

EL CONTEO RÁPIDO DE VERACRUZ 2016: ASPECTOS ESTADÍSTICOS Y DE LOGÍSTICA

The Veracruz 2016 Quick Count: Statistical and Logistic Aspects

Sergio Francisco JUÁREZ CERRILLO¹

Luis Enrique NIETO BARAJAS²

Fecha de recepción: 30 de septiembre de 2017.

Fecha de aceptación: 11 de agosto de 2018.

RESUMEN: El pasado 5 de junio de 2016 se celebraron elecciones para gobernador en el estado de Veracruz, en este proceso electoral se llevó a cabo el conteo rápido respectivo.³ Dicho ejercicio fue implementado y organizado por el Organismo Público Local Electoral de Veracruz, integrándose por un comité técnico asesor que tuvo entre sus funciones proponer los criterios científicos, logísticos y operativos para estimar los resultados de las elecciones con el conteo rápido. El comité técnico, con una muestra de 917 casillas extraída de una población de 10,421, estimó mediante intervalos las votaciones para siete fuerzas políticas. Las estimaciones acertaron en seis de ellas, incluyendo al ganador (coalición pan-prd) quien obtuvo una votación oficial de 34.40 por ciento. El intervalo de estimación al 95 por ciento de confianza proporcionado por el comité técnico para este candidato fue [33.32 por ciento, 34.79 por ciento]. El presente artículo muestra los detalles técnicos relevantes con respecto al diseño muestral y los estimadores que se utilizaron en el conteo rápido, así como los aspectos logísticos del mismo.

1 Investigador de tiempo completo en la Facultad de Estadística e Informática de la Universidad Veracruzana. Correo electrónico: sejuarez@uv.mx.

2 Profesor de tiempo completo. Instituto Tecnológico Autónomo de México: Correo electrónico: lnieto@itam.mx.

3 Nota del equipo editorial: anterior a la reforma político-electoral de 2014, la autoridad administrativa electoral de la entidad era el Instituto Electoral Veracruzano, posterior a esta recibió el nombre de Organismo Público Local de Veracruz, siendo entonces la primera vez que en el estado fue llevado a cabo el conteo rápido bajo la nueva lógica estructural electoral nacional-local.

Palabras clave: elecciones, diseño muestral, metodología estadística para conteos rápidos, estimación de resultados electorales.

ABSTRACT: The past June 5, 2016, there was governor election in Veracruz state, Mexico. For this election a quick count was conducted for first time in the electoral history of Veracruz.⁴ This exercise was implemented by the Local Public Electoral Organism of Veracruz. A technical committee of expert advisers was integrated. Among the obligations of this committee were to propose scientific, logistic, and operations criteria to estimate the results of the elections via a quick count. The technical committee used a sample of 917 poll stations from a population of 10,421. Using confidence intervals, the results of the elections were estimated for seven political forces. Six of them were correctly predicted including the winner candidate (PAN-PRD coalition) who had a 34.40 percent of official voting. The interval with 95 percent of confidence degree provided by the committee for this candidate was [33.32%, 34.79%]. In this article we present some of the details of the sampling design and the estimators that were used in the quick count. We finally mention the general aspects of the logistics involved in this quick count.

Keywords: elections, sampling design, statistical methods for quick counts, estimation of elections results.

INTRODUCCIÓN

En la democracia los procesos electorales tienen en la jornada electoral la etapa más importante ante el escrutinio público, en razón de lo anterior es que se han desarrollado mecanismos que generan certidumbre y confianza de que la voluntad ciudadana manifestada en las urnas se reconoce y respeta por las autoridades (electorales y no electorales) respectivas. Algunos de los mecanismos de observación electoral que se han diseñado son las encuestas de salida, los programas de resultados preliminares y los conteos rápidos.

Un conteo rápido (CR) es un ejercicio estadístico para predecir con oportunidad y prontitud los resultados de la elección con base en los conteos de votos observados en una muestra probabilística de casillas.

4 Nota del equipo editorial: by Organismo Público Local de Veracruz.

En la actualidad los cr han evolucionado tecnológicamente, convirtiéndose en instrumentos de monitoreo de procesos electorales confiables (Estok *et al.*, 2002). En México se han usado en las elecciones presidenciales de 1994, 2000, 2006 y 2012,⁵ y, recientemente, en elecciones estatales de gobernador. En el caso de Veracruz, en noviembre de 2015 el Consejo General del Organismo Público Local Electoral (OPLE) de Veracruz aprobó la implementación y operación de un cr para la elección de gobernador; posteriormente, el 4 de diciembre de 2015, el ople realizó la integración del comité técnico asesor para el conteo rápido (COTECORA), quedando conformado por tres académicos de la Universidad Veracruzana, uno del Instituto Tecnológico Autónomo de México y uno del Centro de Investigación y Docencia Económica.

El COTECORA propuso un diseño muestral aleatorio estratificado en el cual los 30 distritos electorales estatales configuraron los estratos. Se usaron estimadores de la teoría de muestreo de poblaciones finitas y estimadores bayesianos. Los intervalos de estimación que se obtuvieron para las tres principales fuerzas políticas de Veracruz fueron: coalición PAN-PRD (33.32 por ciento, 34.79 por ciento), coalición PRI-PVEM-PANAL-AVE-PFCRN (29.01 por ciento, 30.40 por ciento) y MORENA (26.48 por ciento, 28.20 por ciento); los respectivos resultados oficiales para estas fuerzas políticas fueron 34.40 por ciento, 30.29 por ciento y 26.40 por ciento.

En este trabajo se hace una relatoría de los aspectos estadísticos y logísticos del CR de Veracruz, con especificaciones breves de su diseño muestral (Sección I), aspectos logísticos (Sección II), procedimientos de estimación (Sección III), los resultados obtenidos (Sección IV) y, finalmente, reflexiones en torno al mismo (Conclusiones).

I. DISEÑO MUESTRAL

El artículo 23 del documento “Lineamientos Generales del Instituto Nacional Electoral para el Diseño, Implementación y Operación de los Conteos Rápidos de Carácter Institucional en los Procesos Electorales Locales 2015-2016” estableció los siguientes alcances de la muestra para el CR:

5 Nota del equipo editorial: de la misma forma, en la elección presidencial de 2018.

Artículo 23. La muestra con la que se inferirán los resultados de la elección deberá cumplir con las siguientes características:

- I. Que todas y cada una de las casillas instaladas en la demarcación electoral de que se trate tengan una probabilidad conocida y mayor que cero de ser seleccionadas.
- II. Que se utilice un procedimiento aleatorio para la selección de la muestra que respete las probabilidades de selección determinadas por el diseño.
- III. Que la muestra se diseñe con el objetivo de obtener estimaciones con una precisión tal que genere certidumbre estadística, en el cumplimiento de los objetivos requeridos por el tipo de elección.
- IV. Que la precisión arriba señalada se obtenga con un nivel de confianza del 95 por ciento.
- V. Que considere la posibilidad de que abarque la mayor dispersión geográfica electoral posible.

Para lograr el cumplimiento de estas disposiciones, el COTECORA diseñó dos estudios de simulación Monte Carlo con el objetivo de establecer comparaciones, ambos se condujeron de manera independiente y paralela, usando como marco muestral los resultados de 9,827 casillas de la elección gubernamental de Veracruz en 2010.

ESTUDIO DE SIMULACIÓN I

Veracruz es un estado con 212 municipios y 30 distritos locales (Figura 1), cada distrito electoral está dividido en secciones electorales, las cuales pueden ser del tipo rural o urbano. Para este estudio se consideraron las siguientes características en las estratificaciones señaladas:

- i. Distritos locales: 30 estratos
- ii. Distritos locales y tipo rural o urbano: 60 estratos
- iii. Municipios: 212 estratos

Sea p_j la proporción efectiva de votos a favor del candidato j . Con base en una muestra aleatoria estratificada de tamaño n se propone el estimador de razón \hat{p}_j de p_j definido como:

$$\hat{p}_j = \frac{\sum_{k=1}^K N_k \bar{X}_{kj}}{\sum_{k=1}^{J-1} \sum_{k=1}^K N_k \bar{X}_{kj}}$$

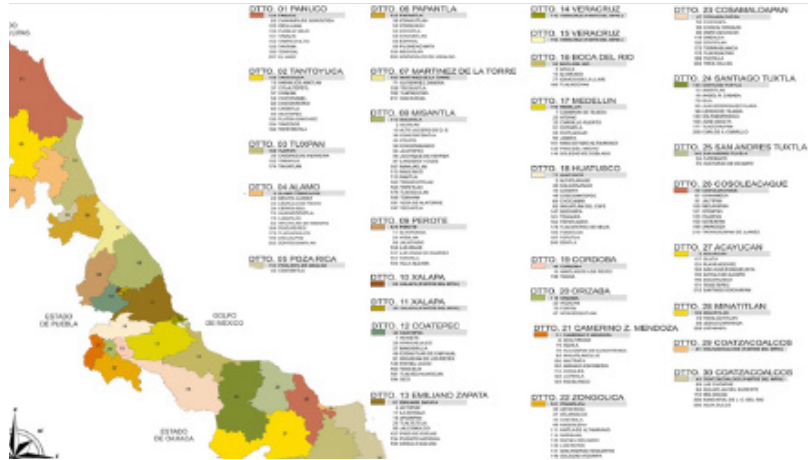
donde K es el número de estratos, N_k es el número de casillas en el estrato k , y \bar{X}_{kj} es el promedio de votos por casilla a favor del candidato j en el estrato k . Para determinar la precisión del estimador \hat{p}_j se estimó su distribución de muestreo mediante un proceso de extracción repetida de muestras estratificadas de tamaño n del marco muestral de la elección de gobernador de 2010 en Veracruz. En la Figura II se muestran los valores del estimador de la proporción de la votación para el PRI en 2000 muestras extraídas.

Una vez fijada la probabilidad $1 - \alpha$, el error de estimación, denotado por B , de una muestra de tamaño n es:

$$P(|\hat{p}_j - p_j| \leq B) = 1 - \alpha.$$

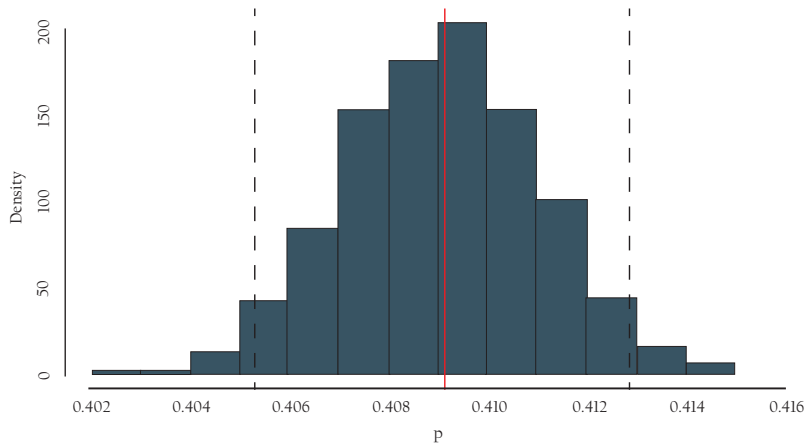
Si $1 - \alpha = 0.95$, entonces $B = \max(p_j - q_{0.025}, q_{0.075} - p_j)$ donde $q_{0.025}$ y $q_{0.075}$ son los cuantiles de orden 2.5 por ciento y 97.5 por ciento de la distribución de muestreo de \hat{p}_j . En la Figura II estos cuantiles están marcados por las líneas punteadas verticales.

FIGURA I. DISTRITOS ELECTORALES LOCALES DEL ESTADO DE VERACRUZ



Fuente: *Geografía electoral*. Disponible en <http://www.oplever.org.mx/geografiaelectoral.php> y *Districtos electorales locales del Estado de Veracruz de Ignacio de la Llave*. Disponible en <http://www.oplever.org.mx/archivos/cartografia/CartografiaVeracruz.pdf>. Ambos consultados el 7 de noviembre de 2018.

FIGURA II. DISTRIBUCIÓN DE MUESTREO DEL ESTIMADOR DE LA PROPORCIÓN EFECTIVA DE VOTOS A FAVOR DEL PRI, ELECCIÓN A GOBERNADOR 2010



Nota: valor real 0.4090 (línea vertical roja) y cuantiles de orden 2.5 por ciento y 97.5 por ciento (líneas verticales punteadas).

Fuente: elaboración propia.

Revista Mexicana de Estudios Electorales. Volumen 3, número 21, primer semestre de 2019 (enero-junio): 91-115. ISSN: 2448-8283.

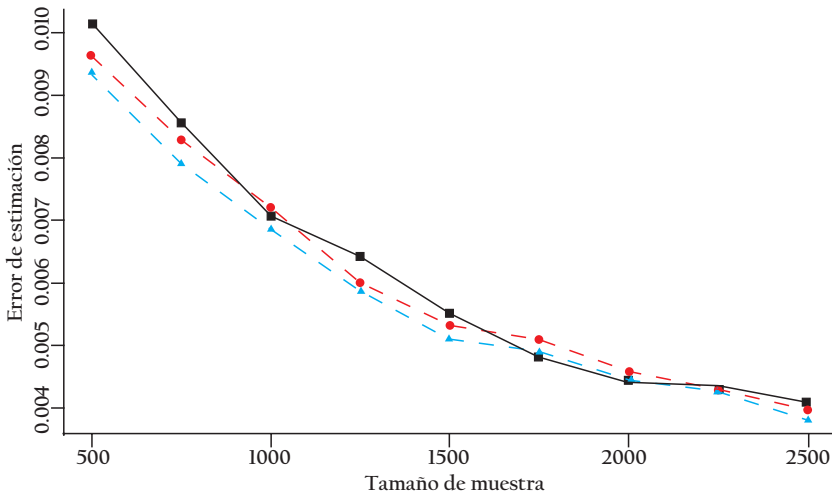
Para determinar cuál de los tres criterios de estratificación i), ii) o iii) proporciona el menor error de estimación, se hizo un estudio de simulación de 2,000 muestras, en el cual se varió el tamaño de muestra en un intervalo de 500 a 2,500, de 250 en 250. La asignación del tamaño del estrato se hizo de manera proporcional al número de casillas. El candidato que se tomó como referencia fue el postulado por el PRI. Los resultados se muestran en la Tabla I y se muestran gráficamente en la Figura III.

TABLA I. ERRORES DE ESTIMACIÓN (EN PORCENTAJES) ALCANZADOS POR CADA CRITERIO DE ESTRATIFICACIÓN PARA DISTINTOS TAMAÑOS DE MUESTRA

Tamaño de muestra	(i) Distritos	(ii) Distritos y tipo	(iii) Municipios
500	1.013	0.964	0.938
750	0.856	0.830	0.792
1000	0.709	0.720	0.687
1250	0.643	0.600	0.587
1500	0.552	0.534	0.511
1750	0.487	0.508	0.492
2000	0.446	0.459	0.446
2250	0.430	0.429	0.425
2500	0.409	0.398	0.382

Fuente: elaboración propia.

FIGURA III. GRÁFICA COMPARATIVA DE LOS ERRORES DE ESTIMACIÓN
ALCANZADOS POR LOS CRITERIOS DE ESTRATIFICACIÓN POR TAMAÑO DE MUESTRA



Nota: distritos locales (cuadrados), distritos locales y tipo (círculos), y municipios (triángulos).
Fuente: elaboración propia.

Los errores de estimación disminuyen al aumentar el tamaño de muestra en todos los casos. Para tamaños de muestra pequeños (500 y 750) la diferencia entre los tres criterios es mayor, mientras que para tamaños de muestra grandes (de 1750 en adelante) las diferencias son mínimas. De manera generalizada, el criterio de estratificación iii) basado en municipios es el que consistentemente produce los menores errores de estimación; seguido del criterio ii) basado en cruce de distrito local y tipo rural o urbano; y, finalmente el criterio i) basado en distritos locales.

Como criterio adicional para la determinación del tamaño de muestra se consideró otro estudio de simulación con la finalidad de verificar la factibilidad del levantamiento de la información dado el número de capacitadores-asistentes electorales (CAE) disponibles. Para este nuevo estudio se usó la información de las zonas de responsabilidad electoral por los CAE en la elección del 2012 en la que había 2,210 CAE asignados para el estado de Veracruz. Se tomaron distintos tamaños de muestra y se obtuvo la proporción de los CAE que tenían asignada al menos una casilla de la muestra, y para aquellos CAE en

muestra la distribución del número de casillas que les tocaría reportar a cada uno. Los resultados se presentan en la Tabla II.

En la Tabla II se observa que con una muestra de 1,250 CAE se involucrarían 44 por ciento de los mismos en el conteo rápido, el 3 por ciento de ellos tendría 3 o más casillas asignadas; mientras que con un tamaño de muestra de 2,000 el 61 por ciento participaría y el 8 por ciento de ellos tendrían 3 o más casillas. Entre más casillas tenga asignado un mismo CAE el levantamiento de la información no está garantizado en cuanto al tamaño de la muestra deseado y un tiempo razonable para dicha actividad.

TABLA II. DISTRIBUCIÓN DE LOS CAE POR NÚMERO DE CASILLAS ASIGNADAS

No. Casillas\n	1250	1500	1750	2000
1	77 %	72 %	68 %	62 %
2	20 %	23 %	24 %	30 %
3	3 %	4 %	7 %	7 %
4	0.3 %	0.5 %	1 %	1 %
5	0 %	0 %	0.1 %	0.1 %
% CAE en muestra	44 %	51 %	56 %	61 %

Fuente: elaboración propia.

ESTUDIO DE SIMULACIÓN 2

En el segundo estudio de simulación sólo se consideraron las estratificaciones i) y ii) del estudio 1. Los parámetros a estimar fueron las proporciones de los cuatro partidos de mayor votación en la elección de gobernador en Veracruz de 2010: PRI, con 40.899 por ciento; PAN, con 38.303 por ciento; PRD, con 6.709 por ciento y Convergencia, con 3.944 por ciento. Para estimar las precisiones de los estimadores primero se determinó el tamaño de muestra óptimo, considerándose la asignación proporcional el tamaño del estrato y la asignación con el procedimiento de Bethel (1989),⁶ estableciendo el estimador de la media poblacional de la variable j , $j = 1, 2, \dots, J$ como:

⁶ Este procedimiento se aplica en muestreo estratificado con variables de interés.

$$\bar{X}_j = \sum_{h=1}^L W_h \left(\frac{1}{n_h} \sum_{i=1}^{n_h} X_{hij} \right),$$

donde $W_h = N_h/N$. La varianza de este estimador es:

$$\text{Var}(\bar{X}_j) = \sum_{h=1}^L W_h^2 \left(\frac{N_h - n_h}{n_h N_h} \right) S_{hj}^2,$$

La función de costo para una unidad muestral en el estrato h cuesta C_h , entonces:

$$C(n_1, \dots, n_n) = \sum_{i=1}^{n_h} C_n n_n.$$

El procedimiento de Bethel minimiza esta función con la restricción de que las J varianzas sean menores a las constantes dadas, es decir: $\text{Var}(\bar{X}_j) \leq kj$, ($j = 1, 2, \dots, J$). Este procedimiento de asignación óptima multivariada es un problema de programación convexa, pues la solución n_1^* , n_2^* , ... n_L^* , no tiene forma “cerrada” pero el algoritmo numérico para encontrarla está implementado en el paquete de R *Bethel*. En el caso concreto, se usaron los mismos estimadores de razón que en el estudio 1. Los tamaños de muestra se calcularon para 5 márgenes de error: 1 punto porcentual, 0.75 puntos porcentuales, 0.5 puntos porcentuales, 0.45 puntos porcentuales y 0.4 puntos porcentuales, mismos que se presentan en la Tabla iii: en esta se aprecia que existe ganancia al aumentar el número de estratos en el sentido de que los tamaños de muestra disminuyen en la estratificación con 60 estratos; sin embargo, la reducción en el tamaño de muestra no se aprecia sustancial, por lo que se decidió considerar sólo a la estratificación (i) de los 30 estratos dados por los distritos locales. Los errores de estimación son muy similares a los obtenidos en el estudio de simulación anterior.

TABLA III. TAMAÑO DE MUESTRA CON LA ASIGNACIÓN DE BETHEL

	Margen de error				
	1 %	0.75 %	0.5 %	0.45 %	0.4 %
30 estratos	387	658	1345	1602	1933
60 estratos	383	643	1298	1543	1857

Fuente: elaboración propia.

La Tabla IV muestra la asignación óptima del tamaño de muestra a cada estrato con el procedimiento de Bethel. Se agregaron dos tamaños más de muestra: 1500 y 2000.

TABLA IV. TAMAÑOS DE MUESTRA ÓPTIMOS Y ASIGNACIÓN ÓPTIMA DE BETHEL

Estrato	387	658	1345	1500	1602	1933	2000
1	16	27	27	61	66	79	82
2	14	24	24	54	57	69	71
3	17	30	30	68	72	87	90
4	10	17	17	39	42	50	52
5	10	17	17	40	42	51	53
6	7	12	12	26	28	34	35
7	16	27	27	61	65	79	82
8	11	19	19	44	47	57	59
9	14	23	23	54	57	69	71
10	13	23	23	52	56	67	69
11	8	13	13	30	31	38	39
12	14	24	24	55	59	71	73
13	12	21	21	47	50	61	63
14	12	21	21	47	51	61	63
15	17	28	28	64	69	83	86
16	12	20	20	47	49	60	62
17	14	24	24	54	58	70	72
18	14	24	24	56	59	72	74

Estrato	387	658	1345	1500	1602	1933	2000
19	17	29	29	67	71	86	89
20	8	13	13	30	32	38	39
21	10	16	16	37	40	48	50
22	12	20	20	45	48	58	60
23	26	45	45	103	110	133	138
24	12	20	20	46	49	59	61
25	12	20	20	44	48	57	59
26	16	28	28	63	67	81	84
27	15	26	26	59	63	76	79
28	9	15	15	34	37	44	46
29	10	16	16	37	40	48	50
30	9	16	16	36	39	47	49

Fuente: elaboración propia.

En la Tabla v se muestran los errores de estimación al 95 por ciento de nivel de confianza para cada tamaño de muestra por partido político (PRI, PAN, PRD y Convergencia) y coalición (PAN con PRD y PRI -PVEM).

TABLA V. PRECISIONES ESTIMADAS AL 95 POR CIENTO DE CONFIANZA

<i>n</i>	PRI	PAN	PRD	CONV	PAN-PRD	PRI-ALIANZA
387 Bethel	1.150	1.260	0.732	0.553	1.263	1.146
Proporcional	1.203	1.320	0.770	0.577	1.311	1.201
658 Bethel	0.899	0.994	0.551	0.426	0.947	0.896
Proporcional	0.948	1.049	0.582	0.451	1.021	0.946
1345 Bethel	0.605	0.648	0.362	0.281	0.633	0.606
Proporcional	0.598	0.696	0.388	0.294	0.675	0.597
1500 Bethel	0.531	0.587	0.347	0.253	0.580	0.529
Proporcional	0.579	0.658	0.364	0.281	0.645	0.575
1602 Bethel	0.522	0.580	0.344	0.255	0.568	0.520
Proporcional	0.559	0.603	0.351	0.263	0.608	0.559

<i>n</i>	PRI	PAN	PRD	CONV	PAN-PRD	PRI-ALIANZA
1933 Bethel	0.479	0.499	0.291	0.227	0.504	0.479
Proporcional	0.485	0.549	0.319	0.240	0.547	0.486
2000 Bethel	0.466	0.509	0.298	0.219	0.510	0.464
Proporcional	0.491	0.537	0.306	0.235	0.525	0.490

Fuente: elaboración propia.

Como se observa en la Tabla v, la asignación óptima de Bethel mejora sustancialmente el error de estimación con respecto a la asignación proporcional de la muestra dentro del estrato.

ELECCIÓN DEL DISEÑO MUESTRAL FINAL

Para la muestral final las dos propuestas de diseño muestral y tamaños de muestra se conciliaron. La decisión del tamaño de muestra y del criterio de estratificación se basó además en las siguientes consideraciones:

- El número de los CAE disponibles para el levantamiento de la muestra para la elección del 2016 (2,186 más 379 supervisores). Cifras similares a las que se tuvieron en las elecciones federales del 2012.
- Los máximos errores de estimación considerados en los conteos rápidos de elección presidencial anteriores (0.3 por ciento en el 2006 y 0.5 por ciento en el 2012).
- Ante la posible falta de información en algún estrato y el consiguiente proceso de estimación, se consideró tener menos estratos.
- Con la finalidad de garantizar estimaciones sustentadas en información suficiente que produjeran certeza, se inició con ellas alrededor de las 22:30 horas.
- Ante mayores tamaños de muestra se logra un menor error de estimación, pero se complica el levantamiento de la información ya que un porcentaje considerable de los CAE tendrían a su cargo 3 o más casillas.

De acuerdo con estas consideraciones y los resultados de ambos estudios de simulación se propuso como criterio de estratificación la (i) basada en distritos locales. Se determinó un tamaño de muestra de 1,250 casillas con asignación

proporcional al tamaño del estrato. Suponiendo que el flujo de información no es uniforme, se determinó que para las 22:30 horas, se tendría un 80 por ciento de la muestra (1,000 casillas), con lo que se esperó observar un máximo error de estimación de 0.7 por ciento con un 95 por ciento de confianza.

II. LOGÍSTICA

Un CR es un proceso cuya complejidad está proporcionalmente determinada por la geografía social, económica y política del lugar donde se llevarán a cabo las elecciones. Por ello requiere de una planeación que el día de la elección logre presentar oportunamente las estimaciones de los resultados finales de la misma.

La operación logística del CR estuvo a cargo de la Dirección Ejecutiva de Organización Electoral del OPLE Veracruz. Contó con un equipo exclusivo de 100 capturistas, 10 supervisores de captura, 10 personas en mesa de ayuda, 3 coordinadores de captura y 4 personas de sistemas; en campo, en el estado de Veracruz, se capacitaron 379 supervisores (SE) y 2,186 CAE. En las instalaciones del OPLE se habilitó el Centro Estatal de Información del Conteo Rápido para el Proceso Electoral 2015-2016 (CEICR) (ver Figura IV).

FIGURA IV. EL CEICR EL DÍA DE LA JORNADA ELECTORAL



Fuente: obtenida por los autores el día de la elección.

En el CEICR se montó el equipo de cómputo y se le instaló el Sistema de Información del Censo Rápido (SICRA). La logística y el funcionamiento del CR se resume en la Tabla VI y la Figura V, respectivamente.

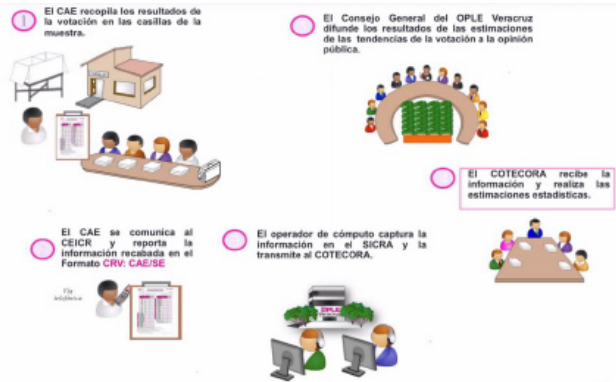
TABLA VI. LOGÍSTICA DEL CONTEO RÁPIDO DEL ESTADO DE VERACRUZ

-
1. Al finalizar la jornada electoral, los conteos de las actas de escrutinio de las casillas en la muestra son recopilados por los CAE o SE que las tienen a cargo.
 2. El CAE o SE transmite al CEICR los resultados del censo.
 3. En el CEICR los capturistas reciben y capturan los datos transmitidos por los CAE o SE, luego se transmiten, por medio del SICRA los datos al COTECORA.
 4. El COTECORA recibe los datos, realiza las estimaciones y las entrega al Consejo General del OPLE.
 5. El Consejo General del OPLE difunde los resultados a la opinión pública.
-

Fuente: elaboración propia.

El CR requiere una adecuada organización logística que permita enfrentar con éxito tanto contingencias anticipadas como las que no. Por ejemplo, que no haya disponibilidad de los recursos humanos, materiales y financieros requeridos para su funcionamiento, la falta de ubicación de casillas pertenecientes a la muestra no reportadas al CEICR, cuestiones de seguridad informática del SICRA —ataques de hackeo, problemas en las telecomunicaciones (falta de conectividad en la telefonía celular o el internet)—, seguridad y continuidad de operación del CEICR (falta de suministro de energía eléctrica, acceso físico al CEICR de personas no autorizadas).

FIGURA V. ESQUEMA GENERAL DEL FUNCIONAMIENTO DEL CR



Fuente: *Censo Rápido Veracruz 16: Guía de Capacitación y Ejecución de la Operación Logística en el ámbito Distrital*. Instituto Nacional Electoral.

Con la finalidad de detectar errores oportunamente y tomar las medidas preventivas pertinentes, se realizaron dos simulacros para evaluar el funcionamiento de los medios y sistemas para la captura: transmisión y recepción de la información electoral, el proceso operativo en campo, el flujo de llegada de la información de las casillas, los medios y sistemas para conocer la cobertura geográfica de la muestra recibida, los métodos de estimación, así como la generación y envío del reporte con las estimaciones de la simulación a los integrantes del OPLE. Los simulacros fueron en mayo, aproximadamente un mes antes de la jornada electoral. Participaron los CAE, los SE, los operadores de cómputo del CEICR y los integrantes del COTECORA. En cada uno se generaron muestras con datos ficticios que se proporcionaron a los vocales de Organización del Consejo Distrital del OPLE quienes a la vez entregaron a los CAE y se para que los reportaran.

Dos días antes de la jornada electoral, personal del OPLE de Veracruz llevó a cabo, ante notario público, la validación del equipo de cómputo en el que se instaló el programa R (R Core Team, 2016) para la obtención de la muestra. Los integrantes del COTECORA entregaron el programa que seleccionaba la muestra junto con sus códigos de integridad. Para la ejecución del programa se requirió de la participación de seis personas seleccionadas entre los asistentes: de una urna cada persona seleccionó un dígito entre 0 y 9 (mismos que fueron anotados en un formato diseñado para tal efecto) formando una cifra

de 6 dígitos —conocido como *semilla*— que fue ingresada al programa sólo a la vista del notario público. Los demás asistentes no conocieron ese número. Un miembro del cotecora ejecutó el programa de selección de muestra, la cual quedó grabada en el disco duro de la computadora. Se generó un código de integridad de la muestra y se entregó a dicho federatario público, así como a los asistentes que lo solicitaron. La muestra se grabó en dos discos compactos no regrabables. Con el fin de que se prepararan los trabajos de acopio de datos de las actas de escrutinio y cómputo correspondientes a la muestra, el primer disco se entregó al Director Ejecutivo de Organización Electoral del OPLE Veracruz, mientras que el segundo, así como los códigos se guardaron en un sobre, siendo entregados ambos al funcionario público mencionado. La computadora se embolsó de forma que no fuera manipulada con posterioridad (y puesta a resguardo del fedatario), junto con la muestra generada (sellada y rubricada previamente, por este último).

III. ESTIMADORES

El COTECORA propuso dos estimadores de razón, uno basado en la teoría del muestreo de poblaciones finitas y otro basado en un modelo bayesiano. Los cálculos de estimación se hicieron en ambos casos en el ambiente de cómputo estadístico R (R Core Team, 2016). Los intervalos de estimación finales y que se reportaron se obtuvieron mediante la unión de los intervalos producidos con los dos métodos que se presentan a continuación:

ESTIMADOR DE RAZÓN COMBINADO

En una población de casillas representada por N , donde en cada una se registra la votación para J candidatos, la población está dividida en L estratos con N_h casillas dentro del estrato h : sea X_{hij} el número de personas que votan por el candidato j en la casilla i del estrato h , la votación total para el candidato j dentro del estrato h es:

$$T_{hj} = \sum_{i=1}^{N_h} X_{hij}$$

de modo que el total de votos en la población para el candidato j es:

$$T_j = \sum_{h=1}^L T_{hj}.$$

Los parámetros de interés son las proporciones poblacionales de votos “efectivos” para cada uno de los J candidatos:

$$\lambda_j = \frac{T_j}{\sum_{j=1}^J T_j}, j = 1, \dots, J.$$

Considerando una muestra aleatoria simple de n_h casillas del estrato h , donde la casilla i del estrato h , sea X_{hij} el número de votos para el candidato j . El estimador del total de votos para el candidato j en la población es:

$$\hat{T}_j = \sum_{h=1}^L N_h \bar{x}_{hj},$$

donde $\bar{X}_{hj} = \sum_{i=1}^{n_h} X_{hij} / n_h$. El estimador de razón combinado de λ_j está dado por:

$$\hat{\lambda}_j = \frac{\hat{T}_j}{\sum_{j=1}^J \hat{T}_j}.$$

Un intervalo de confianza para λ_j se construye del siguiente modo: sea S_{hj}^2 la varianza poblacional de la votación para el candidato j dentro del estrato h :

$$S_{hj}^2 = \sum_{i=1}^{N_h} \frac{(X_{hij} - \bar{X}_{hj})^2}{N_h - 1}$$

donde $\bar{X}_{hj} = \sum_{i=1}^{N_h} X_{hij} / N_h$. Sea $U_{hi} = \sum_{j=1}^J X_{hij}$ el total de la votación de los J candidatos en la casilla i del estrato h :

$$\bar{U}_h = \sum_{i=1}^{N_h} U_{hi}/N_h$$

el promedio de esta votación:

$$\bar{U} = \sum_{h=1}^L \sum_{i=1}^{N_h} U_{hi}/N$$

y, el promedio de la votación para los candidatos en toda la población. Sea $S_{U_h}^2$ la varianza de la votación:

$$S_{U_h}^2 = \sum_{i=1}^{N_h} \frac{(U_{hj} - \bar{U}_h)^2}{N_h - 1}.$$

El coeficiente de correlación entre X_{hij} y U_{hi} es:

$$\rho_{nj} = \frac{1}{N_h} \sum_{i=1}^{N_h} \left(\frac{X_{hij} - \bar{X}_{hj}}{S_{hj}} \right)^2 \left(\frac{U_{hi} - \bar{U}_h}{S_{U_h}} \right)^2.$$

Por lo tanto, la varianza del estimador $\hat{\lambda}_j$ es aproximadamente:

$$\text{Var}(\hat{\lambda}_j) \approx \frac{1}{N^2 \bar{U}^2} \sum_{h=1}^L \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \sigma_{hj}^2$$

donde $\delta_{hj}^2 = S_{hj}^2 + \lambda_j^2 S_{U_h}^2 - 2\lambda_j p_{hj} S_{hj} S_{U_h}$. De modo que un estimador de $\text{Var}(\hat{\lambda}_j)$, $\widehat{\text{Var}}(\hat{\lambda}_j)$, se obtiene estimando las cantidades poblacionales con sus contrapartes muestrales. De esta manera se obtiene un intervalo de confianza asintótico para la proporción de votación para el candidato j , $\hat{\lambda}_j \pm 2\{\widehat{\text{Var}}(\hat{\lambda}_j)\}^{1/2}$. Para detalles se puede consultar Levy y Lemeshow (1991).

ESTIMADOR BAYESIANO

Sean m_{hi} el tamaño de la lista nominal de la casilla del estrato h , $m_h = \sum_{i=1}^{N_h} m_{hi}$ el tamaño de la lista nominal del estrato h , y $m = \sum_{i=1}^L m_{hi}$ el tamaño de la lista

nominal de la población, puede definirse como θ_{jh} la proporción de personas en la lista nominal del estrato h que votan por el candidato j :

$$\theta_{hj} = \frac{1}{m_h} \sum_{i=1}^{N_h} X_{hij} = \frac{T_{hj}}{m_h}.$$

Mientras que la proporción de personas en la lista nominal de toda la población que vota por el candidato j es:

$$\theta_j = \frac{T_j}{\sum_{h=1}^L m_h} = \sum_{h=1}^L \frac{m_h}{m} \theta_{hj}.$$

La propuesta de Mendoza y Nieto-Barajas (2016) supone que el número total de votos para el candidato j en la casilla i del estrato h , X_{hij} se puede modelar mediante una distribución normal de la forma:

$$x_{hij} | \theta_{hj}, \tau_{hj} \sim N \left(m_{hi} \theta_{hj}, \frac{\tau_{hj}}{m_{hi}} \right), i = 1, \dots, n_h$$

donde τ_{hj} es un parámetro de precisión. El modelo bayesiano se complementa con la elección de una distribución inicial no informativa de la forma:

$$f(\theta_{hj}, \tau_{hj}) \propto \tau_{hj}^{-1} I(\tau_{hj} > 0) I(0 < \theta_{hj} < 1).$$

Dado que los parámetros de interés son las proporciones “efectivas” de votos, es necesario renormalizar los θ_j de tal manera que:

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

Debido a la forma de la distribución final se requiere de métodos de simulación para obtener los estimadores de las λ_j , para $j = 1, \dots, J$. Los detalles para

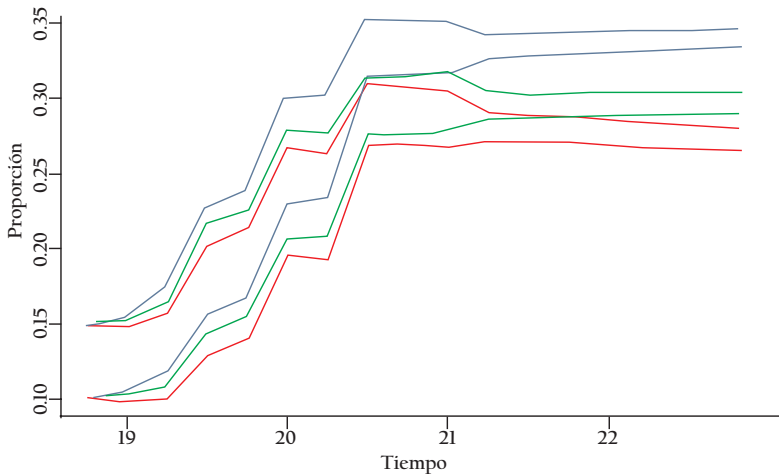
obtener la distribución final y los métodos de simulación se pueden consultar en Mendoza y Nieto-Barajas (2016).

IV. RESULTADOS

La jornada electoral se llevó a cabo el domingo 5 de junio de 2016, inició a las 8:00 horas, fueron instaladas 10,421 casillas, y finalizó alrededor de las 18:00 horas. Los primeros reportes de los conteos de las casillas de la muestra empezaron a llegar al CEICR aproximadamente a las 18:45 horas, comenzándose a generar la primera remesa con la información de 4 casillas.

En la Figura IV se muestran los estimadores parciales para los tres candidatos punteros, al mismo tiempo que se puede observar que a las 20:00 horas, con la información disponible, los tres candidatos se encontraban empatados. Sin embargo, poco tiempo después, alrededor de las 20:30 horas el candidato de la coalición PAN-PRD mostró un comportamiento distinto al de los otros dos que, claramente, le otorgaba el triunfo.

FIGURA IV. ESTIMADORES BAYESIANOS PARCIALES AL 95 POR CIENTO PARA LOS 3 CANDIDATOS PUNTEROS CONFORME LLEGABA LA INFORMACIÓN



Fuente: elaboración propia.

La Tabla VII presenta el tamaño de la muestra con la que se realizaron los cálculos de estimación definitivos para la elección. A las 22:45 horas se contaba con 917 casillas reportadas, lo que representa un avance del 73.34 por ciento de la muestra planeada de 1,000 casillas.

TABLA VII. COBERTURA DEL CR POR DISTRITO LOCAL.
CORTE A LAS 22:50HRS, 5 JUNIO 2016

Distrito Local	Casillas en muestra	Casillas reportadas	% avance
1	48	47	97.92
2	44	35	79.55
3	45	36	80.00
4	48	25	52.08
5	39	37	94.87
6	42	33	78.57
7	41	33	80.49
8	46	35	76.09
9	38	34	89.47
10	36	25	69.44
11	38	11	28.95
12	37	27	72.97
13	42	29	69.05
14	45	26	57.78
15	48	22	45.83
16	45	33	73.33
17	41	31	75.61
18	36	27	75.00
19	42	38	90.48
20	43	37	86.05
21	35	27	77.14
22	35	21	60.00
23	46	34	73.91

Distrito Local	Casillas en muestra	Casillas reportadas	% avance
24	43	37	86.05
25	38	27	71.05
26	44	40	90.91
27	39	31	79.49
28	41	19	46.34
29	42	30	71.43
30	43	30	69.77
Total	1250	917	73.36

Fuente: Castro *et al.*, 2016.

Las estimaciones producidas por ambos métodos se combinaron en un solo intervalo, es decir, el límite inferior quedó determinado por el menor límite inferior de los dos intervalos y el límite superior por el mayor límite superior de los dos intervalos. Como consecuencia, los intervalos reportados tienen un nivel de confianza de al menos el 95 por ciento. A las 23:10 horas el COTECORA entregó al Consejo General del OPLE los resultados que se muestran en la Tabla VIII.

TABLA VIII. RESULTADOS DEL CONTEO RÁPIDO Y CÓMPUTO FINAL

Candidato	Partido	Límite inferior	Límite superior	Real
M.A. Yunes Linares	PAN, PRD	33.32	34.79	34.40
H. Yunes Landa	PRI, PVEM, PANAL, AVE, PFCRN	29.01	30.40	30.29
A.L. Méndez Herrera	PT	1.48	1.71	1.62
A. Méndez de la Luz	MC	1.09	1.28	1.21
C. García Jiménez	MORENA	26.48	28.20	26.40
V.A. Vázquez Cuevas	PES	1.03	1.19	1.11
J. Bueno Torio	Independiente	1.99	2.21	2.00

Fuente: Castro *et al.*, 2016.

CONCLUSIONES

El CR del proceso electoral de Veracruz 2015-2016 proporcionó estimaciones exactas y precisas de la votación. Todos los resultados oficiales, excepto el de un partido (MORENA), fueron capturados por los estimadores por intervalo. Es oportuno señalar que los intervalos proporcionados por ambos métodos de estimación (clásico y bayesiano) coincidieron significativamente.

El diseño muestral que se utilizó es elemental: este fue motivo de discusión entre los integrantes del COTECORA, en la búsqueda de satisfacer los requerimientos del 95 por ciento de confianza, margen de error de 0.07 puntos porcentuales y cobertura geográfica, con un tamaño de muestra menor al que se propuso con el diseño estratificado. Cuestión no resuelta pero que es de interés teórico y práctico investigar con el fin de poder lograr diseños de muestreo complejos con tamaños de muestra más pequeños que el que se usó en el CR de Veracruz. El marco muestral para hacer estos estudios se encuentra disponible al público en <https://oplever.org.mx/miniportales/resultadose/>.

Es importante mencionar que los resultados oficiales de la votación pueden ser el resultado de recuentos de algunas casillas, lo que implica que el dato usado en la estimación difiera del dato utilizado para el cómputo final, provocando que existan errores de estimación no contemplados originalmente por los modelos. Se vuelve necesario proponer modelos más robustos y/o aumentar la confianza/confiabilidad al 99 por ciento en las estimaciones.

Para las elecciones de 2018 se realizó el CR para la elección de presidente, junto con 8 CR para las elecciones de gobernador en los estados de Chiapas, Guanajuato, Jalisco, Morelos, Puebla, Tabasco, Veracruz, y Yucatán; más el CR para la elección de jefe de gobierno de Ciudad de México. Dichas elecciones concurrentes plantearon retos metodológicos en cuanto al diseño muestral y al levantamiento de la información debido a que el personal de campo no es ilimitado, siendo necesario plantear la posibilidad de diseños muestrales de los diversos conteos donde un CAE (en un estado con elección para gobernador) genere la información tanto de la elección gubernamental como presidencial. Sin embargo, el reto recae en los criterios de optimalidad para la definición de los estratos pues estos podrían no coincidir entre sí, además de que los tamaños de muestra necesarios para alcanzar cierto error de estimación en una elección de gobernador son mucho mayores que los que se requieren para la

presidencial dentro de la misma entidad. Cuestión que merece ser analizada una vez finalizada la jornada electoral.

En el caso de Veracruz, el objetivo de proveer con el cr estimaciones oportunas y confiables de la votación se logró. Con este ejercicio se agrega una experiencia de éxito más a lo que ya se ha convertido en una tradición: la participación, desde la academia, de estadísticos en las buenas prácticas electorales, coadyuvando así a tener un México cada vez más democrático.

REFERENCIAS

- BETHEL, James W. 1989. "Sample Allocation in Multivariate Surveys". En *Survey Methodology*, 15: 47-57.
- CASTRO López, Claudio R.; APARICIO Castillo, Francisco J.; MONTANO Rivas, Julia A.; NIETO Barajas, Luis E. y JUÁREZ Cerrillo, Sergio F. 2016. *Informe de Resultados del conteo rápido en la elección de gobernador Veracruz 2016*. Disponible en <http://www.oplever.org.mx/miniportales/resultadose/conteorapido/03.pdf>. Consultado el 7 de noviembre de 2018.
- ESTOK, Melissa, NEVITTE Neil y COWAN, Glenn. 2002. *The quick count and election observation: an NDI handbook for civic organizations and political parties*. Washington DC: National Democratic Institute for International Affairs.
- LEVY, Paul S., y LEMESHOW, Stanley. 1991. *Sampling of Populations. Methods and Applications*. New York: Wiley.
- MENDOZA, Manuel y NIETO-BARAJAS, Luis E. 2016. "Quick Counts in the Mexican Presidential Elections: A Bayesian Approach". En *Electoral Studies*, 43: 124-132.
- R CORE TEAM. 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Disponible en <https://www.R-project.org/>. Consultado el 7 de noviembre de 2018.