

CRITERIOS CIENTÍFICOS, LOGÍSTICOS Y OPERATIVOS PARA LA REALIZACIÓN DEL CONTEO RÁPIDO Y PROTOCOLO PARA LA SELECCIÓN DE LA MUESTRA.

I. ANTECEDENTES.

El 4 de diciembre de 2015, el Consejo General del Organismo Público Local Electoral del estado de Veracruz aprobó, mediante acuerdo OPLE-VER/CG-45/2015, la creación e integración del Comité Técnico Asesor para el Conteo Rápido (COTECORA) a efectuarse el día de la jornada electoral del proceso electoral ordinario 2015-2016 para el estado de Veracruz.

Dicho Comité está conformado por cinco integrantes:

- 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

De acuerdo a lo dispuesto por el Considerando 11, las funciones de Secretario Técnico del Comité Técnico Asesor del Conteo Rápido las asume el Lic. Gregorio Arellano Rocha, Director Ejecutivo de Organización Electoral del Organismo Público Local Electoral del estado de Veracruz.

En cumplimiento a los Lineamientos para la instrumentación y operación del conteo rápido para los procesos electorales del estado de Veracruz, el COTECORA desarrolló los criterios científicos, además conoció el diseño de los procedimientos logísticos y operativos para el acopio de las Actas de Escrutinio y Cómputo de Casilla, así como el protocolo para la selección de la muestra. Los Criterios científicos, logísticos y operativos para la implementación del Conteo Rápido a desarrollarse en el Proceso Electoral 2015-2016 se describen en el presente documento.

II. CRITERIOS CIENTÍFICOS PARA LA SELECCIÓN DE LA MUESTRA

De acuerdo a los “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 y 2016-2017”, la muestra deberá cumplir con las siguientes características:

- I. Que todas y cada una de las casillas instaladas en la entidad 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.

La propuesta de diseño muestral, que cumple con los lineamientos antes mencionados, consistirá en un muestreo estratificado de casillas con asignación proporcional al número de casillas por estrato. Se usará un nivel de confianza del 95% para determinar el tamaño de muestra y el error de estimación alcanzado por dicho tamaño de muestra.

Los estratos estarán definidos por los distritos locales, lo que nos lleva a tener 30 estratos. El tamaño de muestra propuesto es de 1,250 casillas distribuidas de manera óptima por distrito local. Este tamaño de muestra contempla la demora en la llegada de la información y se espera que a las 22:30 hrs. se tenga un 80% de la muestra (i.e., 1,000 casillas). Con 1,000 casillas se alcanzaría un error de estimación de 0.7% con un 95% de confianza.

Los detalles de la selección del esquema de muestreo se encuentran en el apéndice.

III. SELECCIÓN Y RESGUARDO DE LA MUESTRA

3.1 Consideraciones generales para la selección de la muestra.

La muestra con la que se estimará la votación a favor de los candidatos a la Gubernatura del estado de Veracruz se realizará en un acto público con los integrantes del Comité Técnico Asesor del Conteo Rápido, los Consejeros Electorales y representantes de los partidos políticos y candidatos independientes entre el día jueves 2 y el domingo 5 de junio de 2016.

El acto se realizará en presencia de un notario público que dará fe del desarrollo del protocolo, desde la instalación del software hasta la obtención y proceso de resguardo de la muestra definitiva.

Para la selección de la muestra se hará uso del software propuesto por los asesores técnicos del Comité Técnico Asesor del Conteo Rápido.

El marco muestral será el listado definitivo de casillas aprobado para las elecciones de la entidad y la selección se realizará de acuerdo al diseño de muestreo establecido por los integrantes del COTECORA.

3.2 Protocolo de selección y resguardo de la muestra.

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

Selección de la muestra

1.- La y los integrantes del Cotecora entregarán el software necesario para la selección de la muestra junto con sus códigos de integridad.

2.- A la vista de los asistentes se instalará 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 obtendrá los códigos de integridad de los archivos entregados y los comparará con los generados.

4.- A la vista de los asistentes se instalará 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 se requiere la participación de seis personas elegidas de entre los asistentes. Cada una de ellas seleccionará un dígito entre 0 y 9, lo anotarán en un formato diseñado para tal efecto.

6.- Se solicitará a un miembro del Cotecora iniciar la etapa de ejecución de la selección de la muestra.

7.- A la vista del notario público y de los asistentes, los números se ingresarán en el programa de selección de la muestra.

8.- Una vez ingresados los números para construir la semilla, se continuará con el programa para seleccionar la muestra. Esta última quedará grabada en el disco duro que se entregará al Director Ejecutivo de Organización Electoral.

9.- 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 así lo soliciten.

Resguardo de la muestra

10.- La muestra será grabada en un disco compacto no regrabable.

11.- La muestra segmentada en los 21 distritos federales se enviará 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 estará en ceros y servirá exclusivamente para el envío de esta muestra. Lo anterior con el fin de que se preparen 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 serán guardados en un sobre. Asimismo el equipo de cómputo en el cual fue generada la muestra se resguardará. Ambos serán sellados y rubricados por el notario público.

13.- El notario público entregará al Director Ejecutivo de Organización Electoral 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.

4.1 Consideraciones generales

La operación logística del conteo rápido 2016 estará a cargo de la Dirección Ejecutiva de Organización Electoral del OPLE Veracruz, misma que instalará un Centro Estatal de Información del Conteo Rápido para el Proceso Electoral 2015-2016 (CEICR).

El CEICR estará ubicado en las instalaciones del OPLE Veracruz, y dispondrá de los recursos materiales necesarios 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 un sistema de captura en línea.

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, y utilizando los formatos diseñados para tal fin, reporten al Centro Estatal de Información del Conteo Rápido (CEICR) los datos de la votación para Gobernador del estado de Veracruz de las casillas seleccionadas en la muestra.

El OPLE Veracruz deberá salvaguardar la seguridad y confidencialidad de la información en todos los procesos de operación del Conteo Rápido

4.2 Objetivo general

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 este elabore las estimaciones estadísticas para conocer las tendencias de la votación de la elección de Gobernador del estado de Veracruz a celebrarse el domingo 5 de junio de 2016.

4.3 Esquema logístico de funcionamiento

Se desarrolla conforme a los siguientes pasos:

- 1.- El CAE, o en su caso el SE, es el encargado de recabar en el formato diseñado para tal fin, los datos de la votación de la elección de Gobernador, contenidos en las actas de escrutinio y cómputo de las casillas seleccionadas de la muestra que se encuentren asignadas en su Área de Responsabilidad Electoral (ARE). De

manera inmediata llama, a través del medio de comunicación que le fue asignado, al Centro Estatal de Información (CEICR) para reportar la información recopilada.

2.- El operador de cómputo o bien el sistema de recepción de datos en el Centro Estatal de Información del Conteo Rápido (CEICR) recibe la llamada del CAE, o en su caso del SE, y captura directamente en el Sistema de Información del Conteo Rápido (SICRA) los datos que éste le comunique.

3.- El COTECORA procesa la información proporcionada por el sistema y realiza las estimaciones estadísticas que considere convenientes.

4.4. Capacitación del personal involucrado

Para llevar a cabo la capacitación del personal involucrado en la ejecución del Conteo Rápido, se elaborarán dos documentos principales: a) Guía de procedimientos de la operación logística del Conteo Rápido Veracruz 2016 y b) Guía de capacitación para la ejecución del Conteo Rápido Veracruz 2016. Ambos documentos se elaborarán con la supervisión del COTECORA.

Ambos documentos son la base para la capacitación de los operadores del Centro Estatal de Información del Conteo Rápido (CEICR), los capacitadores asistentes electorales 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, haciendo énfasis en los siguientes procedimientos:

Proceso de recolección de la información

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

1.- Presentarse a las 18:00 horas en la primera casilla de la muestra asignada en su ARE de la cual deba realizar el reporte de resultados de la votación al Conteo Rápido.

2.- Una vez que los funcionarios de casilla hayan concluido con el escrutinio y cómputo de la elección de Gobernador y llenado el acta correspondiente, transcribirá los datos de la votación de cada partido político, coalición, candidato independiente en su caso, candidato no registrado y votos nulos, en el formato diseñado para tal fin.

3.- Inmediatamente después se comunicará al Centro Estatal de Información del Conteo Rápido (CEICR), a través del medio de comunicación asignado y, en cuanto entre la llamada, procederá a realizar el reporte de los datos incluyendo una clave única de identificación o autenticación.

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

Proceso de reporte y captura de datos

1.- El operador de cómputo o bien el sistema de recepción de datos en el Centro de Estatal de Información (CEICR) 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.

2.- El operador de cómputo, o bien el sistema de recepción de datos, registrará la información dictada por el CAE o SE. Una vez en el sistema de información del conteo rápido, en cada campo irá seleccionando o capturando la información que le sea transmitida. La captura estará sujeta a las siguientes consideraciones:

- El operador de cómputo establecerá el ritmo de dictado y las pautas para que el CAE o SE continúe con el reporte.
- La clave de autenticación es fundamental para el registro de la comunicación y el ingreso de 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 Vocal de Organización del Consejo Distrital del OPLE del estado de Veracruz.

3.- En caso de que el CAE o SE no pueda realizar el reporte de datos, se habilitará un esquema de contingencia para recibir las llamadas o comunicación de datos.

Simulacros

La realización de simulacros tienen la finalidad de detectar oportunamente cualquier posible falla y realizar los aspectos necesarios para garantizar el puntual desarrollo de la operación logística el día de la jornada electoral. Los simulacros permitirán evaluar el funcionamiento de los siguientes aspectos:

- I. Los medios y sistemas para la captura, transmisión y recepción de la información electoral;
- II. El proceso operativo en campo;
- III. El flujo de llegada de la información de las casillas;
- IV. Los medios y sistemas para conocer la cobertura geográfica de la muestra recibida;
- V. Los métodos de estimación, y
- VI. La generación y envío del reporte con las estimaciones de la simulación a los integrantes del Órgano de Dirección Superior del OPL.

Se realizará un primer simulacro el día 8 de mayo de 2016 y, un segundo el día 22 de mayo de 2016. En ellos participarán, los CAE, los SE y los operadores de cómputo del CEICR.

Previo a su realización, se remitirá de parte de la Dirección Ejecutiva de Organización Electoral del OPLE del estado de Veracruz las instrucciones correspondientes, entre las que se deben incluir:

- Cada CAE reportará resultados ficticios de la votación de un número determinado de casillas de su ARE, mismos que previamente le proporcionará el Vocal de Organización del Consejo Distrital del OPLE del estado de Veracruz.
- Los SE también reportaran resultados de algunas de las casillas asignadas a su Zona de Responsabilidad Electoral (ZORE), y se encargarán de dar seguimiento a los reportes de los CAE a su cargo.
- El Vocal de Organización del Consejo Distrital del OPLE del estado de Veracruz entregará a los CAE y los SE los formatos prellenados, los cuales se utilizaran en los simulacros.

V. DESARROLLO Y RESULTADOS EL DÍA DE LA JORNADA ELECTORAL

- Durante la jornada electoral, los CAE conocerán la o las casillas que hubiesen sido seleccionadas en la muestra.
- Los CAE o SE seleccionados deberán establecer contacto con el CEICR al realizar el primer reporte al SIJE.
- A partir de las 18:00 el COTECORA revisará las remesas de datos en el CEICR con una periodicidad de hasta 30 minutos.

- A partir de las 21:00 horas y hasta la entrega de los resultados, con una periodicidad de hasta 60 minutos, el COTECORA informará al OPLE los avances en la recopilación de datos.
- El informe definitivo del COTECORA se entregará al Presidente del Consejo General del OPLE entre las 22 y 23 horas del día de la jornada electoral.
- Durante la jornada electoral las remesas del conteo rápido serán confidenciales y sólo las podrán conocer los miembros del COTECORA.
- El Consejo General estará obligado a difundir el mismo día de la jornada electoral, los resultados numéricos del ejercicio del Conteo Rápido, precisando el avance de la muestra de acuerdo con las cifras proporcionadas por el COTECORA.

VI. DIFUSIÓN DE RESULTADOS

Al día siguiente de la jornada electoral, el COTECORA entregará al OPLE Veracruz la siguiente información, misma que este deberá publicar en su página de internet:

- I. El protocolo de selección de la muestra;
- II. Las fórmulas de cálculo utilizadas para cada método preseleccionado;
- III. El informe de resultados, y
- IV. La base de datos utilizada en las estimaciones del Conteo Rápido, incluyendo las casillas que fueron seleccionadas en la muestra, así como las casillas que se utilizaron en el cálculo final, cada una con el resultado de la elección.

VII. METODOS DE ESTIMACIÓN.

La muestra obtenida para el conteo rápido será procesada por tres métodos de estimación alternativos con el fin de garantizar la confiabilidad de las estimaciones.

Los métodos a utilizar son:

- 1) Método bayesiano
- 2) Método clásico 1: Estimador de razón combinado
- 3) Método clásico 2: Estimador de razón separado

El método de estimación bayesiano hace uso del diseño estratificado, se apoya de un modelo normal y hace uso de distribuciones iniciales no informativas con el fin de que las inferencias dependan exclusivamente de los datos.

El método clásico 1 hace uso del diseño estratificado, usa estimadores clásicos de razón combinados y construye intervalos de confianza asintóticamente mediante un proceso de aproximación de la varianza.

El método clásico 2 hace uso del diseño estratificado y usa estimadores de razón separados usando la información del listado nominal. Finalmente la varianza es aproximada y se construyen intervalos de confianza asintóticos.

Los detalles técnicos de estos tres métodos de estimación se encuentran en el apéndice.

Integración de las estimaciones de los tres métodos

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

De cualquier forma las estimaciones por intervalo reportadas serán de al menos 95% de confianza.

VI. APÉNDICES METODOLÓGICOS

6.1 Selección del diseño muestral

ESTUDIO 1: (Dr. Luis Enrique Nieto Barajas)

El estado de Veracruz es un estado basto cuya geografía política está dividida en 212 municipios. Para fines electorales existen 30 distritos locales formados por la unión de municipios. Cada distrito electoral está dividido en secciones electorales las cuales pueden ser del tipo rural o urbano.

La base de estudio para la elección del diseño muestral es la elección de gobernador del año 2010 y la división distrital local es la del año 2015.

Dada la información disponible, se propusieron tres tipos de estratificación para estudiar los errores alcanzados por cada una de ellas. Las estratificaciones a considerar son:

- i) Basada en distritos locales: 30 estratos
- ii) Basada en distritos locales y tipo rural o urbano: 60 estratos
- iii) Basada en municipios: 212 estratos

El objetivo del conteo rápido es estimar la proporción efectiva de votos a favor de un candidato, digamos p_j . Con base en una muestra aleatoria estratificada de tamaño n se propone el siguiente estimador de razón

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

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 qué tan bueno es el estimador \hat{p}_j es necesario conocer su distribución de muestreo. Esta distribución se estima mediante un proceso de simulación repetida de muestras aleatorias de tamaño n , calculando para cada muestra el estimador \hat{p}_j . Los distintos valores del estimador para las distintas muestras se representan mediante un histograma (ver Figura 1). Este histograma es un estimador de la distribución de muestreo de \hat{p}_j y nos servirá para estudiar sus propiedades.

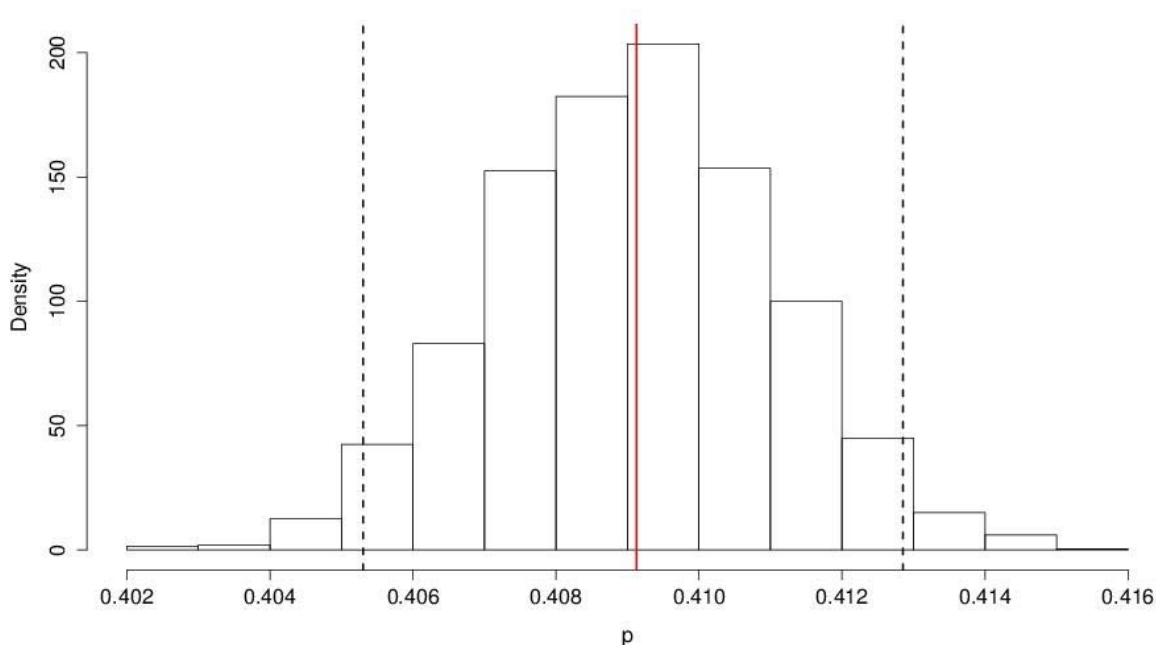


Figura 1. Distribución de muestreo del estimador de la proporción efectiva de votos a favor del PRI. Valor real 0.4090 (línea vertical roja) y cuantiles de orden 2.5% y 97.5% (líneas verticales punteadas).

En específico una característica de interés es la determinación del máximo error de estimación, denotado por B , alcanzado para el tamaño de muestra específico n . La cantidad B satisface

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

Para la determinación de B se fija el valor de la probabilidad $1 - \alpha = 0.95$ y para cada tamaño de muestra n se obtiene B como

$$B = \max(p_j - q_{0.025}, q_{0.975} - p_j)$$

donde $q_{0.025}$ y $q_{0.975}$ son los cuantiles de orden 2.5% y 97.5% de la distribución de muestreo de \hat{p}_j (ver líneas punteadas verticales en la Figura 1).

Para determinar cuál de los tres criterios de estratificación (i), (ii) o (iii) es el mejor, se hizo un estudio de simulación considerando 2,000 réplicas variando el tamaño de muestra en intervalos de 250 casillas desde 500 a 2,500 y asignando la muestra dentro de cada estrato de manera proporcional al número de casillas. El candidato que se tomó como referencia para el estudio es el PRI. En cada caso se calculó el máximo error de estimación B alcanzado. Los resultados se muestran en la Tabla 1.

Tabla 1. Errores de estimación (en porcentajes) alcanzados por cada criterio de estratificación para distintos tamaños de muestra.

Tamaño de muestra	Distritos locales	Distritos locales y tipo	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

Para poder comparar de una mejor manera los errores de la Tabla 1, en la Figura 2 se presentan estos errores para los tres criterios de estratificación en función del tamaño de muestra.

Como se observa en la Figura 2, 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.

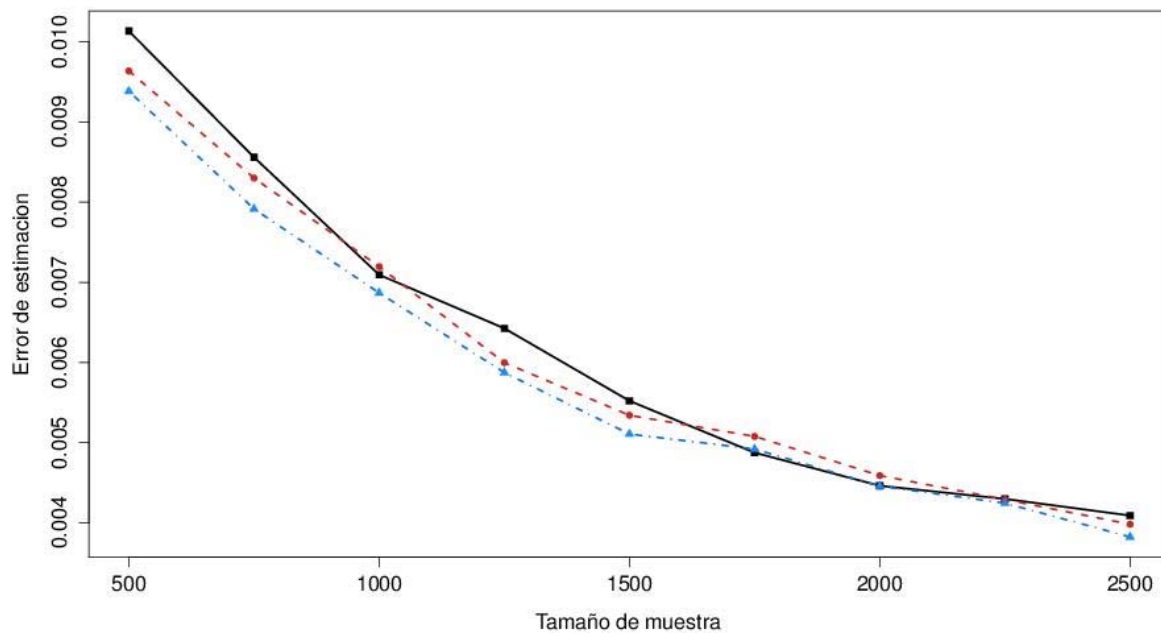


Figura 2. Gráfica comparativa de los errores de estimación alcanzados por los criterios de estratificación por tamaño de muestra. Distritos locales (cuadrados), distritos locales y tipo (círculos), y municipios (triángulos).

Como criterio adicional para la determinación del tamaño de muestra se considera otro estudio de simulación para ver la factibilidad del levantamiento de la información dado el número de CAEs disponibles. Para este nuevo estudio se usa la información de las AREs (áreas de responsabilidad electoral) por CAE en la elección del 2012 en la que había 2,210 CAEs asignados para el estado de

Veracruz. Se tomaron distintos tamaños de muestra y se obtuvo la proporción de CAEs que tienen asignada al menos una casilla de la muestra, y para aquellos CAEs en muestra la distribución del número de casillas que les tocaría reportar a cada uno. Los resultados se presentan en la Tabla 2.

Tabla 2. Distribución de los CAE's 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's en muestra	44%	51%	56%	61%

La decisión del tamaño de muestra y del criterio de estratificación se basará en las siguientes consideraciones:

- El número de CAE's disponibles para el levantamiento de la muestra para esta elección del 2016 será de 2,186 más 379 supervisores. Número similares a los que se tenían en el 2012.
- Los máximos errores de estimación considerados en otros conteos rápidos federales han sido, 0.3% para el 2006 y 0.5% para el 2012.
- El tener falta de información en algún estrato complica los procesos de estimación, por lo que es recomendable tener menos estratos.
- Es importante garantizar que se cuente con información suficiente para producir las estimaciones alrededor de las 22:30 hrs.
- Para mayores tamaños de muestra se logra un menor error de estimación, pero por otro lado se complica el levantamiento de la información ya que un porcentaje considerable de CAEs tendrían a su cargo 3 o más casillas.

Tomando en cuenta las consideraciones anteriores se propone como criterio de estratificación la (i) basada en distritos locales únicamente con un tamaño de muestra de 1,250 casillas y con asignación proporcional al tamaño del estrato. Suponiendo que el flujo de información no es uniforme, se espera que para las 22:30 hrs. se tenga un 80% de la muestra (1,000 casillas) con lo que se tendría un máximo error de estimación de 0.7% con 95% de confianza.

ESTUDIO 2: (Dr. Sergio Francisco Juárez Cerrillo)

Diseño Muestral

Muestreo aleatorio estratificado. La determinación del tamaño de muestra junto con la asignación de la muestra al estrato se hace con el procedimiento de Bethel (1989).

Tamaño de Muestra y Asignación Multivariada

La asignación óptima del tamaño de muestra a los estratos la hacemos con el método de Bethel (1989). Consideremos un muestreo estratificado con J variables de interés. Consideremos al estimador de la media poblacional de la variable j, (j = 1,2, ..., J)

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

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

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

Supongamos que muestrear a una unidad muestral en el estrato h cuesta c_h , de modo que se tiene la función de costo

$$C(n_1, \dots, n_n) = \sum_{h=1}^L c_h n_h.$$

Esta función se desea minimizar con la restricción de que las J varianzas sean menores que constantes dadas, es decir, $\text{Var}(\bar{X}_j) \leq k_j$, ($j = 1, 2, \dots, J$). Este procedimiento de asignación óptima multivariada fue propuesto por Bethel (1989). Es un problema de programación convexa, la solución $n_1^*, n_2^*, \dots, n_L^*$, no tiene forma cerrada pero el algoritmo numérico para encontrarla está implementado en el paquete de R `Bethel`.

Estudio de Simulación para estimar Precisión

El marco muestral está dado por el listado de las 9827 casillas de las elecciones para gobernador 2010. Los parámetros a estimar son las proporciones de los cuatro partidos de mayor votación: PRI con 40.899%, PAN con 38.303%, PRD con 6.709% y Convergencia con 3.944%.

Se consideran dos estratificaciones: la primera es con los estratos dados por cada uno de los 30 distritos estatales; la segunda es con los estratos dados por el cruce entre los 30 distritos estatales y el tipo de casilla (urbana y no urbana).

Para estimar mediante simulación a las precisiones de los métodos de estimación propuestos primero determinamos el tamaño de muestra así como la asignación a los estratos con el procedimiento de Bethel. Consideramos como variables objetivos a las suma de las votaciones de PAN con PRD por un lado y por otro PRI con PVEM. Esto lo hacemos así porque estos partidos han establecido coaliciones para las elecciones de gobernador de este junio de 2016. Suponemos también que el costo de muestrear en cada estrato es fijo, c , de manera que la función a minimizar es el tamaño de muestra total $C(n_1, \dots, n_n) = c(n_1 + \dots + n_L)$. Sin pérdida de generalidad suponemos que $c = 1$. La siguiente tabla muestra las varianzas de las

variables de interés (PRI con PVEM y PAN con PRD), junto con el tamaño del estrato, los c , y la tasa mínima de muestreo dentro de cada estrato.

Tabla 3. Varianzas de las variables objetivo.

strata	VAR(PRI+PVEM)	VAR(PAN+PRD)	CASILLAS	c	n	n_2
1	2521.0818	4132.1043	355	1	3	0.04
2	3461.0056	5838.0945	261	1	3	0.04
3	4325.0488	9317.9140	260	1	3	0.04
4	2397.1938	2607.4008	284	1	3	0.04
5	1875.5525	2104.7766	320	1	3	0.04
6	1573.5342	975.2319	307	1	3	0.04
7	2436.7934	4490.8151	339	1	3	0.04
8	1962.8544	2312.3339	338	1	3	0.04
9	14300.4669	4628.6319	291	1	3	0.04
10	2275.7418	3720.6400	317	1	3	0.04
11	989.6908	1184.7174	316	1	3	0.04
12	904.4387	4558.9880	302	1	3	0.04
13	1950.8813	2838.5372	329	1	3	0.04
14	3686.2990	3258.0689	310	1	3	0.04
15	3464.1924	3582.7646	402	1	3	0.04
16	2607.5939	2030.0428	381	1	3	0.04
17	3106.1158	3330.5540	350	1	3	0.04
18	3064.5582	3894.3523	331	1	3	0.04
19	3362.4976	4014.1909	390	1	3	0.04
20	719.8397	1063.8630	334	1	3	0.04
21	979.4539	1418.4197	367	1	3	0.04
22	1746.1721	2110.9516	366	1	3	0.04
23	2837.8837	12335.3965	346	1	3	0.04
24	1799.0912	3802.7125	276	1	3	0.04
25	2497.3197	2486.4834	331	1	3	0.04
26	3131.0137	4452.9400	352	1	3	0.04
27	2671.3087	4272.7950	335	1	3	0.04
28	2530.8423	1958.0538	287	1	3	0.04
29	1135.6530	1526.0632	351	1	3	0.04
30	2989.6115	2000.3448	299	1	3	0.04

Los tamaños (totales) 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. Los tamaños de muestra óptimos resultantes se presentan en la Tabla 4. Vemos 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.

Tabla 4. 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

Para cada tamaño de muestra simulamos 2000 veces el proceso de muestreo del diseño estratificado. La Tabla 5 muestra la asignación óptima del tamaño de muestra a cada estrato. Se agregaron dos tamaños más de muestra: 1500 y 2000.

Tabla 5. Tamaños de muestra óptimos

	Estr387	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
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

La Tabla 6 muestra las estimaciones del error de estimación al 95% nivel de confianza $B = 2 \times \sqrt{S^2(\hat{p}_j)}$, donde $S^2(\hat{p}_j)$ es la varianza de los valores \hat{p}_j simulados para cada tamaño de muestra y los partidos PRI, PAN, PRD y CONVERGENCIA. Además se incluyen los errores de estimación para las coaliciones PAN con PRD y PRI con PVEM.

Tabla 6. Precisiones estimadas al 95% de confianza

<i>n</i>		PRI	PAN	PRD	CONV	PANPRD	PRIALIANZA
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
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

Propuesta de Muestra

Con base a los resultados anteriores se propone lo siguiente:

- 1) Estratificación con los 30 distritos electorales.
- 2) Una muestra estratificada de tamaño entre 1200 y 1350 para así alcanzar niveles de precisión de alrededor de 0.6 punto porcentual con un nivel de confianza del 95% para estimar las proporciones de votación para el PAN-PRD, PRI-PVEM, PRI, PAN, PRD y CONVERGENCIA en las elecciones del 2010.
- 3) Asignación óptima de la muestra con el método de Bethel tomando en cuenta como variables objetivo las coaliciones PAN-PRD y PRI-PVEM.

Referencias

- Bethel, J.W. (1989). Sample Allocation in Multivariate Surveys. *Survey Methodology*, vol. 15, pp. 47-57.
- 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.

6.2 Nota técnica de los métodos de estimación

METODO 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.

METODO CLASICO 1: ESTIMADOR DE RAZON 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 $S_{U_h}^2$ la varianza de la votación efectiva más los votos anulados, es decir

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

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}

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

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

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

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

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

METODO CLASICO 2: ESTIMADOR DE RAZON SEPARADO

(Propuesta del Dr. Sergio Francisco Juárez Cerrillo)

Sea Y_h el número total de individuos que están en la lista nominal del estrato h . Entonces el total de individuos en la lista nominal es

$$Y = \sum_{h=1}^L Y_h.$$

Sea y_{hi} el número de individuos en la lista nominal en la casilla i de la muestra del estrato h . Consideremos al promedio de votación por casilla para el partido j en la muestra del estrato h

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

y consideremos al promedio de individuos en la lista nominal en la muestra del estrato h

$$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}.$$

El estimador de razón del total de votos para el partido j dentro del estrato h es

$$\hat{t}_{hj} = Y_h \times \frac{\bar{x}_{hj}}{\bar{y}_h}.$$

De modo que un estimador del total de votos para el partido j en la población está dado por $\tilde{t}_j = \sum_{h=1}^L \hat{t}_{hj}$ y el estimador de Razón Separado de R_j es

$$\hat{R}_{js} = \frac{\tilde{t}_j}{Y}.$$

Una estimación de la varianza de \hat{R}_{js} es

$$\widehat{\text{Var}}(\hat{R}_{js}) = \left(\frac{1}{Y^2}\right) \sum_{h=1}^L \widehat{\text{Var}}(\hat{t}_{hj})$$

donde

$$\widehat{\text{Var}}(\hat{t}_{hj}) = \left\{ \left(\frac{Y_h r_{hj}}{\sqrt{n_h}} \right) \times (\widehat{CV}_{Xhj}^2 - \widehat{CV}_{Yh}^2)^{1/2} \sqrt{\frac{N_h - n_h}{N_h - 1}} \right\}^2$$

donde

$$r_{hj} = \frac{\bar{x}_{hj}}{\bar{y}_h},$$

$$\widehat{CV}_{Xhj} = \frac{s_{Xhj}}{\bar{x}_{hj}},$$

$$\widehat{CV}_{Yh} = \frac{s_{Yh}}{\bar{y}_h},$$

$$s_{Xhj}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hij} - \bar{x}_{hj})^2,$$

$$s_{Yh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2.$$

El intervalo de confianza al 95% de confianza para la proporción de votación para el partido j es

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