

APÉNDICE METODOLÓGICO

En términos generales, las encuestas de victimización se llevan a cabo en distintos lugares del mundo debido a la imposibilidad de medir el fenómeno de la criminalidad exclusivamente a partir de los hechos delictivos denunciados al sistema penal, ya que muchos delitos no son denunciados. Precisamente, para contribuir a superar esta problemática, el Centro de Investigación y Docencia Económicas (CIDE) emprendió la realización de la Encuesta de Victimización y Eficacia Institucional (ENVEI).

El objetivo de esta encuesta es proveer información para contribuir a un diagnóstico más certero del problema de la delincuencia que afecta a la población de 15 años y más. Asimismo, la encuesta busca identificar cómo evalúa la ciudadanía la actuación de las instituciones gubernamentales encargadas de la seguridad pública y de la procuración de justicia.

La encuesta tiene una periodicidad semestral, dado que el propósito central de nuestro instrumento es identificar tendencias en la estacionalidad de ciertos fenómenos delictivos, así como las variaciones en la actuación institucional de las organizaciones encargadas de la seguridad pública y de la procuración de justicia, tanto en el Distrito Federal, como en el Estado de México.

De esta manera esta encuesta tiene por objeto realizar mediciones longitudinales. Con este instrumento no se pretende únicamente medir los delitos y su incidencia, sino que fundamentalmente se desea tener una herramienta que mida la Eficacia Institucional, esto es muy difícil de encontrar en el mercado. La fortaleza de este estudio es que mide el delito, las percepciones y actitudes de los individuos frente a él, y la eficacia de las instituciones. Todas estas variables son medidas a través del tiempo para poder compararlas y estimar diversas tendencias de ellas.

Las poblaciones objetivo de las encuestas ENVEI son diferentes. En el diseño inicial (Julio de 2005), la población objetivo fue la población con edad mayor a 15 años que habitaba en viviendas particulares del Distrito Federal o los municipios conurbados del Estado de México, esto es, la Zona Metropolitana de la Ciudad de México (ZMCM). El primer levantamiento (Julio de 2005) y el segundo levantamiento (Enero de 2006) se hicieron sobre dicha población.

Municipios del Estado de México incluidos en el Diseño Inicial

Municipio INEGI	Clave INEGI	Municipio INEGI	Clave INEGI
Acolman	002	Naucalpan de Juárez	057
Atenco	011	Nezahualcóyotl	058
Atizapán de Zaragoza	013	Nicolás Romero	060
Coacalco de Berriozabal	020	Paz, La	070
Cuatitlán	024	Tecamac	081
Chalco	025	Tepotzotlán	095
Chicoloapan	029	Texcoco	099
Chimalhuacán	031	Tlalnepantla de Baz	104
Ecatepec	033	Tultepec	108
Huixquilucan	037	Tultitlán	109
Ixtapaluca	039	Cuautitlán Izcalli	121
Jaltenco	044	Valle de Chalco Solidaridad	122
Melchor Ocampo	053		

Ahora bien, con base en la distribución geográfica detallada, el diseño muestral de la ENVEI es un procedimiento que permite seleccionar muestras probabilísticas de individuos en las áreas mencionadas. Las muestras de individuos son probabilísticas porque todos los individuos tienen una probabilidad conocida y positiva de ser seleccionados. Las muestras probabilísticas tienen la ventaja de permitir conocer la distribución de frecuencia de los estimadores de parámetros. La selección de individuos se hace por medio de la selección de hogares; esto es, se seleccionará una muestra probabilística de hogares de las áreas mencionadas, para después obtener información sobre los incidentes delictivos sufridos por las personas con quince años cumplidos o más.

La identificación del hogar se deja al entendimiento del entrevistado y el entrenamiento del entrevistador. Entonces, es de esperarse que la mayor parte de hogares identificados correspondan en realidad a viviendas, ya que el entrevistador si verifica que las entradas son independientes en el momento de hacer el recorrido sistemático de las manzanas; pero el cuestionario (entrenamiento) no le obliga a identificar a los hogares.

Esto es así dado que existe poca diferencia porcentual entre el número de viviendas particulares y hogares. Por ejemplo, en el conteo de población 2005, el INEGI reporto para el D.F. a 2,292,069 hogares y 2,287,189 viviendas particulares habitadas. En consecuencia, las razones de variables para hogares y viviendas difieren poco como se muestra a continuación. Sea RV una razón de variables para viviendas y RH una razón de variables para hogares

Total de la característica X en la población

RV =

Total de viviendas particulares en la población

Total de la característica X en la población

RH =

Total de hogares en la población

Entonces, $RV = RH \cdot 1.00218$. Así, una proporción calculada sobre hogares no es diferente en más del 0.5 % de una proporción calculada sobre viviendas.

En resumen, la decisión de ignorar a los hogares dentro de las viviendas se justificó por dos razones. En primer lugar, la ENVEI no tenía como objetivo estudiar temas relativos a la economía doméstica de las viviendas. En segundo, las proporciones calculadas sobre hogares o viviendas son similares para el tipo de variables que se miden en la ENVEI.

La información sobre los incidentes delictivos se obtiene por medio de dos tipos de informantes: el informante directo y el informante de hogar. Los incidentes delictivos que son de interés se clasifican en la tabla siguiente, de acuerdo al tipo de informante.

Incidentes delictivos de interés clasificados de acuerdo al tipo de informante

Informante de Vivienda	Informante de Vivienda
(A) Robo de automóvil	(D) Robo personal con o sin violencia
(B) Robo de partes de automóvil	(F) Amenaza
(C) Robo a casa-habitación	
(E) Secuestro	
(X) Indica la sección del cuestionario	

El informante del hogar puede ser cualquier persona con 15 años cumplidos o más; no se hace selección aleatoria, sólo se pregunta por una persona disponible para proporcionar la información. Las fechas de levantamiento de la primera y segunda vueltas y los periodos de los incidentes delictivos estudiados en cada una de ellas se detallan a continuación:

Fechas de levantamiento por vuelta

Vuelta	Fecha de levantamiento	Periodo analizado
1	4-21 de Agosto de 2005	Enero 2005 – Agosto 2005
2	12-27 de Enero de 2006	Julio 2005 – Diciembre 2005

Adicionalmente, el informante del hogar proporciona detalles de los incidentes delictivos de magnitud y características que se suponen conocidos por todos los individuos del hogar. En contraste, los detalles de los incidentes delictivos difícilmente conocidos por todos los individuos del hogar son preguntados directamente a la persona que sufrió el incidente. Cuando una persona sufrió más de un incidente del mismo tipo en el periodo de referencia, se aplica un cuestionario de incidente sólo al incidente más reciente. Es necesario mencionar que los levantamientos se realizan los fines de semana para obtener una mejor representación poblacional.

En un estudio de esta magnitud uno de los elementos claves es el diseño muestral. A continuación, se describen los elementos básicos de éste. En primer lugar el tamaño de la muestra; para después describir la selección de hogares y la formación de estimadores.

A. Tamaño de muestra

Los recursos disponibles en cada uno de los levantamientos permitieron obtener como resultado final el siguiente número de observaciones:

Número de Observaciones por Levantamiento

Levantamiento	Número de Observaciones
1	2,000
2	1,504

Respecto a las incidencias que se esperan observar en las muestras, la Encuesta Nacional sobre Seguridad Pública en las Entidades Federativas (ICESI-2002) permite estimar la incidencia de delitos esperada.

Incidencia delictiva anual por individuo

DF	18%
EDOMEX	5%
NACIONAL	4%

Incidencia delictiva anual por hogar (al menos un delito)

DF	39%
EDOMEX	20%
NACIONAL	14%

En virtud de que tanto las zonas del D.F. como las del Estado de México analizadas en cada una de las vueltas tienen características económicas y demográficas similares, se asume que las incidencias de delitos sean similares en ambas.

Las incidencias delictivas son parámetros estimados a través de encuestas; entonces, para identificar cambios en las incidencias se calcularon los intervalos de confianza por levantamiento y por año, esto para los principales tipos de delito de la encuesta:

- 1) Robo de auto a algún integrante del hogar.
- 2) Robo de partes de automóvil a algún auto del hogar.
- 3) Robo a casa habitación.
- 4) Intento de robo a casa habitación.
- 5) Secuestro de la persona entrevistada.
- 6) Hogares con alguna persona secuestrada.
- 7) Secuestros por habitante.
- 8) Persona entrevistada víctima de robo personal.
- 9) Alguna persona del hogar víctima de robo personal.
- 10) Persona entrevistada víctima de agresión o amenaza.
- 11) Alguna persona del hogar víctima de agresión o amenaza.

Incidencias delictivas por levantamiento

Levantamiento	1	2
Robo de auto a algún integrante del hogar		
Estimador	5	4.4
Límite Inferior	3.6	2.7
Límite Superior	7	7.1
Robo de partes de automóvil a algún auto del hogar		
Estimador	17.4	18.4
Límite Inferior	14.2	15.2
Límite Superior	21.2	22.2
Robo a casa habitación		
Estimador	2.1	2.6
Límite Inferior	1.2	1.4
Límite Superior	3.5	4.8
Intento de robo a casa habitación		
Estimador	0.7	0.9
Límite Inferior	0.3	0.4
Límite Superior	1.7	2.3
Secuestro de la persona entrevistada		
Estimador	0.3	0.3
Límite Inferior	0.1	0
Límite Superior	1.3	1.8
Hogares con alguna persona secuestrada		
Estimador	0.6	0.5
Límite Inferior	0.3	0.1
Límite Superior	1.4	1.8
Secuestros por habitante		
Estimador	0.3	0.2
Límite Inferior	0.1	0
Límite Superior	0.5	0.6
Persona entrevistada víctima de robo personal		
Estimador	12	6.8
Límite Inferior	9.6	5
Límite Superior	15	9.1
Alguna persona del hogar víctima de robo personal		
Estimador	16.9	13.1
Límite Inferior	13.9	10.6
Límite Superior	20.3	16.1

Persona entrevistada víctima de agresión o amenaza		
Estimador	2.9	2.7
Límite Inferior	1.8	1.5
Límite Superior	4.8	5
Alguna persona del hogar víctima de agresión o amenaza		
Estimador	4.1	4.5
Límite Inferior	2.8	2.9
Límite Superior	6	6.9

Incidencias delictivas por año

Levantamiento	1
Robo de auto a algún integrante del hogar	
Estimador	4.7
Límite Inferior	3.6
Límite Superior	6.3
Robo de partes de automóvil a algún auto del hogar	
Estimador	17.9
Límite Inferior	15.6
Límite Superior	20.6
Robo a casa habitación	
Estimador	2.4
Límite Inferior	1.6
Límite Superior	3.6
Intento de robo a casa habitación	
Estimador	0.8
Límite Inferior	0.5
Límite Superior	1.6
Secuestro de la persona entrevistada	
Estimador	0.3
Límite Inferior	0.1
Límite Superior	1.2
Hogares con alguna persona secuestrada	
Estimador	0.6
Límite Inferior	0.3
Límite Superior	1.3
Secuestros por habitante	
Estimador	0.3
Límite Inferior	0.1
Límite Superior	0.5
Persona entrevistada víctima de robo personal	

Estimador	9.4
Límite Inferior	7.9
Límite Superior	11.3
Alguna persona del hogar víctima de robo personal	
Estimador	15
Límite Inferior	13
Límite Superior	17.3
Persona entrevistada víctima de agresión o amenaza	
Estimador	2.8
Límite Inferior	2
Límite Superior	4.3
Alguna persona del hogar víctima de agresión o amenaza	
Estimador	4.3
Límite Inferior	3.3
Límite Superior	5.8

B. Selección de hogares

El esquema de muestreo de selección de hogares es estratificado y en dos etapas. En primer lugar, se construyó un índice de marginalidad por AGEB (Áreas Geo-estadísticas Básicas) inspirado en el índice de marginalidad municipal construido por CONAPO. A partir del índice de marginalidad se construyeron cinco estratos de igual tamaño. El índice de marginalidad se construyó al aplicar componentes principales a las variables siguientes de AGEB.

Variables incluidas en el índice de marginalidad

Z_1	Porcentaje de la población 15 años o mas que es alfabeto
Z_2	Porcentaje de la población 15 años o mas sin primaria completa
Z_3	Porcentaje de viviendas sin drenaje
Z_4	Porcentaje de viviendas sin energía eléctrica
Z_5	Porcentaje de viviendas sin disponibilidad de agua entubada
Z_6	Porcentaje de viviendas con un solo cuarto (Cuarto redondo)
Z_7	Porcentaje de viviendas con piso sin recubrimiento
Z_8	Porcentaje de la población ocupada que recibe hasta 2 salarios mínimos mensuales

El análisis de componentes principales transforma Z_1, \dots, Z_8 en C_1, \dots, C_8 , de tal suerte que los componentes principales son no correlacionados; esto es $Cov(C_i, C_j) = 0$ para i diferente de j ; además, los componentes principales reproducen la varianza de las variables originales $\sum Var(Z_i) = \sum Var(C_i)$.

Los estratos de marginalidad se definieron del modo siguiente. Se ordenaron a las AGEB de acuerdo al primer componente principal. Se formaron cinco estratos de AGEB tales que: a) la población contenida en cada estrato era la misma, y b) las AGEB en los estratos eran *consecutivas* de acuerdo al orden inducido por el primer componente principal.

La importancia de los componentes principales puede medirse por el porcentaje de la varianza explicada con respecto de la suma de las varianzas de Z_1, \dots, Z_8 .

Porcentaje de varianza explicada por los componentes principales

Componente	Porcentaje de Varianza
C ₁	56%
C ₂	19%
C ₃	9%
C ₄	6%
C ₅	4%
C ₆	3%
C ₇	2%
C ₈	1%

En nuestro caso, como el primer componente principal representa al 56% de la variabilidad total de las ocho variables, se considera que el primer componente resume razonablemente a las ocho variables. Después de definir a los estratos de marginalidad se definieron a los estratos geográficos-marginalidad que resultaron de cruzar los estratos de marginalidad con el estado al cual pertenece la vivienda.

Cuando se amplió la población objetivo de la encuesta, se añadieron nuevos estratos geográficos-marginalidad para la población incluida, sin recalcular los componentes principales o redefinir los estratos de marginalidad. Los nuevos estratos geográficos-marginalidad, se formaron del modo siguiente. Cada nueva AGEB se clasificó en uno de los cinco estratos de marginalidad de acuerdo al valor del componente principal, respetando el criterio de que cada estrato tiene la misma población.

Como una muestra de que los nuevos estratos geográficos-marginalidad son comparables con los estratos geográficos-marginalidad iniciales, la tabla siguiente muestra la proporción de población ocupada que gana hasta 2 salarios mínimos para cada estrato geográfico-marginalidad.

Porcentaje de la población ocupada que gana hasta dos salarios mínimos

Estrato Geográfico	Marginalidad	Porcentaje
DF	Muy baja	26.4%
	Baja	43.5%
	Media	52.4%
	Alta	59.4%
	Muy alta	66.6%
Mexico Connurbado	Muy baja	29.7%
	Baja	44.2%
	Media	49.4%
	Alta	56.6%
	Muy alta	62.8%
Resto Mexico	Muy baja	33.5%
	Baja	55.3%
	Media	68.6%
	Alta	76.8%
	Muy alta	80.7%

Después de formar a los estratos se procedió a seleccionar a las viviendas en dos etapas:

- El tamaño de muestra por estrato es proporcional al tamaño relativo del estrato.

- b. El tamaño de muestra por estrato se dividió entre 16, para obtener el número de AGEB a seleccionar por estrato.
- c. En cada uno de los 10 estratos, las AGEB se ordenaron de acuerdo al municipio o delegación e índice de marginalidad.
- d. En cada estrato, las AGEB se seleccionaron con un muestreo sistemático y probabilidad proporcional al tamaño.
- e. Dentro de cada AGEB, las viviendas son seleccionadas con el procedimiento siguiente:
 - i. Se numeran las manzanas de la AGEB y se selecciona con igual probabilidad a una de ellas.
 - ii. Se recorre la manzana iniciando en la esquina noreste y en el sentido de las manecillas del reloj hasta conseguir una entrevista.

Aunque la selección de viviendas no es estrictamente probabilista, la selección aleatoria de la manzana trata de emular una selección aleatoria de viviendas, y a la vez mantiene los costos de muestreo en niveles que permiten realizar la encuesta.

C. Selección de incidentes

Cuando una persona sufrió más de un incidente del mismo tipo en el periodo de referencia, se aplica un cuestionario de incidente sólo al más reciente.

D. Construcción de estimadores

Sea Y una variable medida a las viviendas. Asumiendo que la no respuesta es cero, el total de la variable Y , en el estrato i se estima con:

$$\hat{Y}_i = \sum_j w_{j(i)} \hat{Y}_{j(i)} \quad \hat{Y}_{j(i)} = \sum_k \hat{w}_{k(j)} Y_{k(j)}$$

donde $Y_{k(j)}$ es el valor de Y en la vivienda k , de la AGEB j del estrato i , y

$$w_{j(i)} = \frac{Viv.Estrato_i}{m_i Viv.AGEB_{j(i)}}, \quad \hat{w}_{k(j)} = \frac{Viv.Ageb_{j(i)}}{16},$$

donde m_i es el número de AGEB seleccionadas en el estrato i . En consecuencia, el factor de ponderación para las viviendas del estrato i es

$$\hat{w}_{k(j)} w_{j(i)} = \frac{Viv.Estrato_i}{16 * m_i};$$

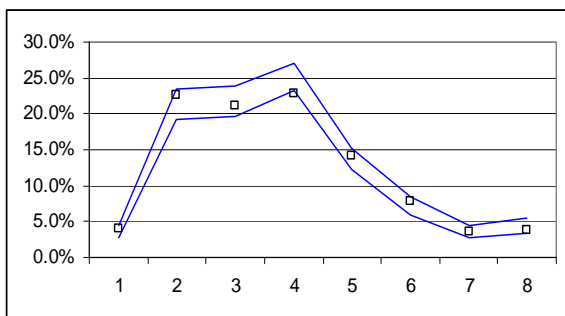
$\hat{w}_{k(j)} w_{j(i)}$ es aproximadamente constante porque el tamaño de muestra por estrato (m_i) es aproximadamente proporcional a las viviendas en el estrato.

El total del Y es estimado con $\hat{Y} = \sum_i \hat{Y}_i$. La razón de las medias de Y y Z se estima con el cociente \hat{Y} / \hat{Z}

En seguida, los ponderadores $w_{j(i)} \hat{w}_{k(j)}$ son modificados para tratar de corregir posibles sesgos inducidos por la no respuesta; en esta encuesta los ponderadores $w_{j(i)} \hat{w}_{k(j)}$ son ajustados en dos etapas:

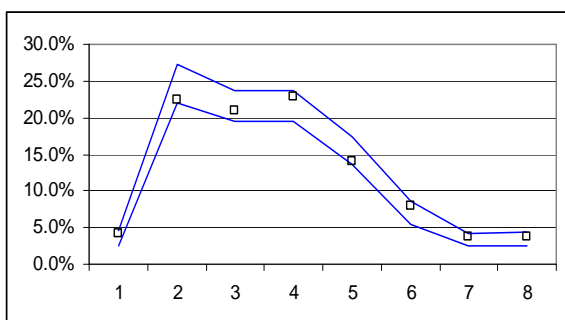
- a) La suma de $w_{j(i)} \hat{w}_{k(j)}$ se estandariza para que reproduzca el total estimado de hogares donde existen habitantes con edad mayor o igual a quince años. Después de que se ha conseguido la aceptación de la persona que abrió la puerta, se pregunta por todas las personas mayores de 15 que habitan en el hogar. Los tabulados siguientes muestran que la distribución de los hogares según el número de personas mayores a 15 años no ha cambiado significativamente. Las graficas muestran los intervalos de confianza comparados con la distribución promedio de cada uno de las encuestas para la primera y la segunda vueltas:

Encuesta 1. Distribución de hogares según número de integrantes. Los intervalos de confianza obtenidos de la encuesta se muestran en azul.



	Proporción	L. Inferior	L Superior
1	3.5%	2.6%	4.5%
2	21.3%	19.2%	23.5%
3	21.7%	19.6%	23.9%
4	25.1%	23.2%	27.1%
5	13.6%	12.2%	15.1%
6	7.2%	6.0%	8.5%
7	3.5%	2.7%	4.5%
>7	4.3%	3.3%	5.4%

Encuesta 2. Distribución de hogares según número de integrantes. Los intervalos de confianza obtenidos de la encuesta se muestran en azul.



	Proporción	L. Inferior	L Superior
1	3.5%	2.6%	4.7%
2	24.5%	22.0%	27.2%
3	21.5%	19.5%	23.6%
4	21.5%	19.5%	23.8%
5	15.5%	13.7%	17.4%
6	6.9%	5.5%	8.6%
7	3.3%	2.5%	4.3%
>7	3.3%	2.5%	4.5%

El hecho de que los intervalos de confianza cubren al valor promedio, es un hecho que fortalece a la encuesta ya que muestra que al menos con respecto al número de habitantes, la encuesta está colectando información sobre el mismo tipo de hogares.

- b) La suma de $w_{j(i)} w_{k(i)}$ se estandariza para que reproduzca de manera aproximada el total de hogares en el Área Metropolitana. El total de hogares se estimó del modo siguiente: a 11 partir de las proyecciones de CONAPO para 2005 se estimó al total de la población 15 y más, total que se dividió por 2.9 para obtener una estimación del número de hogares. Cabe aclarar, que cuando se conocieron los resultados del censo 2005, los ponderadores se ajustaron para obtener las mejores estimaciones de totales poblacionales.

Total de hogares e individuos estimados

Encuesta	Hogares	Individuos 15 y +	Promedio
1	4590	13567	2.96
2	4590	13090	2.85

- c) Aunque la dificultad para obtener una entrevista es dependiente del estrato de marginalidad, este factor no se consideró en la ponderación porque se recorrieron las AGEB hasta completar 16 entrevistas por cada AGEB en muestra.

Promedio de hogares visitados para conseguir una entrevista

Levantamiento	1	2
Hogares	4.02	3.51
Límite Inferior	3.72	3.16
Límite Superior	4.32	3.87

Nota: Intervalos de confianza al 95%.

Por su parte, los cuestionarios de incidentes se ponderaron al multiplicar el ponderador de vivienda por el número de incidentes de un mismo tipo que un respondiente experimentó durante el período de referencia. En consecuencia, la suma de ponderadores de los cuestionarios de incidente estima al total de incidentes ocurridos en el período de referencia.