

Primeros pasos con R-Commander

En esta práctica, aprenderemos el manejo de R-Commander, un paquete adicional de R que proporciona una interfaz gráfica para un uso simple de los comandos básicos de R. Fue diseñado con fines de docencia y permite una primera aproximación a R, pero no se puede obviar que

1. Primeros pasos.

Empezaremos por:

1. Ejecutar R
2. Cambiar el directorio de trabajo al directorio que hayamos escogido (por ejemplo Mis Documentos)
3. En el caso en que no lo tengamos hecho, descargar el fichero `datosprimerasesion.Rdata` desde el Aula Virtual (en Datos necesarios para las prácticas) y lo copiamos en el directorio de trabajo.
4. Limpiar el entorno de trabajo de cualquier objeto heredado de una sesión anterior:

`rm(list=ls())`

5. Cargar el área de trabajo contenido en el fichero `datosprimerasesion.Rdata`
6. Comprobamos los objetos que tenemos definidos con la instrucción `ls()`

2. Lanzamos R-Commander

Suponemos que ya tenemos instalado el paquete `Rcmdr` en nuestro ordenador (se pueden encontrar instrucciones de instalación en el mini-manual llamado **Manual de instalación de R y R Commander**, disponible en el Aula Virtual.)

Para ejecutar R-Commander, basta con introducir en la línea de comandos:

```
library(Rcmdr)
```

Se abre una interfaz gráfica, que usaremos a partir de ahora para interactuar con R.

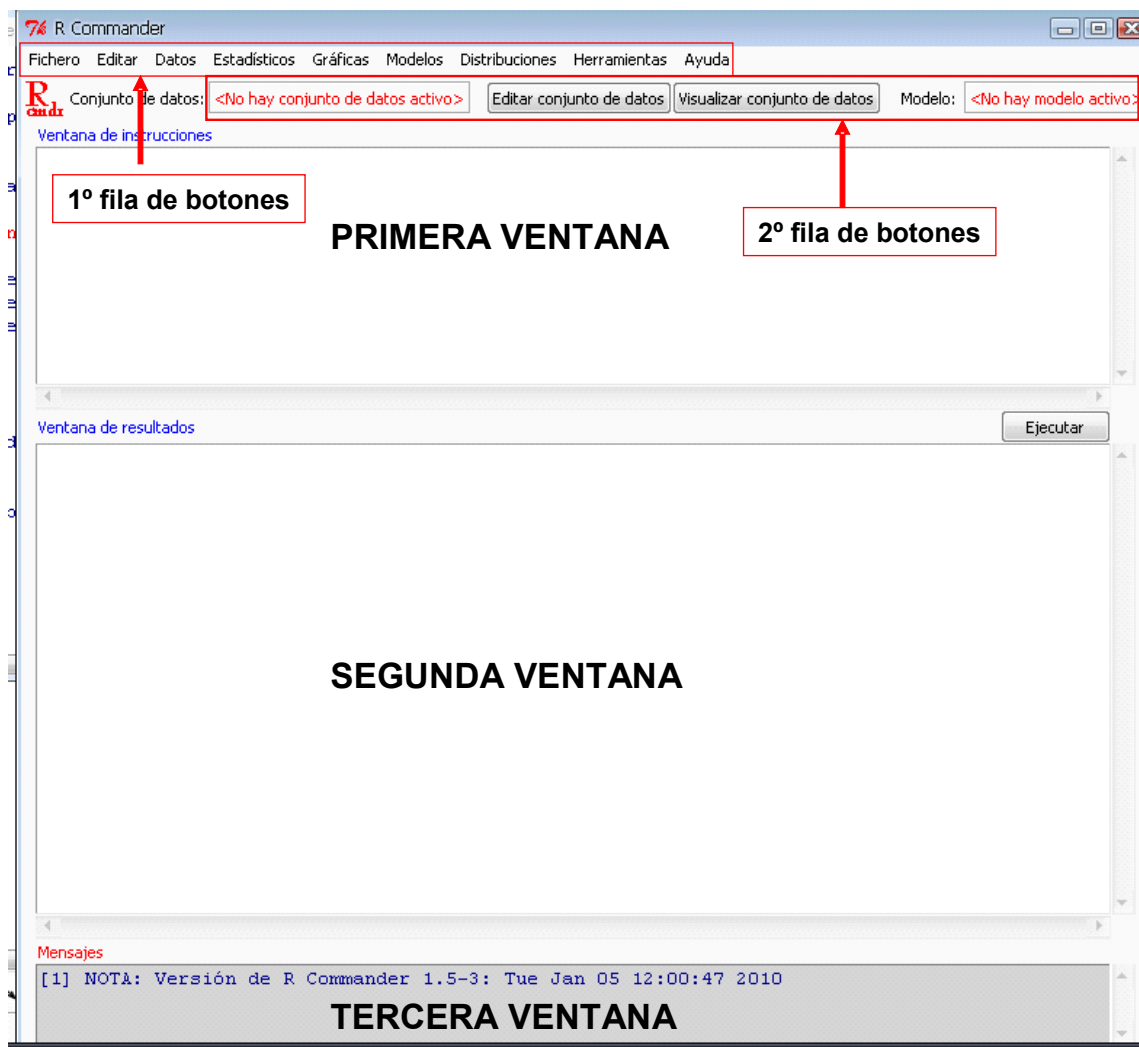
NOTA: Aunque tengamos R-Commander abierto, siempre podemos usar R en modo línea de comandos.

3. Estructura de la interfaz gráfica de R-Commander

La interfaz gráfica está estructurada en 5 elementos:

- Dos filas de botones
- Tres ventanas.

Estos 5 elementos están marcados en la figura siguiente:



Primera Ventana: ventana de instrucciones

En esta ventana, irán apareciendo las **instrucciones** de R que se generarán automáticamente a medida que vayamos escogiendo opciones recorriendo la primera fila de botones.

También cabe la posibilidad de introducir manualmente una instrucción de R como si estuviéramos en la línea de comandos y ejecutarla haciendo click en Ejecutar (Podemos incluso ejecutar varias líneas a la vez si las seleccionamos con el ratón antes de darle a Ejecutar)

Segunda Ventana: ventana de resultados

En esta ventana, irán apareciendo los **resultados** de las instrucciones de R que se hayan generado.

Tercera Ventana: ventana de mensajes

En esta ventana, irán apareciendo mensajes de error o de aviso que genere R.

Primera fila de botones:

Usaremos los distintos menús de esta fila de botones para realizar cálculos y gráficas para un conjunto de datos.

Segunda fila de botones:

En esta fila de botones podemos:

- Seleccionar el conjunto de datos activo, sobre él que realizaremos los cálculos o gráficas que deseemos.
- Visualizar el conjunto de datos activo (no recomendable en el caso de conjuntos grandes)
- Editar el conjunto de datos activo para hacer pequeñas modificaciones (no recomendable para conjuntos grandes)
- Si hemos llevado a cabo varios ajustes (lineales por ejemplo) tenemos la posibilidad de escoger uno de estos ajustes (“modelo”) para obtener información de resumen.

4. Selección y exploración de un conjunto de datos con R-Commander.

4.1 Seleccionamos el conjunto escombrerashoras09:

Basta con escogerlo de la lista desplegable que aparece cuando hacemos clic en el botón de la segunda fila <No hay conjunto de datos activo>.

Al haberlo escogido, aparece en la ventana de Mensajes (tercera ventana) la dimensión del conjunto.

4.2 Obtenemos un resumen numérico de las variables presentes en el conjunto.

Para ello, en la primera fila de botones, usamos el comando

Estadísticos > Resúmenes > Conjunto de datos activo.

El significado de las abreviaturas es:

Min	Valor mínimo de la variable
1st Qu.:	Primer cuartil
Median	Mediana
Mean	Media
3rd Qu.:	Tercer Cuartil
Max.:	Valor máximo de la variable

Contestar a las siguientes preguntas:

- a) ¿Cuál ha sido la temperatura máxima que se dio durante el periodo que duraron las mediciones?
- b) ¿Cuál ha sido el valor medio de SO₂ durante el periodo de medición?

- c) ¿Cuál ha sido el valor mínima de la temperatura que se registró? ¿Algún comentario? ¿Cuál es la influencia de este valor sobre la fiabilidad de la temperatura media que se puede calcular?
- d) Para la velocidad del viento, dar un valor límite debajo del cuál están situados el 25% de los valores medidos.

4.3 Obtenemos un resumen numérico de algunas de las variables pero distinguiendo distintos grupos.

Pasamos ahora a calcular distintos resúmenes numéricos de las variables, pero desglosando sus valores según grupos naturales.

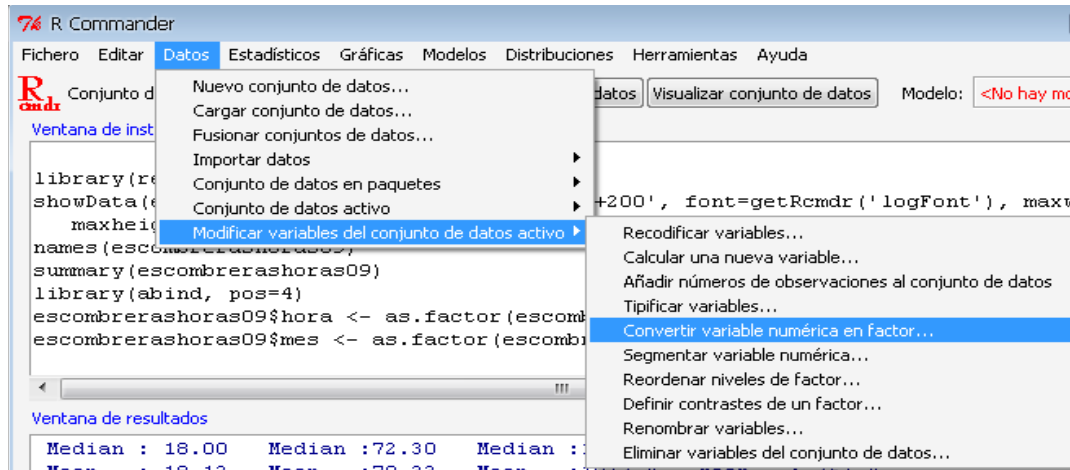
Queremos por ejemplo conocer el valor medio de la temperatura calculado por meses.

Para ello, debemos empezar por especificar a R que algunas de las variables del conjunto son factores (cuyos niveles definen grupos)

4.3.1 Convertimos las variables mes y hora en factores

Para ello, usaremos el comando:

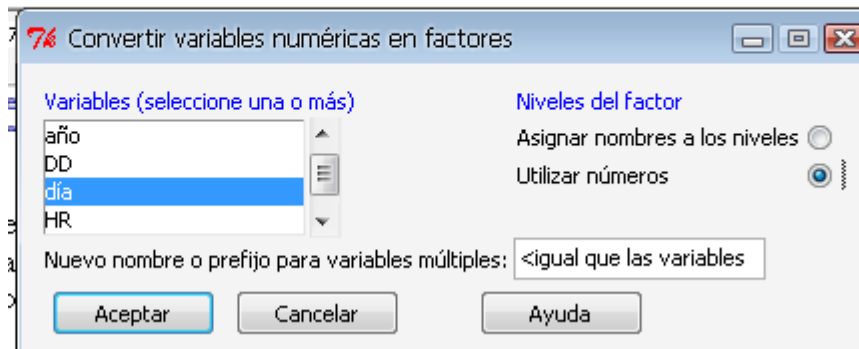
Datos > Modificar variables del conjunto de datos activo > Convertir variables numéricas en factor



Seleccionamos las variables hora y mes en el listado de variables (mantenemos el botón Ctrl pulsado para seleccionar más de una variable a la vez)

Nota: Si no aparecen las variables que queremos convertir, es porque ya están definidas como factores.

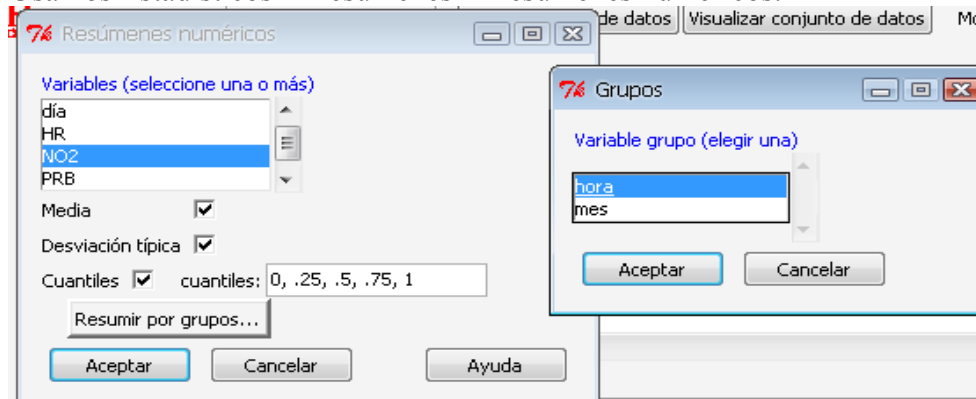
Escogemos el botón Utilizar números, porque no queremos dar nombres personalizados a los niveles:



Aceptamos.

4.3.2 Calculamos resúmenes numéricos desglosando los resultados según los niveles de los factores hora y mes.

Usamos Estadísticos > Resúmenes > Resúmenes numéricos.



Seleccionamos la variable NO2, y al pinchar en Resumir por grupos, escogemos la variable grupo hora.

¿Varía la concentración de NO2

Contestar a las preguntas siguientes:

- ¿Qué hora del día presenta el mayor valor promedio de concentración de NO2?
- ¿Cambia la desviación típica según la hora del día? ¿Cómo se interpretaría?
- ¿En qué horas se produjeron las mayores concentraciones de NO2?

4.4. Obtenemos histogramas.

4.4.1 Histograma básico

Vamos a realizar un histograma de la variable velocidad del viento: VV.

Para ello, usamos el comando, en la primera fila de botones:

Gráficas > Histograma

Escogemos la variable VV, tenemos la opción de especificar el número de clases y además podemos escoger el tipo de frecuencias en el eje Oy: frecuencias absolutas, porcentajes o densidades (la altura es la frecuencia dividida por el ancho de la clase).

4.4.2 Modificar algún aspecto del histograma

R tiene funcionalidades gráficas muy flexibles y potentes, pero R-Commander, en su búsqueda de la simplicidad de uso, implementa las instrucciones gráficas de forma básica con las opciones más elementales.

Sin embargo, es posible modificar, o añadir en la Ventana de instrucciones, cualquier opción, (argumento) que admita el comando de R y volver a ejecutar la línea.

Como ejemplo, vamos a modificar el color de las barras del histograma, cambiar las etiquetas de los ejes, y añadir un título principal al histograma.

La instrucción que generó el histograma tal como aparece en el ventana de instrucciones es:

```
Hist(escombrerashoras09$VV, scale="frequency", breaks="Sturges", col="darkgray")
```

a) Cambio del color

El color se controla en una gráfica con el argumento col. Sustituimos col="darkgray" por col="blue":

```
Hist(escombrerashoras09$VV, scale="frequency", breaks="Sturges", col="blue")
```

y le damos a Ejecutar (debemos asegurarnos de que nuestro cursor está situado en la línea de la instrucción que queremos ejecutar).

Observamos cómo se cambia el color del histograma que realizamos.

b) Cambio de etiquetas de los ejes

Vemos en nuestra gráfica que los ejes del histograma tienen por etiquetas "frequency" y "escombrerashoras09\$VV". Vamos a cambiarlas por algo más entendibles:

Para especificar la etiqueta del eje Ox, el argumento que usamos es xlab.

Para especificar la etiqueta del eje Oy, el argumento que usamos es ylab.

Añadimos por lo tanto estos argumentos al comando Hist en la ventana de instrucciones: (toda la instrucción en una única línea)

```
Hist(escombrerashoras09$VV, scale="frequency", breaks="Sturges",  
col="blue",xlab="Velocidad del viento", ylab="Frecuencias")
```

y ejecutamos la línea.

c) Añadir un título a la gráfica:

Para ello, usamos el argumento main. (toda la instrucción en una única línea):

```
Hist(escombrerashoras09$VV, scale="frequency", breaks="Sturges",  
col="blue",xlab="Velocidad del viento", ylab="Frecuencias",main="Mi  
histograma")
```

4.4.3 Guardar la gráfica para su posterior uso.

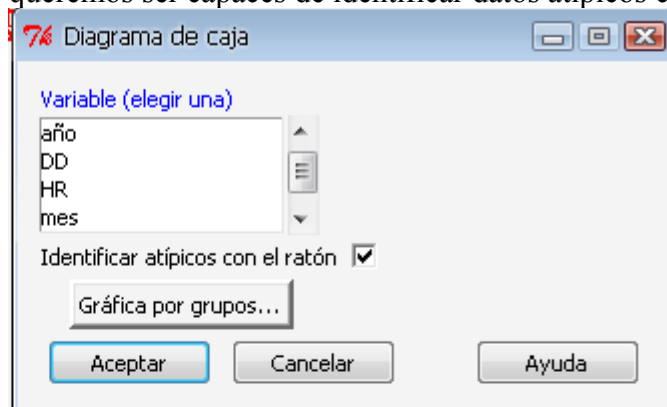
En la ventana de Gráficos, tenemos en el botón Archivo, la opción de guardar el gráfico en una serie de formatos de imagen. También tenemos la posibilidad de copiarlo en el área de transferencia para poder pegarlo directamente en un documento Word por ejemplo.

4.5 Obtenemos diagrama de caja-bigotes.

Para ello, usamos el comando, en la primera fila de botones:

Gráficas > Diagrama de Cajas

Escogemos al igual que en el apartado anterior la velocidad del viento VV. También queremos ser capaces de identificar datos atípicos con el ratón:



Le damos a aceptar.

Basta con pinchar en un dato atípico con el botón izquierdo del ratón para que aparezca el índice de la observación asociada. Para parar la identificación, le damos al botón derecho.

Nota: *Es crucial pensar en parar la identificación cuando hemos acabado porque, de no hacerlo, R-Commander se puede quedar colgado.*

¿Puede obtener por lo menos una fecha en la que se produjo un atípico en la medición de velocidad del viento.

Vamos a realizar ahora diagramas de caja por grupos, distinguiendo dentro del conjunto según las horas del día:

Le damos a Gráfica por grupos, y escogemos la variable hora.

¿Hay alguna tendencia en la evolución de la velocidad del viento a medida que va pasando el día?

EJERCICIOS

1. Con el conjunto de datos prueba_acceso.
 - a. Calcular la nota promedio de expedientes desglosando según los años.
 - b. Comparar, vía una gráfica de caja-bigotes, las calificaciones obtenidas en la convocatoria ordinaria y la convocatoria excepcional.
 - c. Intentando usar el comando Estadísticos > Resúmenes >Tabla de estadísticas, calcular el promedio de las calificaciones desglosando según los años académicos y según el tipo de convocatoria.
2. Con el conjunto de datos cemento,
 - a. Calcular la resistencia promedio según los días de fraguado
 - b. Calcular la desviación típica según los días de fraguado
3. Con el conjunto de datos escombrerashoras09.
 - a. Detectar posibles errores en las mediciones de temperatura. ¿A qué fechas corresponden?
 - b. Realizar un diagrama de cajas por grupos de la variable NO2, usando el factor día. Detectar datos claramente atípicos. ¿A qué fecha/hora corresponden?