

## Estudio Nacional Electoral CIDE-CSES 2009 Nota metodológica



### I) Características del estudio

Encuesta nacional en viviendas (cara a cara) de 2,400 entrevistas efectivas realizadas a los ciudadanos de la República Mexicana en los días posteriores a la elección federal del 5 de julio (13 a 24 de julio). La encuesta se realizó partiendo de una muestra probabilística distribuida en 240 secciones electorales aplicando 10 entrevistas por sección.

Dado que la concurrencia de elecciones locales es un fenómeno que puede incidir en la participación electoral, en la encuesta se incluyen tres dominios de estudio: estados en donde hubo elecciones de gobernador (Campeche, Colima, Nuevo León, Querétaro, San Luis Potosí y Sonora), estados con elecciones de legislaturas locales (D.F., Guanajuato, Jalisco, Estado de México y Morelos) y estados en donde no hubo elecciones concurrentes (los 21 estados restantes).

Es pertinente señalar que la encuesta buscó cubrir una gama amplia de temas, lo que resultó en un cuestionario demasiado extenso imposible de realizar en campo. Para subsanar esto, se diseñaron seis cuestionarios que se aplicaron a submuestras. En estos cuestionarios se mantiene un conjunto de variables básicas y el resto busca unidad temática del cuestionario.

### II) Diseño muestral

El marco de muestreo que se utilizó en la selección de la muestra es el listado de secciones electorales de la República Mexicana. Con este marco se garantiza una división exhaustiva y excluyente de la población bajo estudio. Las secciones forman conglomerados de individuos por lo que el muestreo corresponde a un muestreo polietápico de conglomerados, en donde la unidad primaria de muestreo es la sección, la unidad secundaria es la manzana, la tercera unidad es la vivienda y la unidad final de muestreo es el entrevistado.

Las secciones electorales se eligen dentro de cada dominio-estrato con probabilidad proporcional al listado nominal. Una vez elegidas las secciones la selección de las manzanas y de las viviendas se efectúa durante el operativo de campo a través de un muestreo aleatorio sistemático. En cada sección el encuestador hace una lista de las manzanas que la componen asignándoles un consecutivo de lo cual obtiene el número total de manzanas dentro de la sección ( $k$ ). Dentro de cada sección se deben elegir dos manzanas para lo cual el encuestador divide el número total de manzanas de la sección ( $k$ ) entre 2 lo cual determina el “salto” entre manzanas. Posteriormente el encuestador elige aleatoriamente un número que esté contenido entre 1 y el “salto” ayudándose con una tabla de número aleatorios y el número elegido corresponde a la primera manzana que debe ser seleccionada. Para elegir la segunda manzana se le suma el “salto” al primer número seleccionado.

Una vez seleccionadas las manzanas se procede a seleccionar las viviendas. El proceso de selección de viviendas es muy similar al empleado para seleccionar las manzanas. El encuestador enlista todas las viviendas que hay en la manzana y divide el total entre 5 (salto) ya que se deben elegir 5 viviendas en cada manzana, después el encuestador elige aleatoriamente un número entre 1 y el “salto” y esa es la primer vivienda seleccionada. La segunda vivienda seleccionada es el primer número seleccionado más el salto, la tercera vivienda es el segundo número seleccionado más el salto y así sucesivamente para la cuarta y quinta vivienda.

La última etapa de selección es la del entrevistado. En cada vivienda seleccionada el encuestador enlista a todos los integrantes de la vivienda con su respectiva fecha de cumpleaños y escoge a la persona del último cumpleaños, esto es, aquella cuya fecha de cumpleaños es la más reciente. En caso que la persona seleccionada no se encuentre en la vivienda el encuestador debe hacer 3 visitas para contactar a la persona seleccionada, si al cabo de las 3 visitas el contacto no se pudo realizar entonces se debe sustituir con la vivienda contigua respetando el sentido de las manecillas del reloj.

Bajo este esquema de muestreo todos los ciudadanos de la República Mexicana tienen una probabilidad no nula y conocida de ser seleccionados. La muestra así construida permite, además de generar estimaciones precisas de las variables de interés, hacer comparaciones entre subgrupos de la población, averiguar si existen diferencias entre ellos y sobre todo formular o verificar hipótesis sobre sus causas.

Para asegurar una dispersión de la muestra que considere secciones de todo tipo y obtener así estimadores con menor varianza se forma dentro de cada dominio de estudio 4 estratos de competencia electoral.

Dado el diseño del muestreo es necesario utilizar factores de expansión ( $\pi$ ), los cuales se calculan como el inverso de la probabilidad de selección del entrevistado. Una vez hecho el levantamiento de la encuesta se calculan factores de ajuste por no respuesta y por desviaciones de los parámetros poblacionales de sexo y edad.

$$\pi = \frac{1}{P(\text{individuo esté en muestra})}$$

El tamaño y el diseño de muestra garantizan para un nivel de confianza del 95% un margen de error teórico <sup>(d)</sup> de  $\pm 2.0$  puntos en su conjunto.

El cálculo de este margen de error se presenta a continuación.

$$\begin{aligned}
 d &= \sqrt{\left(\frac{t^2 * P(1-P) * efd}{n}\right)} * 100 \\
 &= \sqrt{\left(\frac{(1.96)^2 * 0.5(1-0.5) * 2}{2,400}\right)} * 100 \\
 &\equiv 2.8
 \end{aligned}$$

en donde,

$d$  = Es el margen de error asociado a la estimación de la proporción  $P$  .

$P$  = es el parámetro poblacional (proporción) que se busca estimar. El cálculo del margen de error se realiza suponiendo una proporción de 0.5 porque este es el valor que maximiza el error, esto es, cualquier otra proporción tendrá un margen de error menor.

$t$  = es el percentil de la distribución normal asociado con la confianza deseada. Para el cálculo del error se supone un nivel de confianza del 95%.

$n = 2400$  , tamaño de muestra.

$efd$  = Efecto de diseño.

$$efd(\hat{p}) = \frac{V(\hat{p}) \text{ Bajo el diseño muestral}}{V(\hat{p}) \text{ Bajo un m.a.s}}$$

$$V(\hat{p}) = \frac{N-n}{(N-1)(n-1)} pq$$

Con

$N$  : El tamaño de la población.

$n$  : El tamaño de la muestra.

$p$  : proporción de interés.

$$q = 1 - p$$

El efecto de diseño se debe incorporar al cálculo del error por tratarse de un esquema de muestreo complejo. Debido a que el muestreo sería de conglomerados (secciones electorales) se considera un efecto de diseño aproximado de 2.