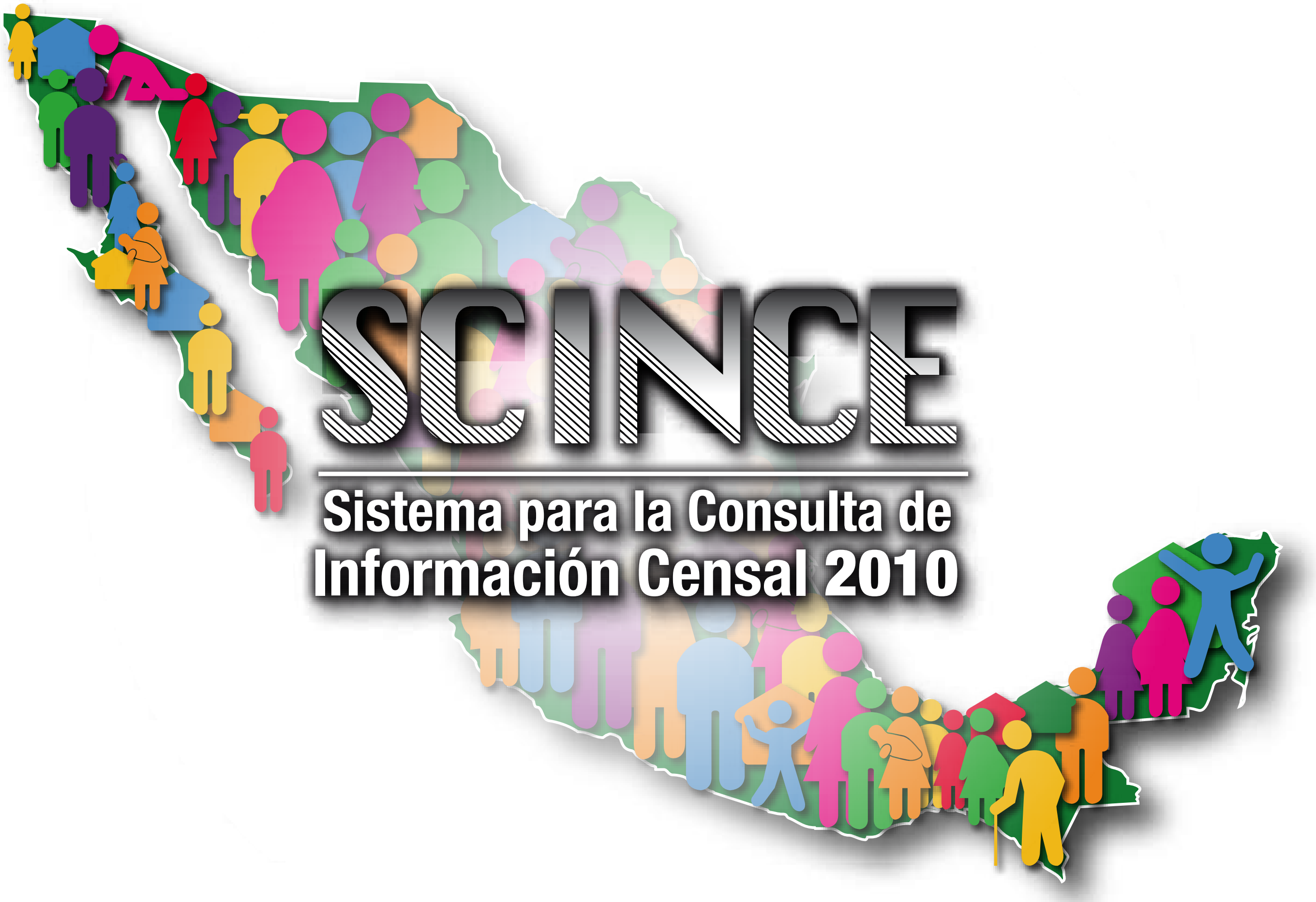
Instituto Nacional de Estadística y Geografía

Nota técnica

Estratificación univariada



Censo de Población y Vivienda 2010

Censo de Población y Vivienda 2010

INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

NOTA TÉCNICA
ESTRATIFICACIÓN UNIVARIADA

Con la finalidad de que el usuario pueda hacer clasificaciones de las unidades geográficas del país utilizando una sola variable, se ha incorporado al Sistema para la Consulta de Información Censal 2010 (SCINCE 2010) una herramienta de estratificación univariada. Es importante que el usuario analice los resultados de la estratificación cuidadosamente antes de utilizar la clasificación obtenida.

1. Cuantiles

Los cuantiles son medidas de posición que definen puntos de corte en una distribución ordenada de datos, de manera que cada estrato definido por estos puntos de corte contenga la misma proporción de valores.

Entre los cuantiles más comunes se encuentran, por ejemplo, los cuartiles que corresponden a los puntos de corte obtenidos al dividir los datos en cuatro partes, cada una equivalente a un veinticinco por ciento.

Dado un conjunto de n observaciones de una variable $x = \{x_1, x_2 ... x_n\}$, el procedimiento para formar h estratos a partir de estas observaciones sería el siguiente:

1. Ordenar los datos de forma ascendente.
2. Obtener los porcentajes requeridos para obtener los puntos de corte en función de los h estratos que se quieren formar:

    - $Q_1 = \frac{1}{h} * 100$
    - $Q_2 = 2\frac{1}{h} * 100$
    - …
    - $Q_{h-1} = h-1\frac{1}{h} * 100$

3. Obtener el porcentaje que cada elemento representa dentro del conjunto de datos, que para la i-ésima observación ordenada está dado por $S_i = \frac{1}{n+1} * 100$. Como puede observarse en la fórmula anterior, en este caso, todas las observaciones representan el mismo porcentaje, independientemente del valor de la variable. Se obtiene también para cada observación su correspondiente porcentaje acumulado que está dado por: $\sum_{k=1}^{i} S_k$.

Observaciones ordenadas	% (S_i)	% Acumulado ($\sum_{k=1}^{i} S_k$)
$x_{(1)}$	$\frac{1}{n+1} * 100$	$\frac{1}{n+1} * 100$
$x_{(2)}$	$\frac{1}{n+1} * 100$	$\frac{2}{n+1} * 100$
…	….	….
$x_{(n)}$	$\frac{1}{n+1} * 100$	$\frac{n}{n+1} * 100$

4. Los puntos de corte para cada estrato (cuantiles) se definen de la siguiente manera:

   - $\xi_1 = \max\{x_{(i)}\}\ tal\ que\ \sum S_i = Q_1$
   - $\xi_2 = \max\{x_{(i)}\}\ tal\ que\ \sum S_i = Q_2$
   - …
   - $\xi_{h-1} = \max\{x_{(i)}\}\ tal\ que\ \sum S_i = Q_{h-1}$

   En caso de no tener una observación que cumpla con la condición $\sum S_i = Q_k$, se realiza una interpolación lineal entre las dos observaciones más cercanas.

5. Finalmente las observaciones se clasifican en h estratos de acuerdo a los limites obtenidos:

   - $E_1 = [x_{(1),}\xi_1]$
   - $E_2 =]\xi_1, \xi_2]$
   - …
   - $E_h =]\xi_{h-1}, x_{(n)}]$

1.1 Ejemplo de estratificación con el método de cuantiles

El objetivo es estratificar las viviendas particulares habitadas por entidad federativa en 4 estratos $(h = 4)$, el nivel de desagregación será a nivel estatal, por tanto $n = 32$.

1. Se ordenan los datos de forma ascendente.
2. Se obtiene los porcentajes que determinarán los puntos de corte en función de los 4 estratos. En este ejemplo se tiene:

- $Q_1 = \frac{1}{4} * 100 = 25$
- $Q_2 = 2\frac{1}{4} * 100 = 50$
- $Q_3 = 3\frac{1}{4} * 100 = 75$

3. Se obtiene el porcentaje que cada elemento representa dentro del conjunto de datos, así como su correspondiente porcentaje acumulado. En este caso cada observación tiene un porcentaje asociado de $S_i = \frac{1}{33} * 100 = 3.03$.

Entidad (Observaciones ordenadas)		Cuantiles		
		Frecuencia	Porcentaje	Porcentaje acumulado
03	Baja California Sur	175,046	3.03	3.03
06	Colima	177,848	3.03	6.06
04	Campeche	211,632	3.03	9.09
29	Tlaxcala	272,507	3.03	12.12
18	Nayarit	288,680	3.03	15.15
01	Aguascalientes	289,575	3.03	18.18
23	Quintana Roo	363,066	3.03	21.21
32	Zacatecas	372,662	3.03	24.24
16	Michoacán de Ocampo	1,066,630	3.03	72.73
07	Chiapas	1,072,560	3.03	75.76
19	Nuevo León	1,191,114	3.03	78.79
11	Guanajuato	1,266,772	3.03	81.82
21	Puebla	1,373,772	3.03	84.85
14	Jalisco	1,802,424	3.03	87.88
30	Veracruz de Ignacio de	1,983,543	3.03	90.91
09	Distrito Federal	2,388,534	3.03	93.94
15	México	3,689,053	3.03	96.97

4. Finalmente, se obtienen los puntos de corte para cada estrato y se clasifican todas las observaciones en el estrato correspondiente (pasos 4 y 5).

El primer punto de corte se establece, con el valor de la variable que acumule el 25%.

Entidad (Observaciones ordenadas)		Cuantiles			
		Frecuencia	Porcentaje	Porcentaje acumulado	Estrato (4 cuantiles)
03	Baja California Sur	175,046	3.03	3.03	1
06	Colima	177,848	3.03	6.06	1
04	Campeche	211,632	3.03	9.09	1
29	Tlaxcala	272,507	3.03	12.12	1
18	Nayarit	288,680	3.03	15.15	1
01	Aguascalientes	289,575	3.03	18.18	1
23	Quintana Roo	363,066	3.03	21.21	1
32	Zacatecas	372,662	3.03	24.24	1
10	Durango	398,471	3.03	27.27	2

Punto de corte del primer estrato

Como se puede ver en el recuadro rojo, no se tiene una observación para la que se acumule exactamente el 25% de los datos. El primer punto de corte se obtiene entonces interpolando los valores de las entidades 32 (Zacatecas) y 10 (Durango). De esta manera se obtiene el primer cuantil $\xi_1 = 379{,}114.25$ y por lo tanto, en la distribución ordenada de los datos, la entidad 32 (Zacatecas) será la última incluida en el primer estrato (punto inmediato inferior a 379,114.25).

El resto de los cuantiles se obtienen de manera análoga, para los porcentajes que determinan los puntos de corte, como se muestra a continuación:

Entidad (Observaciones ordenadas)		Cuantiles			
		Frecuencia	Porcentaje	Porcentaje acumulado	Estrato (4 cuantiles)
03	Baja California Sur	175,046	3.03	3.03	1
06	Colima	177,848	3.03	6.06	1
04	Campeche	211,632	3.03	9.09	1
29	Tlaxcala	272,507	3.03	12.12	1
18	Nayarit	288,680	3.03	15.15	1
01	Aguascalientes	289,575	3.03	18.18	1
23	Quintana Roo	363,066	3.03	21.21	1
32	Zacatecas	372,662	3.03	24.24	1
10	Durango	398,471	3.03	27.27	2
26	Sonora	705,668	3.03	48.48	2
25	Sinaloa	709,960	3.03	51.52	3
16	Michoacán de Ocampo	1,066,630	3.03	72.73	3
07	Chiapas	1,072,560	3.03	75.76	4

Porcentajes que determinan la división de estratos:

- $Q_1 = 25$
- $Q_2 = 50$
- $Q_3 = 75$

5. Una vez obtenidos todos los puntos de corte que definen los límites de los estratos, las observaciones se clasifican en el estrato correspondiente. En este ejemplo los límites de los estratos quedan definidos de la siguiente manera:

- $E_1 = [175046\ ,\ 379114.25\,]$
- $E_2 = \,]\,379114.25\,, 707814\,]$
- $E_3 = \,]707814\,, 1071077.5\,]$
- $E_4 = \,]1077077.5\,,\ 3689053\,]$

Para este ejemplo, al asignar cada entidad a su estrato correspondiente, se obtienen entonces 4 estratos con 8 elementos cada uno como se puede observar en el Anexo del presente documento.

2. Cuantiles relativos

Los cuantiles relativos son una variante de los cuantiles, son medidas de posición que definen puntos de corte en una distribución ordenada de datos. A diferencia de los cuantiles, los cuantiles relativos consideran la proporción que representa el valor de cada observación respecto a la suma total de la variable.

Dado un conjunto de n observaciones de una variable $x=\{x_1, x_2 ... x_n\}$, el procedimiento para formar h estratos a partir de estas observaciones sería el siguiente:

1. Ordenar los datos de forma ascendente.
2. Obtener los porcentajes requeridos para obtener los puntos de corte en función de los h estratos que se quieren formar, esto es de la siguiente forma:
    - $Q_1 = \frac{1}{h} * 100$
    - $Q_2 = 2\frac{1}{h} * 100$
    - …
    - $Q_{h-1} = h - 1\frac{1}{h} * 100$
3. Obtener el porcentaje que cada elemento representa dentro del conjunto de datos, considerando el peso especifico de cada observación: $S_i = \frac{x_{(i)}}{\sum_{i=1}^{n} x_{(i)}} * 100$. Se obtiene también para cada observación su correspondiente porcentaje acumulado que está dado por: $\sum_{k=1}^{i} S_k$.

Observaciones ordenadas	% (S_i)	% Acumulado $(\sum_{k=1}^{i} S_k)$
$x_{(1)}$	$\frac{x_{(1)}}{\sum_{i=1}^{n} x_{(i)}} * 100$	$\frac{x_{(1)}}{\sum_{i=1}^{n} x_{(i)}} * 100$
$x_{(2)}$	$\frac{x_{(2)}}{\sum_{i=1}^{n} x_{(i)}} * 100$	$\frac{x_{(1)} + x_{(2)}}{\sum_{i=1}^{n} x_{(i)}} * 100$
…	….	….
$x_{(n)}$	$\frac{x_{(n)}}{\sum_{i=1}^{n} x_{(i)}} * 100$	100

4. Los puntos de corte para cada estrato (cuantiles relativos) se definen de la siguiente manera:
    - $\xi_1 = \max \{x_{(i)}\} \ tal\ que\ \sum S_i = Q_1$
    - $\xi_2 = \max \{x_{(i)}\} \ tal\ que\ \sum S_i = Q_2$
    - …
    - $\xi_{h-1} = \max \{x_{(i)}\} \ tal\ que\ \sum S_i = Q_{h-1}$

En caso de no tener una observación que cumpla con la condición $\sum S_i = Q_k$, se realiza una interpolación lineal entre las dos observaciones más cercanas.

5. Finalmente las observaciones se clasifican en h estratos de acuerdo a los limites obtenidos:

- $E_1 = [\mathrm{x}_{(1)}, \xi_1]$
- $E_2 =]\xi_1, \xi_2]$
- …
- $E_h =]\xi_{h-1}, \mathrm{x}_{(\mathrm{n})}]$

2.1 Ejemplo de estratificación con el método de cuantiles relativos

El objetivo es estratificar las viviendas particulares habitadas por entidad federativa en 4 estratos $(h = 4)$, el nivel de desagregación será a nivel estatal, por tanto $n = 32$.

1. Se ordenan los datos de forma ascendente.
2. Se obtienen los porcentajes que determinarán los puntos de corte en función de los 4 estratos. En este ejemplo se tiene:
   - $Q_1 = \frac{1}{4} * 100 = 25$
   - $Q_2 = 2\frac{1}{4} * 100 = 50$
   - $Q_3 = 3\frac{1}{4} * 100 = 75$

3. Se obtiene el porcentaje que cada elemento representa respecto al total de viviendas particulares habitadas, así como su correspondiente porcentaje acumulado.

Entidad (Observaciones ordenadas)		Cuantiles relativos		
		Frecuencia	Porcentaje	Porcentaje acumulado
03	Baja California Sur	175,046	0.62	0.62
06	Colima	177,848	0.63	1.25
04	Campeche	211,632	0.75	2.00
29	Tlaxcala	272,507	0.97	2.97
18	Nayarit	288,680	1.03	4.00
01	Aguascalientes	289,575	1.03	5.03
23	Quintana Roo	363,066	1.29	6.32
32	Zacatecas	372,662	1.32	7.64

16	Michoacán de Ocampo	1,066,630	3.79	47.56
07	Chiapas	1,072,560	3.81	51.37
19	Nuevo León	1,191,114	4.23	55.60
11	Guanajuato	1,266,772	4.50	60.09
21	Puebla	1,373,772	4.88	64.97
14	Jalisco	1,802,424	6.40	71.37
30	Veracruz de Ignacio de	1,983,543	7.04	78.42
09	Distrito Federal	2,388,534	8.48	86.90
15	México	3,689,053	13.10	100.00

4. Finalmente, se obtienen los puntos de corte para cada estrato y se clasifican todas las observaciones en el estrato correspondiente (pasos 4 y 5).

El primer punto de corte se establece con el valor de la variable que acumule el 25%.

Entidad (Observaciones ordenadas)		Cuantiles relativos			
		Frecuencia	Porcentaje	Porcentaje acumulado	Estrato (4 cuantiles relativos)
03	Baja California Sur	175,046	0.62	0.62	1
06	Colima	177,848	0.63	1.25	1
04	Campeche	211,632	0.75	2.00	1
29	Tlaxcala	272,507	0.97	2.97	1
18	Nayarit	288,680	1.03	4.00	1
01	Aguascalientes	289,575	1.03	5.03	1
23	Quintana Roo	363,066	1.29	6.32	1
32	Zacatecas	372,662	1.32	7.64	1
10	Durango	398,471	1.42	9.05	1
22	Querétaro	450,104	1.60	10.65	1
17	Morelos	460,868	1.64	12.29	1
31	Yucatán	503,106	1.79	14.08	1
27	Tabasco	559,114	1.99	16.06	1
24	San Luis Potosí	631,587	2.24	18.30	1
13	Hidalgo	662,651	2.35	20.66	1
26	Sonora	705,668	2.51	23.16	1
25	Sinaloa	709,960	2.52	25.68	2

Punto de corte del primer estrato

Como se puede ver en el recuadro rojo, no se tiene una observación para la que se acumule exactamente el 25% de los datos. El primer punto de corte se obtiene entonces interpolando los valores de las entidades 26 (Sonora) y 25 (Sinaloa). De esta manera se obtiene el primer cuantil $\xi_1 = 708{,}795$, por lo tanto, en la distribución ordenada de los datos, la entidad 26 (Sonora) será la última incluida en el primer estrato (punto inmediato inferior a 708,795).

El resto de los cuantiles se obtienen de manera análoga, para los porcentajes que determinan los puntos de corte, como se muestra a continuación:

Entidad (Observaciones ordenadas)		Cuantiles relativos			
		Frecuencia	Porcentaje	Porcentaje acumulado	Estrato (4 cuantiles relativos)
03	Baja California Sur	175,046	0.62	0.62	1
06	Colima	177,848	0.63	1.25	1
04	Campeche	211,632	0.75	2.00	1
29	Tlaxcala	272,507	0.97	2.97	1
18	Nayarit	288,680	1.03	4.00	1
01	Aguascalientes	289,575	1.03	5.03	1
23	Quintana Roo	363,066	1.29	6.32	1
32	Zacatecas	372,662	1.32	7.64	1
10	Durango	398,471	1.42	9.05	1
22	Querétaro	450,104	1.60	10.65	1
17	Morelos	460,868	1.64	12.29	1
31	Yucatán	503,106	1.79	14.08	1
27	Tabasco	559,114	1.99	16.06	1
24	San Luis Potosí	631,587	2.24	18.30	1
13	Hidalgo	662,651	2.35	20.66	1
26	Sonora	705,668	2.51	23.16	1
25	Sinaloa	709,960	2.52	25.68	2
16	Michoacán de Ocampo	1,066,630	3.79	47.56	2
07	Chiapas	1,072,560	3.81	51.37	3
14	Jalisco	1,802,424	6.40	71.37	3
30	Veracruz de Ignacio de	1,983,543	7.04	78.42	4

Porcentajes que determinan la división de estratos:

- $Q_1 = 25$
- $Q_2 = 50$
- $Q_3 = 75$

5. Una vez obtenidos todos los puntos de corte que definen los límites de los estratos, las observaciones se clasifican en el estrato correspondiente. En este ejemplo los límites de los estratos quedan definidos de la siguiente manera:

   - $E_1 = [175046\ ,708795]$
   - $E_2 =]\ 708795,\ 1070434\ ]$
   - $E_3 =]1070434\ ,\ 1895678.56]$
   - $E_4 =]1895678.56\ ,\ 3689053\ ]$

Para este ejemplo, al asignar cada entidad a su estrato correspondiente, se obtienen entonces 4 estratos con 16, 8, 5 y 3 elementos cada uno como se puede observar en el Anexo del presente documento.

3. Dalenius - Hodges

El método de Dalenius-Hodges (1959) consiste en la formación de estratos de manera que la varianza obtenida sea mínima al interior de cada estrato y máxima entre cada uno de ellos, es decir, formar estratos lo más homogéneos posible.

Dado un conjunto de n observaciones de una variable $x = (x_1, x_2 \ldots x_n)$, el procedimiento para formar h estratos a partir de estas observaciones sería el siguiente:

1. Ordenar las observaciones de manera ascendente.

2. Agrupar $x = (x_1, x_2 \ldots x_n)$ en un número J de clases, donde $J = \min\{h * 10, n\}$.

3. Calcular los límites para cada clase de la siguiente manera:

$$\lim\inf C_k = \min\{x_{(i)}\} + (k-1) * \frac{\max\{x_{(i)}\} - \min\{x_{(i)}\}}{J}$$

$$\lim\sup C_k = \min\{x_{(i)}\} + (k) * \frac{\max\{x_{(i)}\} - \min\{x_{(i)}\}}{J}$$

   Los intervalos se tomarán abiertos por la izquierda y cerrados por la derecha, a excepción del primero que será cerrado por ambos lados.

4. A partir de estos límites, obtener la frecuencia de observaciones en cada clase

$$f_i \qquad i \in \{1, \ldots, J\}.$$

5. Calcular la raíz cuadrada de frecuencia en cada clase.

6. Acumular la raíz cuadrada de las frecuencias en cada clase $\sum_{i=1}^{J} \sqrt{f_i}$.

7. Dividir la suma de la raíz cuadrada de las frecuencias por el número de estratos: $Q = \frac{1}{h} \sum_{i=1}^{J} \sqrt{f_i}$.

8. Los puntos de corte de cada estrato se tomarán sobre el acumulado de la raíz cuadrada de las frecuencias en cada clase de acuerdo a lo siguiente: $Q, 2Q, \ldots, (h-1)Q$. Si el valor de Q queda entre dos clases, se tomará como punto de corte aquella clase que presente la mínima distancia a Q. Los límites de los h estratos conformados serán aquellos correspondientes a los límites inferior y superior de las clases comprendidas en cada estrato.

3.1 Ejemplo de estratificación con el método de Dalenius-Hodges

El objetivo es estratificar las viviendas particulares habitadas por entidad federativa en 4 estratos $(h = 4)$, el nivel de desagregación será a nivel estatal, por tanto $n = 32$.

1. Se ordenan los datos de forma ascendente.

Entidad (Observaciones ordenadas)		Dalenius-Hodges
		Frecuencia
03	Baja California Sur	175,046
06	Colima	177,848
04	Campeche	211,632
29	Tlaxcala	272,507
18	Nayarit	288,680
01	Aguascalientes	289,575
23	Quintana Roo	363,066
32	Zacatecas	372,662
11	Guanajuato	1,266,772
21	Puebla	1,373,772
14	Jalisco	1,802,424
30	Veracruz de Ignacio de	1,983,543
09	Distrito Federal	2,388,534
15	México	3,689,053

2. En este caso se generan *J=min(4*10,32)=32* clases en las que se agruparán los datos.

3. Se calculan los límites inferiores y superiores de cada clase y se obtiene la frecuencia de observaciones en cada clase. (Pasos 3 y 4).

Clase	Limites de la clase		Frecuencia (f)
	inf	sup	
1	175046	284859	4
2	284859	394671	4
3	394671	504484	4
4	504484	614297	1
5	614297	724110	5
6	724110	833922	1
7	833922	943735	4
8	943735	1053548	0
25	2810551	2920364	0
26	2920364	3030177	0
27	3030177	3139989	0
28	3139989	3249802	0
29	3249802	3359615	0
30	3359615	3469428	0
31	3469428	3579240	0
32	3579240	3689054	1

4. Se obtiene la raíz cuadrada de la frecuencia en cada clase y se acumula.

Clase	Limites de la clase		Frecuencia (f)	Raiz cuadrada de la frecuencia (√f)	Raiz cuadrada acumulada de la frecuencia (∑√f)
	inf	sup			
1	175046	284859	4	2.00	2.00
2	284859	394671	4	2.00	4.00
3	394671	504484	4	2.00	6.00
4	504484	614297	1	1.00	7.00
5	614297	724110	5	2.24	9.24
6	724110	833922	1	1.00	10.24
7	833922	943735	4	2.00	12.24
8	943735	1053548	0	0.00	12.24

25	2810551	2920364	0	0.00	19.06
26	2920364	3030177	0	0.00	19.06
27	3030177	3139989	0	0.00	19.06
28	3139989	3249802	0	0.00	19.06
29	3249802	3359615	0	0.00	19.06
30	3359615	3469428	0	0.00	19.06
31	3469428	3579240	0	0.00	19.06
32	3579240	3689054	1	1.00	20.06

5. Se obtiene el valor de Q, en este caso:

$$Q = \frac{1}{4}(20.06) = 5.016$$

6. Los puntos de corte estarán entonces determinados por:

- *Q=5.02*
- *2Q=10.03*
- *3Q=15.45*

Clase	Limites de la clase		Frecuencia (f)	Raiz cuadrada de la frecuencia (√f)	Raiz cuadrada acumulada de la frecuencia (∑√f)	
	inf	sup				
1	175046.00	284858.72	4.00	2.00	2.00	
2	284858.72	394671.44	4.00	2.00	4.00	Q
3	394671.44	504484.16	4.00	2.00	6.00	
4	504484.16	614296.88	1.00	1.00	7.00	
5	614296.88	724109.59	5.00	2.24	9.24	$2Q$
6	724109.59	833922.31	1.00	1.00	10.24	
7	833922.31	943735.03	4.00	2.00	12.24	
8	943735.03	1053547.75	0.00	0.00	12.24	
9	1053547.75	1163360.47	2.00	1.41	13.65	$3Q$
10	1163360.47	1273173.19	2.00	1.41	15.06	
11	1273173.19	1382985.91	1.00	1.00	16.06	
12	1382985.91	1492798.63	0.00	0.00	16.06	
13	1492798.63	1602611.34	0.00	0.00	16.06	
14	1602611.34	1712424.06	0.00	0.00	16.06	

En este ejemplo los límites de los estratos quedan definidos de la siguiente manera:

- $E_1 = [175046\ \ ,504484]$
- $E_2 =]\ 504484,\ \ 833922\ ]$
- $E_3 =]833922\ ,\ \ 1{,}273173]$
- $E_4 =]1273173\ ,\ \ 3689053\ ]$

7. A partir de estos límites, se clasifica cada una de las observaciones en alguno de los cuatro estratos obtenidos.

Entidad (Observaciones ordenadas)		Dalenius-Hodges	
		Frecuencia	Estrato
03	Baja California Sur	175,046	1
06	Colima	177,848	1
04	Campeche	211,632	1
29	Tlaxcala	272,507	1
18	Nayarit	288,680	1
01	Aguascalientes	289,575	1
23	Quintana Roo	363,066	1
32	Zacatecas	372,662	1
10	Durango	398,471	1
22	Querétaro	450,104	1
17	Morelos	460,868	1
31	Yucatán	503,106	1
27	Tabasco	559,114	2
12	Guerrero	805,230	2
02	Baja California	858,676	3
11	Guanajuato	1,266,772	3
21	Puebla	1,373,772	4

Para este ejemplo, al asignar cada entidad a su estrato correspondiente, se obtienen entonces 4 estratos con 12, 7, 8 y 4 elementos cada uno, como se puede observar en el Anexo del presente documento.

4. Número de elementos iguales

Este método tiene por objetivo conformar estratos que contengan aproximadamente el mismo número de observaciones.

Dado un conjunto de n observaciones de una variable $x = (x_1, x_2 ... x_n)$, el procedimiento para formar h estratos a partir de estas observaciones sería el siguiente:

1. Ordenar los datos de forma ascendente y asignar a cada elemento el número correspondiente a su posición ($i \in \{1, \dots, n\}$).
2. Obtener el número de observaciones por estrato dividiendo el total de observaciones entre el número de estratos deseados.

$$f = \frac{n}{h}$$

3. Los estratos se conforman de acuerdo a la fracción f y a la posición del elemento. Así cada estrato ($k = 1, \dots, h$) se conformará de la siguiente manera[1]:

    - $E_k = \{x_{(i)}\}\ tales\ que\ \ i \in [[\![(k-1) * f]\!] + 1, [\![k * f]\!]]$

Es importante señalar que dado que en este método de estratificación se considera únicamente la posición del elemento independientemente del valor que tome la variable, los límites de los estratos pueden presentar solapamientos. Por lo general este método arrojará resultados muy similares a los obtenidos con el método de cuantiles.

5. Personalizado

Este método permite al usuario definir el número de estratos a conformar así como determinar los límites de cada estrato de acuerdo a sus necesidades. El usuario deberá tener ciertas consideraciones al determinar los límites de los estratos, por ejemplo:

- El límite inferior del primer estrato no debe ser mayor al mínimo valor registrado en los datos.
- El límite superior del último estrato no debe ser menor al máximo valor registrado en los datos.
- Los intervalos definidos por el usuario deberán abarcar todo el rango de los datos y no deberán presentar solapamientos, o no abarcar el rango de la variable en su totalidad.

Los valores que presenta inicialmente este método corresponden al método de estratificación por cuantiles , el usuario puede modificar estos límites de acuerdo a sus necesidades específicas.

[1] $[\![\quad]\!]$ representa la función máximo entero.

Referencias bibliográficas

Taro Yamare (1974) *Estadística, una análisis introductorio*. Harla

Cáseres J. J. (2007) *Conceptos básicos de estadista para ciencias sociales*. Delta

Sipegel M. and Stephens L. (1999) *Statistics third edition, Shaum's outlines*. McGraw Hill

Dalenius, T. and Hodges, J. L. (1959) *Minimum Variance Stratification*. Journal of the American Statistical Association, 54, 88-101.

Ekman, G. (1959) *An Approximation Useful in Univariate Stratification*. Annals of Mathematical Statistics, 30, 219-229.

Anexo

Ejemplo de estratificación del número de viviendas particulares habitadas por entidad federativa en cuatro estratos, considerando diferentes métodos de estratificación.

Entidad (Observaciones ordenadas)		Frecuencia	Cuantiles	Cuantiles relativos	Dalenius-Hodges	NEI
03	Baja California Sur	175,046	1	1	1	1
06	Colima	177,848	1	1	1	1
04	Campeche	211,632	1	1	1	1
29	Tlaxcala	272,507	1	1	1	1
18	Nayarit	288,680	1	1	1	1
01	Aguascalientes	289,575	1	1	1	1
23	Quintana Roo	363,066	1	1	1	1
32	Zacatecas	372,662	1	1	1	1
10	Durango	398,471	2	1	1	2
22	Querétaro	450,104	2	1	1	2
17	Morelos	460,868	2	1	1	2
31	Yucatán	503,106	2	1	1	2
27	Tabasco	559,114	2	1	2	2
24	San Luis Potosí	631,587	2	1	2	2
13	Hidalgo	662,651	2	1	2	2
26	Sonora	705,668	2	1	2	2
25	Sinaloa	709,960	3	2	2	3
05	Coahuila de Zaragoza	715,158	3	2	2	3
12	Guerrero	805,230	3	2	2	3
02	Baja California	858,676	3	2	3	3
28	Tamaulipas	868,244	3	2	3	3
08	Chihuahua	910,647	3	2	3	3
20	Oaxaca	934,471	3	2	3	3
16	Michoacán de Ocampo	1,066,630	3	2	3	3
07	Chiapas	1,072,560	4	3	3	4
19	Nuevo León	1,191,114	4	3	3	4
11	Guanajuato	1,266,772	4	3	3	4
21	Puebla	1,373,772	4	3	4	4
14	Jalisco	1,802,424	4	3	4	4
30	Veracruz de Ignacio de la Llave	1,983,543	4	4	4	4
09	Distrito Federal	2,388,534	4	4	4	4
15	México	3,689,053	4	4	4	4

Como puede observarse en la tabla anterior, los resultados de estratificación varían en algunos casos considerablemente de acuerdo al método seleccionado para estratificar. Se recomienda al usuario analizar cuidadosamente los resultados antes de utilizar la clasificación obtenida.