



FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

EYP1113 - Probabilidad y Estadística

Laboratorio 04

Pilar Tello Hernández
pitello@uc.cl

Facultad de Matemáticas
Departamento de Estadística
Pontificia Universidad Católica de Chile

Segundo Semestre 2021

Actividad con .dta

1. Desde **CANVAS** descargue la base de datos 'RentasMunich.dta'
2. Guarde bajo el nombre de data la lectura de los datos ya descargados.

¿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.

Comando plot()

- ▶ Se usa el comando `plot(x,y,...)` para graficar un vector versus otro. ¿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=,...)`

Ejemplo:

```
plot(x=data$rent,y=data$rentsqm)
points(x=1500,y=15)
points(x=1000,y=5)
lines(x=c(1000,1500),y=c(5,15))
abline(h=5)
abline(h=15)
abline(v=1000)
abline(v=1500)
abline(a=5,b=10/1500)
plot(x=data$rent,y=data$rentsqm,type="n")
points(x=c(1500,1000),y=c(15,5))
```

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=...`.
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()`

Argumentos opcionales de plot()

Ejemplo:

```
plot(x=data$rent,y=data$rentsqm,  
     xlab="Renta",ylab="Renta por metro cuadrado",  
     main="Renta versus Renta por metro cuadrado",  
     col="darkblue",lwd=2)  
points(x=1500,y=15,col="darkred",lwd=2)  
text(x=1500,y=16,label="Etiqueta del punto")  
  
plot(x=data$rent,y=data$rentsqm,  
     xlab="Renta",ylab="Renta por metro cuadrado",  
     col="darkblue",lwd=2)  
points(x=1500,y=15,col="darkred",lwd=2)  
text(x=1500,y=16,label="Etiqueta del punto")  
title(main="Renta versus Renta por metro cuadrado")
```

Símbolos matemáticos y funciones en gráficos

- ▶ Podemos escribir símbolos matemáticos en R para los títulos, subtítulos, ejes X e Y, entre otros.
- ▶ Usaremos las funciones `expression()`, `bquote()` y `paste()`
- ▶ El comando `curve` nos permite graficar una función con respecto a x . Contiene los argumentos `from` y `to` que permiten indicar los límites de x en donde se quiere graficar la función. Si se quiere añadir esta curva encima de otro gráfico se debe añadir el argumento `add=TRUE`.

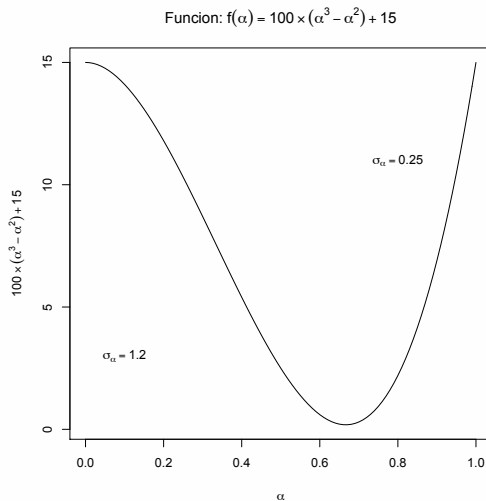


Símbolos matemáticos y funciones en gráficos

Ejemplo:

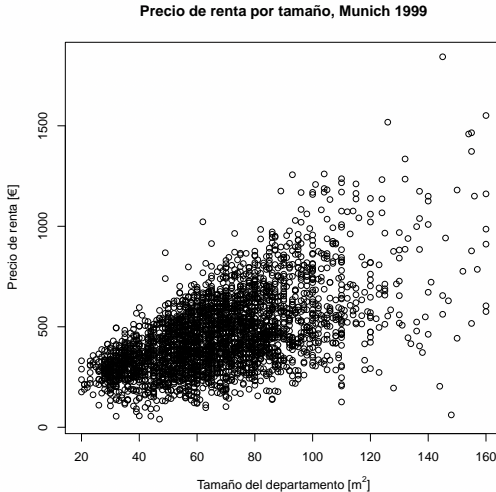
```
par(mar=c(5.1,5.1,4.1,2.1))
curve(expr=100*(x^3-x^2)+15, from=0, to=1,
xlab=expression(alpha),
ylab=expression(100%*(alpha^3-alpha^2)+15),
main=expression(
paste("Función: ",
f(alpha)==100%*(alpha^3-alpha^2)+15)))
sigma1=1.2
text(0.1,3,bquote(sigma[alpha]==.(sigma1)))
text(0.8,11,expression(sigma[alpha]==0.25))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

Símbolos matemáticos en gráficos, Ejemplo



Actividad para la semana con plot()

1. Usando los datos data obtener el siguiente gráfico



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 boxplot()

Ejemplo:

```
boxplot(x=data$rent,main="Boxplot de renta",xlab="",  
ylab="Renta",horizontal=FALSE,  
col="pink")
```

```
boxplot(x=data$rent,main="Boxplot de renta",ylab="",  
xlab="Renta",horizontal=TRUE,  
col="pink")
```

```
boxplot(data$rent~data$localizacion,  
main="Boxplot de renta",  
xlab="Localización",  
ylab="Renta",horizontal=FALSE,  
col="pink")
```

```
boxplot(data$rent~data$localizacion,  
main="Boxplot de renta",  
ylab="Localización",  
xlab="Renta",horizontal=TRUE,  
col="pink")
```

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 = TRUE, xlab=, ylab=, col=)`
- ▶ En este caso el argumento `x` será el vector de los datos.
- ▶ El comando `breaks` puede ser un valor con el cual se indica el número aproximado de clases o un vector cuyos elementos indican los puntos límites entre las clases o intervalos.
- ▶ El comando `freq =` argumento lógico, si se especifica como `TRUE`, el histograma representará las frecuencias absolutas o conteo de datos en cada clase, si se especifica como `FALSE`, se representarán las frecuencias relativas. Por defecto, este argumento toma el valor de `TRUE`.
- ▶ El resto de los comandos son análogos a la función `plot()`.



Comando hist()

Ejemplo:

```
hist(x=data$rent,main="Histograma de renta",  
freq=TRUE,xlab="Renta",ylab="Frecuencia",  
col="yellow")
```

```
hist(x=data$rent,main="Histograma de renta",  
freq=FALSE,xlab="Renta",ylab="Densidad",  
col="yellow")
```

```
hist(x=data$rent,main="Histograma de renta",  
freq=FALSE,xlab="Renta",ylab="Densidad",  
col="yellow",breaks=10)
```

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 terminos 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, legend.text = NULL, beside = FALSE, horiz = FALSE, col= ,...)`
- ▶ En este caso el argumento `height` vector o matriz de valores que describen las barras. En el caso de variables categóricas, entregamos el `table` de la variable. Si es un vector, entonces el gráfico corresponde a un secuencia de barras rectangulares cuyas alturas corresponden a los valores del vector.
- ▶ Si realizamos una tabla de doble entrada entonces tendremos el conteo de dos variables categóricas. Con el argumento `beside` podemos decidir si queremos un gráfico de barras apilado o agrupado (`beside=TRUE`). Con `legend.text=TRUE` podemos activar la leyenda para este gráfico.

Comando barplot()

Ejemplo:

```
table(data$localizacion)  
barplot(height=table(data$localizacion),  
col="darkgreen")
```

```
barplot(height=table(data$localizacion),  
col="darkgreen", horiz=TRUE)
```

```
table(data$localizacion,data$bano)  
barplot(height=table(data$localizacion,  
data$bano),  
col=c("tomato","violetred","turquoise"),  
beside=TRUE,legend.text=TRUE)
```

Más actividades con gráficos para la semana

1. Realizar histograma y diagrama de caja (boxplot) de la Renta por metro cuadrado.
2. Realizar boxplot de la Renta por metro cuadrado con respecto a la Calefacción.
3. Realizar gráfico de barra de la los tipos de Cocina según la Localización.
4. Realizar gráfico de barra de la los tipos de Baño según el tipo de Cocina.
5. Realizar el diagrama de caja de la Renta por metro cuadrado según la presencia de Calefacción.



Parámetros gráficos

- ▶ 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.

Modelos de Probabilidad

Existen diversos modelos de probabilidad para estudiar. Los modelos más utilizados son

Modelo	Comando
Binomial	<code>_binom()</code>
Poisson	<code>_pois()</code>
Uniforme	<code>_unif()</code>
Normal	<code>_norm()</code>
Exponencial	<code>_exp()</code>
Gamma	<code>_gamma()</code>
Chi Cuadrado	<code>_chisq()</code>
t-Student	<code>_t()</code>
Fisher	<code>_f()</code>

Cada comando puede ser utilizado de 4 formas diferentes:

- ▶ `dDISTR(x,...)`. En el caso de las variables discretas, entrega $P(X = x)$, en el caso de las variables continuas entrega $f_X(x)$.
- ▶ `pDISTR(q,...)`. Entrega $P(X \leq q)$.
- ▶ `qDISTR(p,...)`. Entrega el valor de x tal que $P(X \leq x) = p$.
- ▶ `rDISTR(n,...)`. Genera una muestra proveniente de un modelo de distribución.

Exponencial

Si $X \sim \text{Exp}(\nu)$, $x > 0$, entonces:

- ▶ Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dexp(x, rate= ν)`.
- ▶ Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pexp(q, rate= ν)`.
- ▶ Para calcular cuantiles de la distribución se usa el comando `qexp(p, rate= ν)`.
- ▶ Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rexp(n, rate= ν)`.

Exponencial

Ejemplo:

Si $X \sim \text{Exp}(\nu = 3)$:

- ▶ Calcule $f_X(1)$
`dexp(1,rate=3)`
- ▶ Calcule $P(X \leq 1,5) = F_X(1,5)$
`pexp(1.5,rate=3)`
- ▶ Si $P(X \leq k) = 0,5$, obtenga el valor de k .
`qexp(0.5,rate=3)`
- ▶ Genere una muestra de tamaño $n = 1000$.
`rexp(1000,rate=3)`

Uniforme

Si $X \sim \text{Unif}(a, b)$, $a < x < b$, entonces:

- ▶ Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dunif(x, min=a, max=b)`.
- ▶ Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `punif(q, min=a, max=b)`.
- ▶ Para calcular cuantiles de la distribución se usa el comando `qunif(p, min=a, max=b)`.
- ▶ Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `runif(n, min=a, max=b)`.

Exponencial

Ejemplo:

Si $X \sim \text{Uniforme}(a = -2, b = 8)$:

- ▶ Calcule $f_X(0)$
`dunif(0,min=-2,max=8)`
- ▶ Calcule $P(X \leq 0) = F_X(0)$
`punif(0,min=-2,max=8)`
- ▶ Si $P(X \leq k) = 0,3$, obtenga el valor de k .
`qunif(0.3,min=-2,max=8)`
- ▶ Genere una muestra de tamaño $n = 1000$.
`runif(1000,min=-2,max=8)`

La distribución Normal

Si $X \sim \text{Normal}(\mu, \sigma^2)$, $x \in \mathbb{R}$, entonces:

- ▶ Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dnorm(x, mean= μ , sd= σ)`.
- ▶ Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pnorm(q, mean= μ , sd= σ)`.
- ▶ Para calcular cuantiles de la distribución se usa el comando `qnorm(p, mean= μ , sd= σ)`.
- ▶ Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rnorm(n, mean= μ , sd= σ)`.

La distribución Log-Normal

Si $X \sim \text{Log-Normal}(\lambda, \zeta)$, $x > 0$, entonces:

- ▶ Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dlnorm(x, meanlog= λ , sdlog= ζ)`.
- ▶ Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `plnorm(q, meanlog= λ , sdlog= ζ)`.
- ▶ Para calcular cuantiles de la distribución se usa el comando `qlnorm(p, meanlog= λ , sdlog= ζ)`.
- ▶ Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rlnorm(n, meanlog= λ , sdlog= ζ)`.

La distribución Gamma

Si $X \sim \text{Gamma}(k, \nu)$, $x > 0$, entonces:

- ▶ Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dgamma(x, shape=k, rate= ν)`.
- ▶ Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pgamma(q, shape=k, rate= ν)`.
- ▶ Para calcular cuantiles de la distribución se usa el comando `qgamma(p, shape=k, rate= ν)`.
- ▶ Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rgamma(n, shape=k, rate= ν)`.

La distribución Chi-cuadrado

Si $X \sim \chi_n^2, x > 0$, entonces:

- ▶ Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dchisq(x, df=n)`.
- ▶ Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pchisq(q, df=n)`.
- ▶ Para calcular cuantiles de la distribución se usa el comando `qchisq(p, df=n)`.
- ▶ Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rchisq(n, df=n)`.

Ejercicios para la semana

1. Simule 100 variables aleatorias provenientes de una distribución exponencial con tasa 2. Grafique su función de densidad empírica junto con la curva de densidad teórica.
2. Simule 100 variables aleatorias provenientes de una distribución log normal de parámetros 0.5 y 0.02. Grafique su función de densidad empírica junto con la curva de densidad teórica.
3. Simule 100 variables aleatorias provenientes de una distribución normal de parámetros $\mu = 100$ y $\sigma = 15$. Grafique su función de densidad empírica junto con la curva de densidad teórica.
4. Para $n=10, 100, 1000, 10000$, genere muestras de una distribución normal $\mu = 650$ y $\sigma = 50$ y grafique su función de densidad empírica junto con la curva de densidad teórica dentro de una misma ventana.

