

COMISIÓN ESPECIAL PARA LA ATENCIÓN DEL
PROGRAMA DE RESULTADOS ELECTORALES PRELIMINARES
Y CONTEOS RÁPIDOS

Toluca, México, 20 de abril de 2017.

Oficio No. IEEM/CE/PTP/059/2017.

Lic. Pedro Zamudio Godínez
Consejero Presidente del Instituto
Electoral del Estado de México

En atención a su atento oficio No. IEEM/PCG/PZG/910/17 en el que me envía copia simple de los oficios No. INE-JLE-MEX/VE/0808/2017 signado por el Mtro. Joaquín Rubio Sánchez, Vocal Ejecutivo de la Junta Local Ejecutiva del Instituto Nacional Electoral, y No. INE/UNIOCOM/1334/2017 signado por el Ing. Jorge Humberto Torres Antuñano, Coordinador General de la Unidad Técnica de Servicios de Informática, en los que solicitan información complementara relativa a la implementación del conteo rápido para la elección de Gobernador del Estado de México, remito a usted, para dar respuesta, el documento "Criterios Científicos, Protocolo para selección de la Muestra y Proceso Logístico Operativo para la realización del Conteo Rápido", elaborado conjuntamente por los integrantes del Comité Técnico Asesor de los Conteos Rápidos y el Dr. Juan José Rivaud, Jefe de la Unidad de informática y Estadística, para que sea el amable conducto de hacerlo llegar al Dr. Lorenzo Córdova Vianello, Consejero Presidente del Consejo General del Instituto Nacional Electoral.

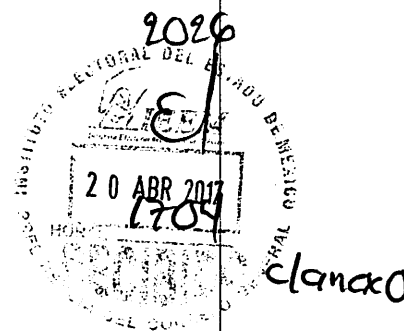
Sin otro particular, reciba un cordial saludo.

"Tú Haces la Mejor Elección"



Mtra. Palmira Tapia Palacios
Consejera Electoral y Presidenta de la Comisión
Especial para la Atención del PREP y Conteos rápidos

c.c.p. M. en A.P. Francisco Javier López Corral. Secretario Ejecutivo. Para su conocimiento
Dr. Juan José Rivaud Gallardo, Jefe de la unidad de Informática y Estadística. Para su conocimiento.





Comité Técnico Asesor para el Conteo Rápido de la Elección para Gobernador del Estado de México 2017

Criterios Científicos, Protocolo para
Selección de la Muestra y Proceso Logístico
Operativo para la realización del Conteo
Rápido

Abril 20, 2017

Contenido

I. Antecedentes.....	2
II. Criterios científicos.....	3
1.1 Objetivos.....	4
1.2 Esquema de muestreo	4
1.3 Estimación e integración de los intervalos de confianza	16
1.4 Jornada Electoral.....	18
1.5 Procedimientos de estimación.....	19
III. Selección y resguardo de la muestra	31
1.1 Consideraciones generales para la selección de la muestra.....	31
1.2 Protocolo de selección y resguardo de la muestra	31
IV. Proceso Logístico Operativo	33
1.1 Estructura general	33
1.2 Contingencias	36
1.3 Funciones del personal del INE involucrado en el ámbito distrital	36
1.4 Funciones del personal involucrado en el ámbito del IEEM.....	37
1.5 Procedimientos.....	38
1.6 Sistema Informático	41

I. Antecedentes

- 1.- El siete de septiembre de dos mil dieciséis, el Consejo General del Instituto Electoral del Estado de México, celebró sesión solemne por la que se dio inicio formalmente al Proceso Electoral Ordinario 2016-2017, para elegir al Gobernador Constitucional del Estado de México, para el periodo comprendido del 16 de septiembre de 2017 al 15 de septiembre de 2023.
- 2.- En sesión extraordinaria del siete de septiembre de dos mil dieciséis, el Consejo General del Instituto Nacional Electoral aprobó el Acuerdo INE/CG661/2016, denominado: "Acuerdo del Consejo General del Instituto Nacional Electoral por el que se aprueba el Reglamento de Elecciones del Instituto Nacional Electoral", mismo que fue publicado en el Diario Oficial de la Federación, el trece de septiembre del mismo año.
- 3.- El doce de septiembre de dos mil dieciséis, la H. "LIX" Legislatura Local, expidió el Decreto número 124, publicado en el Periódico Oficial del Gobierno del Estado Libre y Soberano de México, "Gaceta del Gobierno", en la misma fecha, en donde se convoca a los ciudadanos del Estado de México y a los partidos políticos con derecho a participar en la elección ordinaria para elegir al Gobernador Constitucional del Estado de México, para el periodo comprendido del 16 de septiembre de 2017 al 15 de septiembre de 2023.
- 4.- En sesión ordinaria del veintidós de septiembre del dos mil dieciséis, el Órgano Superior de Dirección de este Instituto, aprobó el Acuerdo IEEM/CG/83/2016, por el que creó la Comisión Especial para la atención del Programa de Resultados Electorales Preliminares y Conteos Rápidos; la cual quedó integrada de la siguiente forma: Presidenta, Consejera Electoral Mtra. Palmira Tapia Palacios; como integrantes, los Consejeros Electorales, Mtro. Miguel Ángel García Hernández y Mtro. Saúl Mandujano Rubio; como Secretario Técnico el Secretario Ejecutivo y con un representante de cada partido político con registro ante este Instituto.

En el Considerando XXVII, se estableció entre otros aspectos, como motivo de creación y dos de sus objetivos, los siguientes:

"Motivo de Creación

Contar con una Comisión que auxilie al Consejo General en la coordinación de las actividades relativas a la implementación y operación del Programa de Resultados Electorales Preliminares (PREP), así como para la realización de los conteos rápidos, en la elección de Titular del Poder Ejecutivo del Estado de México.

Objetivos

- *Proponer al Consejo General la designación de los órganos internos responsables de la coordinación de las actividades del Programa de Resultados Electorales Preliminares (PREP) y de los Conteos Rápidos.*
 - *...*
 - *Supervisar la correcta integración y funcionamiento del Comité Técnico Asesor del Programa de Resultados Electorales Preliminares (COTAPREP) y Conteos Rápidos (COTECORA).*
 - *..."*
- 5.- En sesión extraordinaria celebrada el ocho de diciembre de dos mil dieciséis, la Comisión Especial para la atención del Programa de Resultados Electorales Preliminares y Conteos Rápidos, aprobó el Acuerdo IEEM/CEPREP y CR/03/2016 por el que propone al Consejo General, la integración del Comité Técnico Asesor de los Conteos Rápidos (COTECORA); y
- 6.- El 14 de diciembre de 2016 fue aprobado por el Consejo General del IEEM el acuerdo IEEM/CG/115/2016 por el que se integra el Comité Técnico Asesor de los Conteos Rápidos para el Proceso Electoral 2016-2017.

En apego a las disposiciones normativas en la materia, el COTECORA desarrolló los criterios científicos, así como el protocolo para la selección de la muestra, además conoció el proceso logístico operativo para el acopio de los datos de las Actas de Escrutinio y Cómputo de Casilla, y los consideró pertinentes para los fines del Conteo Rápido del Estado de México 2017. Los Criterios Científicos, el Protocolo para Selección de la Muestra y el Proceso Logístico Operativo propuesto se describen en este documento.

II. Criterios científicos

Los criterios científicos del Conteo Rápido de la Elección del Estado de México son todos los procedimientos que, con base en la probabilidad y en la estadística, se usan para estimar el porcentaje de votos a favor de los candidatos para gobernador del estado.

1.1 Objetivos

El objetivo general de los conteos rápidos es crear certidumbre, confianza y transparencia en el proceso electoral, ofreciendo una estimación de las tendencias de la votación de alta calidad estadística en la noche del mismo día de la elección.

Los objetivos particulares del Conteo Rápido para la Elección de Gobernador en el Estado de México serán:

- Estimar el porcentaje de votos en favor de cada aspirante a gobernador por el Estado de México.
- Estimar el porcentaje de participación en la elección.

Todas las estimaciones del COTECORA estarán dadas en forma de intervalos de confianza del 95%.

1.2 Esquema de muestreo

Diseño de muestreo

Las estrategias de estimación que implementarán los miembros del COTECORA estarán basadas en la misma información muestral. El proceso de selección será un muestreo aleatorio estratificado, en donde al interior de cada estrato se seleccionarán casillas mediante un muestreo aleatorio simple sin reemplazo.

Para definir los estratos, se usaron los 45 Distritos Locales del Estado de México, (ver Tabla 1). La muestra total será de alrededor de 2,000 casillas, esperando recibir al menos 1,200 casillas para realizar la estimación final. Vale la pena resaltar que el tamaño de muestra definitivo se calculará con base en la lista final de casillas a instalarse el 4 de junio y considerando la tasa de no respuesta de cada distrito local observada en elecciones pasadas.

La distribución de casillas en cada estrato se realizará vía afijación proporcional. Suponiendo afijación proporcional y utilizando como referencia la distribución de casillas instaladas para la elección federal del 7 de junio de 2015, la muestra quedaría distribuida como se sugiere en la Tabla 1.

Tabla 1. Distribución preliminar de la muestra para el Conteo Rápido en el Estado de México 2017 (tomando como referencia la Elección Federal de Diputados 2015).

Distrito local	Casillas Instaladas (N_h)	Muestra (n_h)
01	351	39
02	302	33
03	357	39
04	383	42
05	373	41
06	321	35
07	404	46
08	359	40
09	515	57
10	434	48
11	374	41
12	371	41
13	442	49
14	438	48
15	382	42
16	390	43
17	360	40
18	498	55
19	370	41
20	395	44
21	381	42
22	382	42

Distrito local	Casillas Instaladas (N_h)	Muestra (n_h)	
23	403	44	
24	466	51	
25	522	58	
26	428	47	
27	393	43	
28	393	43	
29	461	51	
30	439	48	
31	376	42	
32	415	46	
33	446	49	
34	423	47	
35	384	42	
36	349	39	
37	374	41	
38	412	46	
39	415	46	
40	400	44	
41	505	56	
42	379	42	
43	387	43	
44	407	45	
45	356	39	
Total	18,115 (N)	2,000 (n_{Total})	

Consideraciones para definir el tamaño de muestra y diseño de muestreo

Ideas básicas del muestreo probabilístico para el Conteo Rápido

Como se describió anteriormente, el objetivo particular del conteo rápido es estimar

el porcentaje, p , de votos en favor de un candidato a gobernador¹. 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

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

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 a las exigencias de la Elección y a 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)

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

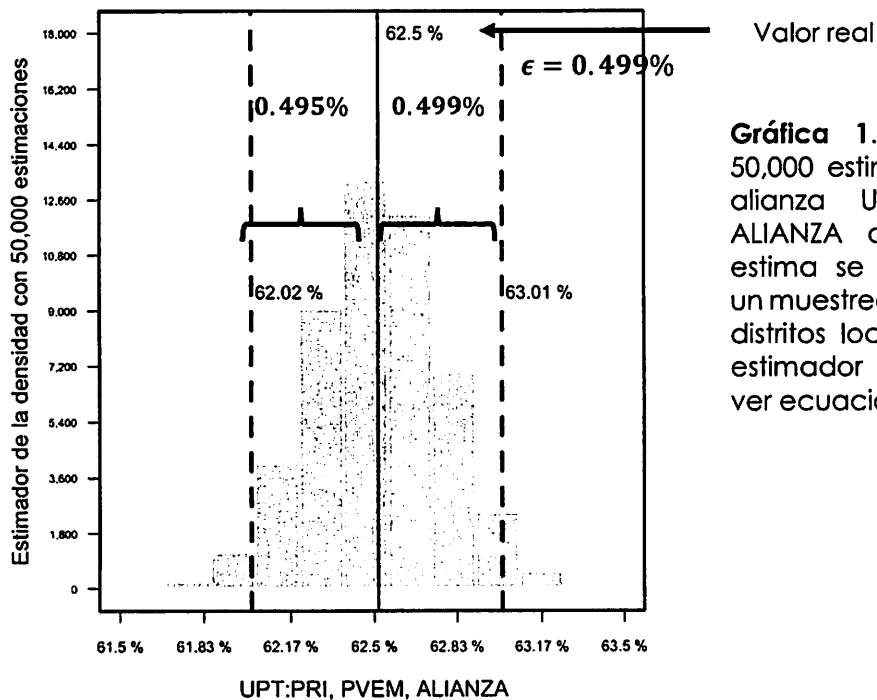
aproximadamente 95 cumplirán con que su distancia a p , es menor o igual a ϵ .

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 México. 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 México en 2011. Cada muestra fue de 1,000 casillas y se siguió un muestreo

¹ 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.

estratificado por distritos locales (45 estratos).



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

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

En donde 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 gobernador por el partido i en el estrato h .

Las estimaciones se realizaron 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, $p = 62.5\%$, 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 $\epsilon = 0.499\%$. 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, en este caso se tiene poco control de las casillas seleccionadas. Para explicar este tipo de estrategia, se recurre a una urna con N objetos numerados y el objetivo es seleccionar una muestra de n , con $n < N$. En cada selección, se revuelve la urna y se toma un objeto al azar. El objeto seleccionado se deja afuera de la urna para la siguiente selección (sin reemplazo), en la selección j -ésima quedan $N - (j - 1)$ objetos en la urna.
- El MASCR es muy similar al MASSR, pero después de cada selección, el objeto se regresa a la urna, de forma que en cada extracción se obtienen N objetos. En este caso es necesario registrar el número del objeto seleccionado.

- En el muestreo estratificado en cambio, se divide a la población en estratos/grupos² y en cada estrato se hace una selección de casillas.³ 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, tenemos 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 México 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.

Poblaciones de referencia para realizar las simulaciones Monte Carlo

En todos los ejercicios realizados se tomaron como referencia los cómputos distritales de tres votaciones:

1. Elección para Gobernador en el Estado de México del 3 de julio de 2011.
 - 17,489 casillas en total \Rightarrow 17,472 casillas para análisis (17 casillas).
2. Elección Presidencial, tomando en cuenta sólo los resultados en el Estado de México, del 1 de julio de 2012.
 - 17,350 casillas en total \Rightarrow 17,229 casillas para análisis (121 casillas).
3. Elección Federal de Diputados del 7 de junio de 2015. En este caso se asumió que los candidatos fueron los mismos en los 40 Distritos Federales y que hubo dos coaliciones: PRI-PVEM y PRD-PT.
 - 18,250 casillas en total \Rightarrow 18,115 casillas para análisis (135 casillas).

² los estratos son ajenos entre sí y su unión conforma la población completa

³ La selección en cada estrato usualmente se realiza mediante un MASS.

La diferencia entre las casillas totales y las casillas para análisis se debe a que no fue posible determinar el DISTRITO LOCAL de algunas casillas. Sin embargo, esto no modifica en manera alguna el análisis para determinar el tamaño de muestra.

Estratificaciones y tamaños de muestra

Las estratificaciones consideradas en los ejercicios de simulación, para cada elección, fueron seis:

1. Distritos Federales \Rightarrow 40 estratos.
2. Distritos Federales y al interior de cada distrito se dividió por sección urbana no urbana \Rightarrow 64 estratos.
3. Distritos Locales \Rightarrow 45 estratos.
4. Distritos Locales y al interior de cada distrito se dividió por sección urbana y no urbana \Rightarrow 75 estratos.
5. Municipios \Rightarrow 125 estratos.
6. Municipios y al interior de cada municipio se dividió por sección urbana y no urbana \Rightarrow 206 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 instaladas⁴ en el Estado de México. En palabras comunes, el tamaño de muestra total, se distribuirá proporcionalmente de acuerdo al 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. Experiencias anteriores indican que una precisión de

$$\epsilon = 0.5 \text{ (medio punto porcentual).}$$

⁴ Se toman las casillas para análisis descritas en la sección anterior.

es una buena alternativa y aunque exige tamaños de muestra grandes, no son prohibitivos y son manejables en campo.

En todas las simulaciones, se utilizó el estimador de razón dado por la expresión (3), los resultados obtenidos con cada una de las tres Elecciones, se muestran en las Tablas 2, 3 y 4 (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 2. Elección para Gobernador 2011: precisiones obtenidas usando el estimador (3), para distintos tamaños de muestra y estratificaciones.

TAMAÑO DE MUESTRA	DISTRITOS FEDERALES	DISTRITOS FEDERALES URBANO - NO URBANO	DISTRITOS LOCALES	DISTRITOS LOCALES URBANO - NO URBANO	MUNICIPIOS	MUNICIPIOS URBANO - NO URBANO
700	0.62	0.62	0.60	0.60	0.58	0.56
750	0.60	0.59	0.58	0.57	0.56	0.54
800	0.58	0.57	0.56	0.55	0.55	0.53
850	0.56	0.55	0.55	0.53	0.53	0.52
900	0.54	0.53	0.53	0.52	0.51	0.50
950	0.53	0.53	0.51	0.51	0.49	0.48
1,000	0.52	0.51	0.50	0.49	0.48	0.47
1,050	0.50	0.50	0.48	0.49	0.47	0.46
1,100	0.49	0.48	0.47	0.47	0.46	0.45
1,150	0.48	0.48	0.46	0.46	0.45	0.44
1,200	0.47	0.47	0.45	0.45	0.44	0.43
1,250	0.46	0.45	0.44	0.43	0.43	0.42
1,300	0.45	0.44	0.43	0.43	0.42	0.41
1,350	0.44	0.44	0.42	0.42	0.42	0.41
1,400	0.44	0.43	0.42	0.41	0.40	0.39

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

TAMAÑO DE MUESTRA	DISTRITOS FEDERALES	DISTRITOS FEDERALES URBANO - NO URBANO	DISTRITOS LOCALES	DISTRITOS LOCALES URBANO - NO URBANO	MUNICIPIOS	MUNICIPIOS URBANO - NO URBANO
700	0.55	0.54	0.52	0.52	0.51	ALGUNOS
750	0.53	0.52	0.50	0.50	0.49	ESTRATOS CON

TAMAÑO DE MUESTRA	DISTRITOS FEDERALES	DISTRITOS FEDERALES URBANO - NO URBANO	DISTRITOS LOCALES	DISTRITOS LOCALES URBANO - NO URBANO	MUNICIPIOS	MUNICIPIOS URBANO - NO URBANO
800	0.51	0.50	0.49	0.48	0.48	1 ÚNICA CASILLA INSTALADA
850	0.49	0.48	0.48	0.46	0.46	
900	0.48	0.47	0.46	0.45	0.45	
950	0.46	0.46	0.44	0.44	0.43	
1,000	0.45	0.45	0.44	0.43	0.42	
1,050	0.44	0.43	0.42	0.42	0.42	
1,100	0.43	0.42	0.41	0.41	0.40	
1,150	0.42	0.41	0.40	0.40	0.40	
1,200	0.41	0.40	0.39	0.39	0.39	
1,250	0.40	0.39	0.38	0.38	0.38	
1,300	0.39	0.39	0.38	0.37	0.37	
1,350	0.38	0.38	0.37	0.36	0.36	
1,400	0.37	0.37	0.36	0.36	0.35	

Tabla 4. Elección Federal de Diputados 2015: precisiones obtenidas usando el estimador (3), para distintos tamaños de muestra y estratificaciones.

TAMAÑO DE MUESTRA	DISTRITOS FEDERALES	DISTRITOS FEDERALES URBANO - NO URBANO	DISTRITOS LOCALES	DISTRITOS LOCALES URBANO - NO URBANO	MUNICIPIOS	MUNICIPIOS URBANO - NO URBANO
700	0.68	0.67	0.66	0.66	0.59	0.56
750	0.67	0.65	0.64	0.63	0.57	0.55
800	0.64	0.62	0.61	0.61	0.55	0.53
850	0.62	0.61	0.59	0.59	0.52	0.51
900	0.60	0.59	0.58	0.57	0.51	0.50
950	0.58	0.58	0.56	0.56	0.50	0.48
1,000	0.57	0.55	0.55	0.54	0.49	0.48
1,050	0.55	0.54	0.54	0.52	0.48	0.47
1,100	0.53	0.53	0.52	0.51	0.47	0.45
1,150	0.52	0.52	0.51	0.50	0.45	0.44
1,200	0.51	0.51	0.50	0.49	0.44	0.44
1,250	0.51	0.49	0.48	0.47	0.43	0.42
1,300	0.50	0.48	0.48	0.47	0.43	0.42
1,350	0.48	0.48	0.46	0.46	0.42	0.40
1,400	0.47	0.47	0.46	0.45	0.41	0.40

Estos resultados muestran que, de las tres elecciones, la que requiere el menor tamaño de muestra para alcanzar una mejor precisión es la Presidencial, le sigue la Elección de Gobernador y la que exige un mayor tamaño de muestra es la Elección de Diputados. Esto es debido a que la varianza de la votación es mayor en la Elección de Diputados, en nuestro análisis supusimos que hubo 8 candidatos (con 2 coaliciones) en los 40 Distritos Federales que componen el Estado de México, sin embargo, en la realidad en cada distrito hubo candidatos distintos. Esto introduce una gran variabilidad en la votación para todo el estado. Entonces, el tamaño de muestra que se derive del análisis de la votación para Diputados necesariamente será suficiente para las otras votaciones.

En la Tabla 3, la mejor estratificación es la que contempla a los municipios, ya que el tamaño de muestra requerido para alcanzar un error de $\epsilon = 0.5$ es de $n = 950$ casillas. Sin embargo, existe un problema práctico con esta estratificación. El detalle se resume en la Tabla 5.

Tabla 5. Elección Federal de Diputados 2015: Estadísticas básicas para cada estratificación y las casillas instaladas en cada estrato.

Variable de interés	DISTRITOS FEDERALES	DISTRITOS FEDERALES URBANO - NO URBANO	DISTRITOS LOCALES	DISTRITOS LOCALES URBANO - NO URBANO	MUNICIPIOS	MUNICIPIOS URBANO - NO URBANO
Número de estratos	40	64	45	75	125	194
Promedio de casillas instaladas	453	283	403	242	145	93
Máximo casillas instaladas	678	606	522	522	1,966	1,966
Mínimo casillas instaladas	350	3	302	2	6	2

El mínimo de casillas instaladas en uno de los municipios es 6, de hecho, en 23 municipios se instalaron menos de 20 casillas. Esto representa un gran riesgo para el conteo rápido, de elegirse los municipios como estratos, de cada municipio se extraería una muestra de casillas y con la información que se recupere de esta muestra, se realiza una estimación. Finalmente, se conjuntan las estimaciones de todos los estratos para obtener la estimación global, ver ecuación (3).

Con estratos con tan pocas casillas instaladas, aunque se extraiga una muestra grande (proporcionalmente), la probabilidad de que no se recupere ninguna casilla en estos estratos es muy alta. Si no se tiene información en algún estrato, será imposible realizar la estimación. Lo mismo ocurre si se utiliza alguna estratificación que contemple la parte urbana y no urbana, el número de casillas instaladas en algunos estratos es muy pequeño. 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 302, el máximo es de 522 y en promedio se instalan 403 casillas por Distrito Local. Además, también se alcanza una buena precisión con un tamaño de muestra aceptable.

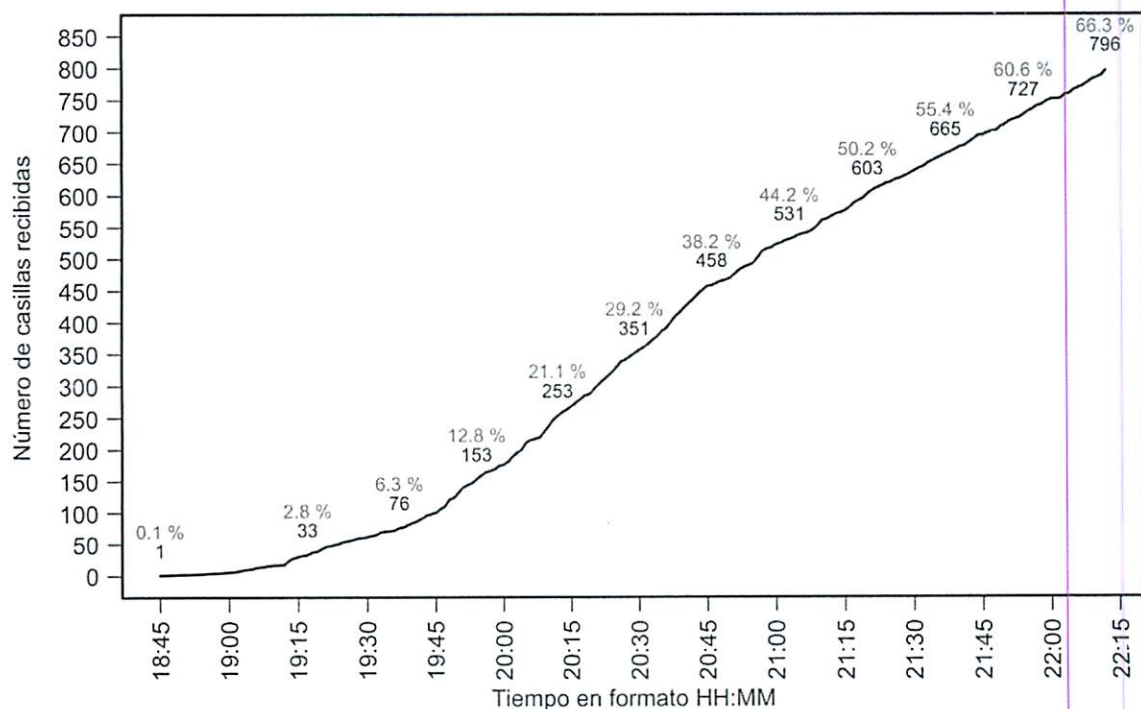
Tomando los Distritos Locales como estratos, se obtendría una precisión de $\epsilon = 0.5$ con $n = 1,200$ casillas, además, utilizando la afijación proporcional el promedio de casillas en muestra por estrato sería de 27, con un máximo de 35 y mínimo de 20.

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 México fue de 1,200 casillas (30 casillas por cada Distrito Federal), sin embargo, a las 10:00 pm sólo se había recibido el 60 % y a las 10:15 pm, cuando se hizo la última estimación, se tenía aproximadamente el 66 % de la muestra total. El número de casillas recibidas, así como la hora de arribo se muestra en la Gráfica 2.

Gráfica 2. Conteo Rápido 2015: Serie de tiempo para el número de casillas recibidas del Estado de México.



Si el mismo comportamiento se mantiene para el Conteo Rápido en la Elección de Gobernador en 2017, para realizar la estimación final a las 10 pm se necesitaría un tamaño de muestra total, considerando la no respuesta, de

$$n_{Total} = \frac{1,200}{0.6} = 2,000 \text{ casillas.}$$

Con este tamaño de muestra total, se esperaría contar con un tamaño de muestra efectivo de $n = 1,200$ casillas alrededor de las 10 pm del 4 de junio.

Consideraciones finales

Por último, es importante tomar en cuenta los siguientes puntos

1. 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.
2. 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.
3. Para la elección en el Estado de México, se espera contar con 3,352 capacitadores asistentes electorales (CAES) que recabarán la información de las casillas seleccionadas en muestra y procederán a su transmisión. Es conveniente apuntar que 2,000 casillas como tamaño de muestra se encuentra dentro de la capacidad operativa de campo.

1.3 Estimación e integración de los intervalos de confianza

Para dar mayor certeza a la estimación, los integrantes del COTECORA utilizarán tres enfoques estadísticos distintos para estimar los resultados de la elección:

- Método Bayesiano, a cargo del Dr. Arturo Erdely Ruiz.
- Método Clásico vía las fórmulas para estimadores de razón, derivadas mediante las técnicas de muestreo probabilístico, a cargo de la Dra. Karla Beatriz Valenzuela Ocaña y el Dr. Alberto Alonso y Coria.
- Método Clásico vía métodos de remuestreo, a cargo del Dr. Carlos E. Rodríguez.

Dado que se aplicarán diversos métodos de estimación, no se espera que los límites inferior y superior de cada intervalo sean estrictamente idénticos, toda vez que los métodos que se utilizarán introducen elementos aleatorios. Sin embargo, el objetivo

de los procedimientos de estimación será el mismo, las estimaciones se consolidarán para presentar resultados únicos en forma de intervalos de confianza.

Las estrategias de estimación que implementarán los miembros del COTECORA estarán basadas en la misma muestra. Durante toda la jornada, el COTECORA monitoreará que las tendencias de votación indicadas por los intervalos de confianza estimados vayan en el mismo sentido.

Para integrar las estimaciones en una sola, se aplicará el siguiente procedimiento: Sea $I_k = [a_k, b_k]$ una estimación por intervalo de un porcentaje de votación, donde el subíndice $k = 1, \dots, r$ representa cada una de las r estimaciones que realizarán los miembros del COTECORA. El intervalo combinado $I = [a, b]$ que reportará el COTECORA se obtendrá mediante las fórmulas:

$$a = \frac{\text{promedio}\{a_1, \dots, a_r\} + \text{mediana}\{a_1, \dots, a_r\}}{2} \quad \text{y} \quad b = \frac{\text{promedio}\{b_1, \dots, b_r\} + \text{mediana}\{b_1, \dots, b_r\}}{2}$$

Utilizar solo promedios tiene la ventaja de incorporar directamente información de todos los intervalos, pero la desventaja de que un solo valor atípico podría tener demasiada influencia. Utilizar solo medianas tiene la ventaja de ser inmune ante un valor atípico, pero la desventaja de que varios intervalos podrían quedar totalmente fuera de consideración. Es por lo anterior que se escoge un término medio entre ambas posibilidades.

1.4 Jornada Electoral

El Sistema de Captura del Conteo Rápido (SICCORA) generará cortes de información cada 10 minutos. Una vez que se tenga información para cada uno de los estratos, es decir al menos 2 casillas por Distrito Local, se realizarán estimaciones con cada una de los tres métodos propuestos para monitorear que éstas vayan en el mismo sentido.

Considerando tanto el avance en la cobertura de la muestra recibida por estrato, como el seguimiento de las estimaciones, el COTECORA determinará el momento oportuno en el que esté en posibilidades de pronunciar sus estimaciones finales. En este momento se generará un informe con formato previamente definido, mismo que se imprimirá y será firmado por cada uno de los miembros del COTECORA para hacerlo llegar al Presidente del Consejo General del IEEM.

El día de la jornada electoral, el COTECORA rendirá un informe de avance de la integración de la muestra al Consejo General del IEEM cada hora a partir de las 21:00 horas y hasta la entrega de los resultados finales.

1.5 Procedimientos de estimación

Propuesta del Dr. Arturo Erdely Ruiz

Estimación bayesiana por intervalos de porcentajes de votación

Dr. Arturo Erdely Ruiz

Abril de 2017

Resumen

Se propone la utilización de la distribución de probabilidad multinomial, y la correspondiente inferencia bayesiana de sus parámetros, para la estimación por intervalo de los porcentajes del voto emitido que obtendrán los candidatos en la elección para gobernador del Estado de México del año 2017, a partir de un conteo rápido de actas electorales por casilla.

1. Introducción

Si una variable o vector aleatorio X con una función de densidad o masa de probabilidades $f(x|\theta)$ cuantifica adecuadamente la incertidumbre sobre un fenómeno aleatorio de interés, pero su parámetro (o vector de parámetros) $\theta \in \mathfrak{S}$ es desconocido, donde \mathfrak{S} es el conjunto de valores admisibles para θ (espacio paramétrico), el enfoque bayesiano de la estadística propone cuantificar la incertidumbre sobre θ mediante una variable aleatoria Θ con una distribución de probabilidad sobre \mathfrak{S} , misma que puede basarse en información inicial con que se cuente en un momento dado y que se denomina *distribución a priori*, y que al combinarse con información muestral se convierte en una *distribución a posteriori*.

Sea (X_1, \dots, X_m) una muestra aleatoria con función de densidad o masa de probabilidades conjunta $f(y_1, \dots, y_m | \theta)$. Si (x_1, \dots, x_m) son los valores observados de dicha muestra, y se denota mediante $\pi(\theta)$ a la función de densidad de probabilidades a priori, y mediante $p(\theta | x_1, \dots, x_m)$ a la función de densidad de probabilidades a posteriori, entonces mediante la *Regla de Bayes* se obtiene la siguiente fórmula:

$$p(\theta | x_1, \dots, x_m) = \frac{1}{B} \pi(\theta) f(x_1, \dots, x_m | \theta), \quad \theta \in \mathfrak{S}, \quad (1)$$

donde la constante $B = \int_{\mathfrak{S}} \pi(\theta) f(x_1, \dots, x_m | \theta) d\theta$. Mediante (1) se realiza cualquier inferencia estadística de interés sobre el parámetro (o vector de parámetros) θ . Para el caso de *estimación puntual*, misma que consiste en escoger un valor $\hat{\theta} \in \mathfrak{S}$ con base en toda la información disponible, puede calcularse alguna medida de tendencia central (la media o la mediana, por ejemplo) por medio de $p(\theta | x_1, \dots, x_m)$.

En el caso de *estimación por intervalo* de probabilidad $0 < \alpha < 1$, si θ es de dimensión 1 entonces se busca un intervalo $I(\alpha) := [a, b] \subset \mathfrak{S}$ tal que:

$$\mathbb{P}[\theta \in I(\alpha)] = \int_a^b p(\theta | x_1, \dots, x_m) d\theta = \alpha. \quad (2)$$

Usualmente la solución para $I(\alpha)$ en (2) no es única, por lo que de entre el conjunto de soluciones posibles $\{[a, b] : \int_a^b p(\theta | x_1, \dots, x_m) d\theta = \alpha\}$ se escoge aquel intervalo cuya longitud $b - a$ sea mínima. En caso de que la dimensión de \mathbb{S} sea $J > 1$ entonces la densidad de probabilidades a posteriori (1) de $\theta = (\theta_1, \dots, \theta_J)$ es multivariada y por tanto deberá marginalizarse para cada componente θ_j ; y mediante $p(\theta_j | x_1, \dots, x_m)$ hacer estimación puntual y por intervalo para cada θ_j con $j \in \{1, \dots, J\}$. Si $p(\theta_j | x_1, \dots, x_m)$ no es simétrica respecto a algún valor $\theta_j = \theta_0$, obtener la estimación por intervalo $I_j(\alpha)$ que sea de longitud mínima pudiera resultar un tanto laborioso o computacionalmente intensivo, y por tanto es común ocupar la solución pseudo-óptima que consiste en $I_j(\alpha) = [a_j, b_j]$ tal que:

$$\int_{-\infty}^{a_j} p(\theta_j | x_1, \dots, x_m) d\theta_j = \frac{1 - \alpha}{2} = \int_{b_j}^{+\infty} p(\theta_j | x_1, \dots, x_m) d\theta_j. \quad (3)$$

2. Notación

Se adopta prácticamente tal cual la notación utilizada en Mendoza y Nieto-Barajas (2016). El *marco muestral* disponible consiste en un total de K casillas electorales que se instalarán el día de la elección para que un total de n posibles votantes (tamaño de la *lista nominal*) estén en condiciones de ejercer su derecho al voto, mismos que pueden decidir sufragar a favor de algún candidato registrado, o a favor de alguno no registrado (NOREG), o anular su voto (NULOS) o simplemente abstenerse de acudir a votar (ABST). Al conjunto de posibilidades anteriores les asignaremos las categorías $1, 2, \dots, J$ reservando siempre $J - 2$ para NOREG, $J - 1$ para NULOS y J para ABST, esto es, los candidatos registrados serán $1, \dots, J - 3$, con $J \geq 4$. En lo sucesivo cuando se haga referencia a un *candidato* $j \in \{1, \dots, J\}$ deberá entenderse en un sentido amplio, es decir considerando como “candidatos” incluso a las categorías NOREG, NULOS y ABST.

El total de K casillas electorales se reparten en N *estratos* o subconjuntos (distritos locales en este caso), y se denotará K_i al número de casillas en el distrito $i \in \{1, \dots, N\}$ de modo que $\sum_{i=1}^N K_i = K$, y mediante n_i al tamaño de la lista nominal en el distrito i de modo que $\sum_{i=1}^N n_i = n$. Al momento del cierre de las casillas se procede al conteo de votos emitidos en cada una y su clasificación en las categorías 1 a la $J - 1$. Se representará mediante las variables aleatorias $X_{i,j}^{(k)}$ el número de votos que obtendrá el candidato $j \in \{1, \dots, J\}$ en el distrito $i \in \{1, \dots, N\}$ y casilla $k \in \{1, \dots, K_i\}$. A su vez, las variables aleatorias $X_{i,j} = \sum_{k=1}^{K_i} X_{i,j}^{(k)}$ representan el número total de votos que obtendrá el candidato j en el distrito i , y las variables aleatorias $\Theta_{i,j} = X_{i,j}/n_i$ representan la proporción de votos a favor del candidato j en el distrito i . De lo anterior es inmediato que $\sum_{j=1}^J X_{i,j} = n_i$ y $\sum_{j=1}^J \Theta_{i,j} = 1$ para todo distrito i .

Con lo anterior es posible definir variables aleatorias X_j como el número total de votos que obtiene el candidato j y en consecuencia $X_j = \sum_{i=1}^N X_{i,j} = \sum_{i=1}^N n_i \Theta_{i,j}$. La proporción total de votos que obtendrá el candidato j se obtiene mediante las variables aleatorias:

$$\Theta_j = \frac{X_j}{n} = \sum_{i=1}^N \frac{n_i}{n} \Theta_{i,j}, \quad j \in \{1, \dots, J\} \quad (4)$$

de donde es inmediato que $\sum_{j=1}^J X_j = n$ y en consecuencia $\sum_{j=1}^J \Theta_j = 1$. La proporción de votos Θ_j es respecto a la lista nominal, pero también es de interés la proporción de votos respecto al total de

votos emitidos en las casillas (esto es, sin considerar la J -ésima categoría ABST), misma que puede representarse mediante las variables aleatorias:

$$\Lambda_j = \frac{X_j}{\sum_{t=1}^{J-1} X_t} = \frac{\Theta_j}{\sum_{t=1}^{J-1} \Theta_t}, \quad j \in \{1, \dots, J-1\}, \quad (5)$$

y la proporción de *participación* en la elección se calcula como complemento del abstencionismo (ABST), esto es como $1 - \Theta_J$.

Se realizará un *muestreo aleatorio estratificado* consistente en muestreo aleatorio simple de casillas por distrito (estrato) proporcional a la lista nominal de cada distrito. Esto es, si en el distrito i hay K_i casillas, de entre éstas se seleccionará al azar un subconjunto de c_i casillas donde $c_i \ll K_i$ y de forma que si $c = \sum_{i=1}^N c_i$ representa el número total de casillas en la muestra estratificada entonces $c_i \approx \frac{n_i}{n} c$.

3. Distribución de probabilidad multinomial

En virtud de que la *unidad de muestreo* para el conteo rápido es el resultado de la votación en cada casilla electoral, se propone representar los resultados de las distintas casillas mediante un vector aleatorio $(X_{i,1}^{(k)}, \dots, X_{i,J}^{(k)})$ donde $i \in \{1, \dots, N\}$ y $k \in \{1, \dots, K_i\}$. Los resultados que se asientan en el acta de cada casilla corresponden a votos efectivamente emitidos, esto es a los valores que se observarán para las variables aleatorias $X_{i,1}^{(k)}, \dots, X_{i,J-1}^{(k)}$, excluyendo la categoría J (ABSTenciones) ya que esta última se deduce restando al tamaño de la lista nominal de la casilla ($n_{i,k}$) el total de la votación efectivamente emitida $\sum_{j=1}^{J-1} X_{i,j}^{(k)}$ de modo que $\sum_{j=1}^J X_{i,j}^{(k)} = n_{i,k}$. Esto último implica una dependencia entre las variables aleatorias $X_{i,j}^{(k)}$ ya que sin importar el resultado de cada casilla su suma siempre debe ser igual (teóricamente) al tamaño de la lista nominal de dicha casilla; así, por ejemplo, un valor muy elevado de alguna de ellas implica necesariamente valores modestos para las demás. Esto es, los valores de unas influyen en las demás porque el máximo número de votos obtenible por casilla ($n_{i,k}$) ya está predefinido (salvo en unas pocas casillas denominadas *especiales*).

La idea de aplicar un muestreo aleatorio estratificado radica en el beneficio inferencial que se obtiene por establecer o escoger una estratificación tal que en cada estrato sea razonable suponer cierta homogeneidad en estas subpoblaciones, en contraste con la heterogeneidad que se tendría al obtener muestras directamente de la población total sin estratificar, en deterioro de la precisión de las inferencias que se desea realizar, véase por ejemplo Särndal *et al.* (1992). Con lo anterior, el vector aleatorio $(X_{i,1}^{(k)}, \dots, X_{i,J}^{(k)})$ puede asociarse a un *experimento aleatorio multinomial* consistente en $n_{i,k}$ repeticiones independientes (votantes en el distrito i y casilla k) de un procedimiento que en cada repetición puede arrojar uno de entre J resultados posibles con una probabilidad $0 < \theta_{i,j} < 1$, $j \in \{1, \dots, J\}$, con $\sum_{j=1}^J \theta_{i,j} = 1$ (nótese que aquí se está aplicando el supuesto de homogeneidad por estrato, ya que las probabilidades teóricas $\theta_{i,j}$ se suponen razonablemente homogéneas para toda casilla k dentro de un mismo distrito i), y por tanto las variables aleatorias $X_{i,j}^{(k)}$ representan el número de veces que se obtuvo el resultado j en las $n_{i,k}$ repeticiones (número de votos a favor del candidato j). Bajo los supuestos anteriores, la función de masa de probabilidades conjunta del vector aleatorio $(X_{i,1}^{(k)}, \dots, X_{i,J}^{(k)})$ para cada distrito i y casilla k , véase por ejemplo Johnson *et al.* (1997),

sería **Multinomial** $(\theta_{i,1}, \dots, \theta_{i,J}, n_{i,k})$:

$$\begin{aligned} f_i^{(k)}(z_1, \dots, z_J | \theta_{i,1}, \dots, \theta_{i,J}, n_{i,k}) &= \mathbb{P}(X_{i,1}^{(k)} = z_1, \dots, X_{i,J}^{(k)} = z_J) \\ &= \binom{n_{i,k}}{z_1, \dots, z_J} \prod_{j=1}^J \theta_{i,j}^{z_j} \mathbf{1}_{\{\sum_{j=1}^J z_j = n_{i,k}\}} \end{aligned} \quad (6)$$

Con lo anterior, el vector aleatorio de votos por candidato en cada distrito se obtiene sumando los votos obtenidos por cada candidato en cada casilla del distrito:

$$(X_{i,1}, \dots, X_{i,J}) = \left(\sum_{k=1}^{K_i} X_{i,1}^{(k)}, \dots, \sum_{k=1}^{K_i} X_{i,J}^{(k)} \right), \quad i \in \{1, \dots, N\} \quad (7)$$

donde de acuerdo a Johnson *et al.*(1997) el vector aleatorio (7) tiene distribución de probabilidad **Multinomial** $(\theta_{i,1}, \dots, \theta_{i,J}, n_i)$ con $n_i = \sum_{k=1}^{K_i} n_{i,k}$.

4. Inferencia bayesiana

En cada distrito (estrato) i con K_i casillas se obtendrá una muestra de $c_i \ll K_i$ casillas por muestreo aleatorio simple, información muestral que puede denotarse:

$$\mathcal{M}_i = \{(x_{i,1}^{(k)}, \dots, x_{i,J}^{(k)}) : k = 1, \dots, c_i\}, \quad i \in \{1, \dots, N\}, \quad (8)$$

a partir de lo cual puede obtenerse la votación total por candidato de cada distrito i , esto es $(x_{i,1}, \dots, x_{i,J})$ donde $x_{i,j} = \sum_{k=1}^{K_i} x_{i,j}^{(k)}$. Con lo anterior se desea realizar inferencias respecto a los vectores de parámetros $(\theta_{i,1}, \dots, \theta_{i,J})$, que desde un enfoque bayesiano requiere una densidad de probabilidades a posteriori como en (1):

$$p(\theta_{i,1}, \dots, \theta_{i,J} | x_{i,1}, \dots, x_{i,J}) = \frac{1}{B} \pi(\theta_{i,1}, \dots, \theta_{i,J}) f_i(x_{i,1}, \dots, x_{i,J} | \theta_{i,1}, \dots, \theta_{i,J}). \quad (9)$$

Para una distribución de probabilidad Multinomial normalmente se utiliza una distribución a priori **Dirichlet**, véase por ejemplo Bernardo y Smith (1994), por dos razones esencialmente: primero, representa un vector de variables aleatorias cada una con soporte en el intervalo abierto $]0, 1[$ y tales que su suma es siempre igual a 1, tal cual se requiere para los parámetros de la distribución Multinomial; segundo, es *familia conjugada* para la distribución de probabilidad Multinomial, esto es que la distribución a posteriori resulta ser también Dirichlet (aunque con parámetros actualizados de acuerdo a la información muestral):

$$\pi(\theta_{i,1}, \dots, \theta_{i,J}) = \frac{\Gamma(\sum_{j=1}^J \alpha_{i,j})}{\prod_{j=1}^J \Gamma(\alpha_{i,j})} \prod_{j=1}^J \theta_{i,j}^{\alpha_{i,j}-1} \mathbf{1}_{\{\sum_{j=1}^J \theta_{i,j} = 1\}} \quad (10)$$

donde los *hiperparámetros* $\alpha_{i,j} > 0$ deben especificarse de acuerdo a *información a priori*. Como lo que se busca es una inferencia basada más en la información muestral que en información a priori puede optarse por utilizar $\alpha_{i,j} = 1$ ya que en tal caso se obtiene una distribución de probabilidad *uniforme* sobre el conjunto $\{(\theta_{i,1}, \dots, \theta_{i,J}) \in]0, 1[^2 : \sum_{j=1}^J \theta_{i,j} = 1\}$:

$$\pi(\theta_{i,1}, \dots, \theta_{i,J}) = (J-1)! \mathbf{1}_{\{\sum_{j=1}^J \theta_{i,j} = 1\}}, \quad (11)$$

y por lo tanto:

$$\pi(\theta_{i,1}, \dots, \theta_{i,J}) f_i(x_{i,1}, \dots, x_{i,J} | \theta_{i,1}, \dots, \theta_{i,J}) \propto \prod_{j=1}^J \theta_{i,j}^{x_{i,j}+1-1} \mathbf{1}_{\{\sum_{j=1}^J \theta_{i,j}=1\}}, \quad (12)$$

de lo cual se deduce que la distribución de probabilidad a posteriori en cada distrito $i \in \{1, \dots, N\}$ para $(\theta_{i,1}, \dots, \theta_{i,J})$ es **Dirichlet** $(1 + x_{i,1}, \dots, 1 + x_{i,J})$, donde la distribución marginal a posteriori para un candidato j^* del distrito i resulta ser **Beta** $(1 + x_{i,j^*}, \sum_{j=1}^J x_{i,j} - x_{i,j^*} + J - 1)$.

5. Estimación por intervalo

Con lo obtenido en la sección anterior se obtiene la distribución de probabilidad a posteriori para cada variable aleatoria $\Theta_{i,j}$, la proporción de votos a favor del candidato j en el distrito i , y por tanto la proporción total de votos a favor del candidato j en (4) es una combinación lineal convexa de la forma $\Theta_j = \sum_{i=1}^N \beta_i \Theta_{i,j}$ donde $\beta_i = n_i/n > 0$ y $\sum_{i=1}^N \beta_i = 1$. Definiendo $\mu_{i,j} := \mathbb{E}(\Theta_{i,j})$ y $\sigma_{i,j}^2 := \mathbb{V}(\Theta_{i,j})$, por la propiedad de linealidad de la esperanza es inmediato obtener que:

$$\mu_j := \mathbb{E}(\Theta_j) = \sum_{i=1}^N \beta_i \mu_{i,j}, \quad j \in \{1, \dots, J\}, \quad (13)$$

lo cual deja claro que la esperanza de Θ_j solo depende de las esperanzas marginales de las variables aleatorias $\{\Theta_{i,j} : i = 1, \dots, N\}$ sin importar la posible dependencia que pudiera existir entre ellas. Sin embargo, en el cálculo de la varianza de Θ_j sí se encuentra implicada dicha dependencia ya que aparecen covarianzas:

$$\sigma_j^2 := \mathbb{V}(\Theta_j) = \sum_{i=1}^N \beta_i^2 \sigma_{i,j}^2 + 2 \sum_{1 \leq i < r \leq N} \beta_i \beta_r \text{Cov}(\Theta_{i,j}, \Theta_{r,j}) \quad (14)$$

$$\begin{aligned} &= \sum_{i=1}^N \sum_{r=1}^N \beta_i \beta_r \text{Cov}(\Theta_{i,j}, \Theta_{r,j}) \\ &\leq \sum_{i=1}^N \sum_{r=1}^N \beta_i \beta_r \sigma_{i,j} \sigma_{r,j} = \left(\sum_{i=1}^N \beta_i \sigma_{i,j} \right)^2 \end{aligned} \quad (15)$$

en donde la desigualdad (15) es consecuencia inmediata de la *Desigualdad de Cauchy-Schwarz*, ver por ejemplo Casella y Berger (2002), y proporciona el valor máximo posible de la varianza σ_j^2 . Sea $I_j(\alpha)$ un intervalo de probabilidad $0 < \alpha < 1$ para Θ_j , esto es tal que $\mathbb{P}[\Theta_j \in I_j(\alpha)] = \alpha$. Supongamos que para algún valor z_α dicho intervalo puede expresarse en la forma:

$$I_j(\alpha) = [\mu_j - z_\alpha \sigma_j, \mu_j + z_\alpha \sigma_j] \quad (16)$$

y lo análogo para $\Theta_{i,j}$ mediante $I_{i,j}(\alpha) = [\mu_{i,j} - z_\alpha \sigma_{i,j}, \mu_{i,j} + z_\alpha \sigma_{i,j}]$. Si se construye un intervalo $I_j^*(\alpha)$ como combinación lineal convexa de los intervalos $I_{i,j}(\alpha)$ mediante las ponderaciones $\beta_i = n_i/n$ se

obtiene lo siguiente:

$$\begin{aligned}
I_j^*(\alpha) &= \sum_{i=1}^N \beta_i I_{i,j}(\alpha) \\
&= \left[\sum_{i=1}^N \beta_i \mu_{i,j} - z_\alpha \sum_{i=1}^N \beta_i \sigma_{i,j}, \sum_{i=1}^N \beta_i \mu_{i,j} + z_\alpha \sum_{i=1}^N \beta_i \sigma_{i,j} \right] \\
&= [\mu_j - z_\alpha \sigma_j^*, \mu_j + z_\alpha \sigma_j^*]
\end{aligned} \tag{17}$$

donde la desviación estándar $\sigma_j^* = \sum_{i=1}^N \beta_i \sigma_{i,j}$ es el valor máximo de σ_j de acuerdo a (15). En caso de que fuese válido un supuesto de independencia entre distritos entonces $\text{Cov}(\Theta_{i,j}, \Theta_{r,j}) = 0$ y de acuerdo a (14) la desviación estándar resultante sería $\sigma_j^\perp = \sqrt{\sum_{i=1}^N \beta_i^2 \sigma_{i,j}^2}$, cantidad que al ser menor que σ_j^* da origen a un intervalo de menor longitud $I_j^\perp(\alpha) = [\mu_j - z_\alpha \sigma_j^\perp, \mu_j + z_\alpha \sigma_j^\perp] \subset I_j^*(\alpha)$. Lo anterior deja en claro que si el supuesto de independencia entre distritos no es correcto entonces el intervalo resultante I_j^\perp podría ser de menor longitud de la que debiera y, por lo tanto, de menor probabilidad al nivel α deseado. Pero utilizar el intervalo I_j^* resultante de suponer covarianzas máximas entre distritos podría ser el otro extremo con una longitud mayor a la que debiera y por tanto con un nivel de probabilidad mucho mayor al deseado. Lo ideal sería obtener $I_j(\alpha)$ como en (16) con el valor preciso de σ_j mediante (14) pero esto requería estimar $N(N-1)/2$ covarianzas distintas con base en una cantidad de información con la que no se cuenta. Suponiendo que el valor buscado σ_j pertenece al intervalo $[\sigma_j^\perp, \sigma_j^*]$ entonces debe existir un valor $0 \leq \delta_j \leq 1$ tal que

$$\sigma_j = (1 - \delta_j) \sigma_j^\perp + \delta_j \sigma_j^* \tag{18}$$

y a partir de lo anterior el intervalo de probabilidad α para Θ_j sería:

$$\begin{aligned}
I_j(\alpha) &= \mu_j \pm z_\alpha [(1 - \delta_j) \sigma_j^\perp + \delta_j \sigma_j^*] \\
&= (1 - \delta_j) \mu_j + \delta_j \mu_j \pm z_\alpha (1 - \delta_j) \sigma_j^\perp \pm z_\alpha \delta_j \sigma_j^* \\
&= (1 - \delta_j) I_j^\perp(\alpha) + \delta_j I_j^*(\alpha).
\end{aligned} \tag{19}$$

En (19) persiste el problema de estimar δ_j pero ya solo serían J valores (uno por candidato) además de que el valor $0 \leq \delta_j \leq 1$ nos proporcionaría una forma de medir del grado de dependencia entre distritos por candidato (con $\delta_j = 0$ representando independencia y $\delta_j = 1$ covarianzas máximas), mismo que al menos podría estimarse respecto a elecciones recientes, por medio del siguiente:

Algoritmo 1

1. Calcular los intervalos extremos $I_j^\perp(\alpha)$ e $I_j^*(\alpha)$ para cada $j \in \{1, \dots, J\}$.
2. Simular una cantidad elevada de muestras estratificadas del tamaño que se haya determinado a partir del marco muestral de una elección pasada.
3. Con base en dichas simulaciones determinar cada valor δ_j de modo que el intervalo (19) efectivamente capture el verdadero valor θ_j en 100% de las simulaciones del paso 2.

Se aplicó el algoritmo anterior a los resultados de las siguientes elecciones:

Candidato	Gob2011	Pre2012	Dip2015	<i>promedio</i>	<i>mediana</i>	<i>combinado</i>
Participación	0.4638	0.3722	0.5609	0.4656	0.4638	0.4647
NULOS	0.2622	0.0976	0.0632	0.1410	0.0976	0.1193
NOREG	0.3000	0.2960	0.1534	0.2498	0.2960	0.2729
PAN	0.3521	0.4474	0.5683	0.4559	0.4474	0.4517
PRI	0.2248	0.2604	0.3345	0.2732	0.2604	0.2668
PRD	0.2987	0.3058	0.5329	0.3791	0.3058	0.3425
MORENA	NA	NA	0.3272	0.3272	0.3272	0.3272
PT	0.2987 ^a	0.3058 ^a	0.3430	0.3158	0.3058	0.3108
MC	0.2987 ^a	0.3058 ^a	0.5868	0.3971	0.3058	0.3515
PANAL	0.2248 ^b	0.0852	0.3856	0.2319	0.2248	0.2283
PES	NA	NA	0.2056	0.2056	0.2056	0.2056
PH	NA	NA	0.1482	0.1482	0.1482	0.1482
Independientes	NA	NA	NA	NA	NA	0.1769^c

Cuadro 1: Valores de δ_j por candidato de acuerdo al Algoritmo 1. NA = no aplica. ^a Valor imputado por su alianza con PRD. ^b Valor imputado por su alianza con PRI. ^c Se imputó el promedio de los valores combinados de PES y PH por ser partidos pequeños de reciente creación y que posiblemente serían el referente más cercano a lo que podría resultar de candidatos independientes que por primera vez participan en una elección de gobernador del Estado de México en la historia reciente.

- Gobernador del Estado de México en el año 2011 (Gob2011),
- Presidente de México en el año 2012, restringiéndose estrictamente a la votación en el Estado de México (Pre2012),
- Diputados federales por el Estado de México en el año 2015 (Dip2015).

En los tres casos anteriores se mapearon los resultados a la estructura de distritos locales que se utilizará en la elección de gobernador del Estado de México en el año 2017, con el objeto de que los resultados sean comparables entre sí y útiles para la elección de 2017, obteniéndose los resultados que se resumen en el Cuadro 1 con muestras estratificadas de tamaño $c = 1,200$ proporcionales a la lista nominal por distrito y con $\alpha = 0.95$ (intervalos de probabilidad 95%) para intervalos de proporción efectiva de votos de acuerdo a (5). La proporción de *participación* se calculó como 1 menos la proporción de abstencionismo (ABST).

En el Cuadro 1 el valor *combinado* es simplemente el punto medio entre el *promedio* y la *mediana*. Como puede apreciarse en todos los casos los valores de δ_j son positivos lo cual significa que un supuesto de independencia entre distritos (estratos) en dichas elecciones no era realista y que en dado caso hubiese conducido a subestimaciones, en mayor o menor grado, de las varianzas σ_j^2 y por tanto a la estimación con intervalos de probabilidad menor a la deseada de 95 %, como se resume en el Cuadro 2 donde se estiman las probabilidades de cobertura de intervalos de probabilidad 95 % bajo el supuesto de independencia entre distritos.

Candidato	Gob2011	Pre2012	Dip2015
Participación	39.23	46.67	33.10
NULOS	65.03	81.17	82.60
NOREG	75.10	32.07	69.63
PAN	47.07	40.27	35.57
PRI	51.37	51.93	47.27
PRD	51.27	50.70	35.33
MORENA			48.87
MC			37.03
PANAL		81.67	47.70
PES			64.13
PH			70.03

Cuadro 2: Probabilidades (en %) de que intervalos de 95 % de probabilidad hubiesen capturado el verdadero resultado de la elección (cobertura) bajo el supuesto de independencia entre distritos.

Referencias

J.M. Bernardo y A.F.M Smith (1994) *Bayesian Theory*. Wiley (Chichester).

G. Casella y R.L. Berger (2002) *Statistical Inference*. Duxbury (Pacific Grove).

N.L. Johnson, S. Kotz y N. Balakrishnan (1997) *Discrete Multivariate Distributions*. Wiley (New York).

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

C.-E. Särndal, B. Swensson y J. Wretman (1992) *Model Assisted Survey Sampling*. Springer (New York).

Propuesta de la Dra. Karla Beatriz Valenzuela Ocaña y el Dr. Alberto Alonso y Coria

Para estimar el porcentaje de votación que corresponde a cada partido y el porcentaje de participación ciudadana en la elección, se utilizarán los procedimientos clásicos de estimación para un muestreo aleatorio estratificado.

Como se menciona en la sección de Diseño de Muestreo, se determinó utilizar como estratos la conformación distrital local, así se tienen 45 estratos a analizar.

I. Análisis por estrato h .

a. Estimación del número de votos por candidato

Como primer paso se estima el número de votos que obtendría cada una de las fuerzas políticas en el estrato h .

Sea $v_{i,j}$ la votación por el candidato i en la casilla j y sea ln_j el listado nominal para la casilla j . El estimado total de participación para el candidato i para el estrato h analizado está dado por:

$$\hat{R}_i = \frac{\sum_{j=1}^n v_{i,j}}{\sum_{j=1}^n ln_j}$$

donde n es el número de casillas en la muestra seleccionadas del estrato h .

El cálculo de la varianza para el estimador \hat{R}_i se obtiene de la siguiente forma:

$$V(\hat{R}_i) \cong \frac{1-f}{n\overline{LN}} \cdot \frac{\sum_{j=1}^n (v_{i,j} - \hat{R}_i ln_j)^2}{n-1}$$

donde $f = \frac{n}{N}$, N representa el número total de casillas en el estrato y \overline{LN} el listado nominal promedio de las casillas en el estrato analizado.

El intervalo de confianza para la estimación del total de participación \hat{R}_i para el candidato i para el estrato analizado está dado por:

$$\hat{R}_i \pm z_{\alpha/2} * \sqrt{V(\hat{R}_i)}$$

donde $1 - \alpha$ es el nivel de confianza establecido. Para la estimación se ha definido el nivel de confianza como 95% por lo que $z_{\alpha/2} = 1.96$

El total de votos estimado para el candidato i en el estrato h es simplemente $\hat{R}_i LN$ donde LN es el listado nominal del estrato h .

b. Estimación de la participación de ciudadanos.

Es necesario estimar la participación de los ciudadanos en la elección para el estrato analizado, esto es, el número total de votos generados en el estrato.

Sea $v_{TV,j}$ el total de votos emitidos en la casilla j incluyendo votos nulos y votos de candidatos no registrados, tenemos que \hat{R}_{TV} , el estimador de la participación ciudadana en el estrato analizado, es:

$$\hat{R}_{TV} = \frac{\sum_{j=1}^n v_{TV,j}}{\sum_{j=1}^n ln_j}$$

Este procedimiento se repite para cada uno de los h estratos.

II. Integración de la información de los estratos.

Para obtener la estimación total para cada candidato, se procede a dar el peso correspondiente a cada estrato analizado. Así pues, se agrega la información correspondiente a cada estrato h , teniendo la estimación total de votos obtenidos por el candidato i :

$$v_{T,i} = \sum_h^L \hat{R}_{i,h} LN_h$$

donde LN_h es la lista nominal del estrato h y $\hat{R}_{i,h}$ es la participación total para el partido i en el estrato h y L es el número total de estratos.

La estimación del total de votos emitidos en el estado se calcula de manera semejante, considerando el peso correspondiente a cada estrato. Así:

$$v_{T, total\ de\ votos} = \sum_h \hat{R}_{TV,h} LN_h$$

Finalmente se tiene el intervalo de confianza para P_i como

$$P_i \pm z_{\alpha/2} * \sqrt{V_{T,i}}$$

El análisis se realiza para cada candidato i .

Referencias

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

Propuesta del Dr. Carlos E. Rodríguez

La estimación se realizará mediante el estimador de razón combinado y técnicas de remuestreo (*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 para la Elección para Gobernador en el Estado de México. 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 de Gobernador en el Estado de México 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 $N_h = n_h k_h \Leftrightarrow k_h = \frac{N_h}{n_h} = 1/f_h$ 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 (Distrito Local del Estado de México) se selecciona una sub-muestra SIN reemplazo de m_1 casillas (de las n_1 recuperadas).
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 a gobernador

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

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 gobernador del partido i -ésimo en todo el estado (sumando sobre todos los estratos)

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

- ✓ Se estiman los porcentajes mediante la siguiente expresión

$$\hat{p}_i = \hat{Y}_i / \sum \hat{Y}_i$$

El paso 5 concluye una iteración Bootstrap, el objetivo es realizar B iteraciones con B suficientemente grande. Para el 4 de junio de 2017 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 $f_h' = \frac{m_h}{n_h} = f_h$, esto se repite k_h 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.

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". The Canadian Journal of Statistics. Vol. 20. No. 2, pp 135-154.

III. Selección y resguardo de la muestra

1.1 Consideraciones generales para la selección de la muestra

La muestra con la que se estimará la votación a favor de los candidatos a la gubernatura del Estado de México se realizará en un acto público el viernes 2 de junio de 2017 a las 10:00 horas. En el acto estará presente un notario público que dará fe del desarrollo del protocolo desde la instalación del software requerido hasta la obtención y resguardo de la muestra definitiva.

Para seleccionar la muestra se hará uso de un equipo de cómputo debidamente blindado y habilitado con el software estadístico. El marco muestral será el listado final de las casillas aprobado para las elecciones y la selección se realizará de acuerdo con el diseño de muestra establecido anteriormente.

1.2 Protocolo de selección y resguardo de la muestra

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

Instalación

1. Personal de la Unidad de Informática y Estadística (UIE) del IEEM realizará ante notario público la validación del equipo de cómputo donde se instalará el programa para la obtención de la muestra.
2. El COTECORA entregará un disco compacto con el software necesario para la selección de la muestra (disco de insumos) junto con sus códigos de integridad (programa para la selección de la muestra y la base de datos de casillas así como sus correspondientes códigos de integridad).

3. Personal de la UIE obtendrá los códigos de integridad los archivos entregados por el COTECORA y los comparará con los entregados. El Notario Público validará que se trata del mismo código de integridad.
4. A la vista de los asistentes se instalarán el software R y R Studio.
5. Se solicitará a un miembro del COTECORA iniciar la etapa de ejecución de la selección de la muestra.

Selección de la muestra

1. Para seleccionar la muestra se requiere de un número aleatorio denominado semilla.
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. Esto permitirá garantizar que el número semilla generado sea igual al ingresada en la computadora.
4. A la vista del Notario Público, los números se ingresarán en el programa de selección de la muestra. 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 la muestra. Esta última quedará grabada en el disco duro de la computadora.
6. Se generará un código de integridad de la muestra. El código de integridad será impreso y se entregará al Notario Público y a los asistentes que lo soliciten.

Resguardo y entrega de la muestra

1. La muestra será grabada en dos discos compactos no regrabables.
2. El disco de insumos junto con uno de los discos con la muestra, los códigos de integridad y los formatos en los que se anotaron los números para construir la semilla serán guardados en un sobre sellado.
3. El segundo disco con la muestra será guardado en un sobre distinto debidamente etiquetado.

4. Asimismo, el equipo de cómputo en el cual fue generada la muestra se resguardará en un sobre.
5. Todos los sobres serán sellados y rubricados por el Notario Público.
6. El Notario Público entregará al Secretario Ejecutivo los sobres a los que se refiere el numeral anterior para su resguardo.
7. El disco con la muestra referido en el punto 3 será entregado al titular de la UIE del IEEM en una fecha y hora definidas por la Comisión Especial para la Atención del PREP y Conteos Rápidos, con el fin de que se preparen los trabajos de acopio de datos de las actas de escrutinio y cómputo el día de la jornada electoral.

IV. Proceso Logístico Operativo

1.1 Estructura general

En el Esquema 1 se presenta la estructura general del funcionamiento del Conteo Rápido 2017, el cual se desarrolla en dos ámbitos fundamentales: en la casilla y en oficinas centrales del IEEM:

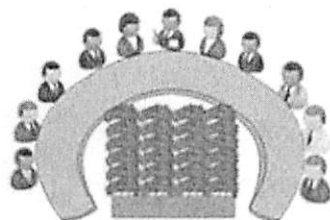
- 1) El CAE es el encargado de usar el teléfono inteligente (smartphone) proporcionado para:
 - a. Tomar al menos una fotografía del Acta de Escrutinio y Cómputo de la Casilla (AECC) con la aplicación de la cámara del teléfono.
 - b. Transcribir la información del AECC en el Formato para recopilación de resultados de la votación (FCORA).
 - c. Ingresar a la aplicación web (SICCORA) mediante su clave de usuario y contraseña.
 - d. Seleccionar la casilla a reportar, capturar la información de la votación por duplicado y adjuntar al menos una fotografía del AECC en el módulo de captura del SICCORA.
 - e. Verificar la conexión con el SICCORA y enviar la información.
- 2) La información capturada en el SICCORA se recibirá en un servidor del IEEM y de ahí se enviará al Centro de Captura del Conteo Rápido con la finalidad de ser validada.

- 3) Para validar la información, los verificadores-capturistas cotejarán la imagen de la fotografía con los datos reportados por el CAE. En caso de ser necesario se realizará la corrección correspondiente. Una vez validada, esta información se pondrá a disposición del COTECORA.
- 4) El Sistema de Captura del Conteo Rápido (SICCORA) generará cortes de información cada 10 minutos.
- 5) El COTECORA recibirá los cortes de información generados por el SICCORA para monitorear el avance de la muestra y realizar las estimaciones estadísticas.
- 6) En su momento, elaborarán un informe final sobre los resultados obtenidos y lo entregarán al Consejero Presidente del IEEM la noche del 4 de junio de 2017.

Esquema 1

Proceso Logístico Operativo del Conteo Rápido 2017 y vinculación del personal

6) Elaboración de un informe final sobre los resultados obtenidos para el Consejero Presidente del IEEM la noche del 4 de junio de 2017.



5) El COTECORA recibe los cortes de información para monitorear el avance de la muestra y realizar las estimaciones estadísticas.

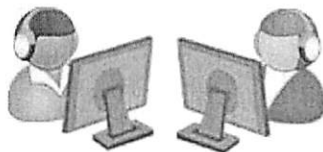
4) El SICCORA genera cortes de información cada 10 minutos.

COTECORA



3) Se valida la información por medio del cotejo de imagen fotográfica con los datos capturados por el CAE, se corrige de ser necesario y se pone a disposición del COTECORA.

Verificadores-capturistas Centro de Captura del Conteo Rápido



2) Se recibe la información en el servidor del IEEM y de ahí se envía al Centro de Captura del Conteo Rápido para su validación.

1) Transcripción del AECC en el FCORA y uso de teléfono inteligente para:

- Toma de fotografía del AECC.
- Ingreso a la aplicación web SICCORA.
- Captura de la información por duplicado y carga de fotografía.

CAE o SE (Casillas electorales)



1.2 Contingencias

En caso de que el teléfono celular no tenga enlace de Internet en la casilla:

- a) Trasladarse a algún lugar cercano a la casilla previamente ubicado donde haya señal de datos para realizar el procedimiento descrito anteriormente.
- b) De no conseguirlo, llamar por teléfono al número de contingencia del Centro de Captura.
- c) En caso de no contar con señal de voz, buscar un teléfono público que le permita llamar al Centro de Captura.

Los verificadores-capturistas del Centro de Captura del Conteo Rápido recibirán las llamadas de los CAE y capturarán la información de la AECC por duplicado y, en caso de haber alguna diferencia, solicitará se verifiquen los casos en los que existe discordancia.

1.3 Funciones del personal del INE involucrado en el ámbito distrital

- a) **Vocal Ejecutivo Distrital (VED):** Con el apoyo del Vocal de Organización Electoral, asegurará la oportuna disponibilidad y funcionamiento de todos los recursos humanos, materiales y financieros que correspondan a la realización del Conteo Rápido en el área de su competencia.
- b) **Vocal de Organización Electoral Distrital (VOED):** Es el coordinador y responsable directo de la operación logística del Conteo Rápido en el área de su competencia:
 - i. Deberá recibir la capacitación sobre el uso del SICCORA que proporcionará el IEEM en la semana del 24 al 28 de abril. Se utilizará un Manual en línea dentro del SICCORA.
 - ii. Distribuirá los teléfonos celulares inteligentes y los formatos para la recopilación de resultados a los CAE. Deberá cerciorarse de que los teléfonos cuenten con el navegador WEB en operación. Se asegurará de que cada CAE cuente con su clave de acceso al SICCORA.
 - iii. Será el responsable, exclusivo, de brindar la capacitación necesaria a los CAE y demás personal involucrado sobre las actividades que deberán desarrollar antes, durante y después de la jornada electoral en el tema.
 - iv. Previo a la Jornada Electoral o durante la misma, entablará comunicación con los CAE para dar aviso de que, al término del escrutinio y

cómputo de los votos, deberán reportar los resultados de la votación de alguna(s) de las casillas de su ARE para el conteo rápido.

- v. Mantendrá comunicación con los supervisores electorales de aquellas ARE donde se ubiquen casillas de la muestra del conteo rápido de las que no se haya recibido el reporte correspondiente, tanto en los simulacros como el día de la jornada electoral del 4 de junio de 2017.

c) Supervisores electorales:

- i. Participarán en la realización de los simulacros que se lleven a cabo, así como en aquellos que prevea realizar el IEEM de forma extraordinaria.
- ii. Verificarán que los CAE bajo su responsabilidad reporten la información de las casillas de la muestra que les corresponda.
- iii. Solicitarán a los CAE los formatos para la recopilación de los resultados de votación de las casillas de la muestra, según les haya correspondido y los resguardará para su posterior entrega al VOED.

d) Capacitadores asistentes electorales:

- i. Tomar al menos una fotografía del Acta de Escrutinio y Cómputo de la Casilla (AECC) con la aplicación de la cámara del teléfono.
- ii. Transcribir la información del AECC en el Formato para recopilación de resultados de la votación (FCORA).
- iii. Ingresar a la aplicación web (SICCORA) mediante su clave de usuario y contraseña.
- iv. Seleccionar la casilla a reportar, capturar la información de la votación por duplicado y adjuntar al menos una fotografía del AECC en el módulo de captura del SICCORA.
- v. Verificar la conexión con el SICCORA y enviar la información.
- vi. Participarán en la realización de los simulacros que se lleven a cabo, conforme a las instrucciones que al respecto les indique el VOED.

1.4 Funciones del personal involucrado en el ámbito del IEEM

a) Verificadores-capturistas del Centro de Captura del Conteo Rápido:

- i. Participarán en la realización de las pruebas de captura y los simulacros.

- ii. Para validar la información, cotejarán la imagen de la fotografía del AECC con los datos reportados por el CAE. En caso de ser necesario se realizará la corrección correspondiente.

En caso de ser necesario:

- iii. Recibirán las llamadas de los CAE para el reporte de resultados del conteo rápido.
- iv. Capturarán los datos de la votación en el SICCORA.

b) Coordinador

- i. Dará seguimiento a las tareas necesarias para la instalación, adecuación y operación en el Centro de Captura en lo que se refiere al personal, equipos y material para la operación del Conteo Rápido.
- ii. Atenderá y pondrá en práctica cada requerimiento e instrucción que reciba de la instancia interna encargada de coordinar el desarrollo de las actividades del Conteo Rápido y es vínculo con las oficinas de la misma.
- iii. Tomará decisiones en el ámbito de operación del Centro de Captura.

c) Asistente

Auxiliará al coordinador en las tareas que le corresponden.

d) Encargado de infraestructura

Realizará las tareas necesarias para la instalación, adecuación y operación del equipo, materiales y redes.

e) Operador de sistemas

Realizará las tareas necesarias para la instalación y el adecuado funcionamiento del sistema.

f) Operador de logística

Ejecutará las acciones necesarias para asegurar la continuidad de la operación en caso de presentarse alguna contingencia.

1.5 Procedimientos

A. Antes de la Jornada Electoral

Con el objetivo de contar con información confiable en la jornada electoral y contemplar esquemas de contingencia en caso de presentarse una situación fortuita,

se llevarán a cabo al menos dos simulacros los días 14 y 21 de mayo de 2017.

Prueba de captura

El propósito de este ejercicio es que los verificadores-capturistas se familiaricen con el SICCORA, así como detectar aspectos que pudieran ocasionar errores al momento de ingresar datos e identificar posibles fallas en el funcionamiento del sistema para proceder a su corrección.

Previo al primer simulacro, se realizarán diversas pruebas de captura, en las que participarán los verificadores-capturistas, bajo la coordinación y supervisión de la UIE.

Asimismo, el SICCORA estará abierto para que tanto VOED como CAE puedan realizar pruebas de captura como parte de su capacitación con el objetivo de que se familiaricen con el SICCORA, los formatos y mecanismos de contingencia.

Simulacros

La realización de los simulacros tiene como objetivo probar el funcionamiento de los medios de comunicación asignados a los CAE desde campo y que éstos ejecuten adecuadamente los procedimientos de captura y/o reporte de datos, verificar la correcta transmisión de la información y comprobar el funcionamiento del SICCORA, con la finalidad de detectar oportunamente cualquier posible falla en esos aspectos y realizar los ajustes necesarios para garantizar el puntual desarrollo de la operación logística el día de la jornada electoral.

Se realizarán al menos dos simulacros, los días 14 y 21 de mayo de 2017. En ellos participarán los CAE, bajo la coordinación y supervisión del VOED, así como los verificadores-capturistas bajo la supervisión de la UIE y el COTECORA.

Previo a la realización de los simulacros, el IEEM definirá los lineamientos con las precisiones correspondientes.

Capacitación del personal involucrado

Para llevar a cabo la capacitación del personal involucrado en la ejecución del Censo Rápido se elaboraron dos documentos principales:

- a) Guía de procedimientos de la operación logística en el ámbito distrital y

- b) Guía de capacitación para la ejecución de la operación logística en el ámbito distrital.

Ambos documentos serán la base para que los VOED lleven a cabo la capacitación del coordinador distrital, los operadores de cómputo, los capacitadores-asistentes y los supervisores electorales, con la finalidad de instruirles sobre las funciones y procedimientos que deberán realizar el día de la jornada electoral en el marco del Conteo Rápido.

B. Durante la Jornada Electoral

El sábado 3 de junio, la DEOE hará del conocimiento de los VOED el listado con las casillas de las muestra y local de su Distrito para que al medio día del 4 de junio se comuniquen con los CAE de las ARE donde aquellas se ubican, para instruirles que deberán realizar el reporte de resultados de la votación para el conteo rápido, según corresponda.

El VOED identificará las casillas de las muestra y el ARE a que pertenecen con la finalidad de detectar aquellos casos en que algún CAE deba reportar resultados de dos o más secciones, a efecto de instruir al SE a cargo de la ZORE respectiva, para que asista al CAE en estas funciones.

El VOED deberá precisar a los CAE que únicamente podrán reportar la información de los resultados de la votación de las casillas las muestras mediante las claves de autenticación que se les hayan proporcionado.

Recopilación de la información

Para realizar la recopilación de los datos de la votación de las casillas de las muestras, el CAE o el SE deberá:

- a. Presentarse a las 18:00 horas en la casilla de la muestra asignada a su ARE o ZORE, de la cual deba realizar el reporte de resultados de la votación del conteo rápido.
- b. Deberá informarle al Presidente de la Mesa Directiva que su casilla fue seleccionada para efectos del ejercicio de conteo rápido.
- c. Procederá como sigue:

Una vez que los funcionarios de casilla hayan concluido con el escrutinio y cómputo de la elección de Gobernador y llenado el acta respectiva, tomará al menos una fotografía del AECC con la aplicación de la cámara del teléfono,

transcribirá la información del AECC en el FCORA, capturarán los datos de votación de cada partido político, coalición, candidatos independientes, candidatos no registrado y votos nulos en la aplicación WEB de su teléfono móvil. De acuerdo al sistema deberán hacerlo dos veces. Finalmente enviará la información. Si no existiera señal de internet inmediatamente deberá proceder a realizar el reporte telefónico conforme se explica en el apartado 1.2 Contingencias.

C. Después de la Jornada Electoral

Los CAE a quienes les haya correspondido reportar datos de una casilla de la muestra para el conteo rápido, deberán hacer entrega del o de los formatos para recopilaron de resultados a su SE, quien se encargará de concentrarlos y entregarlos a la brevedad posible al VOED.

El VOED deberá elaborar una relación de estos documentos y conservarlos debido a que, una vez que concluya la jornada electoral, la DEOE le dará instrucciones precisas para su envío al IEEM.

Estos documentos serán utilizados para realizar una evaluación integral sobre la operación logística implementada para el conteo rápido.

1.6 Sistema Informático

A. Módulos y funcionalidad de la aplicación web

La UIE del IEEM desarrolló el Sistema de Captura del Conteo Rápido (SICCORA) mediante el cual se capturarán los datos de las Actas de Escrutinio y Cómputo de la Casilla. Este sistema informático cuenta con distintos módulos de acuerdo a los perfiles de usuario.

El SICCORA cuenta con tres perfiles, el perfil de administrador, el de usuario (CAE) y el de verificador-capturista. El primero será utilizado por los miembros del COTECORA, el segundo por los CAE y el tercero por los verificadores-capturistas. Los módulos por perfil son los siguientes:

1. Perfil de administrador

- Administración de usuarios.
- Carga de casillas sorteadas y pertenecientes a la muestra.
- Generación de cortes de información.
- Monitoreo de avance.

2. Perfil de usuario (CAE)

- Captura. Este módulo le permitirá al CAE realizar tanto el registro de los datos contenidos en el acta PREP: número de casilla y cantidad de votos por actor político; así como cargar una o dos fotografías del acta.
- Consulta. Por medio de este módulo el CAE puede visualizar las capturas que ha realizado.

3. Perfil de verificador-capturista

Consta de los mismos módulos que el perfil de CAE con la diferencia de que el verificador-capturista podrá realizar modificaciones en las capturas, no así el CAE, quien no tendrá la posibilidad de editarlas.

Todos los perfiles de usuario tienen acceso al manual en línea del SICCORA mediante la liga correspondiente ubicada en la página principal del sistema.

B. Mecanismos de seguridad y continuidad para la operación del sistema

En relación con la seguridad, para aumentar la certeza y confiabilidad en la recolección y transmisión de los datos de las Actas de Escrutinio y Cómputo de la Casilla se han tomado las medidas descritas a continuación.

1. Sobre el SICCORA se generará un usuario distinto asociado a cada CAE.
2. Se generarán lotes de contraseñas (passwords) diferentes para la capacitación de los CAE, para cada simulacro y otro más para el día de la jornada electoral. En este último caso las contraseñas serán informadas al INE al mismo tiempo que se entregue la información de la muestra a utilizarse, es decir, con la menor antelación posible para prevenir fugas de información y así aumentar la seguridad en el acceso al SICCORA.
3. Las contraseñas serán de ocho posiciones mezclando dígitos, letras mayúsculas y minúsculas junto con signos de puntuación y signos especiales.

4. Las conexiones al SICCORA se realizarán mediante protocolo HTTPS, la versión segura del protocolo de transferencia de hipertexto.
5. El INE habrá de utilizar un sistema de Master Data Management (MDM) que permite configurar el software y todas las opciones en los teléfonos de los CAE de forma remota y en tiempo real, lo cual permite prever en una gran medida el uso incorrecto de los aparatos celulares.
6. La plataforma del SICCORA ha sido ampliamente probada por el Sistema Integral de Monitoreo de Espectaculares y Medios Alternos (SIMEMA) que a la fecha cuenta con más de 300,000 imágenes en su repositorio y alrededor de 130,000 registros en la base de datos.

En relación con la continuidad, en adición con las medidas tomadas para casos de contingencia en campo, se habrá de contar con redundancia, tanto en los equipos de cómputo que albergan la aplicación del SICCORA (infraestructura de cómputo), como en los enlaces a Internet que permiten la comunicación con los dispositivos móviles y el Centro de Captura del Conteo Rápido (infraestructura de comunicaciones).