

GUIA TEÓRICO PRÁCTICA 3: Histogramas, Diagramas de tallo hoja

1. Introducción al R Commander

1.1 Iniciar R

Generar 2 data frames, datos1 y datos2:

```
> ls()
character(0) #inicialmente no tenemos objetos

> set.seed(103)
> x <- runif(100)
> hist(x)
> y <- 2*x
> plot(x,y)
> datos1 <- data.frame(x=x,y=y)
>
> y2 <- 2*x + rnorm(100,0,1)
> plot(x,y2)
> datos2 <- data.frame(x=x,y=y2)
> ls()
[1] "datos1" "datos2" "x"    "y"    "y2"
```

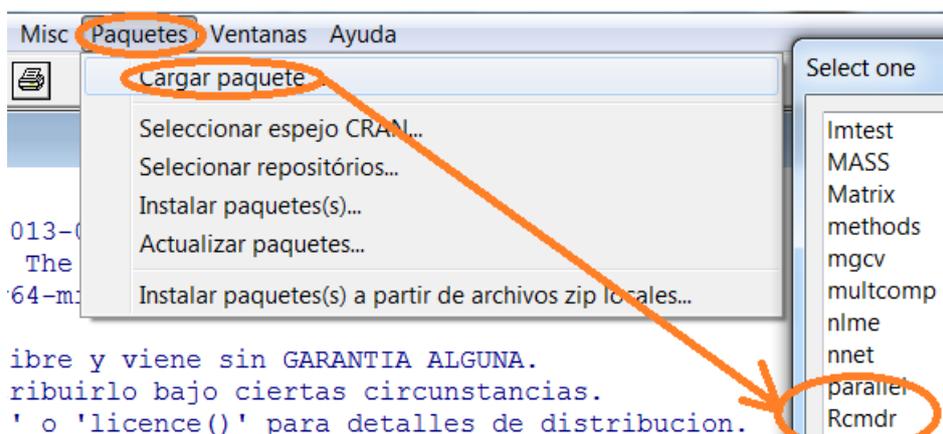
1.2. Iniciar R Commander

El R Commander se inicia como toda biblioteca de R

Opción 1: Desde la consola de R, ejecutar `library("Rcmdr")` o `library(Rcmdr)`

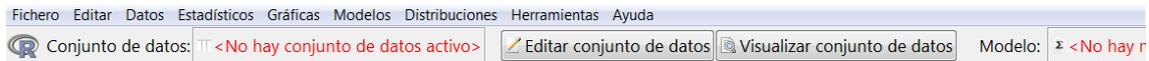
```
> library("Rcmdr")
```

Opción 2: Desde el menú, Paquetes -> Cargar Paquete -> Rcmdr



La ventana del R Commander que se muestra a continuación tiene cuatro partes

1) La barra del menú



2) La ventana de instrucciones. Allí aparecerán las instrucciones de R que generan las opciones seleccionadas en el menú



Es una venta de escritura (script window) en la que **se pueden escribir los comandos directamente y ejecutarlos resaltándolos y presionando el botón "Ejecutar"** que se encuentra a la derecha justo debajo de la ventana.

3) La ventana de resultados



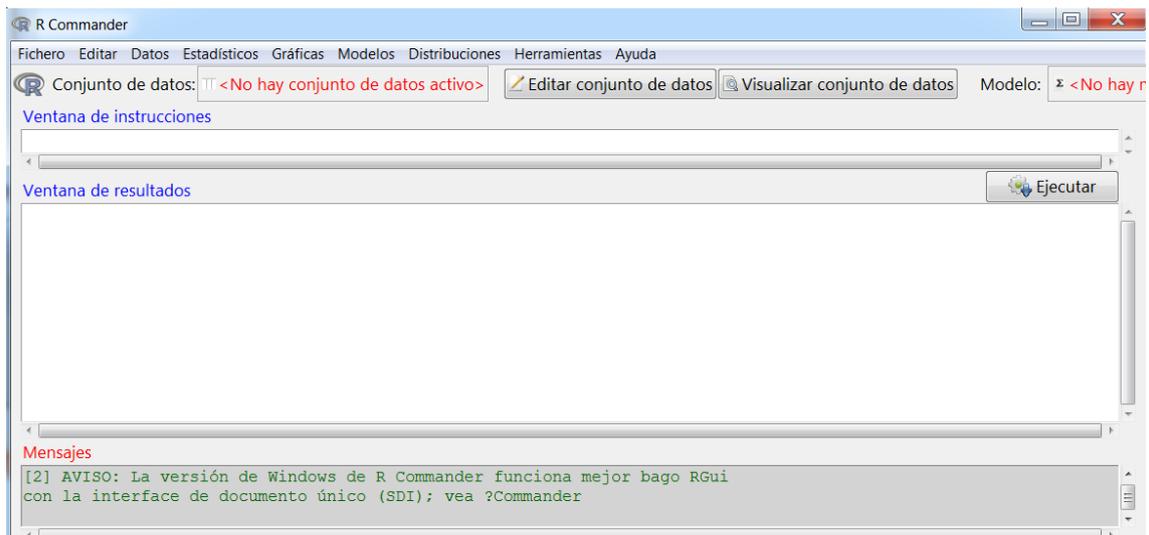
4) La ventana de mensajes donde aparecen los mensajes de error en rojo, de advertencias en verde y azul otras informaciones

Los gráficos aparecerán en una ventana gráfica separada. Solamente se verá el último gráfico pero se podrán recuperar gráficos previos con las teclas "page up" y "page down"

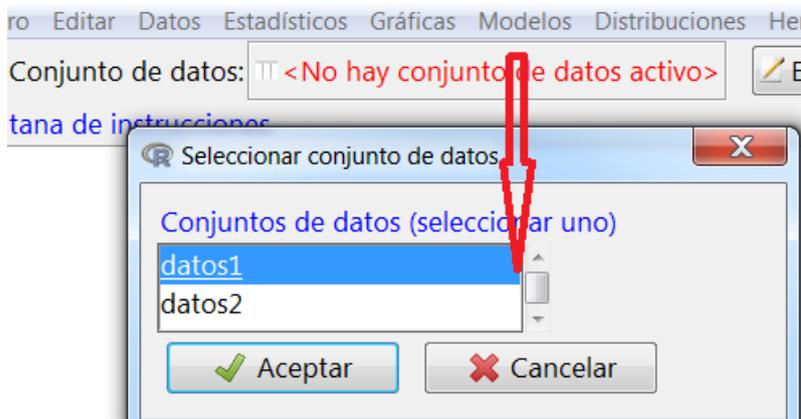
Cerremos el Rcmdr sin cerrar R.

Para volver a abrirlo, ejecutar

> **Commander ()**

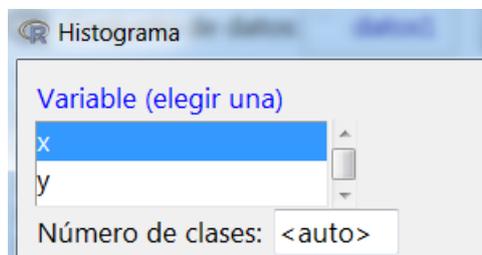


Clickeando sobre la solapa "No hay conjunto de datos activo" el Rcmdr ofrece la opción de seleccionar entre los dos data frames creados: datos1 y datos2. Seleccionamos datos1



2. Histogramas desde Rcmdr

- Obtenga histogramas de frecuencia, de porcentaje y de densidad de x e y (Gráficas-> Histograma). ¿Qué función utiliza para generar el histograma? ¿Qué criterio utiliza para la cantidad de clases por defecto?
- Calcule la cantidad de clases sugeridas por las reglas de Sturges, Scott y Freedman-Diaconis y construya los histogramas anteriores indicando las cantidades calculadas en la opción Gráficas -> Histograma -> Número de clases



- Realice los histogramas anteriores para datos2 y para los puntos de fusión de ceras naturales ("cera.txt")
- Modifique las instrucciones desde la ventana correspondiente para obtener histogramas con diferentes aspectos **presionando el botón "Ejecutar"** que se encuentra a la derecha justo debajo de la ventana.

3. Histogramas desde R

3.1 Para datos2 y cera.txt

Utilizando la función Hist(), ídem con hist()

- cambie la cantidad de clases según las reglas de Scott y Freedman-Diaconis.
- cambie los colores de los rectángulos de clase
- agregue un título
- cambie las etiquetas de los ejes

3.2 Para un conjunto de datos incorporado en "package 'datasets'"

Obtener información sobre el archivo de datos:

```
>?faithful ó >help(faithful)
```

Ver datos incorporados en R

```
> data()
```

Para ver **todos** los datos incorporados en R

```
> data(package = .packages(all.available = TRUE))
```

a) Construya un histograma para las distintas variables del data.frame "faithful"

```
hist(faithful$waiting)
hist(faithful$eruptions)
hist(faithful$eruptions, density =-1)
hist(faithful$eruptions, density =4)
```

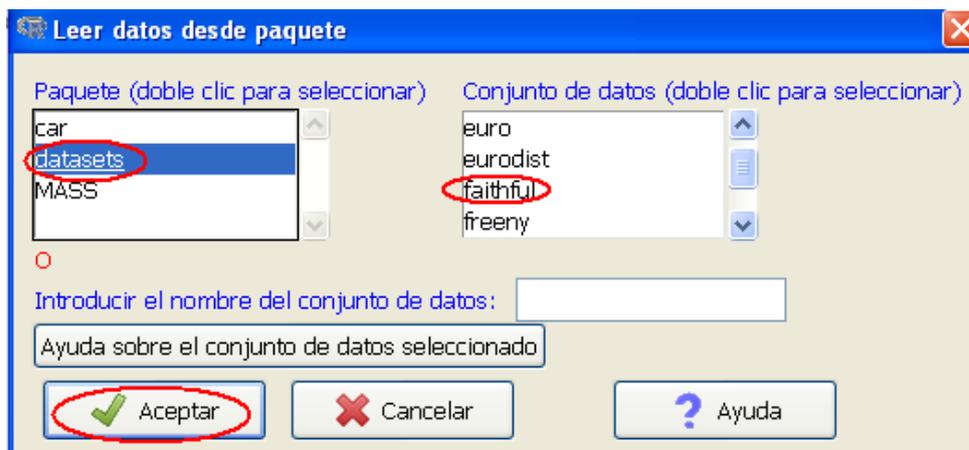
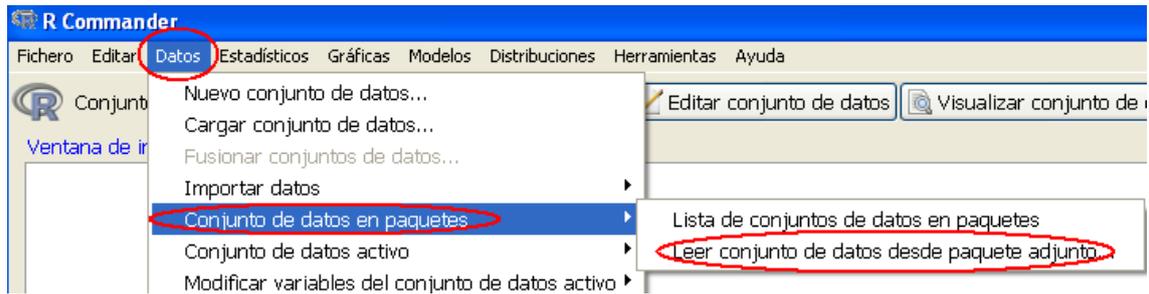
```
hist(faithful$eruptions,probability=T) # histograma en escala de densidad
```

b)?hist # analice las distintas opciones

Obtenga el histograma de acuerdo a los criterios de Scott y Freedman & Diaconis

c) Cambie el valor inicial del histograma y la longitud del intervalo de clase.(Mire el help)

3.3 Desde Rcmdr



Ahora "faithfull" es el conjunto de datos activo en Rcmdr

Construya y modifique los histogramas desde Rcmdr

4. Construcción de un polígono de frecuencias en R

Un polígono de frecuencias se obtiene uniendo los puntos medios de las barras de un histograma.

a) `histo<-hist(faithful$waiting,plot=F) ### no hace el gráfico histo`

```
> histo
$breaks
[1] 40 45 50 55 60 65 70 75 80 85 90 95 100

$count
[1] 4 22 33 24 14 10 27 54 55 23 5 1

$intensities
[1] 0.0029411759 0.0161764706 0.0242647059 0.0176470588 0.0102941176
0.0073529412 0.0198529412 0.0397058824 0.0404411765 0.0169117647
0.0036764706
[12] 0.0007352941
```

```
$density
 [1] 0.0029411759 0.0161764706 0.0242647059 0.0176470588 0.0102941176
     0.0073529412 0.0198529412 0.0397058824 0.0404411765 0.0169117647
     0.0036764706
 [12] 0.0007352941

$mids
 [1] 42.5 47.5 52.5 57.5 62.5 67.5 72.5 77.5 82.5 87.5 92.5 97.5

$xname
 [1] "faithful$waiting"

$equidist
 [1] TRUE

attr(,"class")
 [1] "histogram"

> hist(faithful$waiting,axes=F)
> axis(1, histo$mids)
> axis(2,seq(0,50,10))
> lines (histo$mids,histo$counts)

?axis
```

b) Repita para **faithful\$eruptions**

5. Distribuciones teóricas

El gráfico de la función de densidad de una distribución normal se obtiene con las siguientes instrucciones (ver el help ?dnorm)

```
>x<-seq(-8,8,0.01)
> plot(x,dnorm(x),type="l",ylim=c(0,1),col=1) #normal estandar
> lines(x,dnorm(x,0,2),type="l",ylim=c(0,1),col=2) #media 0 y desvío 1
> lines(x,dnorm(x,0,0.5),type="l",ylim=c(0,1),col=3) #media 0 y desvío 0.5
> lines(x,dnorm(x,2,1),type="l",ylim=c(0,1),col=4) #media 2 y desvío 1
> lines(x,dnorm(x,-2,0.5),type="l",ylim=c(0,1),col=5) #media -2 y desvío 0.5
```

Para obtener una muestra de tamaño 100 de una distribución normal estandar
> rnorm(100)

De una distribución normal con media 10 y desvío 2
> rnorm(100,10,2)

El gráfico de la función de densidad de una distribución t (t de Student) con 30, 10, 3, 2 y 1 grado de libertad se obtiene con las siguientes instrucciones. (ver el help ?dt)

```
> x<-seq(-5,5,0.01)

> plot(x,dt(x,30),type="l",col=1)
> lines(x,dt(x,10),type="l",col=2)
> lines(x,dt(x,3),type="l",col=3)
> lines(x,dt(x,2),type="l",col=4)
> lines(x,dt(x,1),type="l",col=5)
```

Para obtener una muestra de tamaño 100 de una distribución t con 30 grados de libertad

```
>rt(100,30)
```

El gráfico de la función de densidad de una distribución chi- cuadrado con 30, 10, 3, 2 y 1 grado de libertad se obtiene con las siguientes instrucciones. (ver el help ?dchisq)

```
> x<-seq(0,100,0.01)
>plot(x,dchisq(x,50),type="l",col=1,xlim=c(0,100),
ylim=c(0,0.25))
> lines(x,dchisq(x,30),type="l",col=2)
> lines(x,dchisq(x,10),type="l",col=3)
> lines(x,dchisq(x,3),type="l",col=4)
> lines(x,dchisq(x,2),type="l",col=5)
> lines(x,dchisq(x,1),type="l",col=6)
```

f) Para obtener una muestra de tamaño 100 de una distribución chi-cuadrado con 30 grados de libertad

```
>rchisq(100,30)
```

g) Grafique, en la misma pantalla, histogramas con muestras de tamaño 100 de la distribución t con 1, 2, 3, 10, 30 y 50 grados de libertad. Idem para la distribución Chi-cuadrado.

h) Construya un histograma con la mezcla de una muestra de tamaño 100 de una $N(0,1)$ y otra muestra de tamaño 100 de una $N(3,1)$

6- Diagramas tallo hoja.

6.1 Con R

Un diagrama tallo-hoja se obtiene mediante la función **stem**.

Utilice el **help** para hallar las diferentes opciones de **stem**.

a1) Pruebe con los datos de **swiss** de R las diferentes opciones de **stem**.

a2) Construya diagramas tallo-hoja de todas las variables que aparecen en **swiss**. Utilice el **help** para entender de que se tratan las variables. El diagrama de la quinta columna se obtiene:

```
> stem(swiss[,5])
```

b) Pruebe las dos opciones siguientes en el conjunto de datos "abbey" de la librería "MASS" y compare.

```
> library(MASS)
> ?abbey # de qué se tratan los datos?
> stem(abbey)
> stem(abbey,scale=2) # en R
```

c) Estudie mediante un diagrama tallo-hoja las mediciones del cobre en harina integral: datos **chem**

d) Estudie mediante diagramas de tallo-hoja y diferentes histogramas los datos de muertes mensuales por enfermedades de pulmón dados en **mdeaths** (hombres) **fdeaths** (mujeres).
> ?mdeaths

Estudie también los dos conjuntos de datos juntos, utilice histogramas con la misma escala para poder compararlos.

El help del R nos dice:
Monthly Deaths from Lung Diseases in the UK

Description:

Three time series giving the monthly deaths from bronchitis, emphysema and asthma in the UK, 1974-1979, both sexes ('ldeaths'), males ('mdeaths') and females ('fdeaths').

Usage:

ldeaths
fdeaths
mdeaths

6.2 Con Rcmdr

Realice gráficos tallo-hoja para los datos analizados en 6.1. Explore las diferentes opciones del Rcmdr



7 Estimación de densidades de conjuntos de datos con R

a) La función **density** le permite obtener un estimador de densidad de un conjunto de datos. Utilice el **help**.

b) Para los datos de los ejercicios 3 y 6 obtenga estimadores de densidad, con distintos tamaños de ventanas y gráfíquelos. Pruebe modificar las distintas opciones del comando **density**.

c) Superponga la curva de densidad a los histogramas.