



Guía de uso del Estratificador INEGI 2.0

Octubre, 2024

Contenido

Introducción	3
1. Estratificación	3
2. Descripción del Estratificador INEGI 2.0	4
2.1 Variables	6
2.2 Número de estratos	7
2.3 Nivel de desagregación	7
2.4 Transformar datos	8
2.5 Métodos de agrupación	9
3. Resultados	12
3.1 Mapa estratificado	13
3.2 Gráfica de dispersión	14
3.3 Gráfica de centroides	15
3.4 Gráfica de componentes principales	15
4. Múltiples ejercicios durante la misma sesión	16
5. Criterios estadísticos para llevar a cabo comparaciones entre ejercicios alternativos	21
6. Exportación de resultados	23
6.1 Mapas	23
6.2 Gráficas	24
6.3 Resultados numéricos	25

Introducción

Consciente de la creciente necesidad de quienes toman decisiones de contar con información de calidad, oportuna y adecuada, así como de la evolución experimentada por los procedimientos de estratificación y por las infraestructuras de cómputo, el **Instituto nacional de Estadística y Geografía (INEGI)** pone hoy a su disposición una nueva versión del presente servicio llamado Estratificador INEGI 2.0¹ con la finalidad de permitirles construir agrupaciones o estratificaciones de áreas geográficas que se adecuen a sus necesidades, con base en información estadística que refleje semejanzas y diferencias entre dichas áreas.

En general, es posible decir que el propósito principal de los métodos de estratificación² el de formar grupos o estratos claramente diferenciados de unidades tan homogéneas al interior y tan distintas entre grupos como sea posible, a partir de información cuantitativa que toma la forma de una o más mediciones realizadas en todas y cada una de ellas. La asignación de las unidades³ a los estratos se realiza con base en procedimientos numéricos que utilizan esa información. Ahora bien, el logro del propósito planteado dependerá de diversos factores como son la misma información de la que se parte, o los recursos computacionales al alcance, o los métodos y procedimientos disponibles, entre otros, así como de las combinaciones entre todos ellos.

1. Estratificación

Con fines interpretativos, el territorio de un país o de una entidad federativa, o las manzanas dentro de una localidad, pueden ser vistos como la superposición de los territorios de varios "archipiélagos" cada uno formado por unidades geográficas o administrativas pertenecientes al mismo estrato y que, por esa razón, comparten una realidad que es diversa a la de las unidades pertenecientes a otros "archipiélagos", sin consideración a su vecindad geográfica. En consecuencia, cada una de esas realidades requiere, por ejemplo, de acciones diversas para dar atención a sus carencias. En efecto, una pareja de grupos puede mostrar semejanzas en algunos indicadores y diferencias significativas en otros; en tanto que una segunda pareja puede mostrar un comportamiento similar según otro grupo de indicadores. Por lo anterior, la reducción de las desigualdades en cada caso requerirá de esfuerzos no solo cuantitativa sino cualitativamente diferentes. En otras palabras, las acciones requeridas son diversas no solamente a causa de la magnitud del esfuerzo correspondiente, como puede parecer si se mira al ejercicio a través de un resumen en una sola dimensión (ej., "marginación"), sino también a causa de los distintos entornos a los que habrá que dedicar esfuerzos diferenciados, como corresponde cuando se está en presencia de un problema multidimensional.

Sin lugar a duda, los estudios sobre el tema de la marginación elaborados en México por el CONAPO⁴ representan una de las aplicaciones mejor conocidas y más significativas en el país, de los procedimientos orientados a la formación de clases, grupos, conglomerados o estratos bien diferenciados, conformado cada uno de ellos por unidades homogéneas. El innegable impacto de los trabajos del CONAPO en la elaboración e instrumentación de la política social mexicana en apoyo a habitantes de unidades territoriales clasificadas en los estratos de "ALTA" y "MUY ALTA MARGINACIÓN" difícilmente se hubiera alcanzado de no haber sido basados en un sólido cervo de información estadística como la aportada por los censos y conteos de población de los

¹ <https://www.inegi.org.mx/app/estratificador/>.

² Hartigan, J. A. (1975). Clustering algorithms. *John Wiley google schola*, 2, 25-47.

³ En la versión 2.0 del sistema solamente agrupa manzanas, municipios o estados.

⁴ Los documentos metodológicos, así como las bases de datos que dan cuenta del índice de marginación en diversos años, pueden consultarse en: <https://www.gob.mx/conapo/documentos/indices-de-marginacion-2020-284372#:~:text=Conoce%20los%20%C3%ADndices%20de%20marginaci%C3%B3n%20por%20entidad%20federativa.>

últimos años y en un procedimiento que, siendo ya viable para los equipos de cómputo de aquellas épocas, representó en su momento una importante innovación.

Por otro lado, basar en dicho resumen una estratificación, univariada a fin de cuentas, puede no ser la mejor forma de actuar [véase Bustos (2011)⁵]. En efecto, en un principio el CONAPO recurría a la primera componente principal para el cálculo del índice de marginación; hoy en día, el cálculo del índice se basa en el método de la distancia P_2 de Pena⁶. En el primer caso, la experiencia permitía una medición indirecta de la eficiencia del uso de la información como la proporción de la suma de las varianzas de los indicadores originales. Por ejemplo, para el caso municipal dicha proporción oscilaba alrededor de 50 % lo que hace suponer que la información aportada por el indicador compuesto presenta carencias. En otras palabras, el resumen podría ser mejorado. Por su parte, para el caso de la distancia de Pena no es evidente si existe una manera de mejorar la eficiencia con la que el indicador resumen hace uso de la información. En otras palabras, aunque el indicador sintetice la información de múltiples variables y el procedimiento univariado de estratificación sea “óptimo”, el resultado puede ser insatisfactorio⁷. En la medida en que se alcance “mejor” el propósito de un ejercicio multivariado de estratificación, será posible acceder además a una descripción detallada de la estructura del conjunto de datos que permita identificar las causas significativas de las desigualdades entre grupos e instrumentar, en su caso, acciones convenientes de política social. De aquí la importancia de disponer de procedimientos multivariados de estratificación que aprovechen toda la información disponible, y de criterios claros y con consenso que permitan identificar cuando una estratificación es mejor que otras, y en qué sentido.

2. Descripción del Estratificador INEGI 2.0

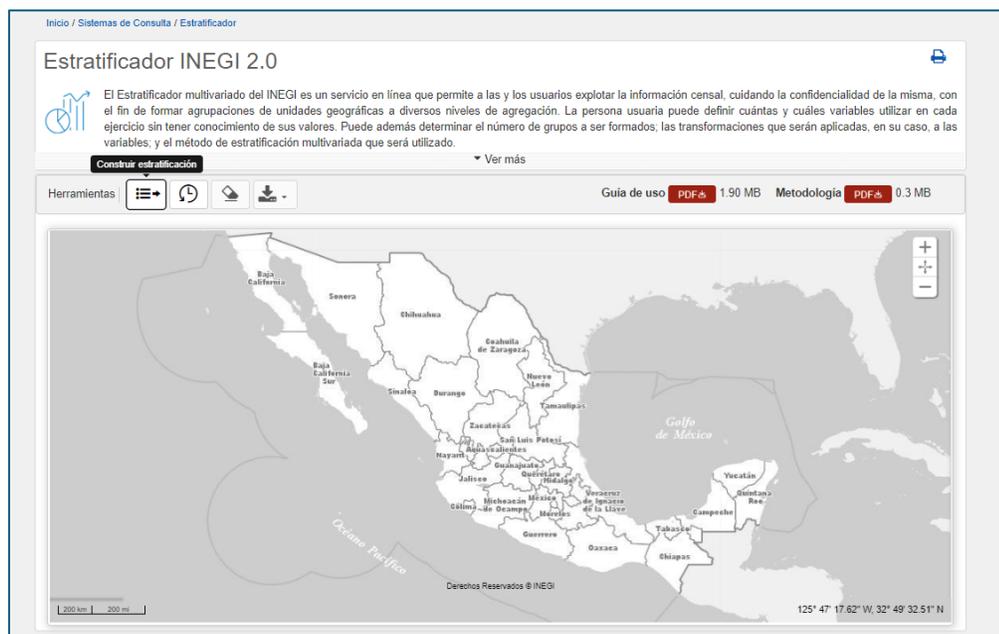
La sección de la pantalla inicial del Estratificador a la que quien hace uso del servicio deberá acudir al iniciar su trabajo, se ubica en la esquina superior izquierda de la pantalla y la identifica el botón  con la leyenda desplegable “Construir estratificación” como se aprecia en la Figura 1.

⁵ Bustos y de la Tijera, V. A., Niveles de marginación: una estrategia multivariada de clasificación, REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA, Edición: Vol.2, Núm.1, enero-abril, 2011. Puede ser consultado en <https://rde.inegi.org.mx/index.php/2011/01/03/niveles-de-marginacion-una-estrategia-multivariada-de-clasificacion/>.

⁶ CONAPO, Anexo C. Metodología de estimación del índice de marginación por localidad, 2011, http://www.conapo.gob.mx/work/models/CONAPO/indices_margina/2010/anexoc/AnexoC.pdf.

⁷ El procedimiento univariado de Dalenius-Hodges [véase Dalenius, T., & Hodges Jr, J. L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54 (285), 88-101.], que es el usado por el CONAPO y del cual no hay una versión multivariada, busca la estratificación que conducirá a la estimación muestral más precisa (o de menor varianza, y este es el sentido en el que es óptima) del promedio poblacional de la variable de estratificación. Ello cuando la selección de muestra se lleva a cabo de manera estratificada, con afijación de la muestra proporcional al tamaño de cada estrato y con selección aleatoria simple de unidades al interior de los estratos. Cabe preguntarse con qué frecuencia quienes diseñan la política social buscan la estimación óptima del promedio poblacional de la primera componente principal (o de variables altamente correlacionadas con ella) como para estar interesados en encontrar la estratificación óptima para ese fin.

FIGURA 1. PANTALLA INICIAL DEL ESTRATIFICADOR INEGI 2.0



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

Haciendo uso de sus cinco secciones, al desarrollar una consulta, se decidirán las condiciones bajo las cuales se generarán sus resultados eligiendo, para los ejercicios que serán considerados (véase Figura 2):

1. Variables de agrupación⁸: 184 indicadores, organizados en once temas expresados, en su mayoría, en términos relativos⁹.
2. El número de estratos a formar.
3. El nivel de desagregación geográfica que alcanza hasta manzanas dentro de la localidad.
4. Las transformaciones que se aplicarán a las variables, en su caso, y
5. El método de estratificación a usar de entre los tres que se incluyen. Estos son: (1) el de las k-medias; (2) el identificado como mulvar; y (3) el conocido como mclust. Más adelante se comentan sus características a mayor profundidad, para los ejercicios que serán considerados.

⁸ Variables de agrupación: conjunto de indicadores en que se basará un ejercicio de estratificación. Al pulsar en el botón “Variables” se despliega la lista de los disponibles en esta versión del Estratificador.

⁹ Debe ejercerse cuidado con la selección de variables y con la interpretación de sus resultados ya que en el conjunto incluido las hay “positivas”, para las cuales “mayor es mejor”, como “% Población de 18 a 24 años que asiste a la escuela” y “negativas”, cuando “mayor es peor”, como “% Población de 8 a 14 años que no saben leer ni escribir”.

FIGURA 2. MENÚ INICIAL PARA EL DESARROLLO DE UNA CONSULTA

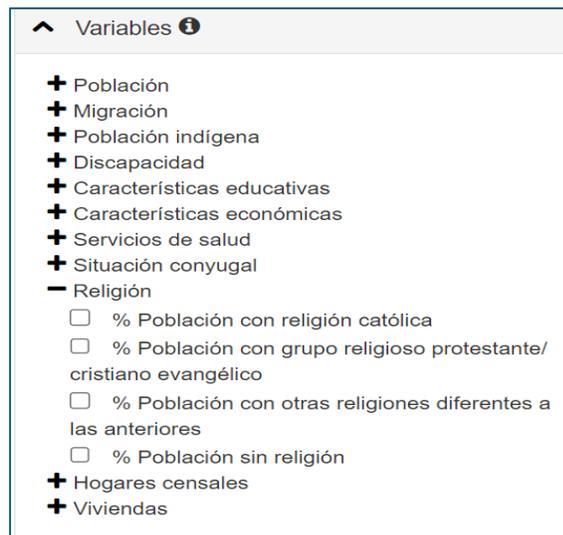


Fuente: <https://www.inegi.org.mx/app/estratificador/>.

2.1 Variables

Al seleccionar “Variables” se presenta un menú, a partir del cual se realiza la selección de tantas variables como se desee. Se ofrecen un total de 184 variables organizadas en once temas, que se muestran enseguida. La única restricción impuesta es la de seleccionar al menos dos variables para cada ejercicio. La mayoría de las variables han sido expresadas como porcentajes.

FIGURA 3. EJEMPLO DE SELECCIÓN DE VARIABLES DENTRO DEL TEMA

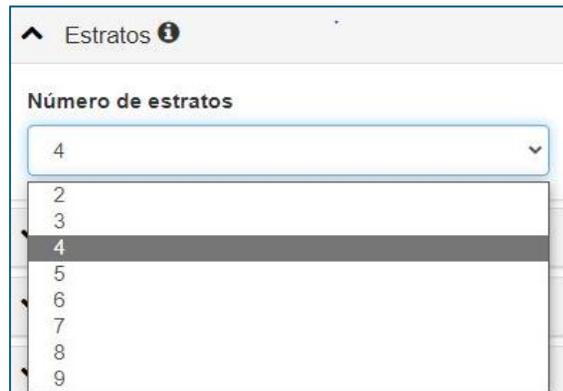


Fuente: <https://www.inegi.org.mx/app/estratificador/>.

2.2 Número de estratos

Permite seleccionar el número de grupos o estratos a formar entre la lista de opciones que se despliega. Pueden ser de dos a nueve estratos.

FIGURA 4. MENÚ DE SELECCIÓN DEL NÚMERO DE ESTRATOS A FORMAR



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

2.3 Nivel de desagregación

Acto seguido es posible seleccionar el Nivel de desagregación geográfica deseado que aparece como “Nivel de desagregación” con las opciones: Nacional (todas las entidades o todos los municipios del país), Municipios en: (todos los municipios dentro de un subconjunto de estados) o por manzanas dentro de una localidad seleccionada. En el primero de estos casos el servicio ofrecerá desagregaciones a nivel de las 32 entidades federativas o los 2 469 municipios del país. En el segundo, se podrá elegir un conjunto de entidades de la lista de estas, y cuyos municipios serán estratificados. Finalmente, al seleccionar la opción localidad, el servicio permitirá determinar la entidad, el municipio y la localidad cuyas manzanas serán estratificadas.

FIGURA 5. OPCIONES OFRECIDAS PARA SELECCIONAR EL NIVEL DE DESAGREGACIÓN GEOGRÁFICA

The figure displays three sequential screenshots of a web application interface for selecting geographical disaggregation levels. Each screenshot is titled 'Nivel de desagregación' with an information icon.

- First Screenshot:** Shows three radio button options: 'Nacional' (selected), 'Municipios en:', and 'Localidad'. Below these are two more radio button options: 'De entidades' (selected) and 'De municipios'.
- Second Screenshot:** Shows the 'Municipios en:' option selected. Below it is a list of Mexican states with checkboxes: Aguascalientes, Baja California, Baja California Sur, Campeche, Coahuila de Zaragoza, Colima, Chiapas, Chihuahua, Ciudad de México, Durango, Guanajuato, and Guerrero.
- Third Screenshot:** Shows the 'Localidad' option selected. Below it are three dropdown menus labeled 'Estado:', 'Municipio:', and 'Localidad:', each with 'Seleccione' as the placeholder text.

Fuente: <https://www.inegi.org.mx/app/estratificador/>.

2.4 Transformar datos

Esta sección concede la posibilidad de aplicar transformaciones usuales a los datos. En la primera de ellas, la estandarización, después de restar su promedio a cada valor, se divide el resultado entre la desviación típica o estándar. El nuevo conjunto de valores tendrá en adelante media igual a cero y desviación estándar igual a uno. Esta transformación debería usarse poco ya que los datos han sido expresados en porcentajes, por lo que carecen de unidades. Se ofrece además la posibilidad de construir indicadores compuestos, denominados componentes principales, cuyas propiedades (no correlación y varianzas decrecientes) permiten reducir la dimensionalidad del problema al considerar solamente algunas de las primeras¹⁰.

¹⁰ De hecho, la primera componente principal calculada a partir de variables estandarizadas fue la propuesta metodológica original del CONAPO para su índice de marginación.

FIGURA 6. SELECCIÓN OPCIONAL DE TRANSFORMACIONES DE DATOS



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

2.5 Métodos de agrupación

Finalmente, es posible decidir si se trabaja con las variables originales o con sus versiones estandarizadas y, para cada caso, con algunas o todas sus componentes principales. Además, se elegirá el procedimiento de estratificación que se aplicará enseguida con base en las elecciones realizadas hasta ahora. En este momento los métodos disponibles son:

1. k-medias¹¹
2. mulvar¹²
3. mclust^{13 14}, al seleccionar esta opción se iniciará el proceso de búsqueda de la “mejor” combinación de número de grupos y de restricciones a imponer sobre las estructuras locales de covarianzas de los datos (véase documento metodológico).

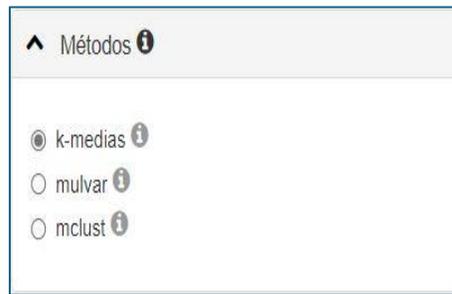
¹¹ Hartigan, J. A. (1975). Clustering algorithms. *John Wiley google schola*, 2, 25-47.

¹² Producto: <http://www.inegi.org.mx/est/contenidos/espanol/sistemas/regsoc/default.asp?s=est&c=11723>, Metodología: http://sc.inegi.org.mx/niveles/datosnbi/reg_soc_mexico.pdf

¹³ C. Fraley and A. E. Raftery, Model-based clustering, discriminant analysis, and density estimation, *Journal of the American Statistical Association*, Vol. 97, pages 611-631 (2002).

¹⁴ C. Fraley and A. E. Raftery, MCLUST Version 3 for R: Normal Mixture Modeling and Model-based Clustering, Technical Report No. 504, Department of Statistics, University of Washington, September 2006 (revised July 2010).

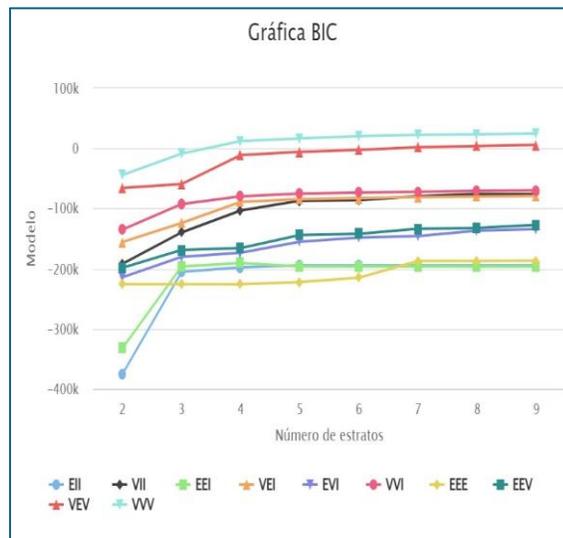
FIGURA 7. MENÚ PARA SELECCIÓN DEL MÉTODO A USAR EN LA CONSULTA



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

El sistema producirá una gráfica del criterio Bayesiano de información (BIC, por sus siglas en inglés) como se muestra en la Figura 8, que penaliza el uso de un mayor número de parámetros, y que ayudará a decidir la mejor combinación, según el mencionado criterio, para ser usada con el método mclust. El procedimiento compara combinaciones y sugiere aquellas a considerar.

FIGURA 8. CRITERIO BAYESIANO DE INFORMACIÓN PARA SELECCIÓN DEL MODELO Y NÚMERO DE ESTRATOS



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

En la sección Métodos del Estratificador es posible seleccionar el o los modelos elegidos al hacer uso del procedimiento “mclust” como se muestra enseguida. Para el caso del ejemplo, el modelo VVV supera a todos los demás sin importar el número de grupos que se considere deseable.

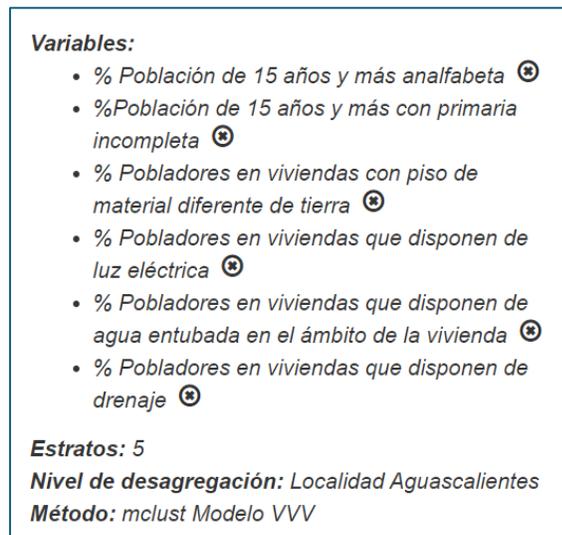
FIGURA 9. SELECCIÓN DEL MODELO PARA SEGUNDOS MOMENTOS DENTRO DE mclust



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

Una vez que se ha decidido cuáles variables usar, cuántos grupos formar, qué nivel de desagregación geográfica considerar, cuáles transformaciones aplicar, en su caso, y el método a utilizar en el primer ejercicio, solamente resta pulsar la tecla “Realizar estratificación”, localizada abajo en la sección “Construir estratificación”. En la ventana “Construir estratificación” se desplegará un resumen de las condiciones bajo las cuales se obtuvieron los resultados.

FIGURA 10. EJEMPLO DE RESUMEN DE LAS OPCIONES SELECCIONADAS



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

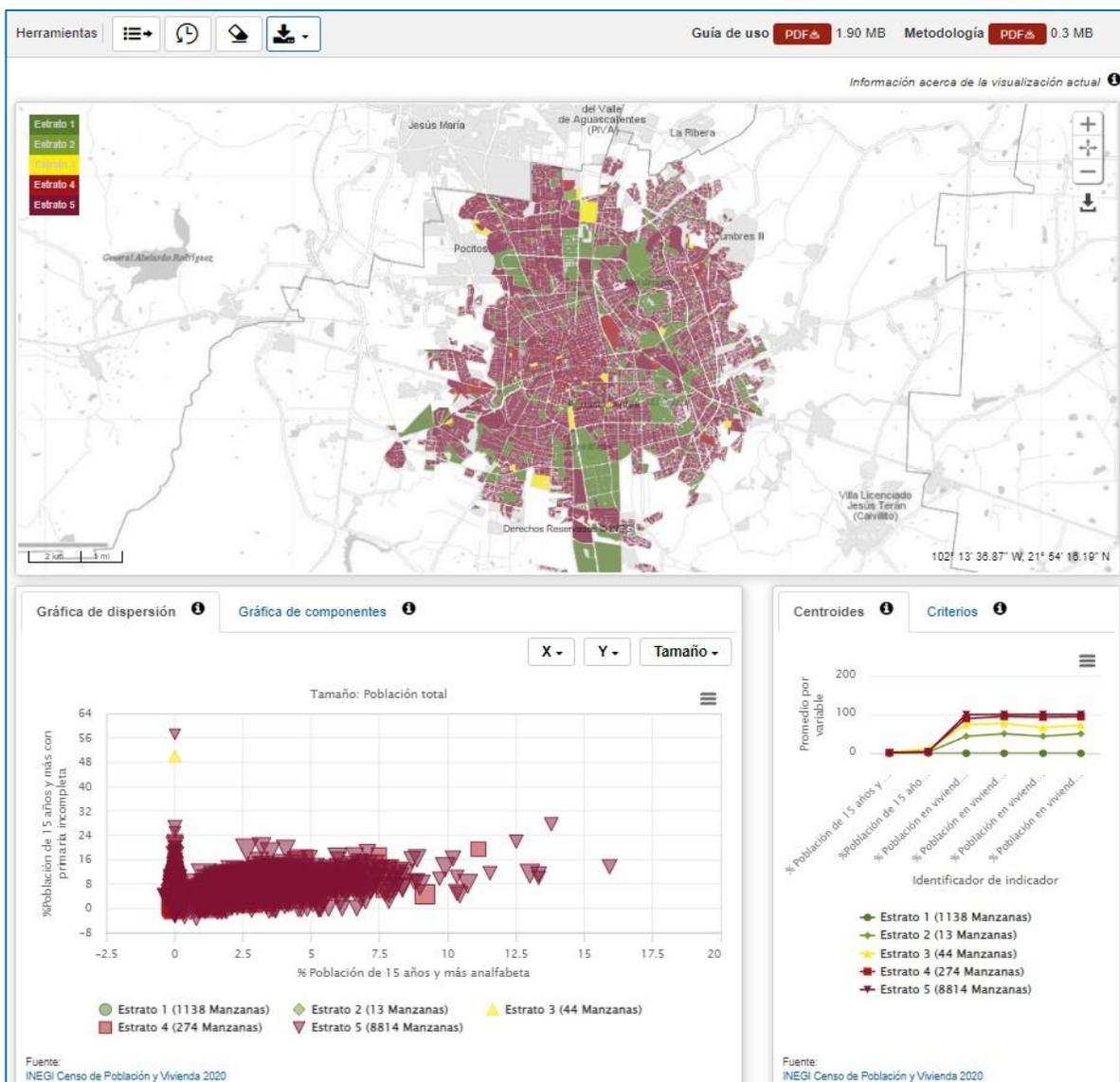
Una vez obtenidos los resultados de una estratificación, esta información podrá ser nuevamente desplegada cuando se hace “click” en el letrero “Información acerca de la visualización actual” que aparece arriba del mapa, a la derecha.

3. Resultados

En la sección anterior indicó todo lo que es necesario a manera de preparación para llevar a cabo un ejercicio de estratificación. Por ello, los resultados de cada estratificación realizada, como se indicó con anterioridad, son desplegados de manera gráfica en la pantalla del Estratificador (Figura 11), que está a su vez subdividida en tres ventanas que pueden ser maximizadas para una mejor visualización. Las gráficas de resultados pueden ser consultadas en alguna de ellas. Dichas ventanas se denominan:

1. Mapas
2. Gráfica de dispersión/Gráfica de componentes
3. Centroides/Criterios

FIGURA 11. ESTRATIFICACIÓN DE MANZANAS DE LA LOCALIDAD AGUASCALIENTES

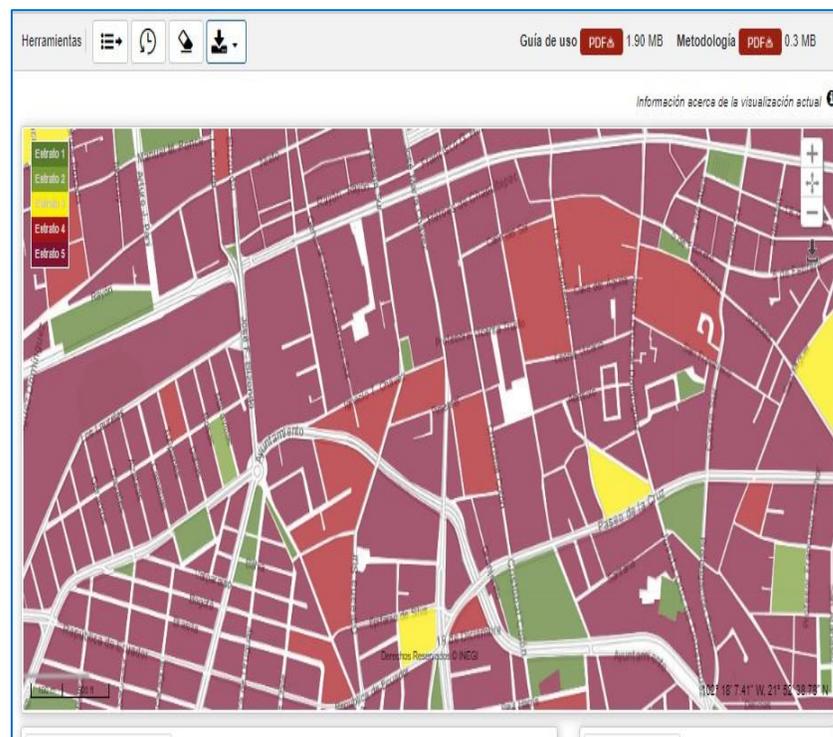


Fuente: <https://www.inegi.org.mx/app/estratificador/>.

3.1 Mapa estratificado

En primer lugar, el sistema arroja un mapa, que se despliega en la parte superior de la pantalla, con los resultados obtenidos coloreando todas y cada una de las unidades geográficas, como se muestra en la Figura 11. Todas aquellas que se muestran en el mismo color pertenecen al mismo estrato. La banda que se despliega en la esquina superior izquierda del mapa señala el color con que se identifica a cada estrato. Por ejemplo, la Figura 12 muestra el despliegue a pantalla completa del resultado de estratificar las 10 283 manzanas de la localidad Aguascalientes. Si se gira la rueda del ratón es posible lograr acercamientos o zoom in que muestren un mayor detalle.

FIGURA 12. ACERCAMIENTO A ZONA CENTRO DE LA LOCALIDAD DE AGUASCALIENTES



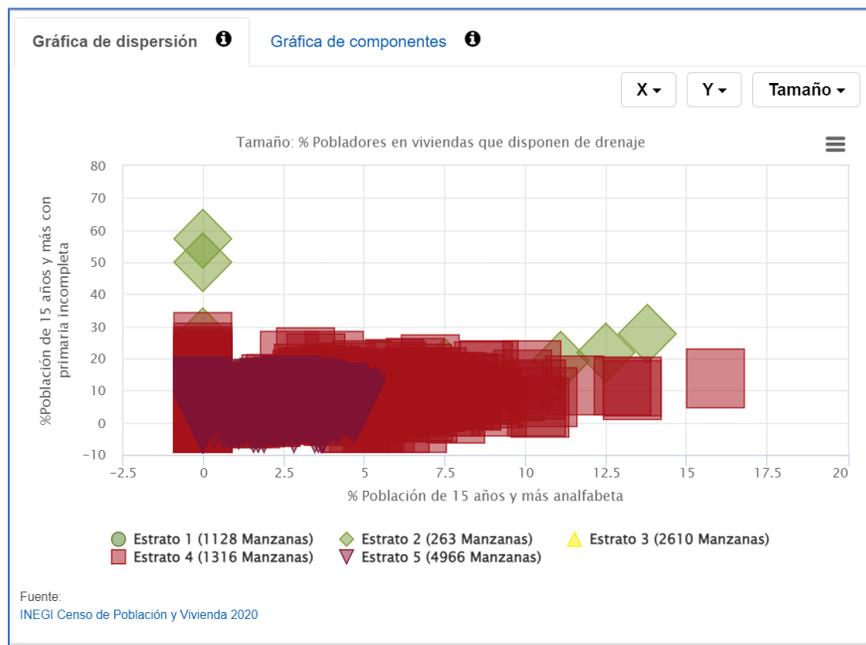
Fuente: <https://www.inegi.org.mx/app/estratificador/>.

Los colores fueron elegidos para realizar una interpretación similar a la que se haría frente a un semáforo. El estrato conformado por unidades que tengan valores menos favorables será representado por un color rojo intenso. Caso contrario, las unidades geográficas con valores más favorables que otras, serán incluidas en el estrato de color verde oscuro. Los demás estratos que no presentan valores extremos tendrán colores dentro de esa gama. Cabe hacer notar que la presencia de indicadores con sentidos opuestos puede dar lugar a otros resultados.

3.2 Gráfica de dispersión

Otra gráfica útil para la visualización de resultados es la que resume la información del ejercicio de estratificación en una proyección sobre un plano definido por dos de los indicadores y que se denomina “Gráfica de dispersión” (véase Figura 13). Esta gráfica incluye, en adición a las identificadas con los ejes horizontal (X) y vertical (Y), información sobre una 3ª dimensión a la que se representa de acuerdo con el tamaño de las burbujas; de hecho, es la única gráfica en la que se permite hacer uso de la variable “Tamaño total de la población” en esta calidad. Es posible elegir al indicador que será representado en el eje “X”, en el “Y” y la que será usada a manera de “Tamaño” en la esquina superior derecha de esta gráfica.

FIGURA 13. GRÁFICA DE DISPERSIÓN DE LA ESTRATIFICACIÓN EN AGUASCALIENTES



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

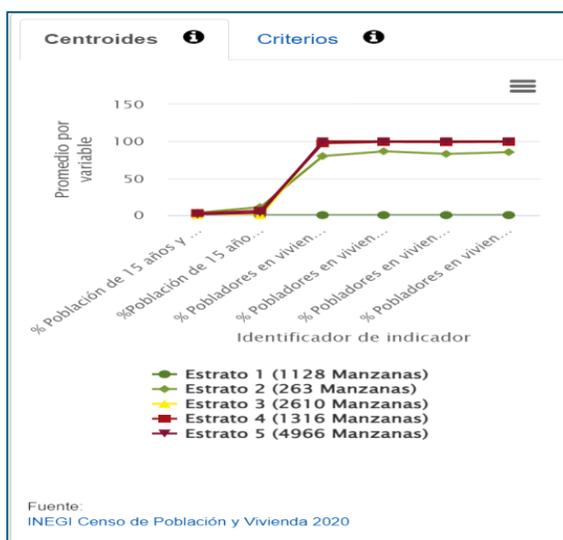
En este caso también es posible “prender” o “apagar” estratos pulsando sobre la o las barras de color que se ubican en la parte inferior de la gráfica. De este modo, por ejemplo, sería posible aislar a las 263 manzanas de la localidad de Aguascalientes ubicadas en el segundo estrato. Al pasar el cursor sobre cada punto, se despliegan tanto la identificación numérica de la unidad según el INEGI, como los valores de las variables en la misma unidad.

Para evitar poner en riesgo la privacidad de las personas informantes, cuando los valores de una variable asociadas a alguna de las áreas sean muy pequeños no se desplegará la figura correspondiente.

3.3 Gráfica de centroides

En la gráfica denominada “centroides”, las líneas del color que identifica a cada estrato unen los valores promedio de cada indicador dentro del estrato; en general, se observará dominancia casi perfecta de uno de los estratos sobre uno o más. Con el fin de permitir una más sencilla interpretación de esta gráfica, las líneas quebradas correspondientes a uno o más estratos pueden ser “apagadas” con solo pulsar sobre el color correspondiente. Se muestra para el ejemplo en la figura 14. Además de la jerarquización de los estratos, esta gráfica permite identificar a los indicadores más importantes en la discriminación entre grupos mediante valores promedio que exhiben diferencias importantes; a diferencia de lo que ocurre en otros casos en los cuales las diferencias entre grupos son imperceptibles. Los indicadores cuyos promedios muestran valores más heterogéneos contribuyen de manera más importante en la definición de los estratos. Cuando estos valores son casi coincidentes para todos los estratos debe tenerse que los indicadores correspondientes no hacen una contribución significativa al mismo fin por lo que podrían ser ignorados para ejercicios subsecuentes. Por su parte, los casos intermedios permiten identificar aquellas dimensiones que hacen diferentes a dos o más grupos y, en consecuencia, permiten asimismo identificar los temas y, de ese modo, la naturaleza de los esfuerzos que deben ser realizados para reducir la desigualdad entre ellos.

FIGURA 14. CENTROIDES DE LA ESTRATIFICACIÓN EN AGUASCALIENTES



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

3.4 Gráfica de componentes principales

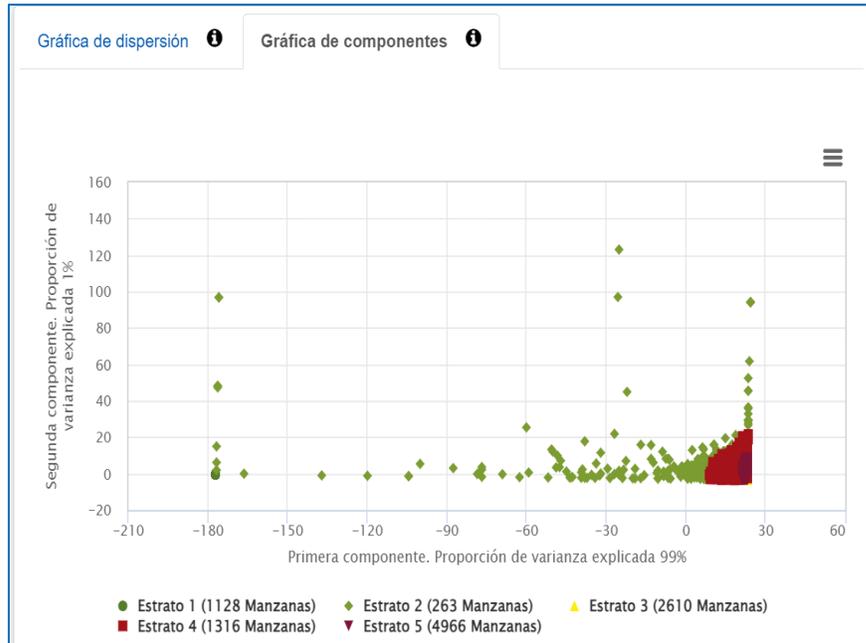
Por supuesto, dependiendo del número de indicadores seleccionados, el número de gráficas de dispersión que es posible elaborar puede ser excesivamente grande. Por ello, y con el fin de mostrar una representación “canónica”, se incluye además la gráfica de dispersión cuando en el eje horizontal se representa a la primera componente principal (CP)¹⁵ y en el vertical a la segunda. En cada caso, se señala además la proporción de la varianza generalizada que cada componente representa.

Para el ejemplo, existe una gran redundancia entre los indicadores seleccionados, la que se expresaría como una correlación significativa entre parejas de ellos. En consecuencia, la primera componente principal explica

¹⁵ Semejante al índice de marginación original de CONAPO.

99 % de la varianza generalizada (i. e., la suma de las varianzas de los indicadores), por lo que las siguientes componentes principales pueden ser ignoradas.

FIGURA 15. PRIMERAS DOS COMPONENTES PRINCIPALES DE LA ESTRATIFICACIÓN EN AGUASCALIENTES



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

4. Múltiples ejercicios durante la misma sesión

Como se ha señalado, es posible realizar durante una misma sesión diversos ejercicios de estratificación., Es posible modificar las condiciones de los ejercicios anteriores para obtener una nueva estratificación. Tanto las condiciones como los resultados de cada ejercicio son almacenados de manera que puedan ser consultados con posterioridad durante la sesión. Para navegar entre todas las distintas estratificaciones realizadas en la misma sesión se incluye el botón Historial  de la sección “Herramientas”, en la parte superior izquierda de la pantalla del Estratificador.

Al pulsar cada una de las anteriores opciones se desplegarán las condiciones bajo las cuales se obtuvieron los resultados, como se muestra en la Figura 16. Al pulsar el botón  se desplegarán los mapas y las gráficas almacenadas relativas a ese ejercicio.

FIGURA 16. CONDICIONES BAJO LAS CUALES SE REALIZARON CUATRO EJERCICIOS EN LA MISMA SESIÓN

<p>^ Estratificación 1</p> <p>Variables:</p> <ul style="list-style-type: none"> • % Población de 15 años y más analfabeta • %Población de 15 años y más con primaria incompleta • % Pobladores en viviendas con piso de material diferente de tierra • % Pobladores en viviendas que disponen de luz eléctrica • % Pobladores en viviendas que disponen de agua entubada en el ámbito de la vivienda • % Pobladores en viviendas que disponen de drenaje <p>Estratos: 5</p> <p>Nivel de desagregación: Localidad Aguascalientes</p> <p>Método: kmedias Componentes principales (1)</p> <p style="text-align: right;">Ver</p>	<p>^ Estratificación 2</p> <p>Variables:</p> <ul style="list-style-type: none"> • % Población de 15 años y más analfabeta • %Población de 15 años y más con primaria incompleta • % Pobladores en viviendas con piso de material diferente de tierra • % Pobladores en viviendas que disponen de luz eléctrica • % Pobladores en viviendas que disponen de agua entubada en el ámbito de la vivienda • % Pobladores en viviendas que disponen de drenaje <p>Estratos: 5</p> <p>Nivel de desagregación: Localidad Aguascalientes</p> <p>Método: kmedias Componentes principales (3)</p> <p style="text-align: right;">Ver</p>
<p>^ Estratificación 3</p> <p>Variables:</p> <ul style="list-style-type: none"> • % Población de 15 años y más analfabeta • %Población de 15 años y más con primaria incompleta • % Pobladores en viviendas con piso de material diferente de tierra • % Pobladores en viviendas que disponen de luz eléctrica • % Pobladores en viviendas que disponen de agua entubada en el ámbito de la vivienda • % Pobladores en viviendas que disponen de drenaje <p>Estratos: 5</p> <p>Nivel de desagregación: Localidad Aguascalientes</p> <p>Método: kmedias Componentes principales (6)</p> <p style="text-align: right;">Ver</p>	<p>^ Estratificación 4</p> <p>Variables:</p> <ul style="list-style-type: none"> • % Población de 15 años y más analfabeta • %Población de 15 años y más con primaria incompleta • % Pobladores en viviendas con piso de material diferente de tierra • % Pobladores en viviendas que disponen de luz eléctrica • % Pobladores en viviendas que disponen de agua entubada en el ámbito de la vivienda • % Pobladores en viviendas que disponen de drenaje <p>Estratos: 5</p> <p>Nivel de desagregación: Localidad Aguascalientes</p> <p>Método: mclust Modelo VEV Componentes principales (6)</p> <p style="text-align: right;">Ver</p>

Fuente: <https://www.inegi.org.mx/app/estratificador/>.

Resta ahora decidir si el resultado obtenido es satisfactorio para sacar conclusiones y basar en ellas diversas decisiones. En la Tabla 1 se presentan los tamaños de cada uno de los cinco estratos para cada uno de los cuatro ejercicios de estratificación. Como puede observarse el “Estrato 1” muestra pequeñas diferencias entre los ejercicios, en tanto que los restantes estratos son más diversos. Entre los cuatro ejercicios de estratificación, el segundo y el tercero resultan muy parecidos, exhibiendo solamente diferencias menores. Las estratificaciones tres y cuatro fueron obtenidas a partir de las seis componentes principales que se pueden calcular a partir del mismo número de variables originales. Cabría esperar que mostrarán semejanzas, pero resultan ser diferentes.

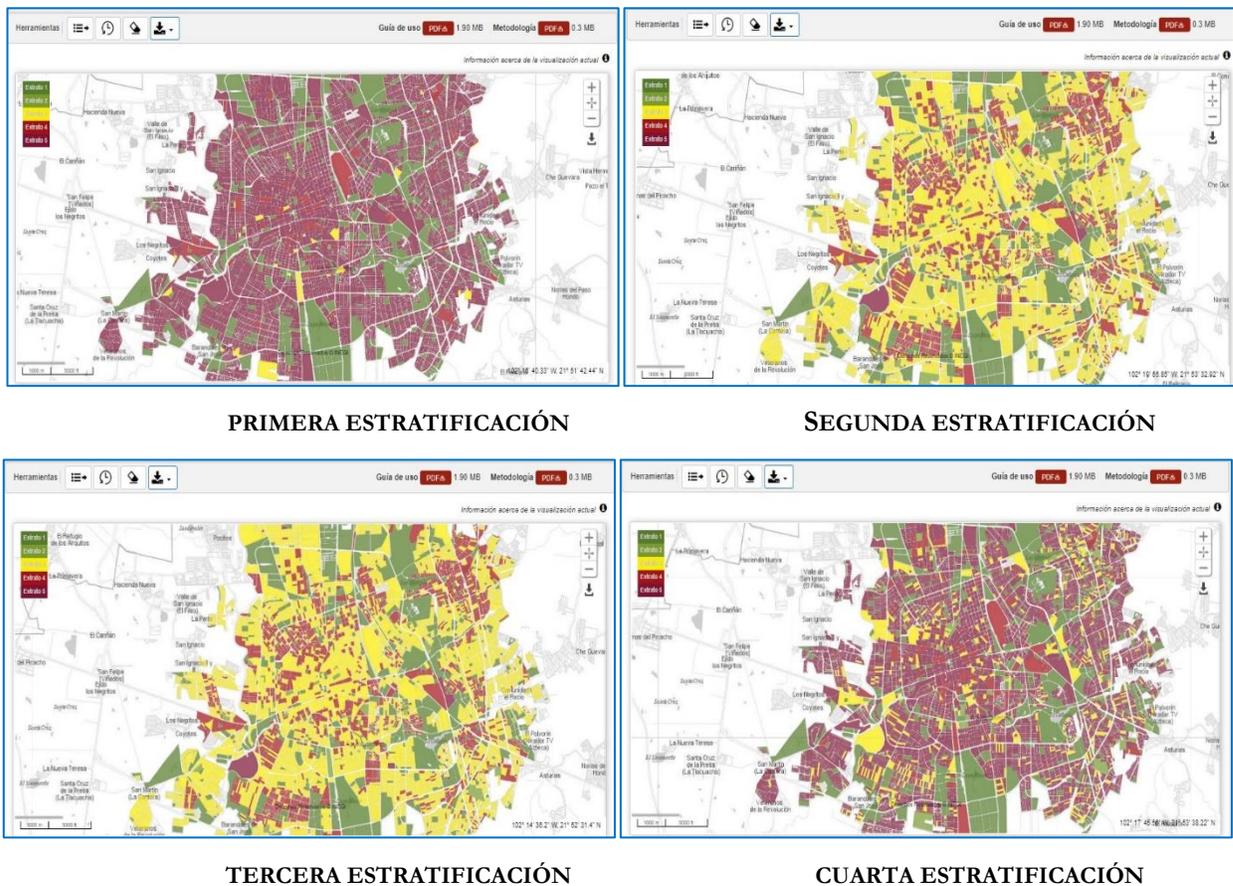
TABLA 1. FRECUENCIAS ABSOLUTAS POR ESTRATO Y ESTRATIFICACIÓN

		Estrato				
		1	2	3	4	5
Estratificación	1	1138	13	44	274	8814
	2	1139	60	6912	2162	10
	3	1139	60	6915	2158	11
	4	1128	442	751	556	7406

Elaboración propia. Fuente: <https://www.inegi.org.mx/app/estratificador/>.

Los elementos gráficos y cartográficos no son suficientes para determinar fácilmente si el resultado de alguno de los ejercicios es mejor que los otros en algún sentido. La Figura 17 se refiere a los resultados gráficos de cuatro ejercicios de estratificación de la localidad de Aguascalientes. A partir de los mapas, en la primera de ellas, no resulta sencillo realizar una evaluación de los resultados alternativos. Nuevamente, entre la segunda y la tercera estratificación es aparente una importante consistencia, pero también una diferencia importante con las otras dos. De hecho, entre estas no resulta evidente a qué se deben sus diferencias, ni si alguna es más adecuada que la otra.

FIGURA 17. MAPAS PARA DIFERENTES TIPOS DE ESTRATIFICACIONES EN AGUASCALIENTES



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

En la Figura 18, que presenta las gráficas de centroides, se posible encontrar alguna evidencia que permita establecer si los objetivos se cumplen de mejor manera en alguno de los casos. Se puede observar que los valores

promedio para las variables % Población de 15 años y más analfabeta y % Población de 15 años y más con primaria incompleta se encuentran muy cercanos para todas las estratificaciones. Los correspondientes a las restantes variables muestran en cambio notorias diferencias y un orden claro para la primera estratificación. En contraste se tiene que para la cuarta estratificación los promedios de todas las variables para cuatro de los cinco grupos son casi indistinguibles. En este caso se puede dar una importante confusión entre estos grupos lo que lleva a sospechar que no son tan heterogéneos como es deseable. Ello hará difícil discernir entre estos. Para la segunda y la tercera estratificación parece estar en presencia de solamente tres grupos adecuadamente diferentes.

FIGURA 18. MAPAS PARA DIFERENTES TIPOS DE ESTRATIFICACIONES EN AGUASCALIENTES



PRIMERA ESTRATIFICACIÓN

SEGUNDA ESTRATIFICACIÓN



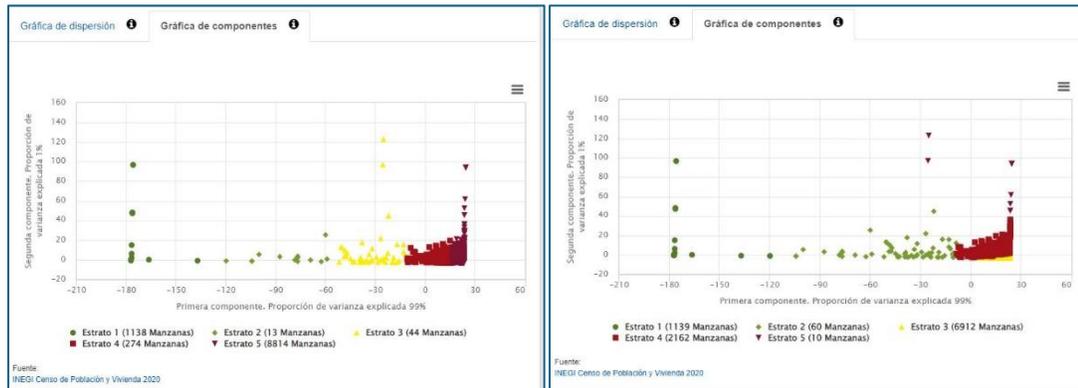
TERCERA ESTRATIFICACIÓN

CUARTA ESTRATIFICACIÓN

Fuente: <https://www.inegi.org.mx/app/estratificador/>

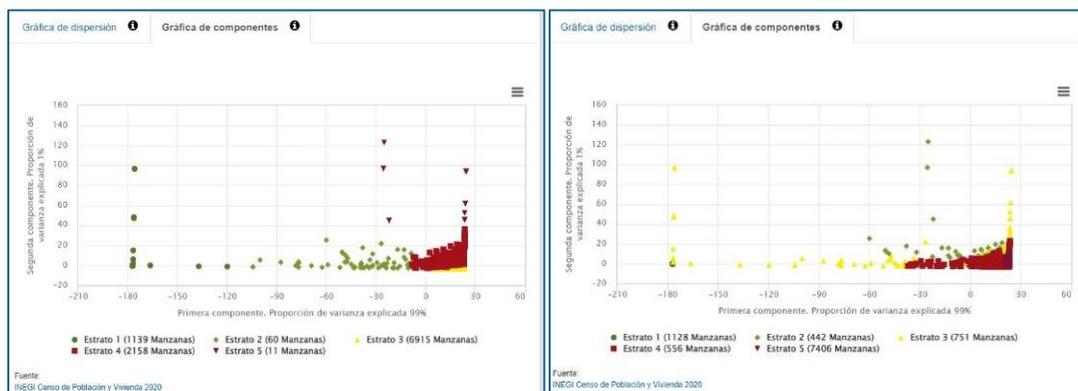
Por su parte, la Figura 19 muestra las cuatro gráficas de componentes principales. Salta inmediatamente a la vista el hecho de que la primera componente principal tiene una varianza equivalente a 99 % de la suma de las varianzas de los indicadores originales, en tanto que la segunda explica casi el 1 % restante. En otras palabras, las otras cuatro componentes principales tienen todas varianzas muy cercanas a cero. Lo anterior implica que existe una gran redundancia entre los indicadores seleccionados, misma que se manifiesta en correlaciones cercanas a uno entre cada pareja de ellos, lo que no es deseable pues el comportamiento de todas las variables en la población es semejante. Sin embargo, esta circunstancia será útil con fines ilustrativos. Como puede observarse en la gráfica correspondiente a la primera estratificación las unidades son clasificadas en grupos a lo largo de los valores de la primera componente principal. Estos grupos son distinguibles entre sí, ratificando lo obtenido a través de la gráfica de centroides. En cambio, las restantes gráficas exhiben grupos que se superponen haciéndolos indistinguibles en uno o más casos. Los grupos así formados contradicen el propósito de contar con grupos tan heterogéneos como sea posible. Bajo otras circunstancias, podría pensarse que de incorporarse una tercera dimensión los grupos exhibirían una separación a lo largo de esta. Sin embargo, ya que para nuestro caso las varianzas de las restantes componentes principales son cercanas a cero, es prácticamente imposible que eso ocurra pues se estará siempre muy cerca del plano definido por las dos primeras.

FIGURA 19. GRÁFICAS DE COMPONENTES PARA DIFERENTES TIPOS DE ESTRATIFICACIONES EN AGUASCALIENTES



PRIMERA ESTRATIFICACIÓN

SEGUNDA ESTRATIFICACIÓN



TERCERA ESTRATIFICACIÓN

CUARTA ESTRATIFICACIÓN

Fuente: <https://www.inegi.org.mx/app/estratificador/>

5. Criterios estadísticos para llevar a cabo comparaciones entre ejercicios alternativos

Una vez realizados varios ejercicios durante la misma sesión, llega el momento de comparar los resultados de todos ellos con el propósito de determinar cuál es el mejor, en algún sentido. Con el fin de auxiliar a quien hace uso del Estratificador, se incluye un conjunto de seis criterios propuestos en la literatura, que permiten evaluar si las unidades dentro de un estrato son más homogéneas, en tanto que los estratos son más heterogéneos. Sus definiciones se presentan enseguida:

TABLA 2. CRITERIOS PROPUESTOS POR LA LITERATURA

Id	Definición	Criterio
A	Suma de cuadrados de las distancias Euclidianas de cada punto al centroide de su grupo [Ward (1963)] ¹⁶	$tr(\mathbf{W}) = \sum_{i=1}^G tr(\mathbf{W}_j)$
B	Determinante de la suma de las matrices locales de sumas de cuadrados y de productos cruzados [Friedman et al. (1967)] ¹⁷	$ \mathbf{W} $
C	Suma ponderada del logaritmo de trazas de matrices locales de covarianzas [Banfield et al. (1993)] ¹⁸	$\sum_{i=1}^G n_i \log \left(tr \left(\frac{\mathbf{W}_i}{n_i} \right) \right)$
D	Suma ponderada del logaritmo de determinantes de matrices locales de covarianzas [Scott et al. (1971)] ¹⁹	$\sum_{i=1}^G n_i \log \left(\left \frac{\mathbf{W}_i}{n_i} \right \right)$
E	Promedio ponderado de las relaciones entre determinantes locales y globales	$\sum_{i=1}^G n_i \frac{ \mathbf{W}_i }{n_i} \frac{1}{ \Sigma }$
F	Cociente de la inercia dentro de los estratos a la inercia entre los estratos (Véase apéndice metodológico)	$\frac{tr(\mathbf{W})}{tr(\mathbf{B})}$

Elaboración propia. Fuente: <https://www.inegi.org.mx/app/estratificador/>.

Para las definiciones y el cálculo²⁰ de las matrices consideradas en los anteriores criterios, se puede demostrar la siguiente descomposición de la suma de cuadrados y productos cruzados de la muestra con respecto al vector de medias muestrales:

$$\begin{aligned} \sum_{j=1}^G \sum_{i=1}^{n_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}})(\mathbf{x}_{ij} - \bar{\mathbf{x}})^t &= \sum_{j=1}^G \sum_{i=1}^{n_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)^t + \sum_{j=1}^G n_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^t \\ &= \sum_{j=1}^G \mathbf{W}_j + \sum_{j=1}^G n_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^t = \mathbf{W} + \mathbf{B}, \end{aligned}$$

¹⁶ Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58 (301), 236-244.

¹⁷ Friedman, H. P., & Rubin, J. (1967). On some invariant criteria for grouping data. *Journal of the American Statistical Association*, 62 (320), 1159-1178.

¹⁸ Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian Clustering. *Biometrics*, 803-821.

¹⁹ Scott, A. J., & Symons, M. J. (1971). Clustering methods based on likelihood ratio criteria. *Biometrics*, 387-397.

²⁰ Expresiones matriciales alternativas de cálculo y más fácilmente programables en R están dadas en el Documento Metodológico.

donde

$n = n_1 + n_2 + \dots + n_G$ es el tamaño de la muestra y n_j el del j -ésimo estrato, $j = 1, \dots, G$.

\mathbf{x}_{ij} es el i -ésimo, $i = 1, \dots, n_j$, vector de observaciones, de dimensión $k \times 1$, dentro del j -ésimo estrato, $j = 1, \dots, G$.

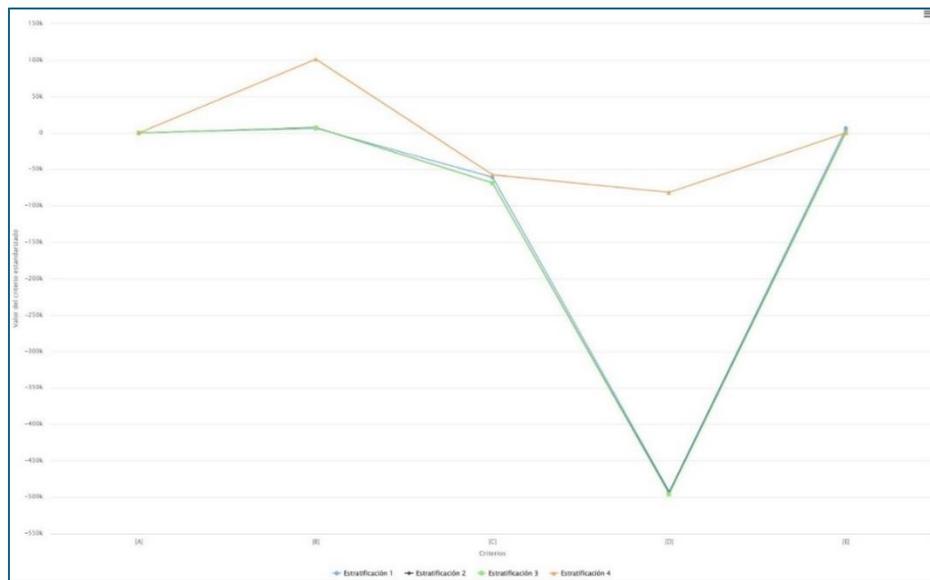
$\bar{\mathbf{x}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbf{x}_{ij}$ es el vector promedio, o centroide, del j -ésimo estrato.

$\bar{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^G \sum_{i=1}^{n_j} \mathbf{x}_{ij}$ es el vector promedio, o centroide, muestral.

Finalmente, se define $\Sigma_j = \frac{1}{n_j} W_j$.

Los valores de los mencionados criterios son desplegados en la gráfica que aparece en la pestaña “Criterios”. En ella se observan líneas quebradas de tantos colores como ejercicios a comparar. En su eje vertical se despliegan los valores alcanzados por los criterios; en el horizontal, se identifica a cada uno de los criterios. Como en los casos anteriores, las líneas pueden ser resaltadas o apagadas al ubicar el cursor sobre una de ellas.

FIGURA 20. GRÁFICA DE CRITERIOS PARA COMPARACIÓN ENTRE ESTRATIFICACIONES



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

Los valores de cada criterio para cada estratificación son desplegados al colocar el cursor sobre el punto correspondiente. Para nuestro ejemplo numérico, se muestran en la Tabla 3. Se destaca que en todos los casos los valores menores son más deseables. De este modo, el primer ejercicio da el mejor resultado de acuerdo con el criterio B (resaltado en color verde), pero el peor según el E (en anaranjado). Los dos ejercicios posteriores dan resultados parecidos para casi todos los criterios, pero no para el E. Finalmente, la cuarta estratificación resulta en los peores valores de cinco criterios, pero el mejor para el quinto. Es claro que no existe unanimidad y que ha de recurrirse a reglas de decisión tales como mayoría de votos, que en este caso resultaría en la elección del tercer ejercicio, o la elección de uno de los criterios como el preferido de la persona que lleva a cabo el análisis, con lo que se seleccionaría el del menor valor para ese criterio; por ejemplo, el segundo o el tercero para el A, el primero para el B, el tercero para C y D, el cuarto para el E, y el tercero para el F.

TABLA 3. VALORES DE LOS CRITERIOS PARA CUATRO ESTRATIFICACIONES

		Criterio					
		A	B	C	D	E	F
Estratificación	1	39	6284	-60743	-492235	6355	0.0097
	2	33	7390	-68752	-494519	1500	0.0082
	3	33	7163	-68887	-496152	400	0.0082
	4	99	101244	-57343	-81618	71	0.0250

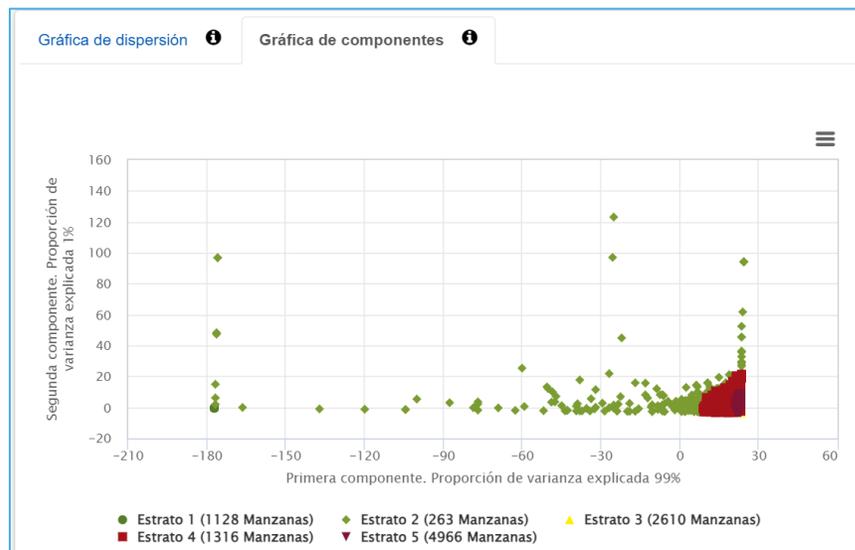
Elaboración propia. Fuente: <https://www.inegi.org.mx/app/estratificador/>

6. Exportación de resultados

6.1 Mapas

Por supuesto, dependiendo del número de indicadores seleccionados, el número de gráficas de dispersión que es posible elaborar puede ser excesivamente grande. Por ello, y con el fin de mostrar una representación “canónica”, se incluye además la gráfica de dispersión cuando en el eje horizontal se representa a la primera componente principal (CP)²¹ y en el vertical a la segunda. En cada caso, se señala además la proporción de la varianza generalizada que cada componente representa.

FIGURA 15. PRIMERAS DOS COMPONENTES PRINCIPALES DE LA ESTRATIFICACIÓN EN AGUASCALIENTES



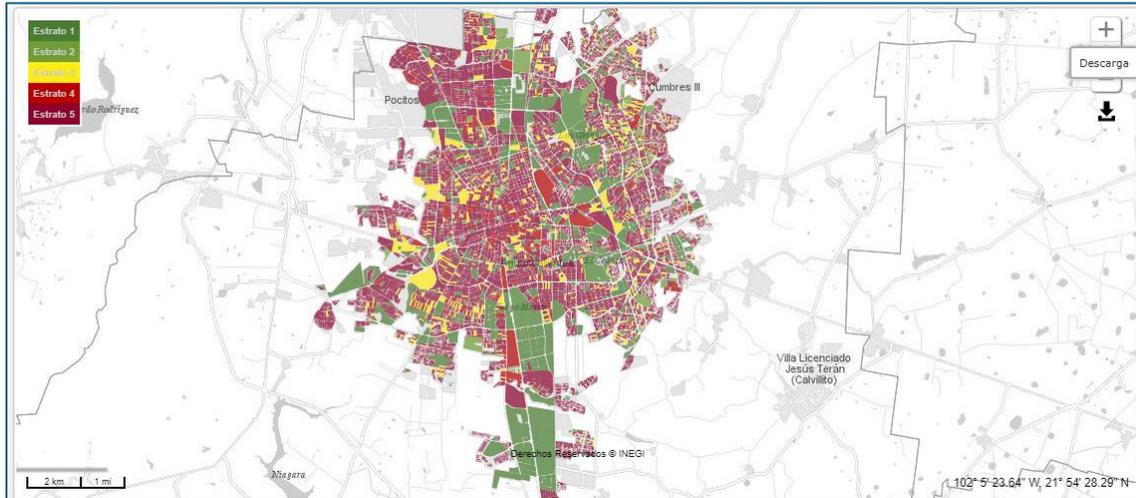
Fuente: <https://www.inegi.org.mx/app/estratificador/>

Para el ejemplo, existe una gran redundancia entre los indicadores seleccionados, la que se expresaría como una correlación significativa entre parejas de ellos. En consecuencia, la primera componente principal explica 99 % de la varianza generalizada (i. e., la suma de las varianzas de los indicadores), por lo que las siguientes componentes principales pueden ser ignoradas.

²¹ Semejante al índice de marginación original de CONAPO.

Con el fin de permitir la utilización de los resultados obtenidos a quienes hacen uso del Estratificador en sus investigaciones y reportes se cuenta con la posibilidad de exportar tanto los resultados gráficos como los numéricos. Entre los primeros se encuentran los mapas, así como las gráficas diferentes. Con respecto a aquellos basta con llevar el cursor a la flecha descendente que aparece a la derecha de cualquiera de los mapas y pulsar sobre ella (véase Fig. 21). El sistema ofrecerá la descarga de un archivo en formato de imagen PNG.

FIGURA 21. ILUSTRACIÓN PARA DESCARGA DE MAPAS

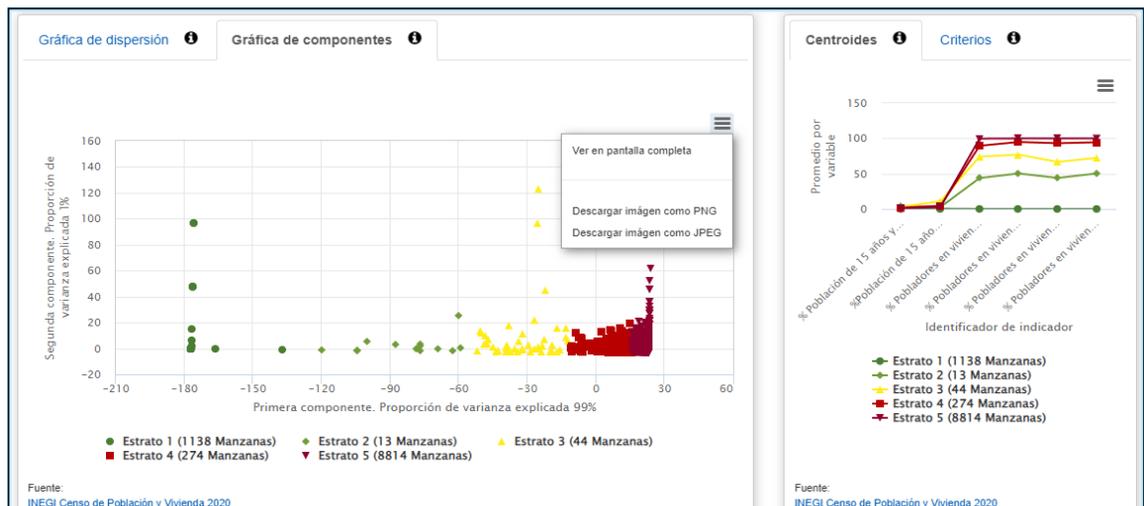


Fuente: <https://www.inegi.org.mx/app/estratificador/>.

6.2 Gráficas

Por su parte, las gráficas de dispersión, de componentes principales, de centroides y de criterios pueden ser descargadas tanto en formato JPEG como en formato PNG, opciones ofrecidas por el servicio al pulsar sobre el botón con tres barras que se encuentra arriba y a la derecha de cada una de las imágenes.

FIGURA 22. ILUSTRACIÓN PARA DESCARGA DE GRÁFICAS



Fuente: <https://www.inegi.org.mx/app/estratificador/>.

6.3 Resultados numéricos

Finalmente, también es posible exportar un archivo en formato Excel que consta de tres pestañas. La primera de ellas, denominada configuración, resume las condiciones bajo las cuales se obtuvieron los resultados; es decir, las variables seleccionadas, el número de grupos, el nivel de desagregación geográfica y el método utilizado. Por ejemplo, en la pestaña configuración se muestra la siguiente información:

TABLA 4. RESULTADO OBTENIDO EN LA PESTAÑA “CONFIGURACIÓN” DEL ARCHIVO XLS DESCARGABLE

Variables:	% Población de 15 años y más analfabeta								
	%Población de 15 años y más con primaria incompleta								
	% Pobladores en viviendas con piso de material diferente de tierra								
	% Pobladores en viviendas que disponen de luz eléctrica								
	% Pobladores en viviendas que disponen de agua entubada en el ámbito de la vivienda								
	% Pobladores en viviendas que disponen de drenaje								
Grupos:	5								
Nivel de desagregación:	Localidad _Aguascalientes								
Método:	kmedias Componentes principales (1)								

Fuente: <https://www.inegi.org.mx/app/estratificador/>.

La segunda, denominada resultados, muestra un conjunto de columnas las primeras de las cuales identifican a cada una de las unidades geográficas clasificadas en el ejercicio a través de la clave geográfica asignada por INEGI; la última contiene las etiquetas numéricas que identifican el estrato al cual pertenece cada una de las unidades. Ejemplo:

TABLA 5. RESULTADO OBTENIDO EN LA PESTAÑA “RESULTADO” DEL ARCHIVO XLS DESCARGABLE

Clave Geográfica	Estrato
100100010017001	5
100100010017002	5
100100010017003	5
100100010017004	5
100100010017005	5
100100010017006	5
100100010017007	5
100100010017008	5
100100010017009	5
100100010017010	5
100100010017011	5
100100010017012	5
100100010017013	5
100100010017014	5
100100010017015	5
100100010017016	5
100100010017017	5
100100010017018	1
100100010017019	1
100100010017020	5

Fuente: <https://www.inegi.org.mx/app/estratificador/>.

La tercera pestaña denominada “promedios” muestra, como su nombre lo indica, los valores promedio de cada una de las variables del ejercicio dentro de cada uno de los estratos.

**TABLA 6. RESULTADO OBTENIDO EN LA PESTAÑA “PROMEDIOS” DEL ARCHIVO XLS
DESCARGABLE**

Estratificador - promedios por variable y estrato							
	% Población de 15 años y más analfabeta	% Población de 15 años y más con primaria incompleta	% Pobladores en viviendas con piso de material diferente de tierra	% Pobladores en viviendas que disponen de luz eléctrica	% Pobladores en viviendas que disponen de agua entubada	% Pobladores en viviendas que disponen de drenaje	Total de unidades geográficas
Estrato 1	0.0483	0.3187	0.0222	0.0222	0.0222	0.0222	1138
Estrato 2	3.8163	2.816	43.8004	50.2798	43.8868	50.2971	13
Estrato 3	3.0434	10.9957	74.1529	76.9272	66.6459	72.3324	44
Estrato 4	1.8747	4.4853	89.4685	94.8344	93.1099	94.4195	274
	0.944	2.88	99.6396	99.9076	99.8872	99.9379	8814

Fuente: <https://www.inegi.org.mx/app/estratificador/>.