|  |  |  |
| --- | --- | --- |
| **EL INGRESO DE LOS HOGARES EN LOS CENSOS DE POBLACIÓN Y VIVIENDA 2011** | 12 de Setiembre  2016 | |
| La encuesta de hogares contiene información complementaria para estimar | |  |

SP-1632 Casos Estadísticos

Pedro Montenegro Masís.

**Introducción**

En Costa Rica se han realizado diez Censos Nacionales de Población y seis Censos Nacionales de Vivienda, los más recientes corresponden al Censo de Población y Vivienda de 2011, que tuvo desarrollo del 30 de mayo al 3 de junio de 2011.

El Censo de 2011 tuvo el adelanto de la Cartografía y el beneficio de la base de datos de toda la población y las viviendas de Costa Rica, que se resume en el sistema de información Geoestadístico. Dicha información conlleva una importante actualización mediante relevantes encuestas, principalmente la Encuesta Nacional de Hogares (ENAHO).

Desde 1986 que se construyó en primer Marco Muestral de Viviendas (MMV) a partir del Censo de Población y Vivienda de 1984, se ha venido actualizando de forma extendida y poco frecuente la información referente al Marco Muestral. En los inicios se conformaron segmentos que representan áreas geográficas que comprendían aproximadamente 60 viviendas en zonas urbanas y 40 viviendas en zonas rurales.

En los periodos entre 1993 y 1997 con un gran lapso de desactualización con respecto al Marco Muestral anterior, se actualizó la información para dar inicio al Censo de Población y Vivienda de 2000, consigo se elaboró el nuevo Marco Muestral con cambios geográficos utilizando la nueva clasificación de regiones de planificación (Región Central, Región Chorotega, Región Pacifico Central, Región Brunca, Región Huetar Atántica y Región Huetar Norte), de igual forma se utilizó la desagregación de zona (urbano, rural).

A partir del Censo de Población y Vivienda del 2011 se construyó el Marco Muestral de Viviendas 2011 (MMV-2011) y con ello se inició el cambio base de las encuestas como la EHANO y la Encuesta Continua de Empleo (ECE).

El Marco Muestral de 2011 se dividió en áreas geográficas que en promedio estas áreas contenían 125 viviendas en la zona urbana y 100 viviendas en la zona rural, cambio drástico en comparación con MMV-2000 que en promedio tenían 60 viviendas en la zona urbana y 40 viviendas en la zona rural. Con esto el principal cambio que se observó fue el aumento de los segmentos urbanos del MMV-2000 al MMV-2011, que representa aproximadamente un incremento del 61,2% a 70,4% respectivamente.

El siguiente trabajo pretende simular el Marco Muestral de 2011, a partir de los parámetros geográficos que corresponden a las regiones de planificación, las zonas urbano-rurales. Por otra parte, considera la proporción de unidades principales de muestreo (UPM) por distrito, así como la variabilidad distintas variables que le corresponden a las viviendas y las personas, con el propósito de representar la información contenida por segmentos.

**ENAHO 2011**

La encuesta de hogares











**CENSO 2011**

Un marco muestral contiene un listado exhaustivo de viviendas que se agrupan en unidades principales de muestreo (UPM). A cada UPM se le incorpora información censal más actualizada que permite la estratificación del marco de acuerdo con las necesidades particulares de cada encuesta. Esta información está relacionada con su ubicación geográfica, cantidad de viviendas, características sociodemográficas, socioeconómicas de la población residente en esas viviendas.

**Cálculo de las variables de estratificación.**

Con el propósito de que el marco muestral contenga la información correspondiente del Censo de Población y Vivienda de 2011, se utilizaron en resumen # indicadores que determinen el ingreso.

Las variables analizadas corresponden a las pertenencias de la vivienda, número de aposentos, número de dormitorios, total de miembros en el hogar, ocupación y escolaridad del jefe de hogar.

Construyéndose 4 indicadores para la estimación del ingreso como el hacinamiento, nivel educativo, ocupación, pertenencias.

**Construcción de indicadores**

***Indicador de hacinamiento***

Según la Comisión Económica para América Latina (CEPAL) el indicador de hacinamiento está conceptualizado como número de personas por pieza de dormitorio. Los resultados de este indicador tratan de caracterizar la forma de vida y tienen implicaciones en la salud, comodidad y particularidades de la vivienda.

Para la construcción del indicador es necesario el total de personas en el hogar y la cantidad de aposentos para dormir. Para el número de aposentos menor a 1, se le asigna un 1, esto es porque se parte del supuesto que todas las viviendas tienen al menos un aposento para dormir.

La encuesta de hogares contiene la variable de hacinamiento, sin embargo, la variable muestre resultados enteros categorizados con 1 y 2 personas por aposento, por lo que es conveniente que dicho indicador presente la misma construcción en la enaho y el censo.

CONTINUAR

**Cálculo de pertenencias de la vivienda.**

Para el cálculo de las pertenencias de la vivienda se incluyen artefactos y servicios, que corresponden a la batería de preguntas del Censo 2011, de manera que se analiza la confección de un indicador que aproxime las pertenencias de las viviendas y consigo mismo se pueda representar que las viviendas que mayor cantidad de artefactos o pertenencias tienden a tener mejores condiciones económicas, inversamente sucede cuando el número de pertenencias es menor, disminuye las capacidades de adquisición, que finalmente se relaciona con las condiciones económicas de un hogar.

Para la construcción del indicador se inició recodificando todas las variables, de forma que se le asignan el valor de 0= sí no tiene dicho artefacto y se asigna el valor de 1=sí tiene el artefacto, esto se replicó para los doce artefactos o servicios:

Los resultados de la tenencia para cada artefacto o servicio son los siguientes:

Cuadro 4. Frecuencias de la tenencia de artefactos o servicios

|  |  |  |
| --- | --- | --- |
| **Id** | **Artefacto o servicio** | **Freq (ni/N)** |
| 1 | Tv convencional | 0,898 |
| 2 | Radio o Equipo de sonido | 0,785 |
| 3 | Teléfono fijo | 0,578 |
| 4 | Tv cable satélite | 0,459 |
| 5 | Carro no de trabajo | 0,378 |
| 6 | Internet | 0,335 |
| 7 | Computadora de escritorio | 0,328 |
| 8 | Computadora portátil | 0,271 |
| 9 | Tv moderno | 0,206 |
| 10 | Tanque de almacenamiento agua | 0,132 |
| 11 | Sistema de agua caliente | 0,119 |
| 12 | Motocicleta no de trabajo | 0,112 |

Para conocer si deben utilizarse todos los artefactos o servicios se realizó la correlación de cada una de las pertenencias con los años de educación del jefe, de esta forma se obtiene el Eta que indica la razón de la correlación de cada variable.

Cuadro 5. Análisis de correlación entre los artefactos y los años escolaridad del jefe

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Años de escolaridad del jefe | | | | | | |
| Artefacto o servicio | Promedio | N | Desviación Estándar | Eta | | Eta Cuadrado |
| Radio o Equipo de sonido | | | | | | |
| No | 7,15 | 260018 | 4,581 |  |  | |
| Sí | 8,68 | 951946 | 4,856 | ,129 | ,017 | |
| Teléfono fijo | | | | | | |
| No | 7,15 | 510987 | 4,235 |  |  | |
| Sí | 9,22 | 700977 | 5,060 | ,211 | ,045 | |
| Tv moderno | | | | | | |
| No | 7,53 | 962317 | 4,466 |  |  | |
| Sí | 11,50 | 249647 | 4,928 | ,332 | ,110 | |
| **Tv convencional** | | | | | | |
| **No** | **8,40** | **123960** | **5,483** |  |  | |
| **Sí** | **8,34** | **1088004** | **4,760** | **,004** | **,000** | |
| Tv cable satélite | | | | | | |
| No | 6,61 | 656121 | 4,089 |  |  | |
| Sí | 10,40 | 555843 | 4,855 | ,390 | ,152 | |
| Sistema de agua caliente | | | | | | |
| No | 7,83 | 1067551 | 4,575 |  |  | |
| Sí | 12,18 | 144413 | 5,011 | ,291 | ,057 | |
| Tanque de almacenamiento agua | | | | | | |
| No | 8,11 | 1052010 | 4,710 |  |  | |
| Sí | 9,90 | 159954 | 5,365 | ,125 | ,016 | |
| Computadora de escritorio | | | | | | |
| No | 7,16 | 815023 | 4,424 |  |  | |
| Sí | 10,80 | 396941 | 4,734 | ,353 | ,125 | |
| Computadora portátil | | | | | | |
| No | 6,99 | 883455 | 4,102 |  |  | |
| Sí | 11,99 | 328509 | 4,788 | ,459 | ,211 | |
| Internet | | | | | | |
| No | 6,66 | 805949 | 3,942 |  |  | |
| Sí | 11,70 | 406015 | 4,707 | ,492 | ,242 | |
| Carro no de trabajo | | | | | | |
| No | 6,91 | 753625 | 4,150 |  |  | |
| Sí | 10,72 | 458339 | 4,958 | ,382 | ,146 | |
| **Moto no de trabajo** | | | | | | |
| **No** | **8,33** | **1075934** | **4,886** |  |  | |
| **Sí** | **8,50** | **136030** | **4,448** | **,011** | **,000** | |

Como puede verse, tanto la tenencia de tv convencional y la de moto no están relacionadas a los años de educación del jefe, por lo tanto no se incluyeron en la construcción del indicador. Por este motivo, de los 12 ítems iniciales quedaron únicamente 9 ítems. Las frecuencias simples para las pertenencias son las siguientes (sin tv convencional ni moto):

Cuadro 6. Frecuencias de la tenencia de artefactos o servicios

|  |  |  |
| --- | --- | --- |
| **Id** | **Artefacto o servicio** | **Freq (ni/N)** |
| 1 | Tv convencional | 0,898 |
| 2 | Radio o Equipo de sonido | 0,785 |
| 3 | Teléfono fijo | 0,578 |
| 4 | Tv cable satélite | 0,459 |
| 5 | Carro no de trabajo | 0,378 |
| 6 | Internet | 0,335 |
| 7 | Computadora de escritorio | 0,328 |
| 8 | Computadora portátil | 0,271 |
| 9 | Tv moderno | 0,206 |
| 10 | Tanque de almacenamiento agua | - |
| 11 | Sistema de agua caliente | - |
| 12 | Motocicleta no de trabajo | - |

Para construir un procedimiento que posteriormente facilite la construcción de la cantidad de artefactos de la vivienda se procedió a realizar un esquema de las Escalas de Guttman. Dado que cada artefacto o pertenencia vale 0 o 1, el máximo es de 9. Es decir, un puntaje de 9 equivale a tener todas las pertenencias. Un puntaje de 7 equivale a tener computadora portátil y todos los que le preceden, pero no los que están después del artefacto 7.

Obsérvese que cada pertenencia tiene un puntaje. Ese puntaje es para establecer las pertenencias que tendrá cada vivienda. Por ejemplo, un puntaje de 9 significa que tiene todos, en cuyo caso todos deberán llevar el código 1. Si el puntaje es 0 entonces tendrá código cero en todas las pertenencias. Si el puntaje es 5 entonces la vivienda tendrá código 1 en internet y todos los que lo preceden, mientras que, carro, tv cable satélite y radio o equipo de sonido tendrán código 0. Así se procede con el resto de los puntajes.

Cuadro 7. Asignación de puntajes para los artefactos

|  |  |  |
| --- | --- | --- |
| Pertenencias | % | Puntaje |
| Tv convencional | 0.22 | 9,0 |
| Tv moderno | 0.21 | 8,0 |
| Computadora portátil | 0.27 | 7,0 |
| Computadora de escritorio | 0.33 | 6,0 |
| Internet | 0.34 | 5,0 |
| Carro no de trabajo | 0.38 | 4,0 |
| Tv cable satélite | 0.46 | 3,0 |
| Teléfono fijo | 0.58 | 2,0 |
| Radio o Equipo de sonido | 0.79 | 1,0 |
| Total | 3.56 |  |

Para generar un puntaje al azar para cada UPM del marco se calculó el puntaje de pertenencias promedio de cada distrito y su desviación estándar. Posteriormente, se calculó un intervalo del promedio±1,5 desviaciones estándar en cada distrito. Es decir, cada distrito en el archivo (zona urbana y rural) tiene un puntaje mínimo (pertenencias\_min) y máximo (pertenencias\_max) del índice de pertenencias con la idea de generar puntajes al azar entre esos dos valores para las viviendas de cada UPM en un distrito determinado.

Si el intervalo tenía valores negativos o superiores a 9, se acoto la cantidad de pertenencias para los valores negativos en 0 y para los valores mayores a 9 en su valor máximo.

Una vez que una vivienda de una UPM tenga asignado al azar un puntaje, se procederá a decidir qué pertenencias específicas posee, bajo el esquema de las Escalas de Guttman.





**NUMERO DE APOSENTOS Y DORMITORIOS**

Para el cálculo del número de aposentos, se considera el promedio de aposentos por zona y distrito, se calcula un intervalo de confianza de 2 veces la desviación estándar de la media, además se obtiene el número de aposentos mínimo (num\_aposentos\_min) y el número de aposentos máximo (num\_aposentos\_max). Si el número de aposentos mínimo es menor que 1, se asigna un 1. Esto es porque todas las viviendas tienen al menos un aposento.

Para cada distrito debe generarse, dentro de cada UPM y dentro de cada vivienda, un número aleatorio entre “num\_aposentos\_min” y “num\_aposentos\_max”.

Por otra parte, para calcular el número de dormitorios se considera el promedio de apliqué el mismo procedimiento, de la siguiente manera:

Posteriormente se corrobora que las diferencias del número de aposentos con el número de dormitorios no deben ser negativas, es decir, el número de dormitorios en ningún caso puede ser superior al número de aposentos.

No obstante, puede suceder que al momento de asignar al azar los dormitorios, su número sea mayor al de aposentos. Por ejemplo, si el mínimo de aposentos es 2 y el máximo es 9, pues ser que por azar salga el 4. En este caso, si el número mínimo de dormitorios es 0 y el máximo es 5, entonces podría salir el 5.

Debido a lo anterior, la asignación aleatoria del número de dormitorios estará restringida al número de aposentos que tenga la vivienda. Por ejemplo, si el número de aposentos fue de 4, entonces el número de dormitorios, que tenía un rango de variación de cero a 5 va a estar restringido para que salga de cero a 4.

**VIVIENDAS QUE COCINAN CON ELECTRICIDAD, GAS O LEÑA**

Para construir la proporción de viviendas que utilizan como combustible para cocinar la electricidad, el gas o leña se debe respetarse la distribución porcentual del combustible utilizado para cocinar en cada distrito.

En general, el promedio de viviendas que utilizan combustibles tienen distribución de uso de combustibles de la siguiente manera:

Cuadro 8. Frecuencia en el uso de combustibles para cocinar

|  |  |
| --- | --- |
| Combustible | Frecuencia oscila |
| Gas | 0,1-0,20 |
| Electricidad | 0,2-0.77 |
| leña | 0.01-0.02 |
| No usa | 0.00-0.02 |
| Total usan | 0.988 |

Por lo tanto, si en el distrito “x” se deben generarse 22 UPMs de aproximadamente 135 viviendas y la información del censo dice que cocinan con electricidad 0,77 de las viviendas, con gas 0,19 y con leña 0,02 y el porcentaje que no está es porque dice “otro o no cocina”. Esto significa que cada una de las 22 UPMs debe respetar, aproximadamente, esa distribución, haciendo que un 77% de las viviendas cocinen con electricidad, 19% con gas y casi ninguna con leña.

**NÚMERO DE MIEMBROS**

Para calcular el número de residentes en la vivienda se calcula el número de hombres y luego el número de mujeres. Para hombres se calculó un intervalo por distrito “total\_hombres\_min” y “total\_hombres\_max”. De la misma forma se realizó para las mujeres. En ambos casos se aplicaron dos desviaciones estándar para la construcción de los intervalos de confianza.

Seguidamente se generan aleatoriamente el número de hombres y de mujeres por vivienda y posteriormente se obtiene el número total de residentes, sumando los hombres y las mujeres.

**ESCOLARIDAD DEL JEFE**

Para construir la escolaridad del jefe se utiliza la variable de jefe de hogar donde se extrae la escolaridad mínima y la máxima escolaridad, posteriormente se aplica una desviación estándar para la construcción de los intervalos de confianza.

Finalmente, se obtendrá un número aleatorio para asignar la escolaridad del jefe de hogar, la ventaja de usar este procedimiento es controlar dicha variable por zona y por distrito.

**Análisis y construcción de la creación del marco muestral.**

Para construir el marco muestral de aproximadamente 1,200,000 viviendas se utilizó el lenguaje de programación R.

**Comparación de resultados.**

Inicialmente se presentó en los cuadros 1 y 3 distribuciones generales de las características del MMV-2011, posterior a la construcción, programación y elaboración de las variables para la simulación del MMV-2011 se presenta la información general de las características del Marco muestral.

**Cuadro 9. Características generales del MMV-2011**

|  |  |  |
| --- | --- | --- |
| **Zona** | **Región** | **Total de viviendas** |
| Urbano |  | **Total** |
| Región Central | 669,003 |
| Región Chorotega | 51,379 |
| Región Pacífico Central | 47,482 |
| Región Brunca | 42,301 |
| Región Huetar Atlántica | 63,973 |
| Región Huetar Norte | 31,589 |
| **Sub-Total** | 905,727 |
| Rural |  | **Total** |
| Región Central | 103,476 |
| Región Chorotega | 42,819 |
| Región Pacífico Central | 25,299 |
| Región Brunca | 52,903 |
| Región Huetar Atlántica | 47,471 |
| Región Huetar Norte | 59,286 |
| **Sub- Total** | 331,254 |
| **Total** | | 1,236,981 |

Fuente: Censo de Población y Vivienda, INEC 2011.

De esta forma se comparan nuevamente las características por zona y región del Marco Muestral simulado para el total de viviendas.

**Cuadro 10. Características generales de la simulación del MMV-2011**

|  |  |  |
| --- | --- | --- |
| **Zona** | **Región** | **Total de viviendas** |
| Urbano |  | **Total** |
| Región Central | 646,501 |
| Región Chorotega | 49,654 |
| Región Pacífico Central | 47,200 |
| Región Brunca | 44,290 |
| Región Huetar Atlántica | 63,557 |
| Región Huetar Norte | 32,501 |
| **Sub-total** | 883,703 |
| Rural |  | **Total** |
| Región Central | 111,386 |
| Región Chorotega | 43,160 |
| Región Pacífico Central | 25,045 |
| Región Brunca | 50,222 |
| Región Huetar Atlántica | 45,227 |
| Región Huetar Norte | 57,070 |
| **Sub-total** | 332,110 |
| **Total** | | 1,212,314 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Nótese que la cantidad de viviendas que posee el MMV-11 “simulado” es similar a la cantidad de viviendas original, con una diferencia de 24,667 menos viviendas que el marco muestral simulado.

A pesar de lo anterior, se tiene que el total de personas en el marco simulado es mayor al total de personas indicado en dicho censo por 634,868 habitantes, sobreestimando los resultados del MMV-11. Otro aspecto importante es que la relación entre hombres y mujeres en ambos casos es similar, simulado (0.984) original (0.959), es decir, se respecta que hay mayor cantidad de mujeres en el país y la estimación es bastante buena.

**Cuadro 11. Comparación de valores absolutos**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Características | Censo 2011 | | | Simulado | | | DIF Total (censo-simulado) |
| Urbano | Rural | Total | Urbano | Rural | Total |
| Total de viviendas | 878,180 | 333,784 | 1,211,964 | 878,478 | 333,836 | 1,212,314 | -350 |
| Viviendas Urbanas | 878,180 | 0 | 878,180 | 878,478 | 0 | 878,478 | -298 |
| Viviendas Rurales | 0 | 333,784 | 333,784 | 0 | 333,836 | 333,836 | -52 |
| Cantidad de dormitorios | 2,235,528 | 802,304 | 3,037,832 | 2,415,790 | 822,827 | 3,238,617 | -200,785 |
| Cantidad de aposentos | 4,306,787 | 1,502,973 | 5,809,760 | 4,321,128 | 1,506,011 | 5,827,139 | -17,379 |
| Total de hombres | 1,463,451 | 595,080 | 2,058,531 | 1,724,331 | 680,359 | 2,404,690 | -346,159 |
| Total de mujeres | 1,574,254 | 580,760 | 2,155,014 | 1,769,719 | 674,004 | 2,443,723 | -288,709 |

Se puede apreciar que las diferencias que se presentan para ambos casos son relativamente pequeñas.

También se muestran las diferencias encontradas con las principales variables para construcción del MMV-2011.

**Cuadro 12. Distribuciones porcentuales de las principales variables de análisis**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Característica | Censo 2011 | | | Simulado | | | DIF Total (censo-simulado) |
| Urbano | Rural | Total | Urbano | Rural | Total |
| Electricidad | 62.0 | 34.6 | 54.4 | 65.0 | 54.1 | 62.0 | -7.6 |
| Gas | 34.3 | 45.1 | 37.2 | 32.1 | 28.8 | 31.2 | 6.0 |
| Leña | 2.9 | 19.0 | 7.3 | 2.9 | 19.4 | 7.5 | -0.1 |
| Teléfono fijo | 62.7 | 45.0 | 57.8 | 85.6 | 69.7 | 81.2 | -23.4 |
| Tv moderno | 24.3 | 10.9 | 20.6 | 5.8 | 0.4 | 4.3 | 16.3 |
| Tv cable satélite | 55.6 | 20.3 | 45.9 | 71.5 | 49.5 | 65.4 | -19.6 |
| Computadora escritorio | 38.1 | 18.8 | 32.8 | 27.3 | 5.9 | 21.4 | 11.4 |
| Computadora portátil | 31.4 | 15.8 | 27.1 | 14.9 | 1.7 | 11.3 | 15.8 |
| La vivienda cuenta con internet | 40.1 | 16.2 | 33.5 | 41.4 | 14.9 | 34.1 | -0.6 |
| La vivienda cuenta con carro propio | 41.1 | 29.1 | 37.8 | 56.3 | 30.0 | 49.1 | -11.2 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Las diferencias más grandes se encuentran en el porcentaje de teléfonos fijos, tv por cable satelital, la tenencia de computadora portátil y si cuenta con carro propio. Sin embargo las demás diferencias son menores al 10% y se consideran bastante aproximadas a las originales.

Además, se presentan resultados de los indicadores construidos, como el total de pertenencias y los años de escolaridad del jefe.

**Cuadro 13. Promedios de los indicadores de análisis**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Características | Censo 2011 | | | Simulado | | | DIF Total (censo-simulado) |
| Urbano | Rural | Total | Urbano | Rural | Total |
| Indicador de pertenencias | 4.0 | 2.5 | 3.6 | 4.0 | 2.6 | 3.6 | -0.1 |
| Años escolaridad | 9.1 | 6.5 | 8.3 | 9.5 | 7.0 | 8.8 | -0.5 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Obsérvese que no existen grandes diferencias entre dichos indicadores, ya que los promedios son muy similares.

En el anexo 2 y 3 se presentan resultados por región de planificación, nótese que se observar que se mantiene la misma tendencia de sobreestimación.

**AJUSTES EN LA SIMULACIÓN DEL MMV-11**

Debido a la sobreestimación en la mayoría de las variables de la simulación del Marco Muestral de Viviendas se procedió a estimar nuevamente el MMV-11, está vez considerando el ajuste de las variables que creaban problemas en las estimaciones.

Para la segunda simulación de marco muestral la cantidad de viviendas calculadas fue subestimada en 35.918, sin embargo, el total de personas en esta ocasión es de 14.738, que se considera que el ajuste es bastante bueno al comparado con la primer simulación de MMV-11 que sobreestimó en 634,868 en total de personas.

Para el nuevo cálculo de hombres y mujeres se generaron números aleatorios con una distribución normal, aprovechando la media y desviación estándar por UPM, para resolver la sobreestimación se disminuyó la media en 0,15.

**Cuadro 14. Comparación de valores absolutos (ajustado)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Características | Censo 2011 | | | Simulado | | | DIF Total (censo-simulado) |
| Urbano | Rural | Total | Urbano | Rural | Total |
| Total de viviendas | 878,180 | 333,784 | 1,211,964 | 879,019 | 297,027 | 1,176,046 | 35,918 |
| Viviendas Urbanas | 878,180 | 0 | 878,180 | 879,019 | 0 | 879,019 | -839 |
| Viviendas Rurales | 0 | 333,784 | 333,784 | 0 | 297,027 | 297,027 | 36,757 |
| Cantidad de dormitorios | 2,235,528 | 802,304 | 3,037,832 | 2,306,923 | 739,118 | 3,046,041 | -8,209 |
| Cantidad de aposentos | 4,306,787 | 1,502,973 | 5,809,760 | 4,319,171 | 1,372,749 | 5,691,920 | 117,840 |
| Total de hombres | 1,463,451 | 595,080 | 2,058,531 | 1,523,087 | 538,381 | 2,061,468 | -2,937 |
| Total de mujeres | 1,574,254 | 580,760 | 2,155,014 | 1,605,431 | 531,908 | 2,137,339 | 17,675 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Para el cálculo de las pertenencias se generaron números aleatorios entre 0 y 0.8 de una distribución uniforme, además, se establecieron intervalos de confianza para asignar la pertenencia del artefacto, siempre manteniendo la relación del esquema de las escalas de Guttman, en donde todos los artefactos están relacionados con la tenencia y su cantidad para las viviendas, restringiendo la posibilidad de que se encuentren viviendas con artefactos que económicamente son más caros pero no posean artefactos básicos, es decir, se mantiene la congruencia.

**Cuadro 15. Ajuste del número de pertenencias**

|  |  |  |  |
| --- | --- | --- | --- |
| **Artefacto** | **Valor Mínimo** | **Valor máximo** | **Escala Guttman** |
| Agua caliente | 0,1 | 0,216 | 9 |
| Tv moderno | 0,3 | 0,206 | 8 |
| Computadora portátil | 0,2 | 0,271 | 7 |
| Computadora de escritorio | 0,2 | 0,328 | 6 |
| Tenencia de internet | 0,1 | 0,335 | 5 |
| Tenencia de Carro | 0,1 | 0,378 | 4 |
| Tv cable satélite | 0,02 | 0,459 | 3 |
| Televisor fijo | 0 | 0,578 | 2 |
| Radio equipo | 0 | 0,785 | 1 |
| **Pertenencias** |  | **3,669** |  |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Asimismo, se calculan y se comparan los porcentajes del censo MMV-2011 contra el marco muestral simulado.

**Cuadro 16. Distribuciones porcentuales de las principales variables de análisis (ajustado)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Característica | Censo 2011 | | | Simulado | | | DIF Total (censo-simulado) |
| Urbano | Rural | Total | Urbano | Rural | Total |
| Electricidad | 62.0 | 34.6 | 54.4 | 42.5 | 11.7 | 54.19 | 0.26 |
| Gas | 34.3 | 45.1 | 37.2 | 30.1 | 9.2 | 39.32 | -2.07 |
| Leña | 2.9 | 19.0 | 7.3 | 2.2 | 4.7 | 6.89 | 0.43 |
| Teléfono fijo | 62.7 | 45.0 | 57.8 | 48.1 | 16.2 | 64.37 | -6.53 |
| Tv moderno | 24.3 | 10.9 | 20.6 | 17.1 | 5.8 | 22.87 | -2.27 |
| Tv cable satélite | 55.6 | 20.3 | 45.9 | 37.9 | 12.8 | 50.66 | -4.79 |
| Computadora escritorio | 38.1 | 18.8 | 32.8 | 27.2 | 9.2 | 36.42 | -3.67 |
| Computadora portátil | 31.4 | 15.8 | 27.1 | 22.5 | 7.6 | 30.09 | -2.99 |
| La vivienda cuenta con internet | 40.1 | 16.2 | 33.5 | 27.8 | 9.4 | 37.26 | -3.76 |
| La vivienda cuenta con carro propio | 41.1 | 29.1 | 37.8 | 31.4 | 10.6 | 42.01 | -4.20 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Obsérvese que las diferencias de los totales mejoraron sustancialmente, en este caso la diferencia más grande es de -6.53% para la tenencia de teléfono fijo, para los demás servicios la diferencia se encuentra entre el -3%, que quiere decir que el porcentaje calculado por la simulación del marco muestral es mayor en promedio el 3% para los artefactos de las viviendas.

**Cuadro 17. Promedios de los indicadores de análisis (ajustado)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Características | Censo 2011 | | | Simulado | | | DIF Total (censo-simulado) |
| Urbano | Rural | Total | Urbano | Rural | Total |
| Indicador de pertenencias | 4.0 | 2.5 | 3.6 | 4.0 | 3.9 | 3.9 | -0.4 |
| Años escolaridad | 9.1 | 6.5 | 8.3 | 9.0 | 6.8 | 8.5 | -0.1 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Entre los resultados más importantes se puede concluir que la nueva forma de calcular el número de pertenencias, entre la primer simulación del marco muestral de viviendas 2011 y el marco muestral ajustado 2011mejoró significativamente.

**Cuadro 18. Diferencias entre los totales del primer MMV y el MMV-Ajustado**

|  |  |  |  |
| --- | --- | --- | --- |
| Característica | DIF Total (censo-simulado) | DIF Total (censo-simulado) | Disminución % |
|
| Electricidad | -7.6 | 0.26 | -3% |
| Gas | 6 | -2.07 | -35% |
| Leña | -0.1 | 0.43 | N.A |
| Teléfono fijo | -23.4 | -6.53 | -28% |
| Tv moderno | 16.3 | -2.27 | -14% |
| Tv cable satélite | -19.6 | -4.79 | -24% |
| Computadora escritorio | 11.4 | -3.67 | -32% |
| Computadora portátil | 15.8 | -2.99 | -19% |
| La vivienda cuenta con internet | -0.6 | -3.76 | N.A |
| La vivienda cuenta con carro propio | -11.2 | -4.2 | -38% |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Se puede observar que en todos los casos las disminuciones están por encima de un 10% de las diferencias obtenidas con respecto al primer marco muestral simulado. En general, la mayor diferencia que se observa en la segunda estimación de MMV-ajustado, se da en telefonía fija con una diferencia porcentual de -6.5%, mientras que anteriormente las diferencias eran mayores al 10% en todos los casos y telefonía fija estaba sobreestimada con un 23%.

**EVALUACIÓN DE DISEÑOS MUESTRALES PARA EL MMV-11 SIMULADO**

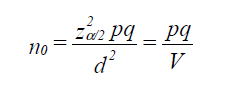
Con el propósito de evaluar el marco muestral 2011 se decidió simular diferentes diseños muestrales y se realizaron las estimaciones de errores para comparar los diseños de muestreo entre sí.

Utilizando el archivo “MM\_2011”, se seleccionó una muestra de viviendas para hacer una encuesta a adultos de 18 años y más. Además, se establecen los siguientes supuestos para el cálculo y selección de la muestra de forma que se utilice el mejor diseño muestral.

* Para la selección de la muestra se conoce que las variables de interés tiene una gran variación (entre 30 y 70%).
* Se estableció para la muestra un nivel de confianza del 95%.
* Un error máximo permitido del 3%.
* Para efectos del trabajo de campo se desean cargas diarias en cada *UPM* seleccionada de alrededor de 5 entrevistas, una entrevista por vivienda.
* Por lo general, en este tipo de encuestas se obtiene 10% de no respuesta. Encuestas similares indican que valores de *rho* cercanos a 0.3 son bastante probables.

Inicialmente se calcula el tamaño de muestra asumiendo un muestreo irrestricto al azar (mia) sin ajustes.

con p= 0.5

(1)

Ajustando el 10% de no respuesta el tamaño de muestra se modifica de la siguiente manera:

* Ajuste del 10% de no respuesta:

Posteriormente, utilizando un diseño complejo se ajusta el tamaño de la muestra del diseño mia obtenido anteriormente, a la vez, se aprovecha el valor del coeficiente de correlación intraclase ρ=0.3 y la selección de 5 entrevistas por UPM. De esta manera se calcula el tamaño de muestra para un diseño complejo:

donde b es la selección de entrevistas por UPM

De esta manera se tiene que el tamaño de muestra para calcular los diseño es aproximadamente n=2608.

Las variables de interés son las siguientes:

* Total de personas: total de personas en cada vivienda.
* Hacinamiento en las viviendas: total de personas en la vivienda / cantidad de dormitorios en la vivienda.
* Tenencia de internet en las viviendas: variable dicotómica que indica si la vivienda posee o no internet (1=tenencia de internet, 0=no posee internet)
* Tenencia de carro en la vivienda: variable dicotómica que indica si la vivienda posee o no internet (1=tenencia de carro, 0=no posee carro)

Para completar el tamaño de muestra se realizó un diseño conglomerado de una etapa donde se seleccionaban UPMs y posteriormente el total de viviendas de la UPM, en promedio el tamaño de muestra es 2608, ya que el promedio de UPM urbano tiene 130 viviendas y el promedio de UPM rural es de 95 viviendas.

|  |  |  |
| --- | --- | --- |
| **Cuadro 19. Diseño conglomerado de una etapa** | | |
| Zona | UPMs | Viviendas b |
| Urbano | 15 | Completas |
| Rural | 7 | Completas |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Se utiliza el mismo diseño de muestreo estratificados por zona y con selección de UPMs completas.

**Cuadro 20. Diferencias entre los totales para cada diseño con el total**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Total | Diseño 1 n=2628 | Dif (total-estimado) | Diseño 2 n=2620 | Dif (total-estimado) |
| internet | 438,170 | 522,686 | -84,516 | 455,623 | -17,453 |
| carro | 494,093 | 592,823 | -98,730 | 508,246 | -14,153 |
| num\_dormitorios | 3,046,041 | 3,742,246 | -696,205 | 3,143,367 | -97,326 |
| num\_miembros | 4,198,807 | 5,136,353 | -937,546 | 4,957,441 | -758,634 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Seguidamente, se obtienen los cálculos de los errores para las variables de análisis, según la réplica del diseño utilizado.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Diseño 1 n=2628 | | | |  | Diseño 1 n=2610 | | | |
|  | Urbano | Rural | Total de viviendas |  |  | Urbano | Rural | Total de viviendas |
| Estimación | 1,024,416 | 416,556 | 1,440,972 |  | Estimación | 878,071 | 357,048 | 1,235,119 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Para el número de miembros por vivienda.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Diseño 1 n=2628  Total de miembros por vivienda | | | |  | Diseño 1 n=2610  Total de miembros por vivienda | | | |
|  | Urbano | Rural | Total de Población |  |  | Urbano | Rural | Total de Población |
| Estimación | 3,616,963 | 1,519,390 | 5,136,353 |  | Estimación | 3,522,320 | 1,435,120 | 4,957,441 |
| Error estándar | 28,579 | 20,453 | 35,144 |  | Error estándar | 27,774 | 17,714 | 32,942 |
| CVE | 0.5531 | 0.9423 | 0.4790 |  | CVE | 0.4731 | 0.7406 | 0.3987 |
| DEFF | 0.7000 | 0.7000 | 0.7066 |  | DEFF | 0.6000 | 0.6000 | 0.6001 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Para el número de viviendas con internet.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Diseño 1 n=2628, Internet | | | |  | Diseño 1 n=2610, Internet | | | |
|  | Urbano | Rural | Total de viviendas con internet |  |  | Urbano | Rural | Total de viviendas con internet |
| Estimación | 370,810 | 151,876 | 522,686 |  | Estimación | 321,210 | 134,413 | 455,623 |
| Error estándar | 11,096 | 7,800 | 13,563 |  | Error estándar | 9,864 | 6,222 | 11,662 |
| CVE | 2.0946 | 3.5950 | 1.8164 |  | CVE | 1.8426 | 2.7774 | 1.5358 |
| DEFF | 0.7000 | 0.7000 | 0.7057 |  | DEFF | 0.6000 | 0.6000 | 0.6002 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Para el total de dormitorios.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Diseño 1 n=2628, Números de dormitorios | | | |  | Diseño 1 n=2610, Números de dormitorios | | | |
|  | Urbano | Rural | Total de dormitorios |  |  | Urbano | Rural | Total de dormitorios |
| Estimación | 2,690,458 | 1,051,788 | 3,742,246 |  | Estimación | 2,309,178 | 834,190 | 3,143,367 |
| Error estándar | 45,848 | 30,535 | 55,086 |  | Error estándar | 40,691 | 21,971 | 46,244 |
| CVE | 1.1929 | 2.0322 | 1.0304 |  | CVE | 1.0573 | 1.5803 | 0.8827 |
| DEFF | 0.7000 | 0.7000 | 0.6996 |  | DEFF | 0.6000 | 0.6000 | 0.5998 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Estimación para el total de carros.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Diseño 1 n=2628. Carros | | | |  | Diseño 1 n=2610, Carros | | | |
|  | Urbano | Rural | Total de viviendas con carro |  |  | Urbano | Rural | Total de viviendas con carro |
| Estimación | 423,932 | 168,891 | 592,823 |  | Estimación | 361,362 | 146,884 | 508,246 |
| Error estándar | 11,371 | 7,957 | 13,879 |  | Error estándar | 10,078 | 6,319 | 11,896 |
| CVE | 1.8777 | 3.2977 | 1.6388 |  | CVE | 1.6734 | 2.5814 | 1.4043 |
| DEFF | 0.7000 | 0.7000 | 0.7052 |  | DEFF | 0.6000 | 0.6000 | 0.6004 |

Fuente: Elaboración propia, a partir de la muestra de la base del Censo-2011.

Se puede observar que los coeficientes de variación para las variables de análisis son muy altos y los efectos de diseño no son mayores a 1 lo cual son presentan inconsistencias, es claro que se debe revisar el procedimiento y comprobar que dicho diseño y análisis es correcto.

Complementariamente se realizaron varias réplicas del mismo diseño y los resultados son similares al diseño 2, además se aumentó el tamaño de muestra a 4100 viviendas y las estimaciones de los coeficientes de variancia disminuyeron muy poco.

**Anexos**

**Anexo 1. Variables y definiciones para construcción del MMV-20111.**

|  |  |  |
| --- | --- | --- |
| **ID** | **Variables** | **Definición** |
| 1 | REGION | Son las 7 regiones administrativas 1=Gran Área Metropolitana, 2= Resto de la Región Central, 3=Región Chorotega, 4=Región Pacifico Central, 5=Región Brunca, 6=Región Huetar Atántica y 7=Región Huetar Norte. |
| 2 | zona | Zona de residencia 1) U=urbana, 2) R= rural. |
| 3 | pcd | Código de provincia, cantón y distrito 7 provincias, 81 cantones (1=20, 2=15, 3=8, 4=10, 5=11, 6=11, 7=6), con 472 distritos |
| 4 | prov | Provincias 1=San José, 2=Alajuela, 3=Cartago, 4=Heredia, 5=Guanacaste, 6=Puntarenas, 7=Limón. |
| 5 | cant | Cantones 81. |
| 6 | dist | Distritos 472 |
| 7 | viv\_total\_sum | Total de viviendas |
| 8 | viv\_urbana\_sum | Total de viviendas urbanas |
| 9 | viv\_rural\_sum | Total de viviendas rural |
| 10 | viv\_por\_upm | Total de viviendas por UPM |
| 11 | numero\_de\_upms | Numero de UPM por distrito |
| 12 | li | Límite inferior de viviendas |
| 13 | ls | Límite superior de viviendas |
| 14 | cant\_dormitorios\_mean | Promedio de la cantidad de dormitorios |
| 15 | cant\_dormitorios\_sd | Desviación estándar de la cantidad de dormitorios |
| 16 | num\_dormitorios\_min | Límite inferior de la cantidad de dormitorios |
| 17 | num\_dormitorios\_max | Límite superior de la cantidad de dormitorios |
| 18 | cant\_aposentos\_mean | Promedio de la cantidad de aposentos |
| 19 | cant\_aposentos\_sd | Desviación estándar de la cantidad de aposentos |
| 20 | num\_aposentos\_min | Límite inferior de aposentos |
| 21 | num\_aposentos\_max | Límite inferior de aposentos |
| 22 | electricidad\_mean | Promedio de cocina en electricidad |
| 23 | gas\_mean | Promedio de cocina en gas |
| 24 | lena\_mean | Promedio de cocina en leña |
| 25 | radio\_equipo\_sum | Total de radios |
| 26 | tel\_fijo\_sum | Total de teléfonos fijos |
| 27 | tv\_moderno\_sum | Total de televisor moderno |
| 28 | tv\_cable\_satelite\_sum | Total de televisor cable |
| 29 | agua\_caliente\_sum | Total de agua caliente |
| 30 | compu\_escritorio\_sum | Total de computadora de escritorio |
| 31 | compu\_portatil\_sum | Total de computadora portatil |
| 32 | internet\_sum | Total de internet |
| 33 | carro\_sum | Total de carro |
| 34 | pertenencias\_mean | Total de pertenecías |
| 35 | pertenencias\_sd | Desviación estándar de pertenencias |
| 36 | pertenencias\_min | Límite inferior de pertenencias |
| 37 | pertenencias\_max | Límite superior de pertenencias |
| 38 | total\_hombres\_mean | Promedio total de hombres |
| 39 | total\_hombres\_sd | Desviación estándar de hombres |
| 40 | total\_hombres\_min | Límite inferior de total de hombres |
| 41 | total\_hombres\_max | Límite superior de total de hombres |
| 42 | total\_mujeres\_mean | Promedio total de mujeres |
| 43 | total\_mujeres\_sd | Desviación estándar de mujeres |
| 44 | total\_mujeres\_min | Límite inferior de total de mujeres |
| 45 | total\_mujeres\_max | Límite superior de total de mujeres |
| 46 | anos\_escolaridad\_mean | Promedio de años de escolaridad |
| 47 | anos\_escolaridad\_sd | Desviación de años de escolaridad |
| 48 | anos\_escolaridad\_min | Límite inferior de años de escolaridad |
| 49 | anos\_escolaridad\_max | Límite superior de años de escolaridad |

Fuente: Elaboración propia a partir del resumen de MMV-2011.

Anexo 2. Distribuciones absolutas por región para el MMV simulado

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| REGIONES | | | | | | | | |
| Características | Gran Área Metropolitana | Resto Región Central | Región Chorotega | Región Pacífico Central | Región Brunca | Región Huetar Atlántica | Región Huertar Norte | Total |
| Total de viviendas | 619,142 | 137,737 | 92,638 | 72,011 | 93,392 | 108,583 | 88,811 | 1,212,314 |
| Viviendas Urbanas | 576,117 | 69,238 | 48,458 | 46,949 | 42,968 | 63,101 | 31,647 | 878,478 |
| Viviendas Rurales | 43,025 | 68,499 | 44,180 | 25,062 | 50,424 | 45,482 | 57,164 | 333,836 |
| Cantidad de dormitorios | 1,776,673 | 409,065 | 218,417 | 161,091 | 227,904 | 240,388 | 205,079 | 3,238,617 |
| Cantidad de aposentos | 3,161,082 | 688,696 | 407,081 | 306,858 | 420,630 | 457,717 | 385,075 | 5,827,139 |
| Total de hombres | 1,212,184 | 269,139 | 185,209 | 143,916 | 187,892 | 221,059 | 185,291 | 2,404,690 |
| Total de mujeres | 1,251,396 | 274,548 | 185,630 | 142,940 | 188,473 | 220,404 | 180,332 | 2,443,723 |

Fuente: Elaboración propia

Anexo 3. Distribuciones absolutas por región para el MMV-2011 INEC.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| REGIONES | | | | | | | | |
| Características | Gran Área Metropolitana | Resto Región Central | Región Chorotega | Región Pacífico Central | Región Brunca | Región Huetar Atlántica | Región Huertar Norte | Total |
| Total de viviendas | 618,761 | 136,730 | 92,584 | 71,690 | 93,978 | 109,316 | 88,905 | 1,211,964 |
| Viviendas Urbanas | 575,646 | 68,367 | 48,801 | 46,651 | 43,377 | 63,545 | 31,793 | 878,180 |
| Viviendas Rurales | 43,115 | 68,363 | 43,783 | 25,039 | 50,601 | 45,771 | 57,112 | 333,784 |
| Cantidad de dormitorios | 1,616,137 | 357,844 | 220,585 | 160,194 | 226,509 | 247,473 | 209,090 | 3,037,832 |
| Cantidad de aposentos | 3,137,365 | 681,351 | 406,953 | 302,975 | 425,724 | 465,551 | 389,841 | 5,809,760 |
| Total de hombres | 1,037,233 | 232,901 | 158,855 | 119,294 | 160,427 | 188,776 | 161,045 | 2,058,531 |
| Total de mujeres | 1,121,188 | 238,032 | 162,444 | 120,115 | 163,849 | 190,078 | 159,308 | 2,155,014 |

Fuente: Proporcionado por el profesor del curso SP-1625 Tópicos de Encuestas por Muestreo

\*\*Sintaxis\*\*

#library(foreign)

#library(car)

#Importar la base de datos#

#De esta manera se puede copiar un conjunto datos recien copiado#

#x=read.table("clipboard",header=T)

#x

#IMPORTACIÓN COMÚN#

#Es importante dirigir el directorio donde se encuentra la base a importar#

#Para direccionar File-> Change dir-> "seleccione la carpeta donde se encuentra la base"#

#Nombro mi base como "MM"

MM <- read.csv("MM.csv", header=T,sep=",")

attach(MM)

#Para verificar que la base se importó correctamente

#MM

#Dimensiones de la base (filas\*columnas)#

dim(MM)

#Etiquetas de las variables#

dimnames(MM)

#La función tiene problema cuando el valor que se debe replicar es cero#

#Para efectos del análisis no hay distritos sin UPMS, por lo tanto no es un problema#

#df=dimensión de las filas#

#dc=dimensión de las columnas#

#sv=suma de las upms#

df=dim(MM)[1]

dc=dim(MM)[2]

#MM es la base, los "[]" indican MM["filas","columnas"], es la posición especifica, en este caso las filas son cero y la columna se nombra entre comillas la variable que quiero replicar, también se puede utilizar el número de la columna donde esta la variable.

sv=sum(MM[,"numero\_de\_upms"])

#Ahora, construyo una matrix de ceros, de sv"filas" y dc"columnas"#

#Lo anterior se da porque voy a replicar el total de upms disponibles#

MN=matrix(0,sv,dc)

#Para programar la función requiero los siguientes números enteros, que inician con valores de 1.#

#En general, la función va a recorrer de forma ordenada las filas, todas las columnas y va a replicar las upms que le corresponde, siempre completando los valores con la base original.#

i=1 #RECORRE FILAS DE LA MATRIZ ORIGINAL.#

j=1 #RECORRE LAS FILAS DE LA MATRIZ NUEVA.#

r=1 #RECORRE LOS CAMPOS DE LAS COLUMNAS ANTES DE TERMINAR LAS REPLICAS DE UPMS.

p=1 #VALOR LOGICO PARA EL FUNCIONAMIENO DEL PRIMER FOR.#

#Rango del números de réplicas que voy replicar de mi columna.#

s=MM[1,"numero\_de\_upms"]#NUMERO DE REPLICAS TOTAL POR FILA DE LA MATRIX ORIGINAL.#

#Inicio con el while, que es el que va a recorrer mis filas de la matrix original, en este caso MM#

#Indico que haga el ciclo de 1 hasta df "dimensión de mis filas", recuérdese que es sobre la matrix original#

while (i <= df)

{

s=MM[i,"numero\_de\_upms"]#sirve para aumentar mi contador

for ( p in 1:s)#la variable p lo uso como valor fijo ya que j me sobrescribe los números de la matrix ya que su contador afecta el for. El "for" me repite la cantidad de upms que obtengo en la fila "#".#

{

for ( r in 1:dc)#este "for" me ayuda a completar las columnas de la posición 1 hasta dc. Lo hace adentro del "for" anterior para que en la fila 1 de la matrix nueva, complete con los mismos valores de la matrix original la dimensión de columnas de ambas matrices.

{

MN[j,r]= MM[i,r]#Indica los valores de (j,r) son iguales que los valores de (i,r)

r=r+1

}

j=j+1 #cuando se sale del "for2" aumenta el j en 1.

}

i = i+1#cuando se sale del "for1" aumenta el j en 1, por lo tanto i=j.

}#Duración de procesar la función es menor a 1 min.#

#MN

#resumen descriptivo de la matrix nueva.

summary(MN)

#Comprobar los cambios de la matrix nueva, verificó num de filas y columnas#

dim(MN)

#La Matriz nueva no tiene nombre en las variables, procedemos a nombrar#

(dimnames(MM)[[2]])#Extraemos los nombres de las columnas de la matrix original.#

dimnames(MN)<-list(NULL,paste(dimnames(MM)[[2]]))#renombramos la matrix nueva para que no pierda la etiqueta.#

dimnames(MN)

detach(MM)#desvincular las variables de la matriz original#

MN=data.frame(MN)#Necesario para realizar el attach, sino va producir error porque no la considera como una matriz.

attach(MN)#ya que vamos a trabajar con las variables de la base MN y que no se confunda con la matriz original "MM".#

#Hasta ahora creamos la base dividida por UPMS#

#Continuamos construyendo la base del MM\_T de viviendas#

#Ahora se debe seleccionar el número de viviendas aleatorias por UPMS#

#Función para números aleatorios enteros de un intervalo de confianza.#

#Voy a llamar a mi función "aleato", es decir, función= aleato(intervalo\_mín,intervalo\_max)

aleato = function(xm, xmax)

{

i=1

B=rep(0,length(xm))#B es mi vector de longitud de cualquiera de los dos vectores li and ls.#

for(i in 1:length(xm))#Indico que A "valor lógico" va de 1 a la longitud de los datos de los vectores límites.#

{

B[i]=runif(1,xm[i],xmax[i])#La función dice que complete el primer elemento del vector, ya que i=1, con números aleatorios uniformes entre el elemento del vector inferior y el vector superior#

i=i+1 #Contador de la función

}

round(B,0)#redondear los valores al mayor sin decimales, ya que necesitamos un número entero de viviendas#

}

#Recordemos que la base originales y la base nueva (expandida a UPMS) tiene las variables (li and ls) del número de viviendas, que se obtuvo del supuesto que las viviendas por zona tiene una desviación estándar de 35 y promedio (135=urbano, 95=rural)#

#El li y ls se llaman las variables de la base original, posteriormente renombradas en la base nueva "MN"..

ale\_viv\_upm=c(aleato(li,ls))

ale\_viv\_upm

length(ale\_viv\_upm)

#Ahora agrego la variable "ale\_viv\_ump" en nuestra base nueva.#

MN1=data.frame(MN,ale\_viv\_upm)

dim(MN1)#MN1 debe tener una columna adicional.#

detach(MN)#deshabilito las variables de MN#

attach(MN1)#habilito las variables de MN1#

####Expansión de las UPMS a viviendas aleatorias por UPMS.#

#Voy a omitir la explicación, porque es la misma función utilizada para replicar la primer base.#

df=dim(MN1)[1]

dc=dim(MN1)[2]

sv=sum(MN1[,"ale\_viv\_upm"])

#Ahora, construyo una matrix de ceros, de sv"filas" y dc"columnas"#

#Lo anterior se da porque voy a replicar el total de upms disponibles#

MN\_final4=matrix(0,sv,dc)

dim(MN\_final4)

#Para programar la función requiero de los siguientes números enteros, que inician mi función en 1.#

i=1

j=1

r=1

p=1

s=MN1[1,"ale\_viv\_upm"]

#Inicio con el while, que es el que va a recorrer mis filas de la matrix original, en este caso MM#

#Indico que haga el ciclo de 1 hasta df" dimensión de mis filas", recuérdese que es sobre la matrix original#

while (i <= df)

{

s=MN1[i,"ale\_viv\_upm"]

for ( p in 1:s)

{

for ( r in 1:dc)

{

MN\_final4[j,r]= MN1[i,r]

r=r+1

}

j=j+1

}

i = i+1

}

#MN\_final4

#summary(MN\_final4)

dim(MN\_final4)

#Repito la renombrar las variables, procedimiento conocido.#

(dimnames(MN1)[[2]])#Extraemos los nombres de las columnas de la matrix original.#

dimnames(MN\_final4)<-list(NULL,paste(dimnames(MN1)[[2]]))#renombramos la matrix nueva para que no pierda la etiqueta.#

dimnames(MN\_final4)

detach(MN1)#desvincular las variables de la matriz original#

MN\_final4=data.frame(MN\_final4)#Necesario para realizar el attach, sino va producir error porque no la considera como una matriz.

attach(MN\_final4)#ya que vamos a trabajar con las variables de la base MN y que no se confunda con la matriz original "MM".#

dimnames(MN\_final4)

#####EXPORTAR LA BASE DE DATOS A FORMATO txt. "Blog de notas" posteriormente se abre en excel.#

#write.table(MN\_final4, 'MN\_final4.txt', sep='\t', dec='.')

#Finalmente obtengo las características aleatorias de todas las variables de análisis.#

#Son 7 variables:1) cantidad de aposentos, 2) cantidad de dormitorios.#

#Para la selección aleatoria de la cantidad de aposentos utilizó los intervalos de confianza respectivos.#

#Recordar que el número aleatorio que obtenga en la cantidad de aposentos va se mi límite superior para la escogencia de cantidad de domitorios.#

#APOSENTOS#

num\_aposentos=aleato(num\_aposentos\_min,num\_aposentos\_max)

#DORMITORIOS#

num\_dormitorios=aleato(num\_dormitorios\_min,num\_aposentos)

#Electricidad,gas,leña#

#En este caso selecciono los números aleatorios de 0 a 0.98, y recodifico según la restricción de cada vivienda sobre el promedio de uso de elect, gas y leña que tenga la UPM.#

cocina=runif(length(gas\_mean),0,0.98)

i=1

lena=matrix("0",length(cocina))

gas=matrix("0",length(cocina))

elec=matrix("0",length(cocina))

for (i in 1:length(cocina))

{

if (cocina[i]<lena\_mean[i])

{

lena[i]=1

}else

lena[i]=0

}

for (i in 1:length(cocina))

{

if (cocina[i]>lena\_mean[i]&cocina[i]<gas\_mean[i])

{

gas[i]=1

}else

gas[i]=0

}

for (i in 1:length(cocina))

{

if (cocina[i]>gas\_mean[i]&cocina[i]<0.98)

{

elec[i]=1

}else

elec[i]=0

}

#PERTENENCIAS#

num\_pertenencias=aleato(pertenencias\_min,pertenencias\_max) ##crear la cantidad de pertennecias

i=1

pertenencias=c(rep(1:num\_pertenencias[1]),rep(0,(9-num\_pertenencias[1])))

for (i in 2:length(num\_pertenencias))

{

r=c(rep(1,num\_pertenencias[i]),rep(0,(9-num\_pertenencias[i])))

pertenencias=c(pertenencias,r)

i=i+1

}

#length(pertenencias)

m\_pertenencias=t(matrix(pertenencias,9))

#TOTAL DE HOMBRES#

tot\_hombres=aleato(total\_hombres\_min,total\_hombres\_max)

#TOTAL DE MUJERES#

tot\_mujeres=aleato(total\_mujeres\_min,total\_mujeres\_max)

#NUM DE MIEMBROS#

num\_miembros=tot\_hombres+tot\_mujeres

#AÑOS ESCOLARIDAD#

anos\_escolaridad=aleato(anos\_escolaridad\_min,anos\_escolaridad\_max)

#####ENUMERAR LAS UPM Y LAS VIVIENDAS#####

i=1

id\_viv=seq(1:ale\_viv\_upm[1])

for (i in 2:length(ale\_viv\_upm))

{

r=seq(1:ale\_viv\_upm[i])

id\_viv=c(id\_viv,r)

i=i+1

}

length(id\_viv)

marco\_muestral4=data.frame(REGION,zona, pcd,prov,id\_viv,cant,numero\_de\_upms,num\_aposentos,num\_dormitorios,cocina,electricidad\_mean,gas\_mean,lena\_mean,lena,gas,elec,num\_pertenencias,tot\_hombres,tot\_mujeres,num\_miembros,anos\_escolaridad,m\_pertenencias)

base=data.frame(m\_pertenencias)

**#AJUSTE DE VARIABLES#**

**#Función para crear números aleatorios de una distribución normal#**

r\_normal = function(xm, xmax)

{

i=1

B=rep(0,length(xm))#B es mi vector de longitud de cualquiera de los dos vectores li and ls.#

for(i in 1:length(xm))#Indico que A "valor lógico" va de 1 a la longitud de los datos de los vectores límites.#

{

B[i]=rnorm(1,xm[i],xmax[i])#La función dice que complete el primer elemento del vector, ya que i=1, con números aleatorios uniformes entre el elemento del vector inferior y el vector superior#

i=i+1 #Contador de la función

}

round(B,0)#redondear los valores al mayor sin decimales, ya que necesitamos un número entero de viviendas#

}

**#CÁLCULO DE PERTENENCIAS#**

perte=runif(length(pertenencias\_mean),0,0.9)

i=1

agua\_caliente=matrix("0",length(perte))#0,216

tv\_moderno=matrix("0",length(perte))#0,206

compu\_portatil=matrix("0",length(perte))#0,271

compu\_escritorio=matrix("0",length(perte))#0,328

internet=matrix("0",length(perte))#0,335

carro=matrix("0",length(perte))#0,378

tv\_cable\_satelite=matrix("0",length(perte))#0,459

tel\_fijo=matrix("0",length(perte))#0,578

radio\_equipo=matrix("0",length(perte))#0,785

#1#

for (i in 1:length(perte))

{

if (perte[i]>0.1&perte[i]<=0.316)

{

agua\_caliente[i]=1

}else

agua\_caliente[i]=0

}

#2#

for (i in 1:length(perte))

{

if (perte[i]>0.3&perte[i]<=0.506)

{

tv\_moderno[i]=1

}else

tv\_moderno[i]=0

}

#3#

for (i in 1:length(perte))

{

if (perte[i]>0.2&perte[i]<0.471)

{

compu\_portatil[i]=1

}else

compu\_portatil[i]=0

}

#4#

for (i in 1:length(perte))

{

if (perte[i]>0.2&perte[i]<=0.528)

{

compu\_escritorio[i]=1

}else

compu\_escritorio[i]=0

}

#5#

for (i in 1:length(perte))

{

if (perte[i]>0.1&perte[i]<=0.435)

{

internet[i]=1

}else

internet[i]=0

}

#6#

for (i in 1:length(perte))

{

if (perte[i]>0.1&perte[i]<=0.478)

{

carro[i]=1

}else

carro[i]=0

}

#7#

for (i in 1:length(perte))

{

if (perte[i]>0.15&perte[i]<=0.606)

{

tv\_cable\_satelite[i]=1

}else

tv\_cable\_satelite[i]=0

}

#8#

for (i in 1:length(perte))

{

if (perte[i]>0.1&perte[i]<=0.679)

{

tel\_fijo[i]=1

}else

tel\_fijo[i]=0

}

#9#

for (i in 1:length(perte))

{

if (perte[i]>0.0&perte[i]<=0.785)

{

radio\_equipo[i]=1

}else

radio\_equipo[i]=0

}

matrix\_perte=data.frame(agua\_caliente,tv\_moderno,compu\_portatil,compu\_escritorio,internet,carro,tv\_cable\_satelite,tel\_fijo,radio\_equipo)

attach(matrix\_perte)

#num\_pertenencias1=agua\_caliente+tv\_moderno+compu\_portatil+compu\_escritorio+internet+carro+tv\_cable\_satelite+tel\_fijo+radio\_equipo

write.table(matrix\_perte, 'perte.txt', sep='\t', dec='.')

#APOSENTOS#

num\_aposentos=r\_normal(cant\_aposentos\_mean,(2\*cant\_aposentos\_sd))

num\_aposentos=replace(num\_aposentos,num\_aposentos<1,1)

num\_aposentos=replace(num\_aposentos,num\_aposentos>9,9)

table(num\_aposentos)

length(num\_aposentos)

#DORMITORIOS#

a=rep(0,length(num\_aposentos))

num\_dormitorios=aleato(a,num\_aposentos)

table(num\_dormitorios)

num\_dormitorios=replace(num\_dormitorios,num\_dormitorios<1,1)

#TOTAL DE HOMBRES#

tot\_hombres=r\_normal(total\_hombres\_mean-0.15,total\_hombres\_sd)

tot\_hombres=abs(tot\_hombres)

tot\_hombres=replace(tot\_hombres,tot\_hombres<1,1)

tot\_hombres[is.na(tot\_hombres)]<-2

#TOTAL DE MUJERES#

tot\_mujeres=r\_normal(total\_mujeres\_mean-0.15,total\_mujeres\_sd)

tot\_mujeres=abs(tot\_mujeres)

tot\_mujeres=replace(tot\_mujeres,tot\_mujeres<1,1)

tot\_mujeres[is.na(tot\_mujeres)]<-2

#NUM DE MIEMBROS#

num\_miembros=tot\_hombres+tot\_mujeres

#AÑOS ESCOLARIDAD#

anos\_escolaridad=r\_normal(anos\_escolaridad\_mean,anos\_escolaridad\_sd)

anos\_escolaridad=replace(anos\_escolaridad,anos\_escolaridad>26,12)

anos\_escolaridad=replace(anos\_escolaridad,anos\_escolaridad>22,13)

anos\_escolaridad=replace(anos\_escolaridad,anos\_escolaridad>21,12)

anos\_escolaridad=replace(anos\_escolaridad,anos\_escolaridad>20,14)

anos\_escolaridad=replace(anos\_escolaridad,anos\_escolaridad<1,1)

anos\_escolaridad[is.na(anos\_escolaridad)]<-1

mmv\_2011\_final=data.frame(MN\_final4,agua\_caliente,tv\_moderno,compu\_portatil,compu\_escritorio,internet,carro,tv\_cable\_satelite,tel\_fijo,radio\_equipo,num\_dormitorios,num\_aposentos)

#####EXPORTAR LA BASE DE DATOS A FORMATO txt. "Blog de notas" posteriormente se abre en excel.#

write.table(mmv\_2011\_final, 'mmv\_2011\_finaln.txt', sep='\t', dec='.')

#####EXPORTAR LA BASE DE DATOS A FORMATO txt. "Blog de notas" posteriormente se abre en excel.#

write.table(marco\_muestral4, 'marco\_muestral4.txt', sep='\t', dec='.')

write.table(base, 'base.txt', sep='\t', dec='.')

#primer diseño#

prueba<- read.table("prueba.txt", header=T,sep="\t")

attach(prueba)

dimnames(prueba)

#INSTALAR PAQUETE#

install.pakages(TeachingSampling)

library(TeachingSampling)

#proporciones de los estratos

table(zona)[[2]]/(table(zona)[[1]]+table(zona)[[2]])

N1=table(zona)[[1]]

N2=table(zona)[[2]]

N1;N2

N=c(N1,N2)

n1=15

n2=7

n=c(n1,n2)

sample=S.STSI(zona,N,n)

muestra=prueba[sample,]

attach(muestra)

muestra

sum(ale\_viv\_upm)

table(seg)

#157, 399, 1421, 1670, 2115, 2139, 2721, 3378, 3483, 3532, 4019, 4091, 4544, 4660, 4679, 5260, 6172, 6480, 6856, 7848, 8472, 9658)

N=2628

(ov=xtabs(ale\_viv\_upm~zona))

1 1967

2 661

#N1=URBANO, N2=RURAL.

N1=1463451

N2=595080

N1;N2

N=c(N1,N2)

#MUESTRA.

n1=2837

n2=1273

n=c(n1,n2)

#estimación#

estima=data.frame(num\_miembros,internet,num\_dormitorios,carro)

0.6\*E.STSI(zona,N,n,estima)

m1=2837

m2=1273

m=c(m1,m2)

res=S.STPPS(zona,num\_miembros,m)

sam=res[,1]

pk=res[,2]

dim(pk)

muestra=diseno4[sam,]

attach(muestra)

dim(muestra)

estima=data.frame(num\_miembros,internet,num\_dormitorios,carro)

E.STPPS(estima,pk,m,zona)