

Encuesta Nacional de Juventud

1. Introducción

1.1 Cobertura Conceptual

Los grandes temas a cubrir por la Encuesta Nacional de Juventud se refieren a nueve de las características principales que se destacan en la población joven de México, que son:

- Educación.
- Trabajo.
- Salud.
- Sexualidad.
- Procreación.
- Vida Privada.
- Vida Pública.
- Valores.
- Acceso a la Justicia y Derechos Humanos.

1.2 Cobertura de personas

La Encuesta generará información sobre la población joven del país, al estar conformada por residentes de las viviendas del país, que en el momento de la encuesta tengan más de 12 años y menos de 30 años.

1.3 Cobertura Geográfica

La encuesta generará información con precisión y confianza medibles con diferentes niveles de cobertura geográfica:

- Nacional.
- Nacional regional.
- Zonas Metropolitanas.
- Estatal y municipal.

Por nacional se entiende al conjunto de todos los individuos entre 12 y 29 años de edad residentes en alguna vivienda del país.

Por zonas metropolitanas, se entiende al conjunto de todos los individuos entre 12 y 29 años de edad, residentes en alguna vivienda localizada en las Ciudades de México, Guadalajara y Monterrey, contemplando en los tres casos sus zonas metropolitanas.

Por estatal y municipal, se entiende al conjunto de todos los individuos entre 12 y 29 años de edad, residentes en alguna vivienda localizada dentro de los estados de: Coahuila, Distrito Federal, Jalisco, Michoacán, Nuevo León, Quintana Roo, Yucatán, Zacatecas o en el municipio de Guadalajara.

2. Diseño Muestral

2.1 Esquema de Muestreo

Se obtuvo un esquema de muestreo probabilístico, estratificado y polietápico.

- Probabilístico: cada elemento de la población bajo estudio tiene una probabilidad conocida y diferente de cero, de ser seleccionado en la muestra.
- Estratificado: las unidades de observación se agrupan con base en características similares.

- Polietápico: las unidades de observación se seleccionan a través de varias etapas.

A continuación se describen el procedimiento de selección de la muestra; el tamaño de la misma y el procedimiento seguido para obtener las estimaciones correspondientes.

2.2 Procedimiento de selección

2.2.1 Estratos Poblacionales

Todos los municipios de cada entidad federativa se estratificaron por tamaño de población, según su número de habitantes. Se establecieron 6 estratos:

Tamaño del Municipio	
Estrato	Habitantes
VI	500,000 y más
V	100,000 a 499,999
IV	20,000 a 99,999
III	5,000 a 19,999
II	2,500 a 4,999
I	menos de 2,500

2.2.2 Regiones de estudio

Se dividió al país, de manera adicional, en 5 regiones geográficas, con características comunes, de acuerdo a las regiones establecidas por el Gobierno Federal, para el desarrollo sustentable, mismas que son:

Región Centro. Comprendida por los estados de Hidalgo, México, Morelos, Puebla, Tlaxcala y el Distrito Federal.

Región Noreste. Comprendida por los estados de Coahuila, Chihuahua, Durango, Nuevo León y Tamaulipas.

Región Noroeste. Comprendida por los estados de Baja California, Baja California Sur, Sinaloa y Sonora.

Región Centro Occidente. Comprendida por los estados de Aguascalientes, Colima, Guanajuato, Jalisco, Michoacán, Nayarit, Querétaro, San Luis Potosí y Zacatecas.

Región Sur–Sureste. Comprendida por los estados de Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz y Yucatán.

2.2.3 Etapas de Selección.

Procurando mantener la representatividad para los diferentes ámbitos de la cobertura geográfica planteada, se manejaron cuatro procesos de selección distintos, el primero para los esquemas nacional, nacional regional, mientras que, se siguió un segundo proceso de selección para el esquema de zonas metropolitanas. Un tercer proceso de selección se siguió para las entidades federativas preseleccionadas. Por último, el cuarto proceso de selección se realizó para el municipio de Guadalajara. A continuación se detallan las etapas de selección de ambos casos.

2.2.3.1 Etapas de Selección. NACIONAL.

La selección de los entrevistados se hace a través de cinco etapas sucesivas.

- Primera etapa: Adentro de las regiones, se hizo una selección por probabilidad proporcional al tamaño (ppt) de población joven adentro de las AGEB's (grupo de manzanas con características similares, determinado por INEGI). La cantidad de Ageb's a seleccionar por región fue de 58. Adentro de cada región, se realizó una subregionalización por entidad federativa y estratos poblacionales. Entonces, si suponemos que n_i representa el número de entidades federativas en la región i , se generan $6n_i$ celdas. Teniendo como primer criterio mantener la uniformidad de Ageb's a seleccionar adentro de cada estrato poblacional, manteniendo una ligera sobrerrepresentatividad en los últimos dos estratos (excepto en la región Sur – Sureste), como segundo criterio mantener la uniformidad de Ageb's a seleccionar adentro de cada entidad federativa, y como tercer criterio mantener la uniformidad de Ageb's a seleccionar en cada entidad federativa en cada estrato poblacional, se realizó la siguiente selección de Ageb's por región:

	REGIÓN CENTRO					
ESTRATOS	DF	Hidalgo	México	Morelos	Puebla	Tlaxcala
1 a 2,499 hab	0	0	0	0	8	1
2,500 a 4,999 hab	0	2	2	0	3	2
5,000 a 19,999 hab	0	2	1	2	2	2
20,000 a 99,999 hab	1	1	2	2	1	2
100,000 a 499,999 hab	3	2	2	3	1	0
500,000 habitantes o más	6	0	4	0	1	0

	REGIÓN NORESTE				
ESTRATOS	Coahuila	Chihuahua	Durango	Nuevo León	Tamaulipas
1 a 2,499 hab	2	2	1	2	2
2,500 a 4,999 hab	1	2	2	2	2
5,000 a 19,999 hab	1	2	2	2	2
20,000 a 99,999 hab	2	2	2	2	1
100,000 a 499,999 hab	2	1	3	2	2
500,000 habitantes o más	4	3	1	4	0

	REGIÓN NOROESTE			
ESTRATOS	Baja California	Baja California Sur	Sinaloa	Sonora
1 a 2,499 hab	0	0	0	9
2,500 a 4,999 hab	0	0	0	9
5,000 a 19,999 hab	0	4	4	1
20,000 a 99,999 hab	3	3	2	1
100,000 a 499,999 hab	2	4	2	1
500,000 habitantes o más	6	0	4	3

	REGIÓN CENTRO OCCIDENTE								
ESTRATOS	Aguas-calientes	Colima	Guana-juato	Jalisco	Micho-acán	Nayarit	Queré-taro	San Luis Potosí	Zaca-tecas
1 a 2,499 hab	0	0	0	3	0	0	0	0	4
2,500 a 4,999 hab	0	0	1	3	1	0	0	2	2
5,000 a 19,999 hab	1	1	1	1	1	1	1	1	1
20,000 a 99,999 hab	2	2	1	1	1	2	1	1	1
100,000 a 499,999 hab	0	2	1	1	1	2	2	1	1
500,000 habitantes o más	2	0	2	2	2	0	1	1	0

ESTRATOS	REGIÓN SUR - SURESTE							
	Campeche	Chiapas	Guerrero	Oaxaca	Quintana Roo	Tabasco	Veracruz	Yucatán
1 a 2,499 hab	0	2	0	4	0	0	2	4
2,500 a 4,999 hab	0	2	0	3	0	0	2	2
5,000 a 19,999 hab	3	1	1	1	3	0	1	1
20,000 a 99,999 hab	1	1	2	1	1	2	1	1
100,000 a 499,999 hab	2	1	1	1	2	1	1	0
500,000 habitantes o más	0	0	2	0	0	3	0	2

En todos los casos, cuando no se seleccionó Ageb's, fue porque no había ninguna que seleccionar

Una vez que se determinó totalmente el número de Ageb's a seleccionar por estrato por entidad federativa, la selección se hace con probabilidad proporcional a su tamaño (PPT), con reemplazo, utilizando como medida el número de habitantes entre 12 y 29 años de edad, según el CGPV/2000.

- Segunda etapa: Dentro de cada AGEB seleccionada se obtuvo una muestra de manzanas, de manera sistemática, y con arranque aleatorio. Se proponen 4 manzanas.
- Tercera etapa: En cada manzana se seleccionaron tres viviendas, su selección se realizará de manera sistemática y con arranque aleatorio.
- Cuarta etapa: En cada vivienda incluida en la muestra, se seleccionó una persona residente de la vivienda que tenga entre 12 y 29 años edad. Se dio igual probabilidad a cada uno de los residentes de la vivienda que cumplan con el requisito de edad mencionado, seleccionando al individuo que tenga la fecha de cumpleaños más próxima. Dicho individuo contestará el cuestionario individual. Adicionalmente, se seleccionó de forma sistemática otro individuo mayor de 15 años de edad (que no necesariamente era diferente al individuo que contestó el cuestionario individual), quien contestó un cuestionario con las características de la vivienda (trh).

Este procedimiento arrojó un total de 3,480 entrevistas en total (696 en cada una de las regiones).

2.2.3.2 Etapas de Selección. ZONAS METROPOLITANAS.

La selección de los entrevistados se hace a través de cuatro etapas sucesivas.

- Primera etapa: Dentro de cada una de las tres ciudades, se seleccionaron 50 Ageb's, la selección se hace con probabilidad proporcional a su tamaño (PPT), sin reemplazo, utilizando como medida el número de habitantes entre 12 y 29 años de edad, según el CGPV/2000.
- Segunda etapa: En cada AGEB seleccionada, se seleccionarán 4 manzanas de manera sistemática con arranque aleatorio.
- Tercera etapa: Dentro de cada manzana seleccionada se seleccionarán 3 viviendas, de manera sistemática, con arranque aleatorio.
- Cuarta etapa: En cada vivienda incluida en la muestra, se seleccionó una persona residente de la vivienda que tenga entre 12 y 29 años edad. Se dio igual probabilidad a cada uno de los residentes de la vivienda que cumplan con el requisito de edad mencionado, seleccionando al individuo que tenga la fecha de cumpleaños más próxima. Dicho individuo contestará el cuestionario individual. Adicionalmente, se seleccionó de forma sistemática otro individuo mayor de 15 años de edad (que no necesariamente era diferente al individuo que contestó el cuestionario individual), quien contestó un cuestionario con las características de la vivienda (trh).

Este procedimiento arrojó un total de 1,800 entrevistas (600 en cada una de las tres ciudades).

2.2.3.3 Etapas de Selección. ESTATAL.

La selección de los entrevistados se hace a través de cinco etapas sucesivas.

- Primera etapa: Adentro de las siete entidades federativas, se hizo una selección por probabilidad proporcional al tamaño (ppt) de población joven adentro de las AGEB's (grupo de manzanas con características similares, determinado por INEGI). La cantidad de Ageb's a seleccionar por entidad fue de 70, aproximadamente. Adentro de cada entidad, se realizó una subregionalización por nivel socioeconómico (de los municipios¹) y estratos poblacionales. Entonces, si consideramos que hay un máximo de siete niveles socioeconómicos y un máximo de seis estratos presentes en cada entidad, se generan máximo 42 celdas. Teniendo como primer criterio mantener la uniformidad de Ageb's a seleccionar adentro de cada nivel socioeconómico, manteniendo una ligera sobrerrepresentatividad en los últimos dos estratos, como segundo criterio mantener la uniformidad de Ageb's a seleccionar adentro de cada estrato poblacional, y como tercer criterio mantener la uniformidad de Ageb's a seleccionar en cada nivel socioeconómico en cada estrato poblacional, se realizó la siguiente selección de Ageb's por entidad:

		COAHUILA					
NSE	Estratos	I	II	III	IV	V	VI
	4		2	11	5	0	0
6		2	0	6	6	5	0
7		0	0	0	5	10	15

		DISTRITO FEDERAL					
NSE	Estratos	I	II	III	IV	V	VI
	6		0	0	0	5	0
7		0	0	0	0	25	40

¹ Determinada por INEGI.

		JALISCO					
NSE	Estratos	I	II	III	IV	V	VI
	1		0	0	1	0	0
2		0	1	1	0	0	0
3		0	2	2	0	0	0
4		3	6	5	3	0	0
5		0	0	2	1	0	0
6		0	2	4	7	8	0
7		0	0	0	2	3	10

		MICHOACÁN					
NSE	Estratos	I	II	III	IV	V	VI
	1		0	0	1	1	0
2		0	0	2	2	0	0
3		0	2	3	3	0	0
4		0	3	5	5	3	0
5		0	0	2	4	3	0
6		0	0	2	4	10	0
7		0	0	0	0	0	15

		NUEVO LEÓN					
NSE	Estratos	I	II	III	IV	V	VI
	2		0	0	2	0	0
3		0	2	2	3	0	0
4		2	0	2	3	0	0
6		3	6	3	3	0	0
7		0	0	2	2	15	15

		QUINTANA ROO					
NSE	Estratos	I	II	III	IV	V	VI
	2		0	0	0	4	0
3		0	0	0	7	0	0
6		0	0	4	8	15	0
7		0	0	0	7	25	0

		YUCATÁN					
NSE	Estratos	I	II	III	IV	V	VI
	1		0	1	0	0	0
2		2	2	2	2	0	0
3		2	5	5	5	0	0
4		0	0	0	0	0	0
5		2	2	3	10	0	0
6		0	0	0	3	0	0
7		0	0	0	0	0	20

		ZACATECAS					
NSE	Estratos	I	II	III	IV	V	VI
	3		0	4	8	4	0
4		5	5	8	8	0	0
6		0	0	4	8	8	0
7		0	0	0	0	8	0

En todos los casos, cuando no se seleccionó Ageb's, fue porque no había ninguna que seleccionar

Una vez que se determinó totalmente el número de Ageb's a seleccionar por estrato por entidad federativa, la selección se hace con probabilidad proporcional a su tamaño (PPT), con reemplazo, utilizando como medida el número de habitantes entre 12 y 29 años de edad, según el CGPV/2000.

- Segunda etapa: Dentro de cada AGEB seleccionada se obtuvo una muestra de manzanas, de manera sistemática, y con arranque aleatorio. Se proponen 4 manzanas.
- Tercera etapa: En cada manzana se seleccionaron tres viviendas, su selección se realizará de manera sistemática y con arranque aleatorio.
- Cuarta etapa: En cada vivienda incluida en la muestra, se seleccionó una persona residente de la vivienda que tenga entre 12 y 29 años edad. Se dio igual probabilidad a cada uno de los residentes de la vivienda que cumplan con el requisito de edad mencionado, seleccionando al

individuo que tenga la fecha de cumpleaños más próxima. Dicho individuo contestará el cuestionario individual. Adicionalmente, se seleccionó de forma sistemática otro individuo mayor de 15 años de edad (que no necesariamente era diferente al individuo que contestó el cuestionario individual), quien contestó un cuestionario con las características de la vivienda (trh).

Este procedimiento arrojó un total de 6,720 entrevistas en total (840 en cada una de las entidades).

2.2.3.4 Etapas de Selección. MUNICIPIO DE GUADALAJARA.

La selección de los entrevistados se hace a través de cuatro etapas sucesivas.

- Primera etapa: Dentro del municipio, se seleccionarán 70 Ageb's, la selección se hace con probabilidad proporcional a su tamaño (PPT), sin reemplazo, utilizando como medida el número de habitantes entre 12 y 29 años de edad, según el CGPV/2000.
- Segunda etapa: En cada AGEB seleccionada, se seleccionarán 4 manzanas de manera sistemática con arranque aleatorio.
- Tercera etapa: Dentro de cada manzana seleccionada se seleccionarán 3 viviendas, de manera sistemática, con arranque aleatorio.
- Cuarta etapa: En cada vivienda incluida en la muestra, se seleccionó una persona residente de la vivienda que tenga entre 12 y 29 años edad. Se dio igual probabilidad a cada uno de los residentes de la vivienda que cumplan con el requisito de edad mencionado, seleccionando al individuo que tenga la fecha de cumpleaños más próxima. Dicho individuo contestará el cuestionario individual. Adicionalmente, se seleccionó de forma sistemática otro individuo mayor de 15 años de edad (que no necesariamente era diferente al individuo que contestó el cuestionario individual), quien contestó un cuestionario con las características de la vivienda (trh).

Este procedimiento arrojó un total de 840 entrevistas.

2.2.4 Tamaño General de Muestra.

Lo expuesto en el apartado 2.2.3, se resume en el siguiente cuadro, que presenta el tamaño total de la muestra nacional.

Tipo de Esquema	Total de Entrevistas
NACIONAL REGIONAL	3,480
ZONAS METROPOLITANAS	1,800
ESTATAL	6,720
MUNICIPAL	840
TOTAL NACIONAL	12,840

2.3 Tamaño de Muestra

La determinación del tamaño de muestra considera las características de cada región. Una vez que se determina el tamaño óptimo por región, el agregado nacional es el tamaño de muestra

La expresión utilizada para determinar el tamaño de muestra es la siguiente:

$$n = (k^2 / d^2) (P*Q) (Deff / (1-Tnr))$$

donde:

n: es el tamaño de muestra de personas de 18 años o más

k: es el valor de tablas que determina la confianza en las estimaciones

d: es el error absoluto, aceptable en las estimaciones

P: es la proporción que se desea estimar

Q=1-P

Deff: es el efecto de diseño

Tnr: es la tasa de no respuesta

El valor asignado a los diferentes componentes de esta expresión es:

$k = 1.96$; lo cual garantiza un 95% de confianza en las estimaciones. Igualmente se utilizan los valores 1.64, lo cual garantiza un 90% de confianza, o bien 2.56, lo cual garantiza un 99% de confianza.

$Deft = 1.75$ promedio de encuestas sociales de inseguridad y de salud.

$Tnr =$ Será el asociado a cada una de las regiones, de acuerdo al levantamiento, tal y como se planteará más adelante.

A continuación se presentan los márgenes de error máximos con un confianza del 95%, considerando un efecto de diseño de 1.75 en las regiones sin sobremuestra, las zonas metropolitanas, las entidades federativas y el municipio de Guadalajara, y las localidades rurales; y un efecto de diseño de 3.5 en las regiones con sobremuestra en entidades y municipios, las localidades semiurbanas y urbanas, y en el total nacional. y una tasa de no respuesta de 0 (por considerar el tamaño levantado de muestra y no el esperado).

REGIÓN O LOCALIDAD	TAMAÑO LEVANTADO DE MUESTRA	Márgenes de error máximos al 95% de confianza
CENTRO	2,117	3.99%
NORESTE	2,995	3.35%
NOROESTE	620	5.20%
CENTRO OCCIDENTE	4,620	2.70%
SUR - SURESTE	2,444	3.71%
Ciudad de México	1,586	3.26%
Ciudad de Guadalajara	1,665	3.18%
Ciudad de Monterrey	1,029	4.04%
COAHUILA	972	4.15%
DISTRITO FEDERAL	1,311	3.58%
JALISCO	2,356	2.67%
MICHOACÁN	922	4.27%
NUEVO LEÓN	1,625	3.22%
QUINTANA ROO	899	4.32%
YUCATÁN	1,001	4.10%
ZACATECAS	949	4.21%
GUADALAJARA	1,209	3.73%
TOTAL NACIONAL	12,796	1.68%

3. Cálculo del ponderador.

El ponderador es el que otorga a cada cuestionario el peso correcto, de acuerdo a la probabilidad de selección del entrevistado. Así, un entrevistado que tenga una menor probabilidad de ser entrevistado tendrá un cuestionario que a su vez, tendrá un mayor peso que otros, debido a la dificultad de la selección.

3.1 Cálculo de la Probabilidad de selección del informante del cuestionario individual. NACIONAL

De acuerdo a lo señalado en 2.2.3.1, la probabilidad de seleccionar al individuo i , será:

$$P[\text{individuo } i] = P[\text{individuo } i | \text{vivienda } j]^* \\ P[\text{vivienda } j | \text{manzana } k]^* \\ P[\text{manzana } k | \text{ageb } h]^* \\ P[\text{ageb } h]$$

Donde:

$$P[\text{individuo } i | \text{vivienda } j] = \frac{1}{\# \text{ personas entre 12 y 29 años en la vivienda } j},$$

$$P[\text{vivienda } j | \text{manzana } k] = \frac{3}{\# \text{ viviendas en la manzana } k},$$

$$P[\text{manzana } k | \text{ageb } h] = \frac{4}{\# \text{ manzanas en la ageb } h},$$

$$P[\text{ageb } h] = P[X > 0], \text{ donde, } X \sim \text{Bin}(n_{r,s,t}, p_{h,s,t})$$

Donde $n_{r,s,t}$ es el número de ageb's a seleccionar originalmente, de la región r , la entidad federativa s y el estrato poblacional t . A su vez, $p_{h,s,t}$ es la probabilidad de seleccionar una vez a la ageb, misma que se calcula dividiendo el número de personas entre 12 y 29 años de edad de la ageb h entre el total de individuos entre 12 y 29 años de edad de la entidad federativa s y el estrato poblacional t .

Por último, debido a la ausencia de información referente al número de viviendas en cada manzana y el número de manzanas en cada ageb, se toma como estimador de ese número al total de viviendas particulares habitadas, proporcionado por INEGI.

3.2 Cálculo de la Probabilidad de selección del informante del cuestionario individual. ZONAS METROPOLITANAS

De acuerdo a las 4 etapas sucesivas, la probabilidad del individuo i , será:

$$P[\text{individuo } i] = P[\text{individuo } i | \text{vivienda } j]^* \\ P[\text{vivienda } j | \text{manzana } k]^* \\ P[\text{manzana } k | \text{ageb } h]^* \\ P[\text{ageb } h]$$

Donde:

$$P[\text{individuo } i | \text{vivienda } j] = \frac{1}{\# \text{ personas entre 12 y 29 años en la vivienda } j},$$

$$P[\text{vivienda } j | \text{manzana } k] = \frac{3}{\# \text{ viviendas en la manzana } k},$$

$$P[\text{manzana } k | \text{ageb } h] = \frac{4}{\# \text{ manzanas en la ageb } h},$$

$$P[\text{ageb } h] = P[X > 0], \text{ donde, } X \sim \text{Bin}(50, p_{h,t})$$

Donde, $p_{h,t}$ es la probabilidad de seleccionar una vez a la ageb, misma que se calcula dividiendo el número de personas entre 12 y 29 años de edad de la ageb h entre el total de individuos entre 12 y 29 años de edad de la zona metropolitana t . Cabe señalar que éste es el cálculo de la probabilidad de selección de una ageb con reemplazo, sin embargo, debido a la gran cantidad de Ageb's en cada una de las tres ciudades, el cálculo de las probabilidades con reemplazo resulta un buen estimador del cálculo de las probabilidades sin reemplazo, mismo que, incluso, resulta conservador.

Por último, debido a la ausencia de información referente al número de viviendas en cada manzana y el número de manzanas en cada ageb, se toma como estimador de ese número al total de viviendas particulares habitadas.

3.3 Cálculo de la Probabilidad de selección del informante del cuestionario individual. ESTATAL

De acuerdo a las 4 etapas sucesivas, la probabilidad del individuo i , será:

$$P[\text{individuo } i] = P[\text{individuo } i | \text{vivienda } j]^* \\ P[\text{vivienda } j | \text{manzana } k]^* \\ P[\text{manzana } k | \text{ageb } h]^* \\ P[\text{ageb } h]$$

Donde:

$$P[\text{individuo } i | \text{vivienda } j] = \frac{1}{\# \text{ personas entre 12 y 29 años en la vivienda } j},$$

$$P[\text{vivienda } j | \text{manzana } k] = \frac{3}{\# \text{ viviendas en la manzana } k},$$

$$P[\text{manzana } k | \text{ageb } h] = \frac{4}{\# \text{ manzanas en la ageb } h},$$

$$P[\text{ageb } h] = P[X > 0], \text{ donde, } X \sim \text{Bin}(n_{r,s,t}, p_{h,s,t})$$

Donde $n_{r,s,t}$ es el número de ageb's a seleccionar originalmente, de la entidad r , el nivel socioeconómico s y el estrato poblacional t . A su vez, $p_{h,s,t}$ es la probabilidad de seleccionar una vez a la ageb, misma que se calcula dividiendo el número de personas entre 12 y 29 años de edad de la ageb h entre el total de individuos entre 12 y 29 años de edad del nivel socioeconómico s y el estrato poblacional t .

Por último, debido a la ausencia de información referente al número de viviendas en cada manzana y el número de manzanas en cada ageb, se toma como estimador de ese número al total de viviendas particulares habitadas.

3.4 Cálculo de la Probabilidad de selección del informante del cuestionario individual. MUNICIPAL

De acuerdo a las 4 etapas sucesivas, la probabilidad del individuo i , será:

$$P[\text{individuo } i] = P[\text{individuo } i | \text{vivienda } j]^* \\ P[\text{vivienda } j | \text{manzana } k]^* \\ P[\text{manzana } k | \text{ageb } h]^* \\ P[\text{ageb } h]$$

Donde:

$$P[\text{individuo } i | \text{vivienda } j] = \frac{1}{\# \text{ personas entre 12 y 29 años en la vivienda } j},$$

$$P[\text{vivienda } j | \text{manzana } k] = \frac{3}{\# \text{ viviendas en la manzana } k},$$

$$P[\text{manzana } k | \text{ageb } h] = \frac{4}{\# \text{ manzanas en la ageb } h},$$

$$P[\text{ageb } h] = P[X > 0], \text{ donde, } X \sim \text{Bin}(70, p_h)$$

Donde, p_h es la probabilidad de seleccionar una vez a la ageb, misma que se calcula dividiendo el número de personas entre 12 y 29 años de edad de la ageb h entre el total de individuos entre 12 y 29 años de edad del municipio de Guadalajara. Cabe señalar que éste es el cálculo de la probabilidad de selección de una ageb con reemplazo, sin embargo, debido a la gran cantidad de Ageb's en el municipio de Guadalajara, el cálculo de las probabilidades con reemplazo resulta un buen estimador del cálculo de las probabilidades sin reemplazo, mismo que, incluso, resulta conservador.

Por último, debido a la ausencia de información referente al número de viviendas en cada manzana y el número de manzanas en cada ageb, se toma como estimador de ese número al total de viviendas particulares habitadas.

3.5 Cálculo de la Probabilidad de selección del individuo conjunta.

Los marcos muestrales son complementarios, es decir, cada individuo, puede provenir de hasta cuatro tipos de diseño muestrales distintos e independientes, por ejemplo, los individuos residentes de las Zonas Metropolitanas pueden provenir tanto de la muestra nacional, de la estatal, de la municipal, o bien, de la muestra de las zonas metropolitanas. Lo anterior, únicamente en la selección de las Ageb's , ya que al momento del levantamiento, las probabilidades permanecen sin cambios, debido a los diseños muestrales.

No obstante lo anterior, la probabilidad de selección de un individuo de las zonas metropolitanas tiene que contemplar todos los diseños. Así, si denotamos por P_{ZM_i} la probabilidad de selección del individuo i , de acuerdo al diseño para las zonas metropolitanas, por P_{NAL_i} la probabilidad de selección del individuo i , de acuerdo al diseño para el levantamiento nacional, por P_{EDO_i} la probabilidad de selección del individuo i , de acuerdo al diseño para el levantamiento estatal, y por P_{GDL_i} la probabilidad de selección del individuo i , de acuerdo al diseño para el levantamiento municipal; y recordamos que todo individuo puede provenir de alguno de los diseños muestrales, de dos, de tres o hasta de cuatro de ellos, se obtiene la probabilidad de selección del individuo, sin importar de cual diseño proviene, de acuerdo a la siguiente fórmula:

$$\begin{aligned}
P[X_i] = & \\
& (P_{NAL_i})(1-P_{ZM_i})(1-P_{EDO_i})(1-P_{GDL_i}) \\
& + (1-P_{NAL_i})(P_{ZM_i})(1-P_{EDO_i})(1-P_{GDL_i}) \\
& + (1-P_{NAL_i})(1-P_{ZM_i})(P_{EDO_i})(1-P_{GDL_i}) \\
& + (1-P_{NAL_i})(1-P_{ZM_i})(1-P_{EDO_i})(P_{GDL_i}) \\
& + (P_{NAL_i})(P_{ZM_i})(1-P_{EDO_i})(1-P_{GDL_i}) \\
& + (1-P_{NAL_i})(P_{ZM_i})(P_{EDO_i})(1-P_{GDL_i}) \\
& + (1-P_{NAL_i})(1-P_{ZM_i})(P_{EDO_i})(P_{GDL_i}) \\
& + (P_{NAL_i})(1-P_{ZM_i})(P_{EDO_i})(1-P_{GDL_i}) \\
& + (1-P_{NAL_i})(P_{ZM_i})(1-P_{EDO_i})(P_{GDL_i}) \\
& + (P_{NAL_i})(1-P_{ZM_i})(1-P_{EDO_i})(P_{GDL_i}) \\
& + (P_{NAL_i})(P_{ZM_i})(P_{EDO_i})(1-P_{GDL_i}) \\
& + (P_{NAL_i})(P_{ZM_i})(1-P_{EDO_i})(P_{GDL_i}) \\
& + (P_{NAL_i})(1-P_{ZM_i})(P_{EDO_i})(P_{GDL_i}) \\
& + (1-P_{NAL_i})(P_{ZM_i})(P_{EDO_i})(P_{GDL_i}) \\
& + (P_{NAL_i})(P_{ZM_i})(P_{EDO_i})(P_{GDL_i}) \\
& \hline
& 1 - \left[(1-P_{NAL_i})(1-P_{ZM_i})(1-P_{EDO_i})(1-P_{GDL_i}) \right]
\end{aligned}$$

Con lo anterior, es posible el cálculo del ponderador para todos los individuos entrevistados, con lo que, se obtiene su correspondiente factor de expansión, al calcular el inverso de la probabilidad de selección.

3.6 Cálculo de la Probabilidad de selección de la vivienda

En este caso, estamos interesados en el cálculo de la probabilidad de selección de la vivienda, para generar el factor de expansión adecuado para la información de las características de la vivienda, proveniente de las preguntas 1 a 10 del cuestionario de hogar.

Dado que la selección de la vivienda resulta idéntica en cualquiera al trabajar por separado a los diversos esquemas de muestreo, dicha probabilidad se calcula de la siguiente forma:

$$P[\text{vivienda } j] = P[\text{vivienda } j | \text{manzana } k] * \\ P[\text{manzana } k | \text{ageb } h] * \\ P[\text{ageb } h]$$

Donde:

$$P[\text{vivienda } j | \text{manzana } k] = \frac{3}{\# \text{ viviendas en la manzana } k},$$

$$P[\text{manzana } k | \text{ageb } h] = \frac{4}{\# \text{ manzanas en la ageb } h},$$

$$P[\text{ageb } h] = P[X > 0], \text{ donde, } X \sim \text{Bin}(n_{Q,r,s,t}, p_{h,s,t})$$

Donde $n_{Q,r,s,t}$ es el número de ageb's a seleccionar originalmente de acuerdo al esquema de muestreo Q , de la entidad r , el nivel socioeconómico s y el estrato poblacional t . A su vez, $p_{h,s,t}$ es la probabilidad de seleccionar una vez a la ageb, misma que se calcula dividiendo el número de personas entre 12 y 29 años de edad de la ageb h entre el total de individuos entre 12 y 29 años de edad del nivel socioeconómico s y el estrato poblacional t . Debido a la ausencia de información referente al número de viviendas en cada manzana y el número de manzanas en cada ageb, se toma como estimador de ese número al total de viviendas particulares habitadas.

Con ello, se obtiene la probabilidad de selección de la vivienda para cada uno de los esquemas de muestreo. Para calcular la probabilidad de selección de la vivienda dado que pudo provenir de hasta cuatro esquemas muestrales diferentes, se utiliza la fórmula presentada en 3.5.

3.7 Cálculo de la Probabilidad de los individuos de la vivienda

En este caso, estamos interesados en el cálculo de la probabilidad de selección de los individuos de la vivienda, para generar el factor de expansión adecuado para la información de las características de los miembros del hogar, proveniente de las preguntas 12 a 49 del cuestionario de hogar.

Dado que se obtiene información para todos los individuos de la vivienda, resulta equivalente a seleccionar con certeza a todos los individuos de una vivienda, con lo que la probabilidad de seleccionar a un individuo adentro de una vivienda toma el valor de 1, por lo que se utiliza la probabilidad de selección de la vivienda, que al resultar idéntica en cualquiera al trabajar por separado a los diversos esquemas de muestreo, dicha probabilidad se calcula de la siguiente forma:

$$P[\text{vivienda } j] = P[\text{vivienda } j \mid \text{manzana } k] * \\ P[\text{manzana } k \mid \text{ageb } h] * \\ P[\text{ageb } h]$$

Donde:

$$P[\text{vivienda } j \mid \text{manzana } k] = \frac{3}{\# \text{ viviendas en la manzana } k},$$

$$P[\text{manzana } k \mid \text{ageb } h] = \frac{4}{\# \text{ manzanas en la ageb } h},$$

$$P[\text{ageb } h] = P[X > 0], \text{ donde, } X \sim \text{Bin}(n_{Q,r,s,t}, p_{h,s,t})$$

Donde $n_{Q,r,s,t}$ es el número de ageb's a seleccionar originalmente de acuerdo al esquema de muestreo Q , de la entidad r , el nivel socioeconómico s y el estrato poblacional t . A su vez, $p_{h,s,t}$ es la probabilidad de seleccionar una vez a la ageb, misma que se calcula dividiendo el número de personas entre 12 y 29 años de edad de la ageb h entre el total de individuos entre 12 y 29 años de edad del nivel socioeconómico s y el estrato poblacional t . Debido a la ausencia de información referente al número de viviendas en cada manzana y el número de manzanas en cada ageb, se toma como estimador de ese número al total de viviendas particulares habitadas.

Con ello, se obtiene la probabilidad de selección de la vivienda para cada uno de los esquemas de muestreo. Para calcular la probabilidad de selección de la vivienda dado que pudo provenir de hasta cuatro esquemas muestrales diferentes, se utiliza la fórmula presentada en 3.5.

La única diferencia con 3.6 radica en que este ponderador es imputado a todos los miembros del hogar, mientras que el ponderador calculado de acuerdo a 3.6 (y los correspondientes al cuestionario individual) únicamente se imputa al renglón donde aparecen las características de la vivienda.

3.8 POSTESTRATIFICACIÓN

Consideremos a $f_{i,a} = 1/P[i]$ el factor de expansión de un individuo i con una característica a , donde a es el vector que indica características propias del ese individuo (por ejemplo, a puede hacer referencia a un grupo de edad, a un sexo, a una región geográfica y a un estrato poblacional de manera simultánea).

Consideremos además, a r_a , la población total de los individuos con la característica a , obtenida de medios confiable, de acuerdo a nuestros criterios.

Ahora consideremos a $F_a = \sum_{i=1}^{n_a} f_{i,a}$ la población total expandida de todos los individuos que pertenecen a la característica a , al utilizar los factores de expansión obtenidos de acuerdo a las secciones anteriores. Dicha población, si difiere de r_a , es inadecuada para la expansión real de esa unidad de muestreo, debido a la incertidumbre del dato. Así, en un esfuerzo para que la expansión de la base llegue a los resultados reales (en algún punto del tiempo) que tenemos, se realiza un proceso de postestratificación, de acuerdo a la siguiente expresión:

$$g_{i,a} = f_{i,a} \left(\frac{r_a}{F_a} \right)$$

Donde $g_{i,a}$ es el nuevo factor de expansión para el individuo i que tiene las características del vector a . Con esta transformación, los resultados expandidos, claramente extrapolan a la población deseada, tal y como se demuestra a continuación:

$$G_a = \sum_{i=1}^{n_a} g_{i,a} = \sum_{i=1}^{n_a} f_{i,a} \left(\frac{r_a}{F_a} \right) = \left(\frac{r_a}{F_a} \right) \sum_{i=1}^{n_a} f_{i,a} = \left(\frac{r_a}{F_a} \right) F_a = r_a.$$