

Uso y procesamiento de datos censales para la generación de indicadores sociodemográficos utilizando REDATAM+SP



NACIONES UNIDAS



Fechas: 21 al 24 de julio de 2015

Instructora: Prof. Alejandra Silva. Geógrafa (Universidad Católica de Chile); Master en Sistemas de Información Geográficos (Ohio State University), Doctorandus en Demografía (U. de Córdoba, Argentina). Actualmente se desempeña en el Área de Demografía e Información sobre Población del Centro Latinoamericano de Demografía (CELADE), División de Población, CEPAL. Principales áreas de trabajo: Asesoría técnica y soporte al usuario en el manejo y procesamiento de datos censales con el software Redatam. Asistencia en verificación de errores, creación de manuales y documentación, capacitación y entrenamiento del programa Redatam. Desarrollo de aplicaciones computacionales de tipo xPlan y de indicadores sociodemográficos a partir de los censos. Uso de sistemas de Información geográfica en temas de vulnerabilidad y pobreza. Creación de cartografía digital censal.

En este curso introductorio se aprenderá a generar indicadores sociodemográficos a partir del procesamiento estadístico utilizando el programa Redatam. Se dará énfasis al procesamiento y análisis de información proveniente de censos de población y vivienda y de encuestas de hogar con prioridad en la generación de indicadores en las siguientes áreas temáticas: déficit en educación, estimación del déficit habitacional, caracterización de grupos específicos de población (tercera edad, jóvenes adolescentes, poblaciones indígenas) y sus características sociodemográficas, migración y estimación de Necesidades Básicas Insatisfechas.

El taller está organizado en 4 (cuatro) reuniones diarias de 8 (ocho) horas de duración, realizándose primero el dictado de clases teóricas y, posteriormente, ejercicios prácticos grupales e individuales. A continuación se detalla el cronograma de las cuatro reuniones:

Día 1

A: TEMA: PRESENTACION DE LA BASE DE DATOS CENSAL, TRABAJANDO CON LAS VARIABLES y EL CUESTIONARIO

OBJETIVOS

- Entender la base de datos del censo
- Analizar la potencialidad de la herramienta.
- Entender el concepto de estructura jerárquica
- Manejo de base de datos
- Entender la estructura geográfica.

El censo de población y vivienda

Como se crea una base de datos a partir del censo

Características de las variables

Como estructuro la geografía de un país en forma jerárquica

Características de Redatam y sus módulos

El censo de población y vivienda (aspecto operativo)

El método utilizado por los Censos de Población y Vivienda para obtener la información es la localización de las viviendas, en las que se identifica a los hogares y personas que las componen. Para cada hogar que compone la vivienda se dispone de un cuestionario censal donde el empadronador registrará los datos de la vivienda a la que pertenece el hogar; datos del hogar en cuestión e información para cada una de las personas que lo integran, todo esto a partir de las respuestas entregadas por el jefe de hogar o, en su defecto, otra persona calificada para responder a las preguntas que se le formulen.

La información censal es relevada considerando a todos quienes pernoctaron en el hogar la noche anterior al censo (Censo de hecho) o a los residentes habituales (Censos de Derecho), modalidad, esta última, que permite extender el periodo de levantamiento censal y que corresponde a la adoptada en el Censo demográfico de 2011.

Código geográfico

Una información de vital importancia para la creación de la base de datos numérica que se construirá después de recolectar y validar toda la información del país y para cada una de sus divisiones político administrativas (Departamentos) y las que corresponden a la organización operativa del evento censal (Secciones y Segmentos censales), es la que da cuenta de la localización en un espacio geográfico determinado, de cada una de las viviendas que fueron objeto de la aplicación del cuestionario. De esta manera es posible, que una vez construida la base de datos censal, se pueda recuperar la información recogida hasta el nivel más bajo de desagregación geográfica, que en el caso de la base censal de Uruguay 2011 corresponde al Segmento Censal. Por razones de ley y de protección del secreto estadístico que asegura la inviolabilidad de la intimidad de las personas, la base de datos no permite la individualización de una vivienda, hogar y personas en particular. Esta información compuesta de códigos numéricos es la que da origen al código geográfico que, según se verá más adelante, posee una relevancia especial al momento de trabajar con dicha base de datos.

Estructura de la base de datos censal

Después de los procesos de validación pertinentes, la información se organiza de manera jerárquica dando origen a una estructura como la que se ilustra en la figura de la izquierda.

“CPV2011” es el nombre que define la base de datos en REDATAM+SP y representa a la entidad padre o “país”. Esta contiene todas las entidades de menor nivel jerárquico desde “Departamento” hasta “persona”. Ahora bien, dada esta relación de dependencia entre una entidad y otra, las variables que definen características de una entidad determinada, definen al mismo tiempo las características de todos los elementos que se encuentran en un nivel de jerarquía menor. Por ejemplo, la variable “VIVIENDA.DV05” pertenece a la entidad “VIVIENDA” y los atributos que respecto a esta variable posea una vivienda o conjunto de viviendas en particular son también atributos que describen características de los elementos que le pertenecen. Si la vivienda posee agua proveniente de la red general de agua potable, entonces todos los hogares y personas que pertenecen a dicha vivienda comparten el mismo origen del agua declarado para la vivienda.

REDATAM+SP hace un uso total de las propiedades que presenta la estructura jerárquica del Censo de Población y Vivienda, posibilitando por ejemplo determinar la distribución de la población según sexo, de acuerdo a características de la vivienda o del hogar, de tal manera que es muy rápido y sencillo conocer cuántos hombres y mujeres habitan en viviendas cuyo origen del agua que usan para beber corresponda a “pozo surgente no protegido”. El aprovechamiento de esta particular organización de los datos censales con fines de análisis

estadísticos depende del conocimiento y manejo que los usuarios tengan de las variables y categorías que forman parte de cada entidad ubicada en la estructura jerárquica, de ahí que antes de continuar con el conocimiento de REDATAM+SP resulte beneficioso detenerse un momento y “navegar” por la base de datos del Censo de 2011 ayudados por la boleta o cuestionario censal.

Ejercicios

Revise en el cuestionario censal las preguntas sobre Actividad Laboral. A quienes se pregunta? Cuantas posibles respuestas tiene la pregunta? Cuales son las categorías? Cómo podríamos determinar la población ocupada?

B: TEMA: REDATAM+ SP COMO UNA HERRAMIENTA DE PROCESAMIENTO ESTADISTICO

OBJETIVOS

- Adquirir la capacidad de analizar los datos censales a partir de tabulados : producir frecuencias, tablas cruzadas y promedios en forma simple
- Entender el uso básico de un programa en REDATAM+SP.
- Poder ejecutar programas simples

Tabulados fáciles en Redatam

Selecciones geográficas

Programación de procesos en Redatam

Comando TABLE

Comando DEFINE: Derivando Nuevas Variables – RECODE

Lo primero que usted debe realizar en una sesión es abrir un diccionario (archivo .dic). Inmediatamente se abre una ventana con la estructura jerárquica de la base de datos a mano izquierda y las variables a mano derecha.

Navegar A Través De REDATAM+SP

1. Teniendo activa la ventana de inicio de REDATAM+SP, aún sin haber abierto un diccionario, el sistema ofrece una barra de menús y funciones relacionadas con su sesión de trabajo:
 - Icono para abrir un diccionario
 - Icono para abrir ventana de preferencias
 - Icono de ayuda
 - Menú archivo
 - Menú diccionario
 - Proceso de Estadístico

- Menú de Herramientas
 - Menú de ventana
 - Menú de ayuda
2. Activar cada icono y desplegar las distintas opciones de menús haciendo un breve comentario de cada opción (No se requiere entrar en detalles por cuanto ello será abordado en los módulos subsiguientes)
 3. Abrir el diccionario de la base de datos de ejemplo Nueva Miranda para continuar la navegación por REDATAM+
 4. Desplegar las opciones de cada menú haciendo un breve comentario sobre ellas.

La ventana de diccionario de REDATAM+SP

Como usted puede observar, la ventana de diccionario se encuentra dividida en dos secciones. En la sección de la derecha (o superior) se despliega toda la estructura de la base de datos del censo. Una barra de color azul en posición horizontal indica la entidad activa. En la sección de la derecha (o inferior) se despliegan las variables que describen características de cada elemento perteneciente a la entidad resaltada por la barra de color en la sección izquierda.

Usted puede, al posicionar el puntero del ratón en cualquier entidad hacer clic sobre ella, verificar cómo el contenido de la sección derecha cambia en relación a la entidad resaltada. Esto le permite verificar rápidamente las variables que pertenecen a cada entidad (la entidad define el nivel de la variable, por lo que podemos decir que la variable "PH02" que describe el sexo de las personas se encuentra al nivel de persona).

Usted puede seleccionar una variable y revisar algunas características de la variable seleccionada activando la opción "propiedades" del menú que se abre con el botón derecho del mouse (rótulo, categorías, rangos mínimo y máximo, disponibilidad de la variable y alguna otra información que le pudiese ser de utilidad).

ARCHIVOS QUE SE MANEJAN EN REDATAM+SP

- **PROGRAMAS:** Conjunto de instrucciones que el usuario a empleado con el fin "pedirle" al sistema la obtención de estadísticas provenientes de la base de datos. Estos pueden ser escritos directamente por el usuario en el lenguaje exclusivo de REDATAM+SP (formato libre), o ser generados por REDATAM+SP utilizando el asistente de programas. Los programas se guardan en archivos con la extensión .spc.
- **SELECCIONES:** Para la ejecución de cualquier programa u obtención de tabulados, REDATAM+SP requiere que el usuario defina, previamente, el área geográfica a ser procesada. Esta área puede corresponder a todo el país o a cualquier agrupación de entidades geográficas seleccionables, como ser Departamento, Secciones o Segmentos censales; que corresponde al nivel más desagregado para el procesamiento de la información censal a partir de la estructura de la base. De manera predeterminada, REDATAM+SP tiene seleccionada toda la base de datos con la que se trabaje, el usuario puede hacer referencia a esta selección con la palabra clave "ALL".
- El usuario tiene la posibilidad de guardar las selecciones que realice durante su trabajo, en cuyo caso REDATAM+SP genera un archivo con la extensión ".slw" o ".sel" el cual se almacena donde el usuario defina, con el fin de que puedan ser utilizados cuando se requieran y sin necesidad de repetir el mismo proceso de selección.

- **COMPOSICIÓN DE MAPAS:** Aquí el sistema guarda todas las indicaciones que le entrega el usuario con el fin de utilizar la capacidad de REDATAM+SP para vincular o empalmar la información de la base de datos con un mapa, permitiendo la obtención de mapas temáticos que faciliten la focalización de sectores o áreas geográficas que presenten características de interés para el investigador. La composición de mapas se guarda en un archivo con la extensión .mxp.
- **BASE DE DATOS:** Compuesta por varios archivos, primero los punteros (ptr) que definen la pertenencia geográfica de los registros en toda la base jerárquica y luego las variables (rbf) que se almacenan una por una en un archivo vectorial. La base de datos se accede a través del diccionario con los metadatos (dic).

Selección geográfica

Es la colección de elementos seleccionados de las entidades jerárquicas. Para la ejecución de cualquier programa u obtención de tabulados, REDATAM+SP requiere que el usuario defina, previamente, el área geográfica que va a procesar.

El nombre “Selección” es usado, de manera obvia como “Selección de Área ” y referida a delimitaciones geográficas, sin embargo una selección no tiene porqué limitarse sólo a límites de carácter espacial porque siempre es posible estructurar una base de datos con criterios no geográficos para definir una selección. Por ejemplo, una base de datos de estadísticas vitales podría estar organizada por año, en cuyo caso se tendría una selección referida a un(os) intervalo(s) de tiempo determinado.

El usuario tiene la posibilidad de guardar las selecciones que realice durante su trabajo, en cuyo caso REDATAM+SP genera un archivo con la extensión .slw, que almacena bajo el directorio de trabajo, con el fin de que puedan ser utilizados cuando se requieran y sin necesidad de repetir el mismo proceso de selección. De manera predeterminada, REDATAM+SP tiene seleccionada toda la Base de Datos con la que se trabaje; el usuario puede hacer referencia a esta selección con la palabra clave “ALL”.

Usando el editor de selección geográfica

Pasos para crear una selección

- Para definir una selección se debe abrir la ventana de selección mediante la opción **Archivo>Nuevo>Selección**.
- En la ventana que se abre aparece la estructura de base de datos, empezando por la rama superior, es decir, el nivel más agregado de la base de datos.
- Utilizando la opción expandir, activada con el botón derecho del ratón, abra esta rama en su segundo nivel.
- Los elementos de este segundo nivel (por lo general corresponden a entidades geográficas), son desplegados a través de su código numérico, o si corresponde, a través de su nombre.
- Repita el proceso de **expandir** los niveles en sus niveles inferiores.
- Localice el cursor sobre el elemento o entidad geográfica que desea seleccionar



- Marque la selección haciendo doble click con el ratón o bien utilizando la opción **seleccionar** desde el menú que se abre con el botón derecho del ratón. El cuadrado al lado del nombre de la entidad seleccionada se debe poner de color amarillo.
- Repita el proceso de selección en todas aquellos elementos que desee incluir en el nuevo conjunto de selección.
- Si se selecciona un elemento que no ha sido expandido en elementos inferiores (no tiene el símbolo a un costado), todos los elementos bajo éste nivel son seleccionadas a su vez.
- Para contraer una rama expandida junto con todos los elementos que contiene, localice el puntero del ratón sobre la esquina del árbol que presenta un signo y haga un click sobre él.
- Para guardar los elementos seleccionados elija, desde el menú principal, la opción Archivo>Guardar. Dentro de la ventana de diálogo que se despliega reemplace el *.slw por el nombre del archivo elegido por usted.

Notas:

1. Para verificar el alcance de la Selección de áreas ejecute una frecuencia para una variable al nivel más bajo posible y verifique el número total de elementos.
2. Las áreas parcialmente seleccionadas aparecen con la mitad de su cuadrado coloreado de amarillo.
3. Sólo las áreas seleccionables pueden ser expandidas.

Ejercicios

1. Crear una Selección de áreas Para el Departamento Canelones.
2. Crear una selección que incluya las dos primeras secciones de los Departamentos de Flores y Río Negro.

PRODUCIENDO RESULTADOS BÁSICOS - TABULACIÓN: LA MANERA SENCILLA

REDATAM+SP incluye unas cuantas herramientas, fáciles de usar, que proveen al usuario de una sencilla y rápida forma para realizar tabulaciones básicas tales como FRECUENCIAS, CRUCES, PROMEDIOS y CONTEOS simples.

Estas herramientas se pueden utilizar sin tener la necesidad de definir, verificar y ejecutar un programa.

Cada una de estas herramientas es en efecto una interfaz que genera automáticamente un programa “detrás de escena”. El programa se ejecuta y retorna automáticamente los resultados al usuario.

Para cargar la herramienta apropiada ya sea para calcular una frecuencia, tabla cruzada , promedio o un simple conteo, haga un clic en el menú de proceso estadístico de la ventana del Espacio de Trabajo.

Demostración

Identificar como se distribuye la población del país según su estado civil, edad, y sexo.

- Desde el menú principal seleccione, bajo el menú **Proceso Estadístico**, la opción que refleje el proceso que se desea ejecutar: FRECUENCIA,
- Desde la ventana del diccionario seleccione con el ratón la variable que desea procesar,
- Tome la variable (el nombre) y arrástrela hacia el casillero en blanco dentro de la ventana de **Definición del Proceso Estadístico**.
- Llene el (los) casillero (s) con la(s) variable(s) que desee procesar, dependiendo si se trata de una frecuencia, cruce de variables o promedio.
- • Ejecute el proceso estadístico haciendo un clic sobre el icono  desde la barra de iconos.
- Espere que el proceso finalice y revise los resultados presentados en la ventana de salida.

Ejercicios

Determine las siguientes distribuciones a nivel del País:

1. Viviendas, según el material de construcción predominante en sus pisos
2. Viviendas, según el material de construcción predominante en la cubierta del techo
3. Viviendas, según origen del agua usada para beber
4. Hogares, según fuente de energía utilizada para cocinar.
5. Personas, según su relación de parentesco con el jefe de hogar
6. Personas, según su estado civil actual

PRODUCIENDO CRUCES Y PROMEDIOS

Demostración

- A) Determinar cómo se distribuye la población según sexo y tipo de vivienda, a nivel de País
- Desde el menú principal seleccione, bajo el menú Proceso Estadístico, la opción que refleje el proceso que se desea ejecutar: CRUCE
 - Desde la ventana del diccionario seleccione con el ratón la variable que desea procesar,
 - Tome la variable (el nombre) y arrástrela hacia el casillero en blanco dentro de la ventana de Definición del Proceso Estadístico.
 - Llene el (los) casillero (s) con la(s) variable(s) que desee procesar, dependiendo si se trata de una frecuencia, cruce de variables o promedio.
 - Ejecute el proceso estadístico haciendo un clic sobre el icono  desde la barra de iconos.

- Espere que el proceso finalice y revise los resultados presentados en la ventana de salida.

Ejercicios

1. Utilizando la ventana de Cruces identifique la relación entre origen del agua en la vivienda y la composición por sexo de sus residentes.
2. Utilizando la ventana de Cruces identifique la relación entre las variables trabajo la semana pasada y máximo nivel educacional alcanzado.
3. Utilizando la ventana de Promedios, determine el promedio de edad de las personas de acuerdo al parentesco con el jefe del hogar

Día 2

TEMA: GENERACIÓN DE INDICADORES DE EDUCACIÓN Y DEMOGRÁFICOS

OBJETIVOS

- Adquirir capacidad de producir indicadores básicos
- Comprender los conceptos básicos del procesamiento jerárquico.
- Comprender y adquirir la capacidad para realizar procesos de conteo de elementos.
- Comprender el proceso de generación y adición de nuevas variables al diccionario de la Base de Datos.

Utilizando UNIVERSE y AREABREAK

Promoviendo Información – El Comando COUNT

Condicionar el conteo con un filtro - FOR

Resultados según Listas de Áreas

Haciendo uso del editor de programas de REDATAM+SP avanzamos en la obtención de los siguientes indicadores:

1. **Índice de masculinidad:** Consiste en la razón de varones entre mujeres en una determinada población o sub-población. Se considera como el principal indicador, para analizar la distribución por sexo en la población.
2. **Proporción de masculinidad:** Como su nombre lo indica, corresponde a la proporción de varones entre la población total. Al igual que la medida anterior, permite conocer la distribución de la población por sexo.
3. **Relación niños-mujer:** Esta relación se utiliza como un indicador aproximado de la fecundidad, especialmente cuando no se disponen datos de calidad sobre los nacimientos. Corresponde a la razón de los niños menores de 5 años, por cada 1.000 mujeres en edad reproductiva (edades cumplidas entre 15 y 49 años).

4. **Tasa de alfabetismo:** Es un índice del nivel educativo de un pueblo. Aunque se le conoce como tasa, en realidad es una proporción. Se obtiene del cociente de la población alfabetada de 10 años y más, entre el total de la población de 10 años y más.
5. **Relación de dependencia:** Esta medida, desde el punto de vista demográfico, se define como el cociente de la población económicamente dependiente, que corresponde a los menores de 15 años y junto con las personas de 65 años o más, sobre la población con edades cumplidas entre 15 y 64 años. Sin embargo, existe otra definición de esta medida, que corresponde al concepto económico del término. En ella, se varían las definiciones de la población económicamente activa (PEA) y de la económicamente dependiente. No obstante, su fórmula de cálculo es similar.
6. **Relación de reemplazo de la población en edad activa:** Esta medida al igual que la anterior, tiene algunos problemas en la definición de los términos. Generalmente, se considera como la relación entre el grupo de edades entrante a la población activa (grupo de 15-19) con respecto al grupo saliente de esta población (grupo 60-64). Su valor representa la capacidad de población económicamente activa de reemplazarse.

Día 3

TEMA: GENERACIÓN DE INDICADORES DE SITUACIÓN LABORAL

OBJETIVO: Conocer la tasa de participación laboral y la tasa de desempleo que presenta la población por grupos de edad según características educativas y área de residencia

Comando DEFINE: Utilizando Expresiones Aritméticas (calcula de tasas y %)

Promoviendo Información - El comando SWITCH

Guardar la variable: SAVE

Agregación de un indicador según niveles geográficos

Definir variables para análisis de la situación laboral

Indicadores:

1. Tasa de participación laboral en jóvenes de 19 a 29 años:

Definición: Mide la proporción de jóvenes que se encuentran trabajando o buscando trabajo, en relación al total de jóvenes. Proporciona un primer acercamiento respecto a la inserción juvenil en el ámbito laboral. Se calcula obteniendo el cociente entre los jóvenes que se encuentran en la PEA dividido por el total de Jóvenes, por cien.

Desagregación: Total y nivel de instrucción por área Urbano - rural

2. Tasa de desempleo en jóvenes de 19 a 29 años:

Definición: Mide la proporción de jóvenes que se encuentran sin trabajo. Permite un acercamiento más refinado a la situación laboral que afecta a la población joven. Se calcula como cociente entre los jóvenes desocupados sobre el total de jóvenes en la PEA, por cien.

Desagregación: Total y nivel de instrucción por área Urbano – rural

Día 4

A: TEMA: CARENCIA DE VIVIENDAS (NBI)

OBJETIVO: Conocer la tasa de participación laboral y la tasa de desempleo que presenta la población por grupos de edad según características educativas y área de residencia

Comando DEFINE: Utilizando Expresiones Aritméticas (calcula de tasas y %)

Promoviendo Información - El comando SWITCH

Guardar la variable: SAVE

Agregación de un indicador según niveles geográficos

Mapeo de los NBI

Indicadores:

1. Viviendas en situación de hacinamiento

Definición: Mide el grado de hacinamiento de las viviendas particulares ocupadas. Se calcula como las viviendas con más de 3 personas por dormitorio dividido por el total de viviendas, multiplicado por 100. Desagregación: por área Urbano – rural

2. Mala calidad de Techos, Paredes o Pisos

Definición: Es un indicador de falta de acceso vivienda de calidad; mide la proporción de viviendas construidas con materiales deficientes en sus paredes, techos o pisos. Se calcula como las viviendas con materiales inadecuados dividido por el total de viviendas, multiplicado por 100. Desagregación: por área Urbano – rural

3. Porcentaje de viviendas sin acceso a agua potable:

Definición: Es un indicador de falta de acceso a agua potable; mide la proporción de viviendas que cuentan con abastecimiento de agua inadecuado. Se calcula como las viviendas con abastecimiento de agua inadecuado dividido por el total de viviendas, multiplicado por 100.

Desagregación: por área Urbano – rural

4. Porcentaje de viviendas sin acceso a sistema de alcantarillado:

Definición: Es un indicador de falta de acceso a algún sistema de saneamiento; mide la proporción de viviendas sin baño, sin sistema o inadecuado sistema de evacuación de excretas. Se calcula como las viviendas sin acceso a saneamiento dividido por el total de viviendas, multiplicado por 100.

Desagregación: por área Urbano – rural

B: TEMA: MAPAS TEMATICOS

OBJETIVO: Conocer como se traspasan los indicadores a mapas temáticos en Redatam y exportar a otro mapeador como gvSIG

Uso básico de gvSIG

Mapas en Redatam

Tablas de exportación DBF

Código de enlace : JOIN

La versión 5 de Redatam permite desplegar en mapas las variables de la base de datos que estén asociadas a un código geográfico. Esta posibilidad es de fundamental importancia para la planificación, ya que una observación cartográfica permite a primera vista saber cuales son las áreas que tienen un comportamiento similar y cual un comportamiento diferente en la variable asociada. Incluso colocando distintas capas de cartografía puede observarse como accidentes geográficos tales como ríos, cadenas de montañas, etc. impactan en el comportamiento de las variables. (Por ejemplo, la dificultad de acceso puede ser relevante a la hora de explicar distintas tasas de escolaridad).

Una vez abierta la base de datos sobre la que se está trabajando (el diccionario de la base de datos), el procedimiento para asociar un mapa a la base consiste en armar una composición de mapa, la que se compone de mapa, título, leyenda y textos mas la conexión interna a la base de datos. Esta composición de mapa se almacena en un archivo con la extensión .mxd.

Trabajar con mapas en R+SP presupone la existencia de estos mapas en el formato Shape de ArcView™. El sistema R+SP NO construye mapas digitalizados, apenas convierte el formato SHAPE al formato interno de trabajo de R+SP.