



Proyecto
para integrar la Línea Base
para Evaluar la Política Social
en el Estado de Guanajuato



Instituto Nacional
de Salud Pública

ANEXO 4.
Encuesta para la Evaluación de la
Política Social de Guanajuato
(EEPS Gto), 2014:
Metodología de Ponderación

Anexo 4

Encuesta para la Evaluación de la Política Social
de Guanajuato (EEPS Gto), 2014:
Metodología de ponderación

1. Objetivo

El procedimiento de ponderación puede resumirse, de modo general, como la asignación de un conjunto de factores de expansión (ponderadores) w_1, w_2, \dots, w_n a las unidades en muestra u_1, u_2, \dots, u_n de tal suerte que el total del atributo Y en la población de interés pueda ser estimado mediante la expresión.

$$\hat{Y} = \sum_{i=1}^n w_i y(u_i)$$

2. Población de interés

La población de interés son las personas con edad de 15 o más años que residen en viviendas particulares del estado de Guanajuato.

3. Consideraciones sobre el procedimiento de ponderación

El procedimiento de ponderación incorpora tres elementos: la no respuesta, las probabilidades de selección y el conocimiento *a priori* de parámetros poblacionales.

La no respuesta es relevante porque implica el desconocimiento de dos elementos: el atributo de interés Y en un subconjunto de la muestra y el procedimiento que causa a la no respuesta. Sí el investigador conociera el procedimiento que causa la no respuesta, entonces dicho procedimiento podría ser incorporado al procedimiento de ponderación en la misma manera que el procedimiento de muestreo.

El procedimiento de muestreo es relevante para la ponderación porque en ausencia de no respuesta, el estimador \hat{Y}_H que define a los ponderadores como el inverso de las probabilidades de selección.

$$\hat{Y}_H = \sum_{i=1}^n \frac{y(u_i)}{\text{Prob}(u_i \text{ sea seleccionado})}$$

es un estimador cuyo valor esperado es el total en la población del atributo Y

$$E\hat{Y}_H = \sum_{i=1}^N Y_i ;$$

esto es, el promedio de \hat{Y}_H sobre todas las posibles muestras que pueden ser generadas es igual al total en la población del atributo Y .

Finalmente, el conocimiento de parámetros poblacionales es relevante para la ponderación porque los ponderadores w_1, w_2, \dots, w_n pueden ser elegidos de tal suerte que a partir de la encuesta sea posible reproducir parámetros poblacionales conocidos, por ejemplo el número de hogares del estado de Guanajuato.

Entonces, de modo general, el procedimiento de ponderación propuesto consta de tres etapas:

- a) Construcción de los ponderadores iniciales mediante el inverso de la probabilidad de selección.
- b) Cuantificación de la no respuesta.
- c) Ajuste de los ponderadores para reproducir parámetros poblacionales y considerar la no respuesta.

A continuación, se describe el detalle del procedimiento de ponderación.

4. Probabilidades de selección

El procedimiento de muestreo fue poli-etápico y estratificado. El muestreo fue poli-etápico porque primero se seleccionaron AGEB (Áreas Geo-estadísticas Básicas) como unidades primarias de muestreo. Luego, en las AGEB urbanas se seleccionaron manzanas y viviendas dentro de las manzanas; por otra parte, en las AGEB rurales, se seleccionaron localidades y dentro de las localidades se seleccionaron viviendas.

El procedimiento de muestreo fue estratificado. Las variables de estratificación que se usaron fueron dos: urbanidad y área geográfica. La variable de urbanidad dividió a la población en urbana, rural y metropolitana. La población rural es aquella que vive en localidades con menos de 2500 habitantes y la población urbana será el resto. La población metropolitana corresponde a las localidades con 100mil habitantes o más, capitales de los estados y localidades con más de 2500 de la áreas metropolitanas según la clasificación del INEGI en 2000. Finalmente, la población urbana es el resto del estado. Por otra parte, las áreas geográficas fueron las usada por el gobierno de Guanajuato.

Figura 1. Regiones geográficas del estado de Guanajuato



Las AGEB, manzanas y localidades se seleccionaron con probabilidad proporcional a su tamaño (población en el censo 2010). Luego, en cada manzana se hizo un croquis de viviendas y se seleccionó, en general, a 6 viviendas mediante un muestreo sistemático; por otra parte, en las localidades rurales, se formaron conglomerados de 12 viviendas y se seleccionó a uno de ellos. En consecuencia, el inverso de la probabilidad de selección de una vivienda es:

$$w_H = \frac{Pob}{mPob_i} \frac{Pob_i}{k_iPob_{ij}} \frac{Viv_{ij}}{l_{ij}} \quad (1)$$

Donde **Pob** es la población del estrato, **Pob_i** es la población del AGEB *i*, **Pob_{ij}** es la población de la manzana o localidad *ij*, **Viv_{ij}** es el número de viviendas en la manzana o localidad *ij*; **m** es el número de AGEB seleccionadas, **k_i** es el número de manzanas o localidades seleccionadas en el AGEB *i* y **l_{ij}** es el número de viviendas seleccionadas en la manzana o localidad *ij*.

En la última etapa de selección, en cada vivienda se seleccionó a una persona de edad 15 o más años. Entonces, el inverso de la probabilidad de selección de un individuo es $w_I = w_H f$ donde **f** es el número de personas en el hogar con 15 o más años.

5. No respuesta

El tamaño de muestra fue de 1,656 hogares seleccionados. Los resultados finales que se obtuvieron fueron los del cuadro 1.

Cuadro 2. Resultados del cuestionario de hogar

Resultado	Frecuencia	Porcentaje
Completa	1,584	90
Informante inadecuado	2	0
Ausencia de ocupantes	29	2
Se negó a dar información	65	4
Vivienda deshabitada*	38	2
Vivienda de uso temporal*	27	2
No habla español	7	0
No es vivienda*	10	1
Total	1,762	100

A partir del cuadro 2 se puede deducir que la tasa de respuesta en hogar fue del 95.7%. La tasa de respuesta se calculó después de excluir a viviendas deshabitadas o temporales, además de las construcciones listadas durante el levantamiento cartográfico que resultaron no ser viviendas. Respecto a los cuestionarios individuales, se hicieron 1,584, de los cuales el 98% fueron completos.

Cuadro 3. Resultados del cuestionario individual

Resultado	Frecuencia	Porcentaje
Completa	1,549	98
Ausencia de persona seleccionada	12	1
Se negó a dar observación	10	1
Otro	13	1
Total	1,584	100

El ajuste por no respuesta se hizo a nivel AGEb; esto es para cada AGEb se calculó r , la tasa de respuesta y los ponderadores se multiplicaron por $1/r$.

6. Ajuste de los ponderadores para reproducir parámetros poblacionales

Los ponderadores (1) fueron ajustados para considerar el crecimiento poblacional entre 2010 y 2014.

7. Validación de la muestra

Dado que la encuesta es probabilística, es de esperarse que la encuesta reproduzca de manera razonable a la distribución de edades de los integrantes del hogar. Entonces, para verificar dicha afirmación. La gráfica siguiente compara la distribución no ponderada por edad del Censo 2010, la ENSANUT 2012 y la EEPS-Gto 2014. En la gráfica se observa que los intervalos de confianza de la ENSANUT 2012 y la EEPS-Gto 2014 se intersecan para todos los grupos de edad.

Figura 2. Distribución por edad del Censo 2010, ENSANUT 2012 y EEPS-Gto 2014. Las barras indican a los intervalos de confianza

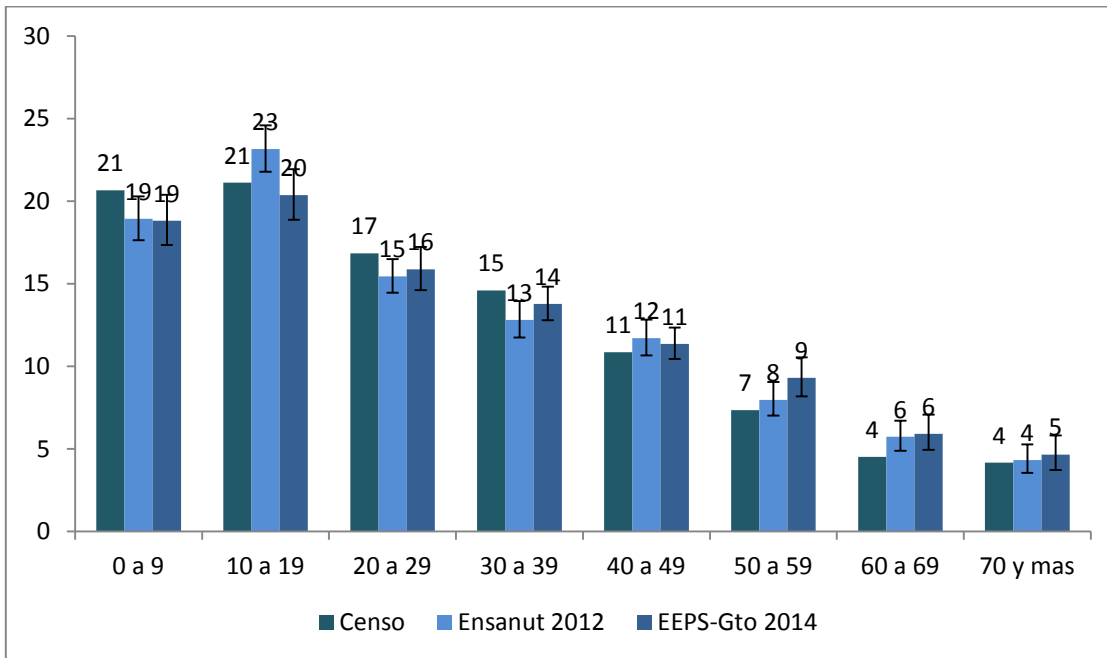
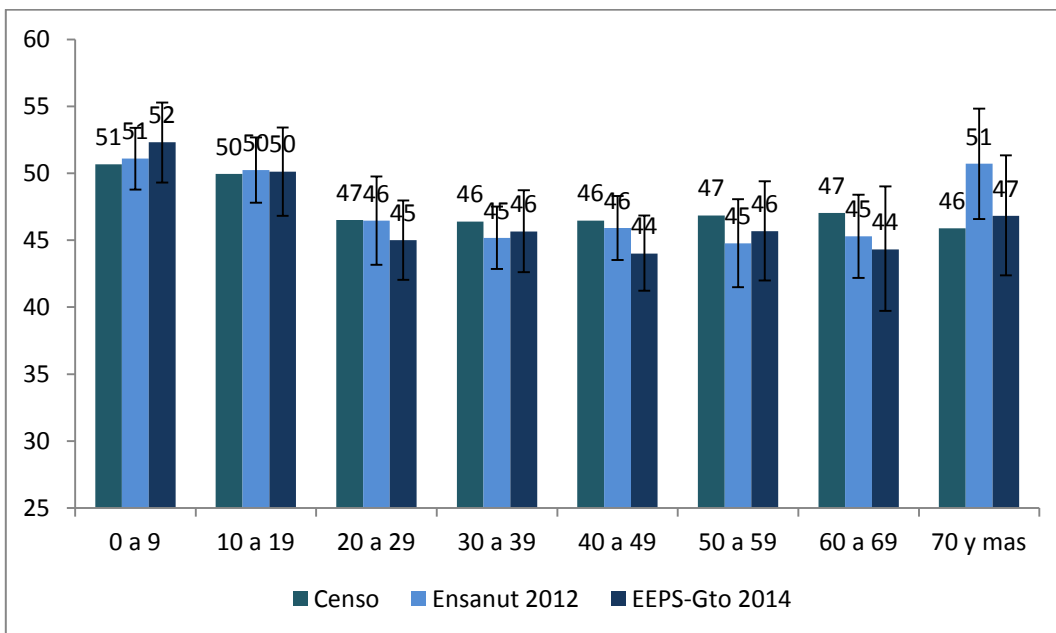


Figura 3. Proporción de hombres según grupo de edad en el Censo 2010, ENSANUT 2010 y la EEPS-Gto 2014. La barra indica a los intervalos de confianza



8. Validación de la selección

Después de que se listan a todos los integrantes del hogar, debe de seleccionarse a un adulto con edad 15 o más años. Entonces, para verificar que el procedimiento de selección funcionó bien, las tablas siguientes comparan la pirámide poblacional de los individuos seleccionados. La pirámide de los adultos seleccionados se presenta ponderada por el número de adultos con edad 15 o más años del hogar porque la probabilidad de selección depende del número de integrantes del hogar.

Figura 4. Distribución por edad del Censo 2010 y EEPS-Gto 2014. Las barras indican a los intervalos de confianza

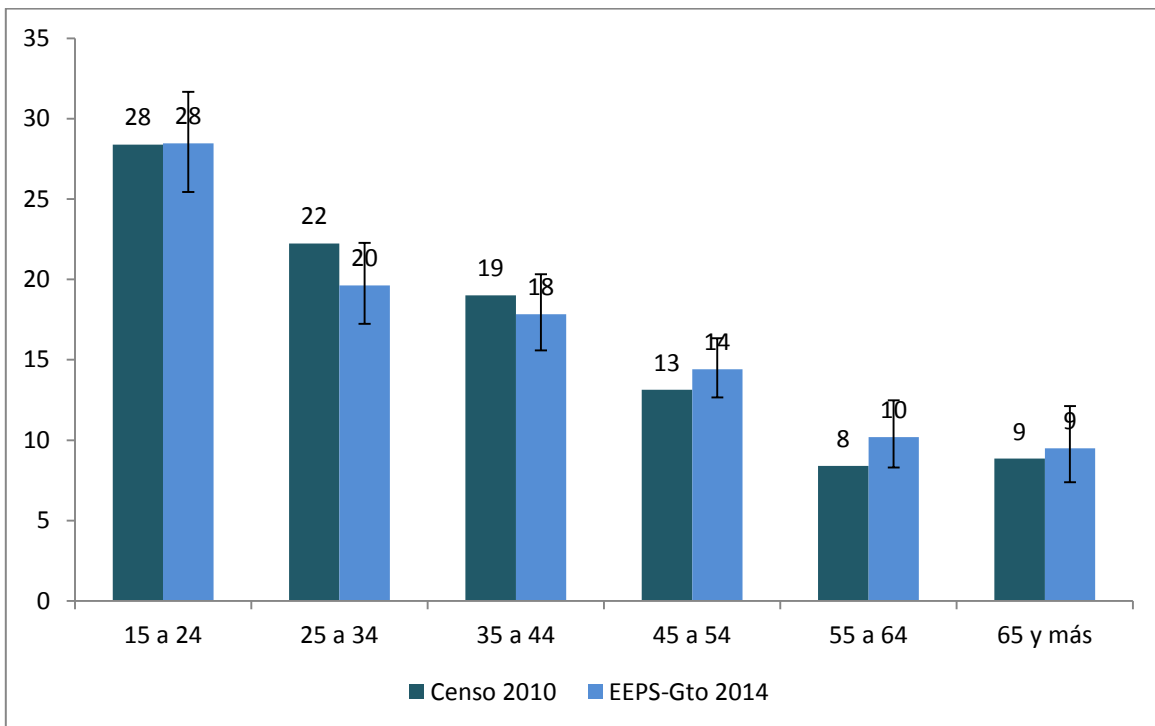
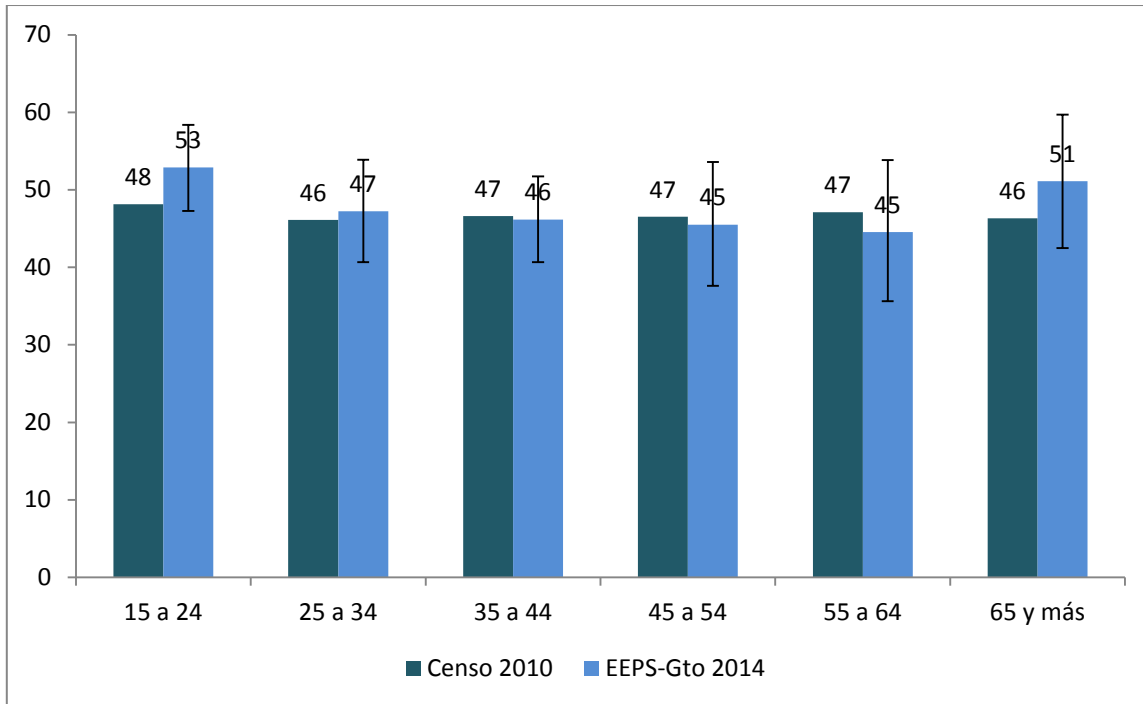


Figura 5. Proporción de hombres según grupo de edad en el Censo 2010 y la EEPS-Gto 2014. La barra indica a los intervalos de confianza



Las figuras 4 y 5 muestran que hay concordancia entre la pirámide de edad de las personas con edad 15 años o más seleccionadas por la EEPS-Gto 2014 y la pirámide del censo de 2010.

9. Resultados

Como resultado de la construcción de los ponderadores se obtuvieron las siguientes estimaciones, las cuales se presentan con el único fin de ilustrar el valor de la suma de los ponderadores.

Cuadro 5. Distribución viviendas según combustible para cocinar en EEPS-Gto 2014

	Estimador	Intervalo de confianza 95%		Conteo Simple
		Inferior	Superior	
Totales				
leña?	135,545	103,276	167,813	161
carbón?	1,670	-740	4,079	2
gas de cilindro o tanque (estacionario)?	1,288,411	1,174,263	1,402,559	1398
gas natural o de tubería?	19,392	-5,596	44,380	17
electricidad?	4,630	648	8,613	5
¿Otro combustible?	1,962	-2,022	5,946	1
Total	1,451,610	1,346,035	1,557,186	1584
Porcentajes				
leña?	9.3	7.2	12.1	161
carbón?	0.1	0.0	0.5	2
gas de cilindro o tanque (estacionario)?	88.8	85.6	91.3	1398
gas natural o de tubería?	1.3	0.4	4.7	17
electricidad?	0.3	0.1	0.7	5
¿Otro combustible?	0.1	0.0	1.0	1
Total	100	100	100	1584

Cuadro 5. Distribución de individuos según derechohabencia (primera opción) en EEPS-Gto 2014

	Estimador	Intervalo de confianza 95%		Conteo Simple
		Inferior	Superior	
Totales				
del Seguro Social (IMSS)?	1,715,626	1,452,138	1,979,114	1,778
del ISSSTE?	248,193	147,041	349,345	233
del ISSSTE estatal?	55,364	8,561	102,167	61
de Pemex?	22,270	-19,965	64,504	15
Defensa o Marina?	18,929	-3,123	40,982	26
del Seguro Popular o para una Nueva Generación?	2,711,414	2,431,768	2,991,059	2,919
de un seguro privado?	46,904	-3,434	97,242	33
entonces, ¿no está afiliado o inscrito a los servicios médi	1,006,524	865,201	1,147,847	1,025
NS/NR	19,697	-176	39,570	17
Total	5,844,921	5,452,716	6,237,126	6,107
Porcentajes				
del Seguro Social (IMSS)?	29.4	25.9	33.0	1,778
del ISSSTE?	4.2	2.8	6.4	233
del ISSSTE estatal?	0.9	0.4	2.3	61
de Pemex?	0.4	0.1	2.5	15
Defensa o Marina?	0.3	0.1	1.0	26
del Seguro Popular o para una Nueva Generación?	46.4	42.6	50.3	2,919
de un seguro privado?	0.8	0.3	2.3	33
entonces, ¿no está afiliado o inscrito a los servicios médi	17.2	15.2	19.5	1,025
NS/NR	0.3	0.1	0.9	17
Total	100	100	100	6,107

Cuadro 6. Distribución individuos 15 y más según opinión sobre la afirmación “Está satisfecho con su vida” en EEPS-Gto 2014

	Estimador	Intervalo de confianza 95%		Conteo
		Inferior	Superior	Simple
Totales				
1. Totalmente en desacuerdo	48,453	25,405	71,501	21
2. Bastante en desacuerdo	67,326	40,262	94,390	30
3. Algo en desacuerdo	161,075	112,199	209,950	65
4. Ni de acuerdo ni en desacuerdo	308,465	233,456	383,474	129
5. Algo acuerdo	947,018	845,800	1,048,237	365
6. Bastante de acuerdo	1,315,496	1,151,980	1,479,013	513
7. Totalmente de acuerdo	1,167,755	1,026,977	1,308,533	426
Total	4,015,588	3,744,348	4,286,829	1,549
Porcentajes				
1. Totalmente en desacuerdo	1.2	0.8	1.9	21
2. Bastante en desacuerdo	1.7	1.1	2.5	30
3. Algo en desacuerdo	4.0	2.9	5.5	65
4. Ni de acuerdo ni en desacuerdo	7.7	6.2	9.5	129
5. Algo acuerdo	23.6	21.4	25.9	365
6. Bastante de acuerdo	32.8	29.8	35.8	513
7. Totalmente de acuerdo	29.1	26.1	32.3	426
Total	100	100	100	1,549

Finalmente, como último resultado del procedimiento de ponderación se anexaron a las base de datos las siguientes variables.

Cuadro 7. Variables del diseño de muestra anexadas a las bases de datos

Variable	Descripción
Pondef	Factor de expansión.
Code_UPM	Identificador de Unidad Primaria de Muestro, se usa para calcular varianzas
Est_var	Variable de estratificación para calcular varianzas
Est_u	<p>Estrato de urbanidad.</p> <p>Rural (1) = Localidades con menos de 2500 habitantes</p> <p>Metropolitana (3) = Localidades con 100mil habitantes o más, capitales de los estados y localidades con más de 2500 de la áreas metropolitanas de INEGI en 2000</p> <p>Urbana (2) = Resto del estado</p>
Región	<p>Región geográfica.</p> <p>Noreste (1), Norte(2), Centro Oeste(3), Suroeste(4), Sur (5), Centro Este (6).</p>