

**Comité Técnico Asesor para el Conteo Rápido de la Elección
Extraordinaria para Gobernador de Colima**

**Criterios científicos, logísticos y operativos para
la realización del Conteo Rápido y protocolo
para selección de muestra**

Diciembre, 2015

Contenido

I. Antecedentes	3
II. Criterios científicos	4
1.1 Esquema de muestreo	4
1.2 Procedimiento de estimación.....	13
1.3 Integración de las estimaciones del COTECORA	21
III. Selección y resguardo de la muestra	22
1.1 Consideraciones generales para la selección de la muestra	22
1.2 Protocolo de selección y resguardo de la muestra	22
IV. Procedimientos logísticos y operativos para el acopio de datos de las actas	25
1.1 Objetivo general de la operación logística	25
1.2 Esquema general de funcionamiento	26
1.3 Simulacros	31

I. Antecedentes

El 11 de noviembre de 2015, el Consejo General del Instituto Nacional Electoral aprobó, mediante acuerdo INE/CG957/2015, la realización del Conteo Rápido, basado en Actas de Escrutinio y Cómputo de Casilla correspondientes a la Elección Extraordinaria de Gobernador en el estado de Colima con el fin de conocer las tendencias de los resultados de la votación el día de la jornada electoral, asimismo aprobó la integración del Comité Técnico Asesor en la materia (COTECORA).

En apego a las disposiciones normativas en la materia, el COTECORA desarrolló los criterios científicos, además conoció el diseño de los procedimientos logísticos y operativos para el acopio de los datos de las Actas de Escrutinio y Cómputo de Casilla, así como el protocolo para la selección de la muestra, y los consideró pertinentes para los fines del Conteo Rápido. Los criterios científicos, logísticos y operativos así como el protocolo de selección de la muestra se describen en este documento.

II. Criterios científicos

Los criterios científicos del Conteo Rápido de la elección extraordinaria de Colima 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 y el porcentaje de ciudadanos que acudieron a votar.

El enfoque estadístico que se usará para la estimación de los resultados de la elección es el Clásico (algunas veces llamado Frecuentista). Cabe mencionar que, dentro de este enfoque, es posible incorporar distintos matices metodológicos.

Se aplicarán diversos procedimientos de estimación, no se esperan resultados idénticos porque 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 referidos al porcentaje de votos a favor de cada candidato a la gubernatura de Colima.

1.1 Esquema de muestreo

Diseño de muestreo

Las tres 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 municipios y distritos locales del estado de Colima: al interior de cada uno de los 10 municipios de Colima se realizó una división por distritos locales obteniendo 22 estratos (ver Tabla 1).

Tabla 1. Estratificación utilizada en el Conteo Rápido para la elección extraordinaria de Colima.

NO	MUNICIPIO	DISTRITO LOCAL	ESTRATO	CASILLAS 2015	DISTRIBUCIÓN DE LA MUESTRA (PRELIMINAR)
1	ARMERÍA	09	ARMERIA_09	41	20
2	COLIMA	01	COLIMA_01	50	25
3	COLIMA	02	COLIMA_02	76	38
4	COLIMA	03	COLIMA_03	58	29
5	COLIMA	05	COLIMA_05	34	17
6	CÓMALA	04	COMALA_04	32	16
7	COQUIMITLÁN	05	COQUIMITLAN_05	33	16
8	CUAUHTEMOC	06	CUAUHEMOC_06	43	21
9	IXTLAHUACÁN	03	XTLAHUACA_03	12	6
10	MANZANILLO	11	MANZANILLO_11	56	28
11	MANZANILLO	12	MANZANILLO_12	54	27
12	MANZANILLO	13	MANZANILLO_13	56	28
13	MANZANILLO	14	MANZANILLO_14	50	25
14	MINATITLÁN	14	MINATITLAN_14	15	9
15	TECOMÁN	09	TECOMAN_09	9	5
16	TECOMÁN	10	TECOMAN_10	43	21
17	TECOMÁN	15	TECOMAN_15	41	20
18	TECOMÁN	16	TECOMAN_16	44	22
19	VILLA DE ÁLVAREZ	04	VILLA_DE_ALVAREZ_04	29	14
20	VILLA DE ÁLVAREZ	06	VILLA_DE_ALVAREZ_06	16	8
21	VILLA DE ÁLVAREZ	07	VILLA_DE_ALVAREZ_07	54	27
22	VILLA DE ÁLVAREZ	08	VILLA_DE_ALVAREZ_08	57	28
TOTAL				903	450

La muestra total será de 450 casillas, esperando recibir información de al menos 350 casillas. La distribución de casillas en cada estrato se realizará vía afijación proporcional u óptima. 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 indica en la Tabla 1.

Consideraciones para definir el tamaño de muestra y diseño de muestreo

Ideas básicas del muestreo probabilístico para el Conteo Rápido de Colima

El objetivo del Conteo Rápido es estimar la proporción p de votos en favor de un candidato a gobernador¹. Entonces, 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, 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 error máximo aceptable en la estimación. La expresión (1) se puede escribir de forma equivalente en términos de intervalos de confianza

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

Esto significa que si, por ejemplo, se extraen 100,000 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

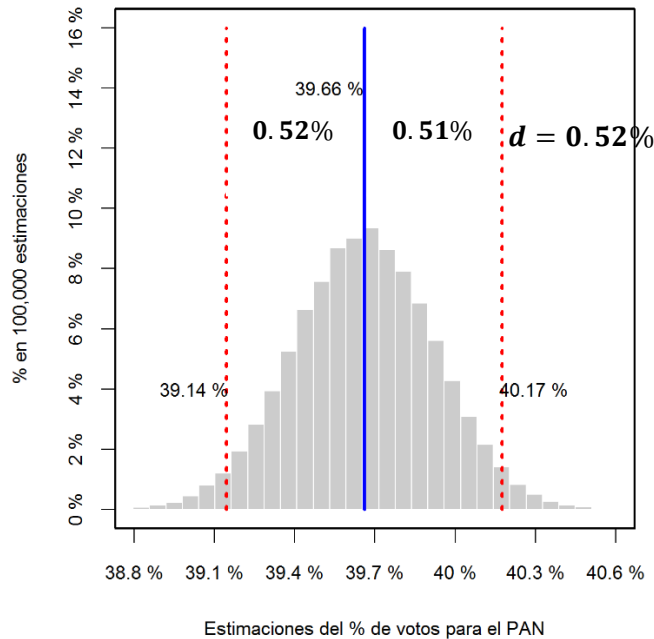
$$\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5, \dots, \hat{p}_{100,000}$$

al menos 95,000 estimaciones cumplirán con la condición (1).

Precisión en la estimación

Lo anterior nos brinda un procedimiento mediante simulación para obtener las precisiones de varias estrategias de selección y estimadores usando diferentes tamaños de muestra y así decidir cuál es el que mejor se ajusta a los objetivos del Conteo Rápido. Por ejemplo, en la Gráfica 1 se seleccionaron 100,000 muestras cada una de 350 casillas siguiendo un muestreo estratificado por municipio y distritos locales (22 estratos).

¹ 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 es mejor concentrarse en la estimación para un solo contendiente.



Gráfica 1. Histograma de 100,000 estimaciones para el porcentaje de votos en favor del PAN, cada estimación se realizó mediante un muestreo estratificado por municipios y distritos locales y usando el estimador común de razón.

Con cada una de las 100,000 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 de estratos, **N_h** es el número de casillas instaladas en el estrato **h** y **y_{hi}** es el número 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 nulos), sin embargo en este apartado sólo se muestra el análisis para el candidato que registró mayor varianza y requería un tamaño de muestra mayor para alcanzar la misma precisión que los demás partidos.

Las líneas punteadas 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%, mientras que por debajo del segundo se encuentra el 97.5% de las estimaciones por arriba el 2.5%. Entre ambos cuantiles se encuentra el 95% de las estimaciones, Entonces, regresando a las expresión (2) tenemos que

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

Y la precisión se obtendría mediante

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

Así, con la estrategia de selección, tamaño de muestra y estimador usado para generar la Gráfica 1 se obtiene una precisión de $d = 0.52\%$.

Siguiendo el mismo procedimiento se obtuvieron las precisiones para varias estrategias de selección, tamaños de muestra y estimadores.

Estrategias de selección

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

El MASS es el muestreo básico y suele usarse como parte de los demás esquemas, en este caso no se tendría control alguno de las casillas seleccionadas. En el muestreo estratificado se divide a la población en estratos/grupos² y en cada estrato se hace una selección de casillas³. 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 el MASS.

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.

Por último, tenemos el muestreo por conglomerados en donde también se divide a la población en conglomerados/grupos y la selección se realiza en varias etapas, primero se seleccionan conglomerados y posteriormente, y sólo de los conglomerados seleccionados, se eligen casillas. El muestreo por conglomerados se usa para reducir costos y facilitar la logística operativa de campo sacrificando precisión en la estimación.

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 7 de junio de 2015 en Colima.

Estratificaciones

Las estratificaciones consideradas en los ejercicios de simulación fueron nueve:

1. 2 estratos ⇒ distritos federales
2. 16 estratos ⇒ distritos locales
3. 10 estratos ⇒ municipios
4. 2 estratos ⇒ ubicación de casillas (urbano y no urbano)

Y combinaciones de las mismas

5. 4 estratos ⇒ distritos federales y ubicación de casillas
6. 29 estratos ⇒ distritos locales y ubicación de casillas
7. 17 estratos ⇒ municipios y ubicación de casillas
8. 22 estratos ⇒ municipios y distritos locales
9. 37 estratos ⇒ municipios, distritos locales y ubicación

Estimadores considerados

Se probaron dos estimadores, el estimador de razón presentado en la expresión (3) y el estimador que incorpora información auxiliar de la elección federal de diputados⁴ del 7 de junio de 2015, el estimador está dado por la siguiente fórmula

$$\hat{p}_{aux\ i} = \hat{\beta}_i X_i / \sum \hat{\beta}_i X_i \quad (4)$$

con

$$\hat{\beta}_i = \frac{\sum_{h=1}^L N_h \bar{y}_{hi}}{\sum_{h=1}^L N_h \bar{x}_{hi}}$$

En donde la información distinta a la usada en la ecuación (3) y que es necesario describir es X_i que esta dado por el número total de votos en favor del candidato a diputado del partido i -ésimo (dato real y total) y \bar{x}_{hi} que es el número promedio de votos en las casillas seleccionadas en muestra para el candidato a diputado por el partido i en el distrito h .

Resultados de las simulaciones para fijar el tamaño de muestra

Las precisiones obtenidas usando el estimador de razón sin información auxiliar, ver expresión (3), se muestran en la Tabla 2 mientras que las precisiones alcanzadas vía el estimador (4) se despliegan en la Tabla 3.

En el caso del estimador (3) las estratificaciones que incluyen a los municipios arrojan mejor precisión y lo mismo sucede en el caso del estimador (4). En la comparación de estimadores es claro que el estimador que incorpora información auxiliar (4) alcanza mejor precisión en todos los escenarios explorados.

En las elecciones presidenciales de 2006 y 2012 la precisión en la estimación que se usó en los conteos rápidos fue de $d = 0.3\%$ y $d = 0.5\%$ respectivamente, por lo que un tamaño de muestra de 350 casillas y estratificando por los municipios de Colima (en general las estratificaciones que incluyen a los

⁴ Obtenida de los cómputos distritales.

municipios) obtiene este nivel de error entre estos porcentajes con información auxiliar o sin ella.

Por último, es importante considerar que

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 se fija este tamaño de muestra tomando como referencia los resultados de una elección previa y a que tendremos un porcentaje de no respuesta que no conocemos, 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 los dos métodos de estimación usados, 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. En el Conteo Rápido para la elección federal de 2015, el tamaño de muestra de casillas ubicadas en Colima fue de 60, de las cuales finalmente llegó información de 48. Observando así un porcentaje de respuesta del 80%, tomando esto en cuenta se aumentó el tamaño de muestra a 450 casillas, esperando recuperar información de al menos 350.
4. Se contará con 231 capacitadores y asistentes electorales (CAES) y con 40 supervisores para total de 271 personas que recabarán la información de las casillas seleccionadas y procederán a su transmisión. Por lo cual es conveniente apuntar que este tamaño muestra se encuentra dentro de los límites de la capacidad operativa de campo.

Tabla 2. Precisiones usando el estimador (3), para distintos tamaños de muestra y estratificaciones (Para obtener cada uno de los valores de la tabla se realizaron 100,000 simulaciones)

n	Distritos federales	Distritos federales y ubicación	Distritos locales	Distritos locales y ubicación	Ubicación	Municipio	Municipio y ubicación	Municipio y distritos locales	Municipio, distritos locales y ubicación
200	0.92	0.92	0.86	0.84	0.99	0.82	0.80	0.79	0.77
250	0.79	0.79	0.74	0.73	0.86	0.72	0.70	0.68	0.67
300	0.70	0.69	0.64	0.64	0.75	0.62	0.61	0.59	0.58
350	0.62	0.61	0.57	0.57	0.66	0.55	0.54	0.52	0.52
400	0.55	0.55	0.51	0.51	0.60	0.49	0.48	0.47	0.46
450	0.49	0.49	0.46	0.46	0.53	0.44	0.44	0.43	0.42
500	0.44	0.44	0.41	0.40	0.48	0.39	0.39	0.37	0.37
550	0.39	0.39	0.36	0.36	0.42	0.35	0.35	0.34	0.33
600	0.35	0.34	0.32	0.32	0.38	0.31	0.31	0.30	0.29
650	0.31	0.30	0.28	0.28	0.33	0.27	0.27	0.26	0.25
700	0.26	0.26	0.24	0.24	0.29	0.24	0.23	0.23	0.22
750	0.22	0.22	0.20	0.20	0.24	0.20	0.19	0.19	0.19
800	0.18	0.17	0.16	0.16	0.19	0.16	0.16	0.15	0.15
850	0.12	0.12	0.11	0.11	0.13	0.11	0.11	0.10	0.10

Tabla 3. Precisiones usando el estimador (4), para distintos tamaños de muestra y estratificaciones (Para obtener cada uno de los valores de la tabla se realizaron 100,000 simulaciones)

n	Distritos federales	Distritos federales y ubicación	Distritos locales	Distritos locales y ubicación	Ubicación	Municipio	Municipio y ubicación	Municipio y distritos locales	Municipio, distritos locales y ubicación
	0.55	0.53	0.49	0.47	0.54	0.47	0.46	0.45	0.44
	0.47	0.46	0.42	0.41	0.46	0.40	0.39	0.39	0.39
	0.41	0.40	0.37	0.36	0.40	0.35	0.35	0.34	0.33
	0.36	0.36	0.33	0.32	0.36	0.31	0.31	0.30	0.30
	0.32	0.32	0.29	0.29	0.32	0.27	0.27	0.27	0.27
	0.29	0.29	0.26	0.26	0.29	0.25	0.25	0.24	0.24
	0.26	0.25	0.23	0.23	0.26	0.22	0.22	0.21	0.21
	0.23	0.23	0.21	0.20	0.23	0.20	0.20	0.19	0.19
	0.21	0.20	0.18	0.18	0.20	0.17	0.17	0.17	0.17
	0.18	0.18	0.16	0.16	0.18	0.15	0.15	0.15	0.14
	0.16	0.15	0.14	0.14	0.16	0.13	0.13	0.13	0.13
	0.13	0.13	0.12	0.12	0.13	0.11	0.11	0.11	0.11
	0.10	0.10	0.09	0.09	0.10	0.09	0.09	0.09	0.09
	0.07	0.07	0.07	0.06	0.07	0.06	0.06	0.06	0.06

1.2 Procedimiento de estimación

Propuesta de Carlos E. Rodríguez con la colaboración de Antonio Soriano

Se realizarán dos estimaciones en paralelo⁵:

- **Estimación de razón combinado.** Se usará sólo la información de la muestra de casillas correspondiente a la votación para gobernador el 17 de enero de 2016.
- **Estimación de razón combinado utilizando información auxiliar.** Se usará la información de la muestra de casillas correspondiente a la votación para gobernador del 17 de enero de 2016 y además, información de las mismas casillas pero correspondiente a la votación de diputados en la elección federal del 7 de junio de 2015. Esta última información se obtendrá de los cómputos distritales.

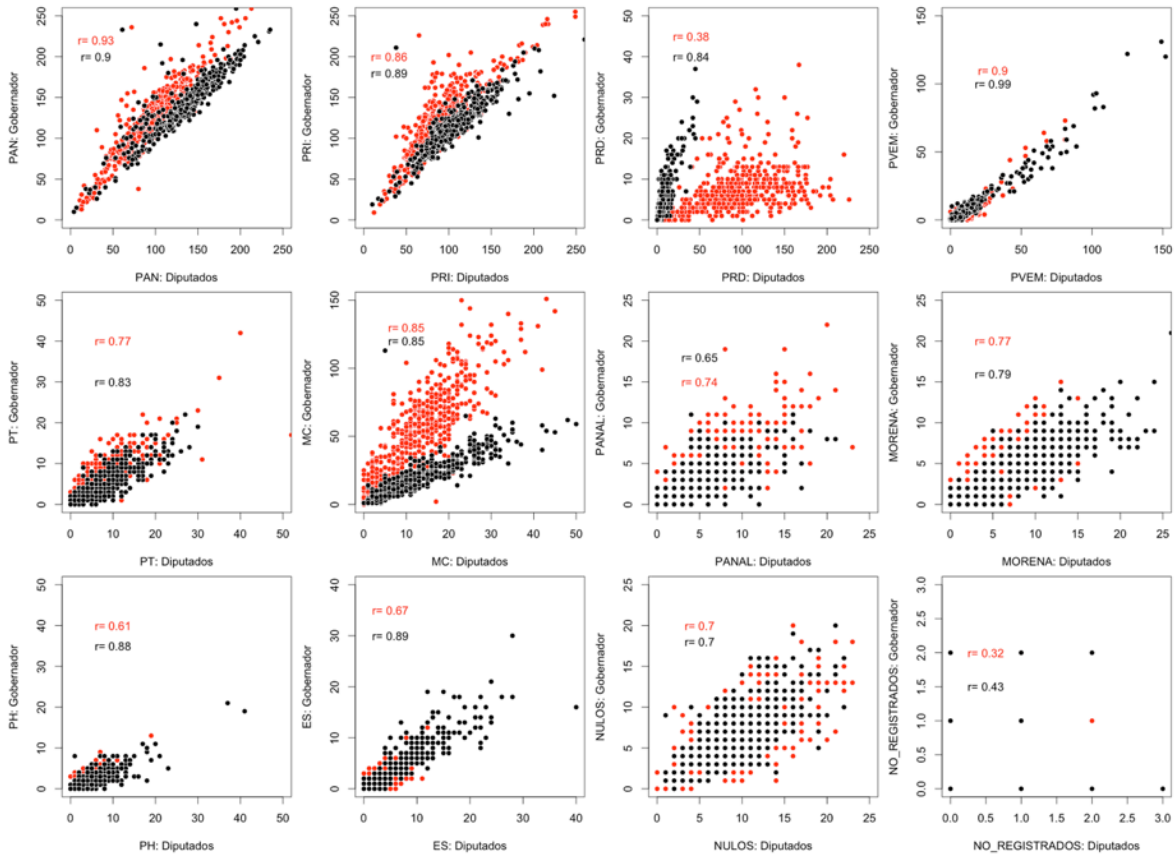
Información auxiliar

El estado de Colima se divide en dos distritos federales electorales y el 7 de junio de 2015 se realizaron elecciones para elegir un diputado en cada uno de estos dos distritos así como al gobernador del estado. En la Gráfica 1 se despliega un diagrama de dispersión que muestra la relación entre el número de votos por casilla que obtuvo cada partido con candidato a gobernador con la votación correspondiente al diputado del mismo partido político. Para diferenciar la votación en los dos distritos federales, se colorea de rojo el distrito 01 y de negro el distrito 02. La gráfica también incluye el coeficiente de correlación lineal de Pearson⁶.

⁵ Estimador de razón: se estima un cociente de dos cantidades ambas desconocidas.

⁶ El valor del coeficiente de correlación de Pearson toma valores en el intervalo $[-1, 1]$:

- Si $r = 1$, existe una correlación positiva perfecta. Existe una dependencia total entre ambas variables, cuando una de ellas aumenta, la otra también lo hace en proporción constante.
- Si $0 < r < 1$, existe una correlación positiva y esta es más fuerte cuando el valor se aproxima a 1.
- Si $r = 0$, no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes; pueden existir todavía relaciones no lineales entre las dos variables.
- Si $-1 < r < 0$, existe una correlación negativa.



Gráfica 1. Diagrama de dispersión gobernador (eje y) Vs diputados federales (eje x): El diagrama se realiza diferenciando la votación de los dos distritos federales electorales que comprende el estado de Colima (rojo y negro).

La Gráfica 1 muestra una asociación lineal muy fuerte entre la votación para diputados federales y gobernador. Esto sugiere que un buen método para estimar los votos en favor del candidato a gobernador, de cualquier partido, es suponer que existe una relación lineal del tipo

$$\hat{y}_j = \hat{\beta} x_j$$

en donde \hat{y}_j será la estimación del número total de votos en favor del candidato a gobernador (de alguno de los partidos) en la casilla j , x_j es el número de votos a favor del candidato a diputado por el mismo partido en la casilla j . Entonces, el problema se reduce a encontrar el $\hat{\beta}$ que produzca el mejor

- Si $r = -1$, existe una correlación negativa perfecta. También existe una dependencia total entre las dos variables pero ahora la relación es inversa: cuando una de ellas aumenta, la otra disminuye en proporción constante.

ajuste. Hay que notar que x_j es el valor real y se obtiene de la base de los cómputos distritales de la elección federal del 7 de junio de 2015.

Remuestreo *Bootstrap*

Para construir intervalos de confianza para el porcentaje de votación en favor de cada candidato a Gobernador, así como para el porcentaje de participación en la elección se utilizarán técnicas de re-muestreo, específicamente el método conocido como *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 extraordinaria para Gobernador en el estado de Colima; 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.

Las ideas de Sitter (1992A) aplicadas a la elección estatal de gobernador en el estado de Colima 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 de Colima 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

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

- ✓ Estimador del total de votos en el primer estrato para i-ésimo candidato a diputado (información auxiliar de los cómputos distritales para la elección federal de diputados de 2015):

$$\hat{X}_{1i} = N_1 \bar{x}_{1i}$$

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)

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

- ✓ El estimado del total de votos a favor del candidato a diputado federal del partido i-ésimo en todo el estado (sumando sobre todos los estratos)

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

- ✓ La razón de cambio entre el voto para gobernador y diputados federales

$$\hat{\beta}_i = \hat{Y}_i / \hat{X}_i$$

Lo anterior se hace para todos los partidos, candidatos no registrados y votos nulos (ver Gráfica 1).

- ✓ Las estimaciones de los porcentajes por ambos estimadores de razón, el que no usa información auxiliar

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

así como el que incorpora información auxiliar (X_i es el total de votos real para el candidato a diputado por el partido i -ésimo se obtiene de los cómputos distritales para la elección federal de diputados de 2015)

$$\hat{p}_{aux\ i} = \hat{\beta}_i X_i / \sum \hat{\beta}_i X_i$$

El paso 5 concluye una iteración *Bootstrap*, el objetivo es realizar B iteraciones con B suficientemente grande. Para el 17 de enero de 2016 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 ejecutar 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.

Propuesta de Carlos Hernández Garciadiego

Estimación

Para estimar los porcentajes de votación obtenidos por cada candidato, se utilizará como información auxiliar los datos de la elección de diputados federales de junio de 2015, agrupando los votos de los partidos de acuerdo a las alianzas que se den para esta elección.

Varianza

Para el caso de los dos candidatos más fuertes, se ordenan, por cada uno de ellos, las casillas de las elecciones de diputados federales de 2015 de acuerdo con la suma de votos obtenidos por los partidos que apoyan a dicho candidato entre el número de votos totales en dicha casilla.

Se agrupan las casillas en estratos (entre 10 y 22) de acuerdo al orden obtenido. Se usan estos estratos para calcular la varianza tanto de manera teórica como mediante simulaciones (10000) de muestreo.

Para el resto de los candidatos, se utilizan como estratos los que se vayan a usar para la selección de la muestra (municipios, distritos locales)

Estimación de la votación

En lo que sigue, para los dos candidatos más fuertes los estratos utilizados son explicados en la sección anterior, para el resto de los candidatos, los municipios.

Para estimar la votación a partir de la muestra que se vaya recibiendo, 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 partido 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 de 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

$$\bar{P}_i = \frac{\sum V_{k,i}}{\sum T_{k,i}}$$

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

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

Y la varianza global

$$V = \sum N_k^2 V_k$$

Y finalmente, el intervalo de confianza del candidato es

$$(\bar{P}_i - 1.96\sqrt{V}, \bar{P}_i + 1.96\sqrt{V})$$

Propuesta de Gabriel Núñez Antonio

Para estimar los porcentajes de votación obtenidos por cada candidato en las próximas elecciones extraordinarias para Gobernador en el Estado de Colima, se considerará una muestra de 450 casillas de las cuales, dada la experiencia de las elecciones federales de junio de 2015, se esperan recibir al menos 350 casillas el día de la elección.

Aprovechando la estratificación geo-política-electoral del Estado de Colima se considera un muestreo aleatorio estratificado donde la estratificación será al menos a nivel municipal.

Tipo de Estimadores Puntuales que se emplearán

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

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

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

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

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

Por otro lado, si se considera como información auxiliar los resultados de la elección de diputados federales de junio de 2015 (X) se puede tomar

$$\hat{T}_j = \frac{\hat{T}_{y,est}^j}{\hat{T}_{x,est}^j} T_x \quad , \quad \hat{T} = \frac{\hat{T}_{y,est}}{\hat{T}_{x,est}} T_x$$

donde $\hat{T}_{y,est}^j$ es el estimador estratificado del total de votos para el candidato j , $\hat{T}_{x,est}^j$ es el estimador estratificado del total de votos para el candidato j usando la información de la elección de diputados federales y T_x es el total de votos en la elección de diputados federales.

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, a partir de ésta, obtener los cuantiles de orden 2.5% y 97.5%, y con ellos construir el intervalo correspondiente. Esta última estrategia es la que se emplearía para construir los correspondientes intervalos de confianza, considerando un proceso de remuestreo con al menos 50,000 simulaciones. Lo anterior derivado de los diferentes estudios y análisis realizados con las bases de datos de la elección de Gobernador y de diputados federales realizadas en junio de 2015.

1.3 Integración de las estimaciones del COTECORA

Durante toda la jornada se monitoreará que las tendencias de votación indicadas por los intervalos de confianza de los tres integrantes del COTECORA vayan en el mismo sentido. Para integrar las estimaciones en una sola, se tomará para cada candidato, la unión de los dos intervalos de confianza de menor longitud.

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 de Colima se realizará en un acto público el jueves 14 de enero de 2016. 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 habilitado con software estadístico. El marco muestral será el listado de las casillas aprobado para las elecciones y la selección se realizará de acuerdo con el diseño de muestreo establecido.

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 Dirección Ejecutiva del Registro Federal de Electores (DERFE) 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 el software necesario para la selección de la muestra 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 DERFE obtendrá los códigos de integridad de 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.
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.
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 de la muestra

1. La muestra será grabada 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 datos de las actas de escrutinio y cómputo.

3. El otro disco 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.

Asimismo el equipo de cómputo en el cual fue generada la muestra se resguardará en un sobre.

Ambos sobres serán sellados y rubricados por el Notario Público.

4. El Notario Público entregará al Secretario Ejecutivo del Instituto Nacional Electoral o a la persona que él designe, los sobres a los que se refiere el numeral anterior para su resguardo.

IV. Procedimientos logísticos y operativos para el acopio de datos de las actas

La operación logística estará a cargo de la Dirección Ejecutiva de Organización Electoral y se desarrollará sobre la estructura definida para el funcionamiento del *Sistema de Información sobre el desarrollo de la Jornada Electoral (SIJE)*.

Este sistema está planeado para que opere desde cada sede distrital a partir de la información recopilada por los capacitadores-asistentes electorales (CAE) en sus recorridos por las casillas. Como parte de este sistema se tiene prevista la instalación de una Sala del SIJE en cada una de las dos Juntas Distritales Ejecutivas (JDE) en las que se dispondrá de los recursos materiales necesarios -líneas telefónicas habilitadas con diademas y equipos de cómputo conectados a la RedINE- para que los operadores de cómputo reciban las llamadas que realizan los CAE, a través del medio de comunicación previamente asignado, y realicen la captura de los datos en el sistema informático correspondiente.

Bajo esta dinámica, se tiene planeado que los CAE o en su caso los supervisores electorales (SE), realicen la recopilación de los datos contenidos en las Actas de Escrutinio y Cómputo de Casilla y, utilizando los formatos diseñados para tal fin, reporten a las Salas del SIJE los datos de la votación para Gobernador del estado de Colima de las casillas de la muestra.

1.1 Objetivo general del operativo logístico

Proveer de manera confiable y oportuna la información de los resultados de la votación de las casillas de la muestra asentados en las actas de escrutinio y cómputo de casilla al Comité Técnico Asesor para el Conteo Rápido (COTECORA), con la finalidad de que elabore las estimaciones estadísticas para

conocer las tendencias de la votación de la elección extraordinaria de Gobernador del estado de Colima, el próximo 17 de enero de 2016.

1.2 Esquema general de funcionamiento

Se desarrolla en dos ámbitos fundamentales, distrital y en oficinas centrales, conforme se explica a continuación:

1. El CAE, o en su caso el SE, es el encargado de recabar en el formato diseñado para tal fin, los datos de votación contenidos en las actas de escrutinio y cómputo de la(s) casilla(s) seleccionada(s) de la muestra que se encuentre(n) asignada(s) a su Área de Responsabilidad Electoral (ARE). De manera inmediata llama, a través del medio de comunicación que le fue asignado (teléfono celular principalmente), a la Sala del SIJE de la sede distrital correspondiente para reportar la información recopilada.
2. El operador de cómputo en la Sala del SIJE recibe la llamada del CAE, o en su caso del SE, y captura directamente en el *Sistema de Información del Conteo Rápido* los datos que éste le comunique, para su transferencia inmediata a la sede del comité.
3. El COTECORA procesa la información proporcionada por el sistema y realiza las estimaciones estadísticas que considere convenientes.

Funciones del personal involucrado en el ámbito distrital

- ◆ *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 se utilizarán para el Conteo Rápido.

- ◆ *Vocal de Organización Electoral Distrital (VOED):* Es el coordinador y responsable directo de la operación logística del Conteo Rápido:
 - ✓ Brindará la capacitación necesaria a los CAE y SE y demás personal involucrado sobre las actividades que deberán desarrollar antes, durante y después de la jornada electoral en el tema.
 - ✓ Distribuirá los formatos para la recopilación de datos a los CAE, y en su caso a los SE.
 - ✓ 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.
 - ✓ Decidirá la participación de los SE cuando algún CAE deba reportar dos o más casillas de la muestra situadas en distintos domicilios y que se encuentren alejados entre sí; en tal caso, se comunicará con el SE que corresponda para indicarle su asistencia en el reporte de datos en un ARE determinada.
 - ✓ Mantendrá comunicación con los SE de aquellas ARE donde se ubiquen casillas de las que no se haya recibido el reporte de Conteo Rápido correspondiente.

- ◆ *Coordinador Distrital del SIJE:* Apoyará al VOED en el cumplimiento de sus funciones:
 - ✓ Apoyará a los operadores de cómputo en caso de que se presenten fallas en el sistema informático.
 - ✓ Verificará la personalidad de los CAE o SE que por alguna razón no cuenten con su clave de autenticación para el reporte de datos.

◆ *Supervisores electorales:*

- ✓ Participarán en la realización de los simulacros.
- ✓ Verificarán que los CAE bajo su responsabilidad reporten la información de las casillas de la muestra que les corresponda.
- ✓ Apoyarán, de ser necesario, en el acopio y reporte de datos de la votación de las casillas incluidas en la muestra, cuando algún CAE tenga que reportar dos o más situadas en diferentes domicilios.

◆ *Capacitadores-asistentes electorales:*

- ✓ Participarán en la realización de los simulacros.
- ✓ Recopilarán los resultados de la votación de la elección extraordinaria de Gobernador en Colima que hayan sido anotados en las actas de escrutinio y cómputo de las casillas de la muestra.
- ✓ Reportarán inmediatamente los resultados a la Sala del SIJE de su respectiva sede distrital, utilizando el medio de comunicación que se le haya asignado.

◆ *Operadores de cómputo:*

- ✓ Participarán en la realización de la(s) prueba(s) de captura y de los simulacros.
- ✓ Recibirán las llamadas de los CAE y SE.
- ✓ Capturarán los datos de la votación en el *Sistema de Información del Conteo Rápido*.

Capacitación del personal involucrado

Para llevar a cabo la capacitación del personal involucrado en la ejecución del Conteo 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, particularmente se hará énfasis en los siguientes procedimientos:

Recopilación de la información

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

- a) Presentarse a las 18:00 horas en la primera casilla de la muestra asignada a su ARE o ZORE de la cual deba realizar el reporte de resultados de la votación al Conteo Rápido.
- b) Una vez que los funcionarios de casilla hayan concluido con el escrutinio y cómputo de la elección extraordinaria de Gobernador y llenado el acta correspondiente, transcribirá los datos de votación de cada partido político, coalición, candidato independiente en su caso, candidato no registrado y votos nulos, en el formato diseñado para tal fin.
- c) Inmediatamente se comunicará a la Sala del SIJE a través del medio de comunicación asignado y, en cuanto entre la llamada, procederá a realizar el reporte de los datos.

- d) En su caso, se trasladará a la siguiente casilla de la muestra y procederá a ejecutar los pasos indicados previamente.

Reporte y captura de los datos

- a) El operador de cómputo en la Sala del SIJE recibirá la llamada del CAE, o en su caso del SE, y registrará directamente en el *Sistema de Información del Conteo Rápido* los datos que éste le transmita, para su transferencia automática a la sede del COTECORA.
- b) El operador de cómputo escuchará la información dictada por el CAE o SE, a través de una diadema telefónica. Una vez en el *Sistema de Información del Conteo Rápido*, en cada campo de datos irá seleccionando o capturando la información que le sea transmitida. Al respecto:
- ✓ El operador de cómputo establecerá el ritmo del dictado y las pautas para que el CAE o SE continúe con el reporte.
 - ✓ La clave de autenticación es fundamental para el ingreso de los datos.
 - ✓ El operador de cómputo indicará al CAE cuándo puede iniciar la repetición de los resultados de la votación, considerando el mismo orden.
 - ✓ El CAE o SE dictará por segunda ocasión los resultados como medida de control.
 - ✓ Concluida la captura de los datos, el operador de cómputo guardará la información en el sistema y, una vez que se asegure que ésta ha sido enviada, indicará al CAE que puede concluir la llamada. Por ningún motivo deberán colgar el teléfono ni el CAE o SE, ni el operador de cómputo, en tanto no se haya enviado la información al Sistema de Información del Conteo Rápido.
 - ✓ El CAE o SE guardará el formato de recopilación de datos para su entrega al VOED.

- c) En caso de que el CAE o SE no pueda realizar el reporte de datos, se habilitará un esquema de contingencia en la Junta Local Ejecutiva para recibir las llamadas única y exclusivamente cuando se presente alguno de los problemas siguientes:
- ✓ Cuando en alguno de los distritos se presente interrupción en la energía eléctrica o en la conexión a la red y el VOED haya dado aviso de ello a la DEOE y al Centro de Atención a Usuarios de la UNICOM. De ser posible, el VOED deberá marcar a los CAE y SE que aún no hayan reportado datos para informarles de la situación e indicarles que se comuniquen directamente a la Junta Local Ejecutiva.
 - ✓ Cuando se presenten fallas en las líneas telefónicas de la Sala del SIJE de manera permanente.

1.3 Simulacros

La realización de estos eventos tiene como objetivo probar el funcionamiento de los medios de comunicación asignados a los CAE o SE desde campo y que éstos ejecuten adecuadamente los procedimientos de reporte de datos, verificar la correcta captura y transmisión de la información y comprobar el funcionamiento del sistema informático, 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 dos simulacros, el primero el día 20 de diciembre de 2015 y el segundo el 10 de enero de 2016, posterior a los simulacros del SIJE. En ellos participarán los CAE, los SE y los operadores de cómputo, bajo la coordinación y supervisión del VOED con apoyo del Coordinador Distrital.

Previo a la realización de estos eventos, se remitirán por parte de la DEOE los lineamientos con las precisiones correspondientes, entre las que se destacan las siguientes:

- ✓ Cada CAE reportará los resultados ficticios de la votación de un número definido de casillas de su ARE que previamente le proporcionará el VOED.
- ✓ Los SE también reportarán resultados de algunas de las casillas asignadas a su Zona de Responsabilidad Electoral (ZORE), y se encargarán de dar seguimiento al reporte de los CAE a su cargo.
- ✓ El VOED entregará a los CAE y SE los formatos prellenados, los cuales se utilizarán en los dos simulacros.