



UNIVERSIDAD DE ATACAMA

FACULTAD DE INGENIERÍA / DEPARTAMENTO DE MATEMÁTICA

ESTADÍSTICA Y PROBABILIDADES

Ejercicio 1: Comenzando con R y R Commander

Profesor: Hugo S. Salinas.

Primer Semestre 2011

1. Entrada de Datos:

La mayoría de los procesos en **R Commander** asumen que hay un conjunto de datos activo. Si hay varios conjuntos de datos en la memoria, podrás elegir entre ellos, pero sólo uno estará activo. Cuando **R Commander** se inicia, no hay ningún conjunto de datos activo. **R Commander** proporciona varias maneras de introducir datos en R:

- Puedes introducir los datos directamente a través de **Datos->Nuevo conjunto de datos...** Esta es una opción razonable para un conjunto de datos pequeño.
- Puedes importar datos desde un archivo de texto sin formato (ascii) o desde el portapapeles, desde otro paquete estadístico (Minitab, SPSS o Stats), o desde un conjunto de datos de Excel, Access o dBase.
- Puedes leer un conjunto de datos que esté incluido en un paquete **R**, ya sea escribiendo el nombre del conjunto de datos (si lo conoce) o seleccionando el conjunto de datos en un cuadro de diálogo.

2. Lectura de datos desde un paquete:

Muchos paquetes **R** contienen datos. Los conjuntos de datos en paquetes pueden listarse en una ventana desplegable a través de **Datos->Conjunto de datos en paquetes->Leer conjunto de datos desde paquete adjunto**. Si sabes el nombre de un conjunto de datos en un paquete, entonces puedes introducir el nombre directamente; en caso contrario, al hacer una doble clic sobre el nombre de un paquete se muestra su conjunto de datos listado en el cuadro de la derecha; y haciendo una doble clic sobre el nombre del conjunto de datos se copia el nombre al campo de entrada del conjunto de datos. Puedes adjuntar paquetes **R** adicionales en **Herramientas->Cargar paquete(s)...**

✓ Cargar la base de datos **Chile** que se encuentra en el paquete **car**.

3. Crear resúmenes numéricos y gráficas:

Una vez que hay un conjunto de datos activo, puedes usar los menús de **R Commander** para generar varios resúmenes numéricos y gráficas. Describiremos simplemente unos ejemplos básicos a continuación. En los ejemplos de más abajo, asumimos que el conjunto de datos activo es **Chile**, leído en la sección anterior.

✓ Seleccionar **Estadísticos->Resúmenes->Conjunto de datos activo**.

Observar que para cada variable numérica en el conjunto de datos (population, age, income y statusquo), **R** informa de los valores máximos y mínimos, el primer y el tercer cuartil, la mediana, la media, así como el número de valores que faltan. Para las variables categóricas (region, sex, education y vote), obtenemos el número de observaciones en cada *nivel* del factor y el número de datos faltantes.

De manera similar, seleccionando **Estadísticos->Resúmenes->Resúmenes numéricos...**, solamente las variables numéricas se muestran en la lista de variables; los factores region, sex, education y vote faltan, porque no es razonable calcular resúmenes numéricos para estas variables.

Por defecto, la instrucción **R** que se ejecuta calcula la media y la desviación estándar (sd) de la variable, junto con los cuantiles correspondientes con el mínimo, el primer cuartil, la mediana, el tercer cuartil y el máximo; *n* es el número de observaciones válidas, y NA es el número de valores que faltan.

El cuadro de diálogo resúmenes numéricos también tiene en cuenta para calcular resúmenes a los grupos definidos por niveles de un factor. Pulsando en el botón **Resumir por grupos...** se abre el diálogo Grupos. Aquí puedes seleccionar una variable (education, region, sex o vote).

✓ Calcular los resúmenes numéricos para la variable income agrupados por region.

Hacer gráficas con **R Commander** también es simple.

Por ejemplo, si seleccionas **Gráficas->Histograma...** en los menús de **R Commander** se abre el cuadro de diálogo Histograma y deberás seleccionar una de las variables numéricas (age, income, population o statusquo).

✓ Hacer un histograma de la variable income.

Si eliges **Gráficas->Diagrama de cajas...** se abre el cuadro de diálogo Diagrama de cajas y deberás seleccionar una de las variables numéricas. Además puedes Identificar los valores atípicos con el ratón o hacer un diagrama de cajas por grupos....

✓ Hacer un diagrama de cajas para la variable income.

✓ Hacer un diagrama de cajas para la variable income separada por género.

4. Renombrar las variables:

Una vez que hay un conjunto de datos activo, puedes utilizar los menús **Datos->Modificar variables del conjunto de datos activo->Renombrar variables...** para cambiar el nombre de las variables, por ejemplo llevarlas al español o simplemente cambiar por un nombre más representativo.

✓ Cambiar el nombre de las variables age, education, income, population, sex y vote, por edad, educacion, renta, poblacion, genero y voto, respectivamente.

5. Eliminar variables:

Una vez que hay un conjunto de datos activo, puedes utilizar los menús **Datos->Modificar variables del conjunto de datos activo->Eliminar variables del conjunto de datos...** para eliminar la o las variables de la base de datos.

✓ Eliminar la variable statusquo de la base de datos Chile.