

Pablo Reyes

Numérica®

2015 MAY 15 PM 3:02

México, D.F., Mayo 15 de 2015

Original / Firmados  
SECRETARÍA EJECUTIVA

le Acreditacion

Lic. Edmundo Jacobo Molina  
Secretario Ejecutivo  
Instituto Nacional Electoral  
Presente:



03061

Por medio del presente y en cumplimiento de las obligaciones derivadas del Acuerdo INE/CG238/2015 del Consejo General del Instituto Nacional Electoral por el que se establecen los lineamientos, así como los criterios de carácter científico que deberán observar las personas físicas o morales que pretendan ordenar y realizar encuestas de salida y/o conteos rápidos para la jornada electoral del día 7 de junio de 2015, me permito presentar a usted la metodología de la empresa MEDICIÓN Y ANÁLISIS ESTADÍSTICO AVANZADO S.C. (*Numérica*), para su **registro** ante la Secretaría Ejecutiva del Instituto Nacional Electoral.

## 1 METODOLOGÍA

A continuación se describe la metodología y la información sobre los criterios de carácter científico que *Numérica* seguirá para realizar encuestas de salida y conteos rápidos durante la jornada electoral del 7 de junio de 2015 de dos estados del país, **Nuevo León** y **Sonora**.

La metodología contempla la **realización de una encuesta de salida y un conteo rápido para cada uno de estos estados** (2 encuestas de salida y 2 conteos rápidos en total).

Las metodologías para ambas elecciones son análogas aunque pueden diferir en los procesos de selección de muestra y en la generación de estimaciones.

### 1.1 ENCUESTAS DE SALIDA

- Objetivos del estudio.** Estimar las preferencias electorales para **gobernador** de *Nuevo León* y *Sonora* durante la jornada electoral del 7 de junio de 2015.

2. **Marco muestral.** El marco muestral estará compuesto por el listado de secciones electorales de ambos estados con detalle de manzanas, listado nominal y referencias cartográficas. Se utilizarán las siguientes bases de datos para conformarlo:

Nombre del insumo	Fuente
Listado Nominal y Padrón Electoral	INE
Concentrado General de Secciones Electorales	INE
Catálogo de Información Geoelectoral	INE
Condensado de Información Geoelectoral Básica	INE
Catálogo de Rangos de Secciones por Municipio	INE
Catálogo de Secciones Electorales por Tipo	INE
Catálogo General de Localidades	INE
Catálogo de Manzanas	INE
Catálogo de Casillas	INE
Planos Urbanos Seccionales	INE
Planos por Secciones Individuales Urbanas	INE
Planos por Secciones Individuales Rurales	INE
Planos por Secciones Individuales Mixtos	INE

### 3. Diseño muestral.

- a. **Población objetivo. Votantes** en la elección de gobernador de cada entidad federativa de estudio (Nuevo León y Sonora).
- b. **Procedimiento de selección de unidades.** El diseño de muestreo para las encuestas de salida será probabilístico, bietápico, por conglomerados y con estratificación. Las etapas de muestreo son:
  - i. **Selección de secciones.** Se elegirán secciones mediante muestreo probabilístico con estratificación. La población objetivo será estratificada conforme a listado nominal utilizando *Hiper-Estratificación Óptima Numérica®*. Esta metodología optimiza el número de estratos a utilizar conservando homogeneidad al interior de los estratos.
  - ii. **Selección de votantes.** Se elegirá un votante de manera sistemática al salir del lugar donde se ubica la casilla básica de las secciones en muestra.

Las unidades de muestreo correspondientes al diseño anterior serán:

Unidades primarias de muestreo	Secciones
Unidades secundarias de muestreo	Votantes

- c. **Procedimiento de estimación.** Se utilizarán estimadores de expansión simple de Narain

(1951) y Horvitz-Thompson (1952). Los factores de expansión correspondientes consideran una expansión al tamaño de la población objetivo acorde con su probabilidad de inclusión en muestra.

También se utilizarán estimadores de razón de Hájek (1971) en aquellos casos en que sea pertinente utilizar estos en lugar de los de expansión simple.

- d. **Tamaño y forma de obtención de muestra.** A continuación se describen los tamaños de muestra para cada estado. Para los cálculos se tomaron en cuenta las expresiones matemáticas que se pueden encontrar en la conocida monografía de Méndez, Eslava & Romero (2004) (pp. 12-15, 44-50) con título *Conceptos Basicos de Muestreo*, editado por el IIMAS-UNAM.

Cabe mencionar que la siguiente fórmula es para la estimación de proporciones bajo un diseño de muestreo aleatorio simple. Esta expresión se utilizará como base, posteriormente se afecta por un efecto de diseño aproximado calculado a partir de datos históricos de *Numérica*.

Sea  $n$  el tamaño de muestra tenemos que,

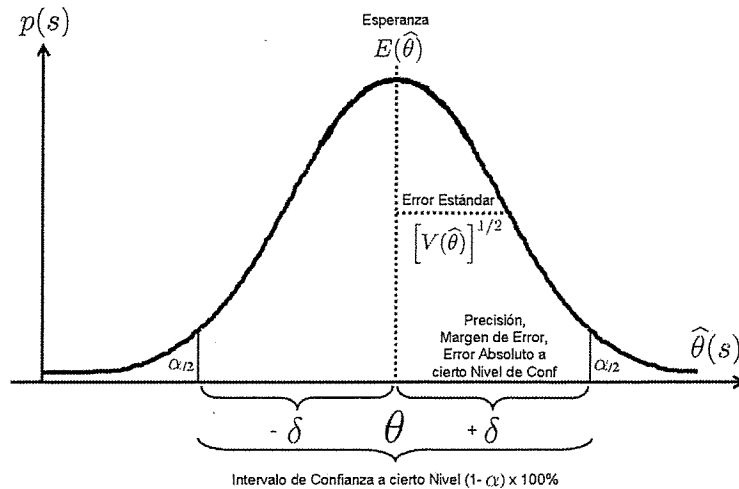

$$\quad \quad \quad (1)$$

donde  $N$  representa el tamaño de la población,  $\alpha$  representa el complemento a 1 del nivel de confianza a utilizar (e.g. si se utiliza un nivel de confianza del 95%, entonces  $\alpha$  será 0.05),  $z_\alpha$  representa el cuantil de una distribución Normal que acumula una probabilidad de  $\alpha$ ,  $\delta$  es el margen de error asociado a cierto nivel de confianza,  $P$  representa la proporción que se quiere estimar, que desconocemos y cuyo valor en la expresión anterior puede ser sustituido por alguno aproximado de estudios anteriores o de una prueba piloto.

Un supuesto conservador es asumir  $P = 0.50$  de modo que  $P(1 - P)$  se maximiza y por lo tanto se obtiene una  $n$  mayor (conservadora).

El siguiente gráfico esquematiza cada uno de los componentes utilizados en la expresión del cálculo de tamaño de muestra. El gráfico contempla en general la distribución muestral de un estimador  $\hat{\theta}$  que estima al parámetro  $\theta$ .

## Distribución Muestral de un Estimador.



Generalmente en ejercicios prácticos se utiliza un nivel de confianza del 95%, de modo que en la expresión anterior (1) se estaría utilizando un  $\alpha = 0.05$ . Posteriormente este tamaño de muestra se multiplica por el efecto de diseño correspondiente para obtener un tamaño de muestra para una proporción bajo cualquier diseño de muestreo. La expresión resultante puede ser despejada para calcular los márgenes de error.

La siguiente tabla muestra la relación de los tamaños de muestra a utilizar y los márgenes de error correspondientes (bajo muestreo aleatorio simple y bajo el diseño de muestreo a utilizar que toma en cuenta la afectación por el diseño de muestreo).

Entidad Federativa (Gobernador)	Proporción Estimada	Tamaño de la Población (Listado Nominal)	Tamaño de Muestra Secciones (Referente)	Tamaño de Muestra Global (Sup. 40 entrevistas por sección)	Nivel de Confianza	Verificación de Normalidad (Variables Dicotómicas)		Cuantil de Normal Estándar a la $\alpha/2$	Margen de Error Absoluto para Muestreo Aleatorio Simple bajo Sup. Normalidad	Efecto de Diseño (aprox. e diseño a usar histórico)	Margen de Error Absoluto para el diseño a usar bajo Sup. Normalidad
	P	N	nSeccs	n	(1-Alpha)	nP>5	n(1-P)>5	Z(alpha/2)	delta MAS	Deff	delta esperado MÁXIMO
Nuevo León	50%	3,559,405	200	8,000	95%	4,000	4,000	1.95996	1.09%	5.68	2.61%
Sonora	50%	1,967,435	180	7,200	95%	3,600	3,600	1.95996	1.15%	5.68	2.75%
Total			380	15,200							

Sobre la forma de obtención de la muestra esta sigue el procedimiento descrito en el apartado anterior donde se especifica el diseño de muestreo. La implementación de tal diseño de muestreo y las estimaciones se realizan utilizando software de código

abierto R. En particular los paquetes `sampling` y `samplingVarEst`.

- e. **Calidad de la estimación: confianza y error máximo implícito en la muestra seleccionada.** En la tabla anterior se muestran los errores teóricos de las estimaciones asumiendo un muestreo aleatorio simple (MAS). Se exhiben los márgenes de error teóricos esperados (realistas máximos) bajo el diseño de muestreo utilizado.

Es importante señalar que cada estimación tiene un error muestral que depende del diseño de muestreo y de la variabilidad en las respuestas. El día de la jornada electoral, los errores muestrales reales observados se estimarán considerando un nivel de confianza del 95% y el diseño de muestra utilizado sin necesidad de asumir un MAS.

Como medida de calidad se propone el uso de coeficientes de variación estimados calculados a partir de la correcta estimación de varianza de los estimadores utilizados, sin asumir un muestreo aleatorio simple.

- f. **Tratamiento de la no respuesta.** La tasa de respuesta *RR1* se calculará con base en los *Standard Definitions de AAPOR*.

Las frecuencias de los votantes que dejen en blanco o anulen la boleta simulada serán reportadas en la estimación de la preferencia electoral *bruta* y se asignarán proporcionalmente a los candidatos para la estimación de la preferencia *efectiva*.

4. **Técnica de recolección de datos.** Entrevistas cara a cara a votantes al salir de las casillas con urna simulada y cuestionario estructurado aplicado en papel por encuestadores.
5. **Cuestionario.** Boleta y urna simulada. Adicionalmente, se aplicará un cuestionario breve que permita perfilar a los votantes.
6. **Forma de procesamiento, estimadores e intervalos de confianza.** Se utiliza la teoría explicada en el apartado anterior sobre el tamaño de muestra. El procesamiento de información incluye varias rutinas de validación de información de campo y depuración de registros que alimentan los algoritmos de estimación.

La implementación de tales algoritmos de estimación se realiza utilizando rutinas propias de estimación, además de los paquetes R: `samplingVarEst` y `samplingEstimates`. Estos están disponibles en las siguientes dos ligas:

<http://cran.r-project.org/web/packages/samplingVarEst/index.html>

<http://cran.r-project.org/web/packages/samplingEstimates/index.html>.

Tales paquetes son de reconocida calidad, e.g. son utilizados para la enseñanza de muestreo en la Universidad de Michigan, una institución de histórico abolengo en temas de muestreo.

Los estimadores a utilizar fueron descritos anteriormente (estimadores para diseños de muestreo sin reemplazo con probabilidades desiguales) y la construcción de intervalos de confian-

za es aquella que se detalla en la monografía de Méndez, Eslava & Romero (2004) con título *Conceptos Basicos de Muestreo*, editado por el IIMAS-UNAM.

7. **Software.** Rutinas propias escritas en R y en C. Además se utilizan los paquetes R: *sampling*, *samplingVarEst* y *samplingEstimates*. Estos dos últimos de autoría propia de Numérika, única empresa mexicana que utiliza paquetes de cómputo R propios publicados en el CRAN. Cabe mencionar que los paquetes fueron realizados por estadísticos muestristas mexicanos premiados por los organismos internacionales más altos de Estadística (*International Statistical Institute* - ISI).
8. **Patrocinio.** ELEKTRA DEL MILENIO S.A. DE C.V.

El monto económico involucrado para este ejercicio es de \$ 1,300,000.00 pesos más I.V.A. Es importante señalar que este monto incluye las estimaciones de las encuestas de salida y conteos rápidos, pero no incluye el monto del trabajo de campo. La recolección de la información en campo la contratará el cliente a otra empresa.

## 1.2 CONTEOS RÁPIDOS

La metodología de los Conteos Rápidos es similar a la descrita anteriormente. Sin embargo, el diseño de muestreo y el método de recolección de información varían ligeramente.

### 1. Diseño muestral para los Conteos Rápidos

- a. **Población objetivo.** **Votantes** en la elección de gobernador de cada estado de estudio.
- b. **Procedimiento de selección de unidades.** El diseño de muestreo para los conteos rápidos será probabilístico con estratificación. En este caso, sólo hay una etapa de muestreo:
  - ▷ **Selección de secciones.** Los Conteos Rápidos se realizarán en las mismas secciones electorales elegidas para las encuestas de salida. En cada sección en muestra, se censan las casillas.
- c. **Procedimiento de estimación.** Se utilizarán estimadores de expansión simple de Narain (1951) y Horvitz-Thompson (1952). Los factores de expansión correspondientes consideraran una expansión al tamaño de la población objetivo acorde con su probabilidad de inclusión en muestra.

También se utilizarán estimadores de razón de Hájek (1971) en aquellos casos en que sea pertinente utilizar estos en lugar de los de expansión simple.

- d. **Tamaño y forma de obtención de muestra.** A continuación se describen los tamaños de muestra para cada estado. **Las expresiones matemáticas utilizadas son las mismas**

que para la Encuesta de Salida aunque con diferentes valores en su evaluación. Estos valores se detallan en la siguiente tabla:

Entidad Federativa (Gobernador)	Proporción Estimada	Tamaño de la Población de Secciones	Tamaño de Muestra Secciones	Nivel de Confianza	Verificación de Normalidad (Variables Dicotómicas)		Cuantil de Normal Estándar a la Alpha/2	Margen de Error Absoluto para Muestreo Aleatorio Simple bajo Sup. Normalidad	Efecto de Diseño (aprox. e histórico)	Margen de Error Absoluto para el diseño a usar bajo Sup. Normalidad
					nP>5	n(1-P)>5				
	P	N	n	(1-Alpha)	nP>5	n(1-P)>5	Z(alpha/2)	delta MAS	Deff	delta esperado MÁXIMO
Nuevo León	50%	2,640	200	95%	100	100	1.959964	6.66%	0.14	2.45%
Sonora	50%	1,518	180	95%	90	90	1.959964	6.66%	0.14	2.52%
Total			380							

e. **Calidad de la estimación: confianza y error máximo implícito en la muestra seleccionada.** En la tabla anterior se muestran los errores teóricos de las estimaciones asumiendo un muestreo aleatorio simple (MAS). Se exhiben los márgenes de error teóricos esperados (realistas máximos) bajo el diseño de muestreo utilizado.

Es importante señalar que cada estimación tiene un error muestral que depende del diseño de muestreo y de la variabilidad en las respuestas. El día de la jornada electoral, los errores muestrales reales observados se estimarán considerando un nivel de confianza del 95 % y el diseño de muestra utilizado sin necesidad de asumir un MAS.

Como medida de calidad se propone el uso de coeficientes de variación estimados calculados a partir de la correcta estimación de varianza de los estimadores utilizados sin asumir un muestreo aleatorio simple.

f. **Tratamiento de la no respuesta.** En este caso no se tiene un cálculo de tasa de respuesta.

Las frecuencias de los votos por candidatos no registrados, votos en blanco o votos nulos serán reportadas en la estimación de la votación *bruta* y se eliminarán para la estimación de la votación *efectiva*.

2. **Método de recolección de la información.** Los encuestadores recabarán la información de los resultados de las elecciones de gobernador publicados en las casillas de las secciones en muestra.

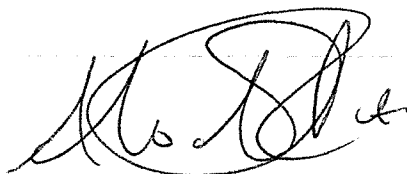
Finalmente, debido a que las encuestas de salida y conteos rápidos serán publicados, *Numérica* también dará cumplimiento al Acuerdo INE/CG220/2015 del Consejo General del Instituto Nacional Elec-

toral. Por lo tanto, se anexan al presente documento lo siguiente:

- ▶ Comprobante de registro de *Numérika* ante el *International Statistical Institute - ISI*,
- ▶ Datos de registro de *Numérika* ante la *World Association for Public Opinión Research - WAPOR*,
- ▶ Curriculum vitae con la experiencia profesional y formación académica del equipo que participará en este proyecto:
  - ▶ Dr. Rubén Hernández Cid,
  - ▶ Dr. Emilio López Escobar,
  - ▶ Mtro. Alvaro Mejía Avilés.

Sin otro particular, quedo a sus órdenes para cualquier duda o comentario sobre la presente metodología y a la espera de su notificación del registro de la misma.

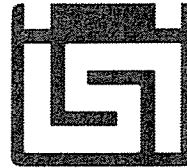
Atentamente,



**Mtro. Álvaro Mejía Avilés**  
**Representante Legal**  
alvaro@numerika.mx



13 November 2014



International Statistical Institute  
Permanent Office  
Henri Faasdreef 312  
2490 AB Den Haag, The Netherlands  
Tel: (31) 70 337 5737 [isi@cbs.nl](mailto:isi@cbs.nl)  
Fax: (31) 70 386 0025 [www.isi-web.org](http://www.isi-web.org)  
ISI VAT No. : NL 0033.85.073.B.01

Dr. Emilio Lopez Escobar  
Numérica - Medición y Análisis Estadístico Avanzado, SC  
Av. Colonia del Valle 511 Piso 1  
Colonia Del Valle  
MÉXICO, C.P. 03100  
MEXICO

## RECEIPT

Membership number: 17222

This is to acknowledge with thanks receipt of €200 for the annual IASS Institutional membership fee for the year 2015.

Yours sincerely,

Margaret de Ruiter-Molloy  
Membership Officer



## DATOS DE MEMBRESÍA A WAPOR

A continuación se listan los datos de la membresía de *Numérica* a la **World Association for Public Opinion Research**.

- ▶ **Nombre de la organización:** *Numérica*.
- ▶ **Número de miembro:** 1319.
- ▶ **Vigencia:** 01/01/16.