



# Metodología para la asociación de áreas geoestadísticas a la cartografía electoral



## Índice

Introducción	5
Antecedentes	7
Metodología	7
Etapa 1	7
Etapa 2	8
Etapa 3	9
Etapa 4	10
Anexos	15
Anexo I	17
Anexo II	18
Anexo III	19
Glosario	24

## Introducción

La necesidad de contar con información que permita elaborar análisis demográficos y socioeconómicos para cada una de las unidades electorales del país, definidos por el Instituto Nacional Electoral, ha conducido a la Dirección Ejecutiva del Registro Federal de Electores (DERFE) a emprender la tarea minuciosa de trasladar la información de la Encuesta Intercensal 2015 a cada una de estas unidades.

El objetivo central de dicha tarea es proporcionar al público en general las herramientas e información geográfica que permitan realizar la consulta estadística, y la elaboración de mapas temáticos a escala de distritos electorales federales en la República Mexicana.

Para obtener las Estadísticas Intercensales a Escalas Geoelectorales, fue necesario definir una metodología con base en el trabajo colegiado entre el Instituto Nacional de Estadística y Geografía (INEGI) y el Instituto Nacional Electoral (INE). Gracias a este trabajo conjunto fue posible establecer los procedimientos para incorporar la información de la Encuesta Intercensal 2015 a las unidades geoelectorales y con ello facilitar el estudio de la población a nivel de Distrito Electoral.

## Antecedentes

El grupo colegiado INEGI - INE, tomó como base las Estadísticas Censales a Escalas Goelectorales, obtenidas del II Conteo de Población y Vivienda, y las más recientes generadas con los resultados del Censo de Población y Vivienda 2010, para realizar, a partir de los resultados de la Encuesta Intercensal 2015, una nueva versión del producto, utilizando para ello el Sistema de Información Geográfica desarrollado por el INEGI, denominada Mapa Digital de México (MDM).

Cabe mencionar que la decisión de presentar en este producto información a nivel de distrito federal electoral, y no a nivel de sección como en los ejercicios anteriores, obedece a que la Encuesta Intercensal 2015 sólo se levantó en una muestra probabilística de manzanas, por lo que el tamaño de muestra a nivel sección no es suficiente para realizar estimaciones con precisión y confianza aceptables.

## Metodología

Para el desarrollo del proyecto se definieron cuatro etapas:

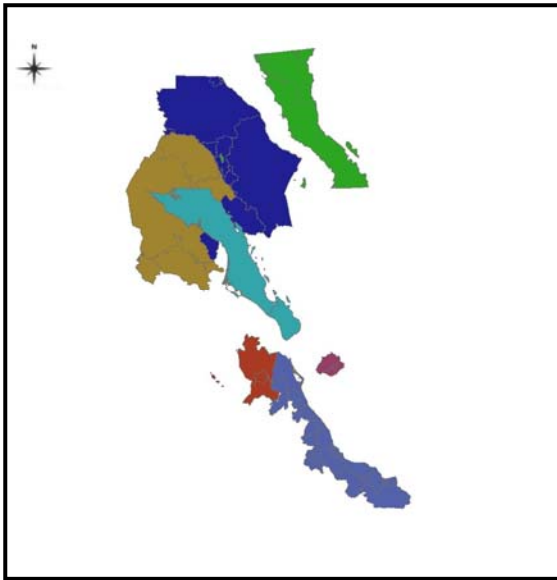
- Etapas 1:** Homogeneizar los sistemas de coordenadas de las cartografías de las dos instituciones y construir los continuos nacionales a nivel manzana y distrito.
- Etapas 2:** Asociar claves de distrito electoral a las manzanas y localidades del Marco Geoestadístico del INEGI.
- Etapas 3:** Analizar espacialmente los resultados de la asociación para garantizar la consistencia en los resultados a nivel estatal.
- Etapas 4:** Asignar espacialmente los resultados de la Encuesta Intercensal 2015 a los Distritos Electorales federales.

### **Etapas 1. Homogeneizar los sistemas de coordenadas de las cartografías de las dos instituciones y construir los continuos nacionales a nivel manzana y distrito.**

La cartografía del INEGI utilizada fue la capa de manzanas y localidades rurales puntuales del Marco Geoestadístico, correspondientes al corte de la Encuesta Intercensal 2015 y del INE la Distritación Federal 2016-2017. La primera tarea importante fue realizar el cambio de proyección de la información proporcionada por el INE, debido a las diferencias en los marcos cartográficos; mientras que el Marco Geoestadístico del INEGI es un continuo nacional en proyección CCL (*Cónica Conforme de Lambert*), la cartografía del INE, correspondiente a la Distritación Federal 2016-2017, se encuentra a nivel estatal

en una proyección UTM (*Universal Transversa de Mercator*). La imagen 1 ilustra una representación nacional empleando ambas proyecciones cartográficas.

Cartografía en proyección UTM



Cartografía en proyección CCL



Imagen 1. Cartografía en proyecciones UTM (izquierda) y CCL (derecha).

## **Etapla 2. Asociar claves de distrito electoral a las manzanas y localidades del Marco Geoestadístico del INEGI.**

Las unidades mínimas del Marco Geoestadístico del INEGI son las manzanas en el caso de las localidades ameznadas; y la localidad rural, cuando estas no cuentan con ameznamiento. En esta etapa se asignaron las claves del Distrito Electoral a cada manzana o localidad del INEGI.

El primer criterio prevaleciente para esta etapa del proceso fue la no división de manzanas o localidades; es decir, una unidad geoestadística debe pertenecer a un distrito y sólo a uno. La asignación se llevó a cabo mediante intersección espacial, esto es, **sobreponiendo a los distritos electorales del INE y los centroides de las manzanas o localidades puntuales** (para las localidades rurales no ameznadas). La imagen 2 ilustra las claves de distrito asignadas espacialmente a los centroides de manzana y/o localidades rurales del INEGI. El resultado generó la primera versión de la relación entre los marcos cartográficos de INE e INEGI. Posteriormente se analizaron los casos especiales donde existían diferencias en la digitalización de los objetos espaciales, por lo que fue necesario hacer ajustes que corresponden a la siguiente etapa del proyecto.

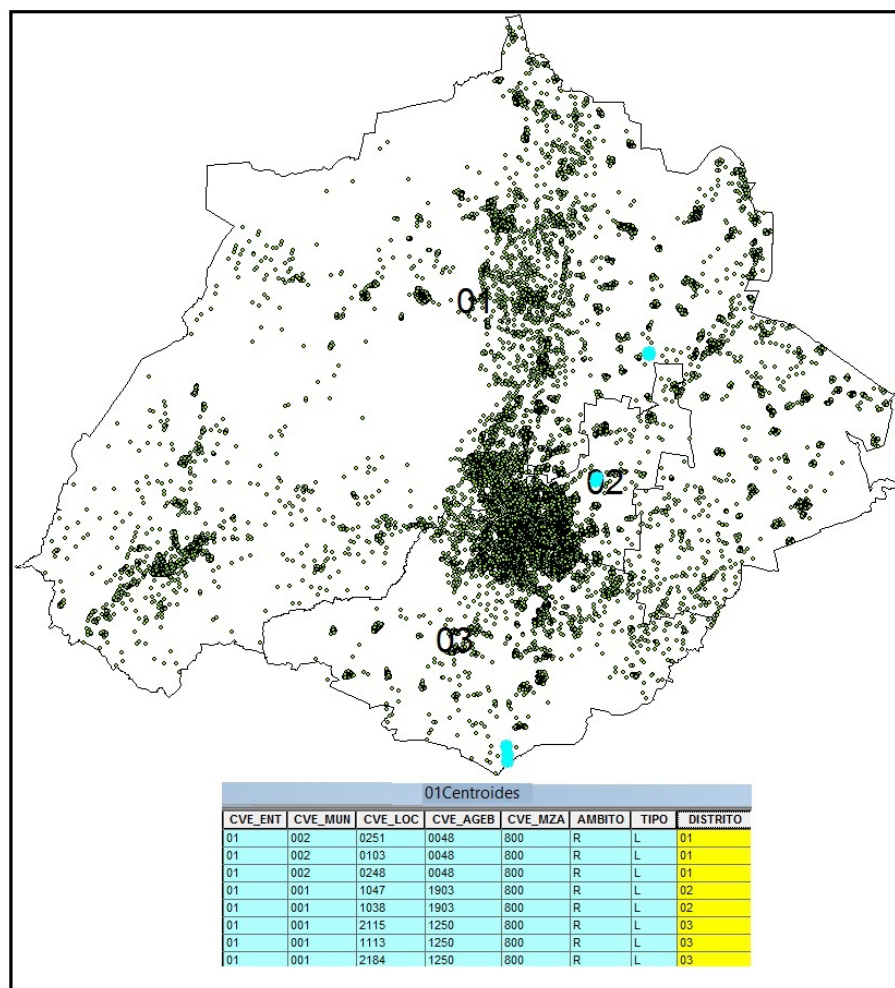


Imagen 2. Centroides de manzanas y localidades rurales, y su relación espacial con los Distritos Electorales.

### **Etap 3. Analizar espacialmente los resultados de la asociación para garantizar consistencia en los resultados a nivel estatal.**

Se partió de un análisis espacial para detectar y corregir, de manera colegiada, los casos en que hubiera centroides de manzanas o localidades rurales puntuales sin distrito electoral, sobre todo en aquellos casos donde las fronteras estatales de uno y otro marco cartográfico presentan diferencias. La conclusión de esta etapa llevó a la asignación completa de claves de Distritos Electorales a las manzanas y localidades puntuales del Marco Geoestadístico del INEGI.

Para los eventos en que los centroides de localidades o manzanas se situaban fuera de los límites distritales de una entidad, pero estaban contenidos en la poligonal del Marco Geoestadístico de la misma entidad, la asignación correspondió al distrito electoral más cercano. Como se aprecia en el ejemplo de la imagen 3 los centroides espacialmente fuera de los distritos electorales, se asignaron de

forma manual al distrito más cercano. No obstante el criterio anterior, prevalecieron 43 localidades rurales cuya ubicación se aproximó a dos distritos; estas situaciones se revisaron puntualmente y asignaron manualmente.

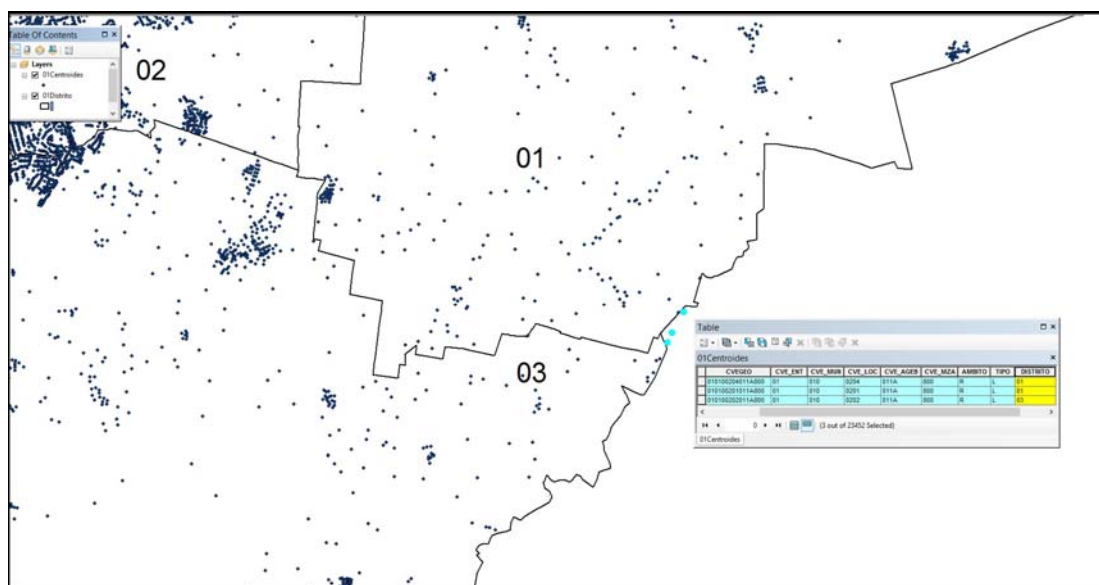


Imagen 3. Centroides asignados de forma manual a los distritos 01 y 03 respectivamente

#### **Etapas 4. Asignar espacialmente los resultados de la Encuesta Intercensal 2015 a los Distritos Electorales federales.**

Como ya se mencionó previamente, dado que la Encuesta Intercensal 2015 sólo se levantó en una muestra probabilística de manzanas y localidades, fue necesario hacer una revisión para asegurar que el tamaño de muestra en cada una de las unidades electorales permitiera hacer estimaciones de calidad para los diferentes indicadores.

Derivado de esta revisión, se determinó que no es posible realizar la asignación de resultados a nivel sección. En cuanto al tamaño de muestra en cada uno de los distritos electorales federales, se tiene lo siguiente:

	Mínimo	Promedio	Máximo
Tamaño de muestra por distrito	4 090	19 515	85 215
Proporción mínima estimable	1.6%	0.9%	0.5%
Confianza	90%	90%	90%
Error relativo máximo	25%	15%	10%

Por lo anterior, se determinó que los tamaños de muestra a nivel distrito electoral federal permiten hacer estimaciones de calidad a partir de la información de la Encuesta Intercensal 2015. Cabe mencionar que se presentaron algunas situaciones derivadas de la asociación entre el marco INEGI y los distritos electorales federales.

La primera situación detectada fue que algunos de los estratos y de las unidades primarias de muestreo (UPM) quedaron divididas en dos o más distritos, las áreas en esta situación son las siguientes:

- 6.95% de los estratos están divididos entre dos o más distritos.
- 0.22% de las unidades primarias de muestreo están divididas entre dos o más distritos.

La distribución de esta situación por entidad federativa se muestra en el Anexo II. Derivado de esta situación, el cálculo de las varianzas de los indicadores se realizó considerando los distritos como subpoblaciones, utilizando para ello el software SAS. La metodología para el cálculo de los indicadores por se encuentra en el Anexo III del presente documento.

La segunda situación a considerar es que, durante el levantamiento de la Encuesta Intercensal 2015, no fue posible captar la información completa para 11 municipios, los cuales se encuentran en los estados de Chihuahua, Oaxaca, Puebla y Sonora, mismos que se encuentran en 8 distritos federales electorales:

Entidad	Distrito	Porcentaje que representan los municipios con muestra insuficiente respecto al total distrital	
		Viviendas	Población
<b>Nacional</b>		<b>0.14</b>	<b>0.13</b>
Chihuahua	02	5.85	5.94
Chihuahua	07	3.35	3.17
Chihuahua	08	0.01	0.01
Chihuahua	09	9.96	8.74
Oaxaca	03	0.22	0.20
Oaxaca	07	16.30	16.52
Puebla	05	2.67	2.83
Sonora	01	4.20	4.43

Si bien se considera que la distribución de los indicadores no se ve afectada de manera significativa a causa de estos municipios con muestra insuficiente, es importante tenerlo en consideración para los análisis que se hagan de dichos distritos.



Para mayor información sobre las características de la Encuesta Intercensal 2015, puede consultarse la síntesis metodológica y conceptual, ubicada en la siguiente liga:

<http://www.beta.inegi.org.mx/app/biblioteca/ficha.html?upc=702825078836>

Teniendo en cuenta estas situaciones, se realizó el cálculo de los indicadores intercensales a para cada uno de los distritos federales electorales. En total, se incluyen 107 indicadores sociodemográficos para las unidades geoelectorales. Para conocer la metodología del cálculo de los indicadores incluidos por distrito federal electoral, se puede consultar el documento “Metadatos para el cálculo de indicadores”, incluido como parte de la documentación del sistema.

Adicional a los indicadores sociodemográficos, se incluye dentro de la información el grado de complejidad electoral. Para entender el concepto y la construcción de este indicador, puede consultarse la metodología en la siguiente página:

<http://cartografia.ife.org.mx/sige7/?mapoteca=geoelect&TDE-DERFE>

Finalmente, se encuentran identificados los distritos clasificados como indígenas por el Instituto Nacional Electoral, que corresponden a las siguientes 28 unidades electorales:

Entidad	Distrito	Cabecera	
		Municipio	Localidad
<b>Chiapas</b>	01	Palenque	Palenque
<b>Chiapas</b>	02	Bochil	Bochil
<b>Chiapas</b>	03	Ocosingo	Ocosingo
<b>Chiapas</b>	05	San Cristóbal de las Casas	San Cristóbal de las Casas
<b>Chiapas</b>	11	Las Margaritas	Las Margaritas
<b>Guerrero</b>	05	Tlapa de Comonfort	Tlapa
<b>Guerrero</b>	06	Chilapa de Álvarez	Chilapa
<b>Hidalgo</b>	01	Huejutla de Reyes	Huejutla de Reyes
<b>Hidalgo</b>	02	Ixmiquilpan	Ixmiquilpan
<b>Oaxaca</b>	01	San Juan Bautista Tuxtepec	San Juan Bautista Tuxtepec
<b>Oaxaca</b>	02	Teotitlán de Flores Magón	Teotitlán de Flores Magón
<b>Oaxaca</b>	04	Tlacolula de Matamoros	Tlacolula de Matamoros
<b>Oaxaca</b>	05	Salina Cruz	Salina Cruz
<b>Oaxaca</b>	06	Heroica Ciudad de Tlaxiaco	Heroica Ciudad de Tlaxiaco
<b>Oaxaca</b>	07	Ciudad Ixtepec	Ciudad Ixtepec
<b>Oaxaca</b>	09	San Pedro Mixtepec	Puerto Escondido
<b>Puebla</b>	01	Huauchinango	Huauchinango de Degollado
<b>Puebla</b>	02	Zacatlán	Cuautlulco Barrio
<b>Puebla</b>	03	Teziutlán	Teziutlán

Entidad	Distrito	Cabecera	
		Municipio	Localidad
<b>Puebla</b>	04	Ajalpan	Ajalpan
<b>Quintana Roo</b>	02	Othón P. Blanco	Chetumal
<b>San Luis Potosí</b>	07	Tamazunchale	Tamazunchale
<b>Veracruz de Ignacio de la Llave</b>	02	Tantoyuca	Tantoyuca
<b>Veracruz de Ignacio de la Llave</b>	06	Papantla	Papantla de Olarte
<b>Veracruz de Ignacio de la Llave</b>	18	Zongolica	Zongolica
<b>Yucatán</b>	01	Valladolid	Valladolid
<b>Yucatán</b>	02	Progreso	Progreso
<b>Yucatán</b>	05	Ticul	Ticul

## Anexos

## Anexo I

El sistema utilizado para difundir el proyecto es la plataforma de Mapa Digital de México (MDM V6.3.0).

Los datos de la proyección en la que se crearon los shapes finales para el MDM fue la definida por el sistema de coordenadas CCL (Conical Conformal Lambert) cuyos parámetros son:

Projected Coordinate System:	North_America_Lambert_Conformal_Conic_ITRF_2008
Projection:	Lambert_Conformal_Conic
False_Easting:	2500000.00000000
False_Northing:	0.00000000
Central_Meridian:	-102.00000000
Standard_Parallel_1:	17.50000000
Standard_Parallel_2:	29.50000000
Latitude_Of_Origin:	12.00000000
Linear Unit:	Meter
Geographic Coordinate System:	GCS_ITRF_2008
Datum:	D_ITRF_2008
Prime Meridian:	Greenwich
Angular Unit:	Degree

**Anexo II**

Estratos y UPM divididas entre estratos por entidad federativa

Entidad federativa	Estratos en muestra			Upm en muestra		
	Total	Divididos	%	Total	Divididos	%
<b>Nacional</b>	<b>17 470</b>	<b>1 215</b>	<b>6.95</b>	<b>273 773</b>	<b>603</b>	<b>0.22</b>
Aguascalientes	125	15	12.00	1 861	5	0.27
Baja California	192	61	31.77	2 423	19	0.78
Baja California Sur	93	13	13.98	1 316	2	0.15
Campeche	143	1	0.70	2 108	0	0.00
Coahuila de Zaragoza	300	42	14.00	6 080	17	0.28
Colima	105	3	2.86	1 629	3	0.18
Chiapas	1 079	91	8.43	17 169	63	0.37
Chihuahua	406	35	8.62	11 617	26	0.22
Ciudad de México	146	54	36.99	3 373	5	0.15
Durango	271	20	7.38	6 457	11	0.17
Guanajuato	655	80	12.21	8 139	35	0.43
Guerrero	716	43	6.01	13 634	26	0.19
Hidalgo	632	31	4.91	9 732	28	0.29
Jalisco	1 048	65	6.20	16 245	20	0.12
México	1 328	97	7.30	15 044	29	0.19
Michoacán de Ocampo	938	41	4.37	14 385	23	0.16
Morelos	317	18	5.68	4 207	14	0.33
Nayarit	174	4	2.30	3 534	6	0.17
Nuevo León	343	37	10.79	7 242	13	0.18
Oaxaca	2 212	65	2.94	28 348	53	0.19
Puebla	1 308	58	4.43	18 388	32	0.17
Querétaro	224	16	7.14	2 918	2	0.07
Quintana Roo	127	12	9.45	2 163	7	0.32
San Luis Potosí	445	33	7.42	8 411	20	0.24
Sinaloa	272	48	17.65	4 429	15	0.34
Sonora	429	31	7.23	8 247	18	0.22
Tabasco	277	19	6.86	2 078	1	0.05
Tamaulipas	343	34	9.91	7 173	11	0.15
Tlaxcala	292	20	6.85	5 021	18	0.36
Veracruz de Ignacio de la Llave	1 649	88	5.34	24 440	58	0.24
Yucatán	445	22	4.94	8 522	10	0.12
Zacatecas	436	18	4.13	7 440	13	0.17

## Anexo III

### 1. Metodología para calcular estimadores

El presente apartado describe las funciones utilizadas para calcular los estimadores de indicadores y las varianzas, basados en un muestreo complejo para un dominio de estudio  $m$ . Para el caso de la Encuesta Intercensal 2015, los dominios de estudio son las Entidades federativas y los municipios.

#### *Estimador de totales*

$$\hat{Y}_m = \sum_{h=1}^H \hat{Y}_h$$

$$\hat{Y}_h = \sum_{j=1}^H \sum_{k=1}^{n_h} \sum_{l=1}^{M_j} F_{hj} Y_{hjk}$$

Donde:

$\hat{Y}_m$  = Es el total estimado de la característica de interés para la subpoblación  $m$

$\hat{Y}_h$  = Es el total estimado de la característica de interés en el estrato  $h$  de la subpoblación  $m$

$F_{hj}$  = Es el factor de expansión de la  $j$ -ésima UPM en el  $h$ -ésimo estrato

$Y_{hjk}$  = Es el valor de la característica de interés en la  $k$ -ésima vivienda de la  $j$ -ésima UPM en el  $h$ -ésimo estrato

$H$  = Número de estratos en la subpoblación  $m$

$M_j$  = Número de elementos en muestra dentro de la  $j$ -ésima UPM en el  $h$ -ésimo estrato

#### *Estimador de la media y proporciones*

$$\hat{\bar{Y}}_m = \frac{\hat{Y}_m}{M_m}$$

$$= \frac{\hat{Y}_m}{\sum_{h=1}^H \sum_{j=1}^{n_h} F_{hj}}$$

Donde:

$M_m$  = Estimador del total de elementos en la población

### 1.1. Error estándar (EE)

Existen expresiones específicas para el cálculo del error estándar acorde con cada tipo de estimador (total, media, razón, proporción, etcétera).

Para la estimación de un total, el estimador de la varianza del estimador, para un dominio dado, está dada por:

$$\hat{V}(\hat{Y}_m) = \sum_{h=1}^H \hat{V}(\hat{Y}_h) = \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{S}_h^2}{n_h}$$

Donde:

$$\hat{S}_h^2 = \sum_{j=1}^{n_h} \frac{(y_{hj} - \bar{y}_h)^2}{n_h - 1}$$

Por ende, el error estándar asociado está dado por:

$$EE_{\hat{Y}_m} = \sqrt{\hat{V}(\hat{Y}_m)}$$

Si lo que se requiere es el estimador de la varianza del estimador de la media de una característica en particular, se usa la fórmula:

$$\hat{V}(\hat{\bar{Y}}_m) = \sum_{h=1}^H \hat{V}\left(\frac{\hat{Y}_h}{M_m}\right) = \sum_{h=1}^H \frac{1}{M_m^2} \hat{V}(\hat{Y}_h)$$

Y el error estándar para la media está dado por:

$$EE_{\hat{\bar{Y}}_m} = \sqrt{\hat{V}(\hat{\bar{Y}}_m)}$$

## 2. Estimadores para una subpoblación

Para una subpoblación  $D$ , sea  $I_D$  la variable indicadora, tal que

$$I_D = \begin{cases} 1 & \text{si la observacion } (h, j, k) \text{ pertenece a } D \\ 0 & \text{de otra forma} \end{cases}$$

y sean

$$z_{hjk} = y_{hjk} I_D(h, j, k) = \begin{cases} y_{hjk} & \text{si la observacion } (h, j, k) \text{ pertenece a } D \\ 0 & \text{de otra forma} \end{cases}$$

$$v_{hjk} = F_{hjk} I_D(h, j, k) = \begin{cases} F_{hjk} & \text{si la observacion } (h, j, k) \text{ pertenece a } D \\ 0 & \text{de otra forma} \end{cases}.$$

Los estimadores para la variable  $y$  en la subpoblación  $D$  se calculan utilizando los nuevos factores  $v_{hjk}$ .

El estimador del total en subpoblaciones está dado por:

$$\hat{Y}_D = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{k=1}^{M_j} v_{hjk} y_{hjk}$$

Mientras que el estimador de la media resulta ser:

$$\widehat{\bar{Y}}_D = \left( \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{k=1}^{M_j} v_{hjk} y_{hjk} \right) / v_{...}$$

Donde:

$$v_{...} = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{k=1}^{M_j} v_{hjk}$$

En otras palabras  $v_{...}$  representa el total de viviendas en la subpoblación  $D$  para el caso de la estimación de indicadores de vivienda.

Para la estimación de un total, la varianza del estimador, para una subpoblación está dada por:

$$\hat{V}(\hat{Y}_D) = \sum_{h=1}^H \hat{V}_h(\hat{Y}_D)$$



Donde:

\* si  $n_h > 1$ , entonces:

$$\hat{V}_h(\hat{Y}_D) = \frac{n_h(1-f_h)}{n_h-1} \sum_{j=1}^{n_h} (z_{hj\cdot} - \bar{z}_{h\cdot\cdot})^2$$

$$z_{hj\cdot} = \sum_{k=1}^{M_j} v_{hjk} z_{hjk}$$

$$\bar{z}_{h\cdot\cdot} = \left( \sum_{j=1}^{n_h} z_{hj\cdot} \right) / n_h$$

\* Si  $n_h = 1$ , entonces  $\hat{V}_h(\hat{Y}_D) = 0$ , es decir, no aporta a la estimación de la varianza global.

Por ende, el error estándar asociado está dado por:

$$EE_{\hat{Y}_m} = \sqrt{\hat{V}(\hat{Y}_m)}$$

La tasa de muestreo  $f_h$  para el estrato  $h$ , que se utiliza en la estimación de la varianza de la serie de Taylor, es la fracción de unidades primarias de muestreo (UPM's) seleccionadas para la muestra. Y  $n_h$  corresponde al número de observaciones en muestra en el estrato  $h$ .

El correspondiente estimador de la varianza del estimador de la media está dado por:

$$\hat{V}(\hat{Y}_D) = \sum_{h=1}^H \hat{V}_h(\hat{Y}_D)$$

Donde:

\* si  $n_h > 1$ , entonces

$$\hat{V}_h(\hat{Y}_D) = \frac{n_h(1-f_h)}{n_h-1} \sum_{j=1}^{n_h} (r_{hj\cdot} - \bar{r}_{h\cdot\cdot})^2$$

$$r_{hj\cdot} = \left( \sum_{k=1}^{M_j} v_{hjk} (y_{hjk} - \hat{Y}_D) \right) / v_{\dots}$$

$$\bar{r}_{h..} = \left( \sum_{j=1}^{n_h} r_{hj.} \right) / n_h$$

Si  $n_h = 1$ , entonces,  $\hat{V}_h(\widehat{\bar{Y}_D}) = 0$ , es decir, no aporta a la estimación de la varianza global.

El error estándar para la media está dado por:

$$EE_{\hat{Y}_m} = \sqrt{\hat{V}(\widehat{\bar{Y}_m})}$$

## Glosario.

*Centroide:* Es un punto referenciado al centro de un polígono espacial.

*CCL:* Cónica Conforme de Lambert, proyección cartográfica de tipo cónica que se utiliza para las cartas de pequeña escala que representan grandes porciones o la totalidad de la República Mexicana.

*Localidad puntual:* Punto espacial que representa una localidad no amanzanada y está colocado aproximadamente al centro de la localidad.

*Shapes:* Archivos vectoriales para la representación cartográfica digital.

*UTM:* Universal Transversal de Mercator, proyección cartográfica de tipo cilíndrica.

*Zonas UTM:* Superficie terrestre representada entre los puntos de corte del cilindro UTM. La representación total de la tierra se obtiene girando este cilindro cada 6°. A cada faja de 6° se le da el nombre de zona UTM. México queda comprendido entre las zonas 11 y 16.