

EYP1113 - PROBABILIDAD Y ESTADÍSTICA

LABORATORIO 2

PROFESORAS: NATALIA VENEGAS Y PILAR TELLO

FACULTAD DE MATEMÁTICAS

DEPARTAMENTO DE ESTADÍSTICA

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

SEGUNDO SEMESTRE 2019

- 1 Lectura de datos .dta
 - [En R] con librería “foreign”
 - [En R Studio] Importar base de datos

- 2 Gráficos en R
 - Comando `plot()`
 - Argumentos opcionales de `plot()`
 - Símbolos matemáticos en gráficos
 - Diagrama de caja
 - Histograma
 - Gráfico de barras

LECTURA DE DATOS .DTA

[EN R] CON LIBRERÍA “FOREIGN”

[En R] Para leer datos en formato .dta se puede usar la librería “foreign” la cual se debe instalar tan solo una vez antes de cargar y empezar a usar.

```
# install.packages("foreign")  
library(foreign)  
data = read.dta(...)
```

[En **R Studio**]

- Environment (Arriba a la derecha)
- Import Dataset
- From Stata
- Browse... (Seleccionar archivo)
- Import

Tarea: Notar qué package y función usa **R Studio** para leer los datos.

1. Desde **webcursos** descargue la base de datos 'RentasMunich.dta'
2. Usando R, guarde bajo el nombre de data la lectura de los datos ya descargados.
3. Usando R Studio, importe los datos. ¿Cómo se llama el objeto que almacena la base de datos?

¿DE QUÉ TRATA LA BASE DE DATOS?

La data contiene 3082 observaciones correspondientes a rentas de apartamento en Munich durante 1999.

Sus variables son:

- rent = valor renta
- rentsqm = renta por metro cuadrado
- area = tamaño/superficie en metros cuadrados
- yearc = año de construcción
- location = localización: 1 Promedio; 2 Bueno; 3 Alta
- bath = tipo de baño: 0 Estándar; 1 Premium
- kitchen = tipo de cocina: 0 Estándar; 1 Premium
- cheating = presencia de calefacción: 0 No; 1 Sí
- district = distrito de Munich donde el apartamento está localizado.

GRÁFICOS EN R

COMANDO `plot()`

- Se usa el comando `plot(x,y,...)` para graficar un vector versus otro. *Tarea:* ¿Qué ocurre al omitir el argumento `y`?
- Si a un gráfico se quiere agregar un punto (x_1, y_1) en particular, se usa el comando `points(x1,y1,...)` una vez ejecutado el comando `plot()`.
- Si a un gráfico se quiere agregar una línea entre dos puntos $(x_1, y_1), (x_2, y_2)$, se usa el comando `lines(c(x1,x2), c(y1,y2), ...)`, luego de haber ejecutado el comando `plot()`.

- Líneas horizontales o verticales pueden ser agregadas al gráfico mediante el comando `abline(...)` con argumentos `h=` o `v=`, respectivamente.
- Si se quiere agregar una línea con intercepto “a” y pendiente “b”, se utiliza `abline(a,b)`
- Si se quiere dibujar un gráfico en blanco, al cual vamos a agregar otros gráficos, se usa el comando `plot(x,y,type="n", ylim=, xlim=,...)`

ARGUMENTOS OPCIONALES DE `plot()`

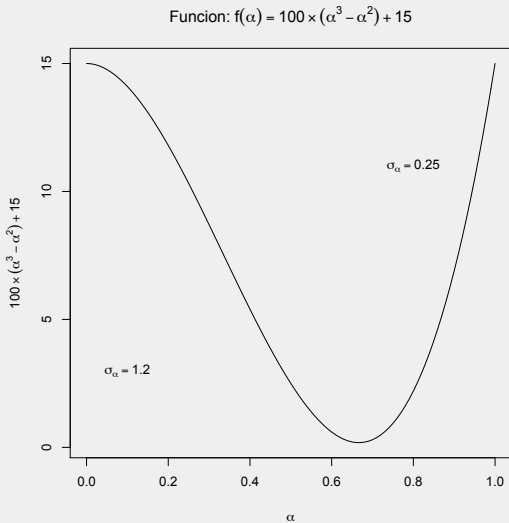
- `xlab=`, `ylab=`, `main=` etiquetan el eje X, eje Y y el título del gráfico, respectivamente.
- Puntos y líneas en el gráfico pueden tener distintos colores, agregando el argumento `col="..."` en la función. También se puede modificar su grosor usando `lwd=...`
Tarea: Ejecute `colors()` ¿Qué aparece en la consola?
- Para agregar etiqueta en un punto (x, y) del gráfico usamos el comando `text(x, y, label=)`.
- Para agregar título a un gráfico ya ejecutado, se usa el comando `title()`

- Podemos escribir símbolos matemáticos en R para los títulos, subtítulos, ejes X e Y, entre otros.
- Usaremos las funciones `expression()`, `bquote()` y `paste()`

SÍMBOLOS MATEMÁTICOS EN GRÁFICOS, EJEMPLO

```
par(mar=c(5.1,5.1,4.1,2.1))  
curve(100*(x^3-x^2)+15, 0, 1,  
      xlab=expression(alpha),  
      ylab=expression(100%*%(alpha^3-alpha^2)+15),  
      main=expression(  
        paste("Función: ",  
              f(alpha)==100%*%(alpha^3-alpha^2)+15)))  
sigma1=1.2  
text(0.1,3,bquote(sigma[alpha]==.(sigma1)))  
text(0.8,11,expression(sigma[alpha]==0.25))
```

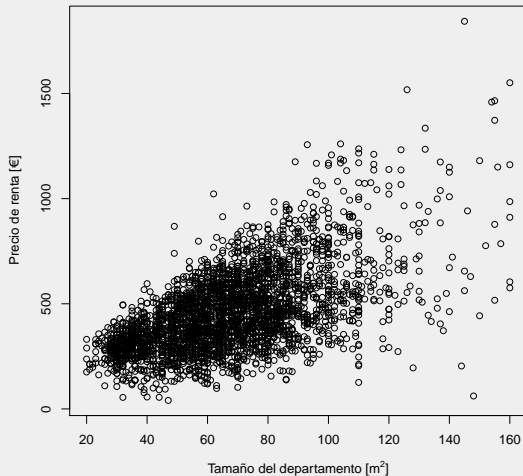
SÍMBOLOS MATEMÁTICOS EN GRÁFICOS, EJEMPLO



ACTIVIDAD CON `plot()`

1. Usando los datos `data` obtener el siguiente gráfico

Precio de renta por tamaño, Munich 1999



COMANDO `boxplot()`

Este comando sirve para graficar diagramas de cajas (representación de la distribución de una muestra a través de sus cuartiles, mínimo y máximo). El comando es el siguiente:

- `boxplot(x=, main=, xlab=, ylab=, horizontal=, col=)`

En este caso el argumento `x=` será el vector de los datos y el comando `horizontal=` puede ser `TRUE` o `FALSE` si se quiere que las cajas vayan de manera horizontal o vertical, respectivamente. El resto de los comandos son análogos a la función `plot()`.

- `boxplot(x~y, main=, xlab=, ylab=, horizontal=, col=)`

A diferencia del comando anterior, éste lleva una variable adicional y que puede asignar los datos de `x` en distintos grupos.

COMANDO `hist()`

El histograma de un conjunto de datos es un gráfico de barras que representan las frecuencias con que aparecen las mediciones agrupadas en ciertos rangos o intervalos. El comando es el siguiente:

- `hist(x=, main=, breaks, freq = NULL, xlab=, ylab=, col=)`

COMANDO `barplot()`

El gráfico de barras permite representar para cada una de las modalidades definidas para una variable cualitativa o los valores de una variable discreta en términos de frecuencias relativas o absolutas. Para obtener este tipo de gráficos en R, la función base es `barplot()`, como se describe a continuación:

- `barplot(height, width = 1, space = NULL, names.arg = NULL, legend.text = NULL, beside = FALSE, horiz = FALSE, col= ,...)`

MÁS ACTIVIDADES CON GRÁFICOS

1. Realizar histograma y diagrama de caja (boxplot) de la Renta por metro cuadrado.
2. Realizar boxplot de la Renta con respecto a la Localización.

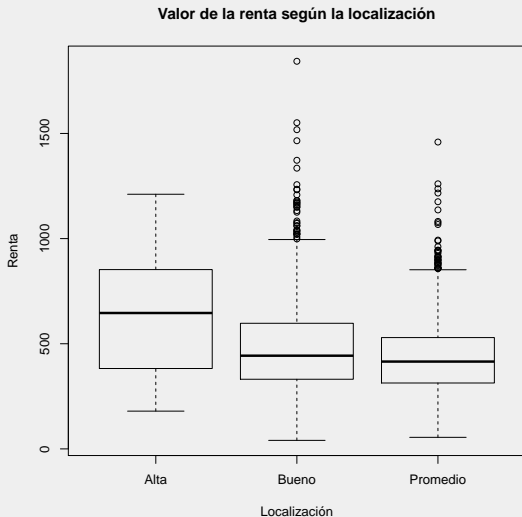
Ojo: La Localización es una variable categórica almacenada como {1,2,3}, pero que significa "Promedio", "Bueno" y "Alto" respectivamente.

```
data$Localización <- ifelse(data$location == 1, "Promedio",  
  ifelse(data$location == 2, "Bueno",  
    "Alta"))
```

3. Realizar gráfico de barra de la los tipos de Baños según la Localización.
4. Realizar gráfico de barra de la los tipos de Cocina según la Localización.
5. Realizar el diagrama de caja de la Renta por metro cuadrado según la presencia de Calefacción.

ACTIVIDAD CON `boxplot()`

Boxplot de la Renta con respecto a la Localización.



- Hay que distinguir entre el gráfico y la ventana donde aparece el gráfico. Un gráfico siempre aparece en una ventana. En una ventana puede haber más de un gráfico.
- Se pueden cambiar los parámetros gráficos de la ventana. Usar la función `par()` en R para ver cuales son los parámetros por defecto. Para obtener mas información de todos los parámetros se puede usar la ayuda de R, `?par`.

GUARDANDO LOS GRÁFICOS

- Para guardar los gráficos generados en R en formato pdf se usan los siguientes comandos:

```
pdf(file="nombre.pdf ", width= , height= )  
plot(...)  
dev.off()
```

- Si se grafica más de un plot aparecerá en el archivo pdf en mas de una página.
- Se pueden guardar los gráficos en formatos `postscript`, `jpeg`, `png` o `bmp` usando los comandos respectivos.
- Haciendo click derecho en el mismo gráfico, se puede guardar como imagen `metafile` o `bmp`.

AYUDA SOBRE FUNCIONES YA EXISTENTES EN R.

Si tiene alguna pregunta sobre algún comando ya existente en R, basta con ejecutar `?NombreDeLaFuncion` y un html le extenderá la ayuda. Ejemplo: `?hist`. Ahí podrá estudiar los parámetros de la función, sus opciones y ejemplos.