



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
INSTITUTO DE INVESTIGACIONES JURÍDICAS
DEPARTAMENTO DE INVESTIGACIÓN APLICADA Y OPINIÓN



GOBIERNO DE LA
CIUDAD DE MÉXICO

SECRETARÍA DE EDUCACIÓN, CIENCIA,
TECNOLOGÍA E INNOVACIÓN

Inventario CDMX. Presente y futuro de su Gente

Diseño Muestral

Introducción

El estudio *“Inventario CDMX. Presente y futuro de su gente”* muestra un panorama profundo de la Ciudad de México desde el punto de vista de mujeres y hombres. Con ello se busca obtener una visión integral de las transformaciones que ha tenido esta Ciudad en los últimos años y así proporcionar recursos para la toma de decisiones.

Para ello, se llevarán a cabo 10 encuestas de 1,000 casos cada una, a personas de 15 años y más distribuidas dentro de las 16 alcaldías que comprende la Ciudad de México. Los temas que comprenderán estas encuestas son las siguientes:

1. Economía y empleo
2. Educación, ciencia y tecnología
3. Salud
4. Niñas, niños, adolescentes y jóvenes
5. Envejecimiento
6. Cultura, lectura y deporte
7. Vivienda
8. Movilidad y transporte
9. Medio ambiente y estrategias para la sustentabilidad
10. Seguridad pública y justicia

El propósito de este apartado es dar a conocer la metodología asociada al diseño muestral general de la investigación. Es importante resaltar que se realizarán 10 muestras independientes que seguirán el mismo diseño muestral, el cual se describe en este documento. En principio se encuentran las fuentes de información que serán utilizadas para la selección de las etapas y unidades de muestreo. Después, se describe el diseño

muestral que se implementará para las encuestas antes mencionadas y finalmente se presentan las características y procedimiento para la obtención del tamaño de muestra.

Objetivos:

- Realizar un diagnóstico amplio y profundo de la Ciudad de México, a través de las percepciones de la gente en materia de: economía y empleo; educación, ciencia y tecnología; salud; niñas, niños, adolescentes y jóvenes; envejecimiento; cultura, lectura y deporte; vivienda; movilidad y transporte; medio ambiente y estrategias para la sustentabilidad, así como en seguridad pública y justicia.
- Generar una línea base que permita establecer un horizonte básico para la medición de opiniones y necesidades de los habitantes de la Ciudad.

Población objetivo

La población objetivo está constituida por todos los habitantes de 15 años o más dentro de las 16 alcaldías que constituyen la Ciudad de México.

Marco Muestral

Se utilizará el software interactivo SINCE 2010 versión 1.1.2 del Instituto Nacional de Estadística y Geografía (INEGI) para la recopilación de los datos de la población desagregados en el ámbito de localidad y de Área Geoestadística Básica (AGEB) urbana.

Estratificación

Para fines de la selección de la muestra, la Ciudad de México será dividida en 4 regiones geográficas con características comunes, así como por una variable de estratificación construida a través de métodos multivariados, que agrupa las principales características socioeconómicas de la población de la Ciudad de México en tres niveles: alto, medio y bajo (Véase *Anexo I*).

En el *Anexo I* de este documento, se describen los métodos para la construcción de la variable de estratificación a través de variables socioeconómicas de la población.

Las regiones se describen en el Cuadro 1.

Cuadro1. Regiones de la Ciudad de México

Región	Alcaldías de la Ciudad de México
Poniente	Azcapotzalco, Cuajimalpa de Morelos, La Magdalena Contreras, Álvaro Obregón, Miguel Hidalgo
Sur	Milpa Alta, Tláhuac, Tlalpan, Xochimilco
Oriente	Gustavo A. Madero, Iztacalco, Iztapalapa, Venustiano Carranza
Centro	Coyoacán, Benito Juárez, Cuauhtémoc

Bajo los criterios anteriormente descritos se obtendrán un total de 11 estratos para la Ciudad de México.

En el Cuadro 2, se describe la distribución de la población de 15 años y más por estrato.

Cuadro 2. Distribución de la población de 15 años y más

Estrato	Región	Clasificación	Número de personas de 15 años y más
1	Centro	Alto	762,630
2	Centro	Medio	444,330
3	Oriente	Alto	524,196
4	Oriente	Medio	1,403,418
5	Oriente	Bajo	941,860
6	Poniente	Alto	594,106
7	Poniente	Medio	694,385
8	Poniente	Bajo	188,806
9	Sur	Alto	241,052
10	Sur	Medio	473,624
11	Sur	Bajo	433,666

Tamaño de la muestra

El tamaño de muestra para cada encuesta será de 1,000 casos. Dicho tamaño de muestra será distribuido entre los diferentes estratos, con el fin de generar una disminución de los márgenes de error estadísticos. Para determinarlo se consideró la siguiente fórmula:

$$n = \frac{(z^2)(p)(1 - p)(Deff)}{d^2(1 - TNR)}$$

donde:

- **n** = Tamaño de la muestra
- **z^2** = Es el valor en tablas de valores de probabilidad acumulada para la distribución normal estándar. Este valor depende del nivel de confianza asignado $(1-\alpha)*100\%$.
- **p** = Probabilidad de éxito del evento. Se refiere a la probabilidad de éxito esperada. El más conservador de los valores se obtiene cuando $p = 0.5$.
- **d^2** = es la diferencia entre el valor estimado y el valor poblacional (en este caso, elevado al cuadrado).
- **TNR** = Tasa de no respuesta esperada.
- **$Deff$** = Efecto de Diseño por utilizar un muestreo diferente al muestreo aleatorio simple.

Considerando una proporción de 0.5, error muestral máximo de 4.2%, nivel de confianza del 95%, efecto de diseño de 1.65 y tasa de no respuesta máxima de 10%, se calculó para cada muestra un total aproximado de 1,000 casos.

Afijación de la muestra

La afijación se realizará entre los diferentes estratos de manera proporcional al tamaño de la población, para lo cual se empleará la siguiente expresión:

$$n_h = \frac{N_h}{N} n$$

Donde:

n_h número de personas de 15 años y más en muestra en el h-ésimo estrato

n tamaño de la muestra

N_h número total de personas de 15 años y más en el h-ésimo estrato

N número total de personas de 15 años y más que habitan en la Ciudad de México.

El Cuadro 3 describe el número de casos que se buscarán dentro de cada estrato.

Cuadro 3. Distribución de casos de cada muestra

Estratos	Unidades de muestreo UPM	Número de casos
1	10	100
2	7	70
3	8	80
4	18	180
5	13	130
6	9	90
7	10	100
8	6	60
9	6	60
10	7	70
11	6	60

Esquema de selección

La selección de unidades corresponde a un diseño muestral probabilístico, estratificado y se realizará de manera independiente mediante etapas sucesivas para cada estrato definido.

- 1) *Unidades primarias de muestreo (UPM)*: Son las unidades geográficas denominadas AGEB, las cuales son definidas por el INEGI. En el caso de las zonas rurales, las UPM estarán conformadas por las localidades rurales.
- 2) *Unidades secundarias de muestreo (USM)*: Son las viviendas particulares habitadas contenidas dentro de cada AGEB seleccionada en la etapa de muestreo anterior.
- 3) *Unidades últimas de muestreo (UUM)*: Es la población de 15 años o más que reside en la vivienda seleccionada en la etapa anterior.

La selección de la muestra se realiza de manera independiente para cada estrato. El procedimiento es el siguiente:

- 1) Para cada uno de los estratos se seleccionarán k UPM's con probabilidad proporcional al tamaño de la población objetivo.
- 2) En cada UPM se seleccionarán 10 viviendas de manera sistemática y con un arranque aleatorio.
- 3) Dentro de cada vivienda se selecciona de manera sistemática y con arranque aleatorio a una persona del total de personas residentes en la vivienda que pertenecen a la población objetivo.

Cálculo de los factores de expansión

Para obtener los factores de expansión utilizados en el cálculo de la estimación de las características de interés de la población en la encuesta, se utilizó la probabilidad de selección correspondiente a cada etapa de muestreo.

La probabilidad de selección del i -ésimo individuo dentro de la j -ésima vivienda en la k -ésima AGEB del h -ésimo estrato es:

$$P[X_{ijkh}] = \frac{n_h P_{kh}}{\sum_{kh} P_{kh}} \frac{10}{M_{kh}} \frac{1}{V_{jkh}}$$

Donde:

X_{ijkh} = Es el i -ésimo individuo en la j -ésima vivienda de la k -ésima AGEB del h -ésimo estrato.

n_k = Es el número de UPM seleccionadas aleatoriamente en el h -ésimo estrato.

P_{kh} = Es el tamaño de la población objetivo perteneciente a la k-ésima AGEB del h-ésimo estrato.

M_{kh} = Es el número de viviendas existentes en la k-ésima AGEB del h-ésimo estrato.

V_{jkh} = Es el número de personas pertenecientes a la población objetivo dentro de la j-ésima vivienda de la k-ésima AGEB en el h-ésimo estrato.

El recíproco de las probabilidades de selección son los factores de expansión correspondientes, es decir:

$$F_{ijkh} = \frac{1}{P[X_{ijkh}]}$$

AJUSTE DE LOS FACTORES DE EXPANSIÓN

Los factores de expansión fueron calibrados con base en la información obtenida durante la encuesta intercensal 2015 realizada por INEGI.

El ajuste se calculó de la siguiente forma:

$$F'_{ijkd} = F_{ijkd} \frac{T'}{T''}$$

Donde:

F'_{ijkd} = Es el factor de expansión corregido para el i-ésimo individuo de la j-ésima vivienda en la k-ésima AGEB del h-ésimo estrato.

F_{ijkd} = Es el factor de expansión natural para el i-ésimo individuo de la j-ésima vivienda en la k-ésima AGEB del h-ésimo estrato.

T'' = Es el total de la población estimada por la encuesta para cada grupo de edad, sexo y estrato.

T' = Es el total de la población estimada por INEGI en 2015 para cada grupo de edad, sexo y estrato.

Anexo I

Construcción de la variable de estratificación

Debido a la particularidad que tiene la Ciudad de México respecto del resto de los estados del país, se buscó elaborar una estratificación que considerara las características económicas, educativas, de vivienda y de accesibilidad a los servicios de salud de la población que habita en la Ciudad. Con el fin de clasificar a la población de acuerdo con su nivel socioeconómico, se utilizaron 56 variables que reflejaran las características socioeconómicas de la zona geográfica de residencia de cada segmento de la población.

La fuente de información utilizada es del Censo de Población y Vivienda 2010, particularmente del SCINCE, versión 2012 del Instituto Nacional de Estadística y Geografía (INEGI).

Transformación de las variables a indicadores

Con el fin de reducir la dimensión de las variables y descartar aquellas que no discriminan a la población, se realizó un análisis de correlaciones entre los indicadores, manteniendo únicamente las variables que aportaban información general. Además, se analizaron los indicadores a través de su variabilidad, descartando aquellos cuyo valor fuera cercano a cero.

Al final de este proceso, se consideraron únicamente 15 indicadores para realizar el proceso de clasificación.

Cálculo de los indicadores

Para los indicadores finales el cálculo es el siguiente:

1. Porcentaje de población de 3 a 5 años que asiste a la escuela

Es la población de 3 a 5 años que asiste a la escuela entre el número de personas de 3 a 5 años.

$$I_EDU1_P = \left(\frac{EDU1}{POB4} \right) \times 100$$

II. Porcentaje de población de 15 años y más con educación básica incompleta

Es la población de 15 años y más con educación básica incompleta entre la población de 15 años y más.

$$EDU34_P = \left(\frac{EDU34}{POB20} \right) \times 100$$

III Porcentaje de población de 18 años y más con al menos un grado aprobado en educación media superior.

Es la población de 18 años y más con al menos un grado aprobado en educación media superior entre la población de 18 años y más.

$$EDU43_P = \left(\frac{EDU43}{POB21} \right) \times 100$$

IV. Porcentaje de población económicamente activa.

Es la población económicamente activa entre la población total.

$$ECO1_P = \left(\frac{ECO1}{POB1} \right) \times 100$$

V. Porcentaje de población ocupada de 12 años y más con secundaria o equivalente completa.

Es la población de 12 años y más con secundaria o equivalente incompleta entre la población de 12 años y más.

$$ECO16_P = \left(\frac{ECO16}{POB19} \right) \times 100$$

VI. Porcentaje de población derechohabiente a los servicios de salud.

Es la población derechohabiente a los servicios de salud entre la población total.

$$SALUD1_P = \left(\frac{SALUD1}{POB1} \right) \times 100$$

VII. Porcentaje de viviendas particulares habitadas con piso diferente de tierra

Es el complemento del porcentaje de las viviendas particulares habitadas con piso diferente de tierra.

$$VIV6_Pc = \left(1 - \frac{VIV6}{VIV2} \right) \times 100$$

VIII. Porcentaje de viviendas particulares habitadas que disponen de luz eléctrica, agua entubada en el ámbito de la vivienda y drenaje.

Son las viviendas particulares habitadas que disponen de luz eléctrica, agua entubada en el ámbito de la vivienda y drenaje dividido por el número de viviendas particulares habitadas.

$$VIV24_P = \left(\frac{VIV24}{VIV2} \right) \times 100$$

IX. Porcentaje de viviendas particulares habitadas que disponen de refrigerador.

Son las viviendas particulares habitadas que disponen de refrigerador entre el número de viviendas particulares habitadas.

$$VIV26_P = \left(\frac{VIV26}{VIV2} \right) \times 100$$

X. Porcentaje de viviendas particulares habitadas que disponen de lavadora

Son las viviendas particulares habitadas que disponen de lavadora entre el número de viviendas particulares habitadas.

$$VIV27_P = \left(\frac{VIV27}{VIV2} \right) \times 100$$

XI. Porcentaje de viviendas particulares habitadas que disponen de automóvil o camioneta

Son las viviendas particulares habitadas que disponen de automóvil o camioneta entre el número de viviendas particulares habitadas.

$$VIV28_P = \left(\frac{VIV28}{VIV2} \right) \times 100$$

XII. Porcentaje de viviendas particulares habitadas que disponen de computadora

Es el número de viviendas particulares habitadas que disponen de computadora entre el número de viviendas particulares habitadas.

$$VIV33_P = \left(\frac{VIV33}{VIV2} \right) \times 100$$

XIII. Porcentaje de viviendas particulares habitadas que disponen de línea telefónica fija.

Es el número de viviendas particulares habitadas que disponen de línea telefónica fija entre el número de viviendas particulares habitadas.

$$VIV34_P = \left(\frac{VIV34}{VIV2} \right) \times 100$$

XIV. Porcentaje de viviendas particulares habitadas que disponen de teléfono celular

Es el número de viviendas particulares habitadas que disponen de teléfono celular entre el número de viviendas particulares habitadas.

$$VIV35_P = \left(\frac{VIV35}{VIV2} \right) \times 100$$

XV. Porcentaje de Viviendas particulares habitadas que disponen de internet

Es el número de viviendas particulares habitadas que disponen de internet entre el número de viviendas particulares habitadas.

Método de estratificación

Finalmente, se aplicó el método de K-medias sobre los 15 indicadores para crear tres agrupaciones de AGEBS con características socioeconómicas similares. Los grupos fueron etiquetados como “alto”, “medio” y “bajo”, de acuerdo al orden de su valor en el centroide.