial y de la elección local de Gobernatura o Jefatura de Gobierno, es probable que en ciertas horas de la noche se genere un mayor flujo de llamadas a las sedes distritales. Para evitar una saturación de las líneas telefónicas, las/los CAE's procederán como sigue:

- Intentarán comunicarse a la Sede Distrital, realizando **dos llamadas** con intervalos de **dos minutos cada una**.
- Si después de los dos intentos no logran establecer comunicación, llamarán al número telefónico de INETEL, conforme a los procedimientos que para ello se definan.
- La llamada a INETEL deberá permanecer en espera hasta que entre la llamada o hasta que ésta sea interrumpida de forma automática.
- Si después de realizar dos intentos a INETEL, la llamada se interrumpe de forma automática, deberá intentar llamar a la Sede Distrital conforme al procedimiento antes referido, y así de forma alternada hasta concretar la llamada en alguna de las dos instancias.

4.2.3.3. Captura en el sistema y transmisión al COTECORA

A continuación se detalla la captura de la información en las sedes distritales. Esta se llevará a cabo mediante el siguiente procedimiento:

- El/la capturista utilizará una diadema telefónica para recibir la llamada de el/la CAE o SE; de esta manera, podrá registrar cada dato al momento en que sea dictado.

- La captura se realizará directamente en el *Sistema de Información del Conteo Rápido 2018* en dos ocasiones. El primer dictado se realizará dígito por dígito y el segundo dictado podrá realizarse en cifras completas omitiendo los ceros a la izquierda. En caso de que hubiera discrepancias entre los registros, se deberá solicitar nuevamente el dictado de los datos respectivos para corregir la información ingresada en el sistema.
- Una vez que son guardados los datos, el sistema generará una clave de confirmación para que el/la capturista la dicte a el/la CAE o SE y, así podrá concluir la llamada telefónica.
- El/la CAE o SE registrará en el formato correspondiente la clave de confirmación, con la cual comprobará que realizó el reporte de resultados de la casilla respectiva.
- La información capturada en el sistema informático se transmitirá de forma automática a la sede del COTECORA, cuyos integrantes se encargarán de procesar la información, así como de realizar las estimaciones estadísticas correspondientes.

4.2.4. Esquemas de contingencia

Como medida preventiva a escenarios adversos que pudieran ocurrir de forma imprevista el día de la Jornada Electoral y obstaculizar el reporte de datos a las Sedes Distritales, se implementarán esquemas de contingencia, para garantizar el flujo de la información y dar continuidad a las acciones que realizará el personal en campo, ante situaciones tales como:

- Fallas en el sistema informático.
- Fallas en la señal de la RedINE.
- Fallas en el funcionamiento de las líneas telefónicas o suspensión del servicio.
- Suspensión del suministro de energía eléctrica.
- Toma de instalaciones o situaciones de violencia que pongan en riesgo las actividades al interior de la junta distrital.

Los esquemas de contingencia serán diferentes en función del número de Conteos Rápidos que atienda el distrito electoral de la entidad federativa, esto es:

- Sólo atienden Conteo Rápido federal.

- Atienden los dos tipos de Conteo Rápido (de la elección federal Presidencial y de la elección local de Gobernatura o Jefatura de Gobierno).

a) Atienden únicamente el Conteo Rápido federal

Para accionar el plan de contingencia se deberá seguir el siguiente procedimiento:

- El/la VOED deberá realizar el reporte de la situación que se presente al Centro de Atención a Usuarios (CAU) y asimismo, notificar a la DPS de la DEOE, a través del personal previamente asignado para brindarle atención. Este procedimiento podrá realizarse vía telefónica o por correo electrónico.
- El personal de la DPS estará en continua comunicación con el/la VOED a efecto de dar seguimiento y, en su caso, proceder a aplicar el esquema de contingencia con la participación de la Junta Local Ejecutiva correspondiente.
- En la Junta Local Ejecutiva se preparará a las/los capturistas para recibir las llamadas de las/los CAE's.
- En la vocalía de Organización Electoral correspondiente, se encargarán de dar aviso a las/los CAE's para que realicen el reporte de resultados a la sede de la Junta Local Ejecutiva. El/la VOED se apoyará en el/la Coordinador/a Distrital para hacer llegar las instrucciones a CAE's y SE sobre el procedimiento a seguir.
- Si el problema se soluciona antes de que concluya el reporte de casillas de la muestra, se retomarán las llamadas en la Sede Distrital correspondiente, para lo cual el/la VOED se encargará de dar aviso al personal en campo pendiente de realizar su reporte.
- El/la VOED tendrá la obligación de revisar el reporte de avance del sistema informático a efecto de dar seguimiento y, en su caso, retomar la comunicación con los/las CAE's y SE para asegurar el adecuado flujo de la información.

b) Atienden los dos tipos de Conteo Rápido (federal y local)

Para accionar el plan de contingencia se deberá seguir el siguiente procedimiento:

- El/la VOED deberá realizar el reporte de la situación que se presente al CAU y asimismo, notificar a la DPS de la DEOE, a través del personal

previamente asignado para brindarle atención. Este procedimiento podrá realizarse vía telefónica o por correo electrónico.

- El personal de la DPS estará en continua comunicación con el/la VOED a efecto de dar seguimiento y, en su caso, proceder a aplicar el esquema de contingencia con la participación del personal de INETEL.
- En la vocalía de Organización Electoral correspondiente, se encargarán de dar aviso a las/los CAE's para que realicen el reporte de resultados directamente a INETEL. El/la VOED se apoyará en el/la Coordinador/a Distrital para hacer llegar las instrucciones a CAE's y SE sobre el procedimiento a seguir.
- Si el problema se soluciona antes de que concluya el reporte de casillas de la muestra, se retomarán las llamadas en la Sede Distrital correspondiente, para lo cual el/la VOED se encargará de dar aviso al personal en campo pendiente de realizar su reporte.
- El/la VOED tendrá la obligación de revisar el reporte de avance del sistema informático a efecto de dar seguimiento y, en su caso, retomar la comunicación con los/las CAE's y SE para asegurar el adecuado flujo de la información, y evitar saturar las llamadas a INETEL.

4.2.5. Esquema de seguimiento

Con la finalidad de asegurar el flujo en el reporte de información, se prevé que en el ámbito distrital se aplique un esquema de seguimiento con apoyo de los datos que proporcione el *Sistema de Información del Conteo Rápido 2018*.

Para ello, el sistema dispondrá de un reporte de las casillas pertenecientes a la muestra cuya información no haya sido reportada. Con base en dicho reporte el/la VOED, con apoyo de el/la Coordinador/a Distrital, tendrá la obligación de verificar las casillas de la elección que se encuentren en esa situación, a partir de las 20:00 horas⁶ (horario local) para que a través de los/las SE a cargo del ARE a que éstas pertenezcan⁷, se comuniquen con los/las CAE's respectivos a efecto de conocer el avance en el escrutinio y cómputo o, en su caso, conocer si se presentan contratiempos que

⁶ Considerando que las casillas cierran en punto de las 18:00 horas y el escrutinio y cómputo de la casilla de la elección de Presidencial dure aproximadamente 1 hora 30 minutos.

⁷ Para ello será necesario que previo a la Jornada Electoral, el/la VOED defina un esquema para estar en comunicación constante con los/las SE y ellos/as, a su vez, con los/las CAE's a su cargo. Por ejemplo, si la mayoría cuenta con teléfono celular, se podría implementar un grupo de Chat para una comunicación más expedita.

obstaculicen el reporte de resultados para así, instruir las acciones a que haya lugar a fin de garantizar el flujo de la información.

Una vez que se lleve a cabo la difusión de las estimaciones a la opinión pública, los reportes de resultados que se reciban posteriormente no serán considerados por el COTECORA, motivo por el cual **es fundamental que el trabajo de supervisión por parte de los/las VOED promueva el mayor reporte de información antes de las 22:30 horas (hora del centro)** o, en su defecto, hasta el momento que le sea indicado, exclusivamente, por la vía de comunicación oficial definida para el Conteo Rápido.

Luego de que este evento tenga lugar, previa indicación oficial, el/la VOED podrá dar aviso al personal en campo que no hubiere concretado el reporte de datos que se podrá prescindir de hacerlo; no obstante, se tendrá la obligación de asentar en el formato correspondiente los motivos que impidieron llevar a cabo dicho reporte, conforme a las indicaciones que previamente se hayan brindado.

Se deberá describir el procedimiento de entrega de los reportes de las estimaciones de los resultados de las elecciones a los OPL respectivos, en el correspondiente informe del COTECORA.

4.3. Medidas de Seguridad en la sede del COTECORA

El día 1 de julio de 2018, el recinto acondicionado para la estimación de las tendencias de los resultados de la votación de la Presidencia de los Estados Unidos Mexicanos y de las Gubernaturas de Chiapas, Guanajuato, Jalisco, Morelos, Puebla, Tabasco, Veracruz y Yucatán, así como de 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 tendrán acceso al recinto las siguientes personas:
 - Integrantes del Comité Técnico y colaboradores.
 - Personal de apoyo aprobado por el Secretario Técnico del Comité.

Toda persona que tenga acceso al recinto deberá registrarse 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 segura 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.

5. Anexos. Diseño Muestral y Métodos de Estimación

5.1. Elección Presidencial

5.1.1. Estratificación y tamaño de muestra

Determinación del tamaño de muestra de casillas para la elección Presidencial de 2018⁸

Con base en las elecciones Presidenciales del 2006 y 2012, bajo un diseño aleatorio simple sin reemplazo de casillas y bajo un diseño estratificado aleatorio, con diferentes formas de estratificación, se calcularon precisiones para los tres principales partidos, considerando diferentes tamaños de muestra. Con los resultados de la elección Presidencial del 2012 se calcularon también tamaños de muestra para diferentes precisiones.

Elecciones 2006

La información en que se basan los cálculos, es la del Programa de Resultados Electorales Preliminares (PREP) de votación para Presidente del año 2006, que contiene 117,287 casillas.

Con esta información se calcularon las precisiones teóricas con un 95% de confianza, considerando diferentes tamaños de muestra entre 5,000 y 7,500 casillas y diferentes diseños de muestra. Para el diseño estratificado aleatorio con asignación proporcional de la muestra a los estratos, se consideró al estimador de razón combinado y se trabajó con distritos federales. La información referente a tipos de casilla: urbana y no urbana, se tomó del marco de casillas de mayo de 2006.

La precisión es $\delta = \sqrt{V(\hat{R})} * 1.96 * 100$ con 95% de confianza.

DISEÑO M.A.S. DE CASILLAS	PAN	APM	PBT
5000	0.49%	0.32%	0.50%
5500	0.47%	0.30%	0.48%
6000	0.45%	0.29%	0.46%
6500	0.43%	0.28%	0.44%
7000	0.41%	0.26%	0.42%

⁸ Elaborado por la Mtra. Patricia Isabel Romero Mares con la colaboración de la Mtra. Adriana Ducoing Watty.

DISEÑO M.A.S. DE CASILLAS	PAN	APM	PBT
7500	0.40%	0.26%	0.41%

DISEÑO ESTRATIFICADO (300 DISTritos FEDERALES) DISTRIBUCIÓN PROPORCIONAL	PAN	APM	PBT
5000	0.31%	0.22%	0.27%
5500	0.29%	0.21%	0.26%
6000	0.28%	0.20%	0.25%
6500	0.27%	0.19%	0.24%
7000	0.26%	0.19%	0.23%
7500	0.25%	0.18%	0.22%

DISEÑO ESTRATIFICADO (DISTRITO FEDERAL 300 X TIPO URBANO Y NO URBANO, 481) DISTRIBUCIÓN PROPORCIONAL	PAN	APM	PBT
5000	0.30%	0.22%	0.26%
5500	0.29%	0.21%	0.25%
6000	0.28%	0.20%	0.24%
6500	0.26%	0.19%	0.23%
7000	0.25%	0.18%	0.22%
7500	0.25%	0.17%	0.21%

Elecciones 2012

La información en que se basan los cálculos es la del archivo de votación para Presidente del año 2012 "Computos_distritales_Presidente_2012.txt", que contiene 143,437 casillas.

Con esta información se calcularon las precisiones teóricas con un 95% de confianza, considerando diferentes tamaños de muestra entre 5,000 y 7,500 casillas y diferentes diseños de muestra. Para el diseño estratificado aleatorio con asignación proporcional de la muestra a los estratos, se consideró al estimador de razón combinado y se trabajó con distritos federales. La información referente a tipos de casilla: urbana y no urbana, se tomó del mismo archivo.

La precisión es $\delta = \sqrt{V(\hat{R})} * 1.96 * 100$ con 95% de confianza.

DISEÑO M.A.S. DE CASILLAS	PAN	PRI	PRD
5000	0.37%	0.32%	0.43%
5500	0.36%	0.31%	0.41%
6000	0.34%	0.29%	0.39%
6500	0.33%	0.28%	0.38%
7000	0.31%	0.27%	0.36%
7500	0.30%	0.26%	0.35%

DISEÑO ESTRATIFICADO (300 DISTRITOS FEDERALES) DISTRIBUCIÓN PROPORCIONAL	PAN	PRI	PRD
5000	0.24%	0.24%	0.26%
5500	0.23%	0.23%	0.24%
6000	0.22%	0.22%	0.23%
6500	0.21%	0.21%	0.23%
7000	0.20%	0.20%	0.21%
7500	0.20%	0.20%	0.21%

DISEÑO ESTRATIFICADO (DISTRITO FEDERAL 300 X TIPO URBANO Y NO URBANO, 555) DISTRIBUCIÓN PROPORCIONAL	PAN	PRI	PRD
5000	0.24%	0.24%	0.25%
5500	0.23%	0.22%	0.23%
6000	0.22%	0.21%	0.22%
6500	0.21%	0.21%	0.21%
7000	0.20%	0.20%	0.21%
7500	0.19%	0.19%	0.20%

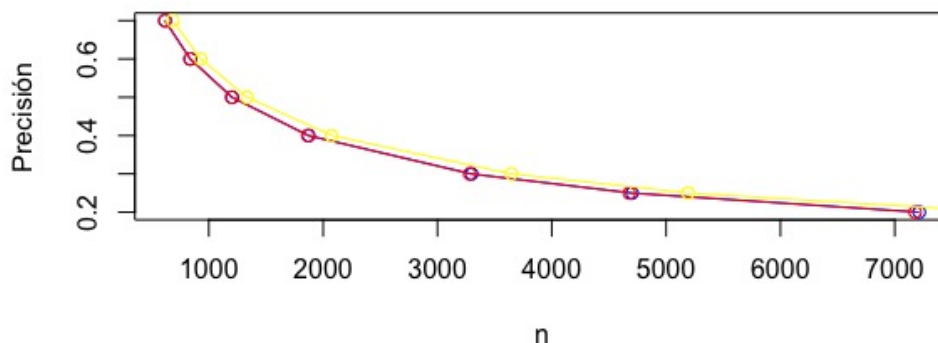
Tamaños de muestra para obtener una precisión específica con una confianza del 95%

Si se consideran 300 estratos (distritos federales), con distribución de la muestra proporcional al número de casillas en el estrato, m.a.s. de casillas en cada estrato y el estimador de razón combinado, el tamaño de muestra para cada uno de los principales partidos, fijando la precisión es:

PRECISIÓN	PAN	PRI	PRD
------------------	------------	------------	------------

0.2%	7214	7178	7950
0.25%	4702	4678	5191
0.3%	3298	3281	3645
0.4%	1874	1864	2073
0.5%	1205	1198	1334
0.6%	839	834	929
0.7%	617	614	683

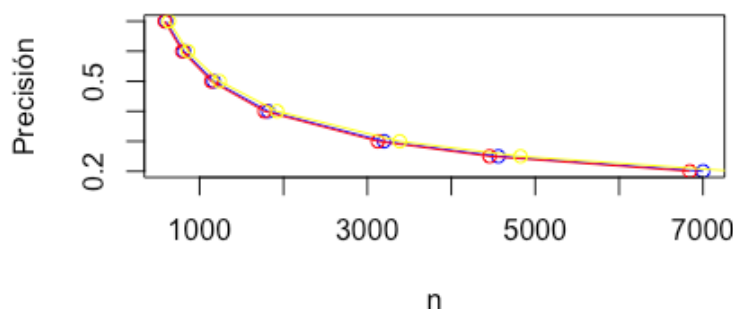
Tamaño de muestra, 300 estratos



Si consideramos 555 estratos (distritos federales x tipo de sección), con distribución de la muestra proporcional al número de casillas en el estrato, m.a.s. de casillas en cada estrato y el estimador de razón combinado, el tamaño de muestra para cada uno de los principales partidos, fijando una precisión es:

PRECISIÓN	PAN	PRI	PRD
0.2%	7001	6843	7400
0.25%	4561	4456	4826
0.30%	3199	3125	3386
0.40%	1817	1775	1925
0.50%	1169	1141	1238
0.60%	814	795	862
0.70%	599	585	635

Tamaño de muestra, 555 estratos



Con respecto a tipo de sección: urbanas y no urbanas, su distribución en el año 2012 y su votación:

	PAN	PRI	PRD	% CASILLAS
Total	0.254	0.382	0.316	100%
Urbanas	0.257	0.359	0.338	66%
No urbanas	0.247	0.434	0.267	34%

Con base en la información presentada se concluyó que la estratificación por distrito federal es la más conveniente porque las precisiones no mejoraron substancialmente con una estratificación más fina; sin embargo, las precisiones sí mejoran respecto a un muestreo aleatorio simple sin reemplazo de casillas.

Considerando la información de los años 2006 y 2012 se decidió una precisión del 0.25% con una confianza de 95 % para la elección presidencial del 2018, lo que implica un tamaño de muestra de 7,500 casillas que se distribuirán de manera proporcional al número de casillas de cada estrato.

Debido a los diferentes usos horarios en el país, y la dificultad que esto implica para recibir información de manera oportuna, se decidió tomar una sobre muestra en Sonora de 174 casillas y en Baja California de 113 casillas.

5.1.2. Procedimiento de estimación

5.1.2.1. Procedimiento del Equipo 1

5.1.2.1.1. Mtra. Patricia Isabel Romero Mares⁹

Procedimiento de estimación

La proporción de votos a favor del candidato p tanto de la elección presidencial, así como de la elección a la gubernatura de Puebla se estimará del siguiente modo:

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

Donde:

- \hat{R}_p : estimador de la proporción de votos a favor del candidato p
- \hat{Y}_p : estimador del total de votos a favor del candidato p
- \hat{X}_p : estimador del total de votos emitidos
- \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
- \hat{y}_{hp} : estimador del promedio de votos por casilla a favor del candidato p, en el estrato h
- \hat{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 será el siguiente:

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

Donde:

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

⁹ Con la colaboración de la Mtra. Adriana Ducoing Watty

- \hat{X} : total estimado de los votos emitidos
- 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 por casilla en el 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 del partido p se calculará como:

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

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

$$\hat{R}_p - precision , \hat{R}_p + precision .$$

Referencias:

- Cochran, W.G. (1971). Técnicas de muestreo. CECSA, México
- Des Raj (1968). Sampling Theory. McGraw-Hill, Inc. Nueva York
- Lohr, S.L. (2011). Sampling: Design and Analysis, 2nd ed. Brooks/Cole, Boston, USA

5.1.2.2. Procedimiento del Equipo 2

5.1.2.2.1. Dr. Manuel Mendoza Ramírez y Dr. Luis Enrique Nieto Barajas¹⁰

Descripción del método de estimación Bayesiano

Inferencia Bayesiana

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 (De Groot, 2004).

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 y la variabilidad de estas muestras se describen 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, es necesario cuantificar la incertidumbre sobre los parámetros a través de modelos de probabilidad. En este caso los modelos no describen variabilidad alguna sino el estado de conocimiento que, en un momento determinado, tiene el científico sobre el fenómeno bajo estudio. Diferentes modelos pueden reflejar desde un conocimiento relativamente preciso hasta un desconocimiento casi completo sobre el valor del parámetro. En cualquier caso, 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 acerca de 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.

¹⁰ Con la colaboración del Lic. Gian Carlo Di-Luvi Martínez y del Mtro. Carlos Samuel Pérez Pérez

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

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

Esta es una expresión matemática simple; sin embargo, en la práctica, es necesario establecer con detalle las características de este nuevo modelo de probabilidad final, con el propósito de evaluar las posibles inferencias a la luz del nuevo estado de conocimiento. En una variedad de circunstancias, 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).

Notación y definiciones

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 han de producir 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 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 o la lista nominal del país si se trata de la elección federal. Finalmente, la proporción efectiva de votos a favor del candidato j se puede calcular como

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

Modelo y proceso de inferencia

Del total de casillas en el país (o estado, 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 $j=1, \dots, J-1$ y producir las inferencias que interesan a partir de ese modelo de probabilidad. Debido a la forma de esa distribución final, es necesario utilizar métodos de simulación para obtener los resultados. Los detalles y los métodos de simulación se pueden consultar en Mendoza y Nieto-Barajas (2016).

Referencias

- Chen, M.-H., Shao, Q.-M. e Ibrahim, J.G. (2000). *Monte Carlo methods in Bayesian computation*. Springer, New York.
- DeGroot, M.H. (2004). *Optimal statistical decisions*. Wiley, New Jersey.

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

5.1.2.3. Procedimiento del Equipo 3

5.1.2.3.1. Dr. Carlos Hernández Garciadiego¹¹

Estimación de la votación

Para estimar la votación a partir de la muestra que se vaya recibiendo, primero se estima la votación del candidato i obtenida en cada estrato k

$$V_{k,i} = \frac{\sum v_j}{\sum \ln_j} \ln_k ,$$

donde $\sum v_j$ es la suma de los votos obtenidos por el candidato en las casillas que se han recibido hasta el momento, $\sum \ln_j$ es el listado nominal de las casillas recibidas y \ln_k es el listado nominal del estrato k .

De la misma manera se estima la votación total obtenida en el estrato

$$T_{k,i} = \frac{\sum t_j}{\sum \ln_j} \ln_k ,$$

donde $\sum t_j$ es la suma de todos los votos obtenidos en las casillas recibidas, incluyendo los nulos y votos a candidatos no registrados.

Para estimar el porcentaje de votos obtenidos por cada candidato se suman los estimados en cada estrato y se divide entre la suma de los totales de votos estimados

$$\hat{p}_i = \frac{\sum V_{k,i}}{\sum T_{k,i}} .$$

Para obtener los intervalos de confianza de cada candidato, se utilizarán como varianzas poblacionales en cada estrato las varianzas calculadas en la sección anterior.

Se calcula primero la varianza en cada estrato

$$V_k = \frac{s_k^2}{n_k} \left(\frac{N_k - n_k}{N_k} \right) ,$$

donde s_k^2 es la varianza de la muestra de casillas que se ha recibido y la varianza global:

11 Con la colaboración de Mtro. Emiliano Geneyro Squarzon

$$V = \frac{1}{N^2} \sum N_k^2 V_k .$$

Y finalmente, el intervalo de confianza del candidato es

$$\left(\hat{P}_i - 1.96 \sqrt{V}, \hat{P}_i + 1.96 \sqrt{V} \right) .$$

Explicación del procedimiento

Para estimar el porcentaje de votación de cada candidato, lo que se hace grosso modo es lo siguiente:

Se divide el total de casillas en estratos, es decir, en colecciones de casillas que tienen algo en común, como podrían ser los mismos distritos federales, distritos locales, municipios o tipos de casillas. En este caso, consideramos estratos tomando en cuenta los distritos locales.

Se elige aleatoriamente una muestra de casillas de tamaño n seleccionando en cada estrato un número proporcional al número total de casillas en dicho estrato.

Cuando se tiene una remesa de casillas que han llegado, se clasifican de acuerdo con su estrato, y en cada estrato se estima el número total de votos de cada candidato tomando en cuenta cuántas casillas han llegado y cuántas hay en total en el estrato. Esto se hace básicamente, haciendo una regla de tres. Se estima también de manera similar el número total de votos emitidos y se calcula el porcentaje de votos de cada candidato dividiendo el estimado del número de votos de ese candidato entre el estimado del número total de votos.

Después hay que analizar qué tan certera es la estimación anterior. Para ello se utiliza lo que se conoce como varianza, es decir, en cada estrato se analiza qué tanto varían los porcentajes de votos de cada candidato en las casillas de ese estrato y se combinan estas varianzas tomando en cuenta el número de casillas de la muestra que han llegado para obtener un número llamado radio o error estándar. Este radio irá decreciendo conforme vayan llegando más casillas de la muestra.

Finalmente, el intervalo del porcentaje de votos que se espera que obtenga cada candidato se calcula sumando y restando el radio al estimado del porcentaje de votos de dicho candidato.

5.1.2.3.2. Dr. Gabriel Núñez Antonio¹²

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 tanto para la presidencia como para la Gubernatura del estado de Veracruz en las elecciones del 1° de julio de 2018 consideran el diseño de muestreo estratificado definidos en la sección 2.1.1 y 5.2.1.8, respectivamente; con selección de casillas aleatoria sin reemplazo

12 Con la colaboración del Mat. Ernesto Silva Sartorius

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,

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

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 Veracruz de 2016.

5.1.2.3.3. Dr. Carlos Erwin Rodríguez Hernández-Vela¹³

Método de estimación para el Conteo Rápido

La estimación se realizará mediante el estimador de razón combinado y técnicas de re-muestreo (*Bootstrap*).

El Bootstrap es un método estadístico para estimar la distribución de muestreo de un estimador, se obtiene seleccionando B sub-muestras de la muestra original y calculando el estimador de interés con cada sub-muestra. Si se realiza un histograma con los B estimadores generados (en el caso de un estimador de una dimensión), se observaría la aproximación resultante. El número de réplicas B , debe ser suficientemente grande para lograr una buena aproximación. Una vez obtenidas las réplicas Bootstrap es posible calcular los intervalos de confianza.

Existen muchas variantes del Bootstrap, el trabajo de Efron (1979) sentó las bases; sin embargo, esta estrategia ha sido explorada por muchos autores. A continuación, se describe el enfoque del Bootstrap que se implementará en el Conteo Rápido tanto para la elección federal, como para la elección de Gubernatura del estado de Yucatán. En particular se usarán las ideas de Sitter (1992A) y (1992B). El primer artículo describe el método que se implementará el día de la elección, mientras que el segundo describe varias comparaciones, incluso con re-muestreo Jackknife. Estas ideas ya han sido aplicadas con éxito para realizar la estimación de la composición de la cámara de diputados en la elección federal de 2015, así como para la elección extraordinaria de Colima en 2016.

Las ideas de Sitter (1992A) aplicadas a la elección Presidencial y de Gubernatura del estado de Yucatán se describen a continuación. Sean N_h el número total de casillas instaladas en el estrato h y n_h la muestra de casillas seleccionadas (o recuperadas) en el mismo estrato. Entonces, la probabilidad de selección de una casilla en el estrato h está dada por

$$f_h = \frac{n_h}{N_h} .$$

Esbozo del algoritmo

Si $N_h = n_h k_h$ y $k_h = \frac{N_h}{n_h} = 1/f_h$ y $m_h = f_h n_h$ (suponiendo que k_h y m_h son enteros), entonces:

1. De la muestra recuperada del primer estrato se selecciona una sub-muestra SIN reemplazo de m_1 casillas (de las n_1 recuperadas).

¹³ Con la colaboración del Mtro. David Chaffrey Moreno Fernández

2. Se repite el paso 1, k_1 veces, obteniendo una muestra de $n_1 = m_1 k_1$ casillas en total.

3. Se calcularán las siguientes estadísticas:

- ✓ Estimador común del total de votos en el primer estrato para el i -ésimo candidato

$$\hat{Y}_{1i} = N_1 \hat{y}_{1i} .$$

4. Se repiten los pasos 1 a 3 para cada uno de los L estratos restantes.

5. Se calculan

- ✓ El estimador del total de votos a favor del candidato a la Presidencia/Gubernatura del partido i -ésimo en todo el país/estado (sumando sobre todos los estratos)

$$\hat{Y}_i = \sum_{h=1}^L \hat{Y}_{hi} .$$

- ✓ Se estiman los porcentajes mediante la siguiente expresión

$$\hat{p}_i = \frac{\hat{Y}_i}{\sum \hat{Y}_i} .$$

El paso 5 concluye una iteración Bootstrap, el objetivo es realizar B iteraciones con B suficientemente grande. Para el 1 de julio de 2018 y con cada remesa de información que se reciba se tiene pensado realizar $B=5,000$ iteraciones.

Es importante observar que en el paso 1, en el muestreo SIN reemplazo, la probabilidad de selección es la misma que en el diseño de muestreo original

$$f'_h = \frac{m_h}{n_h} = f_h , \text{ esto se repite } k_h \text{ veces de manera independiente para}$$

obtener una muestra de n_h casillas. Esta estrategia asegura obtener estimaciones in-sesgadas para la varianza y para el tercer momento.

En caso de que k_h o m_h no sean enteros se sigue un proceso de aleatorización para alcanzar la consistencia del algoritmo.

El algoritmo anterior es computacionalmente demandante. Sin embargo, las iteraciones del Bootstrap son independientes, así que es posible correr el algoritmo mediante procesamiento paralelo.

Consideraciones adicionales

Existen fórmulas analíticas para calcular la varianza y por lo tanto el intervalo de confianza para el estimador de razón combinado. Sin embargo, siempre se utiliza la aproximación normal para construir el intervalo de confianza y esto puede ocasionar problemas cuando se estiman porcentajes muy pequeños. En estos casos, el intervalo inferior suele ser negativo y se debe truncar el intervalo en el cero. Esta solución, no tiene ningún fundamento técnico. El Bootstrap soluciona este tipo de problemas de manera automática.

Referencias

- Efron, B. (1979). "Bootstrap methods: Another look at the jackknife". The Annals of Statistics, Vol 7. No 1, pp. 1-26.
- Sitter, R. R. (1992A). "A Resampling algorithm for complex survey data". Journal of the American Statistical Association. Vol. 87. No. 419, pp. 755-765.
- Sitter, R. R. (1992B). "Comparing Three Bootstrap Methods for Survey Data".

5.1.2.3.4. Integración de los intervalos de confianza del Equipo 3

El procedimiento para reportar los intervalos será tomando la mediana de los límites inferiores y la mediana de los límites superiores de los intervalos producidos por cada uno de los tres miembros del COTECORA del equipo 3

5.2. Elecciones de Gobernatura y Jefatura de Gobierno

5.2.1. Estratificación y tamaño de muestra

5.2.1.1. Chiapas¹⁴

El objetivo del Conteo Rápido Chiapas 2018, es obtener estimaciones del porcentaje de votos emitidos a favor de los candidatos a la Gobernatura del estado, así como el porcentaje de ciudadanos que acudan a votar. Para tal fin, se desarrollarán diversos procedimientos técnicos que tienen sustento en la Teoría de la Inferencia Estadística.

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 a la Teoría de Muestreo para Poblaciones Finitas, con el propósito de establecer un diseño eficiente de la muestra que proveerá la información para la estimación, y a los Métodos de Estimación Paramétrica para definir el mecanismo que habrá de aplicarse a la información muestral para obtener las estimaciones.

Esquema de muestreo

En este caso, la Población Estadística está formada por la totalidad de las casillas que se instalarán en la entidad, el día de la Jornada Electoral. En esa población es de interés estimar, por una parte, el número total de votos que se recibirán para la elección de Gobernatura y consecuentemente, el porcentaje de ciudadanos, en la lista nominal, que acudirán a ejercer ese derecho. Además, interesa también estimar el porcentaje de votos emitidos a favor de cada candidato a la Gobernatura.

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 RE, debe abarcar la mayor dispersión geográfica posible. Para tal fin se utilizará un esquema de diseño muestral estratificado, esto significa que la entidad se dividirá en regiones o 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.

¹⁴ Elaborado por el Dr. Manuel Mendoza Ramírez con la colaboración del Lic. Gian Carlo Di-Luvi Martínez

Este diseño muestral, además de responder al requerimiento de la normativa electoral, es muy común en la práctica del muestreo y se sabe que produce estimaciones más precisas que las correspondientes al caso en que la muestra se seleccionase completamente al azar y sin reemplazo del conjunto de todas las casillas instaladas en la entidad sin ninguna restricción geográfica. Una vez que se ha optado por un diseño estratificado, como el que se ha indicado, es necesario establecer, por una parte, el criterio con que se dividirá la entidad en estratos y, en segundo lugar, el tamaño total de la muestra.

La determinación del criterio de estratificación se lleva a efecto tomando en cuenta que un diseño estatificado produce mejores resultados en la medida en que las unidades en 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 un nivel aceptable de confiabilidad y precisión. En el caso del Conteo Rápido de Chiapas, el nivel de confiabilidad, como lo establece el RE del INE, 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, para cada una de las propuestas de estratificación, considerando este nivel de confiabilidad.

Criterios de estratificación

Para el Conteo Rápido es natural considerar como posible estratificación la división de la entidad en sus distritos electorales. De hecho, en ejercicios previos de Conteo Rápido federal, se han definido como estratos, en una primera instancia, a los 300 distritos electorales federales. Adicionalmente, se ha considerado la división de esos distritos en sus componentes urbana y no urbana, utilizando para tal fin la información del tipo de sección. Así se procedió en los Conteos Rápidos Presidenciales de 2006 y 2012. Para la elección de Gubernatura en una entidad, se pueden considerar esas dos alternativas para la estratificación: solamente distrito electoral federal; o bien, distrito electoral federal combinado con tipo de sección. Es conveniente mencionar que también existen los distritos electorales locales, que en número pueden superar a los distritos electorales federales. Además, otro posible criterio puede ser el municipio o, más genéricamente, el tipo de municipio (de acuerdo con su cantidad de población, por ejemplo).

En el caso de Chiapas se consideraron distintas propuestas de estratificación a partir de estos criterios: distrito electoral federal (DF), distrito electoral local (DL), tipo de sección (TS) y tipo de municipio (TM). En total, se examinaron 26

alternativas, cada una de las cuales resulta de una combinación de estos criterios.

Evaluación de las alternativas de estratificación

Una vez establecidas las estratificaciones posibles, para seleccionar la alternativa óptima se utilizaron los resultados de la elección local para gobernador de 2012. 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.

De esta manera, para una estratificación fija y un tamaño de muestra también establecido M , 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 con el esquema de estratificación.
2. Para el i -ésimo candidato a la Gubernatura, se calcula la estimación del porcentaje de votación a su favor \hat{p}_i , a partir de esa muestra.
3. Considerando el valor verdadero del porcentaje por estimar p_i , se calcula el error de estimación observado $p_i - \hat{p}_i$.

Los pasos 1 a 3 se repiten un número K de veces, a fin de poder establecer el valor de la constante ϵ , tal que

$$P(|p_i - \hat{p}_i| \leq \epsilon) = 0.95 .$$

El valor de ϵ es el error de estimación máximo, correspondiente al nivel de confiabilidad 0.95, asociado a una muestra de tamaño M , seleccionada de acuerdo con la estratificación considerada.

Este mecanismo se puede aplicar para el porcentaje de votos de cada uno de los candidatos a la Gubernatura. A partir del comportamiento que sigue este error, como función del tamaño de muestra para las diferentes propuestas de estratificación, se determina la alternativa óptima y una vez con esta decisión, se selecciona el tamaño de muestra que, para ese esquema, asegure un error razonable, en este caso en torno a 1%.

Para efectos de este procedimiento, la estimación \hat{p} , se calcula con la expresión convencional para los estimadores de razón. Es decir,

$$\hat{p}_i = \frac{\sum_{j=1}^H N_j \bar{y}_{ji}}{\sum_{j=1}^H \sum_{i=1}^C N_j \bar{y}_{ji}},$$

donde \bar{y}_{ji} es el promedio de votos en el estrato j a favor del candidato i , N_j es el tamaño del estrato j , H es el número de estratos y C es el número de categorías en que se clasifican los votos emitidos (una por cada candidato a la Gubernatura, otra para los candidatos no registrados y una más para los votos nulos).

Para establecer un criterio conservador, las decisiones de diseño se toman a partir de los resultados para la categoría (candidato) donde se producen los errores de estimación más grandes.

Resultados de la elección para gobernador en 2012

En la elección de ese año contendieron cuatro candidatos. (MVC) Manuel Velasco Coello, por la Coalición Compromiso por Chiapas, formada por los partidos Verde Ecologista de México, Revolucionario Institucional y Nueva Alianza; (MEOL) María Elena Orantes López, por la Coalición Movimiento Progresista por Chiapas, integrada con los partidos de la Revolución Democrática, del Trabajo y Movimiento Ciudadano; (ENG) Emmanuel Nivón González por el partido Acción Nacional y finalmente, (MBG) Marcela Bonilla Grajales por el partido Orgullo por Chiapas.

Estructura del estudio de estratificación

En una primera etapa, se consideraron los siguientes criterios para definir la estratificación.

1. Distrito electoral federal (DF).
2. Tipo de sección (TS).
3. Distrito electoral local (DL).
4. Tamaño de municipio 1 (TM1).
5. Tamaño de municipio 2 (TM2).
6. Tamaño de municipio 3 (TM3).
7. Tamaño de municipio 4 (TM4).

Todas con la información vigente al momento de la elección de 2012. Como punto de partida, se consideraron dos estratificaciones básicas: B1, la que se produce con el criterio DF, de forma que la entidad simplemente se divide en los doce distritos electorales federales vigentes en 2012; en segundo lugar, B2 para la que se aplicó el criterio TS sobre la primera estratificación, separando las componentes urbana, rural y mixta de cada distrito electoral federal, en el entendido de que no todos los distritos cuentan con las tres componentes.

Cada una de esas dos estratificaciones básicas se refinó posteriormente, dividiendo las casillas en cada estrato de acuerdo con el distrito electoral local (DL) al que pertenecen o, de acuerdo con el tamaño del municipio (TM#) que les corresponde o, respecto a estos dos criterios. Al respecto, es oportuno señalar que, en lo que toca al tamaño de los municipios, cada uno de estos criterios clasifica a los municipios en *grandes* o *pequeños*, dependiendo de la población con que cuenta. Así, un municipio es considerado *grande* si se encuentra entre los de mayor población que sumados acumulan 20% (TM1), 25% (TM2), 30% (TM3) o 50% (TM4) de los habitantes de la entidad.

En consecuencia, inicialmente se evaluaron 20 alternativas de estratificación, cuya estructura se describe en la Tabla 1, y posteriormente, a partir del propio análisis, se incluyeron otras posibilidades que se describen más adelante en este documento.

Tabla1. Estructura de las estratificaciones 1-20.

Estratificación	Estructura
1	B1
2	B1 x DL
3	B1 x TM1
4	B1 x TM2
5	B1 x TM3
6	B1 x TM4
7	B1 x DL x TM1
8	B1 x DL x TM2
9	B1 x DL x TM3
10	B1 x DL x TM4
11	B2
12	B2 x DL
13	B2 x TM1
14	B2 x TM2
15	B2 x TM3
16	B2 x TM4
17	B2 x DL x TM1
18	B2 x DL x TM2
19	B2 x DL x TM3
20	B2 x DL x TM4

Una consideración adicional en la definición de las estratificaciones, se refiere al tamaño de los estratos. No es conveniente que los estratos tengan un número de casillas muy pequeño ya que, en caso de que se reciba una muestra incompleta, se corre el riesgo de que se pierdan estratos completos.

Para evitar casos extremos, provisionalmente para este ejercicio, se fijó un umbral de 30 casillas como el tamaño mínimo de los estratos. Por otra parte, tomando en cuenta que en los Conteos Rápidos para las elecciones federales de 2006 y 2012 los tamaños de la muestra asignados al estado de Chiapas se encontraban entre 200 y 300 casillas y en virtud de que para la estimación de los resultados locales será necesario contar con al menos ese número de casillas, el estudio consideró tamaños de muestra entre 300 y 700 en múltiplos de 100. Finalmente, para determinar el error máximo de estimación, se seleccionaron $K=2000$ muestras de cada tamaño para cada esquema de estratificación.

Resultados del estudio inicial de estratificación

Al calcular los errores máximos de estimación para la votación de cada uno de los candidatos en la elección de 2012, se observó que, sistemáticamente, los niveles más altos de estos errores correspondieron al candidato MVC. En consecuencia, el siguiente análisis se concentra en los errores de esa estimación.

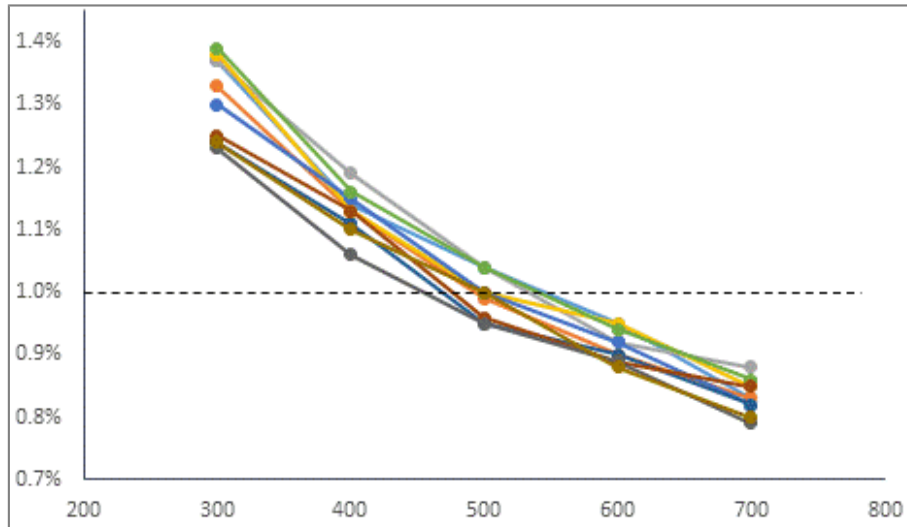
En la Tabla 2 se reportan los errores de estimación y se puede observar que los mismos se encuentran alrededor de 1.00%, para todas las estratificaciones, cuando el tamaño de muestra es de 500 casillas. De hecho, es ese el tamaño de muestra para el cual las estratificaciones empiezan a producir errores por debajo de ese nivel. Como complemento, las Gráficas 1 y 2 exhiben esta información, la primera para las estratificaciones 1 a 10, y en el segundo caso para las estratificaciones 11 a 20.

Tabla 2. Errores máximos de estimación para las estratificaciones 1-20.

Estratificación	Tamaño de Muestra				
	300	400	500	600	700
1	1.37%	1.14%	1.04%	0.95%	0.83%
2	1.33%	1.13%	0.99%	0.90%	0.83%
3	1.37%	1.19%	1.04%	0.92%	0.88%
4	1.38%	1.13%	1.00%	0.95%	0.85%
5	1.30%	1.15%	1.00%	0.92%	0.82%
6	1.39%	1.16%	1.04%	0.94%	0.86%
7	1.24%	1.11%	0.95%	0.90%	0.82%
8	1.25%	1.13%	0.96%	0.89%	0.85%
9	1.23%	1.06%	0.95%	0.89%	0.79%
10	1.24%	1.10%	1.00%	0.88%	0.80%
11	1.31%	1.12%	0.98%	0.93%	0.83%
12	1.30%	1.12%	1.00%	0.88%	0.78%
13	1.35%	1.15%	1.01%	0.95%	0.83%
14	1.37%	1.11%	1.00%	0.92%	0.84%
15	1.31%	1.09%	0.99%	0.87%	0.81%
16	1.36%	1.09%	0.99%	0.90%	0.82%
17	1.30%	1.08%	0.98%	0.88%	0.83%
18	1.28%	1.09%	1.00%	0.90%	0.79%
19	1.29%	1.07%	0.97%	0.88%	0.79%
20	1.27%	1.07%	0.95%	0.85%	0.76%

Un elemento adicional tanto para definir la estratificación, como para establecer el tamaño de muestra, es la carga de trabajo que la muestra resultante representa para el personal en campo, los CAE's. La muestra debería tener asociada una proporción de CAE's a cargo de reportar más de una casilla sensiblemente baja. El COTECORA decidió que la proporción de CAE's con una casilla no fuese menor a 80%. Para el caso que nos ocupa, los porcentajes aparecen en la Tabla 3.

Gráfica 1. Error máximo; estratificaciones 1-10.



Gráfica 2. Error máximo; estratificaciones 11-20.

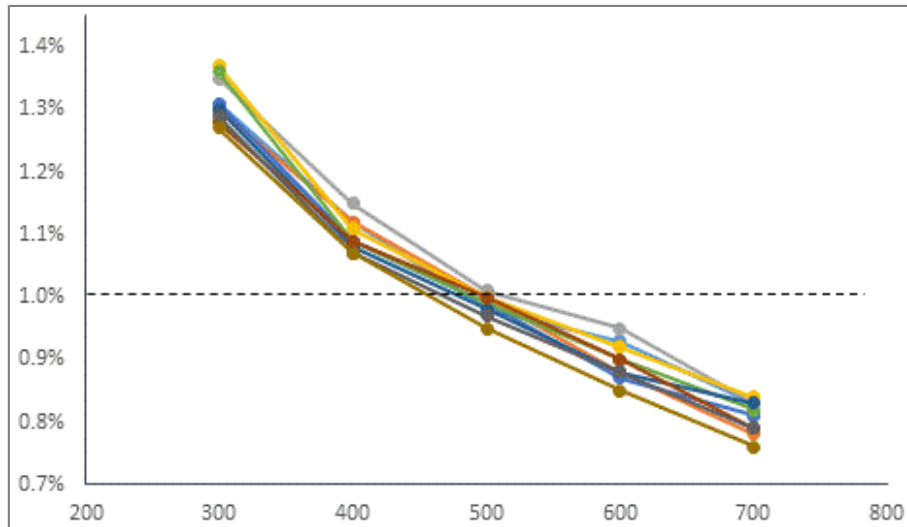


Tabla 3. Porcentaje de CAE's con una casilla; estratificaciones 1-20.

Estratificación	Tamaño de Muestra				
	300	400	500	600	700
1	91.69%	88.79%	85.97%	83.35%	80.69%
2	92.03%	89.15%	86.45%	83.71%	80.99%
3	91.76%	88.82%	86.14%	83.43%	80.76%
4	91.83%	88.88%	86.18%	83.45%	80.78%
5	91.89%	88.91%	86.19%	83.43%	80.76%
6	91.98%	89.11%	86.24%	83.51%	80.79%
7	92.07%	89.25%	86.45%	83.71%	81.01%
8	92.09%	89.21%	86.50%	83.77%	81.05%
9	92.18%	89.26%	86.53%	83.74%	81.04%
10	92.22%	89.36%	86.61%	83.88%	81.13%
11	92.26%	89.19%	86.55%	83.78%	81.04%
12	92.25%	89.42%	86.63%	83.91%	81.07%
13	92.48%	89.40%	86.58%	83.93%	81.09%
14	92.43%	89.40%	86.70%	83.95%	81.12%
15	92.57%	89.45%	86.73%	84.05%	81.11%
16	92.74%	89.80%	86.99%	84.20%	81.29%
17	92.35%	89.45%	86.72%	83.94%	81.15%
18	92.39%	89.44%	86.63%	83.90%	81.11%
19	92.39%	89.49%	86.65%	83.93%	81.19%
20	92.56%	89.68%	86.80%	84.14%	81.27%

De esta forma, como resultado de una primera fase, se estableció que las estratificaciones 7, 9 y 20 eran alternativas razonables en tanto que sus errores máximos eran los más bajos en el estudio y no solamente alrededor de 1% sino incluso inferiores. Por otra parte, también se determinó que un tamaño de muestra de 500 casillas, en cualquiera de estos casos, resultaba aceptable.

Ese tamaño, $M=500$ casillas, además, conduce a un porcentaje de CAE's con sólo una casilla que supera el 85%. En forma complementaria, y puesto que para el Conteo Rápido de la elección Presidencial, Chiapas tendrá asignado un tamaño de muestra no mayor que 300, se consideró que el incremento a 500 era razonable.

Ahora bien, las estratificaciones 7, 9 y 20 están definidas, las tres, como el cruce de distrito local y un tamaño de municipio, aplicado a una de las estratificaciones básicas B1 ó B2 (Tabla 4). En la práctica, el hecho de utilizar todos estos criterios simultáneamente, específicamente en el caso de la estratificación 20, conduce a un número de estratos grande con el consecuente riesgo de que, de no llegar la muestra completa, se puedan perder estratos completos.

Tabla 4. Estructura de las estratificaciones 1-20.

Estratificación	Estructura	No. Estratos
7	DF x DL x TM1	31
9	DF x DL x TM3	34
20	(DF x TS) x DL x TM4	49

Como consecuencia de estas valoraciones, y tomando en cuenta que tanto para B1 como para B2, el refinamiento producto del cruce de distrito local y Tipo de municipio es un elemento común, se planteó la posibilidad de analizar otras estratificaciones con una estructura más simple con sólo dos criterios: distrito local y alguno de los criterios de tamaño de municipio. El objetivo era identificar una alternativa, con esa estructura simplificada, que produjese un error alrededor de 1% y un porcentaje de CAE's con una casilla superior al 85%. Además se espera contar con menos estratos que con la estratificación 20.

Para tal fin, se consideraron 5 posibilidades de estratificación utilizando los distritos locales y las diferentes especificaciones del tamaño de municipio, tal como se muestra en la Tabla 5.

Tabla 5. Estructura de las estratificaciones 21-25.

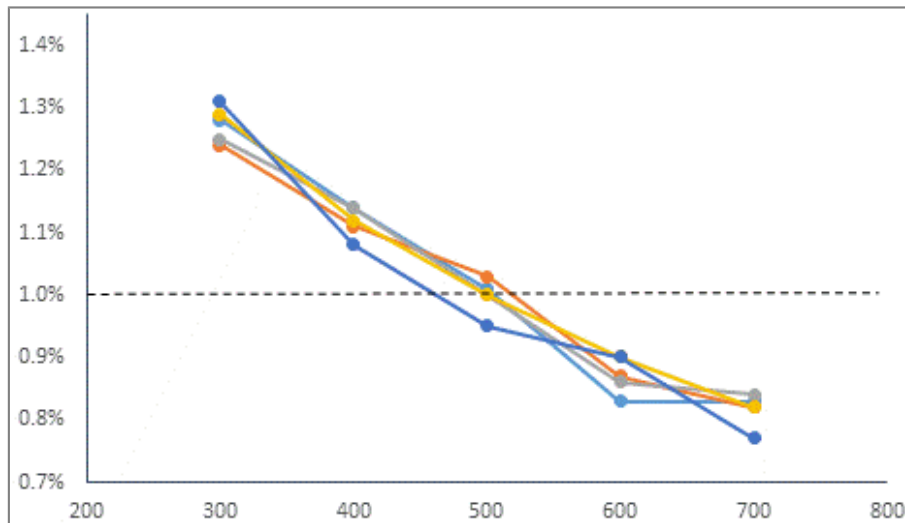
Estratificación	Estructura
21	DL
22	DL x TM1
23	DL x TM2
24	DL x TM3
25	DL x TM4

Los resultados, en términos de los errores máximos, aparecen en la Tabla 6 y en la Gráfica 3. Como se puede observar, la estratificación 25, definida por refinamiento de los distritos electorales locales por tipo de municipio, cuando los municipios grandes se definen como los más habitados que acumulan el 50% de la población, resulta comparable con las alternativas 7, 9 y 20.

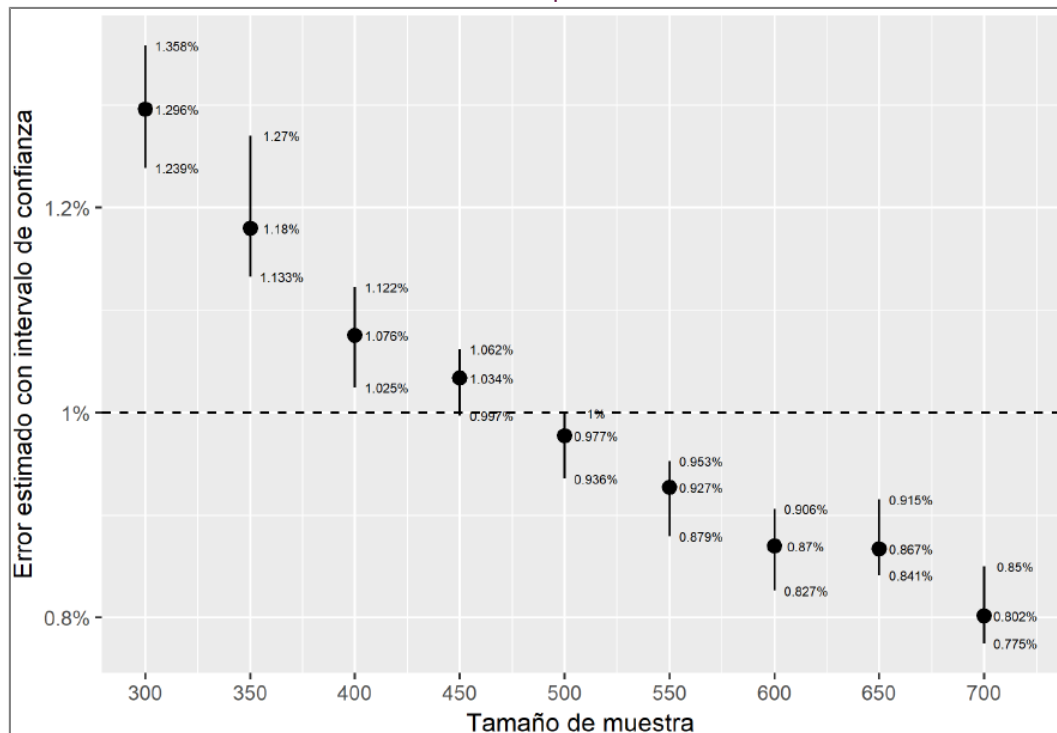
Tabla 6. Errores máximos de estimación para las estratificaciones 20-25.

Estratificación	Tamaño de Muestra				
	300	400	500	600	700
21	1.28%	1.14%	1.01%	0.83%	0.83%
22	1.24%	1.11%	1.03%	0.87%	0.82%
23	1.25%	1.14%	1.00%	0.86%	0.84%
24	1.29%	1.12%	1.00%	0.90%	0.82%
25	1.31%	1.08%	0.95%	0.90%	0.77%

Gráfica 3. Error máximo; estratificaciones 21-25.



Gráfica 4. Errores máximos para la estratificación 25.



Para contar con mayor precisión en la determinación de los errores para esta estratificación, se incrementó el número de muestras para el cálculo. Los resultados, con $K=4000$ muestras, aparecen en la Gráfica 4. Esta evidencia confirma que, con alta confiabilidad, el error máximo para una

muestra de 500 casillas no solo se ubica alrededor del 1% sino que es inferior a ese valor.

Finalmente, en la Tabla 7 se puede comprobar que la estratificación 25 (al igual que las otras cuatro nuevas alternativas) satisface la condición de que más del 85% de los CAE's tiene asignada sólo una casilla. De esta forma, como resultado de esta primera fase del análisis, se seleccionó *en forma preliminar* la alternativa 25, que con dos criterios divide la entidad en 35 estratos, una reducción apreciable, si se compara con la estratificación 20, y equiparable a las estratificaciones 7 y 9 que requerían tres criterios.

Tabla 7. Porcentaje de CAE's con una casilla; estratificaciones 20-25.

Estratificación	Tamaño de Muestra				
	300	400	500	600	700
21	91.95%	89.01%	86.40%	83.49%	80.93%
22	95.91%	89.09%	86.39%	83.55%	80.96%
23	92.02%	89.09%	86.41%	83.62%	80.95%
24	92.13%	89.13%	86.39%	83.54%	80.97%
25	92.17%	89.17%	86.53%	83.72%	81.08%

Segunda fase del estudio de estratificación

Como se indicó, la información que se consideró en la etapa inicial de este estudio fue la que se refiere a la elección local para gobernador de 2012. Asimismo, la definición de distritos electorales federales, distritos electorales locales y municipios es la vigente en el momento de esa elección. Tomando en cuenta que a la fecha de este estudio (2018) se han producido una redistribución federal, una redistribución local y han ocurrido cambios en los municipios de la entidad, se planteó el análisis de los mismos resultados electorales, pero considerando la versión actualizada de la estratificación 25 con los nuevos distritos locales y los nuevos municipios.

En el proceso que se siguió para obtener esos resultados, se identificaron algunas ambigüedades derivadas de la definición de nuevos municipios creados en la entidad a partir de 2012. Tras valorar el impacto que estas podrían tener en la definición de los estratos, se optó por abandonar los criterios relacionados con el tamaño del municipio.

Por otra parte, para el Conteo Rápido federal 2018 se acordó que los estratos estarán definidos como los 300 distritos electorales federales. Así, para la estratificación local se decidió utilizar solo dos criterios (actualizados) y que el resultado constituyera una refinación de la estratificación que se utilizará para el Conteo Rápido federal. De esta forma, se optó por una alternativa que partiera de los 13 distritos electorales federales (DF) y que se refinara

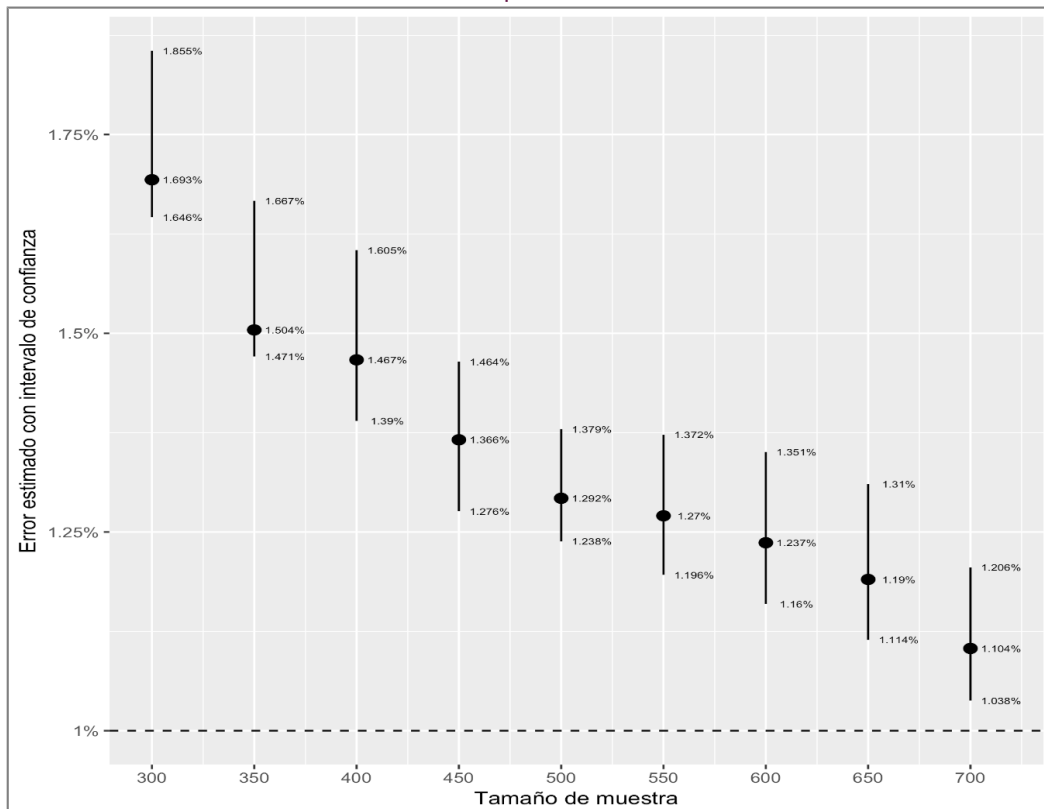
dividiendo a las casillas de cada uno de esos distritos de acuerdo con el distrito electoral local (DL) al que perteneciera. Al respecto, es conveniente señalar que si bien existen 24 nuevos distritos electorales locales, como consecuencia de los factores que en ambas distritaciones se siguen, es de esperarse que el refinamiento no conduzca a un número excesivo de estratos. Esta idea se reforzó, con la experiencia de la primera fase, al establecer que un tamaño mínimo de los estratos de 100 casillas.

La nueva propuesta de estratificación (26) se sometió al mismo tipo de análisis que las anteriores y, por lo que se refiere a los errores máximos, los resultados se presentan en la Tabla 8 y en la Gráfica 5.

Tabla 8. Errores máximos de estimación para la estratificación 26.

Estratificación	Tamaño de Muestra				
	300	400	500	600	700
26	1.69%	1.47%	1.29%	1.24%	1.10%

Gráfica 5. Errores máximos para la estratificación 26 (Final).



Como se puede apreciar, el patrón de los errores es similar al de las propuestas previas si bien para un tamaño de muestra 500, alcanza el valor de 1.29% y con una confiabilidad del 95% se puede asegurar que se encuentra entre 1.24% y 1.38%. Este nivel de error es cercano al 1% si bien resulta superior al que se obtiene con las propuestas previas. Al respecto es oportuno señalar que cuando esta estrategia de estratificación se analiza con la geografía electoral de 2012 (estratificación 2 en la Tabla 1) el error es de 0.99%.

Por otra parte, en lo que respecta al porcentaje de CAE's con sólo una casilla por reportar, como se puede observar en la Tabla 9, es mayor al 85%.

Tabla 9. Porcentaje de CAE's con una casilla para la estratificación 26 (Final).

Estratificación	Tamaño de Muestra				
	300	400	500	600	700
26	91.78%	88.93%	86.18%	83.34%	80.71%

Resultado final del estudio de estratificación

La propuesta definitiva de estratificación para el diseño muestral del Censo Rápido 2018 que se llevará a efecto con motivo de la elección de la Gubernatura en el estado de Chiapas es la propuesta 26, que utiliza definición de distritos electorales federales (DF) y distritos electorales locales (DL) en la entidad. En relación con el tamaño de muestra, éste queda determinado en 500 casillas que se distribuirán entre los 15 estratos resultantes, con una asignación proporcional al tamaño del estrato. En la Tabla 10 se presenta la estructura de esta alternativa de estratificación. Como puede observarse, esencialmente coincide con la división de la entidad en distritos electorales federales con la salvedad de que el distrito 10 se subdivide en tres estratos que corresponden a sus intersecciones con los distritos electorales locales 14, 23 y 15.

Tabla 10. Estructura de la estratificación 26 (Final).

Estrato	Estructura
1	DF1
2	DF2
3	DF3
4	DF4
5	DF5
6	DF6
7	DF7
8	DF8
9	DF9
10	DF10 x DL14
11	DF10 x DL23
12	DF10 x DL15
13	DF11
14	DF12
15	DF13

5.2.1.2. Ciudad de México¹⁵

Tamaño de muestra

Para calcular el tamaño de muestra, se tomó como referencia los datos de la elección de diputados federales de 2015, agrupando los partidos contendientes según las coaliciones que se formaron para la elección de 2018.

Se consideraron varias estratificaciones posibles y la que mejor se adapta a las condiciones de la Ciudad de México es considerar como estratos a los distritos locales.

Para encontrar el tamaño de muestra necesario para obtener un error máximo e , con un nivel de confianza de 95%, para cada estrato se considera su peso

$$w_i = \frac{N_i}{N} .$$

Donde N es el número total de casillas y N_i es el número de casillas en el estrato i .

Por candidato se calcula su varianza σ_i^2 en el estrato i considerando la elección federal de 2015 como si fuera elección del poder ejecutivo, y agrupando los partidos de acuerdo con las alianzas que habrá en 2018.

El tamaño de la muestra requerida está dado por

$$n = \frac{\sum \frac{N_i^2 \sigma_i^2}{w_i}}{\frac{e^2 N^2}{1.96^2} + \sum N_i \sigma_i^2} .$$

Para obtener un error de $e=0.5\%$ el tamaño obtenido por la fórmula anterior es 745 casillas.

Estratificación

Se utilizarán como estratos los distritos locales de la Ciudad de México, ya que con esta estratificación se obtiene un tamaño de muestra muy razonable para poder obtener un error de $e=0.5\%$ con una confianza de 95%.

¹⁵ Elaborado por el Dr. Carlos Hernández Gaciadiego con la colaboración del Mtro. Emiliano Geneyro Squarzon

Ideas básicas del muestreo probabilístico aplicadas al Conteo Rápido de la Ciudad de México

El objetivo es estimar el porcentaje, p , de votos en favor de un candidato a la Jefatura de Gobierno de la Ciudad de México.

Se selecciona una muestra aleatoria de n casillas de un total de N casillas y con la información recabada se calcula \hat{p} .

Mediante el muestreo probabilístico es posible definir estrategias de selección y tamaños de muestra, n , para asegurar que:

$$|p - \hat{p}| \leq \epsilon,$$

con un 95% de confianza.

Estrategias de selección

- Muestreo aleatorio simple sin reemplazo (y muestreo aleatorio simple con reemplazo).
- Muestreo estratificado.
- Muestreo por conglomerados (poli-etápicas).

Se tomaron como referencia las elecciones:

- Jefe de Gobierno de la Ciudad de México 2012.
- Diputados 2015, restringido a la Ciudad de México.
 - Se asumió la misma elección para los distritos federales de la Ciudad de México
 - $FRENTE = PAN + PRD + 0.5 * C_PRD_PT + MC$
 - $PRI = PRI + 0.5 * C_PRI_PVEM$
 - $JHH = MORENA + PT + 0.5 * C_PRD_PT + PS$

En cada caso se probaron las estratificaciones:

1. Tipo sección (TS), urbana y no urbana
2. Distritos federales (DF), 24

3. Distritos locales (DL), 33
4. DF ∩ TS, 31
5. DL ∩ TS, 36
6. DF ∩ DL, 64
7. DF ∩ DL ∩ TS, 69

Cada una con distintos tamaños de muestra

Resultados para la elección de diputados de 2015 (Precisiones)

n	MASSR	TS	DF	DL	DF ∩ TS	DL ∩ TS	DF ∩ DL	DF ∩ DL ∩ TS
400	0.835	0.834	0.703	0.694	0.683	0.694	0.655	0.655
425	0.809	0.808	0.677	0.675	0.667	0.670	0.633	0.631
450	0.785	0.785	0.677	0.652	0.645	0.650	0.614	0.612
475	0.764	0.763	0.647	0.635	0.626	0.634	0.596	0.595
500	0.744	0.743	0.636	0.617	0.609	0.615	0.581	0.579
525	0.725	0.724	0.618	0.603	0.592	0.601	0.567	0.567
550	0.707	0.707	0.588	0.587	0.579	0.587	0.554	0.553
575	0.691	0.691	0.581	0.574	0.566	0.571	0.542	0.542
600	0.676	0.676	0.565	0.561	0.556	0.559	0.530	0.529
625	0.662	0.661	0.564	0.550	0.543	0.547	0.519	0.518
650	0.648	0.648	0.559	0.539	0.532	0.538	0.509	0.508
675	0.635	0.635	0.533	0.528	0.521	0.527	0.495	0.495
700	0.623	0.623	0.537	0.518	0.512	0.516	0.490	0.490
725	0.612	0.611	0.522	0.507	0.502	0.506	0.480	0.479
750	0.601	0.600	0.519	0.500	0.494	0.497	0.470	0.470
775	0.590	0.590	0.508	0.491	0.484	0.489	0.463	0.463

*Precisiones para la coalición con mayor varianza: FRENTE

No respuesta basada en el Conteo Rápido para la elección de diputados en 2015

Para obtener el tamaño de muestra total y que después de la no respuesta se alcance el tamaño de muestra necesario es

$$(n_{Tot})(TR) = n_{\epsilon} \rightarrow n_{Tot} = \frac{n_{\epsilon}}{TR}$$

Distrito Local	Nh	nh	% respuesta 2015	Afijación de la no respuesta	nh total	Distrito Local	Nh	nh	% respuesta 2015	Afijación de la no respuesta	nh total
1	349	19	100%	0	19	16	412	23	76%	7	30
2	609	34	63%	20	54	17	279	15	80%	4	19
3	402	22	63%	13	35	18	418	23	52%	21	44
4	450	25	80%	6	31	19	311	17	64%	10	27
5	454	25	72%	10	35	20	378	21	67%	10	31
6	399	22	43%	29	51	21	434	24	85%	4	28
7	300	17	60%	11	28	22	275	16	83%	3	19
8	297	16	67%	8	24	23	450	25	78%	7	32
9	390	22	85%	4	26	24	439	24	59%	17	41
10	428	24	61%	15	39	25	334	18	45%	22	40
11	441	24	61%	15	39	26	480	27	79%	7	34
12	426	24	83%	5	29	27	320	18	82%	4	22
13	370	20	50%	20	40	28	325	18	83%	4	22
14	341	19	29%	47	66	29	268	15	47%	17	32
15	437	24	58%	17	41	30	383	21	64%	12	33
						31	324	18	54%	15	33
						32	396	22	59%	15	37
						33	331	18	67%	9	27
TOTAL							12,650	700		408	1,108

Presión sobre los CAE's

- Diseño de muestreo estratificado, por distrito local.
- Tamaño de muestra total, 1,108 casillas.
- En el año 2015, en la Ciudad de México, se contó con 2,357 CAE's.
- Con el diseño de muestra definido anteriormente:
 - Aproximadamente 915 CAE's apoyarían al Conteo Rápido (1,442 no participarían).
 - De los 915 CAE's que participarían, a 742 (81.1%) les tocaría reportar sólo 1 casilla

Estimación del número de casillas por CAE

Número de casillas	Número de CAES	Muestra
1	742	742
2	154	308
3	18	54
4	1	4
Total	915	1,108

En resumen: Diseño muestral para la estimación en la Ciudad de México

- Estratificado por **distrito local**
- Tamaño de **muestra total, n = 1,108 casillas** (12,650 totales en 2015)
- Al repartir proporcionalmente nos queda que
 - min. casillas x estrato= 19
 - max. casillas x estrato= 66
 - promedio casillas x estrato= 34
- Si se observa una no respuesta similar a la del Conteo Rápido de 2015 para la elección de diputados, sólo en la Ciudad de México, se recibirá un tamaño de **muestra efectivo de 700 casillas**
- Con el tamaño de muestra efectivo se alcanzaría una **precisión de 0.51% en la estimación**
- Participarán (aproximadamente) **915 CAE's** y de estos el **81.1% apoyará con máximo 1 casilla**

5.2.1.3. Guanajuato¹⁶

Ideas básicas del muestreo probabilístico

El objetivo del Conteo Rápido es estimar la proporción de votos p en favor de un candidato. Para cumplir esto, se selecciona una muestra aleatoria de n casillas de un total de N y con la información recuperada se calcula el estimador \hat{p} .

Mediante el muestreo probabilístico es posible definir estrategias de selección, tamaño de muestra (n) y estimadores, para asegurar que la diferencia entre el estimador \hat{p} y la verdadera proporción p , no difiera en valor absoluto en más de una cierta cantidad d , con un 95% de confianza. A la cantidad d se le conoce como precisión.

Para el siguiente ejercicio utilizaremos el estimador de razón, definido como:

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

donde, para cada estrato h , n_h es el tamaño de la muestra, N_h es el número total de casillas en el estrato, L es el número total de estratos, J 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 .

Consideraciones para definir el tamaño de muestra y diseño de muestreo

El diseño muestral en Guanajuato será estratificado, donde al interior de cada estrato se seleccionarán casillas mediante un muestreo aleatorio simple sin reemplazo. Para definir los estratos a emplear, se consideraron cinco posibles combinaciones de las siguientes variables:

- Distrito federal
- Distrito local
- Tipo de sección

¹⁶ Elaborado por la Dra. Michelle Anzarut Chacalo con la colaboración de la Mtra. Maria Teresa Ortiz Mancera

Además, como punto comparativo, se consideró un muestreo aleatorio simple sin estratificación. Para cada combinación, se evaluaron los siguientes criterios:

- **Tabla de estratos:** Indica el número total de estratos, el promedio y mediana del número de casillas dentro de cada estrato y la cantidad de estratos con menos de 50 casillas.
- **Tabla de precisiones:** Precisiones con distintos tamaños de muestra calculadas con el estimador de razón. Los tamaños de muestra considerados fueron 50, 100, 200, 300, 358, 500, 600, 700. Usando asignación proporcional al tamaño de cada estrato en la elección federal, se obtuvo que a Guanajuato le tocan 358 casillas. Por esta razón, se les dio preferencia a tamaños de muestra cercanos 358 casillas.
- **Tabla de CAE's:** Consiste en un estudio de la distribución de casillas por CAE. Las estadísticas se calculan a lo largo de 1,000 replicaciones de la muestra, para cada tamaño de muestra propuesto, se calcula en promedio cuántas casillas reportaría cada CAE, se calcula también la mediana, el porcentaje de CAE's encargados de más de una casilla y el porcentaje de CAE's encargados de más de dos casillas.

Entre más variables se usan en la estratificación, mayor es la precisión. Sucede lo mismo con el tamaño de muestra, entre mayor es la muestra, mayor es la precisión. Sin embargo, utilizar demasiadas variables en la estratificación resulta en algunos estratos con pocas casillas y al faltar información de algún estrato se complican los procesos de estimación. Así mismo, considerar muestras demasiado grandes, conlleva un mayor porcentaje de CAE's con más de una casilla a reportar, esto dificulta el trabajo de los CAE's y puede resultar en que no reporten alguna de las casillas. Por esta razón, el diseño muestral evalúa en conjunto todos estos criterios, con el objetivo de encontrar un balance que maximice la precisión cuidando no rebasar la capacidad operativa de campo.

Para el análisis, PRI representa la coalición del PRI con PVEM, PAN representa la coalición del PAN con PANAL y Otros incluye, PT, MC, votos nulos y no registrados.

Resultados

1. Muestreo aleatorio simple

Tabla de precisiones:

PARTIDO	50	100	200	300	358	500	600	700
Otros	1.37	0.94	0.65	0.51	0.45	0.39	0.36	0.33
PAN	2.85	1.98	1.37	1.11	1.04	0.79	0.77	0.73
PRD	1.61	1.14	0.76	0.65	0.58	0.47	0.44	0.39
PRI	3.02	2.24	1.53	1.2	1.17	0.86	0.83	0.77

Tabla de CAE's:

NÚMERO DE CASILLAS	MEDIA	MEDIANA	% MÁS DE 1 CASILLA:	% MÁS DE 2 CASILLAS
50	1	1	2	0
100	1	1	3	0
200	1.1	1	6	0
300	1.1	1	10	1
358	1.1	1	11	1
500	1.2	1	16	2
600	1.2	1	19	2
700	1.2	1	22	3

2. Estratificación por distrito federal.

TABLA ESTRATOS:	
Número de estratos	14
Media número casillas	482
Mediana número casillas	480
Número de estratos < 50	0

Tabla de precisiones:

PARTIDO	50	100	200	300	358	500	600	700
---------	----	-----	-----	-----	-----	-----	-----	-----

Otros	0.97	0.92	0.56	0.47	0.45	0.33	0.32	0.26
PAN	2.74	2.11	1.2	1.12	1.07	0.92	0.65	0.69
PRD	1.35	0.88	0.65	0.54	0.46	0.4	0.36	0.35
PRI	2.8	2.01	1.28	1.14	1.02	0.93	0.65	0.65

Tabla de CAE's:

NÚMERO DE CASILLAS	MEDIA	MEDIANA	% MÁS DE 1 CASILLA:	% MÁS DE 2 CASILLAS
50	1	1	1	0
100	1	1	3	0
200	1.1	1	6	0
300	1.1	1	9	0
358	1.1	1	11	1
500	1.2	1	16	1
600	1.2	1	19	2
700	1.2	1	21	3

3. Estratificación por distrito local.

TABLA ESTRATOS:	
Número de estratos	22
Media número casillas	307
Mediana número casillas	308
Número de estratos < 50	0

Tabla de precisiones:

PARTIDO	50	100	200	300	358	500	600	700
Otros	1.01	0.76	0.59	0.46	0.41	0.36	0.34	0.27
PAN	2.8	1.9	1.24	1.1	0.92	0.75	0.76	0.62
PRD	1.15	0.84	0.54	0.48	0.44	0.35	0.3	0.34
PRI	2.94	1.76	1.27	1.09	0.93	0.73	0.67	0.61

Tabla de CAE's:

NÚMERO DE CASILLAS	MEDIA	MEDIANA	% MÁS DE 1 CASILLA:	% MÁS DE 2 CASILLAS
50	1	1	1	0
100	1	1	2	0
200	1.1	1	6	0
300	1.1	1	9	0
358	1.1	1	11	1
500	1.2	1	15	1
600	1.2	1	19	2
700	1.2	1	21	3

4. Estratificación por distrito local y tipo de sección

TABLA ESTRATOS:

Número de estratos	44
Media número casillas	153
Mediana número casillas	156
Número de estratos < 50	4

Tabla de precisiones:

PARTIDO	50	100	200	300	358	500	600	700
Otros	1	0.75	0.52	0.4	0.42	0.38	0.29	0.27
PAN	2.16	1.57	1.23	0.96	0.9	0.77	0.73	0.69
PRD	0.96	0.85	0.64	0.46	0.42	0.33	0.34	0.33
PRI	1.96	1.58	1.25	0.94	0.96	0.69	0.7	0.6

Tabla de CAE's:

NÚMERO DE CASILLAS	MEDIA	MEDIANA	% MÁS DE 1 CASILLA:	% MÁS DE 2 CASILLAS
50	1	1	3	0
100	1	1	3	0
200	1.1	1	6	0

NÚMERO DE CASILLAS	MEDIA	MEDIANA	% MÁS DE 1 CASILLA:	% MÁS DE 2 CASILLAS
300	1.1	1	9	0
358	1.1	1	11	1
500	1.2	1	15	1
600	1.2	1	18	2
700	1.2	1	21	3

5. Estratificación por distrito local y distrito federal

TABLA ESTRATOS:

Número de estratos	38
Media número casillas	178
Mediana número casillas	177
Número de estratos < 50	3

Tabla de precisiones:

PARTIDO	50	100	200	300	358	500	600	700
Otros	0.91	0.8	0.53	0.4	0.39	0.37	0.32	0.31
PAN	2.12	1.86	1.3	1.11	0.94	0.85	0.71	0.62
PRD	1.06	0.83	0.57	0.47	0.44	0.35	0.33	0.3
PRI	2.13	1.82	1.22	1.09	0.92	0.81	0.74	0.61

Tabla de CAE's:

NÚMERO DE CASILLAS	MEDIA	MEDIANA	% MÁS DE 1 CASILLA:	% MÁS DE 2 CASILLAS
50	1	1	2	0
100	1	1	3	0
200	1.1	1	6	0
300	1.1	1	9	0
358	1.1	1	11	1
500	1.2	1	15	1
600	1.2	1	18	2

NÚMERO DE CASILLAS	MEDIA	MEDIANA	% MÁS DE 1 CASILLA:	% MÁS DE 2 CASILLAS
700	1.2	1	22	3

6. Estratificación por distrito local, distrito federal y tipo de sección

TABLA ESTRATOS:

Número de estratos	75
Media número casillas	90
Mediana número casillas	71
Número de estratos < 50	25

Tabla de precisiones:

PARTIDO	50	100	200	300	358	500	600	700
Otros	0.66	0.71	0.53	0.5	0.32	0.32	0.28	0.29
PAN	1.63	1.41	1.03	1.01	0.93	0.8	0.67	0.69
PRD	0.87	0.71	0.5	0.46	0.43	0.33	0.32	0.29
PRI	1.7	1.49	1.11	0.94	0.92	0.76	0.61	0.67

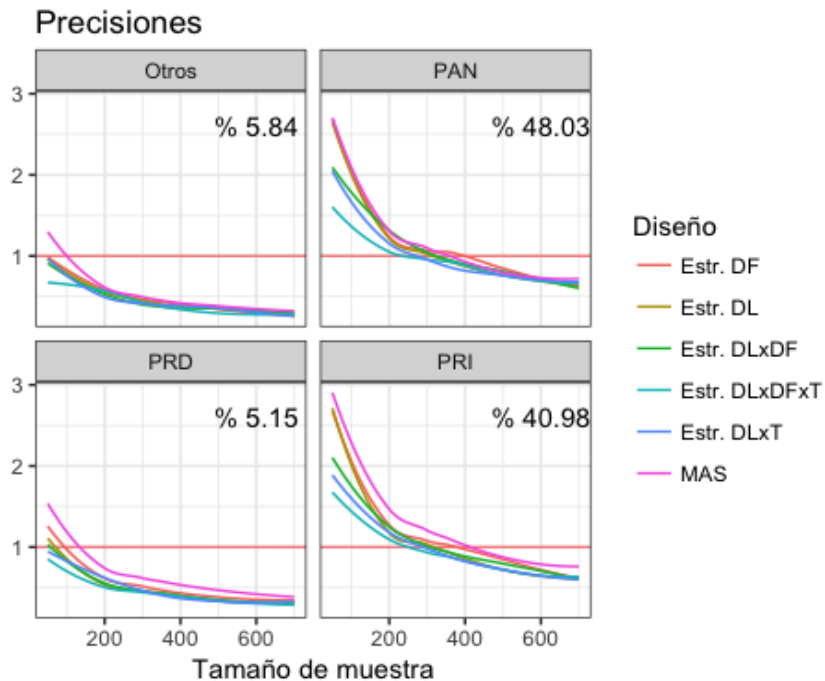
Tabla de CAE's:

NÚMERO DE CASILLAS	MEDIA	MEDIANA	% MÁS DE 1 CASILLA:	% MÁS DE 2 CASILLAS
50	1.1	1	7	1
100	1.1	1	7	1
200	1.1	1	8	0
300	1.1	1	10	1
358	1.1	1	11	1
500	1.2	1	15	1
600	1.2	1	18	2
700	1.2	1	21	2

Para tamaños de muestra variando de 50 hasta 600, en todas las estratificaciones consideradas resulta que más del 80% de los CAE's reportan

únicamente una casilla, por lo que ese criterio es fácil de cumplir en este estado.

A manera de resumen, en la siguiente gráfica vemos las precisiones para distintos tamaños de muestra y diseños. La línea horizontal indica una precisión del 1%. En la esquina superior derecha aparecen los porcentajes de votos a favor recibidos de cada partido en la elección estatal del 2012, datos con los cuáles se realizaron todas las pruebas.



Conclusión

Se eligió la estratificación por distrito local, ya que para muestras mayores a 358 casillas, la precisión se asemeja mucho a la mejor precisión obtenida, que proviene del cruce de todas las variables. Finalmente, se eligió un tamaño de muestra de 500, con lo cual se espera obtener precisiones menores a 0.75 puntos porcentuales, con al menos 95% de confianza. Considerando lo anterior, el diseño muestral corresponde a un muestreo estratificado simple de casillas con 22 estratos.

5.2.1.4. Jalisco¹⁷

Diseño muestral

El estado de Jalisco tiene la particularidad de que para estas elecciones, los distritos electorales federales coinciden con los distritos electorales locales, de manera que en forma natural, se considerarán a los distritos federales como estratos. La proporción de muestra nacional que será asignada al estado de Jalisco es de, aproximadamente, 467 casillas.

Se hicieron varios ejercicios para tener una idea de la precisión que se alcanzaba con este tamaño de muestra, suponiendo que todos los intervalos se construían con un 95% de confianza.

Estos ejercicios consideraron elecciones anteriores, tanto presidenciales como locales y algunas simulaciones en las que se probaron varios escenarios, desde una victoria holgada por parte de un partido, hasta diferentes tipos de elecciones competidas. En todos los casos, se obtuvieron precisiones muy razonables. Para el caso local, esta no fue menor al 0.85% y para el nacional fue del orden del 0.25%. Así, el diseño muestral para el estado de Jalisco, será considerar a los distritos federales como estratos.

¹⁷ Elaborado por el Dr. Raúl Rueda Díaz del Campo con la colaboración del Lic. Javier Santibáñez Cortés

5.2.1.5. Morelos¹⁸

Introducción

Como se acordó previamente en el COTECORA, el diseño muestral para cada uno de los estados será aleatorio estratificado, quedando pendiente la definición de los estratos, el tamaño de muestra y el error de estimación (precisión) alcanzada por el diseño.

En este documento se presenta un estudio de simulación que considera distintos criterios de estratificación y distintos tamaños de muestra para el Conteo Rápido en Morelos. La base de estudio es el listado de las 2,240 casillas junto con su votación reportada en la elección a gobernador en el estado de Morelos en el año 2012.

Los resultados oficiales se encuentran en https://es.wikipedia.org/wiki/Elecciones_estatales_de_Morelos_de_2012. Los porcentajes reportados corresponden al cociente del número de votos totales a favor de un candidato dividido entre el número de votos totales emitidos (que incluye a todos los candidatos, más nulos y no registrados).

Se reportan los errores de estimación alcanzados para cada uno de los tres candidatos punteros en el 2012, para cada criterio de estratificación considerado y los distintos tamaños de muestra.

Adicionalmente, se calcula la distribución de casillas por ARE's en muestra. Es decir, para las 439 ARE's registradas (+1 que corresponde a las casillas que no tienen identificador), se determina primero cuántas de ellas tienen al menos una casilla en muestra y de aquellas ARE's en muestra cuántas tienen cero casillas, cuántas tienen una casilla, cuántas tienen dos casillas, etc.

Método

Se realiza un estudio de simulación en el cual para cada tamaño de muestra n se seleccionan de manera repetida (5,000 veces) muestras estratificadas con asignación proporcional al tamaño del estrato. Es decir se asigna dentro de cada estrato un tamaño de muestra proporcional al tamaño del estrato (número de casillas en el estrato) de tal manera que la suma de las casillas en muestra de cada estrato sumen el tamaño de muestra total n .

Para cada muestra simulada de tamaño n se calcula el estimador de razón de la proporción a favor de un candidato. Considerando una muestra aleatoria simple de n_h casillas del estrato h . En la casilla i del estrato

¹⁸ Elaborado por el Dr. Luis Enrique Nieto Barajas con la colaboración del Mtro. Carlos Samuel Pérez Pérez

h se denota x_{hij} el total de votos del candidato j de modo que se propone el estimador de la proporción efectiva de votos p_j para cada candidato:

$$\hat{p}_j = \frac{\sum_{h=1}^L N_h \left(\frac{1}{n_h} \sum_{i=1}^{n_h} x_{hij} \right)}{\sum_{j=1}^J \sum_{h=1}^L N_h \left(\frac{1}{n_h} \sum_{i=1}^{n_h} x_{hij} \right)} .$$

En la cual N_h es el número de casillas en el estrato h , L es el número total de estratos y J es el número total de candidatos incluyendo los nulos y los no registrados.

En este ejercicio de Morelos 2012 se considera los tres principales partidos/coaliciones que son:

- PAN
- PRD-PT-MC
- PRI-PVEM-PNA

Para una n fija se tienen 5,000 valores de los estimadores de razón para los 3 candidatos y 5,000 distribuciones de casillas por ARE's. Estos 5,000 valores de los estimadores de razón se grafican en histogramas y se utilizan para calcular los errores de estimación a través de los valores correspondientes a los percentiles 2.5 y 97.5. El error de estimación B se calcula como $B = \max\{\xi_{0.975} - p_j, p_j - \xi_{0.025}\}$.

Por otro lado, las 5,000 distribuciones de casillas por ARE's se promedian y se reportan en una sola distribución promedio para cada tamaño de muestra. Se define una nueva variable que tome valores del 1 al 440 que son el total de ARE's reales. Los tamaños de muestra a considerar son $n \in \{100, 125, 150, \dots, 275, 300\}$ es decir de 100 a 300 con incrementos de 25 en 25.

Estratificación

Los criterios que se consideraron para definir los estratos son:

- Criterio I: Muestreo Aleatorio Simple (MAS)
- Criterio II: Tipo de sección (TipoSec)

- Criterio III: Distrito federal (DisFed)
- Criterio IV: Distrito local (DisLoc)
- Criterio V: Tipo de sección + distrito federal (TipoSec_DisFed)
- Criterio VI: Distrito local + distrito federal (DisLoc_DisFed)
- Criterio VII: Tipo de sección + distrito local (TipoSec_DisLoc)
- Criterio VIII: Tipo de sección + distrito local + distrito federal (TipoSec_DisLoc_DisFed)

A continuación se presenta una tabla con distribución del total de estratos para cada criterio de estratificación:

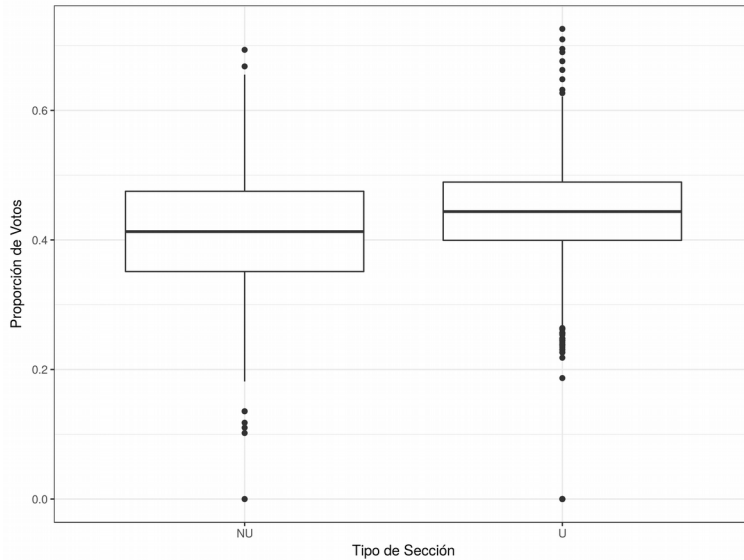
Tabla 1. Número de Estratos por criterio de estratificación

CRITERIO	NÚMERO ESTRATOS
Muestreo Aleatorio Simple	1
TipoSec	2
DisFed	5
DisLoc	12
TipoSec_DisFed	10
DisLoc_DisFed	18
TipoSec_DisLoc	23
TipoSec_DisLoc_DisFed	32

Para no tener problemas en la simulación, se considera la regla de que se seleccione al menos una casilla en cada estrato. Para los estratos con muy pocas casillas se pueden colapsar con algún otro estrato que compartan una o dos de las categorías. En este caso se revisan cuántas casillas caen en cada estrato para ver si son muy pocas por estrato y se deben unificar.

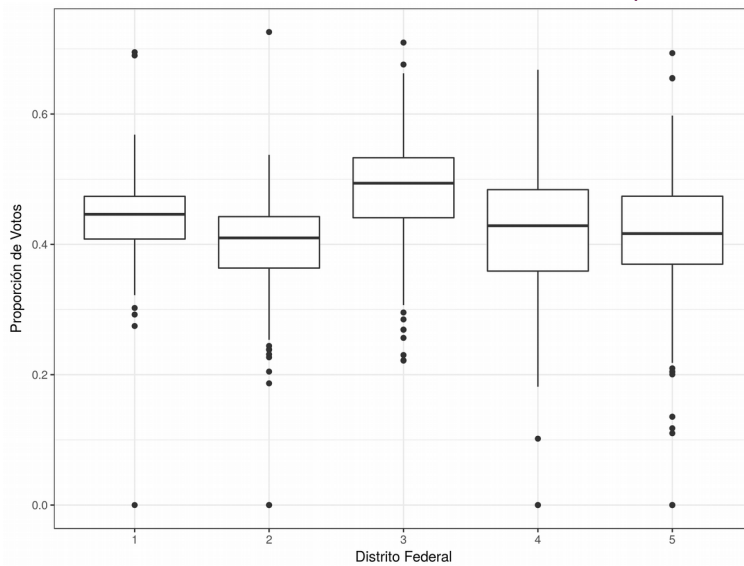
Antes de realizar el estudio de simulación, se presentan diagramas de caja de las proporciones de votos a favor de la coalición ganadora de la elección PRD-PT-MC, por tipo de variable de estratificación. En la Figura 1 se considera la variable tipo de sección en la cual no se aprecia mucha heterogeneidad en las tres categorías de esta variable.

Figura 1. Proporciones de votos a favor de la coalición PRD-PT-MC por tipo de sección



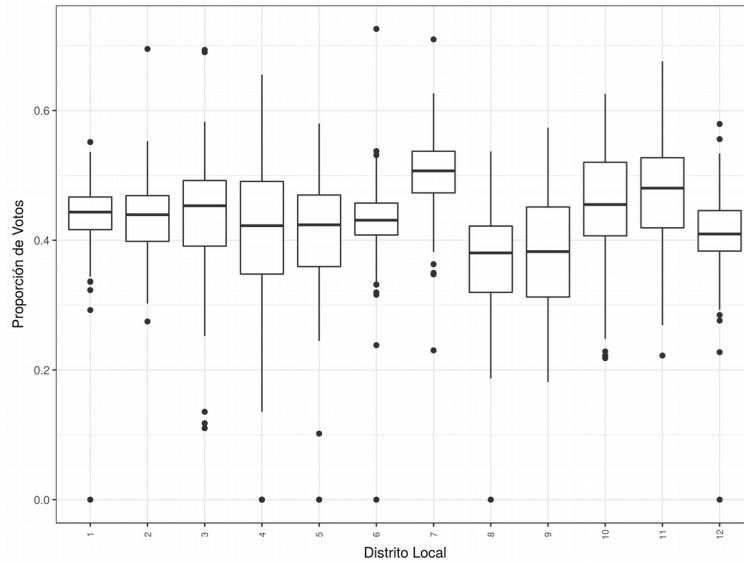
En la Figura 2 se considera el distrito federal en la cual los diagramas de caja tampoco son muy distintos entre los cinco distritos.

Figura 2. Proporciones de votos a favor de la coalición PRD-PT-MC por distrito federal 2012



Finalmente la Figura 3 considera la variable de distrito local. Los diagramas de caja sí cambian considerablemente de un distrito local a otro, lo que indica que esta variable de estratificación podría reducir importantemente la varianza del estimador.

Figura 3. Proporciones de votos a favor de la coalición PRD-PT-MC por distrito local 2012



Optimización de cómputo para simulación

Se implementaron funciones auxiliares que permiten realizar un procesamiento en paralelo para las 5,000 repeticiones para cada tamaño de muestra y cada criterio de estratificación.

Resultados

A continuación se presentan los resultados del estudio de simulación para cada uno de los tres principales candidatos. Se presentan tablas con los errores para cada tamaño de muestra y cada criterio de estratificación, así como una gráfica resumen de los errores de todos los criterios de estratificación considerados. Al final se incluye una tabla con el número de ARE's en muestra y la distribución de casillas por ARE's, también para cada tamaño de muestra.

Candidato: PAN

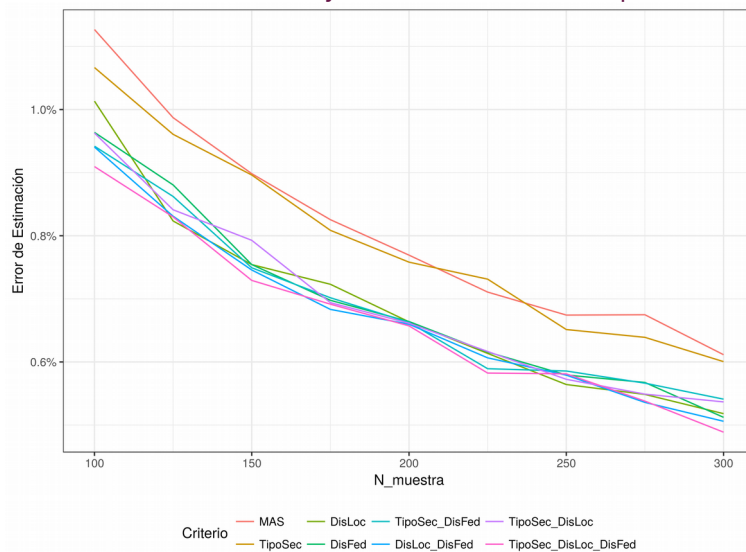
Errores de Estimación

Los resultados para el PAN considerando los distintos criterios de estratificación por tamaño de muestra se presentan a continuación:

Tabla 2. Distribución de errores de estimación para el PAN (en puntos porcentuales)

N	MAS	TIPOSEC	DisLoc	DisFe D	TIPOSE	DisLoc DisFED	TIPOSEC DisLoc	TIPOSEC
					C			DisFED
100	1.127	1.067	1.013	0.964	0.942	0.941	0.963	0.910
125	0.987	0.961	0.823	0.881	0.862	0.831	0.841	0.830
150	0.898	0.896	0.754	0.754	0.749	0.746	0.793	0.729
175	0.826	0.809	0.723	0.698	0.702	0.683	0.694	0.692
200	0.769	0.758	0.664	0.664	0.663	0.660	0.661	0.657
225	0.711	0.731	0.614	0.616	0.589	0.606	0.617	0.582
250	0.674	0.651	0.564	0.579	0.586	0.579	0.572	0.581
275	0.675	0.639	0.549	0.567	0.566	0.536	0.549	0.538
300	0.611	0.600	0.518	0.512	0.541	0.506	0.537	0.489

Figura 4. Relación del error de estimación y el tamaño de muestra para cada criterio para PAN



Candidato: PRD-PT-MC

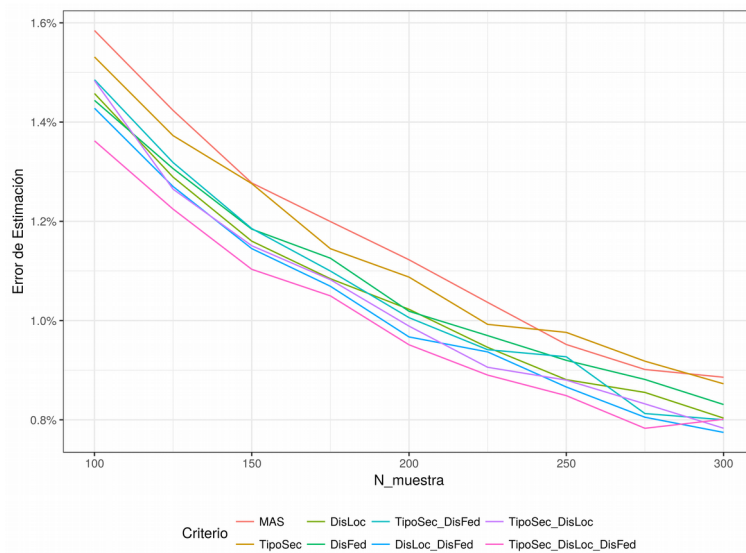
Errores de Estimación

Los resultados para la coalición PRD-PT-MC considerando los distintos criterios de estratificación por tamaño de muestra se presentan a continuación:

Tabla 3. Distribución de errores de estimación para la coalición PRD-PT-MC (en puntos porcentuales)

N	MAS	TIPOSEC	DisLoc	DisFED	TIPOSEC DisFED	DisLoc DisFED	TIPOSEC DisLoc	TIPOSEC DisLoc DisFED
100	1.585	1.531	1.458	1.444	1.485	1.428	1.484	1.362
125	1.423	1.373	1.289	1.306	1.318	1.270	1.265	1.225
150	1.277	1.276	1.160	1.184	1.185	1.145	1.151	1.103
175	1.200	1.145	1.084	1.126	1.100	1.069	1.082	1.050
200	1.123	1.088	1.023	1.019	1.006	0.967	0.989	0.951
225	1.037	0.992	0.946	0.970	0.941	0.937	0.906	0.890
250	0.952	0.976	0.881	0.920	0.927	0.866	0.880	0.849
275	0.901	0.918	0.855	0.881	0.813	0.805	0.832	0.783
300	0.886	0.873	0.803	0.831	0.800	0.775	0.783	0.801

Figura 5. Relación del error de estimación y el tamaño de muestra para cada criterio para la coalición PRD-PT-MC



Candidato: PRI-PVEM-PNA

Errores de Estimación

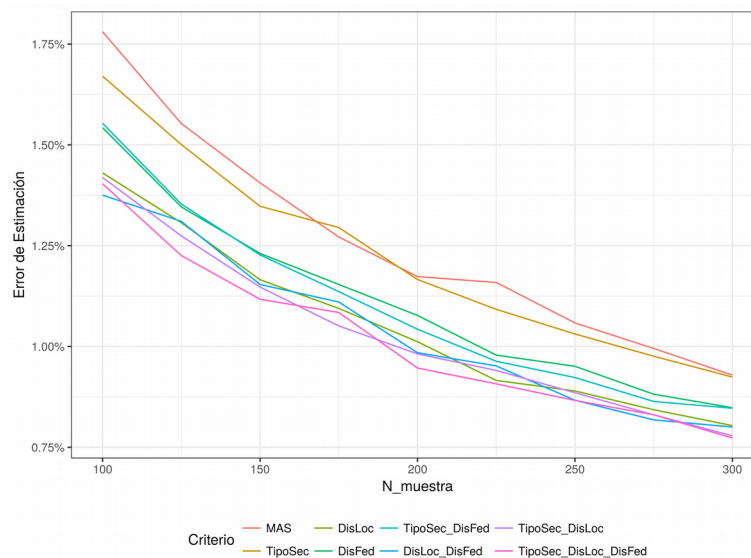
COTECORA | Criterios científicos, logísticos y operativos para la realización de los Conteos Rápidos 2018 y protocolo para la selección de las muestras

Los resultados para la coalición PRI-PVEM-PNA considerando los distintos criterios de estratificación por tamaño de muestra se presentan a continuación:

Tabla 4. Distribución de errores para la coalición PRI-PVEM-PNA (en puntos porcentuales)

N	MAS	TIPOSEC	DISLOC	DISFED	TIPOSEC	DISLOC	TIPOSEC	TIPOSEC
					DISFED	DISFED	DISLOC	DISLOC
100	1.78 1	1.670	1.430	1.543	1.554	1.375	1.419	1.404
125	1.55 3	1.501	1.307	1.346	1.353	1.310	1.274	1.226
150	1.40 6	1.348	1.166	1.231	1.227	1.154	1.147	1.117
175	1.27 2	1.295	1.094	1.154	1.136	1.110	1.051	1.084
200	1.17 3	1.166	1.012	1.077	1.043	0.984	0.981	0.946
225	1.15 9	1.092	0.916	0.979	0.963	0.952	0.941	0.907
250	1.05 8	1.031	0.890	0.951	0.923	0.867	0.885	0.866
275	0.99 5	0.976	0.843	0.882	0.864	0.818	0.830	0.831
300	0.92 9	0.924	0.804	0.848	0.847	0.800	0.778	0.773

Figura 6. Relación del error de estimación y el tamaño de muestra para cada criterio para PRI-PVEM-PNA



Total de ARE's en muestra

En este estudio de simulación también se calculó el total de ARE's en muestra considerando cada criterio de estratificación para cada tamaño de muestra.

Tabla 5. Número ARE's en muestra por criterio y tamaño de muestra

N	MAS	TIPOSEC	DISLOC	DISFED	TIPOSE C DISFED	DISLOC DISFED	TIPOSE C DISLOC	TIPOSE C DISLOC DISFED
100	90	90	93	91	91	95	93	97
125	111	111	113	111	112	113	114	116
150	130	130	130	130	132	132	131	135
175	148	148	149	149	150	150	151	153
200	165	165	166	167	165	167	167	168
225	181	181	182	183	182	184	183	185
250	198	198	199	198	196	201	198	200
275	212	212	214	213	213	216	216	216
300	226	225	228	226	226	228	229	232

Distribución de casillas en muestra por ARE's

Finalmente se presenta la distribución de casillas en muestra por ARE's para cada tamaño de muestra.

Tabla 6. Porcentaje de ARE's por número de casillas

N	CASILLAS	MAS	TIPOSEC	DISLOC	DISFED	TIPOSEC DISFED	DISLOC DISFED	TIPOSEC DISLOC	TIPOSEC DISLOC DISFED
100	1	90.45	90.53	91.10	90.76	91.02	91.64	91.91	92.14
100	2	9.04	8.97	8.51	8.78	8.56	8.02	7.77	7.58
100	3	0.49	0.48	0.38	0.45	0.42	0.33	0.31	0.28
100	4	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.00

N	CASILLAS	MAS	TIPOS EC	DISLOC	DISFED	TIPOS EC DISFED	DISLOC DISFED	TIPOS EC DISLOC	TIPOS EC DISLOC DISFED
100	5	0.00	0.00	NA	0.00	0.00	0.00	NA	NA
125	1	88.18	88.09	88.91	88.51	88.68	89.44	89.47	90.01
125	2	11.02	11.11	10.45	10.76	10.64	9.99	9.98	9.50
125	3	0.79	0.77	0.62	0.71	0.66	0.56	0.54	0.48
125	4	0.03	0.04	0.02	0.03	0.02	0.01	0.02	0.01
125	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA
150	1	85.79	85.80	86.63	86.04	86.15	86.96	87.30	87.59
150	2	13.07	13.05	12.44	12.89	12.78	12.18	11.86	11.66
150	3	1.12	1.11	0.90	1.04	1.03	0.84	0.81	0.74
150	4	0.05	0.05	0.03	0.04	0.04	0.03	0.03	0.02
150	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
150	6	NA	NA	NA	NA	NA	0.00	NA	NA
175	1	83.42	83.41	84.23	83.79	83.89	84.60	84.78	85.32
175	2	15.03	15.03	14.43	14.72	14.68	14.17	13.99	13.59
175	3	1.48	1.50	1.29	1.44	1.38	1.19	1.19	1.06
175	4	0.08	0.09	0.07	0.08	0.07	0.05	0.06	0.05
175	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
175	6	0.00	NA	NA	NA	NA	NA	NA	NA

COTECORA | Criterios científicos, logísticos y operativos para la realización de los Conteos Rápidos 2018 y protocolo para la selección de las muestras

N	CASILLAS	MAS	TIPOS EC	DisLo C	DisF ED	TIPOS EC DISFED	DisLo C DISFED	TIPOS EC DisLo C	TIPOS EC DisLo C DISFED
200	1	81.11	81.21	81.81	81.31	81.74	82.33	82.56	82.98
200	2	16.82	16.77	16.38	16.72	16.40	16.02	15.82	15.52
200	3	1.95	1.91	1.72	1.85	1.77	1.58	1.55	1.44
200	4	0.13	0.13	0.10	0.12	0.11	0.08	0.08	0.08
200	5	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
200	6	0.00	NA	NA	NA	NA	NA	0.00	0.00
225	1	78.83	78.87	79.63	78.99	79.31	80.07	80.18	80.66
225	2	18.60	18.55	18.08	18.51	18.27	17.80	17.71	17.34
225	3	2.39	2.40	2.15	2.33	2.25	2.02	2.00	1.89
225	4	0.18	0.19	0.14	0.17	0.16	0.13	0.13	0.12
225	5	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00
225	6	0.00	NA	NA	0.00	0.00	0.00	NA	0.00
250	1	76.59	76.69	77.02	76.85	77.10	77.43	77.89	78.37
250	2	20.23	20.15	20.03	20.07	19.90	19.76	19.41	19.12
250	3	2.90	2.90	2.73	2.85	2.76	2.61	2.51	2.35
250	4	0.27	0.26	0.21	0.23	0.23	0.20	0.19	0.16
250	5	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
250	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

COTECORA | Criterios científicos, logísticos y operativos para la realización de los Conteos Rápidos 2018 y protocolo para la selección de las muestras

N	CASILLAS	MAS	TIPOS EC	DISLOC	DISFED	TIPOS EC DISFED	DISLOC DISFED	TIPOS EC DISLOC	TIPOS EC DISLOC DISFED
275	1	74.37	74.36	74.92	74.58	74.63	75.36	75.35	76.00
275	2	21.78	21.76	21.51	21.70	21.67	21.28	21.26	20.83
275	3	3.50	3.52	3.27	3.39	3.37	3.09	3.12	2.92
275	4	0.34	0.34	0.28	0.31	0.32	0.26	0.26	0.24
275	5	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01
275	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
300	1	72.10	72.20	72.84	72.46	72.48	73.16	73.20	73.36
300	2	23.32	23.26	22.97	23.12	23.13	22.78	22.73	22.72
300	3	4.13	4.09	3.81	4.00	3.97	3.70	3.71	3.59
300	4	0.44	0.44	0.36	0.41	0.40	0.34	0.35	0.32
300	5	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01
300	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
300	7	NA	NA	NA	0.00	NA	NA	NA	NA

Conclusiones sobre el diseño muestral

Después de haber realizado este estudio de simulación basado en la elección de gobernador más reciente que se registró en el estado de Morelos en el año 2012, se sugiere lo siguiente:

- Que el diseño muestral para la elección a la Gubernatura en el estado de Morelos para el año 2018 sea aleatorio estratificado de casillas con asignación proporcional al tamaño del estrato.

- Que el criterio de estratificación esté basado en los distritos locales vigentes en el estado en el 2018.
- Que el tamaño de muestra sea de $n=200$ casillas repartidas en cada estrato de manera proporcional al número de casillas por estrato.
- El error de estimación (precisión) alcanzada por este diseño y con este tamaño de muestra es de 0.664% (PAN), 1.012% (PRI-PVEM-PNA) y 1.023% (PRD-PT-MC), para los tres principales partidos/coaliciones. Es decir, el error de estimación global del diseño es de aproximadamente 1%.
- Con este diseño, la distribución de casillas por ARE's indica que el 81.8% de ellos tienen asignada una sola casilla, lo que garantiza que si todos los CAE's en muestra reportan una sola casilla se tendría disponible información de aproximadamente el 83% de la muestra.
- Se sugiere que este criterio de estratificación, distritos locales, sea el utilizado en el estado de Morelos para el diseño muestral de la elección Presidencial.
- Como dato informativo, usando la información de la elección Presidencial del 2012 y considerando una muestra nacional de 7,500 casillas, de estas, únicamente 118 caerían el Morelos con una asignación proporcional.

5.2.1.6. Puebla¹⁹

Diseño de muestra para la elección de la Gubernatura en el estado de Puebla

Se realizará un muestreo estratificado, donde los estratos son los distritos locales y en cada uno de ellos se seleccionará una muestra aleatoria simple sin reemplazo de casillas. La distribución de la muestra a los estratos se hará proporcionalmente al número de casillas del estrato. La proporción de votos para cada partido se estimará mediante un estimador de razón combinado.

Determinación del tamaño de muestra y estratificación

Para definir la conformación de los estratos a utilizar en la elección del 2018, utilizando la información de la elección de gobernador para Puebla en el año 2016, se estudiaron varias estratificaciones:

1. Sin estratos
2. Distritos federales (son 15)
3. Distritos locales (son 26)
4. Cruce de distritos federales por tipo de sección (urbana, no urbana) (son 25)
5. Cruce de distritos locales por tipo de sección (urbana, no urbana) (son 41)

La comparación entre estos tipos de estratificación se realizó examinando el tamaño de muestra resultante para estimar la proporción de votos de cada uno de los contendientes mediante un estimador de razón combinado y considerando que la selección de casillas se realizará mediante un muestreo aleatorio simple sin reemplazo, que asigna la misma probabilidad de estar en muestra a cada una de las casillas del estrato.

$$n = \frac{z_{(1-\frac{\alpha}{2})}^2 * N \sum_{h=1}^L N_h S_h^2}{X^2 \delta^2 + z_{(1-\frac{\alpha}{2})}^2 * \sum_{h=1}^L N_h S_h^2} .$$

¹⁹ Elaborado por la Mtra. Patricia Isabel Romero Mares con la colaboración de la Mtra. Adriana Ducoing Watty

Donde:

δ : precisión deseada

X: total de votos emitidos

L: Número de estratos

$z_{(1-\alpha/2)}$: cuantil de la distribución normal estándar para una confianza de $(1-\alpha)$

$$S_h^2 = \frac{\sum_{i=1}^{N_h} [(Y_{hi}^p - \hat{Y}_h^p) - R(X_{hi} - \hat{X}_h)]^2}{N_h - 1}$$

Y_{hi}^p : total de votos en la casilla i del estrato h para el candidato p

\hat{Y}_h^p : promedio de votos en el estrato h para el candidato p

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

\hat{X}_h : promedio de votos emitidos por casilla en el estrato h

Como se mencionó anteriormente, se calculó el tamaño de muestra necesario con diferentes precisiones para cada una de las estratificaciones descritas, pero además se utilizó un nivel de confianza del 95%.

Con una precisión del 1%, el tamaño de muestra para cada estratificación y para cada partido o coalición, con una confianza del 95% es:

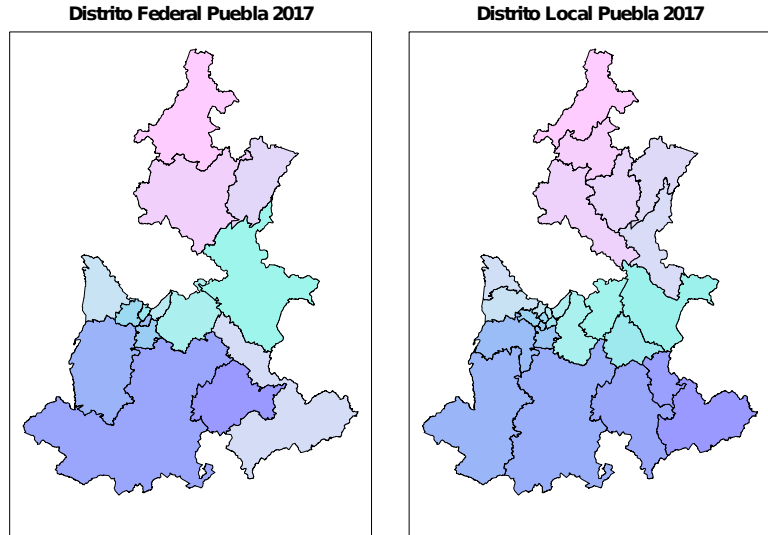
Tabla 1. Tamaños de muestra para diferentes estratificaciones

ESTRATIFICACIÓN	PARTIDO O COALICIÓN				
	PAN	PRI	PRD	MORENA	IND
Sin estratos	493	590	81	160	65
DF	434	422	71	107	32
DL	424	394	70	109	29
DFxTS	430	412	71	98	30
DLxTS	418	384	70	98	28

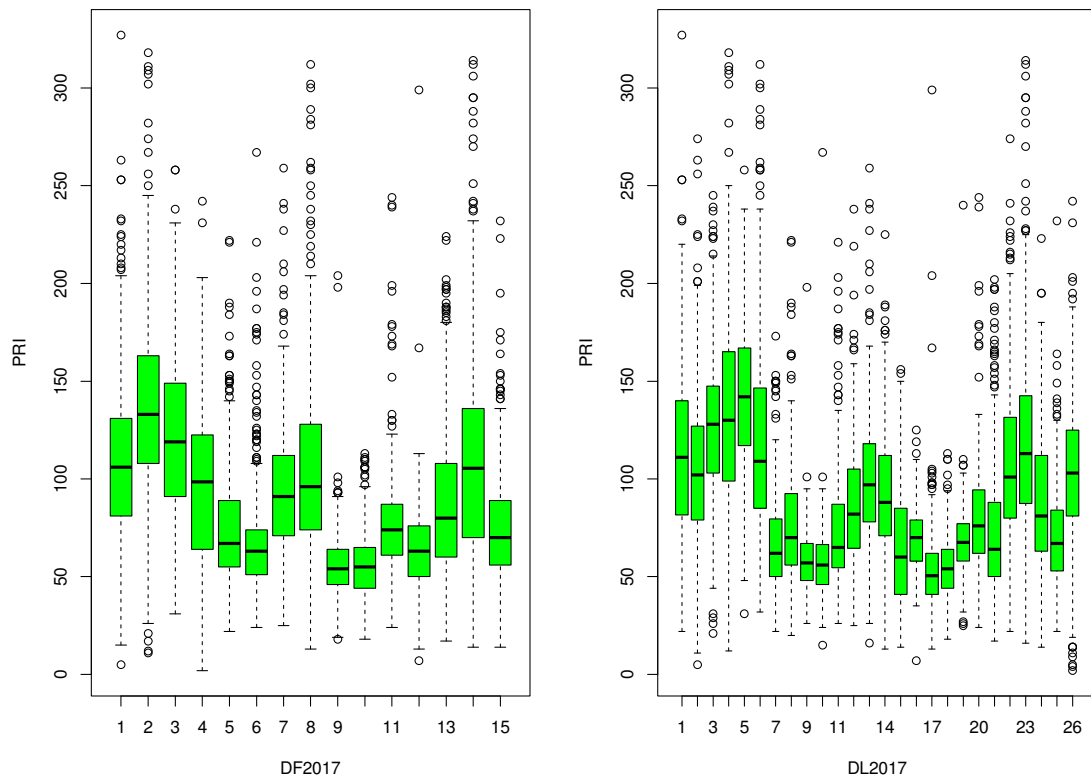
En la Tabla 1 se observa el gran efecto que tiene el usar estratos en el diseño de muestra, ya que se reducen mucho los tamaños de muestra al considerar estratos. Además, se ve que hay mayor variabilidad en la coalición del PAN, debido a que se necesitan tamaños de muestra más grandes. Así mismo, la coalición para la que la última estratificación tiene mayor efecto, es la del PRI. En los demás casos se observa que no hay mucha ganancia en la reducción del tamaño de muestra al hacer más fina la estratificación considerando el tipo de sección.

Comparando DF y DL, hay una ganancia marginal al tener más estratos utilizando la distribución local, y además si se observan los mapas a continuación, los distritos locales permiten tener mayor control en la

distribución geográfica de la muestra, por lo cual se decidió estratificar por distrito local.



Además, la distritación local capta la variabilidad que tiene la votación a la coalición del PRI, como se observa en la siguiente gráfica, comparando las estratificaciones por distrito federal y local.



Tomando en cuenta lo anterior, el diseño de muestra que se utilizará en el Conteo Rápido del estado de Puebla, será un Muestreo Estratificado Aleatorio, donde los estratos son los distritos locales, la selección de la muestra en cada estrato será con muestreo aleatorio simple (sin reemplazo) y se distribuirá el tamaño de muestra estatal ($n=424$) a los estratos en forma proporcional al número de casillas que los compongan.

Por otro lado, debemos considerar tres aspectos adicionales:

1. Habrá dos coaliciones, una de cinco partidos y otra de tres, así como los demás partidos, lo que resulta en 42 formas de votación no nula que podría dificultar el Conteo de los votos en las casillas y por lo tanto, retrasar la transmisión de la información al COTECORA.
2. No se ha realizado un Conteo Rápido en el estado de Puebla, por lo que no se cuenta con información del comportamiento de la llegada de la información en cuanto a tiempo de arribo y votaciones.
3. Con base en la información del Conteo Rápido para la Elección Presidencial del año 2012, se determinó que el porcentaje de casillas de la muestra correspondientes al estado de Puebla, cuya información llegó antes de las 10:30 pm fue de 80.6%.

Por lo anterior, se requiere un aumento al tamaño de muestra de 20%, es decir, el tamaño de muestra sería:

$$n = 424 * 1.2 = 509$$

Paralelamente, con el objetivo de estudiar el comportamiento de la estratificación por distrito local y el tamaño de muestra definido ($n=509$ casillas), así como estimar el porcentaje de CAE's a los que les corresponde entre 1 y 5 casillas en muestra, se realizó un estudio por simulación, con la información de la elección para gobernador de Puebla 2016.

Se tenía disponible la información de la votación de las $N=7,262$ casillas electorales del estado, con las cuales se seleccionaron 10,000 muestras con el diseño de muestra definido anteriormente.

A partir de las muestras seleccionadas se calculó el promedio de los estimadores de razón combinados y de la precisión, además se calculó la cobertura de los intervalos con 95% de confianza (la cobertura es el % de los intervalos que contienen el valor poblacional que se está estimando y que debe ser semejante al nivel de confianza requerido).

Por otro lado, se determinó el porcentaje de CAE's a los que les correspondió entre 1 y 5 casillas.

Tabla 2. Resultado de simulación con 10,000 muestras

PARTIDO O COALICIÓN	\hat{R}	PRECISIÓN OBSERVADA	COBERTURA	VALOR REAL DEL PARÁMETRO R
PAN	45.19	.91	95.2	45.19
PRI	33.42	.87	95.2	33.42
PRD	3.91	.36	94.7	3.91
MORENA	9.69	.45	94.8	9.69
IND	3.87	.23	95.4	3.86

Tabla 3. Porcentaje de CAE's a los que les correspondió entre 1 y 5 casillas en las 10,000 muestras

CAE's	CASILLAS				
	1	2	3	4	5
	79.57%	17.17%	2.80%	0.41%	0.05%

5.2.1.7. Tabasco²⁰

Elementos para diseño muestral

Primeros escenarios explorados

Para establecer el diseño muestral se utilizaron los resultados de 2012 tanto para la Gubernatura y Presidencia de la República. En particular se exploraron los siguientes esquemas de estratificación:

- Distritos federales.
- Distritos locales.
- Municipios.
- Distritos locales y federales separando la parte rural y urbana. (explorándose como criterio de separación el 20%, 30% y 40%, esto es, separando sólo si el distrito contaba con al menos ese porcentaje de casillas rurales o urbanas).
- Intersección de municipios con distritos federales.
- Intersección de distritos locales y distritos federales.

Número de casillas por estrato

Para el establecimiento del diseño muestral un elemento de importancia es el número de casillas que comprendería cada estrato. Este criterio es relevante en razón de que si el diseño considera estratos con un número muy pequeño, es susceptible a generar un punto de debilidad por la posibilidad del no arribo de suficiente información producto de algún incidente el día de la Jornada Electoral. La situación para diversos escenarios se presenta en la siguiente tabla. Como ya se mencionó, en los casos de considerar los distritos federal y local, se estudió el caso de su división en dos estratos dependiendo de su porcentaje de secciones rurales. Como puede observarse la división de los distritos en rurales y no rurales generaría estratos con un número pequeño de casillas, amén del crecimiento del número de ellos con la repercusión del caso federal que se verá más adelante.

²⁰ Elaborado por el Dr. Alberto Alonso y Coria con la colaboración del Fis. Joaquín Morales Bolio

Caso	Número de estratos	CASILLAS EN ELECCION			SOBRE LA MUESTRA		
		Casillas promedio por estrato	máximo numero de casillas	mínimo número de casillas	450		
					Casillas promedio por estrato	máximo numero de casillas en muestra	mínimo número de casillas en muestra
Locales y Rural .4	27	96.93	146	47	17	25	8
Locales y Rural .3	35	74.77	133	38	13	23	7
Locales y Rural .2	37	70.73	133	29	12	23	5
Distritos locales	21	124.62	158	97	21	27	17
Distritos Federales	6	436.17	507	365	75	87	63
Municipio	17	153.94	756	39	26	130	7
MunicipiopyDF	18	145.39	507	39	25	87	7
Federales interseccion locales	28	93.46	158	19	16	27	3
Federales y Rural .4	10	261.70	507	160	45	87	28
Federales y Rural .3	11	237.91	507	160	41	87	28
Federales y Rural .2	11	237.91	507	160	41	87	28

Sobre el tamaño de la muestra: Primeros resultados

Utilizando los resultados de la elección de gobernador 2012 se obtuvieron los siguientes resultados:

Caso	Número de estratos	RESULTADOS (S)					
		500		450		400	
		Máximo error promedio	distancia máxima promedio	Máximo error promedio	distancia máxima promedio	Máximo error promedio	distancia máxima promedio
Locales y Rural .4	27	0.773	0.275	0.834	0.326	0.881	0.307
Locales y Rural .3	35	0.743	0.265	0.809	0.245	0.851	0.283
Locales y Rural .2	37	0.765	0.246	0.976	0.225	0.859	0.316
Distritos locales	21	0.771	0.246	0.834	0.281	0.880	0.280
Distritos Federales	6	0.822	0.256	0.958	0.285	0.951	0.338
Municipio	17	0.777	0.243	0.831	0.259	0.872	0.286
MunicipiopyDF	18	0.770	0.253	0.820	0.258	0.889	0.319
Federales y Rural .4	10	0.852	0.256	0.948	0.283	0.951	0.337
Federales y Rural .3	11	0.855	0.256	0.948	0.272	0.948	0.334
Federales y Rural .2	11	0.855	0.256	0.879	0.274	0.948	0.334
Federales interseccion locales	28	0.755	0.227	0.808	0.272	0.868	0.249

De la tabla anterior se pueden derivar las siguientes conclusiones: La separación en rurales y no rurales en los distritos no disminuye notablemente el error promedio, pero, como ya se mencionó disminuye sensiblemente el número de casillas contenidas. Asimismo, el esquema de distritos locales arroja un escenario mucho mejor que la utilización de municipios o de distritos federales.

Comparación entre dos elecciones.

En la siguiente tabla se consideran dos tipos de elección y el resultado al realizar mil muestras para cada uno de los tamaños indicados, y utilizando distritos locales.

Distritos locales. Última elección gobernador				
Tamaño	Máximo error promedio (%)	distancia máxima promedio (%)	Máximo error en corridas (%)	distancia máxima máxima (%)
350	0.9468	0.2793	1.1458	0.8877
400	0.8774	0.2663	1.0533	1.1995
450	0.8287	0.2745	0.9996	0.8215
500	0.7740	0.2373	0.9143	0.7011

Distritos locales. Última elección presidencial				
Tamaño	Máximo error promedio (%)	distancia máxima promedio (%)	Máximo error en corridas (%)	distancia máxima máxima (%)
350	0.9648	0.3279	1.0950	0.7659
400	0.8959	0.2885	1.0053	0.9032
450	0.8263	0.2631	0.9445	0.7800
500	0.7780	0.2352	0.8653	0.6900

La situación con CAE's

Una problemática que habrá que atender es la referente al flujo real de llegada de información el día de la Jornada Electoral. En los ejercicios de Conteo Rápido en las últimas elecciones a gobernador se ha observado el arribo no oportuno de la totalidad de la información de las casillas de la muestra. Una de las hipótesis es el atribuir este fenómeno a la existencia de CAE's con la responsabilidad de reportar más de una casilla. Así pues, se procedió al estudio de cómo se afectaría la estimación bajo la suposición de que el CAE reportara sólo una de sus casillas asignadas o todas las casillas de una sección pero no de otras secciones. Esto último escenario se hizo en razón de que en este caso no existiría necesidad de traslado entre las casillas y por tanto no habría dificultad en el reporte de los resultados de las mismas.

Primeramente se estudia:

- El número de casillas (promedio) que atendería un CAE

- Si el CAE tendría todas sus casillas en una sola sección y en caso contrario, en cuantas.

Tamaño Muestra	CAES							
	Con una casilla	con dos	tres	cuatro o mas	todas las casillas en una sección	en dos secciones	en tres	en cuatro o más
353	208	58	9	1	235	38	2	0
400	220	69	12	1	253	47	2	0
450	229	81	17	2	269	57	3	0
501	231	95	23	3	280	68	4	0
550	231	109	27	5	287	80	5	0

Arribo de una sola casilla por CAE

Las siguientes tablas muestran la situación de arribo de una casilla por CAE.

Llegada de toda la muestra				
Tamaño	Error promedio (%)	Distancia promedio (%)	Máximo error en corridas (%)	distancia máxima máxima (%)
445/445	0.8323	0.2674	1.0477	1.1483
450/450	0.8279	0.2539	1.0152	1.3209
454/454	0.8260	0.2628	1.0040	1.0854
461/461	0.8178	0.2555	1.0138	1.2005

Una sola casilla por CAE				
Tamaño	Error promedio (%)	Distancia promedio (%)	Máximo error en corridas (%)	distancia máxima máxima (%)
445/332	1.0049	0.2837	1.3821	1.4189
450/325	1.0018	0.2731	1.3789	1.5434
454/339	1.0014	0.2746	1.4335	1.4373
461/338	0.9966	0.2740	1.4047	1.1079

Utilización de la estratificación para la elección Presidencial

- Como hipótesis, se supone que para el caso de la elección Presidencial a Tabasco le corresponderían 150 casillas en la muestra.
- Como primer paso, se analiza cómo quedarían poblados los estratos.
- Se comparan tres escenarios: Estratificación por distrito federal, por distrito local y por municipios

Si bien la estratificación por distritos locales arroja buenos resultados, un elemento de diseño es el que se utilizará la misma estratificación para la elección Presidencial. En razón de que el tamaño de la muestra para esta elección es menor que la necesaria para la local, implicaría una disminución en el número de casillas en los diferentes estratos. En la siguiente tabla se muestra la situación que prevalecería en los distintos escenarios de estratificación.

Distrito Local			Municipio			
Estrato-DL 2017	LN	Casillas	Estrato-Municipio	ID_Municipio	LN	Casillas
1	82,807	7	Balancan	1	40,077	4
2	76,245	7	Cárdenas	2	172,392	15
3	70,649	6	Centla	3	71,504	6
4	75,134	7	Centro	4	505,944	45
5	71,504	6	Comalcalco	5	148,908	13
6	83,254	7	Cunduacan	6	92,183	8
7	74,803	7	Emiliano Zapata	7	22,962	2
8	80,752	7	Huimanguillo	8	128,784	11
9	79,391	7	Jalapa	9	27,873	2
10	81,865	7	Jaipa de Méndez	10	62,384	6
11	73,474	7	Jonuta	11	22,870	2
12	83,113	7	Macuspana	12	116,199	10
13	94,738	8	Nacajuca	13	92,191	8
14	73,447	7	Paraiso	14	68,310	6
15	74,543	7	Tacotalpa	15	32,975	3
16	79,148	7	Teapa	16	39,603	4
17	91,340	8	Tenosique	17	42,730	4
18	74,862	7	Total		1,687,889	149
19	92,191	8				
20	93,524	8				
21	81,105	7				
Total	1687889	149				

Como puede observarse, las estratificaciones consideradas generan estratos con un número de casillas muy reducido. Así pues, se exploró realizar agrupaciones entre los dos distintos escenarios. En particular se exploró:

- Fusionando dos distritos locales (en razón de ser un número impar, en un caso se fusionaron tres). La fusión se realizó entre distritos contiguos de forma que no hubiera un incremento significativo de la varianza.
- Fusionando municipios.

Búsqueda de solución. Agrupación por distritos locales o de municipios

Integrando distritos locales					
ESTRATO	LN Estrato	Casillas	Casillas en Muestra		
1	151,267	290	39	45	50
2	137,255	243	32	37	42
3	142,056	241	32	37	41
4	138,282	245	33	38	42
5	148,830	240	32	37	41
6	218,357	360	48	55	62
7	146,931	264	35	40	45
8	156,106	255	34	39	44
9	132,836	221	30	34	38
10	148,742	258	35	40	44
			350	402	449

Integrando municipios					
ESTRATO	LN Estrato	Casillas	Casillas en Muestra		
1	124,491	239	32	36	41
2	198,962	355	47	54	61
3	459,203	756	101	116	130
4	124,486	224	30	35	39
5	131,349	212	28	32	36
6	131,367	229	30	35	39
7	195,989	336	45	52	58
8	154,815	266	36	41	46
			349	401	450

Debido a que el estrato 3 generado por el municipio “centro”, en comparación con los otros, es muy grande, se procedió a dividirlo obteniendo dos posibles escenarios:

1. A través de su intersección con distritos federales (2 estratos) y;
2. Con la intersección de grupos de distritos locales (3 estratos)

ESTRATO	LNestrato	Casillas	Casillas en Muestra		
3	296,966	507	68	77	87
9	162,237	249	33	38	43

ESTRATO	LNestrato	Casillas	Casillas en Muestra		
3	156,536	252	34	39	44
9	153,837	264	35	40	45
10	148,830	240	32	37	41

Agrupación Municipios dividiendo Centro en dos con DF				Agrupación Municipios dividiendo Centro en tres con DL			
Agrupación	Estrato	LN	Casillas	Agrupación	Estrato	LN	Casillas
1, 7, 11, 17	1	128,639	11	1, 7, 11, 17	1	128,639	11
9, 12, 15, 16	2	216,650	19	9, 12, 15, 16	2	216,650	19
4 y dist 4 federal	3	310,395	28	4 y 7, 8, 12, 21 locales	3	178,321	16
3, 14	4	139,814	12	3, 14	4	139,814	12
10, 13	5	154,575	14	10, 13	5	154,575	14
5	6	148,908	13	5	6	148,908	13
6, 8	7	220,967	20	6, 8	7	220,967	20
2	8	172,392	15	2	8	172,392	15
4 y dist 6 federal	9	195,549	17	4 y dists 6 y 12 locales	9	166,367	15
				4 y dists 9 y 10 locales	10	161,256	14

Resultados de las simulaciones

A continuación se muestran los resultados de las simulaciones.

	Estratos	Máximo error promedio (%)	distancia máxima promedio (%)	Máximo error en comidas (%)	distancia máxima (%)	Tamaño muestra
Colapsando distritos locales	10	0.838	0.268	0.969	0.434	449/449
Colapsando municipios	8	0.877	0.295	1.017	0.646	450/450
Colapsando municipios y dividiendo el municipio de Centro por distritos federales	9	0.856	0.369	0.976	0.910	450/450
Colapsando municipios y dividiendo el municipio de Centro por distritos locales	10	0.873	0.235	0.971	0.632	449/449

Balance de los Estratos con agrupación

	Distancia entre el LN mayor y	Mínimo número de casillas en	Desviación estandar de la LN
Agrupación distritos	89,117	13	26,305
Agrupación Municipios	377,305	12	123,926
Agrupación Municipios dividiendo Centro en dos con DF	181,756	12	56,584
Agrupación Municipios dividiendo Centro en tres con DL	92,328	12	30,237

Conclusión

1. Los resultados fueron alentadores si se utilizaba una estratificación con base en distritos locales.
2. Sin embargo, el utilizar este esquema para el caso de la elección Presidencial arroja estratos con un número muy reducido de casillas.
3. Se exploró la agrupación de distritos locales y de municipios.
4. De los resultados obtenidos se desprende que el error máximo entre las diferentes opciones no supera el 0.067%, de aquí que se pueden considerar equivalentes.
5. Considerando el criterio de balance entre los estratos definidos, observamos que la menor separación entre el estrato con mayor LN y el estrato con menor LN lo presenta el caso de agrupación de distritos locales. Sucede lo mismo si se calcula la desviación estándar de su LN. Este caso cuenta además con la ventaja de que el estrato con menor

número de casillas en la muestra para la elección Presidencial es superior al de las otras agrupaciones.

6. De lo anterior se concluye que la mejor propuesta de estratificación es la de agrupar distritos locales.
7. En cuanto al tamaño de la muestra, en razón del ejercicio de arribo de una casilla por CAE, se recomienda que sea de 450 casillas.
8. La agrupación propuesta estaría conformada por:

Conformación de los estratos		
Estrato	Agrupación distritos locales	LN
1	1, 15	157,350
2	4, 16	154,282
3	12, 18	157,975
4	11, 21	154,579
5	9, 10	161,256
6	6, 7, 8	238,809
7	5, 20	165,028
8	17, 19	183,531
9	2, 14	149,692
10	13, 3	165,387

5.2.1.8. Veracruz²¹

Diseño de muestreo para la elección de Gubernatura en el estado de Veracruz.

Las tres estrategias de estimación que implementarán los tres miembros del COTECORA de este grupo, estarán basadas en la misma información muestral. Considerando el Artículo 373 del RE donde se establece que la muestra debe abarcar la mayor dispersión geográfica posible, se concluye que el diseño muestral sea estratificado, donde al interior de cada estrato se seleccionarán casillas mediante un muestreo aleatorio simple sin reemplazo.

Diseño de muestreo

Para definir los estratos a emplear, se consideraron diferentes combinaciones entre las cuales se encuentran:

- Distritos federales.
- Distritos federales y distritos locales.
- Distritos federales y tipo de sección.
- Distritos locales.
- Distritos locales y tipo de sección.

Para decidir cuál es la mejor estratificación a emplear, se deben considerar factores como la precisión (límite máximo deseable para el error de estimación) que se pretende obtener, el tamaño de la información muestral (tamaño de muestra), el nivel de confianza deseado en las estimaciones y la forma en la que el tamaño de muestra (total) es repartido en cada uno de los estratos.

Ideas básicas del muestreo probabilístico para el Conteo Rápido

El objetivo del Conteo Rápido es estimar la proporción p de votos a favor de un candidato a la Gubernatura²². Para lo anterior, se selecciona una muestra aleatoria (bajo algún esquema de muestreo) de n casillas de un total de N y con la información muestral disponible se calcula un

21 Elaborado por el Dr. Gabriel Núñez Antonio con la colaboración del Mat. Ernesto Silva Sartorius

22 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.

estimador \hat{p} . Mediante el muestreo probabilístico es posible analizar diferentes *estrategias de selección, tamaños de muestra n y estimadores*, para asegurar que

$$|p - \hat{p}| \leq d, \quad (1)$$

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 $(1-\alpha) \times 100\%$, como

$$\hat{p} - d \leq p \leq \hat{p} + d \quad (2)$$

Esto significa que si, por ejemplo, $(1-\alpha) = 0.95$ 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

$$\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5, \dots, \hat{p}_{1000},$$

se esperarían que 950 de estas estimaciones cumplan con la condición (1).

Precisión en las estimaciones

El procedimiento anterior permite analizar diferentes diseños de muestreo para obtener las precisiones bajo varias estrategias de selección, diferentes tamaños de muestra y diferentes estimadores. En consecuencia, poder decidir cuál es el diseño de muestreo más adecuado para llevar a cabo el procedimiento correspondiente del Conteo Rápido en la elección para Gubernatura del estado de Veracruz.

A continuación se presentan y analizan las precisiones para varias estrategias de estratificación y diferentes tamaños de muestra.

Estrategias de selección

Aunque existen diversas estrategias de selección (muestreo aleatorio simple sin reemplazo (MASS), muestreo estratificado, muestreo por conglomerados o polietápico, etc.), dada la experiencia acumulada de otros ejercicios estadístico-electorales, en el caso del Conteo Rápido para Gubernatura del estado de Veracruz se enfocará en el muestreo aleatorio estratificado. En este esquema de muestreo probabilístico se divide a la población en

estratos/grupos²³ y en cada estrato se hace una selección de casillas²⁴. En este caso se tiene mejor control de la muestra pues en cada estrato tendremos casillas seleccionadas, lo que permite tener una mejor distribución de la muestra. Otra ventaja del muestreo estratificado es que permite lograr una mejor estimación si los estratos se definen de manera que al interior de cada estrato, la variable de interés, tenga un comportamiento más homogéneo que si se considerara la población completa, entonces la estimación tendrá más precisión que un MASS.

Por lo descrito en el párrafo anterior, lo que se hizo fue probar distintas estratificaciones y tamaños de muestra para seleccionar la que permitiera obtener la mejor precisión, pero también cuidando no rebasar la capacidad operativa de campo.

Población de referencia para fijar el tamaño de muestra

En todos los ejercicios realizados se tomaron como referencia los cómputos distritales de la elección para gobernador del estado de Veracruz en 2016 (Archivo: `_30_Veracruz_2016.txt`).

Estratificaciones iniciales

Las estratificaciones consideradas en los ejercicios de estimación fueron cinco:

1. Distritos federales
⇒ 20 estratos
2. Distritos federales y distritos locales
⇒ 61 estratos
3. Distritos federales y tipo de sección
⇒ 40 estratos
4. Distritos locales
⇒ 30 estratos
5. Distritos locales y tipo sección
⇒ 60 estratos

Estimador considerado

Se consideró el estimador de razón

23 Los estratos son ajenos entre sí y su unión conforma la población completa.

24 La selección en cada estrato usualmente se realiza mediante un MASS.

$$\hat{p}_i = \frac{\sum_{h=1}^L N_h \bar{y}_{hi}}{\sum_i \sum_{h=1}^L N_h \bar{y}_{hi}} \quad (3)$$

En donde **i** es el 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, N_h es el número de casillas instaladas en el estrato **h** y \bar{y}_{hi} es el número promedio de votos en las casillas seleccionadas en muestra para el candidato a la Gubernatura por el candidato **i** en el estrato **h**.

Las estimaciones se realizaron para todos los candidatos (incluyendo candidatos no registrados y nulos); sin embargo, en este apartado sólo se muestra el análisis para el candidato de MORENA que registró mayor varianza y requería un tamaño de muestra mayor para alcanzar la misma precisión que los demás candidatos.

Resultados basados en simulaciones para determinar el tamaño de muestra

Utilizando la base de la elección a gobernador en 2016 y considerando los 20 distritos federales definidos en 2017, las precisiones obtenidas usando el estimador de razón (3) se muestran en la Tabla 1.

N	DTO. FED (20 ESTRATOS)	DTO. FED X DTO. LOC (61 ESTRATOS)	DTO. FED. X TIPO SECCIÓN (40 ESTRATOS)	DTO. LOC (30 ESTRATOS)	DTO. LOC X TIPO SECCIÓN (60 ESTRATOS)
900	0.866	0.778	0.736	0.812	0.717
1000	0.809	0.764	0.667	0.806	0.671
1100	0.769	0.719	0.665	0.754	0.622
1200	0.745	0.683	0.608	0.697	0.604
1300	0.750	0.661	0.582	0.679	0.580
1400	0.686	0.610	0.590	0.640	0.581
1500	0.646	0.616	0.574	0.621	0.527

Tabla 1. Precisiones usando el estimador de razón, para distintos tamaños de muestra y estratificaciones (para obtener cada uno de los valores de la tabla se realizaron 5,000 simulaciones)

Se puede observar que las estrategias de selección que incluyen a los distritos locales por tipo de sección arrojan las mejores precisiones. Sin embargo, bajo esta estratificación se tienen 60 estratos, los cuales por la experiencia en otras elecciones se consideran demasiados, ya que pueden traer problemas en la proporción asignada a cada uno de estos estratos tanto de la muestra estatal como de la muestra Presidencial.

Por otro lado, dada la inquietud en la carga que han tenido los CAE's en los últimos ejercicio de Conteo Rápido, se analizó también la proporción de CAE's a los que les tocaría reportar 1, 2 y 3 o más casillas. Para este análisis se consideró la información de la elección de gobernador en 2016, así como la información de las ARE's definidas en 2016 y el tipo de sección definido al 31 de enero de 2016. Los resultados se muestran en la Tabla 2.

Tabla 2. Porcentaje de CAE's según el número de casillas (para obtener cada uno de los valores de la tabla se realizaron 5,000 simulaciones en cada escenario).

	CASILLAS POR CAE	DTO. FED (21 ESTRATOS)	DTO. FED X DTO. LOC (58 ESTRATOS)	DTO. FED. X TIPO SECCIÓN (42 ESTRATOS)	DTO. LOC (30 ESTRATOS)	DTO. LOC X TIPO SECCIÓN (60 ESTRATOS)
900	1	83.0%	83.6%	84.8%	83.2%	85.0%
	2	15.3%	14.9%	13.9%	15.2%	13.7%
	3 o más	1.7%	1.5%	1.3%	1.6%	1.3%
1000	1	81.1%	81.7%	83.0%	81.3%	83.2%
	2	16.8%	16.4%	15.3%	16.7%	15.2%
	3 o más	2.1%	1.9%	1.7%	2.0%	1.6%
1200	1	77.5%	78.0%	79.6%	77.6%	79.8%
	2	19.6%	19.2%	17.9%	19.5%	17.8%
	3 o más	2.9%	2.8%	2.5%	2.9%	2.4%

Analizando los resultados proporcionados por la Tabla 1 y la Tabla 2, considerando la recomendación inicial de los miembros del COTECORA, de tomar en cuenta los distritos federales como primera estratificación y considerando el acuerdo de los miembros del COTECORA, de seleccionar un diseño muestral en el que se tenga alrededor del 80% de CAE's asignados a 1 casilla, se consideró una muestra de tamaño $n=1,100$ y una estratificación por distrito federal y tipo de sección.

Con la estratificación anterior y considerando una asignación de la muestra proporcional al tamaño de los estratos, existían estratos con tamaños de muestra pequeños, tanto para la elección local como para la elección federal (esto último basado en el pronóstico de casillas, escenario Alto, proporcionado por la Dirección de Planeación y seguimiento de la DEOE, en marzo de 2018). Después de un análisis conjunto, se decidió colapsar aquellos estratos ubicados en el primer cuartil de acuerdo con el número de casillas para la elección estatal y colapsar aquellos estratos con menos de 10 casillas en muestra para la elección Presidencial. Así, se colapsaron finalmente 12 estratos.

Con esta estratificación y usando nuevamente el estimador de razón definido en (3) se obtuvo una precisión de 0.694 en las estimaciones a nivel estatal, con un porcentaje de CAE's con una sola casilla asignada. Los resultados basados en ejercicios con 5,000 simulaciones se muestran en la Tabla 3. Adicionalmente, a manera de comparación se muestran las precisiones alcanzadas bajo una estratificación por distrito federal.

Tabla 3. Precisiones usando un estimador de razón, para distintos tamaños de muestra y una estratificación de distrito federal por tipo de sección colapsando 12 de los 40 estratos definidos originalmente (resultados basados en 5,000 simulaciones)

N	D.T.O. FED.	
	D.T.O. FED (20 ESTRATOS)	X TIPO SECCIÓN (28 ESTRATOS)
900	0.866	0.772
1000	0.809	0.755
1100	0.769	0.694
1200	0.745	0.672
1300	0.750	0.662
1400	0.686	0.646
1500	0.646	0.599

Los 28 estratos resultantes, así como sus denominaciones se muestran en la Tabla 4. La columna Nh se refiere al total de casillas en cada estrato h. Los dos dígitos en las columnas Nombre, se refieren al número de distrito federal, la U y la NU se refieren al tipo de sección: urbana y no urbana. Así, por ejemplo, el Estrato 3, 03_U, se refiere a la intersección del distrito federal 3 y el tipo de sección urbana, en consecuencia su nombre se construye como 03_U. El Estrato 5, 04_U_NU, se refiere al distrito federal 4 colapsando la parte urbana y la no urbana, en consecuencia su nombre se construye como 04_U_NU. La columna nh se refiere al tamaño de la muestra estatal (n=1,100) para la elección a Gubernatura que se asignaría al estrato h. Por otro lado, la columna nh* se refiere al tamaño de la muestra nacional que se

asignaría al estrato h. Se debe mencionar que el tamaño de muestra nacional será de 7,500 casillas y con base al pronóstico de casillas, escenario Alto, proporcionado por la Dirección de Planeación y seguimiento de la DEOE, en marzo de 2018, al estado de Veracruz le tocarían aproximadamente 506 casillas.

Tabla 4. Estratificación y tamaños de muestra por distrito federal y tipo de sección.

ESTRATO	NOMBRE	Nh	nh	nh*	ESTRATO	NOMBRE	Nh	nh	nh*
1	01_U_N	492	52	24	15	11_U_N	555	59	27
2	02_U_N	481	51	23	16	12_U_N	609	64	30
3	03_U	200	21	10	17	13_U_N	556	59	27
4	03_NU	247	26	12	18	14_U	271	29	13
5	04_U_N	512	54	25	19	14_NU	264	28	13
6	05_U_N	436	46	21	20	15_U_N	512	54	25
7	06_U_N	535	56	26	21	16_U_N	525	55	25
8	07_U	218	23	11	22	17_U	238	25	12
9	07_NU	318	34	15	23	17_NU	307	32	15
10	08_U	206	22	10	24	18_U_N	487	51	24
11	08_NU	368	39	18	25	19_U	216	23	10
12	09_U	225	24	11	26	19_NU	325	34	16
13	09_NU	233	25	11	27	20_U	291	31	14
14	10_U_N	548	58	27	28	20_NU	245	26	12

Se debe mencionar que con la información actualizada de casillas, al 27 de abril de 2018, se realizó la distribución de casillas asignadas a cada estrato. La información actualizada se presenta en la Tabla 5.

Tabla 5. Estratificación y tamaños de muestra por distrito federal y tipo de sección.

ESTRATO	NOMBRE	Nh	nh	nh*	ESTRATO	NOMBRE	Nh	nh	nh*
1	01_U_N	494	51	24	15	11_U_N	566	59	27

ESTRAT O	NOMBRE	Nh	nh	nh*	ESTRAT O	NOMBRE	Nh	nh	nh*
2	02_U_N U	486	50	23	16	12_U_N U	614	64	29
3	03_U	204	21	10	17	13_U_N U	566	59	27
4	03_NU	252	26	12	18	14_U	275	29	13
5	04_U_N U	537	56	26	19	14_NU	273	28	13
6	05_U_N U	438	45	21	20	15_U_N U	521	54	25
7	06_U_N U	541	56	26	21	16_U_N U	534	55	25
8	07_U	225	23	11	22	17_U	235	24	11
9	07_NU	325	34	16	23	17_NU	303	31	14
10	08_U	199	21	10	24	18_U_N U	502	52	24
11	08_NU	386	40	18	25	19_U	217	23	10
12	09_U	230	24	11	26	19_NU	333	35	16
13	09_NU	243	25	12	27	20_U	285	30	14
14	10_U_N U	559	58	27	28	20_NU	256	27	12

1. Se debe señalar que la precisión 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 capaz (teóricamente) de alcanzar el nivel de precisión y confianza deseadas en la estimación. Sin embargo, debido a que se fija este tamaño de muestra tomando como referencia los resultados de una elección previa y a que se tendrá un porcentaje de no respuesta que no conocemos, una vez concluida la elección y con toda la información disponible se determinará la precisión realmente obtenida.
2. Los resultados de las simulaciones presentadas son únicamente para el método de estimación usado, y tienen como objetivo fijar un tamaño de muestra. Sin embargo, se pueden emplear métodos de estimación alternativos que arrojarán precisiones distintas.

3. También se debe señalar que con la información con la que se cuenta hasta el momento (la de marzo de 2017) se cuenta con 2,598 CAE's para la elección a la Gubernatura en el estado de Veracruz. Por lo cual es conveniente apuntar que un tamaño de muestra de 1,100 casillas se encuentra dentro de los límites de la capacidad operativa de campo.
4. Hay que mencionar que si se recibe un tamaño de muestra menor a 1,100 casillas, el día de la Jornada Electoral, todavía se pueden tener estimaciones confiables si la distribución de la muestra recibida es adecuada. Sin embargo, las precisiones serán menores como se puede ver en la Tabla 3. Por ejemplo, aún para un tamaño de muestra de 900 casillas (adecuadamente distribuidas) se tendría una precisión de 0.772% que está dentro del límite establecido del 1%.

5.2.1.9. Yucatán²⁵

Consideraciones para definir el diseño de muestreo para el Conteo Rápido en la elección de Gobernatura en el estado de Yucatán

Para el Conteo Rápido en la elección de Gobernatura en el estado de Yucatán, se seguirá un diseño de muestreo estratificado:

- Los estratos se definen colapsando distritos locales, tal como se muestra en la Tabla 5.
- Tomando como referencia el número de casillas instaladas para la elección de diputados en 2015, el número mínimo de casillas instaladas, por estrato, será de 306 (ver Tabla 8).
- El tamaño de muestra efectivo, para realizar la estimación, será de 205 casillas y con éste se obtendría un margen de error en la estimación de 0.96% (ver Tabla 7).
- Considerando el nivel de respuesta observado en el Conteo Rápido para la elección de diputados de 2015, sólo en Yucatán. Para alcanzar una muestra efectiva de 205 casillas (alrededor de las 10 pm), se necesitan aproximadamente 300 casillas. Las 300 casillas serán la muestra total y se obtienen a partir del nivel de respuesta por estrato mostrado en la Tabla 8.
- El tamaño de muestra mínimo por estrato, para la elección de Gobernatura, es de 28 casillas (ver Tabla 8).
- Tomando como referencia la elección de diputados de 2015, en donde en Yucatán se contó con 601 CAE's en total. Con el diseño de muestreo propuesto y las 300 casillas de muestra total, aproximadamente 246 CAE's participarían en el Conteo Rápido y de estos, el 80% tendría que reportar máximo una casilla.
- Armonización del diseño de muestreo para el Conteo Rápido local con el diseño de muestreo para el Conteo Rápido de la elección federal:
 - El diseño de muestreo para el Conteo Rápido de la elección federal será estratificado, en donde:
 - En los estados en donde haya elección local, se usará la misma estratificación que la usada el Conteo Rápido local.
 - En los estados en los que no haya elección local, se tomarán los distritos federales como estratos.

²⁵ Elaborado por el Dr. Carlos Erwin Rodríguez Hernández-Vela con la colaboración del Mtro. David Chaffrey Moreno Fernández

- o La muestra para el Conteo Rápido federal será de 7,500 casillas (más una sobre muestra en los 2 estados con diferencia de 2 horas con la hora del centro. Estas 7,500 casillas, se repartirán de manera proporcional en cada una de las 32 entidades federativas en las que se divide al país.
- o A Yucatán le corresponderían 127 casillas (tomando como referencia los cómputos distritales para diputados 2015), estas se repartirán de manera proporcional, en cada uno de los estratos definidos en la Tabla 5. El tamaño de muestra mínimo en los estratos será de 15 casillas.

Ideas básicas del muestreo probabilístico para el Conteo Rápido

El objetivo particular del Conteo Rápido es estimar el porcentaje, p , de votos en favor de un candidato a Gobernatura²⁶. Esta cantidad es el valor real y se obtiene una vez finalizados los cómputos distritales de la elección. Sin embargo, mediante la teoría del muestreo probabilístico, se selecciona una muestra aleatoria de n casillas, de un total de N instaladas, y con la información recuperada en la tarde-noche del día de la elección se calcula el estimador \hat{p} . Utilizando las herramientas del muestreo, es posible definir **estrategias de selección, tamaños de muestra y estimadores** para asegurar que

$$|p - \hat{p}| \leq \epsilon, (1)$$

con un 95% de confianza. A la cantidad ϵ se le conoce comúnmente como **precisión o error máximo aceptable** en la estimación y se fija de acuerdo con las exigencias de la elección y con la capacidad operativa de campo. La expresión (1) se puede escribir de forma equivalente en términos de intervalos de confianza

$$\hat{p} - \epsilon \leq p \leq \hat{p} + \epsilon . \quad (2)$$

Lo anterior, se pueden interpretar tomando como referencia el siguiente ejemplo: si se extraen 100 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 genera una estimación, de tal manera que se obtengan 100 estimaciones (independientes).

$$\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5, \dots, \hat{p}_{100},$$

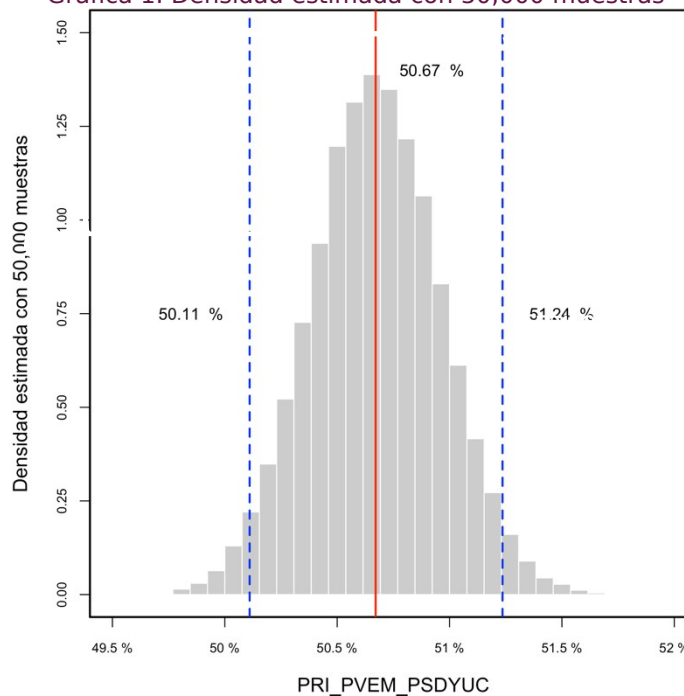
²⁶ Esto se hará para cada contendiente así como para la estimación de la votación en favor de candidatos no registrados, votación nula y participación en la elección; sin embargo, para efectos de la explicación nos concentraremos en la estimación para un solo contendiente.

aproximadamente 95 cumplirán con que su distancia a p , es menor o igual a ϵ .

Precisión en la estimación

La idea anterior nos brinda un procedimiento, mediante simulación Monte Carlo, para obtener la precisión de varias estrategias de selección, con diferentes tamaños de muestra, y así decidir cuál es el que mejor se ajusta a los objetivos y condiciones del Conteo Rápido en el estado de Yucatán. Por ejemplo, para generar la Gráfica 1 se seleccionaron 50,000 muestras aleatorias independientes de casillas de la base de datos de los cómputos distritales para la elección de gobernador en el estado de Yucatán en 2012. Cada muestra fue de 200 casillas y se siguió un muestreo estratificado por distritos locales (15 estratos).

Gráfica 1. Densidad estimada con 50,000 muestras



Con cada una de las 50,000 muestras se calculó el valor del estimador de razón combinado para el muestreo estratificado:

$$\hat{p}_i = \frac{\sum_{h=1}^L N_h \hat{y}_{hi}}{\sum_i \sum_{h=1}^L N_h \hat{y}_{hi}}. (3)$$

En donde i es el candidato i -ésimo (incluyendo también el número de votos por candidatos no registrados y nulos), h es el número de estrato, L el número total de estratos, N_h es el número de casillas instaladas en el estrato h y \bar{y}_{hi} es el promedio de votos, en las casillas seleccionadas en muestra, para el candidato a Gobernatura por el partido i en el estrato h .

Las estimaciones se realizan para todos los candidatos (incluyendo candidatos no registrados y votos nulos); sin embargo, en este apartado sólo se mostrará el análisis para el candidato que registró mayor varianza en su votación y que por lo tanto requería un tamaño de muestra mayor para alcanzar la misma precisión que las estimaciones para los demás candidatos.

Las líneas punteadas (azules), en la Gráfica 1, señalan los cuantiles $q_{0.025}$ y $q_{0.975}$:

- Por debajo del primero se encuentra el 2.5% de las estimaciones y por arriba el 97.5%.
- Por debajo del segundo se encuentra el 97.5% de las estimaciones y por arriba el 2.5%.
- Entre ambos cuantiles se encuentra el 95% de las estimaciones.

El valor real, $p=50.67\%$, se señala con la línea roja, por lo que las 50,000 estimaciones ajustan de forma adecuada este valor.

Regresando a las expresiones (1) y (2), se tiene que

$$q_{0.025} = p - \epsilon \text{ y } q_{0.975} = p + \epsilon.$$

Por lo tanto, la precisión se obtiene mediante la expresión

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

Entonces, con la estratificación y tamaño de muestra para generar la Gráfica 1, se obtiene una precisión de $\epsilon=0.564\%$. Siguiendo este procedimiento se obtuvieron las precisiones para varias estratificaciones y tamaños de muestra.

Estrategias de selección

Existen muchas estrategias de selección, pero de forma general podemos hablar de cuatro que son la base para las demás: el muestreo aleatorio simple sin reemplazo (MASSR), muestreo aleatorio simple con reemplazo (MASCR), el muestreo estratificado y el muestreo por conglomerados o polietápico.

- El MASSR es el muestreo básico y suele usarse como parte de los demás esquemas. Para explicar este tipo de estrategia, se recurre a una urna con N esferas numeradas y en donde el objetivo es seleccionar una muestra de n , con $n < N$. En cada selección, se revuelve la urna y se toma una esfera al azar. La esfera seleccionada se deja afuera de la urna para la siguiente selección (sin reemplazo), en la selección j -ésima quedan $N - (j - 1)$ esferas en la urna.
- El MASCR es muy similar al MASSR, pero después de cada selección, la esfera se regresa a la urna, de forma que en cada extracción se tienen N esferas en la urna. En este caso es necesario registrar el número de la esfera seleccionada.
- En el muestreo estratificado se divide a la población en estratos/grupos²⁷ y en cada estrato se hace una selección de casillas.²⁸ En este caso se tiene mayor control de la muestra, ya que en cada estrato siempre se tendrán casillas seleccionadas, logrando una mejor distribución de la misma. Otra ventaja, es que permite obtener una mejor estimación: si los estratos se definen de manera que, al interior de cada estrato, la variable de interés tenga un comportamiento más homogéneo que si se considerara la población completa, entonces la estimación tendrá más precisión que el MASSR o el MSCR.
- Por último, se tiene el muestreo por conglomerados, en donde también se divide a la población en conglomerados/grupos y la selección se realiza en etapas, primero se seleccionan conglomerados (generalmente mediante MASCR) y posteriormente, y sólo de los conglomerados seleccionados en la primera etapa, se eligen casillas (generalmente mediante MSSR). El muestreo por conglomerados se usa para reducir costos y facilitar la logística operativa de campo; sin embargo, se sacrifica precisión en la estimación.

Por lo descrito en el párrafo anterior, lo que se hizo para determinar el tamaño de muestra en para el Conteo Rápido del estado de Yucatán fue probar distintas estratificaciones y tamaños de muestra eligiendo la combinación que permitiera obtener la mejor precisión, pero cuidando no rebasar la capacidad operativa de campo. Además, también se consideró que el tamaño de muestra, en cada estrato, fuera mayor a un cierto valor. Lo anterior para aumentar la probabilidad de que llegue información en todos los estratos.

Bases de datos de referencia

²⁷ los estratos son ajenos entre sí y su unión conforma la población completa

²⁸ La selección en cada estrato usualmente se realiza mediante un MASS.

Se tomaron tres bases de datos como referencia para los análisis que se presentarán en los siguientes apartados:

- **Cómputos distritales de la elección para gobernador en el estado de Yucatán del 1 de julio de 2012.** Esta base de datos se usó para estimar las precisiones y cuenta con información de la votación en 2,391 casillas. Para esta elección se contó con 475 CAE's.
- Es importante mencionar que el resultado de esa elección no es relevante para estimar el tamaño de muestra. Lo que se busca es tener una estimación de la varianza de la votación, que, junto con la precisión y el nivel de confianza, determina el tamaño de muestra.
- **Cómputos distritales para la elección de diputados de 2015, considerando sólo a Yucatán.** Esta base de datos se usó para estimar la presión sobre los CAE's. Cuenta con información de la votación para 2,539 casillas. En esta elección, sólo en Yucatán, se contó con 601 CAE's.
- **Muestra recibida para realizar la estimación para el Conteo Rápido en la elección de diputados de 2015, considerando sólo la parte correspondiente a Yucatán.** Esta base de datos se usó para medir el nivel de respuesta de un Conteo Rápido previo en la misma entidad.

Estratificaciones y tamaños de muestra

Las estratificaciones consideradas en los ejercicios de simulación, fueron 7:

1. Sin estratos (**MASSR**).
2. Tipo sección (**TS**) \Rightarrow urbana y no urbana
3. Distritos federales (**DF**) \Rightarrow 5 estratos.
4. Distritos federales y al interior de cada distrito se dividió por sección urbana no urbana (**DF \cap TS**) \Rightarrow 10 estratos.
5. Distritos locales (**DL**) \Rightarrow 15 estratos.
6. Distritos locales y al interior de cada distrito se dividió por sección urbana y no urbana (**DL \cap TS**) \Rightarrow 27 estratos.
7. Intersección entre distritos federales y locales (**DF \cap DL**) \Rightarrow 21 estratos.

Para distribuir el tamaño de muestra en cada uno de los estratos, se utilizó lo que se conoce como “afijación proporcional”:

$$n_h = n \frac{N_h}{N},$$

en donde n_h es el tamaño de muestra que se seleccionará en el estrato h , n el tamaño de muestra total, N_h el número total de casillas instaladas en el estrato h y N el número total de casillas instaladas. En palabras comunes, el tamaño de muestra, se distribuirá proporcionalmente de acuerdo con el número de casillas instaladas en cada estrato. En estratos con mayor número de casillas se seleccionará mayor muestra.

Resultados de las simulaciones

Lo que se busca es la estratificación que alcance la mejor precisión (o error máximo aceptable), con el menor tamaño de muestra posible. En todas las simulaciones se utilizó el estimador de razón dado por la expresión (3), los resultados obtenidos, se muestran en la Tabla 1 (para obtener las precisiones en cada celda de la tabla se seleccionaron 50,000 muestras independientes y con cada muestra se hizo una estimación, ver Gráfica 1).

Tabla 1. Elección para gobernador 2012: precisiones obtenidas usando el estimador (3), para distintos tamaños de muestra y estratificaciones.

N	MASSR	TS	DF	DF n TS	DL	DL n TS	DF n DL
175	1.094	1.102	1.073	1.085	1.033	1.028	1.030
200	1.021	1.018	1.005	0.996	0.954	0.946	0.950
225	0.949	0.961	0.948	0.944	0.894	0.891	0.892
250	0.902	0.895	0.888	0.884	0.849	0.841	0.855
275	0.857	0.854	0.837	0.841	0.802	0.808	0.795
300	0.807	0.817	0.799	0.796	0.765	0.763	0.771
325	0.780	0.776	0.765	0.764	0.727	0.729	0.732
350	0.740	0.747	0.733	0.734	0.699	0.696	0.695
375	0.710	0.714	0.698	0.698	0.673	0.665	0.668
400	0.691	0.687	0.680	0.671	0.646	0.639	0.640
425	0.660	0.662	0.650	0.643	0.624	0.620	0.624
450	0.643	0.637	0.628	0.630	0.599	0.604	0.596
475	0.617	0.619	0.609	0.604	0.582	0.579	0.580
500	0.600	0.603	0.586	0.592	0.561	0.564	0.559

Las precisiones se calculan para la coalición PRI-PVEM-PSDYUC, que fue la que tuvo mayor varianza.

En la Tabla 1, claramente se observa que, al aumentar el tamaño de muestra o estratificaciones con mayor número de estratos, la precisión aumenta. Sin embargo, estratificaciones muy finas conllevan a un problema práctico. Es factible que haya estratos con muy pocas casillas, por lo que la muestra en esos estratos tendrá que ser pequeña y por lo tanto existirá un gran riesgo de que al final, considerando la no respuesta, no llegue información de esos estratos. En la Tabla 2 se muestran las estadísticas básicas para el número de casillas por estrato de cada estratificación considerada.

Tabla 2. Cómputos distritales elección de gobernador 2012: Estadísticas básicas para cada estratificación y las casillas instaladas en cada estrato.

ESTRATIFICACIÓN	NÚMERO DE ESTRATOS	MÍNIMO DE CASILLAS X ESTRATO	PROMEDIO DE CASILLAS X ESTRATO	MÁXIMO DE CASILLAS X ESTRATO
TS	2	663	1,195.5	1,728
DF	5	448	478.2	532
DF ∩ TS	10	30	239.1	502
DL	15	120	159.4	227
DL ∩ TS	27	18	88.6	227
DF ∩ DL	21	13	113.9	196

Las estratificaciones que tienen más de 15 estratos representan un riesgo. Ya que se tienen estratos con muy pocas casillas instaladas, aunque se defina una muestra grande (proporcionalmente), la probabilidad de que no se recupere ninguna casilla en estos estratos es muy alta. Este riesgo se reduce de manera sustancial, si se toman los **distrito locales** como estratos, en este caso el mínimo de casillas instaladas es de 120, el máximo es de 227 y en promedio se instalan 159 casillas por distrito local. Además, también se alcanza una buena precisión con un tamaño de muestra aceptable.

Presión sobre los CAE's

Los CAE's son los encargados de recolectar la información de la votación de las casillas seleccionadas en muestra. Debido a que la elección de 2018 será la más grande en la historia de nuestro país, poco menos de 3,400 cargos públicos estarán en juego y a que los CAE's tienen que realizar muchas actividades el día de la Jornada Electoral, no sólo enviar la muestra para el Conteo Rápido, se tiene que plantear una estrategia para no sobrecargar de trabajo a los CAE's.

Para no saturar a los CAE's, el COTECORA acordó establecer un tamaño de muestra tal que, del total de CAE's necesarios para recolectar la muestra del Conteo Rápido, alrededor del 80 % reporte 1 casilla como máximo. Esta es una restricción para evitar tamaños de muestra muy grandes, con respecto al número total de CAE's. Así, disminuyendo los casos en los que un solo CAE reporte 2 o más casillas. En la Tabla 3 se presenta el porcentaje de CAE's con 1 casilla considerando un diseño de muestra estratificado por distrito local.

Tabla 3. Porcentaje de CAE's con 1 casilla considerando un diseño de muestra estratificado por distrito local

TAMAÑO DE MUESTRA	TOTAL DE CAE'S QUE TENDRÍAN QUE APOYAR AL CONTEO RÁPIDO	PORCENTAJE DE CAE'S CON 1 CASILLA
175	155	88.9
200	176	87.1
225	194	85.4
250	211	83.7
275	229	82
300	246	80.2
325	262	78.5
350	278	76.8
375	292	75.2
400	306	73.7
425	320	72
450	333	70.4
475	345	68.9
500	358	67.2

Al aumentar el tamaño de muestra, el número de CAE's que tendría que apoyar al Conteo Rápido iría aumentando, además, el porcentaje de CAE's con 1 casilla decrece. El tamaño de muestra, con el mayor número de casillas y que cumple con la restricción es de **300 casillas**.

Armonización del diseño de muestreo local con el diseño de muestreo para la elección Presidencial

En el Conteo Rápido 2018 para la elección Presidencial, se definió que en los estados en donde habrá elección local, la muestra federal será una submuestra de la muestra para la elección local. Siguiendo exactamente la

misma estratificación que en la local. En los estados en los que no habrá elección local, se tomarán los distritos federales como estratos.

El tamaño de muestra será de 7,500 casillas repartidas proporcionalmente en cada uno de los 32 estados. Más una sobre muestra considerada únicamente para los 2 estados con huso horario con diferencia de 2 horas, con respecto a la hora del centro. De estas 7,500 casillas a Yucatán le corresponderían 127 casillas. Al repartir proporcionalmente las 127 casillas, entre los 15 distritos locales de Yucatán, en algunos estratos se obtienen tamaños de muestra de 7 casillas. Esto también representa un riesgo ya que la probabilidad de que en estos estratos, con tamaños de muestra tan pequeños, no llegue información para el Conteo Rápido de la elección presidencial es alta. En la Tabla 4, se muestra la distribución de las 127 casillas entre los 15 distritos locales.

Tabla 4. Distribución de muestra por distrito local para la elección de Gobernatura y para la elección Presidencial

DISTRITO LOCAL	MUESTRA ELECCIÓN PRESIDENCIAL
1	10
2	9
3	8
4	9
5	9
6	7
7	8
8	8
9	9
10	8
11	7
12	8
13	8
14	8
15	10
Total	127

Para resolver el problema anterior, se decidió construir una nueva estratificación, a partir de los distritos locales. La estrategia fue colapsar distritos locales para lograr una buena precisión, pero además evitando

estratos con un tamaño de muestra pequeño en el caso del Conteo Rápido federal.

La mejor estratificación se obtuvo mediante la técnica de optimización numérica conocida como **recocido simulado**. La idea detrás de este procedimiento se basa en:

1. Fijar un tamaño de muestra específico, se trabajó con $n=200$, 225 y 250.
2. Se fija el número de estratos que se desean, en nuestro caso buscamos 7 estratos. Entonces pasaríamos de 15 estratos a menos de la mitad.
3. Se inicializa el algoritmo con una estratificación cualquiera, de 7 estratos. Esta será la estratificación actual “inicial”.
4. Se propone una estratificación nueva y se compara con la estratificación actual. Si la nueva estratificación es mejor que la actual (de acuerdo con algún criterio), entonces el algoritmo se queda con la nueva estratificación. En caso contrario, se mantiene la estratificación actual.
5. El paso 4 se repite un número muy grande de iteraciones. La estratificación final es la óptima.

El criterio para comparar estratificaciones se basó en las precisiones asociadas a cada estratificación, así como a una penalización por estratos con pocas casillas. Además, se realizaba una verificación para constatar continuidad geográfica.

Siguiendo esta estrategia, se encontró que la mejor estratificación se obtiene colapsando los distritos locales como se describe en la Tabla 5.

Tabla 5. Estratificación óptima colapsando distritos locales y muestra para la elección Presidencial.

ESTRATO	DISTRITOS LOCALES	MUESTRA ELECCIÓN PRESIDENCIAL
1	12 13	16
2	7 1	18
3	6 9	15
4	10	17

ESTRATO	DISTRITOS LOCALES	MUESTRA ELECCIÓN PRESIDENCIAL
	15	
5	8	16
	3	
6	11	16
	14	
7	4	29
	5	
	2	
Total		127

Con esta nueva estratificación, al distribuir las 127 casillas para la elección Presidencial, el estrato con el menor tamaño de muestra tiene 15 casillas. Esto representa un poco más del doble que cuando se consideran los distritos locales.

El porcentaje de CAE's con 1 casilla considerando un diseño de muestra estratificado, con estratos definidos de acuerdo con la Tabla 5, se presenta en la Tabla 6.

Tabla 6. Porcentaje de CAE's con 1 casilla considerando un diseño de muestra estratificado con los estratos definidos en la Tabla 5

TAMAÑO DE MUESTRA	TOTAL DE CAE'S QUE TENDRÍAN QUE APOYAR AL CONTEO RÁPIDO	PORCENTAJE DE CAE'S CON 1 CASILLA
175	156	88.3
200	175	86.6
225	193	85
250	211	83.2
275	228	81.5
300	244	79.9
325	261	78.2
350	276	76.6
375	291	75
400	304	73.4
425	319	71.7

TAMAÑO DE MUESTRA	TOTAL DE CAE'S QUE TENDRÍAN QUE APOYAR AL CONTEO RÁPIDO	PORCENTAJE DE CAE'S CON 1 CASILLA
450	332	70.1
475	344	68.6
500	356	66.9

Se sigue manteniendo el tamaño de muestra de 300 casillas como un buen candidato para lograr que alrededor del 80 % de los CAE's, que apoyarán al Conteo Rápido, reporte 1 casilla como máximo.

En la Tabla 7 se comparan las precisiones obtenidas usando la estratificación por distrito local y la estratificación definida en la Tabla 5, para distintos tamaños de muestra.

Tabla 7. Elección para gobernador 2012: precisiones obtenidas usando el estimador (3), para distintos tamaños de muestra y comparando la estratificación por distrito locales y la definida en la Tabla 5

TAMAÑO DE MUESTRA	DL	ESTRATIFICACIÓN ÓPTIMA CON 7 ESTRATOS
175	1.033	1.045
200	0.954	0.961
225	0.894	0.906
250	0.849	0.860
275	0.802	0.810
300	0.765	0.775
325	0.727	0.740
350	0.699	0.705
375	0.673	0.678
400	0.646	0.653
425	0.624	0.633
450	0.599	0.606
475	0.582	0.592
500	0.561	0.572

Se aprecia que aunque la estratificación por distrito local alcanza mejor precisión, la diferencia es poco significativa. Entonces, con 7 estratos se

estaría obteniendo un nivel de error muy aproximado al que se obtendría con 15.

No respuesta o casillas no reportadas

En cualquier encuesta existe un porcentaje de no respuesta, en el caso del Conteo Rápido serían casillas seleccionadas en muestra, pero que por diversos motivos su información no se recibe a tiempo para hacer la estimación. En el Conteo Rápido para la elección federal de diputados de 2015, la muestra en el estado de Yucatán fue de 150 casillas (30 casillas por cada uno de los 5 distritos federales). De las 150 casillas, alrededor de las 10 pm se reportó información de 102 casillas para el Conteo Rápido, lo que representa el 68%.

Si el mismo comportamiento se mantiene para el Conteo Rápido en la elección de Gubernatura en 2018. Para realizar la estimación final alrededor de las 10 pm, con una muestra efectiva de 205 casillas, se necesitaría un tamaño de muestra total, considerando la no respuesta, de 300 casillas. En este caso se calcula el nivel de respuesta por estrato para asignar mayor sobre muestra a los estratos que tuvieron menor nivel de respuesta. La expresión para asignar la sobre muestra es la siguiente.

$$n_{Total} = \frac{n_{efectiva}}{Tasa\ de\ respuesta}$$

La sobre muestra, así como la muestra total, por estrato, se despliega en la Tabla 8.

Tabla 8. Muestra total y sobre muestra por estrato

ESTRATO	TOTAL DE CASILLAS INSTALADAS EN LA ELECCIÓN DE 2015	TAMAÑO DE MUESTRA EFECTIVO	RESPUESTA OBSERVADA EN EL CONTEO RÁPIDO DE 2015	SOBRE MUESTRA	MUESTRA TOTAL
1	320	26	48.1%	28	54
2	350	28	68.2%	13	41
3	306	25	63.6%	14	39
4	349	28	78.3%	8	36
5	316	26	64.7%	14	40
6	313	25	90.9%	3	28
7	585	47	75.0%	16	63
	2,539	205		95	300

Con 300 casillas de muestra total y la estratificación presentada en la Tabla 5, alrededor de las 10 pm del 1 de julio, se esperaba contar con un tamaño de muestra efectivo de $n=205$ casillas (después de la no respuesta). Con el cual, se alcanzaría una precisión para las estimaciones de 0.96 %. (ver Tabla 7).

Consideraciones finales

Por último, es importante tomar en cuenta los siguientes puntos

- La precisión 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 capaz (teóricamente) de alcanzar el nivel de precisión y confianza deseadas en la estimación. Sin embargo, debido a que este tamaño de muestra se fija tomando como referencia los resultados de una o varias elecciones previas, una vez concluida la elección y con toda la información disponible se determina la precisión realmente obtenida.
- Los resultados de las simulaciones presentadas son únicamente para el método de estimación usado (ecuación 3), y tienen como objetivo fijar un tamaño de muestra. Sin embargo, se pueden emplear métodos de estimación alternativos (usando la misma estrategia de selección) que arrojarán precisiones distintas.

5.2.2. Procedimiento de estimación

5.2.2.1. Dr. Manuel Mendoza Ramírez

Con base en la distribución de las actividades del COTECORA, establecidas en el apartado 2.2, el Dr. Manuel Mendoza Ramírez en colaboración con el Lic. Gian Carlo Di-Luvi Martínez realizarán las estimaciones de las tendencias de la votación y el porcentaje de votación de los estados de **Chiapas** y **Guanajuato**.

Para tal efecto emplearán el procedimiento descrito en el apartado 5.1.2.2.1.

5.2.2.2. Dr. Carlos Hernández Garciadiego

Con base en la distribución de las actividades del COTECORA, establecidas en el apartado 2.2, el Dr. Carlos Hernández Garciadiego en colaboración con el Mtro. Emiliano Geneyro Squarzon realizarán las estimaciones de las tendencias de la votación y el porcentaje de votación de la **Ciudad de México, Veracruz y Yucatán.**

Para tal efecto emplearán el procedimiento descrito en el apartado 5.1.2.3.1.

5.2.2.3. Dra. Michelle Anzarut Chacalo²⁹

El modelo utilizado para estimar la proporción de votos a favor de cada uno de los candidatos a la Gobernatura de Guanajuato es un modelo de postestratificación con regresión multinivel (Park, Gelman, and Bafumi 2004), que es un método Bayesiano para corregir diferencias conocidas entre la muestra y las poblaciones objetivo. La idea central es dividir a la población en celdas basadas en combinaciones de diversos atributos, usar la muestra para estimar la variable de respuesta dentro de cada celda, y finalmente agregar las estimaciones de cada celda para estimar a nivel poblacional, ponderando cada celda por su proporción relativa en la población.

En este escenario, el primer objetivo es estimar el total de votos a favor de cada candidato, así pues, para cada candidato, el método se compone de dos pasos:

1. Para cada casilla se estima el número de votos en favor del candidato utilizando una regresión multinivel en función de ciertas covariables. Con esto, se estima una respuesta promedio θ_l para cada combinación de covariables.
2. Ponderamos la respuesta promedio θ_l de cada combinación de covariables para obtener la respuesta promedio de la población total, θ . Supongamos que tenemos L categorías de covariables, entonces

$$\theta = \frac{\sum_{l=1}^L N_l \theta_l}{\sum_{l=1}^L N_l},$$

donde $N_l, l=1, \dots, L$, indica el número de casillas en cada categoría.

Una regresión multinivel es una regresión en la que los parámetros reciben un modelo de probabilidad. Este modelo de segundo nivel tiene sus propios parámetros, conocidos como hiperparámetros, que también se estiman a partir de los datos. La ganancia principal de utilizar regresión multinivel está en que se pueden estudiar los efectos que varían según la celda. Adicionalmente, ante la falta de respuesta es posible ajustar con mayor exactitud, ya que la regresión atrae los parámetros de un grupo hacia la media grupal, lo hará en mayor medida cuando la desviación estándar del grupo sea pequeña o suavizará para grupos con pocas observaciones.

Las covariables utilizadas son:

²⁹ Con la colaboración de la Mtra. Maria Teresa Ortiz Mancera

1. Tipo de sección (rural o urbana/mixta).
2. Tipo de casilla (básica/contigua/especial o extraordinaria).
3. Tamaño de sección (chica que es menor a 1,000 votantes en la lista nominal, mediana, de 1,000 a 5,000 votantes o grande, con más de 5,000).
4. Región (oriente u occidente).
5. Distrito local.

Además de cada una de las variables, es natural considerar la interacción de tipo de sección con tamaño de sección, pues la relación entre el tipo de sección y el número de votos por un candidato es diferente en secciones de distinto tamaño.

Cuando las variables tienen menos de tres categorías, no hay ninguna ganancia en usar un modelo multinivel contra usar distribuciones iniciales no informativas (véase, por ejemplo, Gelman 2003), por lo que, por simplicidad, se modelarán las variables 1) a 4) como coeficientes de regresión sin estructura multinivel.

Supongamos que se quiere estimar primero el número total de votos para uno de los candidatos, por lo que en lo que sigue se omitirá el superíndice que indicaría el candidato. Sea X_k el número de votos en favor del candidato en la k -ésima casilla, en el primer nivel el modelo asume que:

$$X_k \sim \begin{cases} T(\mu_k, \tau_k, \nu_k) I_{[0,750]} & , \text{ si } b_k = 1 \\ 0 & , \text{ si } b_k = 0 \end{cases}$$

donde $\mu_k = n_k \theta_k$, $\tau_k = n_k^{-1} \tau_k^{\text{distrito}}$ y $\nu_k = \nu_k^{\text{distrito}}$.

La distribución $T(\mu_k, \tau_k, \nu_k)$ se refiere a la distribución t de Student no estandarizada, donde μ_k es el parámetro de localización, τ_k es el parámetro de precisión y ν_k son los grados de libertad, la función indicadora $I_{[0,750]}$ vale uno si el valor se encuentra en el intervalo $[0,750]$ (que son los posibles valores que puede tomar X_k) o cero en otro caso, el término n_k es el tamaño de la lista nominal de la casilla, θ_k representa la proporción de personas en la lista nominal de la casilla k que votaron por el candidato y b_k es una variable aleatoria que toma los valores cero o uno dependiendo de la probabilidad de votos del candidato dentro de la casilla, es decir,

$$b_k \sim \text{Bernoulli}(p_k).$$

Tanto para la proporción θ_k , como para el parámetro p_k , se ajusta un modelo de regresión multinivel

$$\beta^0 + \beta^{rural} \cdot rural_k + \beta^{rural_{tamañoMd}} \cdot rural_k \cdot tamañoMd_k + \zeta$$

$$\theta_k = \text{logit}^{-1} \zeta$$

$$\beta^{tamañoMd} \cdot tamañoMd_k + \beta^{tamañoGd} \cdot tamañoGd_k + \beta_{distrito(k)}^{distrito} + \beta^{tipoEx} \cdot tipoEx_k \zeta,$$

$$p_k = \text{logit}^{-1} \left(\beta_p^0 + \beta_p^{rural} \cdot rural_k + \beta_p^{rural_{tamañoMd}} \cdot rural_k \cdot tamañoMd_k \right)$$

$$\beta_p^{tamañoMd} \cdot tamañoMd_k + \beta_p^{tamañoGd} \cdot tamañoGd_k + \beta_{distrito(k)_p}^{distrito_p} + \beta_p^{tipoEx} \cdot tipoEx_k \zeta,$$

donde $rural_k$ indica si la casilla es rural, $distrito$ indica el distrito local y $tamaño_k$ indica la clase de tamaño de la sección a la que pertenece la casilla.

Finalmente, ajustamos un modelo multinivel al nivel distrito local:

$$\beta_j^{distrito} \sim N \left(\mu^{distrito}, \sigma_{distrito}^2 \right),$$

$$\beta_j^{distrito_p} \sim N \left(\mu_p^{distrito}, \sigma_{distrito_p}^2 \right).$$

Al añadir predictores a nivel distrito local se reduce la variación no explicada dentro de cada distrito local y, como consecuencia, se reduce la variación total, produciendo estimaciones más precisas. El resto de las variables se estiman de los datos dadas distribuciones iniciales no informativas.

Se implementa el modelo usando el software bayesiano JAGS: Just Another Gibbs Sampler, llamado desde R (R Development Core Team 2003). El método de estimación es completamente reproducible y puede encontrarse en el paquete de R 'quickcountmx' (Ortiz 2017).

Ejecutando el modelo anterior para todos los posibles candidatos se estima la distribución de probabilidad del número total de votos por cada candidato en cualquier casilla del estado. Con esta distribución, es directo estimar la distribución de la proporción de votos a favor de cada candidato, la proporción *efectiva* de votos a favor de cada candidato, que excluye la categoría de abstenciones, y la distribución de la participación ciudadana. Los intervalos de credibilidad presentados como resultado serán una evaluación simple de estas distribuciones, es decir, cada intervalo I es una región de 95% de credibilidad para ψ si $P(\psi \in I) = 0.95$. Para esto se tomaron 1,000 observaciones simuladas de las distribuciones y se calcularon los cuantiles de orden 0.025 y 0.975.

Referencias

- Gelman, Andrew. 2003. “Bugs. R: functions for calling Bugs from R”. <http://www.stat.columbia.edu/~gelman/BugsR>.
- Ortiz, Teresa. 2017. “quickcountmx: an R package to estimate election results from Mexico.” <https://github.com/tereom/quickcountmx>.
- Park, David K, Andrew Gelman, and Joseph Bafumi. 2004. “Bayesian multilevel estimation with poststratification: state-level estimates from national polls.” *Political Analysis* 12 (4). Oxford University Press: 375–85.

5.2.2.4. Dr. Raúl Rueda Díaz del Campo³⁰

Estimación de las tendencias de la votación.

En ambos conteos, nacional y local de Jalisco, el objetivo es estimar el número de votos totales emitidos a favor de cada partido. Para esto, la población se divide en S subpoblaciones y en cada una de ellas, se estima este total.

Para esto, suponiendo que se tienen M casillas³¹, distribuidas en S subpoblaciones, de manera que

$$M = \sum_{i=1}^S M_i,$$

con M_i el total de casillas contenidas en la i -ésima subpoblación $\forall i \in J_S$.

La información relevante, es la votación emitida en cada casilla a favor de cada uno de p partidos³². Así, sea

$$X_{ij} = (X_{ij1}, \dots, X_{ijp}) \in R^p, j \in J_{M_i}, i \in J_S,$$

el total de votos en urna en la j -ésima casilla de la i -ésima subpoblación.

El objetivo es estimar la tendencia en la votación el día de la Jornada Electoral, basándose en una muestra de casillas elegidas al azar en cada una de las S subpoblaciones.

Es decir, observando

$$x_{ij} = (x_{ij1}, \dots, x_{ijp}) \in R^p, j \in J_{m_i}, i \in J_S.$$

Suponiendo que los resultados **entre** subpoblaciones son independientes y **dentro** de cada una, podemos modelar como si fueran parte de una sucesión infinita de variables intercambiables.

Esto es, suponiendo que

30 Con la colaboración del Lic. Javier Santibáñez Cortés

31 En una elección presidencial, M es el total de casillas a nivel nacional; mientras que en una elección local, M es el total de casillas en el estado.

32 Incluimos los votos a candidatos no registrados y votos nulos. Es decir, **todos** los votos que son depositados en la urna correspondiente.

$$p(X|D) = \prod_{i=1}^S \int_{\theta_i} p(x_{i_1}^i, \dots, x_{i_{m_i-m_i}}^i | \theta_i) p(\theta_i | x_{i_1}, \dots, x_{i_{m_i}}) d\theta_i,$$

donde $X = \{X_{ij} : j \in J_{M_i}, i \in J_S\}$, $p(\theta_i | x_{i_1}, \dots, x_{i_{m_i}})$ es la distribución de θ_i dada la muestra $s_{m_i} = \{x_{i_1}, \dots, x_{i_{m_i}}\}$, $\forall i \in J_S$, y $p(x_{i_1}^i, \dots, x_{i_{m_i-m_i}}^i | \theta_i)$ es la "verosimilitud" de lo no observado en cada subpoblación, $i \in J_S$.

Suponiendo que

$$X_{ij} \sim \perp N_p(X_i | \theta_i, H_i), \forall i \in J_S \quad \text{y} \quad \pi(\theta_i, H_i) \propto |H_i|^{-\frac{v}{2}}, v \in \mathbb{R}^+$$

y entonces

$$p(\hat{X}_k | \theta_i, H_i) = N_p(\hat{X}_k | \theta_i, k H_i), \quad \text{con} \quad \hat{X}_k = \frac{1}{k} \sum_{j=1}^k X_{ij}$$

y

$$p(\theta_i, H_i | S_{m_i}) \propto |H_i|^{\frac{m_i-v}{2}} \exp\left\{-\frac{m_i}{2} (\hat{x}_i - \theta_i)' H_i (\hat{x}_i - \theta_i) - \frac{m_i}{2} \text{tr}(S_i H_i)\right\}$$

en donde

$$\hat{x}_i = \frac{1}{m_i} \sum_{j=1}^{S_i} x_{ij} \quad \text{y} \quad S_i = \frac{1}{m_i} \sum_{j=1}^{m_i} (x_{ij} - \hat{x}_i)(x_{ij} - \hat{x}_i)', \quad \forall i \in J_S.$$

Usando las ecuaciones anteriores, puede demostrarse que

$$p(\hat{X}_i^i | s_{m_i}) = \text{st } u_p\left(\hat{X}_i^i | m_i - v + 1, \hat{x}_i, \frac{(m_i - v + 1)(M_i - m_i)}{M_i} S_i^{-1}\right).$$

Si $X_{i\cdot} = \sum_{j=1}^{M_i} X_{ij}$ es el vector total de votos emitidos en la i -ésima subpoblación, entonces

$$X_{i\cdot} = M_i \hat{X}_i = m_i \hat{x}_i + (M_i - m_i) \hat{X}_i^i, \forall i \in J_S.$$

por lo que

$$X_{\square} = \sum_{i=1}^S X_{i\cdot}.$$

será el total de votos a cada partido en la población. Así, lo único que se necesita estimar es $\hat{X}_i^i, \forall i \in J_S$.

Recordando que $X_{i\cdot} = (X_{i\cdot 1}, \dots, X_{i\cdot p})$ es el vector total de votos emitidos en cada subpoblación, es decir

$$X_{i\cdot q} = \sum_{j=1}^{M_i} X_{ijq},$$

con X_{ijq} el total de votos emitidos a favor del partido q ($q \in J_p$) en la casilla j ($j \in J_{M_i}$) de la subpoblación i ($i \in J_S$), por lo que

$$X_{i\cdot\cdot} = \sum_{q=1}^p X_{i\cdot q},$$

es el total de votos emitidos en la subpoblación i -ésima, para toda $i \in J_n$.

El procedimiento se basa en simular muestras de $p(X_{\cdot\cdot} \vee D)$, donde $D = \{s_{m_1}, \dots, s_{m_S}\}$, es la información de las muestras en cada subpoblación. Esto se logra con el siguiente algoritmo:

Dada la muestra s_{m_i} de la subpoblación i -ésima

Calcular los estadísticos suficientes de la muestra \hat{x}_i y S_i .

Generar una muestra $\{\hat{X}_i^1, \dots, \hat{X}_i^n\}$ de $p(\hat{X}_i^i | s_{m_i})$.

Generar una muestra $\{X_i^1, \dots, X_i^n\}$ de $p(X_i \cdot | s_{m_i})$.

Repetir (i) para cada $i \in J_S$, con lo que obtenemos

$$\{X_{1\cdot}^l, \dots, X_{S\cdot}^l\}, l \in J_n.$$

Generar una muestra de $p(X_{\cdot\cdot} \vee D)$

$$X_{\cdot\cdot}^l = \sum_{i=1}^S X_{i\cdot}^l, l \in J_n.$$

Finalmente, hacemos $y^l = \sum_{q=1}^p X_{\cdot\cdot q}^l$, la suma total de votos emitidos a nivel poblacional para todos los partidos y calculamos

$$\theta^l = \frac{1}{y^l} X^l, l \in J_n,$$

la proporción de votos a nivel poblacional para cada partido. La participación se calcula como la proporción de votos emitidos con respecto a la lista nominal

$$\tau^l = \frac{y^l}{\ln}, l \in J_n,$$

donde \ln representa la lista nominal nacional.

De esta forma obtenemos una muestra de tamaño n de la distribución final de las proporciones de votos a cada partido y de la participación a nivel nacional o local, según el caso. El cuantil 0.50, la mediana, es un estimador puntual natural de la verdadera proporción y los cuantiles (0.025, 0.975), estiman un intervalo de probabilidad 0.95.

5.2.2.5. Mtra. Patricia Isabel Romero Mares

Con base en la distribución de las actividades del COTECORA, establecidas en el apartado 2.2, la Mtra. Patricia Isabel Romero Mares en colaboración con la Mtra. Adriana Ducoing Watty realizarán las estimaciones de las tendencias de la votación y el porcentaje de votación de los estados de **Jalisco, Puebla y Tabasco**.

Para tal efecto emplearán el procedimiento descrito en el apartado 5.1.2.1.1.

5.2.2.6. Dr. Alberto Alonso y Coria³³

Procedimiento clásico para la estimación del resultado de la Jornada Electoral.

Para estimar el porcentaje de participación electoral y el porcentaje de votos que cada candidato a la Gubernatura de Tabasco obtendrá la próxima Jornada Electoral, se utilizará el método estadístico clásico vía estimadores de razón, utilizándose para ello un muestreo aleatorio estratificado. En este procedimiento, primero se procede a dividir al estado o país en estratos, en cada uno de los cuales se genera una muestra aleatoria simple de casillas y es sobre los resultados obtenidos en esta muestra que se realiza el procedimiento de estimación.

a) Estimación de razón separada.

La principal característica de este procedimiento es el tratar primero cada estrato en forma independiente, obteniendo tanto el número de votos por cada candidato como su varianza, para posteriormente realizar una suma ponderada y obtener así la estimación final y el intervalo de confianza asociado.

Como primer paso se estima el número de votos que obtendría cada una de las fuerzas políticas. El análisis se realiza primero para un determinado estrato h . Si $v_{i,j,h}$ representa la votación por el candidato i , en la casilla j del estrato h , $\hat{v}_{i,h}$ su promedio en el estrato y $\ln_{j,h}$ al listado nominal de esa casilla, el estimador de razón de votación para el candidato i estaría dada por:

$$\hat{R}_{i,h,S} = \frac{\sum_{j=1}^{n_h} v_{i,j,h}}{\sum_{j=1}^{n_h} \ln_{j,h}} .$$

En donde n_h es el número de casillas en la muestra que pertenecen al estrato h .

Una estimación puntual del número total de votos que obtendría un partido, alianza o candidato independiente estaría dada por:

$$\hat{Y}_{i,R_s} = \sum_h \hat{R}_{i,h,S} \ln_h .$$

En donde \ln_h representa la totalidad del listado nominal del estrato.

³³ Con la colaboración del Fis. Joaquín Morales Bolio

Para la obtención del estimador de la varianza y de ahí poder inferir la precisión obtenida, se calcula:

$$v(\hat{Y}_{i,R_s}) = \sum_h \frac{N_h^2(1-f_h)}{n_h} \left[\sum_{j=1}^{n_h} \frac{(v_{i,j,h} - \hat{v}_{i,h} - \hat{R}_{i,h,S}(\ln_{i,j,h} - \ln'_{i,h}))^2}{n_h - 1} \right].$$

Los intervalos de confianza para el número de votos obtenidos por el partido i estarían entonces dados por:

$$\text{Valor m\u00ednimo}_i = \hat{Y}_{i,R_s} - 1.96 \sqrt{v(\hat{Y}_{i,R_s})}$$

$$\text{Valor m\u00e1ximo}_i = \hat{Y}_{i,R_s} + 1.96 \sqrt{v(\hat{Y}_{i,R_s})}$$

b) Estimaci\u00f3n de raz\u00f3n combinada.

En raz\u00f3n de la posibilidad de tener algunos estratos poco poblados en la muestra obtenida, se realiza una segunda estimaci\u00f3n del n\u00famero de votos recibidos por los partidos, pero ahora con el esquema combinado. Esto es, se combinan los estratos para construir un solo estimador de raz\u00f3n y este estar\u00eda dado por:

$$\hat{R}_{i,C} = \frac{\sum_h N_h \hat{v}_{i,h}}{\sum_h N_h \ln'_h}.$$

Con una estimaci\u00f3n puntual de: $\hat{Y}_{i,R_s} = \hat{v}_{i,h}$ y $\hat{R}_{i,C} \ln$. La notaci\u00f3n es semejante al caso anterior, con la observaci\u00f3n de que N_h representa al n\u00famero de casillas del estrato h y \ln el total del listado nominal.

En este caso y en forma muy semejante al caso anterior, la varianza estar\u00eda dada por:

$$v(\hat{Y}_{i,R_c}) = \sum_h \frac{N_h^2(1-f_h)}{n_h} \left[\sum_{j=1}^{n_h} \frac{(v_{i,j,h} - \hat{v}_{i,h} - \hat{R}_{i,C}(\ln_{i,j,h} - \ln'_{i,h}))^2}{n_h - 1} \right].$$

Los intervalos de confianza se generar\u00edan en forma semejante al inciso anterior.

c) Estimaci\u00f3n final.

La manera de integrar el resultado de las dos estimaciones anteriores es simplemente seleccionar la que genera la mayor precisi\u00f3n, esto es

la menor $\sum_i v(\hat{Y}_i)$ en donde la suma se realiza sobre la totalidad de candidatos a elegir.

Referencias

- Cochran, W. G. (1977) Sampling Techniques. 3ed. Wiley&Sons, Inc.

5.2.2.7. Dr. Gabriel Núñez Antonio

Con base en la distribución de las actividades del COTECORA, establecidas en el apartado 2.2, el Dr. Gabriel Núñez Antonio en colaboración con el Mat. Ernesto Silva Sartorius realizarán las estimaciones de las tendencias de la votación y el porcentaje de votación de la **Ciudad de México, Veracruz y Yucatán.**

Para tal efecto emplearán el procedimiento descrito en el apartado 5.1.2.3.2.

5.2.2.8. Dr. Luis Enrique Nieto Barajas

Con base en la distribución de las actividades del COTECORA, establecidas en el apartado 2.2, el Dr. Luis Enrique Nieto Barajas en colaboración con el Mtro. Carlos Samuel Pérez Pérez Martínez realizarán las estimaciones de las tendencias de la votación y el porcentaje de votación de los estados de **Morelos** y **Chiapas**.

Para tal efecto emplearán el procedimiento descrito en el apartado 5.1.2.2.1.

5.2.2.9. Dr. Carlos Erwin Rodríguez Hernández-Vela

Con base en la distribución de las actividades del COTECORA, establecidas en el apartado 2.2, el Dr. Carlos Erwin Rodríguez Hernández-Vela en colaboración con el Mtro. David Chaffrey Moreno Fernández realizarán las estimaciones de las tendencias de la votación y el porcentaje de votación de la **Ciudad de México, Veracruz y Yucatán**.

Para tal efecto emplearán el procedimiento descrito en el apartado 5.1.2.3.3.