

Primera Encuesta Estatal de Movilidad Social en Puebla (EMOVIP, 2015)

Diseño de Muestreo

Contenido

Antecedentes	1
Objetivos de la encuesta	1
Premisas para el diseño de la muestra.....	2
Población objetivo.....	2
Marco Muestral.....	2
Tipo de indicadores	3
Subconjuntos poblacionales para los que interesa generar estimaciones propias	3
Tamaño de la muestra y precisión esperada	4
Dominios de estudio.....	4
El proceso de selección de las unidades	5
Primera etapa.....	8
Segunda etapa.....	11
Resto de las Etapas.....	11
Probabilidad de selección y factores de expansión	13

Antecedentes

En 2006 se levantó la primera encuesta estatal para captar información acerca de la movilidad social en México, y en 2011 se llevó a cabo la segunda encuesta estatal sobre el mismo tema, aunque con mayor alcance analítico. Ahora la Universidad Popular Autónoma del Estado de Puebla (UPAEP) y el Centro de Estudios Espinosa Yglesias (CEEY) establecieron un convenio para realizar por primera vez un estudio de esta naturaleza pero a nivel estatal. El presente documento refiere la metodología aplicada para realizar la selección de la muestra probabilística correspondiente.

Objetivos de la encuesta

- Disponer de información para cubrir conceptos relacionados con la movilidad social a nivel del Estado de Puebla.
- Captar información de corte inter-generacional, para conocer las relaciones entre los recursos económicos y sociales de los padres y el logro socioeconómico de sus hijos.
- Obtener información para identificar las posibles barreras que más limitan a las familias mexicanas para que experimenten movilidad social, ya sea absoluta o relativa, y para observar el logro del esfuerzo personal y de las posibilidades con que pueden disponer para establecer, desarrollar y lograr los proyectos laborales y de vida que vislumbran.

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Premisas para el diseño de la muestra

La muestra debe armonizar con los objetivos de la encuesta y con los siguientes elementos:

- la población objetivo
- el marco de muestreo a partir del cual se intentará poder llegar a las personas de interés
- el tipo de indicadores que interesa puedan estimarse a través de la información recolectada en la encuesta
- los subconjuntos poblacionales para los que interesa generar resultados diferenciados
- el tamaño de la muestra, acorde al alcance de la encuesta
- el proceso de selección de las unidades y
- los factores que permitan generar las estimaciones derivadas de la encuesta.

Población objetivo

La **población objeto de estudio** es mujeres y hombres de entre 30 y 59 años de edad, residentes habituales de los ámbitos urbano y rural en el Estado de Puebla.

Para el estudio, hay interés particular por conocer los patrones de movilidad social de personas que tienen ocupaciones laborales, pero también interesa conocer los patrones de movilidad social de subgrupos poblacionales que pertenecen al grupo de 30-59, como amas de casa y estudiantes.

Marco Muestral

Para seleccionar una muestra, entre otras cosas, se necesita disponer de un marco de las unidades de interés, para que pueda procederse a la selección.

Una premisa para que un marco funcione de manera adecuada es primeramente que la población objetivo esté claramente identificada y que el listado satisfaga los principios que todo marco de muestreo debe contener:

- a) Que las unidades sean plenamente identificables;
- b) Que las unidades pueden ser encontradas;
- c) Que el marco está organizado sistemáticamente;
- d) Que se incluyan otras variables, que pueden llamarse auxiliares, para que pueden ser usadas para diseñar y conformar estratos;
- e) Que se puede realizar una identificación explícita de dominios, en caso de que se requieran;
- f) Que remitan sin duda a cubrir el principio de unicidad, es decir, que cada elemento de la población se presente sólo una vez;
- g) Que sea exhaustivo, es decir que cada elemento de la población está presente en el marco, y
- h) Que sea posible excluir los elementos que no son de interés o no forman parte del universo de estudio.

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Para el estudio que nos ocupa, no se dispone de un listado actualizado y nominal de la población objeto de estudio, residente habitual en el Estado de Puebla. Por lo cual, la muestra tendrá que elegirse mediante la selección de unidades por aproximación sucesiva cada vez más específicas, hasta llegar a las personas que interesa aplicar la entrevista cara a cara. Es decir, se tratará de un muestreo de conglomerados, que se elegirá por etapas en unidades cada vez más específicas. Primero se elegirán municipios, luego áreas geoestadísticas básicas (AGEBS), luego manzanas o caseríos de zonas rurales, luego viviendas y, por últimos, residentes que cumplan con el perfil requerido en la población objetivo.

Si por el contrario, se hubiera dispuesto de un listado actualizado de viviendas donde se supiese que reside alguna persona de las poblaciones de interés para el estudio, se habría realizado un muestreo aleatorio simple a partir de dicho listado. Como no fue el caso, se empleó un esquema de muestreo de conglomerados y selección por etapas.

Tipo de indicadores

Todo tamaño de muestra está condicionado por el tipo de indicadores que interesa estimar con la encuesta. En general, ante premisas de confianza y precisión equivalentes entre sí, una muestra para estimar proporciones es más reducida que una muestra para estimar medias y a su vez, una para estimar medias requiere menor tamaño que una para estimar totales. Si bien hay otro tipo de indicadores que pueden obtenerse a través de encuesta, salen fuera del entorno de este documento.

Dada la temática del estudio, se concibe que la mayor parte de los indicadores que interesa derivar con la encuesta son proporciones, por lo que el tamaño de muestra deberá considerar esta premisa de cálculo.

Subconjuntos poblacionales para los que interesa generar estimaciones propias

Para el estudio interesa disponer de resultados diferenciados para la zona metropolitana de la Ciudad de Puebla y la entidad en conjunto, es decir que interesaba que fueran lo que técnicamente se conoce como dominios de estudio.

Una entidad como Puebla se caracteriza por un mosaico cultural y social diverso, por lo que para seleccionar la muestra es menester tomar en cuenta esta característica. En este sentido, se deberá elegir la muestra de forma tal que se asegure que se capta dicha diversidad. Si bien más adelante se detalla el proceso de selección, en este momento baste con señalar que se contemplarán las distintas regiones socioeconómicas que conforman la entidad, y de cada una se elegirán muestras de forma independiente, para asegurar que se tienen personas elegidas aleatoriamente de toda región contemplada. Además, se considerarán criterios que aún dentro de esas regiones permitan asegurar que se eligen personas de diferentes condiciones de vida. Dentro de la misma zona metropolitana de Puebla también hay diferencias que deberán contemplarse durante el proceso de selección.

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Tamaño de la muestra y precisión esperada

El tamaño de muestra debe responder a los objetivos de la investigación y a las condicionantes que se experimentarán para poder contactar a la población de estudio. Como se dijo antes, no se dispone de un listado actualizado de las unidades de interés, por lo que la muestra deberá ser elegida en aproximaciones sucesivas de especificidad, hasta llegar a las personas entrevistadas. Por tanto se tratará de una muestra de unidades conglomeradas, con selección polietápica. Esto obliga a que se contemple un factor de cálculo denominado efecto de diseño, que corresponde a la compensación que debe hacerse por elegir unidades de un mismo grupo conglomerado, en lugar de unidades simples, sin pertenecer a un mismo grupo de unidades, para contemplar lo que se conoce como correlación intraclase, que mide la similitud que existe entre las unidades que pertenecen a una misma unidad conglomerada. En muestras conglomeradas el efecto diseño es por lo general superior a 1, para compensar el tamaño de muestra que se habría tenido si la muestra hubiera sido elegida como muestra aleatoria simple.

Por otro lado, como se mencionó antes, interesa tener estimaciones diferenciadas para el área metropolitana de Puebla, de las correspondientes a la entidad en conjunto, lo que obliga a realizar cálculos de tamaño de muestra por separado, para lo que se conoce como dominios de estudio.

Dominios de estudio

Los subgrupos poblacionales para los cuales se procuraron estimaciones con confianza y precisión estadística propias son:

- a. Nivel Estatal
- b. Zona metropolitana de Puebla
- c. Hombres
- d. Mujeres

A la luz de lo anterior, se hicieron cálculos por separado ante la necesidad de disponer de tamaños de muestra con confianza y precisión propias para la zona metropolitana de Puebla y el estado en su conjunto. Adicionalmente se tomó en cuenta que el estudio requeriría de indicadores que no fuesen proporciones sino medias, lo que dio lugar a una mayor varianza anticipada.

Se calculó que para la zona metropolitana de Puebla predominantemente interesaba estimar proporciones de 0.15 y superiores con una confianza de 95 por ciento y un error absoluto esperado máximo de ± 3.08 puntos porcentuales. También se consideró un efecto de diseño de 1.6¹ y una tasa de no respuesta máxima de 10 por ciento. El resultado de estas consideraciones es que se seleccionarían 1,030 viviendas aproximadamente.

Para el nivel estatal, se esperaba disponer de una muestra de aproximadamente tres mil entrevistas efectivas, que diera lugar a estimaciones de mayor nivel de precisión, con la misma confianza de 95%. El tamaño de muestra para el nivel estatal fue con la premisa de estimar proporciones de 0.13 o

¹ Estimación propia a partir los datos de la primera encuesta de Movilidad Social, 2006.

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

superiores, con una confianza de 95%, un error de estimación de ± 1.7 puntos porcentuales, un efecto de diseño de 1.8 y una tasa de no respuesta máxima esperada de 10%, y por ello el tamaño de muestra sería de 3,000 entrevistas.

Además, con los tamaños de muestra así calculados se espera poder realizar análisis estadísticos de diferencias entre el ámbito urbano y rural.

La fórmula para calcular el tamaño de la muestra fue la siguiente:

$$n = \frac{Z_{\alpha/2}^2 P(1-P) deff}{\delta^2 TR}$$

donde:

n = tamaño de la muestral

P = proporción a estimar

$Z_{\alpha/2}$ = quantil de la distribución normal asociado al 95% de confianza

δ = margen del error máximo absoluto, $\Pr(|P - \hat{P}| \leq \delta) = 0.95$

TR = tasa de respuesta esperada

$deff$ = efecto de diseño

El proceso de selección de las unidades

Para cumplir con el requerimiento de poder derivar inferencias estadísticas válidas, es requisito indispensable que la muestra sea de corte probabilístico, es decir, que todas y cada una de las unidades en muestra tengan de antemano una probabilidad no nula y conocida de ser seleccionadas.

Con base en lo anterior, inicialmente se utilizó como marco muestral la información que reporta el Instituto Nacional de Estadística y Geografía (INEGI) de los 217 municipios que integran el estado de Puebla. La base utilizada contiene la información agregada de las 190 variables sociodemográficas que lo integran, del censo de población y vivienda 2010. Esto en el afán de considerar a los municipios como la unidad primaria de muestro.

Por otro lado, se debía buscar que cualquier vivienda de la zona geográfica de interés pudiera ser alcanzada, probabilísticamente hablando. Por ello se consideró una regionalización que permitiera dar cuenta de los contrastes existentes en el Estado, a la vez que potenciara la comparabilidad de resultados con otros estudios. Por tanto, se tomaron en cuenta las siete regiones socioeconómicas en la que se encuentra dividida la entidad. Como contexto debe decirse que la construcción de las regiones las hizo el gobierno del estado y obedece a la distribución socioeconómica utilizada para planeación el desarrollo de la entidad a partir de identificar unidades homogéneas. Esta

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

regionalización está vigente y rige oficialmente desde 1986 y responde a las interacciones económicas, sociales y políticas que se dan entre los Municipios.

Cada región será considerada en lo que técnicamente se conoce como estratos de muestreo, de cada uno de los cuales se hará la selección de muestra independiente y, con ello, se asegura que la muestra se conforma de unidades primarias de todas y cada una de las regiones reconocidas para la entidad.

Adicionalmente, a la lista de municipios del INEGI que conforman las regiones socioeconómicas se les integró el grado de rezago social por municipio generado por el CONEVAL, con la finalidad de asegurar que quedaban contenidos puntos geográficos de la diversidad de condiciones de vida que contiene el Estado de Puebla pues dentro de las mismas regiones hay características también diversas.

Ahora bien, en el siguiente cuadro se ofrecen cifras del último censo de población y viviendas (2010) publicado por el INEGI, respecto de la clasificación de la población de 30 a 59 años², según se tratara de población económicamente activa (PEA) y económicamente inactiva (PEI). De los casi dos millones de población de entre 30 y 59 años de la Entidad, según su condición económica la mayoría son personas que el censo refirió como población ocupada, una de cada tres como amas de casa y las demás categorías corresponden a diversas categorías de ocupación.

Cuadro 1. Población de 30 a 59 años residentes habituales en Puebla, según principales condiciones de ocupación

Concepto	Cantidad	%
Económicamente activa	1,190,348	65.8
Población ocupada	1,155,084	63.9
Población desocupada	35,264	2.0
Económicamente NO activa	611,219	33.8
Jubilado(a) o pensionado(a)	13,452	0.7
Estudiantes	3,776	0.2
Quehaceres del hogar	561,330	31.0
Limitación física o mental permanente para trabajar	11,764	0.7
Otro tipo de actividad no económica	20,897	1.2
Sin especificar condición de actividad	6,613	0.4
Total	1,808,180	100.0

Fuente: Elaboración propia a partir de los datos publicados en la página web del INEGI.

² Esta cifra contempla la eliminación de cinco municipios que fueron considerados inseguros para labores de los encuestadores, como más adelante se refiere.

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Para establecer los mecanismos de selección específica que debían aplicarse en las etapas subsecuentes, primeramente se optó por realizar un ejercicio de afijación directa que permitiera vislumbrar la cobertura alcanzada, y la revisión respecto de los objetivos de cumplimiento de subgrupos poblacionales requeridos. Este ejercicio se muestra a continuación:

Cuadro 2: Población de 30 A 59 años por Entidad y Condición de actividad económica, según Sexo

Clave	Nombre de la Entidad	Concepto	Hombres	%	Mujeres	%	Subtotal	Dn total	Dn Hombres	DN Mujeres	Dn de la muestra
21	Puebla	Total	834,108	100.0	974,072	100.0	1,808,180				3000
21	Puebla	Económicamente activa									
		Población ocupada	765,150	91.7%	389,934	40.0%	1,155,084	63.9%	91.7%	40.0%	19.2
		Población desocupada	29,066	3.5%	6,198	0.6%	35,264	2.0%	3.5%	0.6%	0.6
21	Puebla	Económicamente NO activa		0.0%	0	0.0%	0	0.0%	0.0%	0.0%	
		Jubilado(a) o pensionado(a)	6,934	0.8%	6,518	0.7%	13,452	0.7%	0.8%	0.7%	0.2
		Estudiantes	1,089	0.1%	2,687	0.3%	3,776	0.2%	0.1%	0.3%	0.1
		Quehaceres del hogar	4,595	0.6%	556,735	57.2%	561,330	31.0%	0.6%	57.2%	9.3
		Limitación física o mental permanente para trabajar	7,504	0.9%	4,260	0.4%	11,764	0.7%	0.9%	0.4%	0.2
		Otro tipo de actividad no económica	16,369	2.0%	4,528	0.5%	20,897	1.2%	2.0%	0.5%	0.3
21	Puebla	Sin especificar condición de actividad	3,401	0.4%	3,212	0.3%	6,613	0.4%	0.4%	0.3%	0.1

Fuente: Elaboración propia a partir de los datos publicados en la página web del INEGI.

Al valorar que la selección directa no cubriría la población objetivo, se implementó un mecanismo de selección automatizado que permitiera al entrevistador identificar el perfil de entrevista requerido y verificar si debía o no debía entrevistarse, ya que si se hacía de forma proporcional se tendrían mucho menos mujeres con actividad laboral que las requeridas. Por ello, se realizó un cálculo específico y la cantidad esperada de entrevistas fue:

Cuadro 3: Tamaño esperado de muestra, según condición de actividad económica y Sexo, bajo una premisa de asignación proporcional

Clave	Nombre de la Entidad	Concepto	Cantidad esperada de hombres	Cantidad esperada de mujeres
21	Puebla	Económicamente activa		
		Población ocupada	1269.5	646.9
		Población desocupada	48.2	10.3
21	Puebla	Económicamente NO activa	0.0	0.0
		Jubilado(a) o pensionado(a)	11.5	10.8
		Estudiantes	1.8	4.5
		Quehaceres del hogar	7.6	923.7
		Limitación física o mental permanente para trabajar	12.5	7.1
		Otro tipo de actividad no económica	27.2	7.5
21	Puebla	Sin especificar condición de actividad	5.6	5.3
21	Puebla	Total	1383.9	1616.1

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Para los casos en donde hubiese más de una persona con el perfil requerido en la vivienda, el criterio de selección fue que se escogiera a la persona de la fecha de cumpleaños más próxima.

El paso natural siguiente era la afijación de la muestra, pero debido a las condiciones de inseguridad que experimentan varios municipios del estado (en particular la muerte de encuestadores en Ajalpan), para no poner en riesgo la integridad de los entrevistadores, se optó por trabajar con una muestra en la que se excluyeron los municipios de la entidad que presentaban mayores problemas. La identificación se hizo a partir de publicación “Incidencia Delictiva del Fuero Común” municipal 2015-2015” que difunde el Secretariado Ejecutivo del Sistema Estatal de Seguridad Pública, donde reportan la cantidad de delitos en cada uno de los municipios del Estado. A partir de ello se identificaron y excluyeron los municipios considerados como de mayor peligrosidad. De esta forma se seleccionó una muestra sin considerar los municipios que tenían una experiencia delictiva abiertamente reconocida.

Con la afijación establecida, se procedió a la aplicación de las distintas etapas de selección sucesiva, hasta llegar a disponer de la muestra que en campo se complementaría con los procesos de selección que permitieran llegar finalmente a las personas a entrevistar, probabilísticamente hablando.

Dada la naturaleza del levantamiento: por un lado interesaba tener resultados estadísticamente válidos para la zona metropolitana de Puebla, y también el resto del estado. Para la Zona Metropolitana de Puebla (ZMP) la unidad primaria de muestreo (UMP) fueron las Áreas Geoestadísticas Básicas (AGEB) con el que se integran en cada uno de los municipios. Para el resto del estado, la UPM fueron los municipios de cada una de las regiones económicas. De esta forma, la muestra se dividió en ZMP y resto del estado.

Cuadro 4: Municipios y población de la ZMP y resto del estado

Segmentos	Municipios	%	Total de la población	%	Población 30 a 59 años	%
ZMP	19	9	2,390,477	42.5	839,302	46.4
Resto del estado	193	91	3,238,942	57.5	968,878	53.6
Total	212	100	5,629,419	100.0	1,808,180	100.0

Fuente: Elaboración propia a partir de los datos publicados en la página web del INEGI.

Primera etapa

Como se mencionó, para los municipios fuera de la zona metropolitana de Puebla, las unidades primarias de muestreo (UPM) fueron los municipios de la región a la que están circunscritos. Con la clasificación de los municipios según la categorización de rezago social, se integraron dentro de las regiones económicas los municipios con distintos niveles de rezago, pues se buscó que la estratificación incluyera también las condiciones sociales de la población.

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Cuadro 5: Afijación de la muestra en los municipios fuera de la Zona Metropolitana de Puebla

Estrato	Municipios	%	Total de la población	%	Población 30 a 59 años	%	Afijación de Municipios	Total de Entrevistas
Región I	32	16.6	648,708	20.0	194,889	20.1	10	300
Región II	30	15.5	546,434	16.9	165,382	17.1	10	300
Región III	20	10.4	307,776	9.5	88,030	9.1	10	300
Región IV	12	6.2	286,614	8.8	89,178	9.2	8	240
Región V	4	2.1	110,755	3.4	32,355	3.3	2	60
Región VI	45	23.3	345,881	10.7	103,401	10.7	10	300
Región VII	50	25.9	992,774	30.7	295,643	30.5	16	480
Total	193	100.0	3,238,942	100.0	968,878	100.0	66	1980

Fuente: Elaboración propia a partir de los datos publicados en la página web del INEGI.

En el caso de la ZMP, se buscó contemplar a todos los municipios que la integran, por lo cual cada uno fue considerado un estrato y, por ende, las UPM fueron las Áreas Geoestadísticas Básicas (AGEB) que construyen cada uno de los 19 municipios.

Cuadro 6: Afijación de la muestra en los municipios de la Zona Metropolitana de Puebla

Clave	Municipio	Total de la población	%	Población de 30 a 59 años	%	Total de AGEBS
21001	Acajete	60,353	2.5	16,849	2.0	43
21015	Amozoc	100,964	4.2	32,060	3.8	37
21034	Coronango	34,596	1.4	10,877	1.3	17
21041	Cuautlancingo	79,153	3.3	28,279	3.4	25
21048	Chiautzingo	18,762	0.8	5,840	0.7	14
21060	Domingo Arenas	6,946	0.3	2,014	0.2	3
21074	Huejotzingo	63,457	2.7	20,290	2.4	22
21090	Juan C. Bonilla	18,540	0.8	5,768	0.7	9
21106	Ocoyucan	25,720	1.1	6,946	0.8	10
21119	San Andrés Cholula	100,439	4.2	36,404	4.3	45
21122	San Felipe Teotlancingo	9,426	0.4	2,948	0.4	4
21125	San Gregorio Atzompa	8,170	0.3	2,803	0.3	3
21132	San Martín Texmelucan	141,112	5.9	48,061	5.7	52
21136	San Miguel Xoxtla	11,598	0.5	3,871	0.5	2
21140	San Pedro Cholula	120,459	5.0	43,002	5.1	39
21143	San Salvador el Verde	28,419	1.2	8,897	1.1	11
21163	Tepatlatxco de Hidalgo	16,275	0.7	4,775	0.6	6
21181	Tlaltenango	6,269	0.3	1,839	0.2	3
21114	Puebla	1,539,819	64.4	557,779	66.5	490
21000	Región ZMP	2,390,477	100.0	839,302	100.0	835

Fuente: Elaboración propia a partir de los datos publicados en la página web del INEGI.

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

La selección de las UPM fue con probabilidad proporcional al tamaño (PPT) de manera sistemática, donde la medida de tamaño fue la población total registrada el censo de población y vivienda 2010; esto quiere decir que una unidad con mil habitantes del grupo de interés tuvo una doble probabilidad de ser elegida que una con 500 habitantes del mismo grupo.

Para el caso de resto del estado, como se pudo ver en cuadro 7, se seleccionaron 66 de las 193 unidades primarias. Como se dijo antes, por razones de seguridad del personal de campo, dadas la situaciones prevaecientes al momento del levantamiento, se descartaron del marco muestral los municipios de Ajalpan, Aljojuca, San José Chiapa, Libres y Chalchicomula de Sesma.

Cuadro 7 Afijación de la muestra para los municipios fuera de la Zona Metropolitana de Puebla

Estrato: Región y Rezago Social	Total de Municipios	Total de la población	%	Población 30 a 59 años	%	Afijación de Municipios	Total de Entrevistas
R-I + Alto	23	268,471	8.29	76,711	7.92	4	120
R-I + Medio	2	24,118	0.74	7,724	0.80	2	60
R-I + Bajo	7	356,119	10.99	110,454	11.40	4	120
R-II + Alto	13	180,652	5.58	51,809	5.35	4	120
R-II + Medio	14	252,403	7.79	75,318	7.77	4	120
R-II + Bajo	3	113,379	3.50	38,255	3.95	2	60
R- III + Alto y Medio	12	198,105	6.12	55,623	5.74	6	180
R- III + Bajo	8	109,671	3.39	32,407	3.34	4	120
R- IV + Alto/Medio/Bajo	12	286,614	8.85	89,178	9.20	8	240
R- V + Alto/Medio/Bajo	4	110,755	3.42	32,355	3.34	2	60
R- VI + Alto	6	24,748	0.76	7,252	0.75	2	60
R- VI + Medio	17	92,407	2.85	26,580	2.74	4	120
R- VI + Bajo	22	228,726	7.06	69,569	7.18	4	120
R- VII + Alto	10	109,365	3.38	28,744	2.97	4	120
R- VII + Medio	27	463,858	14.32	132,397	13.66	8	240
R- VII + Bajo	13	419,551	12.95	134,502	13.88	4	120
Total	193	3,238,942	100.00	968,878	100.00	66	1980

Fuente: Elaboración propia a partir de los datos publicados en la página web del INEGI.

En el caso del municipio de Puebla, a diferencia del resto de los municipios que integran la ZMP, se estratificaron sus AGEBS de acuerdo a un aproximado de nivel de marginación. De esta forma, se pudieron seleccionar unidades de distintos niveles socioeconómicos.

Para el caso de los municipios de la ZMP se seleccionaron 103 unidades primarias (AGEBs), de los cuales 61 correspondieron al municipio de Puebla. La distribución fue la siguiente:

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Cuadro 8: Afijación de la muestra para los municipios de la Zona Metropolitana de Puebla

Clave	Municipio	Total de la población	%	Población de 30 a 59 años	%	Total de AGEBS	Afijación de AGEBS	Total de Entrevistas
21001	Acajete	60,353	2.5	16,849	2.0	43	3	30
21015	Amozoc	100,964	4.2	32,060	3.8	37	3	30
21034	Coronango	34,596	1.4	10,877	1.3	17	3	30
21041	Cuautlancingo	79,153	3.3	28,279	3.4	25	3	30
21048	Chiautzingo	18,762	0.8	5,840	0.7	14	2	20
21060	Domingo Arenas	6,946	0.3	2,014	0.2	3	2	20
21074	Huejotzingo	63,457	2.7	20,290	2.4	22	2	20
21090	Juan C. Bonilla	18,540	0.8	5,768	0.7	9	2	20
21106	Ocoyucan	25,720	1.1	6,946	0.8	10	3	30
21119	San Andrés Cholula	100,439	4.2	36,404	4.3	45	2	20
21122	San Felipe Teotlancingo	9,426	0.4	2,948	0.4	4	2	20
21125	San Gregorio Atzompa	8,170	0.3	2,803	0.3	3	2	20
21132	San Martín Texmelucan	141,112	5.9	48,061	5.7	52	2	20
21136	San Miguel Xoxtla	11,598	0.5	3,871	0.5	2	2	20
21140	San Pedro Cholula	120,459	5.0	43,002	5.1	39	3	30
21143	San Salvador el Verde	28,419	1.2	8,897	1.1	11	2	20
21163	Tepatlatxco de Hidalgo	16,275	0.7	4,775	0.6	6	2	20
21181	Tlaltenango	6,269	0.3	1,839	0.2	3	2	20
21114	Puebla	1,539,819	64.4	557,779	66.5	490	61	610
21000	Región ZMP	2,390,477	100.0	839,302	100.0	835	103	1030

Fuente: Elaboración propia a partir de los datos publicados en la página web del INEGI.

Segunda etapa

Para el caso de los municipios fuera de la ZMP, la segunda etapa de selección consideró como unidades secundarias de muestreo (USM), para el ámbito urbano, las AGEB y localidades para el ámbito rural, según las afijaciones antes referidas.

Resto de las Etapas

De aquí en adelante, los criterios de selección para las etapas subsecuentes, para los municipios dentro y fuera del ZMP fueron similares.

La selección de las manzanas dentro de cada AGEB se realizó a partir de la cartografía, se hizo la selección de tres AGEBS con igual probabilidad de selección. Por su parte, en cada localidad no urbana, cada una se dividió en cuadrantes y se seleccionaron 2 de ellos para recorrerlos en forma de espiral y levantar el total de entrevistas.

Para la selección de viviendas, en cada manzana de la muestra fueron seleccionadas cuatro viviendas, una vivienda por cada acera, suponiendo como arreglo general de las manzanas, polígonos de cuatro lados. En total fueron seleccionadas 3,010 viviendas. En los casos de formas irregulares, se establecieron criterios específicos con los supervisores respectivos.



Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Cabe mencionar que en esta etapa se distinguió el total de hogares por vivienda. Sin embargo, dado el bajo índice de hogares por vivienda, 1004 hogares por cada 1000 viviendas según el censo del 2010, se decidió que podía hacerse la entrevista directamente sin considerar una etapa de selección *ad hoc* para hogares dentro de las viviendas.

La selección de las viviendas siguió este procedimiento: se ubicó la primera vivienda de la acera y se corroboró que estuviera habitada, de no estarlo el entrevistador se cambió a la siguiente vivienda de la misma acera. Luego, se indagó si residía población dentro del rango de edad definido; nuevamente, de no haber población objetivo, el entrevistador cambió a la siguiente vivienda de la acera. Una vez que se confirmó que reside población objetivo, se aplicó la “tabla del hogar”, mecanismo implementado para asegurar tener información de los subgrupos poblacionales requeridos; en el caso de no encontrar el perfil buscado, el entrevistador cambió de vivienda, hasta encontrar el perfil esperado. Cada uno de los acercamientos a las viviendas quedó registrado en la carátula del cuestionario correspondiente a la entrevista, para contabilizar las incidencias y poder analizar la no respuesta observada.

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Probabilidad de selección y factores de expansión

Para obtener los factores de expansión apropiados, se describen las probabilidades de selección para cada ámbito y etapa de selección y se propone una fórmula general de ponderación.

Ámbito Urbano

La probabilidad de selección de las UPM es:

$$\pi_{ij} = \frac{N_{ij}}{N_j} m_j$$

donde:

π_{ij} es la probabilidad de seleccionar en una extracción con reemplazo el Municipio i del estrato j

N_{ij} es el total de la población de 30 a 59 años del Municipio ij

N_j es el total de la población de 30 a 59 años del estrato j

m_j es el número de Municipios seleccionados en el estrato j

La probabilidad de selección de las AGEB como segunda etapa de muestreo es:

$$\pi_{ijk} = \frac{N_{ijk}}{N_{ij}} 5$$

donde:

π_{ijk} es la probabilidad proporcional al tamaño de seleccionar el AGEB k en el Municipio ij

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

N_{ijk} es el total de la población objetivo en el AGEB ijk

La probabilidad de selección en las Manzanas en la tercera etapa de muestreo es:

$$\pi_{ijkl} = \frac{2}{M}$$

donde:

π_{ijkl} es la probabilidad de seleccionar la manzana l en el AGEB ijk

M es el total de manzanas en el AGEB ijk

La probabilidad de selección en las viviendas en la cuarta etapa es:

$$\pi_{ijklm} = \frac{4}{V}$$

donde:

π_{ijklm} es la probabilidad de seleccionar la vivienda m en la manzana $ijkl$

V es el total de viviendas en la manzana $ijkl$

Así, el factor de expansión, que resulta ser el inverso de la probabilidad de selección, se obtiene con:

$$W_{ijklm} = \frac{1}{\pi_{ij} \pi_{ijk} \pi_{ijkl} \pi_{ijklm}}$$

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

Ámbito No Urbano

Para el ámbito no urbano la probabilidad de muestreo de las UPM es:

$$\pi_{jh} = \frac{N_{hj}}{N_j} n_j$$

donde:

π_{jh} es la probabilidad de seleccionar en una extracción con reemplazo la localidad h del estrato j

N_{ij} es el total de la población de 30 a 59 años de la localidad hj

N_j es el total de la población de 30 a 59 años del estrato j

n_j es el número de localidades seleccionadas en el estrato j

La probabilidad de selección de las AGEB como segunda etapa de muestreo es:

$$\pi_{jhm} = \frac{40}{N_{hj}}$$

Así, el factor de expansión para el ámbito no urbano es:

$$w_{jhm} = \frac{1}{\pi_{jh} \pi_{jhm}}$$

Ahora bien, para seleccionar al respondiente se necesitaba conjugar dos elementos: por un lado asegurar que se elegía de manera aleatoria y no bajo criterios de los propios encuestadores, por el otro se necesitaba asegurar también que se cubrían los subgrupos poblacionales de interés; es

Segunda Encuesta de Movilidad Social en México, EMOVIP – 2015

decir, personas de 30 a 59 años de edad: ocupados hombres, ocupados mujeres, amas de casa y reto poblacional.

Para lograrlo, y disponer de tamaños de muestra que permitieran realizar análisis por subgrupos con suficiente cantidad de información, se optó por “sobre representar” aleatoriamente a unos grupos respecto de los otros. El caso concreto es el de mujeres ocupadas, pues si se permitía que sólo fuera selección aleatoria se reproduciría la proporcionalidad del grupo y, por ende, se tendrían muy poca cantidad de entrevistas de este grupo de mujeres. Por tal motivo, se operacionalizó mediante el uso de números aleatorios entre 1 y 100 de tal manera que las proporciones de levantamiento para cada grupo fueron:

Subgrupo	%
Hombres ocupados	40
Mujeres ocupadas	33.4
Mujeres amas de casa	20
Resto	6.6
Total	100

A partir de estas proporciones se obtuvo el factor de ajuste de la sobre muestra, el cual se obtiene dada la proporción poblacional de cada grupo³ dividido entre la proporción levantada.

Subgrupo	Factor de Ajuste
Hombres ocupados	0.40/0.42
Mujeres ocupadas	0.33/0.21
Mujeres amas de casa	0.20/0.31
Resto	0.07/0.05

Así, el factor de expansión, que resulta ser el inverso de la probabilidad de selección, se obtiene mediante:

$$W_{ijkltg} = \frac{1}{\pi_{ijk} \pi_{ijkl} \pi_{ijkltg} \pi_{ijkltgf}} P_{ijklmf}$$

³ A partir del Censo de Población y Viviendas de 2010
Diseño de muestreo, Emovip - 2015