

Informe que presenta la Dirección Ejecutiva de Capacitación Electoral y Educación Cívica sobre la imposibilidad material para realizar el Estudio muestral de la participación ciudadana en las elecciones federales de 2006

Antecedentes

En la sesión extraordinaria del 20 de marzo de 2013, el Consejo General del IFE aprobó el Acuerdo CG92/2013, mediante el cual se determinó la realización de dos estudios de la documentación electoral utilizada durante el Proceso Electoral Federal 2005-2006, uno de los cuales corresponde al *Estudio muestral de la participación ciudadana en las elecciones federales de 2006*.¹

Dicho estudio, con inferencia a nivel nacional y por entidad federativa, permitiría identificar el perfil de los ciudadanos que sufragaron y de los que no lo hicieron, considerando variables como sexo, edad y tipo de sección electoral, entre otras, a partir de una muestra de las listas nominales utilizadas en el Proceso Electoral Federal 2005-2006. El Estudio también permitiría realizar un análisis comparativo con respecto a los resultados de los estudios de participación ciudadana de 2003, 2009 y 2012.

Descripción

Durante los meses de septiembre y octubre de 2013, la Dirección Ejecutiva de Capacitación Electoral y Educación Cívica (DECEyEC), responsable del Estudio, contó con el apoyo de la Dirección Ejecutiva del Registro Federal de Electores (DERFE) para el diseño de la muestra de los paquetes electorales seleccionados para recuperar de ellos las listas nominales, y de la Dirección Ejecutiva de Organización Electoral (DEOE) para su separación y resguardo.²

¹ El otro corresponde al Estudio muestral sobre las características de los votos nulos en la elección presidencial federal de 2006, coordinado por la Dirección Ejecutiva de Organización Electoral.

² Simultáneamente se realizó la destrucción de la documentación contenida en los paquetes electorales que no formaban parte de la muestra, conforme a los Lineamientos para la destrucción de los votos válidos, los votos nulos, las boletas sobrantes y la lista nominal del proceso Electoral Federal 2005-2006 (Acuerdo CG/238/2013).

La muestra a nivel nacional estuvo integrada por 8,623 paquetes, de los cuales sólo se recuperaron 5,901 listas nominales (68.4%), es decir, no se hallaron 31.6%³ de las listas nominales utilizadas en las casillas instaladas en 2006. Destacan casos como Aguascalientes, en donde sólo se localizaron dos listados nominales de un total de 333 pertenecientes a la muestra. En el resto de las entidades, los porcentajes más altos de listas nominales faltantes oscilan entre 40% y 65%, tal como se puede apreciar en el siguiente cuadro.

Tamaño de la muestra y casillas sin Lista Nominal

Entidad	Casillas muestra	Casillas sin Lista Nominal	
		Cantidad	%
Aguascalientes	333	331	99.4
Baja California	422	269	63.7
Baja California Sur	365	26	7.1
Campeche	187	118	63.1
Chiapas	355	160	45.1
Chihuahua	341	30	8.8
Coahuila	331	106	32
Colima	175	31	17.7
Distrito Federal	275	22	8
Durango	366	206	56.3
Guanajuato	285	116	40.7
Guerrero	199	48	24.1
Hidalgo	190	45	23.7
Jalisco	269	104	38.7
México	239	27	11.3
Michoacán	276	76	27.5
Morelos	223	13	5.8
Nayarit	307	43	14
Nuevo León	218	75	34.4
Oaxaca	208	56	26.9
Puebla	179	39	21.8
Querétaro	197	35	17.8
Quintana Roo	254	53	20.9
San Luis Potosí	371	81	21.8
Sinaloa	170	53	31.2
Sonora	349	232	66.5

³ La DEOE presentó un informe sobre el estado en el que se encontraban los paquetes electorales del proceso electoral federal 2005-2006, en la segunda sesión extraordinaria de las Comisiones Unidas de Capacitación Electoral y Educación Cívica, Organización Electoral y Registro Federal de Electores del 8 de marzo de 2013, que señalaba que el 26.29% de los paquetes no contaban con listados nominales, ya que fueron requeridos a diversas juntas distritales ejecutivas por el Tribunal Electoral del Poder Judicial de la Federación.

Entidad	Casillas muestra	Casillas sin Lista Nominal	
		Cantidad	%
Tabasco	262	67	25.6
Tamaulipas	314	88	28
Tlaxcala	217	41	18.9
Veracruz	201	62	30.8
Yucatán	239	39	16.3
Zacatecas	306	30	9.8
Total	8,623	2,722	31.6

Cuadro elaborado por la Dirección de Estadística de la DERFE

Conclusión

A partir de los hallazgos, la DECEyEC solicitó a la DERFE una opinión técnica respecto a la viabilidad de llevar a cabo el Estudio con los listados disponibles. Dicha dirección determinó que, considerando la gran proporción de listas nominales faltantes, no es posible llegar a conclusiones estadísticamente confiables a nivel nacional o por entidad federativa sobre las variables que se pretendían analizar en el estudio de participación ciudadana, así como encontrar resultados comparables con los estudios de este tipo que se han realizado para otros procesos electorales; además de que no existe herramienta estadística que permita sustituir la información faltante. El detalle de las consideraciones técnicas emitidas por la DERFE puede consultarse en el Anexo del presente informe.

En virtud de lo anterior, se informa que ante la carencia de elementos objetivos, no se puede llevar a cabo el estudio muestral de la participación ciudadana en las elecciones federales de 2006, por lo que se recomienda que una vez concluido el Estudio coordinado por la DEOE, se realice la destrucción de las listas nominales que fueron recuperadas junto con la documentación contenida en los paquetes de la muestra correspondiente, y que aún se encuentran bajo resguardo de la DEOE.

Anexo. Consideraciones estadísticas de la Dirección Ejecutiva del Registro Federal de Electores sobre la disponibilidad de datos y por las que no es posible realizar el estudio muestral de la participación ciudadana en las elecciones federales de 2006

Antecedentes

En el *Acuerdo CG92/2013 del Consejo General del Instituto Federal Electoral por el que se determina la realización de estudios de la documentación electoral utilizada durante el proceso electoral federal 2005-2006*, se estableció la realización de dos estudios:

- Uno sobre la participación ciudadana en las elecciones federales de 2006, coordinado por la Dirección Ejecutiva de Capacitación Electoral y Educación Cívica.
- Otro sobre características de los votos nulos de las elecciones federales de 2006, coordinado por la Dirección Ejecutiva de Organización Electoral.

Como especificaciones técnicas, en el mismo acuerdo se estableció que las estimaciones deberían realizarse al 95% de confianza y con el menor margen de error posible de acuerdo con las necesidades requeridas por las dos direcciones ejecutivas mencionadas. Asimismo, se estableció que el apoyo en la aplicación de técnicas estadísticas estaría a cargo de la Dirección Ejecutiva del Registro Federal de Electores a través de su Dirección de Estadística.

En tanto que la destrucción de la documentación electoral estaba próxima a realizarse, se reservarían los paquetes electorales para obtener los datos que conformarían la muestra.

El estudio de la participación ciudadana, tenía como propósito formar parte de los análisis que se habían realizado en 2003, 2009 y 2012, lo que

complementaría la serie histórica de las elecciones federales para el periodo 2003-2012.

Diseño muestral

Como consideración inicial, se determinó que el tamaño de muestra cubriría los requerimientos de datos de los dos temas de estudio. Por un lado, estimar el porcentaje de participación electoral por edad y sexo en cada una de las 32 entidades federativas y, por otro, analizar las características del voto nulo.

El cálculo del tamaño de muestra se basó en un diseño muestral que cumple con el objetivo de todo análisis estadístico, que es formular inferencias sobre la población basadas en la información contenida en una muestra⁴ y esta muestra es un subconjunto de la población que refleja las características esenciales de la población de la cual se obtuvo.

El diseño muestral tiene las siguientes características:

- Es probabilístico porque todas las unidades de estudio, en este caso los paquetes electorales, tienen posibilidad de formar parte de la muestra, es decir, tienen asociada una probabilidad diferente de cero de ser seleccionadas para formar parte de la muestra. Sin embargo, en cuestiones prácticas, se asume como una limitación ineludible que algunas unidades no podrán ser seleccionadas (en este caso las listas nominales no disponibles).
- El objetivo del diseño de muestreo es el de obtener inferencias a nivel nacional y por entidad federativa.
- El total de paquetes para el estudio fue conformado por 129,640 casillas. Este número excluye las casillas que no tuvieron votos, las

⁴ Scheffer, Mendenhall & Ott "*Elementary Survey Sampling*". Duxbury Press.

correspondientes al voto de los mexicanos en el extranjero y las especiales.

- El tamaño de muestra para cada estudio y entidad federativa se determinó (ver cuadro1) de acuerdo con un diseño de muestreo aleatorio simple al cual se le agregó un posible porcentaje de no respuesta expresado en la siguiente fórmula⁵:

$$n = \frac{p(1-p)Z^2}{\partial^2} (1 + TNR)$$

donde

p: proporción de la característica de interés (participación electoral o voto nulo)

Z: valor de la variable aleatoria que se distribuye Normal y corresponde a un nivel de 95% de confianza

∂: precisión de 2.5%

TNR: tasa de no respuesta

- La muestra nacional estuvo conformada por 8,623 paquetes, el 6.6% del total de paquetes disponibles.
- Los paquetes fueron seleccionados en la bodega de acuerdo a un procedimiento denominado “sistemático”, con arranque aleatorio, ordenado por estado, distrito y porcentaje de voto nulo⁶. Este procedimiento es ampliamente utilizado en el muestreo aleatorio simple, ya que al ordenar de manera apropiada a las unidades en estudio, se obtiene una aportación para mejorar la precisión de las estimaciones.

⁵ Morales Vallejo. “Tamaño necesario de la muestra: ¿Cuántos sujetos necesitamos?” Universidad Pontificia Comillas. Madrid.

<http://www.upcomillas.es/personal/peter/investigacion/Tama%F1oMuestra.pdf>

⁶ El ordenamiento de los paquetes dio preferencia al voto nulo ya que por su reducida magnitud, representaba mayores dificultades de estimación.

Recolección de datos

Con base en el diseño muestral, se calcularon para cada entidad los tamaños de muestra para el estudio sobre la participación ciudadana y para el estudio sobre las características de los votos nulos. Con la intención de facilitar el trabajo operativo de separar en la bodega los paquetes a revisar, se combinaron las muestras de los dos estudios optando siempre, en cada entidad, por el tamaño de muestra más grande. Así, se llegó a un tamaño de muestra final de 8,623 paquetes para realizar ambos estudios (ver Cuadro 1).

Cuadro 1: Tamaño de muestra por tipo de estudio y por entidad federativa

Entidad	Estudio de Participación	Estudio de Voto Nulo	Tamaño de muestra final para ambos estudios
Total	5,823	8,527	8,623
Aguascalientes	201	333	333
Baja California	255	422	422
Baja California Sur	200	365	365
Campeche	187	176	187
Coahuila	165	355	355
Colima	190	341	341
Chiapas	175	158	175
Chihuahua	201	331	331
Distrito Federal	141	275	275
Durango	213	366	366
Guanajuato	154	199	199
Guerrero	204	285	285
Hidalgo	172	190	190
Jalisco	176	269	269
México	143	239	239
Michoacán	192	276	276
Morelos	149	223	223
Nayarit	186	307	307
Nuevo León	152	218	218
Oaxaca	208	164	208
Puebla	159	179	179
Querétaro	162	197	197
Quintana Roo	228	254	254
San Luis Potosí	170	146	170
Sinaloa	224	371	371
Sonora	190	349	349
Tabasco	122	262	262

Entidad	Estudio de Participación	Estudio de Voto Nulo	Tamaño de muestra final para ambos estudios
Tamaulipas	193	314	314
Tlaxcala	150	217	217
Veracruz	170	201	201
Yucatán	144	239	239
Zacatecas	247	306	306

Cabe destacar que en el procedimiento que se llevó a cabo, el tamaño de muestra estaba excedido con respecto a los cálculos teóricos⁷. Esta situación se consideró conveniente debido a que se conocía la imposibilidad de recuperar todos los cuadernillos de lista nominal que serían la base del estudio.

De una revisión efectuada al material de los paquetes seleccionados, se encontró que no se dispone de la lista nominal (es decir, los cuadernillos que fueron utilizados en cada casilla en los que registra si el ciudadano votó o no votó) en 31.6% de los paquetes que conforman la muestra total a nivel nacional (ver Cuadro 2).

Cuadro 2: Número de paquetes en muestra y número de paquetes en los que no se dispone de la Lista Nominal

Entidad	Paquetes en muestra	Paquetes sin Lista Nominal	
		Abs	%
Total	8,623	2,722	31.6
Morelos	223	13	5.8
Baja California Sur	365	26	7.1
Distrito Federal	275	22	8.0
Chihuahua	341	30	8.8
Zacatecas	306	30	9.8
México	239	27	11.3
Nayarit	307	43	14.0
Yucatán	239	39	16.3
Colima	175	31	17.7

⁷ El tamaño de muestra teórico nacional, resultado de la suma de los tamaños de muestra para cada entidad, por concepto de no respuesta, se aumentó en 9% aproximadamente. Después de esto, por concepto de combinar las muestras, el tamaño de muestra se aumentó nuevamente en un 48%.

Entidad	Paquetes en muestra	Paquetes sin Lista Nominal	
		Abs	%
Querétaro	197	35	17.8
Tlaxcala	217	41	18.9
Quintana Roo	254	53	20.9
Puebla	179	39	21.8
San Luis Potosí	371	81	21.8
Hidalgo	190	45	23.7
Guerrero	199	48	24.1
Tabasco	262	67	25.6
Oaxaca	208	56	26.9
Michoacán	276	76	27.5
Tamaulipas	314	88	28.0
Veracruz	201	62	30.8
Sinaloa	170	53	31.2
Coahuila	331	106	32.0
Nuevo León	218	75	34.4
Jalisco	269	104	38.7
Guanajuato	285	116	40.7
Chiapas	355	160	45.1
Durango	366	206	56.3
Campeche	187	118	63.1
Baja California	422	269	63.7
Sonora	349	232	66.5
Aguascalientes	333	331	99.4

La ausencia de cuadernillos se denomina estadísticamente como “datos faltantes”, es decir, aquellos que no han podido ser observados y limitan el análisis de datos, ya que ha disminuido el tamaño de la muestra⁸. La ausencia de datos no permite conocer la forma y la magnitud en la que se podrían modificar las estimaciones.

Hay diversas técnicas estadísticas para la imputación de datos faltantes que están basadas en modelos estadísticos como son la eliminación de datos (*listwise*), el pareo de observaciones (*pairwise*), el método de medias o los algoritmos de imputación múltiple (IM)⁹. Para aplicar estas técnicas es

⁸ “Valores faltantes” Universidad de Valencia - España
http://www.uv.es/webgid/Descriptiva/23_valores_faltantes.html

⁹ Medina, Galván “Imputación de datos: teoría y práctica”. Serie estudios estadísticos y prospectivos. CEPAL(2007) <http://www.eclac.org/publicaciones/xml/9/29949/LCL2772e.pdf>

necesario conocer, entre otras condiciones, el patrón de comportamiento de los datos omitidos y datos cuantitativos de otras variables relacionadas. Por ejemplo, para sustituir datos faltantes en las encuestas, se utilizan las respuestas de personas con características similares a las que no respondieron.

Para el caso particular de este estudio, no existen las condiciones estadísticas para aplicar estas técnicas. Además, está ampliamente documentado que la aplicación de procedimientos inapropiados de sustitución de información introduce sesgos y reduce el poder explicativo de los métodos estadísticos, le resta eficiencia a la fase de inferencia e invalida las conclusiones del estudio.

Por otro lado, aunque a nivel nacional se calcula un 31.6% de no respuesta, el porcentaje de datos faltantes por entidad federativa no es homogéneo: pues va desde el 99.4% para el caso de Aguascalientes, al 5.8% para Morelos, por lo que no es factible establecer inferencias sobre el valor de los parámetros a nivel nacional.

En caso de que se decidiera minimizar la importancia que tiene el porcentaje de datos faltantes, nos encontraríamos con una dificultad de interpretación: el nivel de precisión (margen de error) con el que se diseñó la muestra es de 2.5%, pero en todas las entidades federativas el porcentaje de no respuesta es mayor al 2.5%. Incluso en 27 de ellas, el porcentaje de datos faltantes es mayor al 10% (ver cuadro 2).

Es decir, el porcentaje de ausencia de información para cada entidad federativa es mucho mayor al porcentaje de error del diseño de la muestra, por ejemplo, en el caso de Yucatán el porcentaje de datos faltantes es de 16.3% y para esta entidad no tiene sentido formular inferencias a partir de un diseño muestral con un 2.5% de precisión. Por lo tanto, no es recomendable hacer estudios sobre los paquetes de alguna entidad federativa porque ni siquiera en aquella que se cuenta con más información (Morelos con 5.8% de paquetes sin listados nominales) se podrían obtener estimaciones estadísticamente confiables.

Es importante aclarar que aunque el tamaño de la muestra se hubiera incrementado, incluso al grado de seleccionar todo el universo de paquetes (es decir, mediante la realización de un estudio censal a los 129,640 paquetes) la ausencia de cuadernillos es de tal magnitud, que no se podría haber contado con suficiente información para cumplir con los objetivos del Estudio.

Conclusiones

Considerando la gran proporción de listas nominales faltantes, no es posible obtener estimaciones estadísticamente confiables a nivel nacional o por entidad federativa. Por lo tanto, no es posible realizar el Estudio de participación ciudadana 2006 para cubrir el faltante en la serie histórica 2003 - 2012 de la participación electoral por edad, sexo, tamaño de la localidad, entre otros resultados.

Adicionalmente, no existe herramienta estadística que permita sustituir la información faltante.

Por lo anterior, se concluye que no es posible elaborar inferencia estadística alguna que cumpla con lo ordenado por el punto de acuerdo primero del *Acuerdo CG92/2013*.

Bibliografía

- Morales Vallejo. "Tamaño necesario de la muestra: ¿Cuántos sujetos necesitamos?" Universidad Pontificia Comillas. Madrid. <http://www.upcomillas.es/personal/peter/investigacion/Tama%F1oMuestra.pdf>.
- Muñoz Rosas, Álvarez Verdejo "Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus". Revista de Métodos Cuantitativos para la Economía <http://www.upo.es/RevMetQuant/pdf/vol7/art25.pdf>.
- P. V. Sukhatme "Teoría de Encuestas por Muestreo". Fondo de Cultura Económica.

- Scheaffer, Mendenhall & Ott "*Elementary Survey Sampling*". Duxbury Press.
- "*Valores faltantes*" Universidad de Valencia - España
http://www.uv.es/webgid/Descriptiva/23_valores_faltantes.html.
- William G. Cochran "*Técnicas de Muestreo*". Compañía Editorial Continental, S.A. de C.V.