

Informe de resultados

Comité Técnico del Conteo Rápido
Cotecora

Dr. Claudio Rafael Castro López
Dr. Francisco Javier Aparicio Castillo
Dra. Julia Aurora Montano Rivas
Dr. Luis Enrique Nieto Barajas
Dr. Sergio Francisco Juárez Cerrillo

I. Protocolo de Selección de la Muestra

El proceso de selección y resguardo de la muestra, estuvo formado por las siguientes etapas:

Selección de la muestra

- 1.- Reunidos en la sala de sesiones del OPLE la y los integrantes del Comité Técnico del Coteo Rápido (Cotecora) entregamos el software necesario para la selección de la muestra.
- 2.- A la vista de los asistentes se instaló el software “*Microsoft File Checksum Integrity Verifier*” disponible en:
<http://www.microsoft.com/en-us/download/details.aspx?id=11533>.
- 3.- Un integrante del Cotecora obtuvo los códigos de integridad de los archivos entregados y los comparó con los generados.
- 4.- A la vista de los asistentes se instaló el software “*R: A language and environment for statistical computing*” disponible en:
<https://cran.r-project.org/bin/windows/base/>.
- 5.- Para construir y capturar los números semilla el notario público seleccionó seis números entre 0 y 9 y los anotó en un formato diseñado para tal efecto.
- 6.- Un miembro del Cotecora inició la etapa de ejecución de la selección de la muestra.
- 7.- A la vista del notario público y de los asistentes, los números que seleccionó se ingresaron en el programa de selección de la muestra.
- 8.- Una vez ingresados los números para construir la semilla, se continuó con el programa para seleccionar la muestra. La muestra quedó grabada en el disco duro que se entregó al Director Ejecutivo de Organización Electoral.
- 9.- Se generó un código de integridad de la muestra. El código de integridad se imprimió y se entregó al notario público.

Resguardo de la muestra

- 10.- La muestra se grabó en un disco compacto no regrabable.

11.- La muestra segmentada en los 21 distritos federales se envió por correo electrónico a las y los vocales ejecutivos distritales del INE a través del correo institucional determinado para tal efecto. La cuenta de correo estaba en ceros y sirvió exclusivamente para el envío de esta muestra. Lo anterior con el fin de que se prepararán los trabajos de acopio de datos de las actas de escrutinio y cómputo correspondientes a la muestra.

12.- El disco con la muestra, los códigos de integridad y los formatos en los que se anotaron los números para construir la semilla se guardaron en un sobre. El equipo de cómputo en el cual fue generada la muestra se resguardó. Ambos sellados y rubricados por el notario público.

13.- El notario público entregó al Director Ejecutivo de Organización Electoral los sobres a los que se refiere el numeral anterior para su resguardo.

II. Fórmulas de cálculo utilizadas para cada método de Estimación

Se usaron dos métodos de estimación: el Bayesiano y el clásico. Los intervalos que se reportaron en el informe final se construyeron mediante la unión de los intervalos de estos dos métodos. Los métodos de estimación se especifican a continuación.

MÉTODO BAYESIANO

(Propuesta del Dr. Luis Enrique Nieto Barajas)

Inferencia bayesiana

La estadística bayesiana es una forma alternativa a la estadística frecuentista de hacer inferencia sobre los parámetros desconocidos de un modelo. Tiene sus bases en la teoría de decisión, lo cual fundamenta de manera axiomática los procesos inferenciales al ser planteados como problemas de decisión (De Groot, 2004).

Como parte del proceso inferencial, es necesario cuantificar la incertidumbre sobre los parámetros desconocidos del modelo mediante distribuciones de probabilidad. Esta cuantificación puede reflejar las creencias del estadístico, si es que las hay, o ser el reflejo del desconocimiento por completo de los valores posibles del parámetro. Como consecuencia de esta cuantificación, tanto las variables aleatorias observables como los parámetros fijos son descritos mediante distribuciones de probabilidad, lo que simplifica el proceso inferencial.

La cuantificación inicial (antes de observar la muestra) que se hace sobre los parámetros $f(\theta)$, debe de ser actualizada mediante la información muestral $\underline{X} = (X_1, X_2, \dots, X_n)$ proveniente del modelo $f(x|\theta)$. Esto se hace mediante el teorema de Bayes obteniéndose así la cuantificación final o posterior $f(\theta|\underline{x})$ que combina y resume la información inicial y la información muestral, es decir, $f(\theta|\underline{x}) = f(\underline{x}|\theta)f(\theta)/f(\underline{x})$.

El teorema de Bayes tienen una expresión matemática simple, sin embargo, la obtención analítica de la distribución final de los parámetros se puede complicar debido al cálculo de la constante de normalización $f(\underline{x})$. Gracias a los avances computacionales recientes y en especial a los algoritmos de simulación Monte Carlo vía cadenas de Markov (MCMC), es posible obtener características de cualquier distribución final mediante métodos de simulación sin necesidad de calcular la constante de normalización (Chen et al. 2000).

Notación y definiciones

El objetivo del conteo rápido es producir estimaciones confiables de la proporción efectiva de votos a favor de cada uno de los candidatos. Esta proporción se calcula como el cociente entre el número de votos a favor de un candidato y el número total de votos emitidos por todos los candidatos sin considerar los votos nulos.

Suponiendo que el diseño muestral es un diseño estratificado de casillas, sea N el número de estratos, cada uno con una lista nominal de tamaño n_i para $i = 1, \dots, N$. Definimos θ_{ij} como la proporción de personas en la lista nominal del estrato i cuya preferencia es por el candidato j , donde $j = 1, \dots, J$ con J el número total de candidatos incluyendo los candidatos no registrados y los nulos. En este caso la proporción de personas en la lista nominal de todo el estado cuya preferencia es por el candidato j es

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

donde n es el tamaño de la lista nominal del estado. Finalmente la proporción efectiva de votos a favor del candidato j es

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

Modelo y proceso de inferencia

Del total de casillas en el estado se seleccionan c de ellas, repartidas en c_i para cada uno de los estratos $i = 1, \dots, N$. Sea X_{ij}^k el número de personas a favor del candidato j en el estrato i de la casilla k para $k = 1, \dots, K_i$ con K_i el número total de casillas en el estrato i .

Nuestro modelo asume que

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

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

El modelo bayesiano se complementa con la elección de distribuciones iniciales no informativas para θ_{ij} y τ_{ij} . Debido a la forma de la distribución final se requiere de métodos de simulación para obtener los estimadores de las λ_j , para $j = 1, \dots, J - 1$. Los detalles para obtener la distribución final y los métodos de simulación se pueden consultar en Mendoza y Nieto-Barajas (2012).

Referencias

- Chen, M.-H., Shao, Q.-M. e Ibrahim, J.G. (2000). *Monte Carlo methods in Bayesian computation*. Springer, New York.
- DeGroot, M.H. (2004). *Optimal statistical decisions*. Wiley, New Jersey.
- Mendoza, M. y Nieto-Barajas, L.E. (2012). *Quick counts in the Mexican presidential elections: A Bayesian approach*. Reporte Técnico DE-C12.11, Departamento de Estadística, ITAM.

MÉTODO CLÁSICO: ESTIMADOR DE RAZÓN COMBINADO

(Propuesta del Dr. Sergio Francisco Juárez Cerrillo)

Tenemos una población finita de N casillas y en cada casilla se va a registrar la votación para J partidos. La población de casillas está dividida en L estratos con N_h casillas dentro del estrato h , ($h = 1, \dots, L$). Sea X_{hij} la votación para el partido j en la casilla i del estrato h ; y sean A_{hi} los votos anulados en la casilla i del estrato h . La votación total para el partido j en el estrato h es $t_{hj} = \sum_{i=1}^{N_h} X_{hij}$. De modo que el total de votos en la población para el partido j es

$$t_j = \sum_{h=1}^L t_{hj}.$$

El total de votos anulados en la población es

$$A = \sum_{h=1}^L \sum_{i=1}^{N_h} A_{hi}$$

Los parámetros de interés son las proporciones poblacionales de votos para cada uno de los J partidos

$$R_j = \frac{t_j}{\sum_{j=1}^J t_j + A}, \quad j = 1, 2, \dots, J.$$

Para estimar a las proporciones poblacionales R_j se proponen los siguientes dos estimadores de razón. Las derivaciones teóricas y detalles de estos estimadores se pueden ver en Levy y Lemeshow (1991) o Särndal et al. (1992).

Consideremos una muestra aleatoria simple de n_h casillas del estrato h . En la casilla i del estrato h sean x_{hij} el número de votos para el partido j y sea a_{hi} el número de votos anulados. El estimador del total votos para el partido j en la población es

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

donde

$$\bar{x}_{hj} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hij}.$$

El estimador del total de votos anulados es

$$\hat{A} = \sum_{h=1}^L N_h \left(\frac{1}{n_h} \sum_{i=1}^{n_h} a_{hi} \right).$$

El estimador de Razón Combinado de R_j es

$$\hat{R}_{jc} = \frac{\hat{t}_j}{\sum_{j=1}^J \hat{t}_j + \hat{A}}.$$

Sea S_{hj}^2 la varianza dentro del estrato j de la votación para el partido j

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

donde

$$\bar{X}_{hj} = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hij}.$$

Sea S_{Uhj}^2 la varianza de la votación efectiva más los votos anulados, es decir

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

donde $U_{hi} = \sum_{j=1}^J X_{hij} + A_{hi}$, y además

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

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

Sea ρ_{hj} el coeficiente de correlación entre X_{hij} y U_{hi}

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

La varianza del estimador \hat{R}_{jc} se aproxima por

$$\text{Var}(\hat{R}_{jc}) = \left(\frac{1}{N^2 \bar{U}^2} \right) \sum_{h=1}^L \frac{N_h^2 (N_h - n_h)}{n_h (N_h - 1)} \sigma_{hj}^2$$

donde

$$\sigma_{hj}^2 = S_{hj}^2 + R_j^2 S_{Uh}^2 - 2R_j \rho_{hj} S_{hj} S_{Uh}.$$

Una estimación de $\text{Var}(\hat{R}_{jc})$, $\widehat{\text{Var}}(\hat{R}_{jc})$, se obtiene estimando las cantidades poblacionales en $\text{Var}(\hat{R}_{jc})$ con sus contrapartes muestrales. De esta manera se obtiene el intervalo de confianza al 95% de confianza para la proporción de votación para el partido j

$$\hat{R}_{jc} \pm 2 \sqrt{\widehat{\text{Var}}(\hat{R}_{jc})}.$$

- Levy, P.S., and Lemeshow, S. (1991). *Sampling of Populations. Methods and Applications*. Second Ed. New York: Wiley.
- Särndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

III. Informe de Resultados

1. INFORMACIÓN GENERAL DE LA MUESTRA RECUPERADA

- De las 1250 casillas seleccionadas en muestra, se recibió información de 917. Esta cantidad representa el 73.36 % de la muestra total.
- De los 30 estratos (distritos electorales locales) considerados para definir el diseño muestral, se recibió información de 30 estratos.

2. ESTIMACIÓN DE LAS TENDENCIAS DE LA VOTACIÓN

- Con la información recibida hasta el momento, y con un nivel de confianza del 95%, se estima una participación ciudadana en la elección de entre 52.89 % y 55.20%.
- Las estimaciones con un 95% de confianza del porcentaje en favor de cada candidato a gobernador se presentan a continuación.

Partido	Candidato	Límite inferior	Límite superior
	Miguel Ángel Yunes Linares	33.32	34.79
	Héctor Yunes Landa	29.01	30.40
	Alba Leonila Méndez Herrera	1.48	1.71
	Armando Méndez de la Luz	1.09	1.28
morena	Cuitláhuac García Jiménez	26.48	28.20
	Víctor Alejandro Vázquez Cuevas	1.03	1.19
	Juan Bueno Torio	1.99	2.21

IV. Base de Datos Utilizada en las estimaciones del Conteo Rápido

Entre las 18:00 y las 22:50 horas del domingo 5 de junio, el conteo rápido recabó información de 917 casillas de una muestra de 1250 casillas, es decir 73.3% del total. La distribución de casillas reportadas en la remesa final por distrito local fue así:

Cobertura del Conteo Rápido 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
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
30	1250	917	73.36

La distribución de casillas reportadas en la remesa final por distrito federal fue así:

Cobertura del Conteo Rápido por distrito federal

Corte a las 22:50hrs, 5 junio 2016

Distrito Federal	Casillas en muestra	Casillas reportadas	% avance
1	52	51	98.08
2	66	36	54.55
3	48	39	81.25
4	57	38	66.67
5	58	54	93.10
6	50	39	78.00
7	62	53	85.48
8	75	47	62.67
9	59	47	79.66
10	60	30	50.00
11	68	50	73.53
12	64	33	51.56
13	56	49	87.50
14	58	29	50.00
15	69	54	78.26
16	58	51	87.93
17	61	42	68.85
18	61	37	60.66
19	54	41	75.93
20	54	47	87.04
21	60	50	83.33
TOTALES	1250	917	73.36

Se anexa a este reporte, en archivo `Remesa de informe.csv` la base de datos utilizada para realizar las estimaciones del cálculo final reportadas en el conteo rápido. Se anexa también la huella digital de esta base de datos.

Finalmente, el archivo `casillas en muestra.xlsx` es la a base de datos con casillas en muestra, indicando cuales fueron reportadas y cuales no en la remesa final usada para el reporte final.