

PONENCIAS MESA 5

.....

José Mendoza

COVARRUBIAS-SOFRES, TV AZTECA

Encuesta de Salida y Conteo Rápido de la Elección Presidencial

Covarrubias-Sofres para TV Azteca

INTRODUCCIÓN

El 2 de julio de 2000 Covarrubias y Asociados realizó encuestas de salida y conteo rápido para estimar los resultados de votación de las elecciones para:

- Presidente de la República
- Jefe de Gobierno del Distrito Federal
- Gobernador de Morelos
- Gobernador de Guanajuato

Adicionalmente a la producción de estimaciones tempranas de voto, la encuesta de salida tuvo otros tres objetivos: uno, describir el perfil de los votantes en función de su edad, sexo, escolaridad y ocupación. Dos, explorar las razones que motivaron el voto y, tres, conocer el momento en que los votantes tomaron la decisión de votar por quien votaron para la Presidencia de la República.

A fin de garantizar una adecuada ejecución y desarrollo del proyecto, Covarrubias y Asociados trabajó durante la jornada electoral conjuntamente con Sofres, empresa de reconocido prestigio internacional, líder en Francia en el ramo de encuestas de salida y conteos rápidos.

Nuestro compromiso fue cumplido con oportunidad, según consta en actas del Notario Público No. 218 del Distrito Federal, Lic. José Luis Villavicencio.

La presentación se divide en dos partes. La primera, donde se da cuenta de los resultados de la encuesta de salida y el conteo rápido incluyendo los principales aspectos metodológicos del diseño de muestra y la segunda, en la

que se presentan los resultados de una simulación cuya finalidad es aportar algunos elementos que sirvan de base en estudios futuros de esta índole con relación a la selección de muestras y cálculo de estimadores.

ENFOQUE METODOLÓGICO

El conteo rápido es una técnica más confiable para estimar los resultados de una elección en comparación con la información de las encuestas de salida. En el conteo rápido se tiene información de todos los electores en comparación a una muestra en las encuestas

de salida. Adicionalmente, el conteo rápido tiene la ventaja de no depender de la imprecisión de las entrevistas. En este sentido, la encuesta de salida resulta un “valor agregado” de gran utilidad para el cliente por un costo marginal.

La encuesta de salida fue una encuesta corta de seis variables, cuyos resultados se transmitieron y procesaron el mismo día de la elección. La pregunta sobre el voto se hizo con réplicas de boletas marcadas ostensiblemente como inválidas y urnas, siendo la última pregunta de la entrevista. El cuestionario y la boleta utilizados son los siguientes:

Nacional

Julio 2, 2000

1. Edad 1. 18-25 3. 36-45 2. 26-35 4. 46 y más	2. Sexo 1. Masculino 2. Femenino
3. Escolaridad, ¿hasta qué año estudió? (Si no es completa se haya hasta la categoría anterior) () 1. Sin escolaridad: primaria incompleta 2. Primaria completa 3. Secundaria completa (técnica equivalente) 4. Preparatoria completa (técnica equivalente) 5. Profesional completa 6. Posgrado	4. Ocupación principal () 1. Ama de casa 2. Estudiante 3. Trabajador del campo 4. Trabaja por su cuenta 5. Empleado de Gobierno 6. Empleado empresa privada 7. Dueño de su negocio 8. Profesionalista independiente 9. Desempleado/jubilado
5. Usted acaba de votar por algún partido/candidato para la Presidencia de la República. En relación a eso ¿Cuál fue la principal razón que lo decidió a votar por quien votó? (Escriba la frase que vaya más de acuerdo con su manera de pensar) () 1. Por el mejor partido 2. Por la esperanza de tener una mejoría económica 3. Por el mejor candidato 4. Por sacar al PRI de la Presidencia 5. Por la estabilidad económica 6. Más vale malo por conocido 7. Por acabar con la corrupción 8. Por la paz, la estabilidad social 9. Por la defensa de los pobres	6. ¿En qué momento tomó la decisión de votar por quién votó? (Leer o mostrar alternativas) () 1. Hoy mismo 2. En la última semana 3. En el último mes, después del 2do. debate 4. Hace dos meses, después del 1er. debate 5. Hace más de dos meses

ELECCIÓN PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS	
Alianza por el Cambio  Vicente Fox Quesada ①	Partido Revolucionario Institucional  Labastida Ochoa ②
Alianza por México  Cuauhtémoc Cárdenas Solórzano ③	Partido de Centro Democrático  Manuel Camacho Solís ④
Partido Auténtico de la Revolución Mexicana  Porfirio Muñoz Ledo ⑤	Democracia Social Partido Político Nacional  Gilberto Rincón Gallardo ⑥
Si votó por algún candidato no registrado, escriba aquí el nombre ⑦	

BOLETA INVÁLIDA

El diseño de muestra

El diseño muestral fue el mismo para el conteo rápido y las encuestas de salida, dado que el objetivo principal del proyecto fue elaborar estimaciones tempranas de los resultados de la votación con alto nivel de precisión. El diseño muestral se estableció en función de maximizar la precisión de las estimaciones en el conteo rápido.

El diseño de la muestra es estratificado por estado y tipo de sección electoral, con asignación proporcio-

nal del número de secciones electorales respecto a la lista nominal. Cada estrato se integra por conglomerados de ciudadanos electores agrupados en las secciones electorales. El método de selección de secciones electorales al interior de cada estado es PPT respecto a la lista nominal.

Para el conteo de salida, en cada sección electoral se consideró una segunda etapa de muestreo realizado de manera sistemática, a fin de determinar cuáles de los ciudadanos que acuden a votar a la sección electoral se entrevistarían. En esta encuesta se estudiaron las

preferencias electorales, algunas variables relacionadas con el perfil de los votantes, las razones que impulsaron el voto y el momento de su decisión.

Para el cálculo de los estimadores se emplearon dos métodos: el clásico (atado al diseño de muestra) en el que se utilizaron ponderadores para ajustar por totales estatales y tipo de sección, y otro basado en el comportamiento electoral pasado, en este caso, las elecciones federales de 1997. El método clásico es más conveniente cuando se tiene colectada la información de prácticamente la totalidad de la muestra, en tanto que las estimaciones basadas en comportamiento histórico resultan más convenientes cuando se cuenta con escasa información de la muestra, lo que ocurre en los primeros minutos después del cierre oficial de casillas.

El tamaño de muestra

Por requerimientos fijados por TV Azteca, respecto a la precisión de las estimaciones de porcentajes de votación, el número de secciones electorales consideradas tanto para la encuesta de salida como el conteo rápido se fijó en mil 280. De éstas, mil 34 corresponden a una muestra base para estimar los resultados de la elección presidencial, más 246 utilizadas para controlar la precisión de las estimaciones en las tres contiendas por gubernaturas. El cuadro siguiente ilustra la distribución.

El total de secciones consideradas en la muestra es de mil 280 con un total de lista nominal aproximado de un millón 456 mil ciudadanos. Considerando una participación ciudadana del 65 por ciento, a la salida de casillas se esperaba aplicar un promedio de 50 entrevistas por sección electoral; resultando entre 63 mil y 66 mil entrevistas totales en las mil 280 secciones electorales incluidas en la muestra. El último corte de la encuesta de salida realizado a las 18:00 horas reportó resultados de mil 235 secciones electorales con un total de 67 mil 43 entrevistas. Esto significó un promedio final aproximado de 54 entrevistas por sección electoral.

TAMAÑO Y DISTRIBUCIÓN DE MUESTRA

Estrato	Muestra base	Muestra adicional	Total
Distrito Federal	90	20	110
Guanajuato	49	111	160
Morelos	15	115	130
Resto estados	880	-	880
Total nacional	1034	246	1280

Resultados de la encuesta de salida

Con relación a las encuestas de salida, el trabajo de campo fue realizado conforme a lo previsto, fluyendo la información en promedio a razón de seis entrevistas por hora, por sección. A las 13:00 horas se tenía información del 87 por ciento de las mil 280 secciones electorales en muestra con un total de 26 mil 618 entrevistas. A partir de este momento, los resultados de las entrevistas permitieron concluir que la ventaja de Fox era significativa.

Covarrubias y Asociados decidió publicar los resultados de la encuesta de salida incluyendo en las estimaciones los casos de no información, mismos que tuvieron una importancia que osciló alrededor de los 12 puntos porcentuales. Esto se hizo atendiendo a recomendaciones internacionales de la European Society of Marketing Research (ESOMAR) que establece la importancia de considerar la publicación de los resultados de encuestas incluyendo las tasas de no respuesta.

La calidad de las estimaciones de voto en los diferentes cortes de información puede constatarse al compararse con los resultados oficiales definitivos del IFE, eliminando la no respuesta.

RESULTADOS DE LA ENCUESTA DE SALIDA
Datos directos

Candidato	Corte 1:00 pm	Corte 2:00 pm	Corte 4:00 pm	Corte 6:00 pm
Vicente Fox	39.7	39.4	39.5	38.8
Francisco Labastida	30.4	30.2	30.1	30.5
Cuahtémoc Cárdenas	15.0	15.4	15.3	15.7
Otros	3.0	2.9	3.1	3.3
No información	11.9	12.1	12.0	11.7
Total	100.0	100.0	100.0	100.0
% de muestra	87%	92%	95%	98%
Total de entrevistas	26,618	33,918	50,363	67,043

RESULTADOS DE LA ENCUESTA DE SALIDA
Datos ajustados comparados con resultados definitivos IFE

Candidato	Corte 1:00 pm	Corte 2:00 pm	Corte 4:00 pm	Corte 6:00 pm	Resultados definitivos IFE
Vicente Fox	45.1	44.8	44.9	43.9	43.5
Francisco Labastida	34.5	34.4	34.2	34.6	36.9
Cuahtémoc Cárdenas	17.0	17.5	17.4	17.8	17.0
Otros	3.4	3.3	3.5	3.7	2.6
Total	100.0	100.0	100.0	100.0	100.0
Error de estimación	3.1%	3.0%	2.8%	2.5%	

Resultados del conteo rápido

Con respecto al conteo rápido hicimos varios cortes de información, de los cuales se entregaron tres cortes a TV Azteca, a las 20:30, 21:20 y 22:50 horas. Por primera vez en México se publicaron los avances de un conteo rápido utilizando varios cortes de información en función del arribo de los datos.

En el corte de las 20:30 horas, con base en 120 secciones electorales, se calcularon las estimaciones con una desviación promedio de 1.4 puntos porcentuales. TV Azteca difundió estas estimaciones en la Entrevista con Sarmiento, incluyendo su actualización basada en las 232 secciones electorales correspondientes al corte de las 21:20 horas. El corte de las 22:50 horas fue el definitivo y fue entregado ante notario a TV Azteca.

RESULTADOS DEL CONTEO RÁPIDO
Versión corta

Candidato	Corte 8:30 pm	Corte 9:20 pm	Corte 10:50 pm	Resultados definitivos IFE
Vicente Fox	44.2	43.7	43.5	43.5
Francisco Labastida	34.2	35.2	35.5	36.9
Cauhtémoc Cárdenas	17.8	17.7	17.5	17.0
Otros	3.8	3.4	3.5	2.6
Total	100.0	100.0	100.0	100.0
Número de secciones	120	232	892	
Error de estimación	3.1%	2.3%	1.6%	

La conveniencia de diversos diseños de muestra

Las compañías que realizaron ejercicios de estimaciones de conteos rápidos en estas elecciones utilizaron una amplia gama de diseños de muestra. Al respecto, existe una serie de cuestiones por resolver sobre las conveniencias o desventajas de uno u otro método. Por ejemplo, ¿Es PPT un diseño conveniente de muestra para un conteo rápido?, ¿Qué tan convenientes

son los muestreos aleatorios simples en este contexto?, ¿Qué ventajas tienen los muestreos sistemáticos?, ¿Cuáles son las ventajas de diferentes ordenamientos de secciones para muestrear sistemáticamente las secciones?, etc.

Covarrubias y Asociados elaboró una serie de simulaciones a fin de comparar la conveniencia de algunos de los principales métodos de selección:

- Comparación PPT vs AS
- PPT vs PPT sistemático por lista nominal
- PPT vs PPT sistemático por voto PRI
- PPT vs PPT sistemático por porcentaje PRI
- PPT vs PPT sistemático por índices de competencia:
- Votos primer lugar sobre votos segundo (IC12)
- Votos primer lugar sobre votos segundo más tercero (IC123)

Diseño y resultados de la simulación

Para las simulaciones se utilizó la base de datos de la elección presidencial del PREP del IFE, que fue prestada a Covarrubias y Asociados por el IFE para elaborar esta parte de la presentación.

En cada simulación asociada a un particular tipo de muestreo se generaron dos mil muestras para cada uno de los cinco tamaños de muestra investigados: $n = 100, 200, 300, 500, 1000$ y 2000 .

En la selección de muestras se consideró en todos los casos un muestreo estratificado por tipo de sección electoral: urbano, mixto y rural. Esto se debe a que típicamente las estimaciones de voto en estos estratos presentan fuertes diferencias que permiten reducir la variabilidad de las estimaciones. Con base en los datos del PREP, los porcentajes de voto por tipo de sección se muestran en la siguiente tabla, donde se observa que en secciones urbanas el voto por Fox es de 49 por ciento, disminuyendo hasta prácticamente la mitad (25 por ciento) en secciones rurales. El voto por Labastida presenta una fuerte tendencia a la inversa.

RESULTADOS DE VOTACIÓN POR TIPO DE SECCIÓN

	AC	PRI	AM	PCD	PARM	DS	No. reg.	Total
Urbana	49.23	32.34	15.44	0.55	0.37	2.01	0.07	100
Mixta	34.75	43.67	19.54	0.56	0.48	0.85	0.15	100
Rural	24.79	51.53	21.91	0.62	0.64	0.39	0.13	100
Total general	43.43	36.89	17.00	0.56	0.43	1.61	0.09	100

Para efectos de comparación de la precisión de los métodos de selección de muestras, se calculó la desviación media y el error cuadrático medio de los porcentajes de votación asociados a cada categoría. Estas medidas se obtuvieron utilizando las estimaciones de voto reportadas en siete categorías asociadas a los seis partidos registrados en la contienda: AC, PRI, AM, PCD, PARM, DS y el voto por candidatos no registrados.

Denotando las proporciones de voto obtenidas del PREP por:

$$P_1, P_2, \dots, P_7,$$

y las obtenidas con base en una muestra aleatoria de secciones por:

$$\hat{P}_1, \hat{P}_2, \dots, \hat{P}_7,$$

se tiene que la desviación absoluta o media y el error cuadrático medio del *i*-ésimo estimador están dados por:

$$DM_j = E|P_j - \hat{P}_j| \quad \text{y} \quad ECM_j = E(P_j - \hat{P}_j)^2 \quad j = 1, 2, \dots, 7.$$

Estos parámetros pueden ser fácilmente estimados de manera precisa utilizando dos mil réplicas de muestra con estimaciones de voto denotadas por:

$$P_{i1}, P_{i2}, \dots, P_{i7}, \quad i = 1, 2, \dots, 2000,$$

donde los estimadores de la desviación media y error cuadrático medio del *j*-ésimo porcentaje de voto están dados por:

$$\widehat{DM}_j = \frac{1}{2000} \sum_{i=1}^{2000} |P_{ij} - \hat{P}_{ij}| \quad \widehat{ECM}_j = \frac{1}{2000} \sum_{i=1}^{2000} (P_{ij} - \hat{P}_{ij})^2, \\ j = 1, 2, \dots, 7.$$

Con base en estas estimaciones se estimó la desviación media y error cuadrático medio máximos sobre las siete estimaciones de porcentajes como:

$$\widehat{DM} = \max \{ \widehat{DM}_1, \widehat{DM}_2, \dots, \widehat{DM}_7 \}$$

y

$$\widehat{ECM} = \max \{ \widehat{ECM}_1, \widehat{ECM}_2, \dots, \widehat{ECM}_7 \}$$

Para efectos de comparación gráfica de los métodos de selección, se calculó la raíz del error cuadrático medio a fin de tener la medida de dispersión en las mismas unidades que el estimador. En las gráficas, las estimaciones se denotan como sigue:

- AS: Muestreo aleatorio simple sin reemplazo.
- PPT: Muestreo aleatorio PPT sin reemplazo respecto a lista nominal.
- PPT-LN: Muestreo aleatorio sistemático PPT respecto a lista nominal, ordenando las secciones decrecientemente respecto a lista nominal.
- PPT-VPRI: Muestreo aleatorio sistemático PPT respecto a lista nominal, ordenando las secciones decrecientemente respecto al volumen del voto priísta.
- PPT-PPRI: Muestreo aleatorio sistemático PPT respecto a lista nominal, ordenando las secciones decrecientemente respecto al porcentaje del voto priísta.
- PPT-IC12: Muestreo aleatorio sistemático PPT respecto a lista nominal, ordenando las secciones decrecientemente respecto al valor del índice de competitividad dado por el cociente de votos primer lugar entre votos segundo lugar.
- PPT-IC123: Muestreo aleatorio sistemático PPT respecto a lista nominal, ordenando las secciones decrecientemente respecto al valor del índice de competitividad dado por el cociente de votos primer lugar entre votos segundo lugar más votos tercer lugar.

En las Gráficas 1 y 2 se presentan los resultados de la simulación de los seis métodos de extracción de muestra en el caso de la desviación media y raíz del error cuadrático medio, respectivamente. En primer plano se observa que las diferencias en error son, en general, inferiores a 0.5 por ciento, disminuyendo las diferencias conforme aumenta el tamaño de muestra. Se observa que el peor método de selección (el que produce mayores niveles de error) es el aleatorio simple, en tanto que el mejor es el PPT-IC123. Para tamaños de muestra inferiores a 500 secciones, el muestreo PPT tiene una ligera ventaja sobre el aleatorio simple. No existe diferencia entre el muestreo PPT y el PPT-LN. Esto es, ordenar las secciones no representa una ganancia.

Las Gráficas 3 y 4 muestran los niveles de error que se generan al extraer muestras aleatorias simples de secciones electorales y elaborar las estimaciones con base en el diseño aleatorio simple y con base en estimadores correspondientes a muestreos PPT. Ambos gráficos señalan que el ignorar que la muestra fue diseñada aleatoria simple y utilizar los cálculos asociados a estimadores PPT produce un incremento muy considerable en la variabilidad de las estimaciones. Más aún, en el caso de estimadores PPT el error no decrece a cero, lo que implica que aún con tamaños de mues-

tra enormes como es el de cinco mil secciones electorales, se esperan altos márgenes de error.

Por último, las Gráficas 5 y 6 muestran el comportamiento de las estimaciones aleatorias simples (datos crudos) y estimaciones PPT cuando la muestra fue diseñada PPT. Se observa que en este caso, el ignorar que el diseño de muestra es PPT para utilizar estimaciones aleatorias simples no produce incrementos considerables en los márgenes de error. Sin embargo, resulta claro que es conveniente utilizar el estimador PPT acorde al diseño de muestra.

En síntesis, y como resultado de esta simulación, pueden establecerse las siguientes observaciones generales para ejercicios futuros de esta naturaleza:

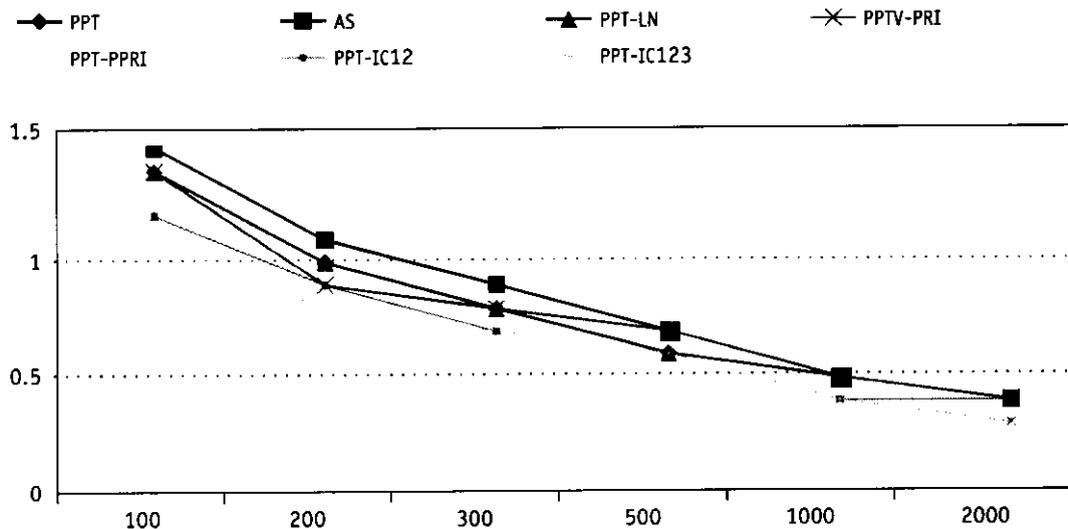
- ✓ El mejor método de selección consiste en ordenar decrecientemente las secciones electorales utilizando el índice de competencia IC123 para, posteriormente, obtener una muestra sistemática PPT respecto a la

lista nominal. Este método, sin embargo, no produce ganancias considerables en comparación al muestreo PPT sin reemplazo respecto a la lista nominal. Además, para implementarlo se requiere de una base de datos histórica con resultados recientes de votación, cosa que no siempre es posible tener.

- ✓ La estimación PPT es ligeramente superior que la aleatoria simple.
- ✓ El estudio nos viene a recordar lo fundamental de la congruencia entre diseño de muestra y ponderación, particularmente si las muestras se diseñan con un esquema aleatorio simple, donde se observó un considerable aumento en los errores al ponderar los datos mediante un muestreo PPT.
- ✓ La selección PPT tiene la ventaja de permitir estimaciones crudas (no ponderadas) con razonable precisión y facilitar la labor de campo debido a que la muestra tiende a tener secciones más grandes que, en general, son más accesibles.

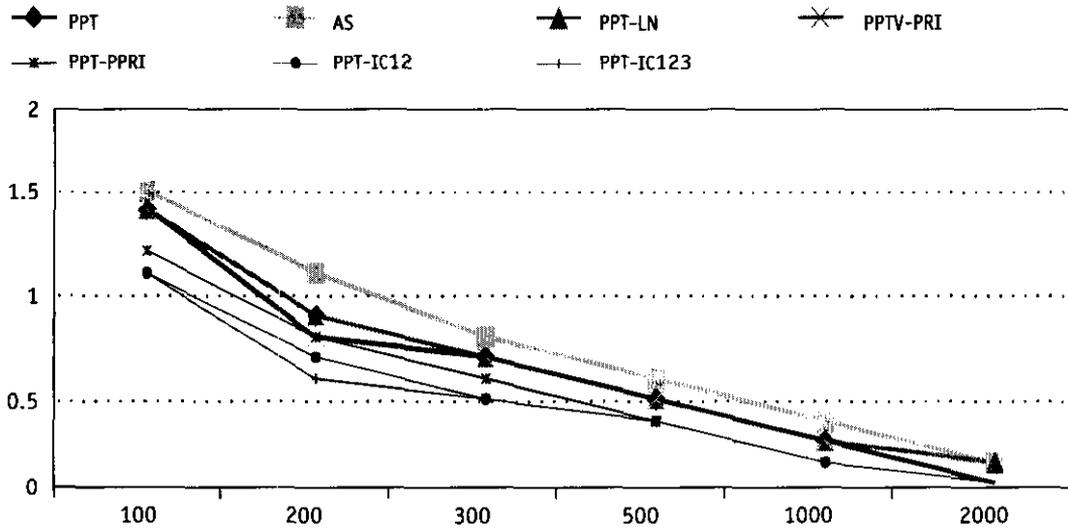
GRÁFICA 1
PPT VS. ALEATORIO SIMPLE

Desviación media



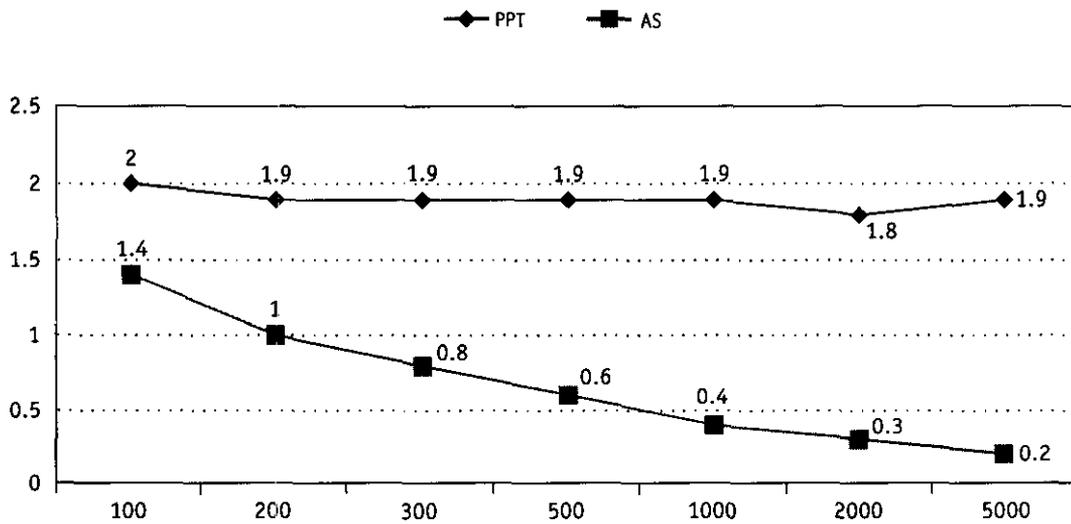
GRÁFICA 2
PPT VS. ALEATORIO SIMPLE

Raíz del error cuadrático medio



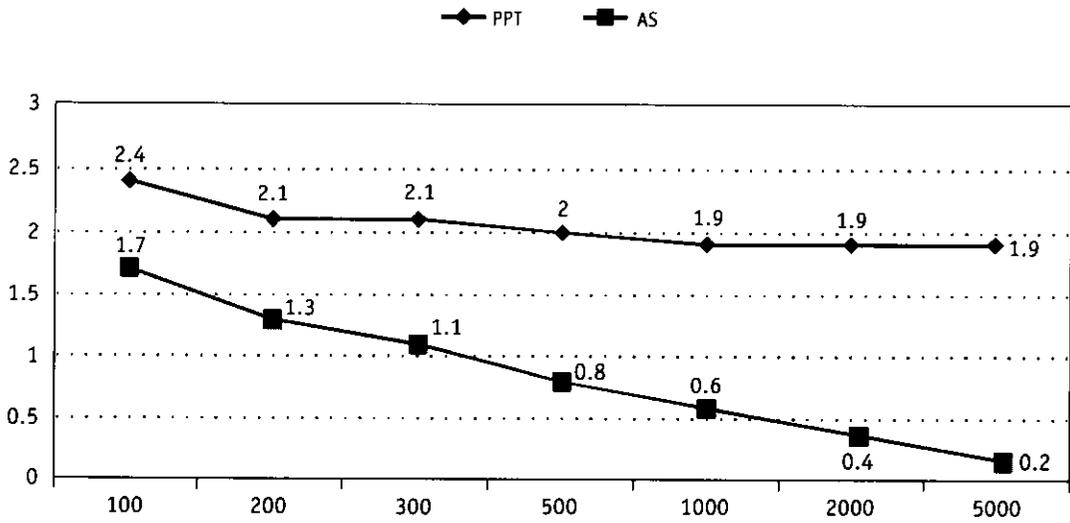
GRÁFICA 3
DISEÑO DE MUESTRA: ALEATORIA SIMPLE

Desviación media



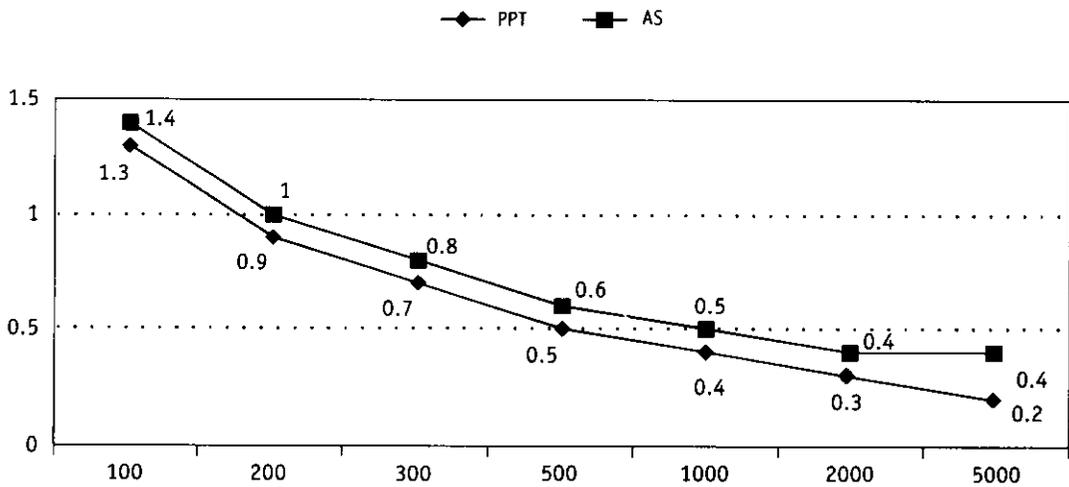
GRÁFICA 4
DISEÑO DE MUESTRA: ALEATORIO SIMPLE

Raíz del error cuadrático medio



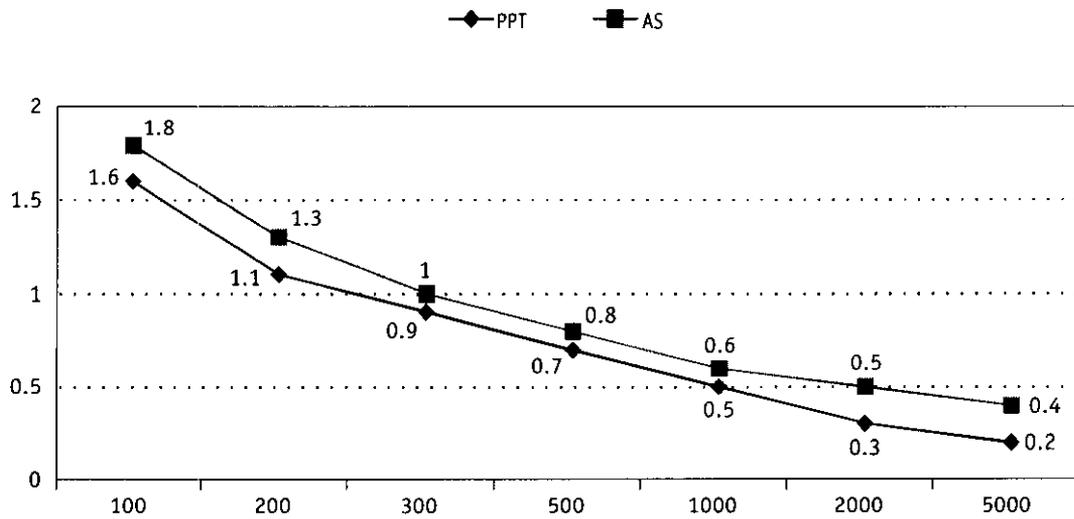
GRÁFICA 5
DISEÑO DE MUESTRA: PPT POR LISTA NOMINAL

Desviación media



GRÁFICA 6
DISEÑO DE MUESTRA: PPT POR LISTA NOMINAL

Raíz del error cuadrático medio



Recomendaciones finales

- ✓ *Uso de datos históricos.* Es recomendable la utilización de datos históricos para mejorar la precisión de las estimaciones, particularmente en el caso de contar con escasa información; esto es, cuando aún no ha llegado suficiente información de las secciones electorales a fin de elaborar estimaciones de razonable precisión.
- ✓ *Uso de métodos mixtos de estimación.* En el caso de los conteos rápidos cabe estudiar la conveniencia de utilizar métodos mixtos de estimación que en cada corte de información utilicen los datos de las secciones con que cuente y, en su ausencia, la información de las encuestas de salida. Esto permitiría mejorar considerablemente la calidad de las estimaciones, particularmente durante los minutos inmediatos posteriores al cierre oficial de casillas.
- ✓ *Formatos de publicación de resultados.* A fin de facilitar la comparación de los resultados que publican las diferentes empresas que participan en este tipo de ejercicios, es conveniente estandarizar los formatos de publicación de estimaciones. En conteos de salida se propone la publicación de resultados eliminando la categoría de no respuesta, pero incluyendo por separado este importante indicador, ya que dependiendo de su tamaño podrían alterarse, incluso, el orden de los candidatos. Recordemos que no necesariamente las preferencias de los respondientes son iguales a las de los no respondientes,

por lo que la simple eliminación de los no respondientes de la muestra puede crear sesgos considerables en las estimaciones e incluso en las pruebas de significancia.

- ✓ *Orientar a comunicadores.* Es también importante señalar la necesidad de orientar a los comunicadores en el trabajo de difusión de los resultados de votación. Se debe tener en cuenta que sólo debe compararse lo comparable. No deben mezclarse resultados de encuestas de salida con los de conteos rápidos, ya que estos métodos tienen fuertes diferencias en precisión.
- ✓ *Orientar a analistas.* Cuando se publiquen cuadros comparativos a fin de evaluar la calidad de las estimaciones obtenidas por varias empresas es fundamental determinar los aspectos que pueden ser comparables. No comparar estimaciones que tienen diferentes escalas o categorizaciones; no mezclar resultados de encuestas de salida con los de conteos rápidos y en cuanto estén disponibles los resultados definitivos de los organismos oficiales, las comparaciones deben realizarse respecto a esta información y no respecto a resultados de programas de estimaciones preliminares. Las muestras se diseñan para estimar resultados finales y no los de estimaciones preliminares.
- ✓ *Constancia de resultados generados.* Por último, se señala la conveniencia de generalizar la práctica de notariar las estimaciones a fin de establecer bases más objetivas en la evaluación de los trabajos realizados. ■