

MESA 5

---

# CONTEOS RÁPIDOS Y ENCUESTAS DE SALIDA-MEDIOS

⋮

## **RELATORÍA**

Juan Manuel Herrero

## **PONENCIAS**

José Mendoza

Roy Campos

Alejandro Moreno

Francisco Báez

Javier Alagón

Enrique Calderón Alzati

# RELATORÍA

.....  
Juan Manuel Herrero A.

**E**n esta mesa se presentaron seis trabajos, dos de los cuales fueron financiados por medios electrónicos, otros dos por medios impresos y los restantes sufragados por una organización empresarial y por una organización social.

La primera participación la hizo José Mendoza con comentarios de Ana Cristina Covarrubias, de Covarrubias/Sofres, trabajo solicitado por TV Azteca.

En este ejercicio se comprometió no solamente un conteo rápido, sino también una encuesta de salida como valor agregado al cliente. Los resultados de la encuesta fueron señalados para las 18:00 horas y los del conteo a las 20:00 horas.

Como elemento adicional se contó con los servicios de un notario público para dar fe de los resultados.

El cliente solicitó un alto nivel de precisión (1.1 por ciento). La muestra base fue de mil 34 secciones y se utilizó el método PPT a nivel estado.

Para el cálculo de los ponderadores se consideraron elementos clásicos (atado al diseño de la muestra) y experiencia electoral previa (la elección federal de 1997).

Las entrevistas fueron con urna. En la encuesta de salida se preguntó por edad, sexo, escolaridad, ocupación y momento en que se tomó la decisión sobre el voto. Sí se tomó en consideración la no información que fue del orden del 11 por ciento. Se realizaron 50 entrevistas por sección electoral.

Para validar que el muestreo PPT es el método idóneo para este tipo de ejercicios, se llevó a cabo una simulación con el PREP de 1997 (Programa de Resultados Electorales Preliminares de la elección federal de 1997).

La presentación concluyó dando cinco recomendaciones:

- Utilidad de usar datos históricos
- Conveniencia de métodos mixtos de estimación
- Estandarizar los formatos de publicación
- Necesidad de orientar a los comunicadores
- Generalizar la práctica de dar las estimaciones ante Notario Público.

La segunda intervención estuvo a cargo de Roy Campos, de Consulta-Mitofsky, quien trabajó para Televisa.

Manifestó que "las prioridades, objetivos y forma de trabajo cambian cuando los conteos rápidos son para los medios electrónicos". Que duda que los conteos rápidos tengan una larga vida y que su cliente no aceptaría resultados de encuestas ante Notario.

Comentó las diferencias en objetivos entre una encuesta de salida y un conteo rápido.

Mientras que para la encuesta la certidumbre y la oportunidad son vitales, para un conteo es no equivocar ganador y no emitir resultados después de la autoridad, utilizar el menor número de recursos a cambio de eficiencia y precisión.

Para el diseño de la muestra consideró información electoral de 1991, 1994 y 1997. El tamaño fue de 450 secciones electorales para la encuesta de salida y 630 para el conteo. Levantó al menos 50 cuestionarios por sección en la encuesta.

Dividió el país en cinco regiones; creó estratos dentro de regiones con base en la división urbano-mixto-rural y tamaño de padrón por cortes.

La selección fue también con muestreo PPT a partir de la lista nominal. Se crearon factores de expansión por sección electoral y la no respuesta se corrigió dentro de cada estrato, donde los pesos de cada estrato se construyeron como el promedio del peso del padrón electoral actual y el de la elección de 1997. No presenta decimales en sus resultados.

Alejandro Moreno, del periódico Reforma, cuestionó la utilidad de llevar a cabo encuestas de salida para medios impresos. Sin embargo, Reforma empezó a presentar datos el mismo 2 de julio en Internet.

El tamaño de su muestra fue más pequeño (150 secciones electorales) lo que a lo largo de las distintas mesas pareció generar consenso de que en el futuro no se requerirán muestras de las dimensiones que tuvimos en este proceso.

Su levantamiento fue en tres etapas, encuestando a un máximo de 25 ciudadanos por casilla. Al final, sus resultados mostraron un error promedio de uno por ciento, siendo muy precisos.

Entre los comentarios de interés en el análisis que se hizo de esta información encontramos que de los ciudadanos que en 1994 votaron por Zedillo, ahora, en 2000, el 25 por ciento votó por Fox y que aquellos que en la interna del PRI votaron por Madrazo, el 53 por ciento votó por Fox en el 2000.

Francisco Báez, de La Crónica de Hoy, se refirió a su experiencia personal histórica en materia de ejercicios electorales.

Refirió que en 1982 ya hacía conteos rápidos para el PSUM en poco más de mil casillas. Que en 1985 realizó un conteo a nivel nacional con tan solo 300 casillas. Que en 1988 y para el semanario Punto levantó 80 casillas en el D.F. y que en 1994, trabajando para un consejero del IFE, un ejercicio de otras 80 casillas.

Ahora en el 2000 y para el periódico La Crónica de Hoy, llevó a cabo una interesante simulación con datos del PREP del IFE.

Diseñó dos muestras de 52 casillas cada una con muestreo aleatorio sistemático. Cada sección tenía una sección alterna del mismo municipio y distrito.

Se toma la primera que llegue al PREP y se van cotejando los resultados de las dos muestras, dándose a conocer los resultados si las diferencias entre ambas muestras es menor a 1.5 por ciento.

Otra restricción para dar resultados es que el PREP lleve, cuando menos, el 15 por ciento de la información total.

En este ejercicio se pondera el bajísimo costo, ya que no requiere trabajo de campo.

Javier Alagón trabajó para la Coparmex y presentó los principales elementos del diseño de su ejercicio y los resultados, pero no quiso desaprovechar la ocasión para hacer algunas reflexiones de carácter general en cuanto al futuro de estos trabajos.

Considera que es fácil hacer un conteo rápido "si nos dan tiempo".

Alagón tomó 600 secciones (mil 315 casillas), estratificó resultados de acuerdo a datos de elecciones pasadas.

Sus estratos se construyeron con base en la preferencia partidista de las secciones electorales, creando siete estratos, a saber:

- Muy priístas
- Priístas
- Panistas
- Muy perredistas
- Perredistas
- Competidas
- Secciones nuevas

A las 10:15 de la noche tenía la información del 74 por ciento de las casillas y resultados muy cercanos a los oficiales.

En cuanto a las reflexiones, manifiesta que él es amigo de los enfoques ortodoxos, que de esta serie de reuniones quedan asuntos pendientes en la agenda entre los que destacan:

- ¿Cuándo se cruzaron las preferencias electorales?
- ¿A qué se deben las diferencias entre encuestas? Sólo encontrando las causas se podrá aportar información "creíble" a la opinión pública.
- Dilucidar las diferencias por:
- Diseño
- Campo
- Filtros/ Votantes posibles
- Posición en el cuestionario de la pregunta sobre el voto.

Propone, finalmente, juntar y analizar la información disponible y pregunta: ¿Estaremos abiertos?

La última participación de la mesa estuvo a cargo de Enrique Calderón, de la Fundación Rosenblueth, quien desarrolló una encuesta de salida para medir los resultados correspondientes al Senado, lo que implicó diseñar una muestra que permitiera tener cifras por entidad federativa.

El objeto se definió como “usando información electoral previa, obtener resultados confiables con muestras pequeñas”.

Se estratificó por tamaño de las localidades, calificadas como rurales o urbanas. También por escolaridad.

El país se dividió en 136 regiones y se generaron tres ponderadores:

- P1 = Regional. Proporción por región.
- P2 = Municipal. Representando grupos sociales.
- P3 = Seccional. Niveles de preferencia electoral.

El ponderador se construyó multiplicando los tres definidos.

P1 x P2 x P3.

La muestra se integró por 433 municipios y 783 secciones electorales.

El sesgo por estimación fue del 1.2 por ciento respecto a 1997.

Enrique Calderón comentó que las encuestas de salida miden la votación “no vergonzante”, pues aún en este método hay quienes no quieren contestar y que su voto puede ser distinto al de aquellos que sí contestan estas encuestas. ■

## CONCLUSIONES

*Los expositores estuvieron de acuerdo en lo fundamental de los elementos técnicos que deben ser considerados en los conteos rápidos y en las encuestas de salida. Sin embargo, no entraron mucho en tratar de responder a los cuestionamientos que planteó Javier Alagón y que son parte medular del seminario. Tampoco se extendieron en las particularidades de los conteos contratados por medios vis a vis los que encargan las fuerzas políticas.*

# 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

**E**l 12 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		2. Sexo	
1. 18-25	3. 36-45	1. Masculino	2. Femenino
2. 26-35	4. 46 y más		
3. Escolaridad. ¿hasta qué año estudió? (Si no es completa se baja hasta la categoría anterior)		4. Ocupación principal ( )	
1. Sin escolaridad/primaria incompleta		1. Ama de casa	
2. Primaria completa		2. Estudiante	
3. Secundaria completa (técnica equivalente)		3. Trabajador del campo	
4. Preparatoria completa (técnica equivalente)		4. Trabaja por su cuenta	
5. Profesional completa		5. Empleado de Gobierno	
6. Posgrado		6. Empleado empresa privada	
		7. Dueño de su negocio	
		8. Profesionista 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 de la lista la frase que vaya más de acuerdo con su manera de pensar)		6. ¿En qué momento tomó la decisión de votar por quién votó? (Leer o mostrar alternativas)	
1. Por el mejor partido		1. Hoy mismo	
2. Por la esperanza de tener una mejoría económica		2. En la última semana	
3. Por el mejor candidato		3. En el último mes, después del 2do. debate	
4. Por sacar al PRI de la Presidencia		4. Hace dos meses, después del 1er. debate	
5. Por la estabilidad económica		5. Hace más de dos meses	
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			

ELECCIÓN PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS	
<b>Alianza por el Cambio</b>  Vicente Fox Quesada ①	<b>Partido Revolucionario Institucional</b>  Labastida Ochoa ②
<b>Alianza por México</b>  Cusúbertméc Cárdenas Solórzano ③	<b>P. tid. de Centro Democrático</b>  Manuel Camacho Solís ④
<b>Partido Auténtico de la Revolución Mexicana</b>  Porfirio Muñoz Ledo ⑤	<b>Democracia Social Partido Político Nacional</b>  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 constatare 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
Cuauhtémoc 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
Cuauhtémoc 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
Cuauhtémoc 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:

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

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:

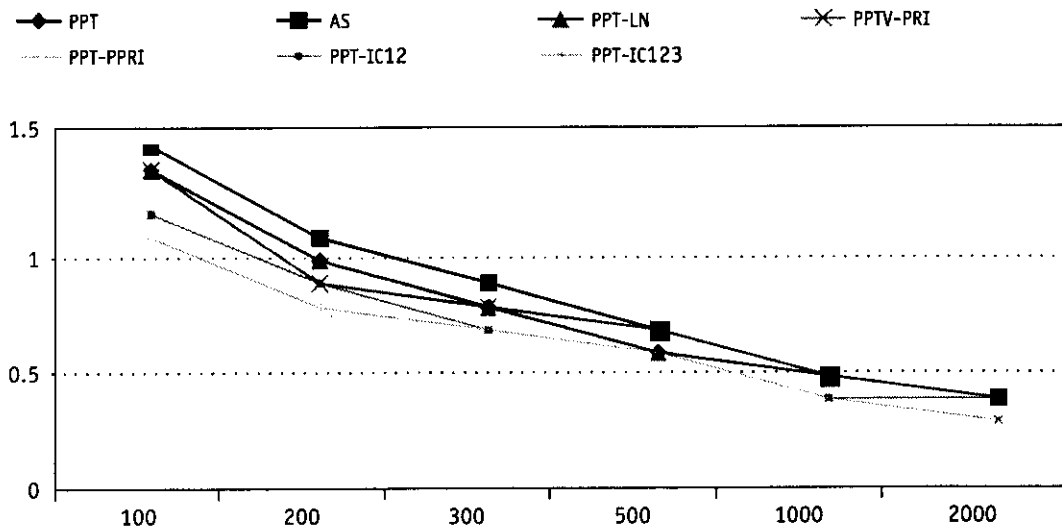
- ✓ El mejor método de selección consiste en ordenar decrecientemente las secciones electorales utilizando el índice de competencia IC123 para, posteriormente, obtener una muestra sistemática PPT respecto a la

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.

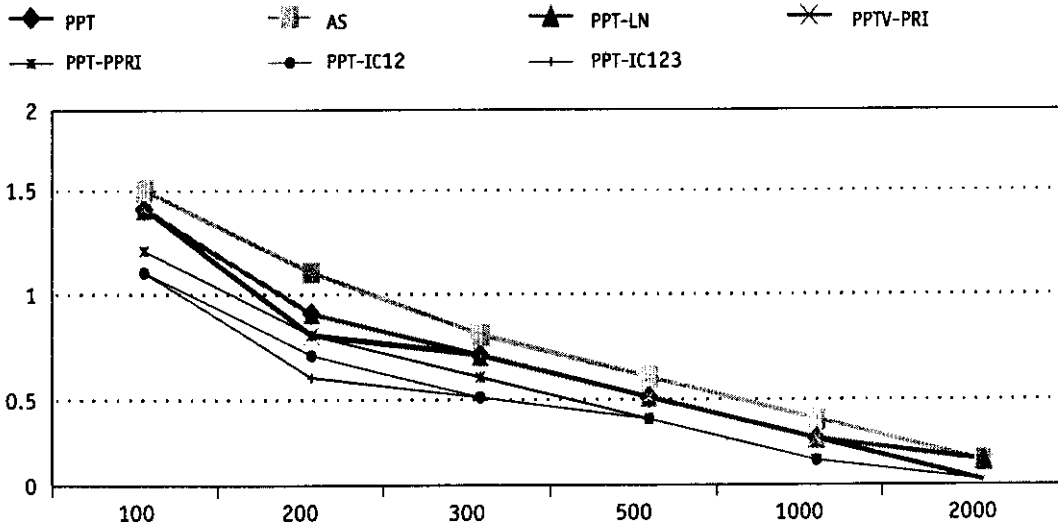
**GRÁFICA 1**  
PPT VS. ALEATORIO SIMPLE

*Desviación media*



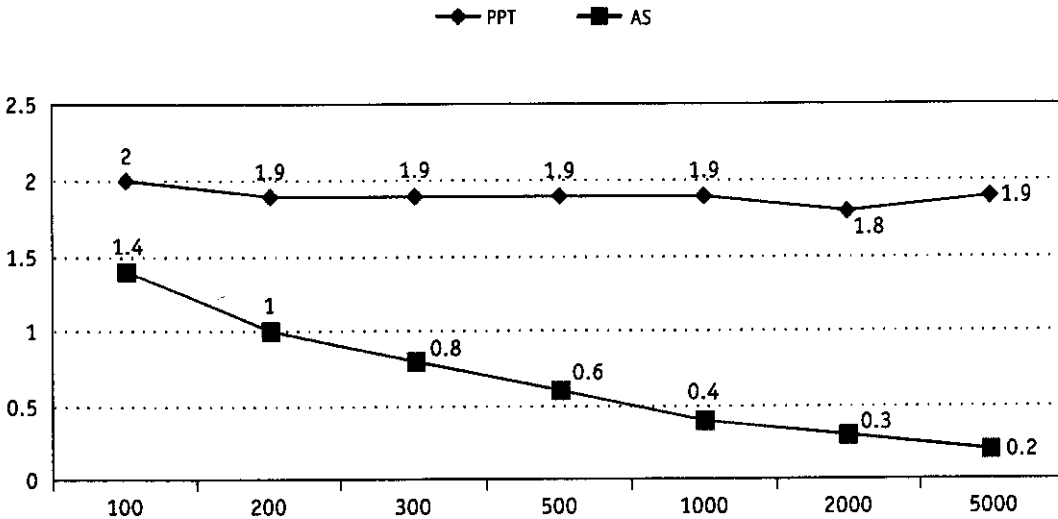
**GRÁFICA 2**  
PPT VS. ALEATORIO SIMPLE

*Raíz del error cuadrático medio*



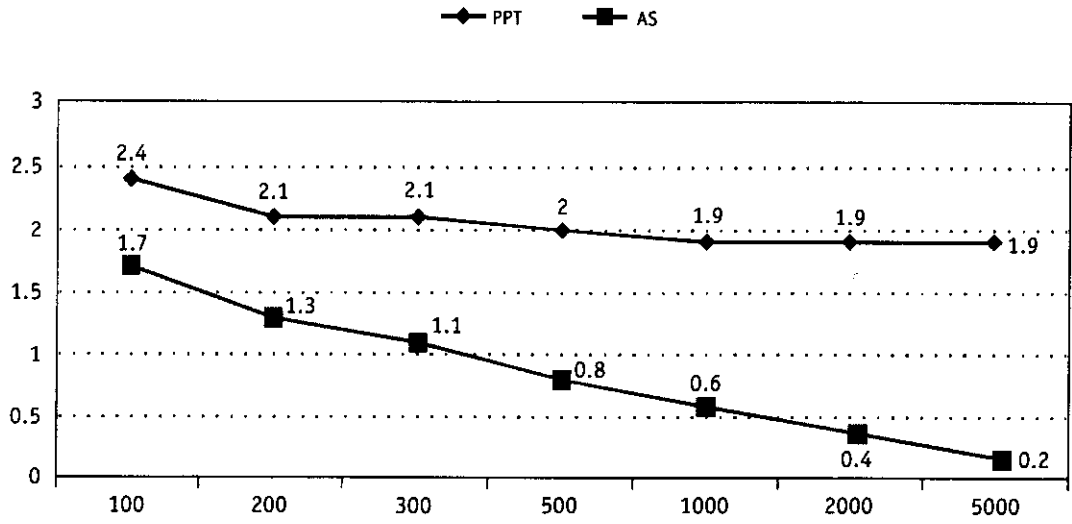
**GRÁFICA 3**  
DISEÑO DE MUESTRA: ALEATORIA SIMPLE

*Desviación media*



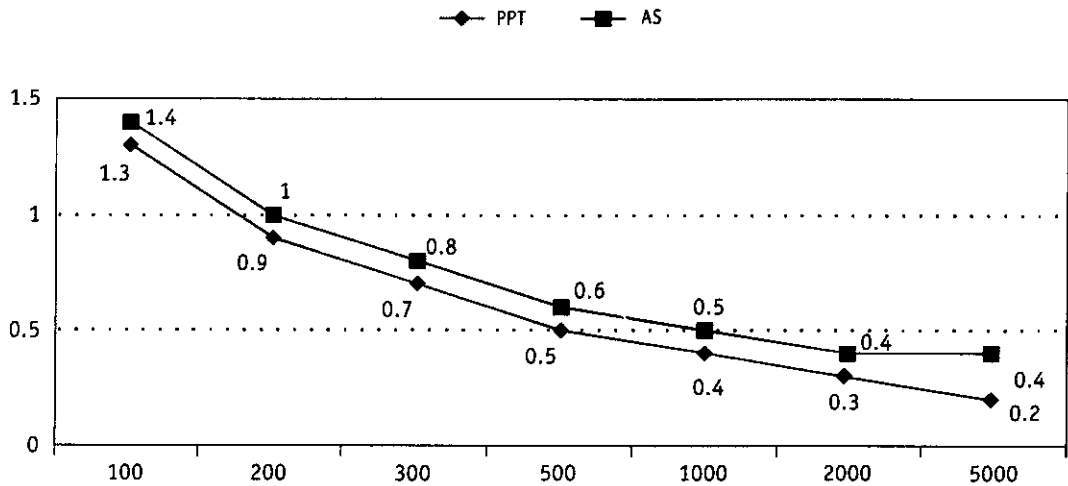
**GRÁFICA 4**  
DISEÑO DE MUESTRA: ALEATORIO SIMPLE

*Raíz del error cuadrático medio*



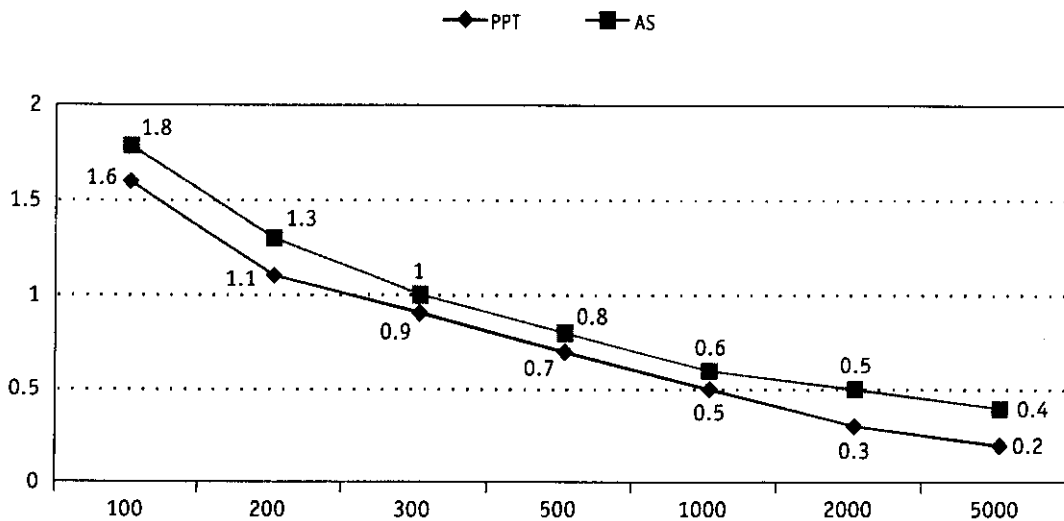
**GRÁFICA 5**  
DISEÑO DE MUESTRA: PPT POR LISTA NOMINAL

*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. ■

# PONENCIAS MESA 5

.....

Roy Campos

CONSULTA-MITOFSKY PARA TELEvisa

---

## ENCUESTA DE SALIDA Y CONTEO RÁPIDO PARA TELEvisa EN LAS ELECCIONES DEL 2 DE JULIO DE 2000

### OBJETIVOS

#### ENCUESTA A LA SALIDA DE LAS CASILLAS

1. **Certidumbre:** No equivocarse en el candidato o partido ganador
2. **Oportunidad:** Nunca después de la autoridad
3. **Análisis:** Proporcionar elementos sólidos para el entendimiento del resultado electoral
4. **Eficiencia:** Utilizar la menor cantidad de recursos para cumplir los objetivos 1, 2 y 3
5. **Precisión:** Sólo es importante como anécdota, no sirve de nada si no se cumplen los objetivos previos

#### CONTEO RÁPIDO

1. **Certidumbre:** No equivocarse en el candidato o partido ganador
2. **Oportunidad:** Nunca después de la autoridad
3. **Eficiencia:** Utilizar la menor cantidad de recursos para cumplir los objetivos 1 y 2
4. **Precisión:** Sólo es importante como anécdota, no sirve de nada si no se cumplen los objetivos 1, 2 y 3

### ASPECTOS GENERALES

**Resultados a generar:** Ganadores en las elecciones federales para Presidente de la República, para jefe de gobierno del Distrito Federal y para gobernadores de Guanajuato y Morelos

**Marco muestral:** Secciones electorales definidas por el IFE, y para las cuales se cuenta con información sobre su votación anterior en 1991, 1994 y 1997

## TAMAÑO DE MUESTRA

## DISEÑO

<i>Dominio</i>	<i>Encuesta de salida</i>	<i>Conteo rápido</i>
Guanajuato	40	70
Morelos	70	100
Distrito Federal	70	100
Resto del país	270	360
Total	450	630

*Estratificación explícita:*

- 8 regiones geográficas
- De 4 a 6 estratos por comportamiento político en cada región (al final fueron 39 estratos)

*Estratificación implícita:* En cada estrato se ordenaron las secciones por clasificación urbano-mixto-rural y por tamaño de padrón en cada corte.

## BASE DE CÁLCULO:

- A) Estimación del RHO en 1991, 1994 y 1997 para PAN, PRI y PRD, considerando la sección como conglomerado tomamos el mayor de los 9 valores.
- B) Para 99% de confianza y 2% de error máximo se calcularon los tamaños de muestra, considerando una participación promedio de 65% . (Y 50 respuestas efectivas por sección en la encuesta de salida)

*Afijación de la muestra:* Tres regiones (D.F., Morelos y Guanajuato) ya estaban definidas. En el resto la quinta parte en cada región.

*Selección:* Probabilidad proporcional al padrón y después submuestra con igual probabilidad para la encuesta de salida.

*Estimación:* Tradicional, utilizando factores de expansión por sección electoral.

*No respuesta:* Se corrigió dentro de cada estrato.

*Pesos de los estratos:* Se promedió el peso que daba su padrón electoral y el peso en la elección federal de 1997.

# PONENCIAS MESA 5

.....

Alejandro Moreno

DIARIO REFORMA

**L**a experiencia del grupo Reforma con conteos y con estadísticas el día de la elección, según entiendo, data de principios o mediados de los ochenta cuando contaban a los votantes, literalmente, con contadores. De ese tiempo para acá los ejercicios se han sofisticado más, se han hecho más complejos, quizás más útiles e indudablemente mucho más rigurosos, metodológicamente hablando.

Para un medio impreso, sin embargo, en una época en la que existen, no solamente como ya lo acabamos de escuchar en las dos ponencias previas, medios de comunicación electrónica, en especial televisoras que hacen este tipo de encuestas de salida (*exit polls*) para poder decir la noche de la elección quién ganó. Pues pareciera y, sobre todo más importante, no solamente las televisoras, sino que la autoridad electoral misma sea capaz de decir la noche de la elección quién ganó y que, hoy en día, en estos nuevos tiempos no solamente la autoridad electoral sino los candidatos y los gobernantes nos digan quién ganó; reconozcan victorias y derrotas, para un periódico que sale al día siguiente con un dato de una encuesta de salida (*exit poll*), es una historia demasiado conocida por todos ya.

Entonces, cuál es el sentido de la encuesta de salida para un medio impreso en estos días y que, seguramente, se dificultará más en años por venir sobre todo porque creemos, o al menos yo soy uno de los que creen, que esto no va a quedar ahí, vamos a saber mucho más o vamos a manejar todavía más información en procesos electorales próximos.

Yo haría tres justificaciones. La primera tiene que ver, y esto es algo que no se comentó en las sesiones de ayer sobre encuestas preelectorales, que el dato puro, en sí, es noticia, sobre todo cuando es una encuesta electoral y, sobre todo, cuando como bien decía Roy Campos, se dice en la noche de la elección quién ganó, quién es el ganador, eso es noticia en sí mismo.

La noticia no termina ahí, la noticia va también en las historias del por qué en lo que los periodistas norteamericanos y quizá en muchos otros países llaman la historia de las noticias (*news history*), cuál es la historia; por qué ganó tal candidato; quién lo apoyó; para qué propósito; cuál es el mandato y, de ahí que, tanto las encuestas preelectorales, las cuales se pueden trabajar con mucho más detalle, como dije, no solamente el dato sino los porqués. Tratar las relaciones entre las variables que, finalmente, juegan en un resultado u otro, es importante. Hay otros más, en el caso del periódico, una encuesta de salida (*exit poll*) o los resultados que van llegando a lo largo del día a través de una encuesta de salida, permiten también, en el sentido político, a partidos y candidatos, asunto que ya se tomó, permiten tomar decisiones de cobertura, qué cubrir, qué no



cubrir. Los encuestadores sirven también, de alguna manera, como enlace con el campo, como enlace de lo que está sucediendo el día de la jornada electoral, no al dejar de hacer su trabajo o al dejar de recopilar la información que están finalmente recopilando ese día a través de cuestionarios, pero sí a través de una serie de quejas que se reportan como parte del mismo ejercicio de la encuesta de salida.

Como parte de esto, quizás una finalidad más en el caso específico de Reforma como medio impreso, se ha entrado también en el mundo de Internet y ahí a partir, quizá me equivoque, pero hasta donde recuerdo es en 1999, en la elección interna del PRI, cuando por primera vez se suben los datos de una encuesta de salida a la Internet del periódico. De hecho, en la elección interna del PRI, dado que también es una competencia interna de medios, nos tocó ser los primeros en subir un dato de cómo venía una encuesta de salida (*exit poll*) a Internet. En el caso del 2 de julio no teníamos mucha prisa por hacerlo, teníamos cuidado de salir con tiempo y de hecho el dato que se sube finalmente a Internet ya después de las televisoras, es un dato que como muchos otros, es un dato incompleto. No tenía todas las entrevistas ya reportadas a la base de datos y, finalmente, lo que se hace al día siguiente con el periódico, esto nos fuerza a defender de alguna manera el dato y hacer el análisis sobre ese dato.

Las características metodológicas de la encuesta son muy parecidas a las de las encuestas preelectorales, mismos tamaños, no de muestra en este caso, pero sí de secciones electorales seleccionadas, 150 en el caso de la encuesta de salida (*exit poll*) que hicimos. Creo que alrededor de 100, en el caso del Distrito Federal y mucho menos puntos en el caso de encuestas de salida que se hicieron a nivel municipal en el estado de Nuevo León y coordinado, sobre todo, por el periódico El Norte.

Yo creo que no hay mucho más que añadir, las metodologías son, con márgenes de error si me lo permiten, muy parecidas a lo que hacen otros medios, me parece que no hay nada que añadir en ese sentido. Seleccionando secciones electorales mediante criterios urbano, mixto, rural; ordenadas también con base en distintos criterios, en esto un criterio muy modesto que nos ha resultado, no ha traído graves fallas. Quizá con los estudios de simulación de Roy Campos pudiéramos refinar nuestras estimaciones un poco más en algunos casos, quizás no refinarlas necesariamente, y el ordenamiento, como dije en ese sentido, es más modesto.

También es más modesta en el tamaño de las muestras y en el tamaño de las secciones electorales, lo cual sí nos indica que a pesar de ser más pequeñas y más modestas, pueden llegar a resultados bastante

confiables y yo creo que en el futuro de las encuestas de salida (*exit polls*) para el periódico vamos a procurar que, incluso, sean ejercicios aún más pequeños porque también simulaciones propias que hemos hecho nos han dado resultados bastante precisos.

Lo que quisiera ahondar un poco más es revisando la metodología que se publicó el 3 de julio ya en la edición impresa del periódico. Me di cuenta de que básicamente todos los detalles, lo que se ha mencionado de alguna manera ya, excepto uno, no me había percatado que habíamos puesto una nota que dice que el levantamiento se dividió en varias etapas, etapas que, de alguna manera, nos permitieron desarrollar otro tipo de ejercicios entre los mismos encuestadores, darles cierto tiempo y con eso, básicamente, me refiero a lo que posiblemente todos ustedes, encuestadores de salida, hacen.

Lo que hicimos fue diseñar tres diferentes horarios de levantamiento bajo los cuales estuviera cubierta toda la jornada electoral, desde las 08:00 horas de la mañana en que se deben abrir las casillas, hasta las 18:00 horas de la tarde en que también, en teoría, se deben cerrar. Lo que hicimos fue dividir en tres horarios. En un primer horario, en las zonas sombreadas, los encuestadores tenían la obligación de estar haciendo entrevistas a través de un salto sistemático, de hecho, también otras de las cosas más modestas en este caso es que nos fuimos con un número relativamente pequeño, comparado con otros, de entrevistas por casilla. En el mayor de los casos hicimos hasta 25 por casilla, en el menor de los casos hicimos hasta 12 ó 15, lo cual no nos trajo consecuencias graves en términos de la estimación.

Entonces lo que hacíamos era, básicamente, tener a nuestros encuestadores brincando votantes sistemáticamente y en cierto periodo no aplicar encuestas, sino dedicarse a alguna otra cosa, entre ellas, obviamente, transmitir su propia información o dársela a la gente que coordina el campo o simplemente tomar un tiempo para cualquier otra actividad que pudieran hacer, entre ellas reportar alguna otra cosa que les pedimos. Qué pasaba en ese tiempo, ese tiempo está cubierto por otro horario y, de alguna manera, no hay un momento de ocio durante el día, sino que, básicamente, estamos empleándolos en todo momento.

El único problema quizás que tuviera el diseño, en este sentido, es que el primer horario difícilmente empezó a las 08:00 horas, dado que las casillas no están en operación a las 08:00 horas. Entonces, éste se redujo un poco y, en algunos de los casos, dado que había encuestadores que no habían logrado la realización del número de cuestionarios dentro de su propio horario, se tuvieron que pasar un poco más allá

que en términos de diseño, tampoco yo veo mucho problema.

Lo que hicimos fue, obviamente, cada uno de los centros de coordinación incluido el periódico Debate de Sinaloa, el Mural de Guadalajara, el Norte de Monterrey y Reforma de la ciudad de México, teníamos una división de las secciones electorales que fueron básicamente asignadas por regiones, pero también asignadas aleatoriamente, de manera que en ciertas secciones se iba a aplicar alguno de los tres horarios y que funcionara de acuerdo con las expectativas que tuviéramos también en operatividad de captura. Es decir, también vemos que en cuestiones de captura podíamos tener algunos momentos libres para poder refrescar un poco a la gente que estaba en este trabajo. Básicamente éste es un asunto organizacional, no tiene nada que ver con la metodología en sí misma.

El resultado publicado fue el siguiente, más o menos. Yo estoy de acuerdo con Roy Campos, yo soy una de las personas que rara vez publica puntos decimales, yo creo que en ejercicios muestrales donde tenemos un margen de error mucho mayor que lo que implican puntos decimales, me parece que es demasiado arrogante hacerlo, no son tan precisos como para decir, este punto sale así, a lo mejor sí. En términos periodísticos no lo hemos querido manejar así, sé que pudiera haber gente que está en desacuerdo con ese sentido por la importancia de la estimación y, sobre todo, porque esperábamos, al menos, que ésta pudo haber sido una elección mucho más cerrada de lo que en realidad fue y, por lo tanto, había que estar preparado para poder decir si dentro de los márgenes de error definidos para el ejercicio estadístico, alguna diferencia entre primero y segundo lugar era una diferencia real o era una diferencia estadística.

Ahora en cuanto a la publicación, creo que aquí viene el segundo pero quizás más importante uso en este sentido, que es darle una cara un poco más explicativa; manejamos la historia de que el 2 de julio había ganado un México urbano y educado. Que finalmente fueron espacios especialmente importantes en el triunfo de Vicente Fox, es algo que de alguna manera ya se había previsto también en las encuestas electorales. Se habían manejado algunas historias en este sentido, se había manejado ya la fuerte relación que había entre estas variables y las intenciones de voto, pero

me parece que de la encuesta de salida (*exit poll*) surgen precisamente algunas otras cosas que pudieran haber sido de bastante interés para quienes finalmente leerían o tendrían el interés de leer estos datos.

Por ejemplo, con el voto reportado de 1994, que no es el voto de hecho de 1994 pero si preguntando ¿usted votó? y ¿por quién votó? en 1994, el 25 por ciento en ese momento había votado por Ernesto Zedillo en 1994 y en el 2000 lo hicieron por Fox. Otro de los hallazgos de la encuesta de salida (*exit poll*) que me parece que fue bastante interesante y de hecho nos llevamos una parte de ello en la primera plana, fue que la mitad de los simpatizantes de Roberto Madrazo, por los cuales nos referimos a aquellos que nos reportaron en la encuesta que habían votado en la elección interna del PRI y que habían votado por Roberto Madrazo, la mitad de ellos habían votado por Vicente Fox, lo cual, de alguna manera, pudo haber sido o no pudo haber sido sorpresivo, porque quizás una buena parte del apoyo de Madrazo en la elección interna había sido de gente que no era simpatizante del PRI, y en la otra, de lo cual abundaré un poco más en la siguiente mesa de trabajo, que efectivamente hubo muchos de los llamados madracistas que le voltearon la espalda al candidato del PRI.

Se lleva todo en términos regionales, no hubo manera de dar datos a nivel estatal, nuestra muestra era de tres mil 380 casos, entonces no nos fuimos a nivel estatal pero sí manejamos las regiones que habíamos venido manejando en términos ya de las publicaciones de las encuestas preelectorales y algunas otras preguntas.

Algo que se me está pasando y que creo que es importante decir, y eso explica la presencia de esta pregunta sobre todo, ¿algún familiar o conocido de usted que radique en Estados Unidos le dijo que votará por algún candidato en particular? Responde al interés de nuestro, en ese momento, socio y compañero de trabajo de la encuesta de salida (*exit poll*) que fue Los Angeles Times. Este fue un ejercicio que hizo en conjunto grupo Reforma con Los Angeles Times, y lo que salió de esto no fue solamente un poco la cooperación metodológica que de hecho casi estuvo en un 90 por ciento a cargo de Reforma, pero incluso también en términos del patrocinio, pero sí para dar una salida todavía más allá del periódico, que fue la publicación en Los Angeles Times. ■

## PONENCIAS MESA 5

.....

Francisco Báez

LA CRÓNICA DE HOY

**R**eseñaré, espero que brevemente, mi experiencia en conteos rápidos en elecciones federales y cómo, a través del método que a todos nos sirve —el de prueba y error—, he llegado a uno que funciona, y a mi juicio funciona muy bien, con rapidez y costos mínimos. Se trata de un ejercicio heterodoxo que es hijo de las dificultades para realizar con éxito y a costos bajos, conteos rápidos con métodos ortodoxos.

El primer conteo rápido en que participé fue en 1982 como parte de un equipo interdisciplinario del PSUM. Se hizo una muestra totalmente aleatoria de mil ocho casillas y se instruyó a los representantes partidistas a enviar a la brevedad los datos de las elecciones presidenciales de ese año.

A la madrugada del día siguiente de la elección habíamos completado apenas, aproximadamente, el 65 por ciento de la muestra. Los datos que nos arrojaba eran consistentes con los que hasta esos momentos había difundido el Consejo Federal Electoral. Calculamos, asimismo, tanto la participación promedio como el número de votos totales que obtendría nuestro candidato a partir del promedio por casilla observado en la muestra: poco más de un millón 100 mil, poco más del cinco por ciento del total de votos válidos.

Días después, los resultados oficiales fueron diferentes. El candidato del PRI tenía cerca de 10 puntos porcentuales más y el del PSUM 250 mil votos menos, por debajo del cuatro por ciento del total. Nos dimos a la tarea de buscar dónde se habían “perdido” esos 250 mil votos. Casi todos estaban en Guerrero, Oaxaca y Chiapas. Hicimos una investigación y resultó que los compañeros de Guerrero y algunos de Oaxaca decidieron enviar al centro de cómputo datos de casillas muy distintas de las de la muestra. Les daba pena el resultado obtenido en las casillas seleccionadas. De Chiapas no había más explicaciones que un fraude o una “buena suerte” extraordinaria en las casillas que habían caído en muestra.

Lecciones de 1982: 1) no hacer una muestra más grande que la capacidad para levantar los datos; no utilizar a militantes (o en todo caso, enfatizar el papel estrictamente estadístico del conteo); 2) buscar un mecanismo sistemático para la distribución de las casillas de muestra, y 3) pensar primero en la posibilidad de error en el conteo que en la posibilidad de manipulación externa.

En 1985, todavía en el PSUM, hicimos dos conteos rápidos que estuvieron a mi cargo, uno para el DF y otro para el país. Aplicamos lo aprendido. En ambos casos utilicé un método aleatorio-sistemático, en el que la unidad es el ciudadano y no la casilla electoral, bajo el supuesto de que el tamaño promedio de cada casilla es el mismo. Igualmente, se redujo el número de casillas a 300 en la muestra nacional y 100 en la capitalina. Se instruyó que se diera

prioridad absoluta a la cobertura de esas casillas y que no se cambiara en ningún caso (tanto es así, que en Pinotepa Nacional, donde el representante del PSUM fue expulsado, éste reportó el resultado: PRI 400, todos los demás 0).

La muestra fluyó con mucha más velocidad. En el DF se completó la misma noche de la elección y sus resultados eran iguales a los de la submuestra capitalina (de 40 casillas) dentro de la muestra nacional. Los resultados de la muestra del DF tuvieron un error máximo de dos décimas de punto porcentual respecto de los oficiales; los de la submuestra, un error máximo de cinco décimas de punto. Era el camino correcto.

A nivel nacional, sin embargo, la muestra tampoco se pudo completar: llegó al 82 por ciento a las 16:00 de la tarde del lunes. Decidimos hacer una ponderación *sui generis*, a partir del hecho evidente de que las casillas que faltaban se encontraban en zonas con escasa o nula presencia del PSUM: calcular el número promedio de votantes por casilla, otorgarle al PRI el 90 por ciento de los votos en cada una de ellas y dividir el restante 10 por ciento entre todos los demás partidos, menos el PSUM, de una manera proporcional a los datos que teníamos en el 82 por ciento capturado. Aún así, frente a los datos oficiales, los nuestros parecían subestimar al PRI y al PPS y sobrestimar al PAN y al PMT. Siempre he tenido la idea de que los datos finales de las elecciones de 1985 fueron abiertamente manipulados.

Lecciones de 1985: 1) el método aleatorio-sistemático funciona bien y puede resultar representativo utilizando un número considerablemente menor de casillas, al menos en entidades con cierta homogeneidad en el voto y en las que existan mecanismos de vigilancia; 2) toda asignación arbitraria de resultados (por más lógica que parezca) en la zona no cubierta de la muestra, crea un área de penumbra que de poco sirve para el análisis (o el alegato), y 3) todo conteo nacional partidista está destinado a tener estas áreas de penumbra, a menos de que se trate –y eso es muy difícil– de un partido muy disciplinado e implantado, efectivamente, a lo largo y ancho del territorio nacional.

En 1988 realicé profesionalmente, bajo los auspicios del semanario Punto, un conteo rápido electoral capitalino. Para eso, se aplicaron las lecciones de 1985: una muestra muy pequeña (40 casillas) con un equipo profesional de campo. Los resultados quedaron dentro de un margen aceptable de error (1.4 puntos porcentuales para el primer lugar, Cuauhtémoc Cárdenas).

Aquí lo interesante fue que en vez de sumar los resultados brutos de la muestra, utilicé un ponderador que igualaba los tamaños de las casillas. En aquella época había una diferencia enorme en el tamaño de

las casillas. Este ponderador, aparentemente lógico en términos estadísticos, (es decir, si utilizo al ciudadano como unidad para la selección de la muestra, automáticamente tiene menos probabilidades de caer en ella el ciudadano que vota en la casilla pequeña: ponderando el tamaño de las casillas le devuelvo a ese ciudadano su porcentaje dentro de la muestra) alejó los resultados brutos, que eran casi exactos, de los resultados finales publicados.

Un análisis *ex-post* me hizo ver que el comportamiento electoral en las casillas pequeñas (ubicadas en colonias de clase media) sí tenía diferencias importantes con el de las casillas gigantes (ubicadas en colonias populares de nueva creación). Opté por regresar al método de sumar directamente los resultados de cada casilla. Esa fue, para mí, la lección de conteos rápidos de 1988.

En 1994 fui invitado por un consejero del IFE a realizar un conteo rápido experimental. Como se recordará, en aquella ocasión se acordó que mostraría datos casilla por casilla, pero que no se darían a conocer datos agregados de las elecciones presidenciales hasta no tener el 15 por ciento del PREP. La idea era tener los porcentajes de cada candidato antes de que el PREP llegara al 15 por ciento. Ahí nació el método que he utilizado desde entonces.

Este método, que no utiliza personal de campo, requiere una muestra pequeña, confiable y flexible. En 1994 utilicé una muestra aleatoria-sistemática basada en la distribución por tamaño del padrón electoral por entidad y municipio. La muestra consistió en 80 secciones divididas en dos submuestras de 40 secciones. En cada una de las secciones se utilizan los datos de la casilla que llegue primero. En el caso de secciones rurales se hacen conglomerados (*Clusters*) intercambiables de cuatro a ocho casillas, tomando en cuenta de que se trate de comunidades rurales con condiciones sociales e historia electoral semejantes (es decir, no se puede meter en el mismo conglomerado (*Clusters*) una casilla rural de Oaxaca y una de Sonora). Si la casilla seleccionada tarda en llegar al sistema se utiliza la última que llegó si está en el mismo distrito, municipio y corresponde a la misma caracterización rural-urbana (en los casos de Oaxaca, Puebla, Veracruz y Yucatán se elimina el requisito de municipio). La suma de cada submuestra se hace sin ponderar. Los resultados de cada submuestra deben tener diferencias mínimas (inferiores a un punto porcentual).

En 1994, cuatro horas antes de que el PREP llegara al 15 por ciento, el conteo rápido experimental no sólo tuvo un margen de error inferior a 1.5 puntos porcentuales para los dos principales candidatos, sino que

fue exacto en el orden que tuvieron los nueve contendientes a la Presidencia.

Este método, repetido en las elecciones de diputados federales de 1997, dio diferencias inferiores a un punto porcentual respecto a los resultados oficiales, para los tres principales partidos.

Bajo esta lógica se decidió difundir la noche del 2 de julio un conteo rápido en la edición en línea de Crónica, cuyos resultados, asimismo, fueron publicados en la edición impresa del lunes 3 de julio.

La metodología usada fue:

- 1) Se escogieron dos muestras de 52 casillas cada una, a través del método aleatorio-sistemático en el que la unidad no es la casilla o sección, sino el ciudadano empadronado (de forma tal que cada elector tenga la misma probabilidad de que su casilla caiga en la muestra y de que la probabilidad de que una casilla sea escogida sea proporcional al tamaño de su padrón).
- 2) En cada una de las muestras, cada sección tiene una sección "alterna", escogida de manera aleatoria, ubicada en el mismo distrito y en el mismo municipio o delegación, si se trata de zonas urbanas; si se trata de zonas rurales, se elaboran conglomerados (*Cluster*) de cuatro secciones con los mismos requerimientos de distrito y municipio. El conteo recoge la primera de las casillas que llegue del PREP del IFE, difundido por Internet. Con esto se obtienen resultados más rápidos, con base en el hecho de que la varianza de la votación entre los ciudada-

nos con apellidos de la A a la L o M y los que tienen apellidos entre M o N y Z (los votantes de la casilla básica y la contigua) es, estadísticamente, insignificante y en el supuesto, comprobado en los hechos, de que la varianza en el comportamiento electoral intramunicipal es pequeña.

- 3) Se cotejan los resultados de las dos muestras que no deben tener diferencias significativas en las principales candidaturas (un punto porcentual).
- 4) Se dan a conocer los resultados del conteo sólo si la diferencia entre el primero y el segundo lugar supera los 2.5 puntos porcentuales.

Con este método, tres conexiones a Internet y un personal mínimo, a pesar de las dificultades que se tuvo con el PREP en Internet para llegar a datos desagregados por casilla, a las 22:30 horas se subió a la edición Web el resultado: Fox, 43.8 por ciento; Labastida, 35.9 por ciento; Cárdenas, 17.3 por ciento; Rincón, 1.9 por ciento; Camacho, 0.7 por ciento; PARM, 0.4 por ciento. Un error máximo inferior a un punto porcentual, un error promedio de tres décimas de punto.

El desarrollo de los sistemas electrónicos de información hará que en un futuro próximo métodos similares al reseñado puedan servir para hacer los conteos con más facilidad y rapidez. Ante este reto deben buscarse nuevas vías para innovar los métodos de proyección rápida de resultados electorales definitivos. La que más posibilidades plantea, por el momento, es la combinación del conteo de resultados de casilla con los datos de encuestas de salida. ■

# PONENCIAS MESA 5

.....

Javier Alagón

ESTADÍSTICA APLICADA PARA COPARMEX

Creo que se ha dicho demasiado con respecto a conteos rápidos y quisiera compartir con ustedes algunas de las reflexiones que me hago al final de los dos días ya de esta jornada. Vine aquí con muchísimas expectativas, algunas se han cumplido, otras no tanto, y quisiera ver hacia dónde vamos con esto que hemos aprendido a lo largo de estos dos días.

Decía Norbert Diner, premio Nobel de química, en su inglés con acento alemán, *"it's difficult to predict in particular the future"*. Esto no parece ser para los conteos rápidos, realmente predecir con conteos rápidos es relativamente fácil sobre todo si se quita la restricción del tiempo, la restricción que tenemos que entregar a las nueve o diez de la noche cualquier cosa.

Antes que otra cosa les quería hacer un comentario con respecto al trabajo que todo el gremio realizó el 2 de julio, y creo que se ha soslayado. Creo que debemos estar realmente de júbilo: Muchísimas felicitaciones y lo quiero hacer en público, al trabajo hecho por el comité técnico, me pareció extraordinario lo que nos presentaron el día de hoy. Realmente no sabíamos que los mexicanos estábamos en tan buenas manos.

Ahora el conteo rápido. Ahí se presenta el objetivo; evidentemente es la estimación confiable, objetiva e independiente de la proporción de votos a cada uno de los partidos.

Nosotros realizamos un conteo rápido en 600 secciones electorales que correspondían a mil 315 casillas, reportamos creo que con 440 por ahí de las 22:15 horas. La selección de las secciones electorales fue de manera estratificada y, como ya se ha mencionado, por resultado de las votaciones en las elecciones presidenciales pasadas y por conglomerados dentro de cada estrato. Aquí sí nos quisimos tomar este riesgo, es una forma de llevar a cabo el muestreo que habíamos utilizado en dos ocasiones anteriores y que nos había funcionado de manera maravillosa, tanto en Jalisco como en Yucatán, en donde habíamos hecho los estratos con base en las votaciones anteriores.

La teoría dice que la estratificación funciona mejor, es decir, se disminuirá la varianza, conforme la variable de estratificación esté más relacionada con la variable que se está midiendo. Y aquí tenemos la situación ideal. No podíamos soslayar eso. Entonces, pues, para eso se realizó la estratificación, formamos siete estratos, creo que pueden llegar. En ocasiones previas habíamos tenido cinco estratos y en esta ocasión hicimos priístas muy priístas, panistas muy panistas y perredistas muy perredistas, competidas y secciones nuevas sin votación previa.

Los resultados del conteo rápido estuvieron listos a partir de las 10:15 con el 74 por ciento de las casillas computadas. Muy pocos de los que están aquí se enteraron de que hicimos un conteo rápido, fue el conteo rápido menos publi-

citado ¿Por qué?, porque además no tan solo este conteo rápido no se le dio tanta difusión, prácticamente se dio a las 20:00 horas con una encuesta de salida como lo hemos visto y, a partir de ahí, todo fueron muchísimas celebraciones y ya que si los números más o menos, la verdad, a la mayoría de los mexicanos no les importaba mucho.

Utilizamos 450 personas para reportar los resultados y nos instalamos en Coparmex en Insurgentes. Esto fue lo que reportamos a las 10:15 con 440.

Tenemos los resultados finales del IFE a la izquierda; a la derecha tenemos, en medio, los resultados que reportamos, nos quedamos 1.2 puntos porcentuales menos, 0.8 para el PAN, etc. El error de estimación en este momento estaba en 1.3 puntos porcentuales, esto es totalmente lo que sucedió: 1.2 por ciento, y tenemos un error de estimación de más menos 1.3; para el caso del PAN de más menos 1.1; para el caso del PRD más menos 1.2 puntos porcentuales.

Lo que habíamos tenido en experiencias anteriores, con un tamaño de muestra prácticamente bueno. Es más, incluso, la tercera parte en Jalisco y 200 secciones electorales en Yucatán, los márgenes de error estaban alrededor del 1.5 por ciento. Por eso pensábamos, junto con muchas simulaciones que hicimos para estimar los márgenes de error antes del día de la elección, todo tipo de pruebas, pensamos que iban a estar alrededor del 1.5 por ciento y se cumplió. Y los errores de estimación varían conforme los diferentes partidos.

Estas son las estimaciones puntuales, ya las vimos. Estas son las estimaciones por intervalo del conteo rápido, ya sin nullos, evidentemente, ahí le suman y le restan. Nos pasamos y aquí estoy usando decimales.

Este diseño metodológico, este diseño muestral como les comento, no son muchas experiencias, dos, pero había funcionado muy bien. En el caso de Yucatán esa sí fue una elección mucho más cerrada que ésta y en aquel caso, las estimaciones fueron más cercanas a lo que terminó el resultado final. Sin embargo, el margen de error fue muy parecido a lo que tuvimos aquí, fue el mismo nivel de precisión. En esta ocasión nos quedamos a un punto del PRI y del PAN, pero eso es simplemente cuestión de azar, no de precisión.

Tenemos, entonces, el marco muestral que tomamos, fue el listado de todas las secciones electorales, según el IFE, creo que esto me lo voy a brincar y pasamos a los siete estratos. Tenemos los tamaños de muestra de los diferentes estratos y el diseño es convencional, se los muestro y en la siguiente vemos que lo que hicimos fue utilizar las ponderaciones (WH) con base en 1994 y dado el carácter de elección presidencial decidimos utilizar 1994.

No creemos que sea un supuesto, a pesar de los comentarios de Edmundo Berumen y la discusión al respecto. Si bien es importante ver cómo se puede tratar el aspecto metodológico, hicimos muchas simulaciones con las muestras que sacamos, variando todo tipo de condiciones, variando tasas de participación dentro de las secciones electorales y sorprende la estabilidad de las estimaciones, es impresionante. Por eso creo yo que si trabajamos el problema, muy seguramente lo que nos va a dar es que a lo mejor no tiene gran efecto y que a lo mejor es un prurito académico y que vale la pena desde luego, da la impresión de que no va a tener mucho impacto.

Este es un diseño totalmente ortodoxo, totalmente documentado, las fórmulas están perfectamente desarrolladas en los libros y el chiste es echarlo a andar, ponerlo a funcionar, hacer que esto funcione y poder estimar los errores la noche de la elección para poder salir con cierto nivel de precisión.

Bien, hasta aquí del conteo rápido, fue la tercera ocasión, pero hasta ahí. Yo quisiera aprovechar estos últimos momentos para hacer unas cuantas reflexiones sobre lo que hemos vivido estos dos días.

Creo que no debíamos estar discutiendo sobre la ortodoxia de los métodos. Simplemente me dieron ayer una referencia (*flash back*) a la discusión del 1994, que si las muestras aleatorias, que si la muestra por conveniencia de no sé quién, que si las 25 ciudades de no sé quién más. Finalmente, estamos convencidos todos de qué debíamos utilizar y no debíamos salirnos del enfoque ortodoxo del muestreo.

Tenemos el asunto de la agenda pendiente, para mí esto es fundamental. ¿Cuándo se cruzaron las preferencias electorales? Nosotros hicimos una encuesta para un partido político de oposición en noviembre, créame que la diferencia entre el PRI y el PAN era enorme y entre el PAN y el PRD era también muy grande, eso fue en noviembre y no creo que haya cambiado mucho a enero. Entonces, eso sí me da la impresión de que en algún momento, y no porque yo haya hecho encuestas, de que por ahí de enero o febrero sí estaba el uno, dos, tres como estaba tradicionalmente, esa es mi impresión. Sin embargo, hemos visto exposiciones que, incluso, cuestionan esta posición, y la hipótesis, la premisa fundamental es, en algún momento, el PRI empezó a caer muy fuertemente y el PAN sube tanto que se cruzan las curvas.

¿Cuándo ocurrió eso?, pues sería lo de menos, ya ganó Fox y es lo de menos, pero no es así, porque para poder saber qué metodologías, qué filtros, qué preguntas, diferentes aspectos, necesitamos poder responder esta pregunta, si no la respondemos vamos a tener

la agenda pendiente y nos vemos en el 2006 y vamos a volver a tocar el mismo tema, y en el 2012 también.

Entonces, la pregunta fundamental es a qué se deben entonces las diferencias entre encuestas electorales. Me mete muchísimo ruido intelectual, pero la exposición que hizo ayer Eduardo Ragasol me mete muchísimo ruido en este sentido. Una agencia con el prestigio de Nielsen, con una contratación impecable, un medio, con la infraestructura de campo adecuada, con los muestristas adecuados, con los resultados que nos presentó, contra las otras posiciones que las otras encuestas que nos mostraban equis tendencia.

Eso desde luego que nos mete una interrogante, ¿qué es lo que no está funcionando?, sólo de esta manera si contestamos, vamos a poder proveer de información creíble a la sociedad mexicana, ya no digamos sólida, confiable, creíble, simplemente en términos de encuestas electorales.

Tenemos que dilucidar esto, son diferencias por diseños muestrales.

Sabemos que si empezaron en equis localidades y ahí se mantuvieron y esa pregunta no me la respondió más que Manolo, el punto de los que hicieron encuestas con rotación de muestras (*trackings*) espero que hayan hecho sus muestras independientes y no por abatir costos nos hayamos quedado o se hayan quedado en un punto geográfico similar. Tenemos que dilucidar estas diferencias por levantamiento de campo, diferencias por filtros, ya ha habido alguna pequeña discusión, lo que le falló a equis o zeta es la forma en que preguntó. Puede ser tan grave el poner totalmente probable, eso es lo que le estaba argumentando a Eduardo Ragasol, a poco eso fue lo único, realmente por eso no funcionó.

Bien, diferencias por posición de pregunta del voto, porque si no respondemos esto entonces el debate se va a ir hacia los indecisos, son los indecisos, estamos

en los indecisos o no, porque a lo mejor ni siquiera tenemos que llegar a los indecisos, a lo mejor las diferencias únicamente se deben a cuestiones de diseño, a cuestiones de campo, a cuestiones de cómo está la muestra y, en este sentido, ahora viene una propuesta.

La propuesta que yo quisiera hacer es, simplemente, juntar y analizar la información disponible; la pregunta es: ¿Estamos abiertos? Ayer Ulises Beltrán pasó un cuestionario, lo tiene Ana Cristina Covarrubias, y yo creo que habría que complementarlo para ver si lo vamos a responder. Pero vayamos un poco más allá, por qué no se sientan en la misma mesa Manolo y Gauss y Rolando Ocampo y ven cómo están las muestras geográficamente distribuidas, ven en qué distritos, en qué secciones y cómo fueron las votaciones en esas secciones, porque si ya desde ahí hay diferencias, caray, pues miren, ni nos peleemos con esa asignación de indecisos, ni siquiera nos vayamos por allá, ni nos peleemos. También por aquí se mencionaron sesgos favorables hacia el PRI, ahora ya parece que van a seguir hacia el PAN.

En la encuesta que nos presentó María de las Heras se había mencionado la parte de los sesgos hacia el PRI. No se ve, incluso, el sesgo hacia el PRI porque antes de la asignación de indecisos traía 11 ó 12 puntos porcentuales y ya después de la asignación de indecisos traemos 10 puntos porcentuales

Entonces, la pregunta es por qué desde 1994 había esta hipótesis de que si había un sesgo o no. Bueno, si hay, ¿lo podemos o no lo podemos quitar? Todo esto es una propuesta o la pregunta es ¿simplemente esperaremos a que la mano invisible del mercado, sabiamente, acomode las posiciones de los diferentes agentes que participan en las encuestas preelectorales? Y lo malo de todo esto es que esto va a ocurrir quizás en 18 ó 20 años, cuando hayamos ido aprendiendo poquito a poquito. ■



## PONENCIAS MESA 5

.....

Enrique Calderón Alzati

FUNDACIÓN ROSENBLUETH PARA EL PRD

**Y**o voy a hablar de la encuesta de salida que hicimos para las elecciones federales. Esta encuesta nosotros la hicimos para el Partido de la Revolución Democrática, aunque los resultados fueron dados a conocer a través de la televisión estatal, alrededor de las 20:00 horas.

Los objetivos de nuestra encuesta eran conocer, desde horas tempranas de la jornada, las tendencias de la elección presidencial a nivel nacional y en cada una de las 32 entidades. Conocer también las tendencias para el Senado en cada una de las 32 entidades y, finalmente, contar con un esquema confiable para analizar *a posteriori* el proceso electoral, que pudiera servir para futuras acciones políticas del partido.

Lo que nosotros hicimos fue utilizar una encuesta ponderada muy *sui generis*, que tenía como característica utilizar pocas secciones electorales relativamente para todos los resultados que nosotros queríamos dar, que era precisamente conocer las tendencias en las 32 entidades.

Lo que nosotros hicimos fue utilizar la información que teníamos de carácter electoral de elecciones anteriores, pero también información muy rica de carácter socioeconómico del país. Nosotros sabemos —creo que es del dominio público— que la población urbana y la población rural tienen formas distintas de votar y no sólo está la estratificación rural y urbana; hay otros elementos sociales que son muy importantes en la votación. Uno de estos son los niveles de escolaridad; la gente con diferentes niveles de escolaridad vota de manera distinta, con diferentes niveles de ingreso y también con diferentes perfiles de ocupación. Entonces, éstos fueron útiles para el modelo que estuvimos trabajando.

También trabajamos utilizando los resultados electorales previos.

Esta gráfica de las elecciones de 1994 muestra, si dividimos el país en seis estratos, desde grandes metrópolis, ciudades grandes de más de 400 mil habitantes; medianas de entre 120 y 400; ciudades pequeñas y municipios semiurbanos y rurales, ¿cómo había estado la tendencia de la votación para el PRI, para el PAN y para el PRD? Entonces, este tipo de información fue la utilizada para construir el modelo.

Lo que vemos aquí es la conformación de acuerdo con los censos de 1990, pero ya manejados en términos de votación en 1994. Tomamos del país 28 municipios metropolitanos; 25 municipios que constituyen las grandes ciudades; 95 las ciudades medianas; 445 las pequeñas, los municipios semiurbanos y los mil 280 rurales, y esta es la composición de los votos de 1994. Lo que nosotros hicimos fue construir una muestra en donde los municipios que escogimos, nuestro primer criterio de selección fueron los municipios para que

nuestra muestra tuviera una composición similar al total del país.

Esto es otra parte: las preferencias en los municipios con diferentes niveles de escolaridad. También tenemos otra información similar en donde lo que tomamos en cuenta son los niveles de ingreso y el perfil por ocupaciones. Entonces sí hay características muy claras en las formas de votación.

Utilizamos también la información a nivel municipal de las elecciones de 1997 de cómo había votado la gente en las elecciones federales de 1997. Este es el caso del PAN, igual tenemos el del PRI y el del PRD. Entonces, con esto lo que nosotros hicimos es construir una muestra tal que se comportara exactamente igual que el total del país en los aspectos rural, urbano, en los aspectos de escolaridad, en los aspectos de ingreso y también en la parte de preferencias políticas.

Lo que hicimos fue dividir el país en 136 regiones electorales; dividimos en promedio cuatro regiones para cada estado. En realidad, los estados más pequeños fueron de dos, los más grandes fueron de seis regiones, y luego para cada una de las regiones asignamos tres municipios y seis secciones, de manera que todo el país estuviera representado. Los objetivos era tener, entonces, un perfil rural, urbano y de escolaridad, también de ingreso, también de ocupación, de la muestra similar al país en su conjunto.

Y una composición de las muestras con preferencias electorales muy cercanas a las del país, y también estado por estado, entre los municipios seleccionados en la muestra y la totalidad de los estados.

Trabajamos en un juego de ponderadores. Es un poco más complejo, pero creamos un ponderador regional que nos asegurara la proporcionalidad entre el tamaño de la población votante, el de la región y el de nuestra muestra. Un segundo ponderador –el primero era para todos los municipios de la región– el segundo era específico para cada municipio y éste estaba orientado a que se comportara como todo el país, en lo escolar, en lo rural, urbano, en lo que tenía que ver con el ingreso y las ocupaciones. Y, finalmente, un ponderador seccional que nos permitía ajustar los niveles de preferencia electoral. El ponderador total era el producto de los tres.

Lo que nosotros logramos es que para las elecciones de 1997 nuestra muestra tenía una diferencia con la totalidad del país del 1.06 por ciento. Esta misma muestra, con estos ponderadores, la aplicamos a la elección de 1994 y nos dio un sesgo del 0.2 por ciento para la elección a nivel nacional.

La muestra construida con estos principios fue de 433 municipios y 783 secciones, y ahí está el nivel

de sesgo por entidad. El sesgo más grande fue de 1.2 por ciento respecto a la elección de 1997 y del 0.06 a nivel nacional.

Estos fueron los municipios como estaban distribuidos en el país. En el caso de Oaxaca los separamos porque son muy pequeños y no se ven en el mapa nacional.

Este es el número de la tabla donde están los municipios y las secciones en cada una de las 32 entidades.

¿Cuáles fueron los resultados que obtuvimos? Bueno, el primero fue y lo veo muy similar a lo que se ha dicho, el primer resultado lo mandamos a nuestro cliente a las 11:44 y, lo que tengo registrado aquí, el último fue a las 19:24. Voy a dar la última, de hecho, desde las 11:44 era claro que Fox iba a la cabeza. En ese momento teníamos solamente 80 secciones con tres mil 285 entrevistas que daban a Fox el 48 por ciento, siguiendo la idea de no utilizar decimales; Labastida, 32; Cárdenas, 17; y otros 2 ó 3. Al final, a las 19:24, se ve que lo que teníamos eran 46 por ciento para Fox, 18 por ciento para Cárdenas y 31 por ciento para Labastida y cuatro para otros.

Luego, para ver la consistencia –qué tanto nuestra gráfica cumplía con las hipótesis con las que habíamos comenzado– dividimos los municipios en cuatro grupos: metrópolis, ciudades grandes, ciudades pequeñas y zonas rurales, y los resultados de las 19:00 horas nos daban estos resultados que se parecieron más a lo que fueron los resultados del IFE.

Esto ya está dividido como estaba originalmente en la muestra: metrópolis, ciudades grandes, pequeñas, semiurbanos y rurales, en donde se ve muy claramente que la hipótesis se cumplía, o sea, la composición era la misma, excepto que en todos los casos la banda del PRI había bajado, la del PAN había subido, la de Cárdenas se había mantenido más o menos similar a la de la elección de 1994.

Aquí están los resultados que obtuvimos ya por entidad. Pero algo que hicimos, que creo que es muy interesante, fue dividir las 32 entidades en tres grupos: las de las entidades gobernadas con gobiernos perredistas, las entidades gobernadas por el PAN y las entidades con gobiernos del PRI.

Esta gráfica es de los resultados en los estados gobernados por el PRD y se puede ver muy claro que lo que tuvimos en la encuesta de salida y los resultados del IFE son idénticos, las variaciones son, en todos los casos, de menos de uno por ciento. Estos resultados se los entregamos al Presidente del IFE para que hubiera una constancia de ellos. En lo que se refiere a los nuestros la comparación es posterior.

Estos son los resultados del PRI en donde hay una diferencia más perceptible. En el caso del estado de

Zacatecas, sin embargo, es del orden del uno por ciento. Finalmente también está la del mismo PRD, donde las tres son prácticamente iguales. En nuestra encuesta no consideramos las no respuestas y todo está ajustado, como en el IFE, a quitar la no respuesta.

En el caso de las entidades gobernadas por el PAN, encontramos una diferencia mayor; ésta anda entre el dos y el tres por ciento para el PAN. El caso se abre más en Querétaro y voy a dar la hipótesis en un momento.

Ésta es la del PRI y quiero indicar que en todos los casos nuestra encuesta de salida da un nivel menor de votación que la que da el IFE, y donde hay un mayor nivel de error es en el caso del PRD. Los niveles de error son pequeños, pero como la votación es muy pequeña, entonces se notan mucho.

Quiero mostrar que la mayor diferencia se da en el caso de Nuevo León, pero también se da en Guanajuato y en Baja California. La hipótesis que nosotros tenemos es que las encuestas de salida, de alguna manera están midiendo algo distinto que los resultados del IFE. Lo que de alguna manera están midiendo es lo que yo diría la votación no vergonzante, o sea, al tipo que se le pregunta y si tiene alguna razón, por ejemplo, que su voto haya estado comprado, coaccionado o como le quieran llamar. Eso no se registra en una

encuesta de salida, de manera que estos resultados lo que nosotros podemos ver es qué tan controlado estuvo este tipo de procesos en la elección.

De aquí lo que surge es que el PRD, teniendo un voto más rural que el PAN, en los estados gobernados por el PRD es bastante más fácil el control de las operaciones de cooptación o de compra de votos que en un estado en donde gobierna el PAN, donde su presencia se da en las zonas urbanas y las zonas rurales no son cuidadas.

Vamos a pasar ahora a los estados gobernados por el PRI y aquí vemos, en la mayor parte de los casos, que la votación de la encuesta de salida es mayor. Por ejemplo, en el caso de Chihuahua o Durango y también de Tabasco. En el caso de Tabasco, lo que nosotros medimos es una diferencia de aproximadamente un ocho por ciento entre la encuesta de salida y los resultados oficiales. Si ustedes ven Chiapas es a la inversa; lo que vemos aquí es que la encuesta de salida es menor que lo que dicen los resultados oficiales. La explicación que esto tiene para nosotros es que hay zonas en donde se da voto de miedo y hay zonas donde se da compra de voto. Para nosotros esto es lo que da la importancia y lo que estamos utilizando.

Ésta es la del PRI y, finalmente, la del PAN. ■

# DEBATE

.....

Manuel Román

MODERADOR DEL DEBATE

**EDMUNDO BERUMEN:** Muy rica la sesión, mucha tentación de hacer muchos comentarios y necesariamente hay que restringirlos y hacer una selección. Primero, una observación de carácter general, en un tema de discusión metodológica la gran ausencia en esta mesa de trabajo fue la discusión de detalles metodológicos sobre cómo operan las encuestas de salida en el terreno, cuántos de los amigos que nos reseñaron sus encuestas de salida usaron un solo encuestador, cuántos dos, cuántos más, cómo eran sus mecánicas, sus intervalos de selección, cómo eran las maneras de expandir, etc.

Es un conocimiento técnico adquirido (*know how*) que tienen ustedes y a subrepticamente bajo llave, porque una cosa que hay que decir abiertamente, a puerta cerrada, es que nuestro personal de campo trabaja para todos nosotros y son el correo para repartirnos nuestros formatos de campo, nuestros manuales de campo, nuestros materiales de capacitación, y que teniéndolos no estemos dispuestos a ponerlos aquí sobre la mesa y discutirlos, se me hace un desperdicio.

Ese era el comentario de carácter general. Entrando a algunas preguntas puntuales sobre algunas de las riquísimas participaciones. En la presentación de José Mendoza seguramente algo no copié bien, pero en alguno de los acetatos que presentó decía que en algún momento en el tiempo de un corte de estimación con el 87 por ciento de la muestra, tenía 26 mil 618 entrevistas de la encuesta de salida, y en ese mismo cuadro en otra columna, que con el 98 por ciento tenía 67 mil 43. Seguramente yo estoy equivocado, anoté mal algún porcentaje, pero si los anoté bien, no entiendo cómo, más que duplico la muestra cuando sólo incrementé 11 puntos porcentuales, y eso es lo que creo que decía el renglón de la gráfica.

Hablaba de un diseño que produjo más o menos 1.1 por ciento de error y también en un renglón aparecían errores del tres por ciento, 2.3 por ciento y 1.6 por ciento y aquí eso no es importante, lo que es importante es que hablaba ya del error estimado y nuevamente algo se me ha de haber escapado porque daba así ese porcentaje de error, pero no se lo imputaba a la estimación de ningún partido y no sé si lo dijo y se me escapó, que era quizás a lo mejor promedios de errores de las estimaciones de todos los partidos, pero no lo decía, nada más decía error en la gráfica.

Cuando nos platicó de sus réplicas, otra vez algo se me ha de haber escapado. Entiendo que de cada tamaño de muestra, tamaños de 100, de 200 y hasta de dos mil, de cada uno de esos tamaños sacó varias muestras e hizo el cálculo con esas varias muestras y que los errores estándar y errores cuadráticos medios que nos enseñó en la gráfica, era el promedio de esos errores, si entendí bien,

y que para hacer este ejercicio usó la base de datos del PREP. Bueno, si fue como lo entendí, a lo mejor estoy mal, hay algo que se me escapa porque si yo tengo el universo que estoy estimando y lo tengo con el PREP, el error estándar o el error cuadrático medio es uno para cada tamaño de muestra; por ilustrar en un muestreo aleatorio simple, el error estándar de la estimación de un porcentaje es "p" por uno menos "p" entre "n". Y ese es un parámetro poblacional y no me cambia de muestra a muestra, y tengo el parámetro poblacional porque tengo el PREP. Entonces, si tengo 50 muestras de tamaño 100 ó 50 muestras de tamaño dos mil, siempre me da el parámetro poblacional de error; no es algo que tenga que estimar y luego promediar, no tengo que tomar "p" minúscula por uno menos "p" minúscula entre "n".

Va a haber distintos errores estándar y distintos errores cuadráticos medios por el solo hecho de cambiar de tamaño de muestra, porque la "p" y la "q" se me quedan, pero lo que me está cambiando es la "n" del tamaño de muestra. Entonces algo se me escapó ahí. Una clarificación.

En los criterios que nos compartió Roy Campos, de cómo juzga los conteos rápidos, certidumbre, oportunidad, eficiencia, precisión, están yo creo que muy interesantes para seguirlos discutiendo y examinando a la luz de distintos clientes, quizás a la luz como él lo dijo, ya de distintos clientes, los órdenes pueden ser cambiados.

Algo que no sé si dijo y se me escapó es el tamaño de muestra: 630 para el conteo rápido y 450 secciones electorales para encuesta de salida. No anoté con qué porcentaje de esas muestras es que salió a dar el resultado, que es algo interesante creo, en esta discusión de diseños a costos mínimos, más eficientes y cómo trabajar al cliente con muestras más pequeñas.

Yo creo que Alejandro Moreno, cuando dice cuál es el objetivo de encuestas de salida para medios impresos, pues el que él está haciendo uso, yo creo que para eso fueron diseñadas. No fueron diseñadas para estimar, como ya lo hemos dicho aquí, el resultado de la elección. Fueron diseñadas para lo que él usó y lo seguirá usando. Son para eso y siguen teniendo excelente oportunidad un día después, no son para estimar el resultado de la elección, sino toda esa riqueza que nos ilustró con algunas cuantas láminas.

**MARÍA DE LAS HERAS:** Una pregunta a los que trabajan los conteos rápidos en medios, ¿por qué no nos dicen participación cuando están dando resultados? Sería un detalle de su parte, si la tienen ahí. De las encuestas de salida (*exit poll*) también yo creo que

una de sus funciones, llevamos dos días platicando, en que no sabemos quién sale a votar, que parece que el problema de las encuestas electorales de saber cuál es la población votante, que cada quién tiene filtros diferentes, si los filtros funcionaron o no. Entonces Alejandro, yo no sé si tú probaste, tú nos diste varias hipótesis de qué podría hacer que tus filtros hicieran más exactas tus proyecciones electorales, yo no sé si tus encuestas de salida (*exit poll*), si yo tuviera una encuesta de salida lo que iría a probar es -si es cierto- si mi hipótesis sobre quién es el que sale a votar y quién no, realmente lo hace. Yo quisiera que me lo correlacionaras con lo que dijiste ayer.

Ayer decías, bueno usabas, qué tan seguro era que fueran a votar y la identificación partidista. No sé si lo estás midiendo porque no lo leí, y es interesante porque Edmundo Berumen, yo cuando digo que lo más seguro es que los seguros de mi modelo salgan a votar, y Edmundo Berumen siempre me regaña porque dice que yo la hice en una hipótesis. Yo no sé si esos son los que votaron o no. Pero ustedes sí lo pueden saber. Entonces, no sé si comprobaron alguna teoría de filtros con las encuestas de salida (*exit polls*), y sería sensacional partir de eso, si están viendo algo que vaya más allá de edad, sexo, que para mí no tiene mucha correlación.

Javier Alagón, un comentario. Sí, es cierto, con las encuestas electorales, y es mi teoría, se sobrestimaba la votación del PRI, y creo que a partir de aquí en adelante va a ser exactamente el problema al revés. La moda en este país es decir que se va a votar por el PAN, todo el mundo está diciendo que va a votar por el PAN. Entonces, ahora los filtros que ya encontramos y que ya los validaron, ya no nos sirven, hay que buscar otros porque ahora lo que nos está sobrestimado es el PAN, tienes toda la razón Javier.

Nada más una cuestión, la de reportar nulos o no en los datos que están dando. Yo no sé si ustedes recuerdan el problema de los votos nulos reportados en las boletas, aún en las actas, es una dificultad porque hay casillas donde el nivel de educación del que está llevando el asunto es tan bajo que no alcanzan a detectar la diferencia entre votos nulos y las boletas que ellos mismos cruzan, porque les dicen tienen que cruzar la boleta para que no se puedan usar. Entonces, de repente esas casillas que vienen reportadas donde votan más personas que el padrón, a veces la gran mayoría se explica por este tipo de prácticas. Entonces, si ustedes se dedican, con base en las actas, a poner nulos van a encontrar dificultades. Si las llevas a los medios de comunicación va a ser muchísimo más problema.

En cuanto al asunto de cuándo cambian o no cambian o cambiaron las preferencias electorales en el tiempo que había dicho Javier Alagón, yo insisto en lo que decía ayer, no se nos tiene que olvidar qué estamos midiendo y lo que estamos midiendo no es algo concreto, no es un vaso, es algo etéreo, y la opinión pública va reaccionando. Si yo mido en noviembre, después de que llevamos cuatro meses oyendo hablar de los candidatos del PRI todo el día, todo el tiempo PRI, PRI, Madrazo, Labastida, evidentemente lo primero que tiene la gente en la cabeza es el PRI y resulta que todo mundo es priísta y noviembre se vota, y después resulta que viene el debate donde hay una serie de conversaciones.

O sea, esos flujos no los puede uno perder de vista, no es un problema de meses o de lejanía de la elección, sino lo que está pasando en el entorno en el momento que estamos midiendo. Yo voy a decir una petulancia: hace un año, con una encuesta de diciembre de 1998, yo publiqué un libro donde doy escenarios electorales y digo: el PAN tiene una máxima de 13 millones, el PRI tiene una máxima de 13 millones y el PRD de seis. Hago tres escenarios, en el primero ganaba el PAN, en el segundo ganaba el PAN y en el tercero ganaba el PRI y no se lo creía nadie.

Yo no creo que no haya pasado nada de diciembre de 1998 al 2 de julio de 2000. Bueno, entre otras cosas, cuando regresé a medir al PAN, en lugar de 13 millones tenía 15, pero lo que sí creo es que en el pronóstico, ya viendo hacia la elección y hasta observando qué es lo que pasa y tratando de explicar las fluctuaciones durante una campaña, uno tiene que pensar en qué momento está midiendo, o sea, la volatilidad de la gente no cambia; no se duerme panista y se despierta perredista y al mediodía es priísta. Hay cambios, pero a veces los que nos están cambiando son esos que además ni siquiera van a ir a votar.

Los que realmente acuden a votar suelen ser, primero, personas más consistentes en su forma de pensar, pero son los menos los que van haciendo los pequeños cambios. Yo creo que uno de los problemas con las encuestas fue que por primera vez hubo una elección donde los medios intervinieron, donde todos los candidatos tenían todo el dinero del mundo metido en medios y que nosotros fuimos midiendo ese ruido que a lo mejor no eran tanto preferencias electorales y que en esos filtros que andamos todos buscando, hay que ver cómo quitar esos ruidos, porque eso hace que si fue en abril, que si fue en mayo, sinceramente pienso que Fox fue arriba desde el año anterior y que lo que se estaba midiendo era la salida de la primaria del PRI.

Yo lo pienso y lo publiqué, lo sigo pensando, o sea, ese cruce fue un efecto de camino de campaña, no realmente los votantes que estaban más probables a salir a votar. Yo creo que siempre hubo, desde antes lo dije cuando lo publiqué, más panistas que priístas.

**JORGE BUENDÍA:** Estoy de acuerdo que las encuestas de salida (*exit polls*) no se deben usar para predecir resultados, porque si se están usando para qué les sirve reportarlos con no asignados o indecisos o como sea. Si la diferencia de los indecisos es mayor que la diferencia entre los ganadores, entonces no sirve para nada. Si los anulados o indecisos eran 13 ó 14 y la diferencia era 10, tú estás diciendo: lo más seguro es que ganó quién sabe. Es lo que estabas diciendo en ese resultado, o sea, no están diciendo en el resultado quién va a ganar.

Lo que dijo ahora María de las Heras es muy importante –se acaba de ver– los republicanos estaban arriba y en un día cambió a los demócratas después de la convención. O sea, una noche te cambia los resultados, entonces dónde estás midiendo, estás tomando, se nos olvida que es una foto de un momento equis.

Los indecisos creo que es muy importante ver modelos, o sea, yo me declaro el primero que cometí un error en 1988, yo lo digo, yo los asigné proporcionales y estaba mal. Yo hice un estudio en esta elección y también hice, como María de las Heras, a lo mejor no ortodoxo, mi fórmula de filtrar y todo y tuve 41 por ciento en mayo. Por ahí, de repente, alguien lo empezó a circular. Entonces, sí hay que estudiar esos indecisos para ver otra forma de hacerlo, lo malo es que como dice María de las Heras, cuando ya no sabemos la respuesta, nos cambiaron la pregunta. Entonces tenemos que ir estudiando eso para más.

Y lo de la posición, nosotros hemos probado varias veces posición de la pregunta del voto al principio, en medio o al final y, por lo general, no tiene un efecto.

**FRANCISCO SÁNCHEZ:** Creo que la comparación de los resultados oficiales con los resultados del conteo rápido y con los de encuestas de salida nos proporcionan un instrumento muy útil para investigar cuáles son las fuentes de error de las encuestas. Ya se mencionaba hace rato que los conteos rápidos, básica o únicamente, tienen error muestral. Comparando el resultado oficial con los resultados de un conteo rápido, las diferencias las atribuimos a un error muestral, pero también si comparamos cómo, por ejemplo, en casos de celebración simultánea de conteo rápido con encuesta de salida, aquí las diferencias que vamos a encontrar son básicamente error no muestral, porque están asumiendo la misma muestra y yo creo que la compa-

ración que se puede hacer entre los conteos rápidos y las encuestas de salida pueden ser muy útiles para ahondar en los problemas de medición, que es algo que encontramos en este caso particular, selección de respondente, qué pasa con la no respuesta, es decir, aquellos que rechazan nuestra entrevista.

En ese sentido sí nos ha mostrado datos acerca de que los resultados de la encuesta de casilla son muy cercanos a los resultados globales, pero qué pasa a nivel casilla, cuál es la comparación, cuáles son las diferencias, los sesgos que encontramos entre los resultados de la encuesta de salida a nivel casilla con los resultados oficiales de esa misma casilla. Cuáles son las diferencias por segmentos, pudiera ser que efectivamente no encontramos sesgos a nivel global, a nivel nacional, porque se cancelan estas diferencias a nivel casilla. A lo mejor los estratos urbanos sobrerrepresentan a la oposición y los rurales sobrerrepresentan al PRI. El global, efectivamente, no encontramos grandes diferencias porque se cancelan estos sesgos.

También quisiera ahondar en un punto que se mencionaba. Esta es una pregunta a todos los que hicieron encuestas de salida: si han comparado cuáles son sus sesgos con los resultados oficiales de las casillas. Y en cuanto a la asignación de indecisos, en las encuestas de salida aquellos que no nos contestan la preferencia electoral son votantes, a diferencia de las encuestas preelectorales donde podemos asumir que muchos de ellos no van a votar. En las encuestas de salida los que no nos dan preferencia electoral son votantes y creo que aquí sí hay que modelarlos y quisiera oír cuáles son sus experiencias y cómo manejan ustedes la no respuesta en la preferencia electoral.

**ALEJANDRO MORENO:** Un poco en el mismo sentido que lo expresaba Jorge Buendía, las encuestas de salida (*exit poll*). Si bien no estamos buscando una estimación, porque eso lo vamos a conseguir con el conteo rápido, si nos da luz sobre qué efectos podríamos tener sobre algunos de los instrumentos que estamos usando para captar la preferencia electoral. Entonces, en ese sentido la falta de precisión en una encuesta de salida (*exit poll*) nos puede tratar de iluminar por dónde debemos de buscar los problemas que tenemos en encuestas preelectorales. Entonces, en ese sentido a mí me gustaría saber si a la vez que preguntaban la preferencia electoral para Presidente, si trataban de captar la preferencia para los otros tipos de elección federal: senador y diputados. ¿Por qué?, porque un poco se ha planteado que algunas de las diferencias son debidas a votos vergonzantes, voto de miedo, en fin, y lo que no dudo que puedan llegarse a dar.

Sin embargo, algunas experiencias que al menos en mi caso he tenido, es que a veces, incluso el partido que vemos sobrestimado es el PRI, que es el caso más común. El subestimado a veces es el tercer partido, cosas así, y pareciera reflejar más que a la hora que captamos la diferencia, era la identidad, la simpatía de la persona que entrevistamos y no tanto lo que fue a votar, porque hay que distinguir que en ocasiones el candidato puede hacer que su partido salga mucho más alto que el promedio de las otras votaciones.

**GUILLERMO BAZ:** Retomando un poco las oportunidades y el tiempo cuando se presentan los resultados. Yo tengo los resultados que aparecieron en la jornada del lunes, al menos los que quedaron registrados para la historia, realmente los que valieron, digamos, al menos para los medios y para todos los demás. Y ahí sí noto unas diferencias enormes contra los resultados finales. Yo creo que la hipótesis del voto vergonzante no es válida de entrada porque si uno ve otras encuestas de salida, la aproximación es buenísima. Está Televisa, Consulta, Presencia Ciudadana, Alianza Cívica, Covarrubias, que tienen muy buenas estimaciones, o sea, yo creo que difícilmente se puede pensar que una diferencia tan fuerte o los resultados que estás encontrando, pueden atribuirse a ese segmento.

Y quiero insistir un poco en lo de ayer, respecto a las encuestas de salida (*exit polls*) del encabezado de Alejandro Moreno, de Reforma, "ganó el México urbano y educado". Perfectamente cierto, tendrán, los que tengan encuestas de salida (*exit poll*), que trabajar justamente en eso para resolver el problema de toda la vida que es la sobreestimación del PRI antes de las elecciones.

**FRANCISCO SÁNCHEZ:** El comentario es el siguiente: a lo largo de todas las presentaciones y discusiones que hemos observado y escuchado, se ha hecho énfasis, prácticamente, en la única variable que aparentemente es la que importa, y es únicamente la proporción de votos para cada uno de los partidos. Las características del elector, las características demográficas son utilizadas únicamente, precisamente, subyugadas a esta variable de todas maneras, para ver de qué manera nos ayudan a efectuar mejores estimaciones.

Sin embargo, hay otras aplicaciones que a lo largo de este día no hemos escuchado, posiblemente un tanto cuanto soslayada en el tercer objetivo que mencionó el doctor Enrique Alzati, y es en cuanto a la calidad misma del ejercicio electoral. La propia ley lo prevé, pues hay diferentes momentos durante la jornada electoral, la apertura, el desarrollo y el momento del escru-

tinio, al menos estos identificados legalmente, etapas que son factibles de medición y, sobre todo, de identificar con un concepto de calidad a través de los muestreos que se realizan prácticamente simultáneos a la encuesta de salida o conteos rápidos, con el mismo personal y hay antecedentes. En 1994, Manuel Román y yo tuvimos oportunidad de colaborar con la organización nacional de observación electoral del magisterio, de la cual se conoció únicamente un conteo rápido que, para nuestra fortuna, en aquella ocasión fue bastante preciso, y los resultados de la observación electoral, ellos estaban registrados como organización no gubernamental que podía entrar a la casilla como consecuencia, que son la hora en que abren la casilla, los materiales que van a integrar esa casilla, tinta, urnas, elementos de identificación, presencia de propaganda a lo largo de la jornada, grupos de electores que puedan llegar de manera sospechosa por ahí, etc.

Todos estos aspectos fueron captados, aunque desafortunadamente no difundidos, y creo que la mayor parte de los ejercicios a los que se han referido este día, no hacen mención de esta otra factible utilización, y hasta lo que yo me doy cuenta, poco desafortunadamente, utilidad que el propio instituto teniendo sus observadores o bien entrevistadores dentro de la propia casilla, podría realizar con un esfuerzo marginal relativamente fácil y lograr otra faceta que estos mismos ejercicios hagan mucho más complementaria y nos quiten de una visión extraordinariamente estrecha, que es la que siento que hasta este punto nos ha normado.

Y la pregunta concreta es si ese tercer elemento que también soslayó el doctor Enrique Alzati, correspondería a este tipo de situaciones o a qué se refería cuando hace mención de un objetivo de calidad.

**RAFAEL GIMÉNEZ:** Me ha parecido muy interesante ver cómo ha evolucionado la discusión de encuestas de salida y conteos rápidos desde 1994. Aspectos muy específicos, de hecho yo encuentro un aspecto muy específico en lo que hemos visto hoy, que es cómo manejar la información incompleta durante el momento en que se necesitan tener los resultados, en el que se necesitan presentar los resultados y básicamente para el caso de conteos y de encuestas, tenemos la situación de muestras incompletas y en el caso de las encuestas de salida hay que agregarle la no respuesta. Cómo manejar esta información en el periodo de tiempo en que los clientes, los que pagan estos ejercicios, requieren la información. Creo que es ya muy centrado el tema, me parece que eso ha sido lo que hemos visto hoy, quienes quisieron presentar algo valioso, porque creo que

hay mucho sobre cómo hicieron las muestras, y creo que ahí ya han avanzado mucho los muestristas, sobre todo, de conteo rápido.

Yo agregaría un aspecto más. A mí me parece que están enormes las muestras de los conteos rápidos y de las encuestas de salida, y creo que difícilmente en una situación de más normalidad democrática digamos, alguien pagará esas muestras. Entonces yo creo que deberíamos de empezar a centrar aparte de esta situación de información incompleta, la reducción de los tamaños de muestra para tener ejercicios de esta naturaleza que sean viables, porque si no, no van a ser viables muy pronto.

Encuentro la posibilidad de que nos concentremos también en estudiar este asunto de los híbridos que mencionaban en algún momento. Los híbridos entre encuestas de salida y conteos rápidos para tener más precisión en la información, que me da la impresión que eso es lo que hace Consulta, pero Consulta Mitofsky, sabiendo que eso es lo que hace Mitofsky en otros lados donde sí ha presentado su información.

Me parece que ahí están los temas de las encuestas de salida y de los conteos rápidos que sería interesante empezar a conocer un poco más. La otra parte que mencionaba Javier Alagón es a qué se deben las diferencias entre encuestas electorales. Noto que ahí sí tenemos muchas más divergencias en los aspectos técnicos. Noto que casi ya lo de los conteos rápidos y encuestas de salida es un asunto logístico, más estos temas que acabo de mencionar, pero en el de las encuestas preelectorales, bueno hay diferencias entre las encuestas preelectorales, más allá de lo que los errores muestrales reportan para las encuestas simultáneas, las 12 ó 13 encuestas simultáneas finales.

Entonces ahí quiere decir que no nada más está siendo un asunto muestral, un asunto estadístico, pero también algo me quedó claro, es que tres o cuatro variantes en el diseño de muestra, la utilización de marcos muestrales, tres o cuatro variantes en el asunto de los filtros, cuáles filtros, tres o cuatro variantes en la posición de la pregunta electoral, una gran variación en la tasa de no respuesta o de indecisos, o ni siquiera tenemos bien claro lo de la tasa de no respuesta total en una encuesta.

Entonces me parece, con todas esas, tenemos una matriz inmensa de diferencias en este caso, o de diferencias entre las encuestas preelectorales, que fácilmente empiezan a perfilar que por ahí están los problemas, que en cuanto encontremos esos elementos comunes, vamos a ir viendo cómo mejorar esto, sobre todo, porque no fueron iguales las encuestas preelectorales que se hicieron al mismo tiempo.



La última cosa, perdón, no me parece este comentario de que las encuestas de salida no sirven para estimar resultados electorales precisos, si no, no veo una de las utilidades principales de estas encuestas, es decir, estamos pidiéndole a las preelectorales que lo hagan y a las de salida no. Yo creo que tienen que ser mucho más precisas que las preelectorales.

**JUAN MANUEL HERRERO:** Bueno, siendo relator que alguien me sustituya durante este tiempo. En estos dos días surgió una competencia, ¿cuál es el conteo rápido menos plublicitado? Y yo me quiero subir a esa competencia y además creo que se las gana. Mi cliente para nada quiso dar a conocer los resultados, lo cual es totalmente su derecho.

Sin embargo, yo creo que sí se vale hacer algunos comentarios que creo que son pertinentes. Voy a tomar una parte del comentario de Rafael Giménez cuando dice: "ya discutimos mucho la muestra, ya discutimos mucho de esto" pero hay algunas cosas que yo insisto en que no se han discutido lo suficiente, a lo mejor aquí cada quien viene a sacar por delante sus deformaciones profesionales. Entonces, como aquí nos han ganado los muestristas, nos llevamos un buen rato discutiendo de muestras y, bueno, a algunos nos gusta más analizar otra parte de este tipo de ejercicios, como es las cuestiones de campo.

Al comentario que hace Jorge Buendía, precisamente nosotros, Jorge, sí hicimos un ejercicio de evaluación de qué tamaño es nuestro error de campo, porque lo podemos medir en el conteo y en la encuesta de salida y de hecho a nuestros 32 enlaces les entregamos su reporte, diciéndoles, sección por sección, en dónde hubo algún problema, de qué tipo era el problema, y nosotros internamente nos beneficiamos de saber que nuestra muestra de pronto no tuvo el problema, salió bien, pero me refiero, las divergencias no son producto del efecto del diseño de la muestra, cuando menos tengo que quitar esta parte que ya conocí, y que es un efecto real y que es un efecto de campo.

Entonces básicamente mi comentario, si, sí hicimos eso, sí nos parece valiosísimo hacerlo porque nos ha permitido ver algunas diferencias en el comportamiento de los operativos, muy interesante, eso no nos llevó a tener una estimación equivocada, sin embargo, nos permite entrar al análisis y discutir en su dimensión otro tipo de componentes de ejercicios de esta naturaleza.

**ANA CRISTINA COVARRUBIAS:** Si me permiten yo daría contestación a alguno de los temas que atañen a la presentación hecha por José Mendoza y le dejaría las partes estrictamente técnicas y de muestreo a él.

La primera cosa que quisiera contestar y que de hecho está interconectada con otras preguntas es la observación que hizo Edmundo Berumen, en el sentido de que cómo es posible que a las 13:00 horas tuviéramos -no me acuerdo de las cifras- pero era aproximadamente 85 por ciento de la muestra, 26 mil encuestas y que dos horas después tuviéramos 90 por ciento de la muestra y ya hubiéramos aumentado considerablemente.

Creo que es obvia la contestación y es muy sencilla, lo que quiere decir y por eso lo pusimos, es que tuvimos una eficiencia muy grande en el trabajo de campo y ya que quieren que ese tema se toque un poquito, explicaré sobre él. Quiere decir que a la 13:00 horas el 85 por ciento de la gente que teníamos apostada en cada una de las secciones electorales, ya se había reportado dando la cantidad casi exacta de encuestas que le habíamos pedido que levantara. Nosotros le habíamos pedido que levantara seis por hora, iba a trabajar ocho horas seguidas, de 09:00 a 17:00 horas. Entonces eso nos daba 50 por persona pero con la consigna de, no se vale que hagas 50 en tres horas en la mañana y te vayas a tu casa y después regresas para el conteo rápido, no, o sea, tú te tienes que estar ahí todo el día porque tienes que hacer seis por hora.

Esos datos justamente dan el testimonio de eso, de que traíamos 26 mil porque ya se había reportado a las 13:00 horas el ochenta y tantos por ciento de la muestra. Los aumentos posteriores de muestra son irrelevantes en el asunto de la encuesta de salida, ya que finalmente llegamos al noventa y tantos por ciento de la muestra. Se está hablando de la eficiencia del trabajo de campo, pero si nos hubiéramos quedado con el ochenta y algo, de todas maneras traíamos ya un nivel de eficiencia muy grande en esta materia.

Entonces esa es la explicación. Ahora esto se conecta con los otros datos que decía Mundo y que también de alguna manera Juan Manuel Herrero ha mencionado, de cuéntenos un poquito más de la cosa de campo. No se reserven ese conocimiento técnico adquirido (*know how*). Creo que la intención nunca fue, tenemos 20 minutos para hablar, y quisimos hablar de lo que considerábamos que era más relevante, pero bueno, tenemos ahorita la oportunidad. Cuántos encuestadores traíamos, los mismísimos, uno por sección electoral y en las áreas rurales traíamos, sobre todo en comunidades alejadas, dos. Esto independientemente de la supervisión.

La persona que estaba levantando la encuesta sí tenía que suspender cada 45 minutos o a criterio podía ser cada dos horas, o sea, tenía que transmitir cada hora o cada dos horas. La realidad es que como el pro-

medio que tenía que levantar era seis, pues no tenía mucho sentido que lo hiciera cada hora, mejor que lo hiciera cada dos horas, llevando 12 encuestas por transmitir. Entonces ese es el número. Ahora, en las áreas rurales sí es muy importante que vayan dos porque la distancia para el teléfono más cercano es muy grande y entonces no es posible porque se tardan 45 minutos de ida y 45 minutos de regreso y no puede estar desatendiendo su puesto de estar levantando encuestas. También tiene que ver con esto de los costos marginales que se mencionó algo en la sesión en la Mesa 4, que sí es mucho más caro hacer las dos cosas que nada más hacer el conteo rápido. No es muchísimo más caro, es un costo verdaderamente marginal porque el verdadero costo está en poner en cada casilla y, sobre todo, en cada sección electoral y a nivel nacional, el estar poniendo ahí a la persona adecuada para que cumpla esta función.

Además hemos visto en las experiencias, no sólo esta vez sino de los otros que se puede a través de la capacitación crear una mística de responsabilidad en los encuestadores que hacen este trabajo. O sea, que es producto de una buena capacitación en donde se hace mucha conciencia y mucha motivación de la trascendencia del trabajo que se está llevando a cabo. Y, además, los encuestadores están mucho más involucrados si permanecen fuera de las casillas todo el día a que si sólo van a recoger los datos del conteo rápido, después de las 18:00 horas. Hacer encuestas de salida y conteo rápido es lo más conveniente y el incremento del costo es marginal.

Por otro lado, lo que mencioné en la mañana, ante lo que había comentado María de las Heras, el asunto de tu respuesta a la autoridad electoral, por encima de cualquier cosa, pero sí es posible hacerte amigo y que te admita la autoridad electoral, pues trata de estar dentro de la casilla para que rápidamente transmitas la información. Eso se da porque está todo el día ahí.

Entonces, creo que esas son las observaciones y quisiera también contestar en relación a la cuestión esta, plantear una reflexión, un debate sobre eso que hicimos nosotros y que no se había hecho antes que es la cuestión de reportar la encuesta de salida dando la no respuesta. Creo que esto da para comentarios a favor o en contra. Nosotros lo hicimos por un prurito ortodoxo de seguir estas normas internacionales, pero está abierta a la discusión y también por el hecho de que siempre va a haber este cuestionamiento de si alguna manera se le debe de dar un tratamiento o no, o simplemente como se hace la mayoría de las veces, asignación proporcional y borrón de la no respuesta.

Esto, yo no tengo la respuesta, es lo que nosotros

hicimos y es tema de reflexión y de debate. Y, por último, en cuanto a la observación de Rafael Giménez, de las muestras grandísimas. Sí es cierto, la muestra que nosotros trabajamos fue grandísima. Nosotros siempre hemos estado por la presentación de muestras más pequeñas, sin embargo, en esta ocasión puedo decir que la razón fundamental por la que se usó una muestra así es porque se esperaba una competencia sumamente reñida y se requerían niveles altos de precisión. Esa fue la razón por la que el propio cliente pidió, no es que nosotros se la propusiéramos, una muestra grande. Además, debemos admitir que si bien la contienda no fue lo cerrada que se esperaba, la muestra grande tuvo una utilidad práctica importante, que fue la velocidad con la que pudimos responder tanto en el caso de la encuesta de salida como en el del conteo rápido. Porque 26 mil encuestas levantadas a las 13:00 horas, es producto de traer una muestra del tamaño que tuvimos.

Ahora, que para qué sirve tener 26 mil encuestas a las 13:00 horas si no las puedes publicar hasta las 20:00 horas, te puedes ir más lento. Bueno, pues sí tuvo utilidad práctica según me han comentado porque en los medios hay una comunicación con partidos políticos, con autoridades, etc., y desde las 13:00 horas los candidatos, así me lo ha manifestado la gente que estuvo cercana a ellos, aproximadamente 13:00 o 14:00 horas, supieron como iban cada uno de los candidatos en función de esta información.

*JOSÉ MENDOZA:* En las simulaciones que di, dos mil simulaciones fueron realizadas para cualquier tamaño de muestra, o sea, tomé muestras de tamaño 100, respeté dos mil veces el proceso de tomar la muestra y hacer las estimaciones, y la estimación tenía que ver con los porcentajes estimados para todos los partidos, y sobre las dos mil muestras se puede calcular cualquier tipo de error. Por ejemplo, el error de estimación de un partido respecto al verdadero valor podemos, sobre cada una de las dos mil muestras, hacer las diferencias, elevar al cuadrado y promediar.

Lo que hice al final es tomar sobre cada partido la estimación y tomé el máximo error. Es como cuando hacemos el error con  $p$  igual a 0.5, es el máximo error, lo que hice con todos los partidos, calculé el error y tomé la estimación máxima.

Efectivamente, podemos calcular la varianza exacta. En muchos muestreos en dos o tres etapas se empieza a complicar. Yo creo que en particular, cuando se toman muestras sistemáticas y hacemos una selección, sólo una selección, estrictamente no podemos estimar la varianza, la propuesta que se hace en muchos de los

textos es partir la muestra en varias sistemáticas y estimar la varianza. Para efectos de la simulación no tengo que hacer eso, simulo muchas veces una extracción en una sola muestra y calculo el valor, pero por supuesto que para efectos de una sola realización, si la voy a hacer sistemática, podría ser conveniente partir en varios muestreos sistemáticos.

Mi intención es con esto, sólo estimar la tasa de error y al tomar muestras sistemáticas y repetir el proceso estoy calculando un estimador del parámetro poblacional, el error que tendría ese método de estimación respecto a los valores verdaderos.

Un comentario adicional de lo que ha planteado el doctor Edmundo Berumen también durante la tarde. Yo creo que todos estamos de acuerdo en que estimar totales de votación no es problema, ya se haya usado PPT, aleatorio simple, cualquier otro método. El problema es cómo lo convertimos a porcentajes. En algunos casos los totales de votación ya suman la lista nominal, por ejemplo en el PPT; en otros como aleatorio simple no necesariamente suman la lista nominal. Al convertirlo en porcentajes entonces estamos, efectivamente, como dice el doctor Edmundo Berumen, usando estimados de razón que yo tampoco veo en todos los planteamientos que se han hecho de fórmulas.

Sin embargo, yo creo que si nos preguntamos cómo le haría el IFE para eliminar la abstención, los no registrados, los nulos, pues simplemente los quita, suma y hace cocientes. Lo mismo podemos hacer nosotros si tenemos totales. Por ejemplo, si nos dieran la tarea de estimar sólo dos partidos, digamos PAN, PRI, cuál es el porcentaje que tiene cada uno, pues los sumamos, tomamos los votos de uno y dividimos entre el total de los dos, y ese es un estimador que tiene un sesgo pequeñísimo de acuerdo con mis estimaciones, es prácticamente nulo, aunque como bien anotaba el doctor Edmundo Berumen, si trata uno de obtener matemáticamente el sesgo, se mete en una serie de problemas, no es fácil demostrar que el sesgo es pequeño.

Sin embargo, las varianzas de esos cocientes son perfectamente calculables, es una varianza en un estimado de razón, la varianza, por ejemplo, en el caso que decía de un partido sobre el total de dos partidos es una varianza de la forma usual de un estimado de razón, no le veo ningún problema. De hecho yo usé ese método en las simulaciones, quité la abstención y los votos nulos y en todos me salió que el sesgo era despreciable.

Con respecto a la pregunta de María de las Heras, estoy totalmente de acuerdo. Todo conteo rápido debería dar una estimación de la participación, lo cual no se hace, tiene que ver con los formatos de publicación, así como creo yo que tienen que ver con publicar la no

información en los conteos de salida, creo que debía de ser también un estándar, en el sentido que hablaba Ian Reader, nosotros traíamos 11.7 por ciento de no respuesta, la diferencia entre primero y segundo lugar andaba entre siete u ocho puntos porcentuales, qué pasa si los indecisos, los que no informan, no se distribuyen proporcionalmente. En las tablas que yo di simplemente los quité y es como asignarlos de manera proporcional, pero si no fuera el caso, esa diferencia que anotábamos, como significativamente entre primero y segundo lugar, podría no serlo.

Si alguien usa un modelo de asignación que no es el proporcional, entonces tendrá que usar parte de ese modelo para hacer también el cálculo de las varianzas y las pruebas. Si usa una asignación proporcional es fácil porque de nuevo vuelve a ser un estimador de razón, en el que es fácil, si no, tiene que ver hasta la calidad del modelo, el cálculo de errores y, por tanto, la declaración de ganador o no.

Creo, a diferencia de los conteos de salida, en este caso sí podemos verificar el supuesto porque no podría haber cambiado un voto saliendo de la casilla y en cuanto nos dice sí y nos dice PRI, yo creo que deberíamos de creerle. A diferencia de cuando tenemos una encuesta tres semanas antes, en el que pudo haber cambios que provocaron un viraje en la decisión del votante. En el caso de la encuesta de salida, como tenemos los datos reales y tenemos los que sí respondieron, es fácil por diferencias de distribuciones, ver si es un supuesto razonable el asignar de manera proporcional. Creo que no hay ningún espacio de tiempo entre la salida de la casilla y el momento de la encuesta, como para que digamos, hubo un suceso importante ahí que cambió su decisión.

Respecto al comentario de Francisco Sánchez, él anotaba que a lo mejor el tener esas encuestas de salida y características demográficas serviría para mejorar la calidad de las estimaciones. Estoy totalmente de acuerdo, pero creo que más que de las encuestas de salida también debería considerarse el banco del IFE, que tengamos 10 ó 20 mil entrevistas las compañías, creo que el IFE tiene todo capturado, podría ignorar nombres y direcciones y tener una base excelente para realizar ponderaciones en conteos de salida. Pudiera ser, me imagino yo, por ejemplo, no he anotado y en ese sentido pensaba a raíz del comentario de Francisco Sánchez, que los que no informan sean más hombres que mujeres, eso no lo sé y sabemos que votan distinto. Entonces, eso podría ser un modelo que nos permitiera estar -incluso- corrigiendo de manera más adecuada, que simplemente redistribuir.

Y un último comentario haría. En muchas definiciones de tamaño de muestra que me ha tocado ver, usamos una cosa que se llama efecto diseño (*Deff*), y tengo la impresión que lo usamos cuando el diseño de muestra es suficientemente complejo, dos o tres etapas donde decimos la varianza quién sabe como se calcula, pero saco una estimación de una varianza en un muestreo simple con 100 unidades, calculo el *Deff* y multiplico la "n".

Mi experiencia en ese sentido es que cuando uno grafica varianzas contra "n", efectivamente van decreciendo, pero dos métodos no tienen por qué ser paralelos, o sea, pueden ir cerrándose y cuando uno hace un cociente de varianzas para una "n" de 100 y dice, tengo que doblar la muestra, a lo mejor para la "n" de 500 era multiplicar simplemente por 13. Entonces, mientras no sean estrictamente paralelas las curvas, el tomar un efecto diseño (*Deff*) a 100 y el tomar un *Deff* a 200 puede dar resultados muy diferentes y en mi opinión sólo sirve el *Deff* en la medida en que el cociente de varianzas sea igual al cociente de los tamaños muestrales, multiplicado por cualquier factor. En ese caso podemos despejar una "n" y sustituirla de manera adecuada.

**ROY CAMPOS:** Al doctor Edmundo Berumen. Solamente te faltó decirle que era infértil explicarle, pero voy a tratar de responder algunas cosas, a lo mejor no en el orden.

Estoy de acuerdo con Rafael Giménez en el razonamiento de que una encuesta preelectoral de tamaño más pequeño cuando tenemos que aplicar filtros o asignar indecisos y nos exigen dar resultado y una encuesta de salida en donde levantamos más en ese momento, recién votados, por qué no va a dar resultados, o sea, no tiene la lógica pues tendría que darlo, de hecho lo está dando. Sin embargo, el argumento era la encuesta de salida en sí es como valor noticioso, tan da datos que luego resulta que son bastante precisos. Estoy de acuerdo que lo puede hacer, la cuestión aquí era qué se anuncia. Si seguimos anunciando números, vamos a seguir anunciando números, no vamos a convencer a los medios que ya no les vamos a dar números, pero que para nosotros sería mejor que no nos pidieran números, que les diéramos ganador y ya, pero bueno, si estoy de acuerdo contigo que no tiene lógica exigirle a la preelectoral y aparte precisión y no exigirle a las encuestas de salida.

Francisco Sánchez hizo una observación electoral. Nosotros, incluso como instrucción, le decimos al encuestador ni te involucres, aléjate, no veas de qué partido es el que le pegó a qué partido o qué partido fue el que

tomó la urna y corrió. No hacemos eso, al contrario tratamos de no mezclar y no hacemos observación electoral, que pudiera servir –bueno– pudiera servir porque al fin de cuentas tenemos 300, 500 ó mil encuestadores que podrían hacer labor además de observación electoral, en esos mil puntos.

Alejandro Cruz, sobre si medí otro voto. Yo no medí voto por diputado ni por senador, pero en los tres estados que tenía dominio, además de presidente tenía gobernador, entonces ahí sí hay un voto diferenciado. Ayer, por ejemplo, dio una cifra muy parecida Enrique Alduncin. En el Distrito Federal, si uno analiza los datos agrupados, puede decir que hubo cruzamiento de algo, pero yo en el Distrito Federal, 22 por ciento de los votantes que entrevisté, tenían diferente voto en términos de partido para Jefe de Gobierno y para Presidente, o sea, en el DF hubo una diferenciación de 22 por ciento, en Morelos 17 por ciento y en Guanajuato de 10 por ciento. Y no estoy diciendo qué autoridad arrastró a cual, no estoy diciendo que Fox arrastró el voto hacia Creel o que López Obrador arrastró el voto de Cárdenas, lo único que se encuentra es que es voto diferenciado. Lo demás son hipótesis que luego prueban, ahí Francisco Abundis y Jorge Buendía con algunas preguntas de para usted cuál es la autoridad que más tiene, o algo por el estilo que ustedes hacen, pero bueno, esto simplemente muestra que hay diferenciación de voto.

María de las Heras, que hizo más preguntas, lo de la participación. Efectivamente, aquí tenemos un problema, a las 09:30 horas viene el cliente y dice, cómo está la participación, pues bien, es una cosa de mucha presión, que quieren que demos la participación. Entonces primero, con encuesta de salida (*exit poll*) no se puede, definitivamente, apenas que lleves un conteo de cuántos entran y salen con lo negrito en el dedo, pero con conteo rápido se podría dar. Al momento cuando ya está el conteo rápido, esa cifra, digamos que no les importa mucho, a algo aproximado, pero ya lo que querían era la noticia de ganador.

Y la otra se puede dar cuando siempre en cada elección tenemos el listado nominal que no en todas las elecciones podemos conseguir el listado nominal actualizado. Si en todas las elecciones lo tuviéramos, sí, si no, no podemos hacerlo, listado nominal, por sección actualizado. Bueno cuando lo tengan, ahora sí podíamos.

Sobre el uso de filtros, efectivamente, uno de los usos de la encuesta de salida es tratar de buscarle qué, otro es lo que nos va a presentar Alejandro, en la Mesa 6, el panel, ahí sí efectivamente aquellos que nos dijeron muy probablemente voté, votaron, porque ahí preguntarle de salida, oiga a usted cuando le pregunta-

ban hace un mes, creía que era muy probable que iba a votar. No, tienen que ser filtros más concretos, por ejemplo, te puedo decir que de los 100 puntos, 11 fueron de votantes que habiendo tenido edad, nunca se habían presentado a votar, fueron convencidos a ir a votar cuando ya habían tenido oportunidad de votar antes. Esos, por ejemplo, en el filtro de Greenberg se les hubieran ido, o sea, en ese filtro se les hubieran ido, bueno, fueron 11 puntos de los 100, ese tipo de cosas sí se pueden encontrar en las encuestas de salida (*exit poll*). Se van publicar.

Sobre la publicación de nulos, efectivamente, la diferenciación en las casillas y en nuestros encuestadores, por más capacitaciones, siempre hay problemas con la diferenciación de nulos, no registrados, en blanco, una serie de cosas. Entonces, por eso nosotros reconsideramos siempre a 100 es imposible.

Al doctor Edmundo Berumen. Cuando dimos los resultados de la encuesta de salida (*exit poll*) a las 20:00 horas, en términos de cumplimiento de muestra era el 95 por ciento de las casillas, que habían reportado de conteo rápido, seguramente era menos del 30 por ciento, o sea, dimos en ese momento. Después quisimos corregir el conteo rápido con el 60 por ciento, se movía un punto la estimación y ya Televisa no nos tomó en cuenta, ya no la difundió. No les importaba la precisión, que nosotros nos peleemos, pues bájale un poco, no, no, ya dimos ganador y aquí son las entrevistas y que vengan y el panel y ya estar corrigiendo datos eso es lo de menos para ellos, eso es lo que a mí me importaría; pero no es su agenda, es mi agenda ahí ese momento, la agenda de ellos es ya estar dando noticias y haciendo entrevistas.

Y, por último, lo de Javier Alagón, es la misma cuestión que yo hice ayer y que creo que es muy importante en qué momento se cruzaron. Hay alguna evidencia que por lo menos durante junio ya iba ganando Fox, en mayo, aparentemente, según las encuestas de salida, ya iba ganando Fox y según las encuestas de rotación (*tracking*) de Rafael Giménez y según las encuestas de Rafael y de Gauss, traían esos datos. En abril no es tan claro y yo creo ahorita la discusión es si en abril ya o todavía no, y si en abril es después del 25 o después del debate. Entonces, la pregunta yo creo que es, por qué en mayo todavía y en junio todavía no lo detectamos.

Sobre lo que dice María de las Heras, ese es un argumento que según el libro hubiera ganado un candidato con el carisma de Fox en cualquier momento en que hubiera sido la elección en febrero, marzo o abril. Está bien, yo creo que sí, tal vez, pero aquí la cuestión es, ¿qué debieron haber medido las encuestas en ese

momento? Es decir, a lo mejor si es cierto que las condiciones estaban dadas para que los votos que hubiera llevado un partido u otro ya hubiera ganado la oposición, pero bien, ya lo debieron haber captado las encuestas, ya debieron los ejercicios demoscópicos ver, es en enero, en febrero haber captado que iba a ganar la oposición. Mi hipótesis es que no, que ni siquiera las estrategias de encuesta del PAN por lo menos no la llevaban, las encuestas en general tampoco lo llevaban.

Yo creo que abril es el punto de discusión, si en abril ya estaba arriba Fox o en abril estaban empataados, y según las encuestas de rotación (*trackings*) de Rafael, depende de los días que los movieran, depende de la pauta que tuvieran los comerciales negativos o algo, podía uno haber captado algo arriba o algo abajo, pero yo creo que en mayo ya se debía haber captado que iba arriba Fox.

**ENRIQUE CALDERÓN:** En relación a las preguntas o las referencias que se hicieron a mi intervención, yo lo que quisiera decir es que en la elección del 2 de julio hay algo que se ha perdido y es que se da la impresión que las elecciones fueron totalmente limpias y en esto fueron totalmente distintas a todas las elecciones anteriores, y nuestra impresión es que no. Nuestra impresión es que en el sentido de votos acarreados, votos comprados, etc., las hubo. Simplemente con el triunfo de Fox no fueron tan relevantes, o sea, la sociedad no compró este tipo de discurso, pero lo hubo.

Lo que yo quiero insistir aquí es que la encuesta de salida es un elemento que no permite comprobar, pero sí permite de alguna manera establecer el tamaño de este voto y en qué parte se da. Yo quisiera decirles, en los resultados que les mostré, es que casualmente es en el estado de Tabasco en donde nosotros detectamos una diferencia entre la encuesta de salida y lo que dice el IFE, y esto para el caso del PRI es de 11 por ciento. Entonces, lo quiero dejar ahí, esto no es una prueba de que hubo fraude ni mucho menos, pero pienso que para los partidos políticos el tener instrumentos que les permitan medir la capacidad de manipulación es algo que es valioso y es algo que la encuesta de salida puede decir.

Creo que efectivamente al medir el número de no respuestas, vamos que se negaron por parte de los entrevistados, da una mejor aproximación. Sin embargo, también me da la impresión que esto se da a distintas horas del día y muy especialmente al final de la jornada, y en este sentido uno no va a medir con encuesta de salida lo que sucedió en los últimos momentos, pero la diferencia entre la encuesta de salida y los resultados formales, sí permiten de alguna manera establecer

una diferencia entre el voto que aparece registrado en actas, con el voto que la gente quiso hablar de él y pudo hablar de él sin ninguna coacción.

Creo que también es importante señalar que en nuestra encuesta de salida hubo zonas muy claras en donde se habla de un voto de miedo. Les puedo señalar el caso de Chiapas en donde el PAN queda arriba en la encuesta de salida respecto al resultado oficial del IFE, mientras que el PRD queda abajo en la encuesta de salida respecto a esos resultados, lo que nos está marcando es que el voto del PAN viene de zonas urbanas o de zonas donde la gente no se sentía víctima de la violencia, cosa que sí sucedía en las zonas perredistas.

Entonces, sólo quiero dejar sobre la mesa que la encuesta de salida puede ser un instrumento de análisis muy importante para estudiar la calidad de la elección.

**ALEJANDRO MORENO:** Javier Alagón, nuestras muestras preelectorales todas son independientes y, por supuesto, también la encuesta de salida (*exit poll*). De hecho, una de las cosas que mencionabas me llamó la atención y haría la pregunta a la inversa, de hecho abrirla es un tema un poco fuera de lo que estamos discutiendo, pero me parece un planteamiento muy interesante, sobre si la mano invisible del mercado va o no a ajustar los defectos que tenemos en la industria del gremio. Mi pregunta sería abierta dado que yo no estoy, y no he formado mucho parte de la comunidad encuestadora, buscando clientes sino que más bien trabajando para un medio que produce sus propias encuestas.

Esto ha funcionado en el pasado, o no ha funcionado, y si no, ¿por qué?

**EDMUNDO BERUMEN:** Los detalles metodológicos sobre encuestas de salida, a lo mejor ya no tengo mucho tiempo para ello. Me gustaría saber si hay algo en específico, ya que comentas que nuestros encuestadores se prestan y que el trabajo es más o menos común, si no te ha llegado nuestro manual de la encuesta de salida (*exit poll*) te lo mando esta semana, no viene nada técnico, viene simplemente qué debe hacer el encuestador. En el caso del 2 de julio tuvimos un encuestador por casilla, de hecho esto quizá se diferencie de muchos de ustedes que están a veces a la merced del cliente y de cerrar contratos y demás.

Nosotros prácticamente desde hace un año sabíamos que íbamos a hacer una encuesta de salida (*exit poll*). Nos fuimos organizando poco a poco, el día de la elección salvo un incidente que fue algo muy sencillo, todo bien, sin ninguna preocupación, salvo que en cuanto empezaron a llegar cuestionarios nos dimos

cuenta que había capturistas que estaban capturando el DF en la nacional, la nacional en el DF. Nos dimos cuenta porque son los cuestionarios de tamaño diferente, hay muchas otras maneras de darse cuenta. Pero me imagino al día siguiente, gana Cárdenas la presidencia y con una impresionante participación rural, Creel se lleva el Distrito Federal.

Pero bueno, lo corregimos a tiempo, y nada de preocuparse. Tuvimos otro percance dentro de este largo proceso. Teníamos previsto utilizar teléfonos satelitales para zonas rurales, fue un convenio que se cayó por alguna razón que desconozco y tuvimos que usar cosas de emergencias, en zonas rurales emplear dos o tres personas. Pero dado el esquema que utilizamos y que ya mostré fue también el campo muy sencillo. De hecho, este número quizá va a sonar muy alto, pero dado nuestra supervisión electoral se pudo supervisar 75 por ciento de la gente y no implicó mucho mayor esfuerzo, sabíamos a qué hora el encuestador iba a estar en su casilla y a qué hora no para ir a visitarlo.

Entonces esta cuestión de jugar con los horarios nos hizo que todo funcionara relativamente bien en general, como experiencia organizacional. Ahora nuevamente aquí está el manual. Mencionabas una cosa más acerca de la riqueza de análisis que nos permite tener una encuesta de salida (*exit poll*) y aquí no se ha hablado de lo que sí se habló ayer, del tamaño del cuestionario. ¿Qué conviene en una encuesta de salida (*exit poll*)? La verdad, en nuestra experiencia y dado que hicimos muy pocas entrevistas por casilla, hicimos un cuestionario que a lo mejor es inusualmente largo. No recuerdo las preguntas porque a diferencia de nuestras preelectorales las de salida (*exit poll*) para que sea fácil las numeramos por letras, pero teníamos un par de hojas de preguntas más una de datos sociodemográficos.

En este respecto yo estoy de acuerdo con Rafael Giménez en que sí se debería de hacer un esfuerzo por reducir los tamaños de muestra, pero siempre hay algo que hay que evaluar. Por ejemplo, nosotros ya empezamos a reducir los tamaños de muestra en las electorales. Nuestra serie nacional preelectoral tuvo una constante de mil 500, mil 520, salvo la de abril en que hicimos un diseño con motivo del Día del Niño, el 30 de abril, un diseño en los hogares en donde hicimos las encuestas de los adultos le siguió una encuesta de niños, al niño que habitaba ahí, ya sea el hijo o cualquier niño.

Entonces, tenemos de alguna manera un diseño que nos permite evaluar las opiniones de los niños en su ambiente de socialización. Entonces, en ese caso como queríamos tener suficientes niños, dado que no sabíamos ni teníamos idea de la tasa de no respuesta por los rechazos, pero nos subimos a mil 650, es el único salto

en tamaños muestrales en nuestra serie, pero el resto de los diseños es prácticamente el mismo.

En el caso de Chiapas que ya hicimos después del 2 de julio, ya redujimos la muestra con 800 casos efectivos. Fue una muestra relativamente buena. Al final no estimamos la diferencia tan reducida que se dio dadas las expectativas que teníamos, pero creo que fue una encuesta bastante cercana y, en este sentido, los cuestionarios sí nos dan esa riqueza que como medio de comunicación necesitamos. Incluso, también, para trabajo académico pudieran ofrecer bastante. Me remito a lo que dijo Ulises Beltrán, que a lo mejor el último cuestionario preelectoral debería ser muy cortito, para que el fin sea realmente la estimación, por ejemplo, la experiencia de nosotros en el Estado de México que fue en nuestra última encuesta y, de hecho, es la más precisa, aunque igual puede ser algo tan relativo como lo decía Rafael ayer. Si mal no recuerdo teníamos un error promedio de 0.5 ó 0.6 y era uno de los cuestionarios más largos que hemos hecho en una encuesta electoral, dado que lo utilizamos un poco como laboratorio hacia el 2000. Meter preguntas de todas las teorías conocidas, incluso por conocer y tenemos un cuestionario extensísimo, y de hecho fue una encuesta en términos de eficiencia muy buena, excelente.

A qué se debe esto, bueno la pregunta electoral siempre la dejamos al principio, no de entrada, pero sí después de una serie de preguntas que le den cierta confianza al encuestado, que nos den también información porque no soy de la idea de que hay que desperdiciar preguntas obviamente, y ya después el resto es una serie de asuntos distintos.

Con respecto a la pregunta de María de la Heras, esto me apena un poco, de hecho no sé si pedir disculpas o protestar porque en mis acetatos estaba la respuesta, pero todos los argumentos de los votantes probables se basa precisamente en la similitud de votantes probables con los votantes entrevistados en la encuesta de salida (*exit poll*). De hecho, recuerdo que cuando escribí una nota para el periódico utilicé este adjetivo, el parecido es tan formidable que los editores lo tacharon, es que es muy parecido en términos de identificación partidista, de ideología, de algunas variables sociodemográficas, entre otros. Por ejemplo, hemos hablado mucho de escolaridad, sexo, edad, como variables de comparación y demás, pero uno de los esfuerzos que sí hemos tratado de hacer en las encuestas de Reforma, es tratar de darle una mayor fineza a las mediciones por ocupación, de manera que sepamos no solamente escolaridad, edad y demás, sino realmente las experiencias laborales que como variable sociológica creo que sí pueden llegar a determinar la forma

cómo piensa o las preferencias de una persona, dónde trabajas, cómo vives, con quién te relacionas. Me parece que es bastante coherente explorar esa área. No tenemos mucha idea de cómo andamos en esas mediciones a nivel de encuestas.

Si en algún momento en alguna preelectoral metimos un cruce por una serie de categorías ocupacionales que entre quienes nos leyeron me parece que fue bastante interesante, bastante novedosa en el sentido de que casi no teníamos una relación. En general, decíamos: las amas de casa están con Labastida, muy bien, pero bueno, qué pasa con las amas de casa jóvenes y con las amas de casa que trabajan, es decir, una segmentación un poco más explicativa. Ese es uno de los problemas (*traders*) con reducir las muestras, no nos permite segmentar al nivel que a veces quisiéramos.

Pero esto también se aplica, me dice Rafael Giménez, a lo de la encuesta de salida (*exit poll*), pero creo que también se aplica a las preelectorales.

Creo que eso es todo. Nada más me gustaría cerrar con una pregunta para Roy Campos dado que sí me inquietó el hecho de que no pudiera estimar participación. Sin embargo, en la interna del PRI, los 10 millones de votos vinieron ¿de qué fue? ¿De una encuesta de salida (*exit poll*) o fue conteo rápido?

**ROY CAMPOS:** Te lo dije en un panel en Reforma, sí, lo conteste así muy claro, yo no valido que existan, hicimos un conteo rápido de actas y lo que nosotros decimos, aunque María de las Heras tiene una historia muy interesante al respecto, la suma de los datos que va a haber en las actas va a acercarse a los 10 millones, no sé si existan ni sé cómo llegaron.

**ALEJANDRO MORENO:** Si me permiten, se me pasó una pregunta y creo que aquí hay un tema de discusión que potencialmente nos podría dar más. Alejandro Cruz preguntaba si medimos diputados y senadores también aparte de Presidente, sí. Ambos lo medimos y aquí hay un asunto también metodológico que como que se ha sumado en foros previos, pero no se le ha dado mucha importancia, ¿en una encuesta entregas las tres boletas juntas? ¿Cuántas boletas puedes usar en una encuesta? Se las das, primero Presidente, luego le echas una batería de preguntas, luego diputado y otra batería, y luego senadores, ¿qué hacemos? Ahí sí yo lo dejaría abierto. La verdad es que el acto de votar es, tengo todas las boletas en mis manos, al menos esa fue la experiencia de casi todos y las veo y no sé si va primero Jefe de Gobierno, va Presidente, y en las encuestas si como que cuidamos mucho ver que va primero y que va después. Entonces, yo lo dejaría abierto.

**JAVIER ALAGÓN:** No hubo preguntas específicas con respecto al conteo rápido. Fue tan rápido que ya ni siquiera preguntas generó. Nada más para cerrar, me quedo con la impresión, he disfrutado la reunión, me ha sido muy útil pero creo que como comentaba ayer, nos falta autocrítica y una revisión retrospectiva con las encuestas preelectorales.

Creo que todos los que tienen encuestas preelectorales tienen el compromiso moral con ellos mismos y con nosotros, con la sociedad incluso, de hacer esta revisión retrospectiva. No he visto y no vi, creo que mañana a lo mejor sí podré ver algo de esto, ejercicios retrospectivos. Regrésense con sus encuestas y vean cómo les hubiera ido con esa muestra que tenían en enero, en febrero, en marzo, en abril, cómo les hubiera ido con los resultados finales que dio el IFE en esa sección electoral.

Es lo más fácil, regrésense, reconstruyan la historia y vamos a ver si en efecto, las muestras, es que da la impresión que seguimos un criterio por agencia, de congruencia antes que otra cosa y quizás eran malos, pero congruentes, o buenos, pero congruentes.

Es decir, si yo ya me comprometí con esto, pues ya no le muevo mucho. Si ya dije que el PRI está arriba, pues el PRI está arriba y yo no vi cruzarse líneas en alguna presentación, no sé si me equivoque.

Otra cosa, hay que hacerlo porque ahora ya está surgiendo otra hipótesis. La hipótesis de la volatilidad y entonces queremos saber qué tan volátiles son las encuestas preelectorales, qué significa esta volatilidad. De lo que vimos, de las exposiciones, de las encuestas con rotación (*trackings*), pues no parece haber tal volatilidad, entendiéndose esta volatilidad como variaciones más allá de cuanto les gusta, cinco ó 10 por ciento de una medición a otra, no pareciera haber.

En fin, yo sí quiero que, de alguna manera, se tiene que hacer este análisis retrospectivo. A lo mejor si las empresas no, si las que hicieron encuestas preelectorales y encuestas de salida no lo quieren hacer con alguna persona independiente, vamos ahí esta la propuesta,

porque quién va a analizar esto, quién va a analizar las diferencias, el cuestionario al que me referí hace rato, quién se va a hacer cargo de analizarlo, si es que es analizable.

Entonces yo creo que a lo mejor debiera haber una publicación en donde trate de dar alguna justificación en este sentido. Yo esperaba eso.

**FRANCISCO BÁEZ:** Se ha explorado la idea, y creo que es muy interesante, de que los conteos rápidos van a la baja, que los conteos rápidos van de salida, mientras las encuestas de salida van a la alza. Yo creo que esto tiene que ver, depende de la evolución política y la sensibilidad de la sociedad. Los conteos rápidos son hijos de la desconfianza, los conteos rápidos requieren una precisión, para una sociedad que esta acostumbrada a dudar de la veracidad de los resultados electorales, pueda irse tranquilamente a dormir sabiendo el ganador.

Las encuestas de salida, en cambio, que son las que tienen un peso más grande en otras naciones, sucede lo contrario. Son hijas de la confianza, son hijas de que tú ya sabes cómo está y que si la encuesta de salida va a tener cinco décimas de diferencia, de todos modos hay una autoridad electoral que va a dar toda las cosas claras y simplemente te están dando una información.

Yo, en ese sentido, quiero subrayar esto que yo he estado planteando. Es decir, en la medida en que los conteos rápidos se puedan hacer de una manera muy barata, muy sencilla, pues entonces quienes se dedican a este negocio tienen que pensar en mecanismos alternativos para hacer conteos rápidos o todavía más precisos o todavía más rápidos.

**MANUEL ROMÁN:** Creo que podemos decir mucho, pero ustedes han dicho todo. A todos los participantes les agradecemos su voz, sus reflexiones. Creo que podemos decir que todos y cada uno de nosotros hemos aprendido algo más. ■