

GUÍA TEÓRICO PRÁCTICA 3.

Histogramas, Diagramas de tallo hoja

1-1 Histogramas

Trabajaremos con un conjunto de datos existentes en “la librería” de R *faithful*, en S-Plus *geyser*
Para obtener información sobre el archivo de datos hacemos

```
?faithful
```

```
help(faithful)
```

a) Construya un histograma para las distintas variables

```
hist(faithful$waiting)
hist(faithful$eruptions)
hist(faithful$eruptions, density =-1) # en S-plus obtenemos barras sin relleno
```

```
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)

1-2 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
```

```
$counts
```

```
[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$waiting)
> axis(1, histo$mids)
> lines (histo$mids,histo$counts)
```

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

*1-3.Otras distribuciones.

- a) 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
```
- b) 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)
```
- c) 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)
```
- d) Para obtener una muestra de tamaño 100 de una distribución t con 30 grados de libertad
- ```
>rt(100,30)
```
- e) 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 $N(3,1)$

2- Diagramas tallo hoja.

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

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

a1) Pruebe con los datos de **swiss** en R (**swiss.fertility** en S-Plus) las diferentes opciones de **stem**.

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

```
> stem(swiss[,5]) en R
> stem(swiss.x[,5]) en S-Plus
```

- b) Estudie mediante un diagrama tallo- hoja las mediciones del cobre en harina integral.
- c) 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
> stem(abbey,scale=-1) # en S-plus
```

Observe que cambian las posiciones del punto decimal (colon), pero los parámetros están definidos en forma diferente en R y S-plus

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). 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,

*ANÁLISIS DE DATOS*  
*2do. Cuatrimestre 2009*

emphysema and asthma in the UK, 1974-1979, both sexes ('ldeaths'),  
males ('mdeaths') and females ('fdeaths').

Usage:

ldeaths  
fdeaths  
mdeaths

e) Para los diagramas tallo-hoja obtenidos obtenga la línea de profundidad. Utilice la profundidad obtenida para hallar la mediana, los cuartos y los cuartiles de los datos considerados.