

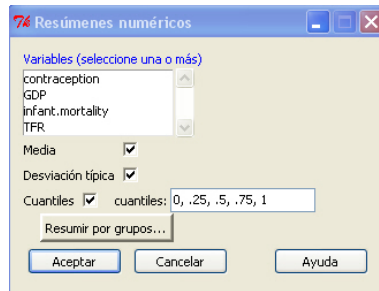
Laboratorio 1: Análisis de datos univariantes

Se toman, por ejemplo, los datos del fichero `Nations.txt`.

Reproduce con la variable `region` el mismo análisis y observa las diferencias.

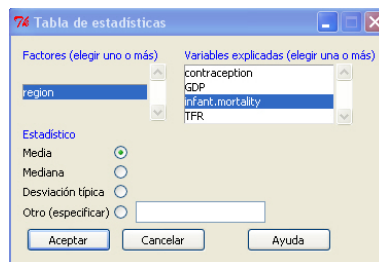
Se calculan los resúmenes de los datos o estadísticos descriptivos básicos:

Estadísticos → **Resúmenes** → **Resúmenes Numéricos**



Se calculan los resúmenes de los datos o estadísticos descriptivos según los niveles de una variable nominal (`region`):

Estadísticos → **Resúmenes** → **Tabla de Estadísticas**



HISTOGRAMAS

Gráficas → **Histograma** (e.g. de la variable `infant.mortality`)

NOTA: Las opciones por defecto del gráfico se pueden cambiar modificando el comando que aparece en la *ventana de instrucciones*:

```
Hist(Naciones$infant.mortality, scale="percent",  
breaks="Sturges", col="darkgray")
```

Basta sustituir las opciones deseadas (pulsar en el **Help** de la ventana *Histograma* para una revisión de las mismas). Una vez hechos los cambios, se marca el comando y se pulsa el botón **Ejecutar**

DIAGRAMA DE BARRAS

Gráficas → **Gráfica de Barras** (e.g. de la variable `region`)

Una vez obtenidos los gráficos se pueden **grabar** en diferentes formatos:

Sobre la ventana gráfica abierta:

Archivo → **Guardar como...**

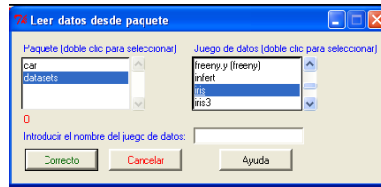
Trabajo con datos incluidos en datasets de R

Muchas librerías de R tienen ficheros de datos.

Se consideran ahora los datos del ejemplo `iris` del paquete `datasets` de R.

Se carga el fichero en Rcmdr:

Datos → Datos en paquetes → Leer datos desde paquete adjunto



En el cuadro de diálogo se elige el paquete `datasets` y dentro de éste el grupo de datos `iris`.

Propuesta: Repetir los análisis anteriores hechos con el fichero `Nations.txt` usando el conjunto de datos `iris`.

Interpretar y comentar todos los resultados obtenidos.

Programación en R

Para facilitar el trabajo y hacer reproducible cualquier análisis, se puede programar todos los procedimientos anteriores mediante un *script* directamente en R (sin la ayuda de RCommander).

Introducir en el *prompt* principal del programa (>):

```
# Lectura de datos

Datos <- read.table("C:/R-2.8.1/library/Rcmdr/etc/Nations.txt",
header=TRUE)

# Se eliminan datos no asignados (ausentes) NA

Datos <- na.omit(Datos)

# Diagrama de barras

X11()

barplot(table(Datos$region), xlab="region", ylab="Frecuencias")

# Diagrama de tarta

X11()

pie(table(Datos$region), labels=levels(Datos$region),
main="Distribucion de region", col=rainbow(length(levels(Datos$region))))
```

```
# Histograma

X11()

hist(Datos$infant.mortality, xlab="Mortalidad Infantil",
ylab="Frecuencias Relativas", main="", col="blue")

# Resúmenes numéricos

summary(Datos)

# Tabla de distribución de datos

Tabla <- table(Datos$region)

100*Tabla/sum(Tabla)

# Media de una variable según los niveles de una variable nominal

tapply(Datos$infant.mortality, list(region=Datos$region),
mean, na.rm=TRUE)
```