

Procedimiento para la obtención de conteos rápidos a partir del PREP

Process to Obtain Quick Counts from PREP

Martínez-Cruz M.Á.

Sección de Estudios de Posgrado e Investigación (SEPI)
Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME)
Unidad Zácatenco, Instituto Politécnico Nacional
E-mail: mamartinezc@ipn.mx

Balankin A.

Sección de Estudios de Posgrado e Investigación (SEPI)
Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME)
Unidad Zácatenco, Instituto Politécnico Nacional
E-mail: abalankin@ipn.mx

Patiño-Ortiz M.

Sección de Estudios de Posgrado e Investigación (SEPI)
Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME)
Unidad Zácatenco, Instituto Politécnico Nacional
E-mail: mpatino2002@yahoo.com.mx

Información del artículo: recibido: mayo de 2008, aceptado: enero de 2011

Resumen

En el presente trabajo se realizó un procedimiento para encontrar muestras representativas del total de casillas instaladas en las elecciones federales a presidente de la República Mexicana de 2006 (conteos rápidos) en diferentes horas, tomando como base de datos el Programa de Resultados Electorales Preliminares (PREP), es decir, partiendo de una muestra no representativa se obtiene una muestra representativa de las elecciones federales. Por sus características de recolección de información, el PREP en las primeras horas no necesariamente forma un muestreo representativo de datos. Se observó que conforme aumenta el número de casillas utilizadas en los procedimientos de conteos, el error tiende a disminuir, de tal modo que 3 horas después de abrirse el Programa de Resultados Preliminares, la precisión del conteo reportado en el presente artículo, es mejor que la precisión lograda por el conteo rápido del Instituto Federal Electoral Mexicano. Con la aplicación del procedimiento de este trabajo, se incidirá en la minimización de costos, el aumento de la confiabilidad en el muestreo de casillas y sobre todo la aplicación de una forma precisa de distinción del candidato ganador, en elecciones federales y estatales, mucho antes de terminar el PREP.

Descriptores

- PREP
- conteo rápido
- casillas
- procedimiento
- distribución uniforme
- muestreo representativo

Abstract

Considering the Preliminary Electoral Results Program (PERP) as a database of the federal elections for president of the Mexican Republic, a methodology was developed in order to find representative samples of ballot boxes installed in the election's day (quick count) in different hours, due to its characteristics of gathering of information, the PREP in the first hours forms a non-representative sample of data. In a particular way, in the election of July 2, 2006, after 3 hours of opening the PREP, it was observed that the accuracy of the process of the quick counts was better than the one obtained by the IFE. Among other things, this allows to lower the cost, to increase the confidentiality of the ballot boxes used in the sampling and to distinguish in a precise moment the winning candidate long before PREP finishes.

Keywords

- PREP
- quick count
- ballot boxes
- process
- uniform distribution
- representative sampling

Introducción

Para contabilizar los votos de las elecciones del 2 de julio del 2006, el IFE utilizó tres distintos mecanismos: el Programa de Resultados Electorales Preliminares (PREP), el conteo rápido y los cómputos distritales.

El PREP tiene la función de informar al público sobre el flujo de los votos conforme van llegando a los distritos electorales, por lo que la información no necesariamente forma una muestra representativa de la población total.

El conteo rápido sirve para pronosticar tendencias en los resultados de la elección la misma noche en que ésta se realiza, y se presenta como una estimación de los intervalos de confianza de las proporciones de dichas preferencias.

Los cómputos distritales sirven para obtener los resultados oficiales de la elección a través del conteo de todas las actas y la suma del total de los votos emitidos en las distintas casillas del país.

En el caso del conteo rápido, se tienen datos duros, es decir, votos contados en mesas y registrados en actas. Aquí no hay opiniones o declaraciones de intención de voto falsas, ni entrevistadores poco atentos o distraídos, por lo que el error de no muestreo se elimina al máximo, quedando básicamente el error de muestreo, que es la probabilidad matemáticamente medible de que la muestra sea sesgada respecto al universo; por ejemplo: que por azar hayan salido sorteados más hogares de algún nivel económico o más mesas de las que favorecen a un candidato en particular.

Para reducir el error total (suma del error no muestral y el error muestral) del conteo rápido, hay que actuar básicamente en el error de muestreo. En el presente artículo se analizaron 30 muestras representativas con base al PREP para cada corte de hora, es decir, 30 conteos rápidos utilizando como base de datos la in-

formación del PREP en tiempo real. A mayor número de casillas en el PREP y mayor número de conteos rápidos derivados del PREP, menor es el error por muestreo que podemos obtener, reduciendo así el error total.

Por lo anterior, nuestro conteo, al basarse en teorías científicas, puede ofrecer tempranamente resultados muy cercanos a los reales.

Procedimiento

La población de estudio considerada para la realización de este conteo rápido, son las personas registradas en el listado nominal del país.

La agrupación del listado nominal está hecha en casillas electorales, considerando las variables de agrupación a la entidad federativa, distrito electoral, sección, tipo de casilla y número de casilla.

Las unidades de observación para la obtención de la muestra son las casillas electorales, que fueron instaladas el 2 de julio de 2006, registradas en el padrón electoral del Instituto Federal Electoral y registradas en el Programa de Resultados Electorales Preliminares (PREP) al momento de realizar el conteo. La población a analizar está conformada por 130,831 casillas.

En la tabla 1 se muestra cómo se procedió a enumerar las 130,831 casillas electorales comenzando por la entidad federativa de Aguascalientes, distrito 1, sección 1 y casilla B1 (correspondiéndole el número 1); terminando en la entidad federativa de Zacatecas, distrito 4, sección 1710, casilla B1 (correspondiéndole el número 130,831).

El PREP difunde en tiempo real los resultados preliminares de las casillas contabilizadas de las elecciones federales al Consejo General, los partidos políticos y la ciudadanía en general, a través de sistemas informáticos, y de manera particular, el día 2 de julio de 2006, comenzó a difundirse la información a partir de las 20:00 horas.

Núm. de casilla	EDO	DTO	SEC	CAS	PAN	APM	PBT	OTROS
1	1	1	337	B1	53	4	18	1
2	1	1	338	B1	202	45	65	41
3	1	1	338	C1	175	38	56	29
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
130829	32	4	1708	B1	6	75	82	3
130830	32	4	1709	B1	13	51	108	12
130831	32	4	1710	B1	23	63	85	21

Tabla 1. Numeración de las casillas electorales instaladas el 2 de julio de 2006

Núm. de casilla	Núm. de casilla	EDO	DTO	SEC	CAS	PAN	APM	PBT	OTROS	Hora registro
1	103687	25	3	3368	E01	12	23	6	2	02/07/06 18:35:04, 251000
2	42245	12	4	95	B	4	11	19	4	02/07/06 19:00:17, 321000
3	40557	12	1	973	E01	6	38	11	4	02/07/06 19:03:15, 059000
-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	
18944	30344	9	24	571	B	83	22	254	22	02/07/06 21:59:59, 605000
18945	40757	12	1	1961	C01	38	65	178	10	02/07/06 21:59:59, 826000
18946	25628	9	13	1799	B	133	43	237	26	02/07/06 21:59:59, 975000

Tabla 2. Casillas contabilizadas en el PREP hasta las 22:00 horas

En la tabla 2 se muestra el número de casillas contabilizadas hasta las 22:00 horas, número con el que relacionó a cada casilla en la tabla 1, el total de votos obtenidos por cada candidato político y la hora de registro en el PREP hasta las 22:00 horas.

En la figura 1, se observan los datos obtenidos por el PREP hasta las 22:00 horas, mostrando la distribución de los números que se le relacionaron a las casillas y a la distribución uniforme que más se acerca a los datos reales (18,946), ésta se representa por una línea horizontal. El comportamiento real es diferente al de una distribución uniforme.

Los intervalos muestran diferente número de casillas, mientras que el número ideal de casillas en todos los intervalos para que el PREP, a cualquier hora, tuviera una distribución uniforme, debería ser el mismo, y en

consecuencia, podríamos obtener una muestra representativa de la población. Por lo tanto, el PREP es una muestra sesgada y no representa una muestra para la población.

Se desarrolló un programa interactivo que considera el histograma de frecuencias de la distribución de casillas en el PREP y la distribución uniforme más cercana a los datos reales. Se observó básicamente el número total de intervalos, el número real de casillas en cada intervalo y el número ideal de casillas que debería haber en cada intervalo para que la distribución de los datos fuera uniforme. Se consideraron en general dos criterios para eliminar datos y conseguir la uniformidad con los datos sobrantes.

- Si el número real de casillas es menor o igual al número ideal de casillas en el intervalo, entonces no se hace nada.
- Si el número real de casillas es mayor al número ideal de casillas en el intervalo, entonces se eliminan casillas aleatoriamente en el intervalo.

Lo anterior se llevó a cabo para cada intervalo existente. Asimismo, se realizó de forma dinámica para cada nueva distribución de datos, mientras no se cumpliera con la uniformidad de los datos reales. Una vez obtenida la uniformidad de los números asociados a las casillas, se procedió a relacionar cada número con su respectiva casilla, de las cuales se sustrajo la cantidad de votantes por cada partido político, se realizó el promedio de la

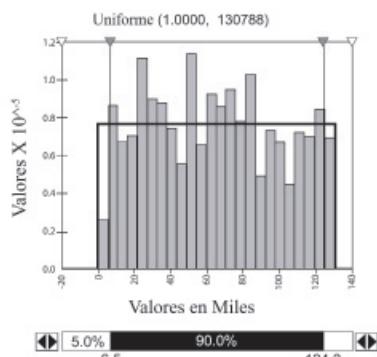


Figura 1. Distribución de “número de casilla” del PREP hasta las 22:00 horas

suma de votos de los correspondientes partidos y finalmente se normalizó para obtener la cantidad proporcional de votantes para cada candidato.

Resultados del proceso dinámico

Se hicieron cortes de datos cada hora, comenzando a las 20:00 horas y terminando a las 03:00 horas del día siguiente, realizando 30 conteos para cada corte de datos.

En la figura 1 se muestra que a las 22:00 horas la distribución inicial cuenta con 18,946 datos y la distribución uniforme que más se acopla a los datos reales no tienen mucha significancia.

La figura 2 muestra la vigésima iteración que consta de 3,045 datos. Existe mejoría, tanto gráficamente como en los valores teóricos; sin embargo, aún no se alcanzan valores aceptables para afirmar que la distribución de los datos corresponde a una distribución uniforme, es decir, no se puede obtener una muestra de datos aceptable para el conteo rápido.

La figura 3 muestra la cuadragésima y última iteración que consta de 985 número de casillas, en la cual se cumple con el criterio de uniformidad; sin embargo, hay que notar que el número de casillas final no es estadísticamente representativo para una estimación de +/-

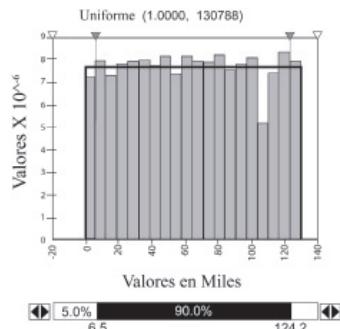


Figura 2. Vigésima iteración del procedimiento para la obtención de conteos rápidos con base al PREP a las 22:00 horas

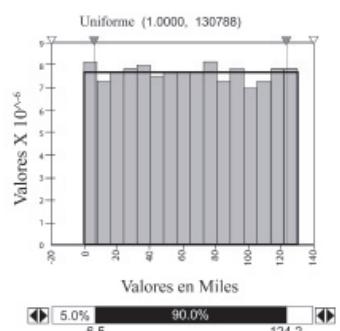


Figura 3. Última iteración de uno de los 30 conteos realizados a las 22:00 horas

0.292% con error de 95% de confiabilidad y una población de 130,831 datos.

Se relacionó la serie "número de casilla" de la última iteración con sus respectivas casillas del PREP, obteniendo el número total de votantes para cada candidato, la suma, la normalización de los datos y finalmente el porcentaje de votantes de cada participante. De esta forma, se comparó el porcentaje de votos de los dos principales contendientes para proporcionar un pronóstico de quién sería el próximo presidente, antes de terminar con el PREP y los cómputos distritales.

La figura 4 muestra la distribución del porcentaje de votación de los dos principales candidatos políticos a la presidencia de la República, uno postulado por el Partido Acción Nacional (PAN) y el otro por la Coalición por el Bien de Todos (PBT) obtenidos por 30 conteos rápidos, considerando que se cumple la condición de uniformidad y teniendo un promedio de 984 datos; también se muestra el porcentaje de votos alcanzado en el PREP a las 22:00 horas por los dos principales competidores a la presidencia de la República Mexicana.

De manera similar, se realizó el conteo rápido a las 20, 21, 23 y 24 horas del 2 de julio de 2006, así como a la 1, 2 y 3 horas del día siguiente.

En la figura 5 se muestra el proceso dinámico que se realizó para llegar a la uniformidad de los datos con el PREP a las 22:00 horas, comenzando en la iteración uno y terminando en la iteración 38, finalizando con la uniformidad de los datos para obtener una muestra representativa de las casillas registradas en el padrón electoral.

La figura 6 muestra cómo va mejorando la uniformidad de los datos conforme pasa el tiempo. La primera figura muestra la uniformidad de los datos a las 20:00 horas, y respectivamente, se muestran los resultados para las 21:00, 22:00, 23:00 y 24:00 horas del día 2 de julio de 2006 y 1:00, 2:00, 3:00 horas del día siguiente.

Los resultados variaron en la misma hora, según el número de casillas aleatorias con las que el programa haya realizado las distribuciones.

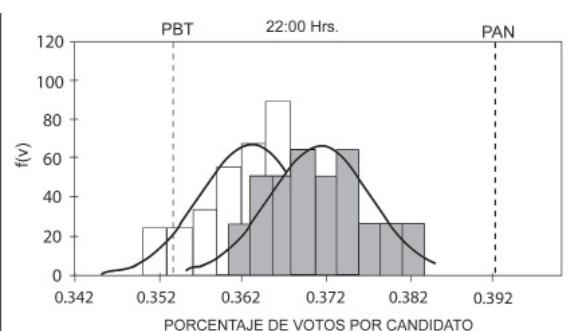


Figura 4. Resultado de 30 procedimientos para la obtención de conteos rápidos con base al PREP a las 22:00 horas

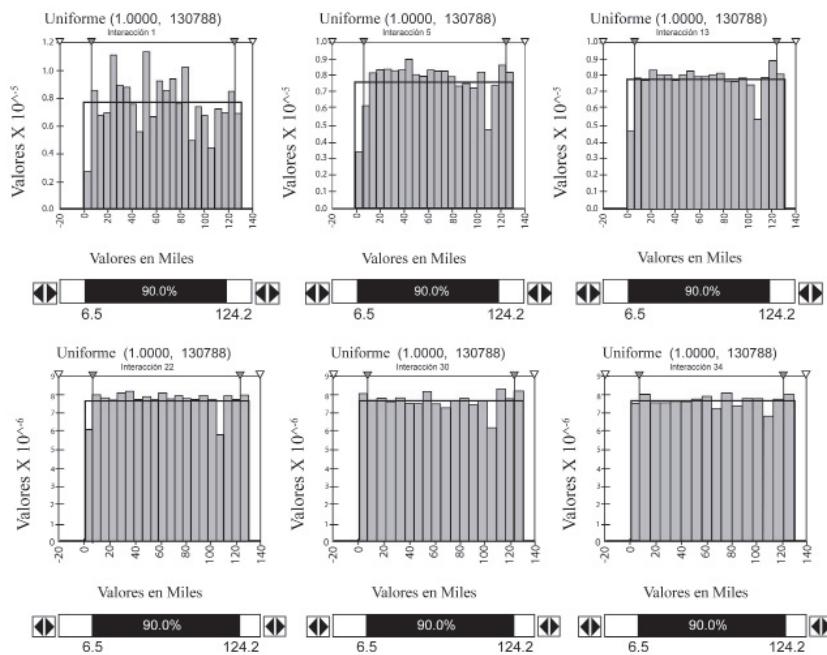


Figura 5. Proceso dinámico para la uniformidad de los datos

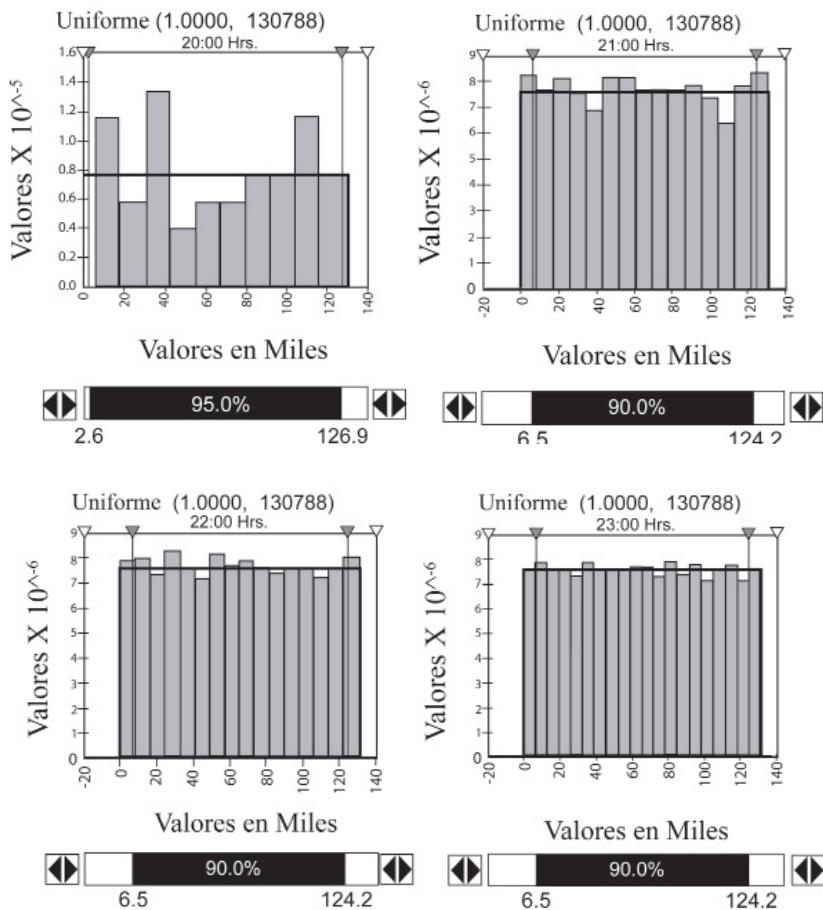


Figura 6. Uniformidad de datos en diferentes horas (continúa)

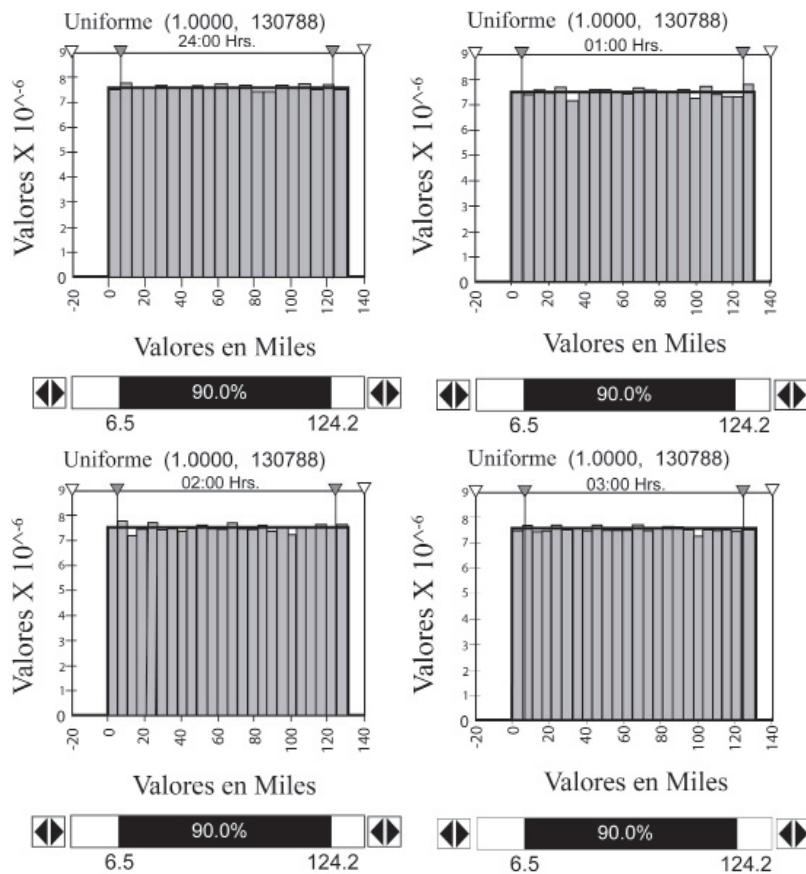
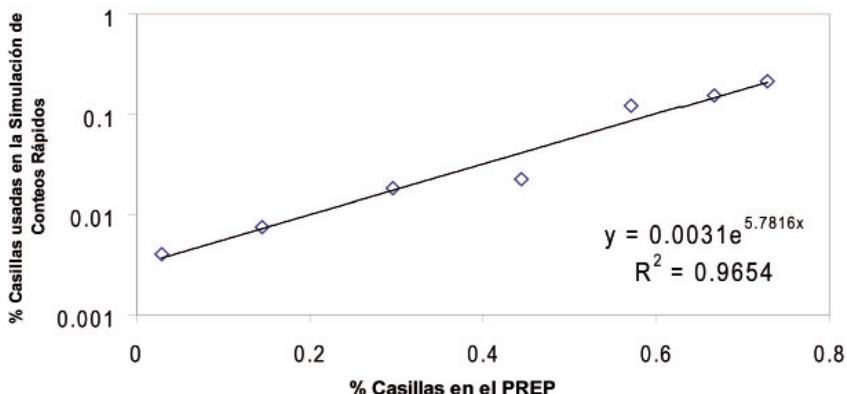


Figura 6 continuación. Uniformidad de datos en diferentes horas



En la figura 7 se muestra la relación que existe entre los datos del PREP y el número de casillas que se utilizaron en los procedimientos de los conteos rápidos a partir de las 21:00 horas y hasta las 3:00 horas del día siguiente.

En la ecuación (1) tenemos y porcentaje máximo de casillas utilizadas para obtener una muestra representativa de datos que depende de x porcentaje de casillas registradas en el PREP del total de casillas contabilizadas en los cómputos distritales.

$$y = 0.0031e^{5.7816x} \quad (1)$$

Observemos que la tendencia se cumple para 100% de las casillas contabilizadas en los cómputos distritales. Es de suponerse que para 100% de los datos, la muestra máxima representativa que podemos obtener es 100% de los datos. Por ejemplo: $1 \cong 0.0031e^{5.7816(1)}$.

Resultados de conteos rápidos en las elecciones presidenciales de 2006

En la figura 8 se muestra el resultado de los 30 procedimientos de conteos rápidos realizados con los datos del

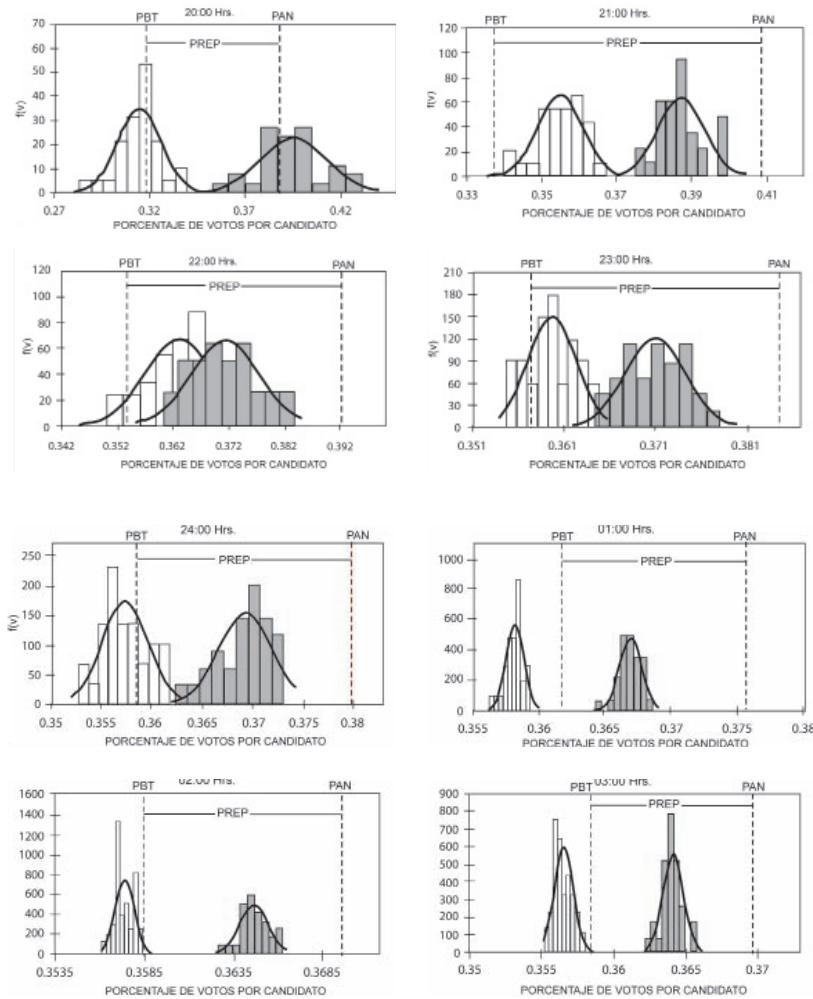


Figura 8. Resultado de 30 procedimientos de obtención del conteo rápido con base al PREP en diferentes horas

PREP a partir de las 20:00 horas y para cada hora, terminando a las 3:00 horas del día siguiente.

En particular, se hace una comparación de los dos principales candidatos a la presidencia de la república, uno postulado por el Partido Acción Nacional (PAN) y el otro por la Coalición por el Bien de Todos (PBT).

Se muestra la diferencia en el porcentaje de votos obtenidos por los dos candidatos en el PREP y las distribuciones de los 30 conteos rápidos realizados cada hora.

Las figuras 9 y 10 muestran la desviación estándar de 30 procedimientos cada hora. Nótese la disminución de la desviación estándar conforme aumenta el número de casillas utilizadas, es decir, los resultados tienden a converger en un punto conforme avanza el tiempo. Segundo los resultados el Candidato de la Coalición PBT tiende a converger primero que el candidato del PAN, por una mínima diferencia.

No hay que confundir los resultados de la desviación estándar de 30 procedimientos realizados cada hora, con

la desviación estándar del porcentaje de votantes obtenido por cada candidato en un procedimiento.

En la figura 11 se muestra la diferencia porcentual entre los dos principales candidatos a la presidencia de la República, con el intervalo de confianza considerando los 30 procedimientos realizados para el conteo rápido, esto con una estimación de confianza de 95%; y con una línea horizontal punteada la diferencia real al cierre de los cómputos distritales. Se observa que hasta las 22:00 horas el intervalo de confianza de 95% cae en diferencias negativas, lo que da a entender que hasta esa hora aun no se definía un ganador. Hasta las 23:00 horas existe una tendencia del intervalo de confianza de caer en diferencias positivas, y ya que se ha trabajado sobre una muestra aleatoria de las casillas electorales que fueron instaladas el 2 de julio de 2006, se concluye que a la hora mencionada tenemos un ganador para la candidatura a presidencia de la República Mexicana.

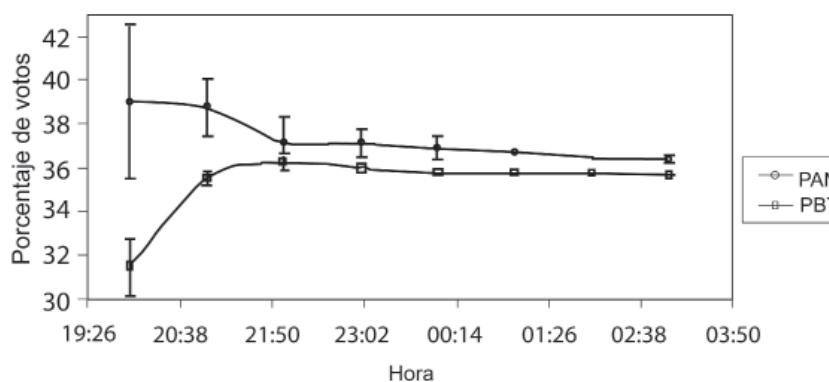


Figura 9. Porcentaje de votos para los 2 candidatos principales y desviación estándar de 30 procedimientos en sus respectivas horas

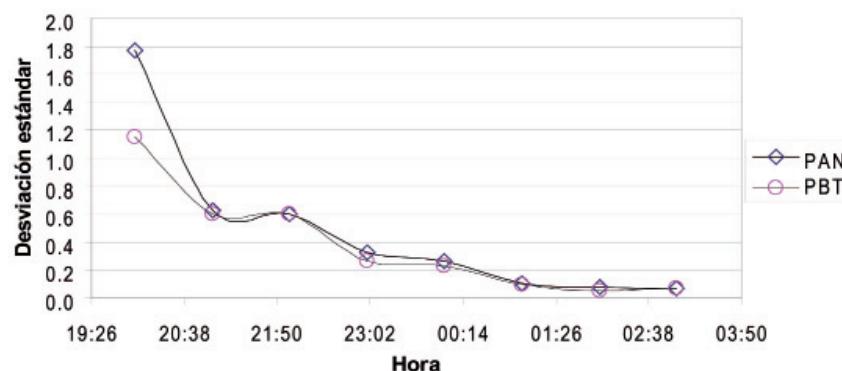


Figura 10. Porcentaje de la desviación estándar de 30 procedimientos para la obtención de conteos rápidos con base al PREP

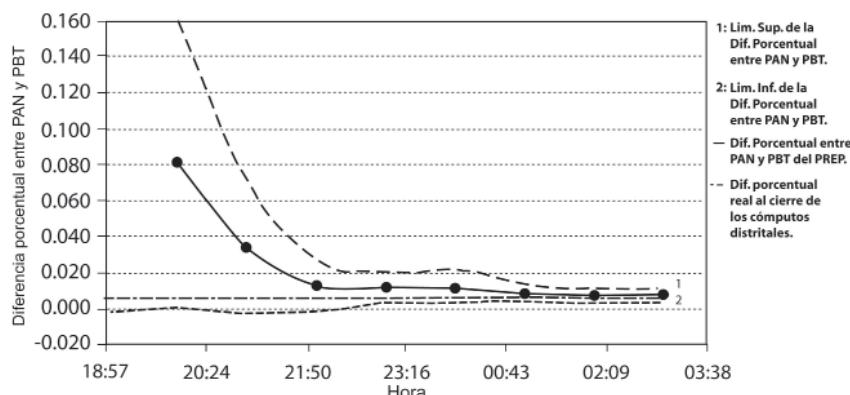


Figura 11. Diferencia porcentual entre los dos principales candidatos a la presidencia e intervalo de confianza de 95%

En la figura 12 se muestra el nivel de confianza de 95% para un procedimiento. Debe notarse que a diferencia de los 30 procedimientos, el error porcentual aumenta a $0.52\% + 0.52\% = 1.04\%$, ya que se considera la suma del error de los dos candidatos políticos; esto con 27,858 casillas contabilizadas a las 3:00 horas del 3 de julio de 2006.

La estadística nos dice que para contar con un error menor a 0.584%, que fue el concluido en los cómputos distritales, hay que dividirlo entre dos $(0.584\%)/2 = 0.292\%$, (ya que se considera la mitad de la diferencia total) que es la diferencia requerida para que no se crucen los porcentajes de votantes de los dos principales

candidatos a la presidencia; y con un nivel de confianza de 95% debemos tener 60,519 casillas en el conteo rápido. Es decir, seguridad = 95% representado con el estadístico $Z_{\alpha}^2 = 1.962$. Precisión $d = 0.584/2 = 0.292$, una proporción esperada $p = 0.5$ (50%) y $q = (1-p) = 0.5$ que maximiza el tamaño muestral. Como la población es finita, conocemos el total de la población N y se desea saber cuántas casillas n del total tendremos que estudiar, la respuesta sería:

$$n = \frac{N * Z_{\alpha}^2 * p * q}{d^2 * (N-1) + Z_{\alpha}^2 * p * q} \quad (2)$$

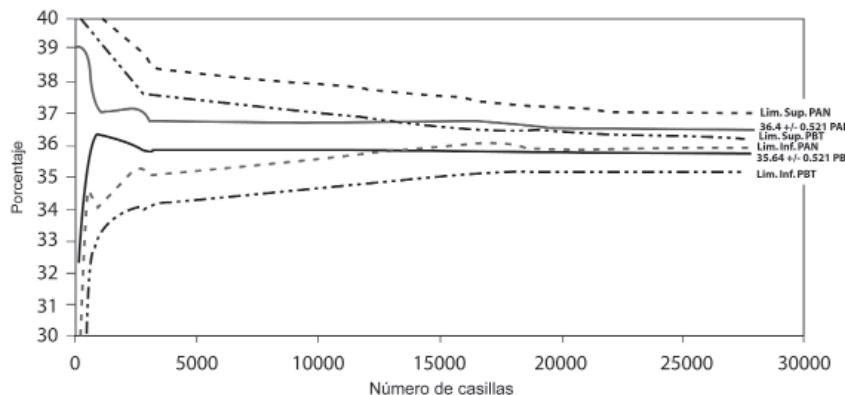


Figura 12. Estimación de error con un procedimiento

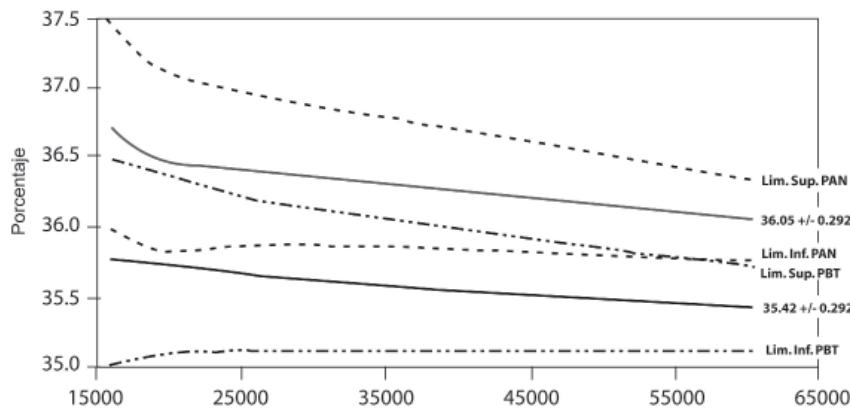


Figura 13. Estimación de error con 30 conteos rápidos

Entonces, $n = 60,519$ casillas que se deben analizar en el conteo rápido.

Considerando en (1) el número n de casillas, se alcanza cuando en el PREP se registre 86.6% de las casillas totales.

En la figura 13 se comprueba la obtención de un conteo rápido, observando que a partir de 60,519 los intervalos de confianza no se cruzan.

Conclusiones

A partir de una muestra no representativa de datos, se obtuvieron muestras representativas, lo que sirvió para la realización de varios conteos rápidos a través de los datos del PREP.

La precisión que se logró con treinta conteos rápidos, obtenidos con base en el PREP después de 3 horas de abierto, fue más certera que la lograda por el comité del conteo rápido del 2 de julio de 2006 reportada por el IFE. Puesto que a esa hora el intervalo de confianza presentado en este artículo cae en diferencias porcentuales positivas, por lo que se puede predecir que la votación

final favorecerá al candidato del PAN, es decir, el candidato ganador de la elección.

Asimismo, se comprobó que con 86.6% de las casillas totales, se logró obtener una muestra con 46.6% de datos, misma que proporciona una precisión de 0.292%.

Bibliografía sugerida

- Berger Y.G., Skinner C.J. A Jackknife Variance Estimator for Unequal Probability Sampling. *J.R. Statist. Soc. B*, (67):79-89, 2005.
- Comité Técnico Asesor para los Conteos Rápidos. *Informe sobre las actividades del comité técnico asesor para la realización de conteos rápidos*, IFE, agosto de 2000.
- Fernández A., Ruiz-Fuentes V. *Muestreo estadístico en poblaciones finitas*, Septem Ediciones, Oviedo, 2006.
- IFE. *Estadísticas de las elecciones federales 2006* [en línea], consultada el 17 de mayo de 2007, disponible en: <http://www.ife.org.mx/documentos/Estadisticas2006/index.html>
- IFE. *Programa de resultados electorales preliminares 2006* [en línea], consultada el 17 de mayo de 2007, disponible en: http://www.ife.org.mx/documentos/PREP/PREP2006/nw_que_es_el_prep.html

Knuth D.E. *The Art of Computer Programming Seminumerical Algorithms*, vol 2, Addison Wesley, Reading, 1973.

Melissa E., Nevitte N., Cowan G. *The Quick Count and Election Observation*, National Democratic Institute, Washington, DC.

Ragasol E., Cárdenas C., Muedano L., Espinosa E. *Elecciones presidenciales 2006: Análisis Estadístico de los Resultados Publicados por el PREP*, ACNielsen, a VNU business, agosto de 2006.

Seminario sobre el desempeño del Programa de Resultados Electorales Preliminares (PREP) en la elección de 2006. *El PREP: ¿Comportamiento aleatorio?*, 16 de noviembre de 2006.

Semblanza de los autores

Miguel Ángel Martínez-Cruz. Obtuvo la licenciatura en física y matemáticas en la ESFM del IPN, la maestría en ingeniería de sistemas y el doctorado en ingeniería mecánica, ambos en la ESIME del IPN, por su tesis de maestría obtuvo el primer lugar en tesis de posgrado por el IPN. Actualmente desarrolla análisis fractal sobre sistemas complejos, se desempeña como profesor investigador en el IPN y forma parte del Sistema Nacional de Investigadores.

Alexander Balankin. Nació en Moscú en 1958, llegó a México en 1992 y se naturalizó mexicano en 2000. Dos años después, obtuvo el Premio Nacional en Ciencias y Artes 2002. En 2005 fue galardonado con UNESCO Science Prize por el desarrollo de mecánica fractal y sus aplicaciones en ingeniería. Es investigador nacional nivel III. Es miembro del Consejo Consultivo en Ciencias de la Presidencia de la República, miembro de la Academia Mexicana de Ciencias y miembro de la Sociedad Mexicana de Física, entre otros. Actualmente es profesor investigador de tiempo completo de la ESIME Zácatenco del IPN.

Miguel Patiño-Ortiz. Egresó de la ESIME-Zacatenco del IPN como ingeniero en comunicaciones y electrónica, con especialidad en computación. Concluyó la maestría en ingeniería de sistemas y el doctorado en ingeniería con mención honorífica, dentro de la Sección de Estudios de Posgrado e Investigación de la misma institución. Realizó un posdoctorado en la Universidad Autónoma Metropolitana en nanotecnología y nanomedicina. Actualmente desarrolla investigación sobre análisis fractal en sistemas complejos. Es profesor investigador en el IPN y miembro del Sistema Nacional de Investigadores.