Script4

Yosune Miquelajauregui

22/12/2017

Creación de base de datos y estadística descriptiva

En ocasiones necesitamos identificar los atributos que coinciden entre dos o más hojas de datos.

```
HD1 <- read.csv(file="Datos2.csv", header = T)
length(HD1[,1])

## [1] 1000

head(HD1)

##   ID V1        V2
## 1  1  1 0.3156189
## 2  2 11 0.6477871
## 3  3 21 0.3560912
## 4  4 31 0.8251590
## 5  5 41 0.4933043
## 6  6 51 0.8434979

HD2 <- read.csv(file="Datos3.csv", header = T)
length(HD2[,1])

## [1] 72

head(HD2)

##   ID       D1 D2
## 1  1 3.867982  5
## 2  2 7.241679  4
## 3  3 6.434363  3
## 4  4 9.706802  8
## 5  5 7.612441  6
## 6  6 6.010280  8

HDMatch <- HD1[HD1$ID %in% HD2$ID, ]
length(HDMatch[,1])

## [1] 72

head(HDMatch)

##   ID V1        V2
## 1  1  1 0.3156189
```

```
## 2   2 11 0.6477871
## 3   3 21 0.3560912
## 4   4 31 0.8251590
## 5   5 41 0.4933043
## 6   6 51 0.8434979
```

```
Todo <- merge(HDMatch,HD2, by="ID")
length(Todo[,1])
```

```
## [1] 72
```

```
head(Todo)
```

```
##   ID V1        V2       D1 D2
## 1  1  1 0.3156189 3.867982  5
## 2  2 11 0.6477871 7.241679  4
## 3  3 21 0.3560912 6.434363  3
## 4  4 31 0.8251590 9.706802  8
## 5  5 41 0.4933043 7.612441  6
## 6  6 51 0.8434979 6.010280  8
```

Estadística descriptiva

```
mean(HD1[,2])
```

```
## [1] 46
```

```
sd(HD1[,2])
```

```
## [1] 28.73719
```

```
range(HD1[,2])
```

```
## [1]  1 91
```

```
min(HD1[,2])
```

```
## [1] 1
```

```
max(HD1[,2])
```

```
## [1] 91
```

```
median(HD1[,2])
```

```
## [1] 46
```

```
mode(HD1[,2])
```

```
## [1] "numeric"
```

```
quantile(HD1[,2])
```

```
##   0%  25%  50%  75% 100%
##    1   21   46   71   91
```

```
x<-seq(131,215,0.1)
plot(x=0,y=0,ylim=c(0,0.016),xlim=c(100,250),xlab=expression(paste(italic
(mu), " = 170 g, ", italic(sigma), " = 25")),ylab="Densidad", type="n",
cex.lab=1.5,cex.axis=1.5,main=expression(paste("Peso des ratones de
laboratorio ", paste(italic(N), "(170,25)"))))
curve(dnorm(x, mean=170, sd=25),100,250,
ylim=c(0,0.16),xlab=expression(paste(mu, " = 170", sigma, " = 25")),
ylab="Densidad", add=TRUE)
```


Peso des ratones de laboratorio N(170,25)
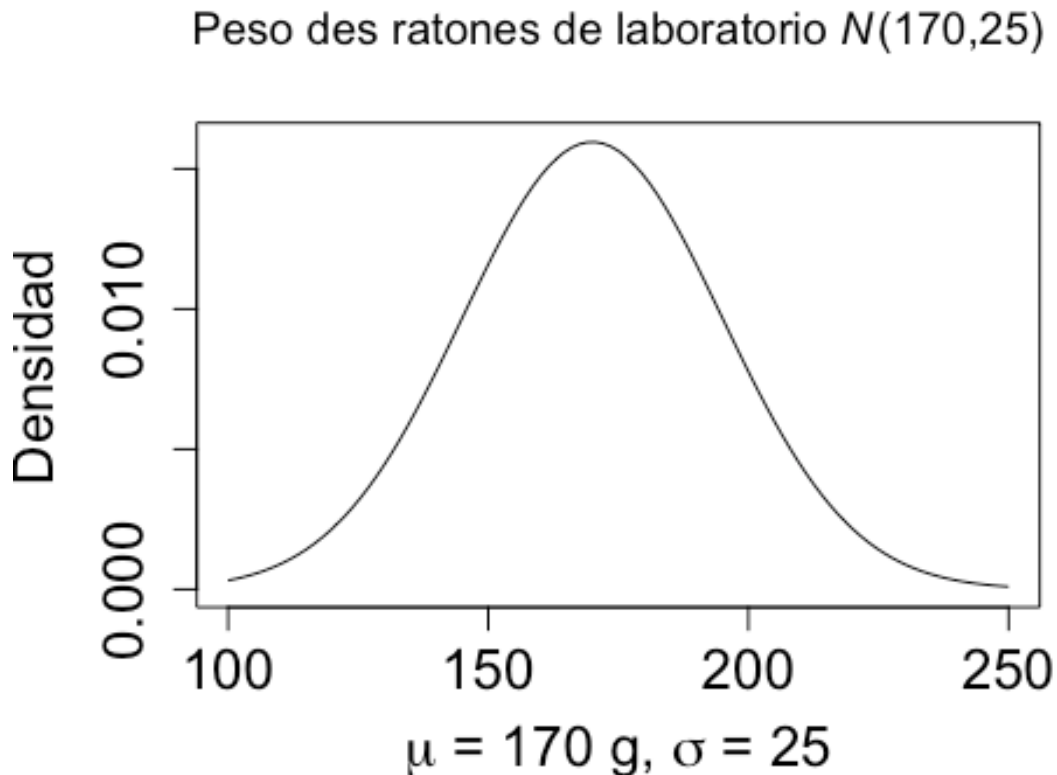
La distribución normal: la media determina la posición

```
x<-seq(-10,10,0.1)
plot(x=0,y=0,ylim=c(0,0.16),xlim=c(-
40,40),xlab=expression(italic(mu)),ylab="Densidad",cex.lab=1.5,
cex.axis=1.5, type="n")
curve(dnorm(x, mean=2, sd=5),-18,20, add=TRUE)
text(-6,0.075, expression(paste(mu, " = 2")), cex=1.5)
curve(dnorm(x, mean=20, sd=5),0,40,
ylim=c(0,0.16),xlab=expression(paste(mu, " = 20")),
ylab="Densidad",cex.lab=1.5,cex.axis=1.5, add=TRUE, col="blue")
text(30,0.075, expression(paste(mu, " = 20")), cex=1.5)
curve(dnorm(x, mean=-20, sd=5),-35,-5, add=TRUE, col="red")
text(-30,0.075, expression(paste(mu, " = -20")), cex=1.5)
```
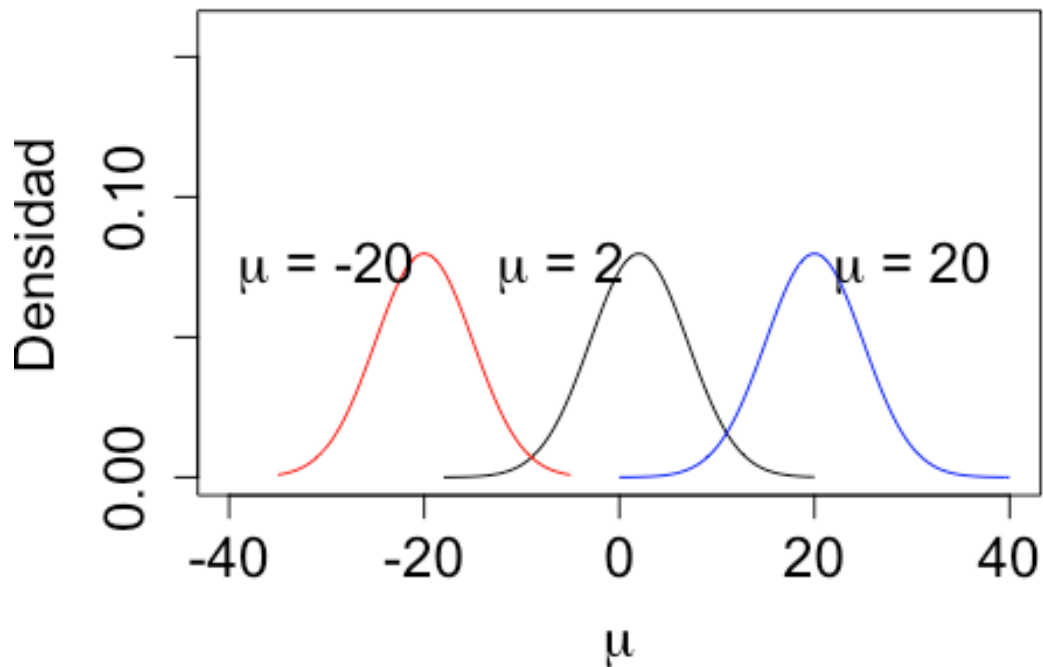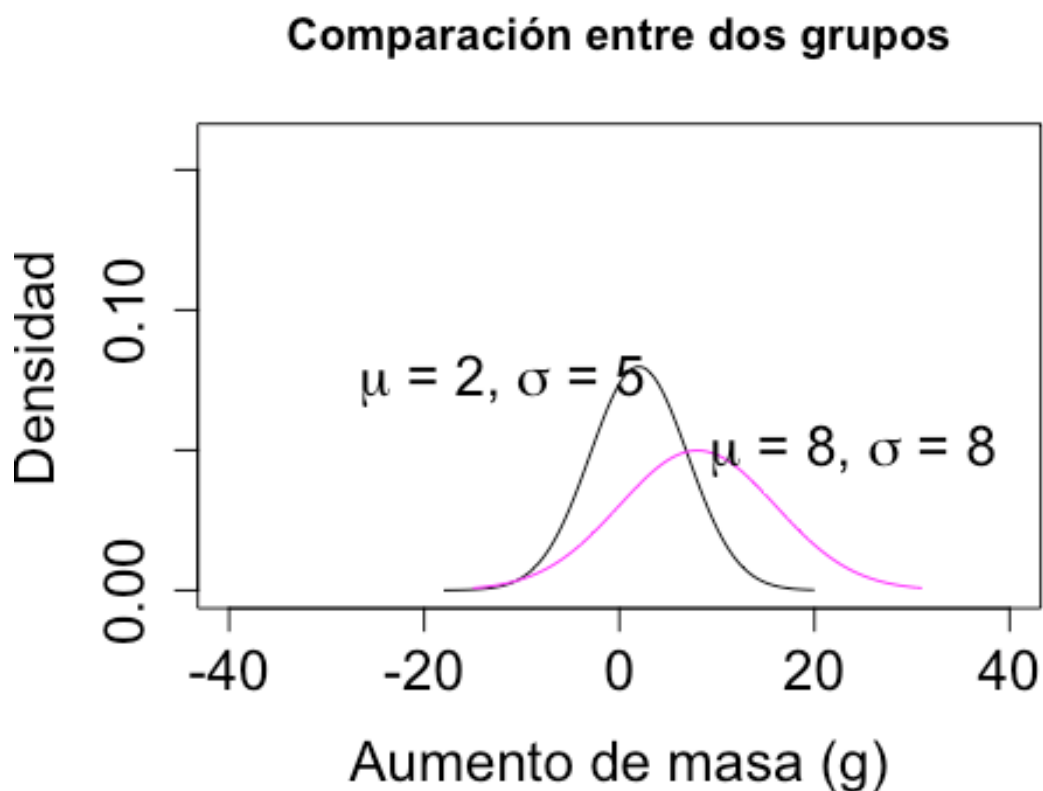
La distribución normal: la varianza determina la forma

```r
x<-seq(-10,10,0.1)
plot(x=0,y=0,ylim=c(0,0.16),xlim=c(-
40,40),xlab=expression(italic(mu)),ylab="Densidad",cex.lab=1.5,
cex.axis=1.5, type="n")
curve(dnorm(x, mean=2, sd=5),-22,26, add=TRUE)
text(-8,0.075, expression(paste(sigma, " = 5")), cex=1.5)
x<-seq(-10,10,0.1)
curve(dnorm(x, mean=2, sd=2.5),-18,20, ylim=c(0,0.16), col="blue",
add=TRUE)
text(-7,0.15, expression(paste(sigma, " = 2.5")), cex=1.5)
```
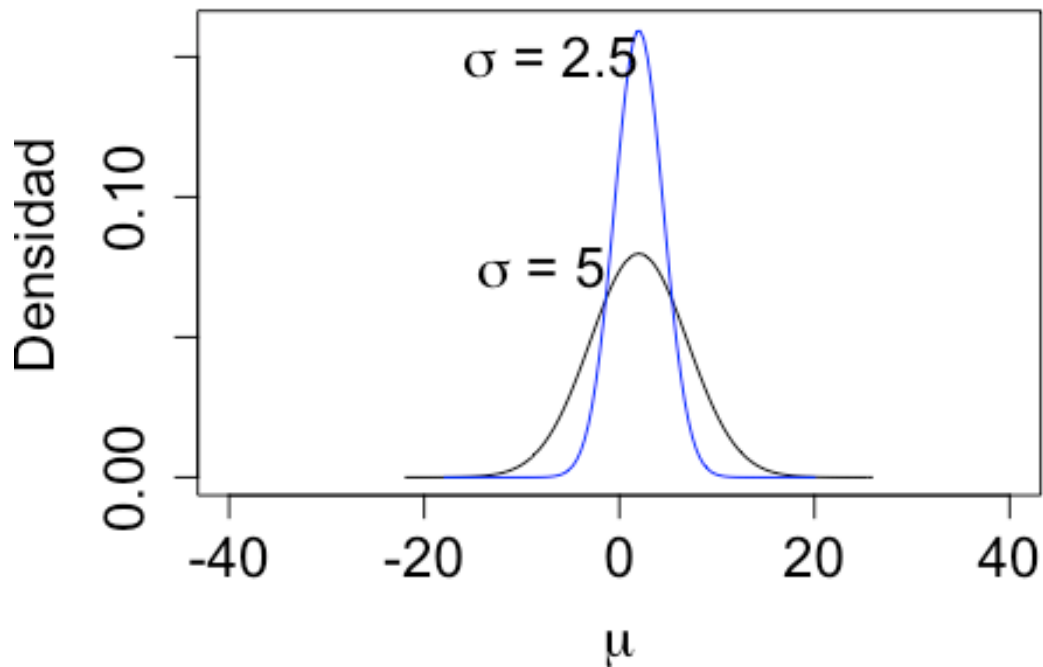
Comparar dos muestras con varianzas distintas

```r
x<-seq(-10,10,0.1)
plot(x=0,y=0,ylim=c(0,0.16),xlim=c(-40,40),xlab="Aumento de masa
(g)",ylab="Densidad",cex.lab=1.5, cex.axis=1.5, type="n",
main="Comparación entre dos grupos")
##abline(v=2,lty=2, lwd=3)
curve(dnorm(x, mean=2, sd=5),-18,20, add=TRUE)
text(-12,0.075, expression(paste(mu, " = 2, ", sigma, " = 5")), cex=1.5)
curve(dnorm(x, mean=8, sd=8),-15,31, add=TRUE, col="magenta")
text(24,0.05, expression(paste(mu, " = 8, ", sigma, " = 8")), cex=1.5)
```

Comparación entre dos grupos

$\mu = 2, \sigma = 5$

$\mu = 8, \sigma = 8$

Comparar dos muestras con misma varianza

```
x<-seq(-10,10,0.1)
plot(x=0,y=0,ylim=c(0,0.16),xlim=c(-40,40),xlab="Aumento de masa
(g)",ylab="Densidad",cex.lab=1.5, cex.axis=1.5, type="n",
main="Comparación entre dos grupos")
curve(dnorm(x, mean=2, sd=5),-18,20, add=TRUE)
text(-12,0.075, expression(paste(mu, " = 2, ", sigma, " = 5")), cex=1.5)
curve(dnorm(x, mean=8, sd=5),-15,31, add=TRUE, col="blue")
text(24,0.05, expression(paste(mu, " = 8, ", sigma, " = 5")), cex=1.5)
```

Comparación entre dos grupos

$\mu = 2, \sigma = 5$

$\mu = 8, \sigma = 5$

Densidad

0.10

0.00

-40   -20   0   20   40

Aumento de masa (g)
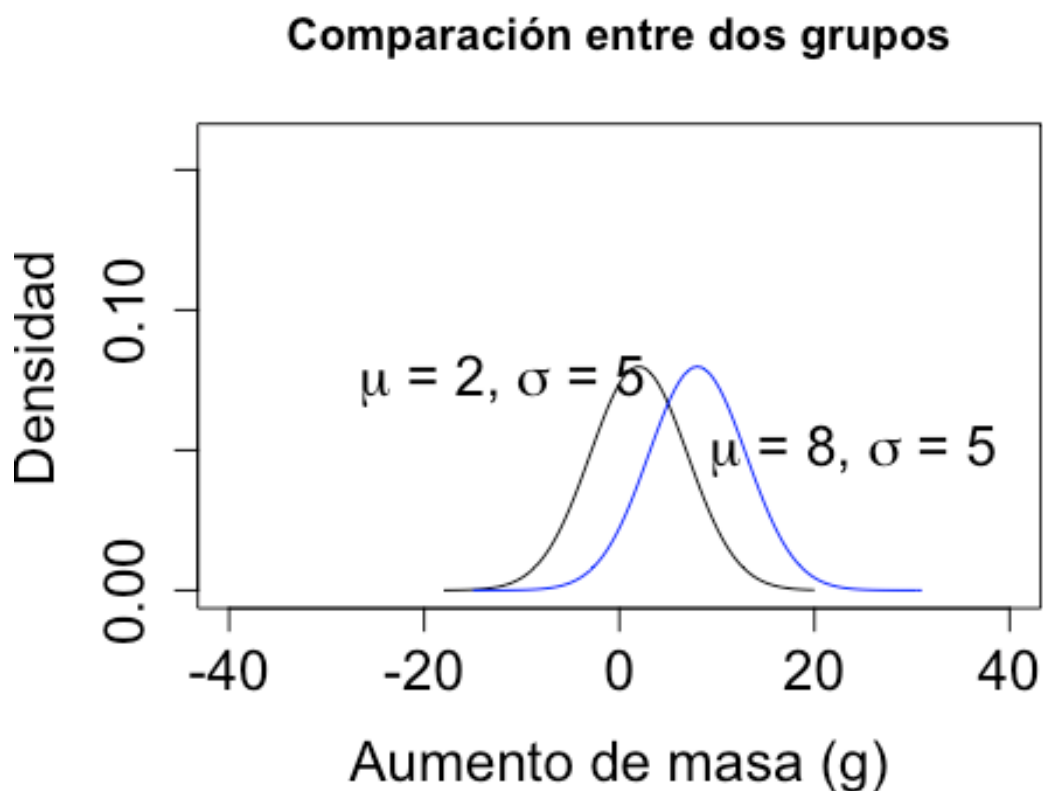
Distribución Poisson

```r
x3<-rpois(100, lambda=3)
x6<-rpois(100, lambda=6)
x9<-rpois(100, lambda=9)
x12<-rpois(100, lambda=12)
data3<-table(x3)
data6 <- table(x6)
data9 <- table(x9)
data12 <- table(x12)
par(mfrow=c(2,2))
plot(data3, ylab="Frecuencia", xlab="Especies", col="red",
main="Lambda=3")
plot(data6, ylab="Frecuencia", xlab="Especies",
col="blue",main="Lambda=6")
plot(data9, ylab="Frecuencia", xlab="Especies",
col="green",main="Lambda=9")
plot(data12, ylab="Frecuencia", xlab="Especies",
col="purple",main="Lambda=12")
```
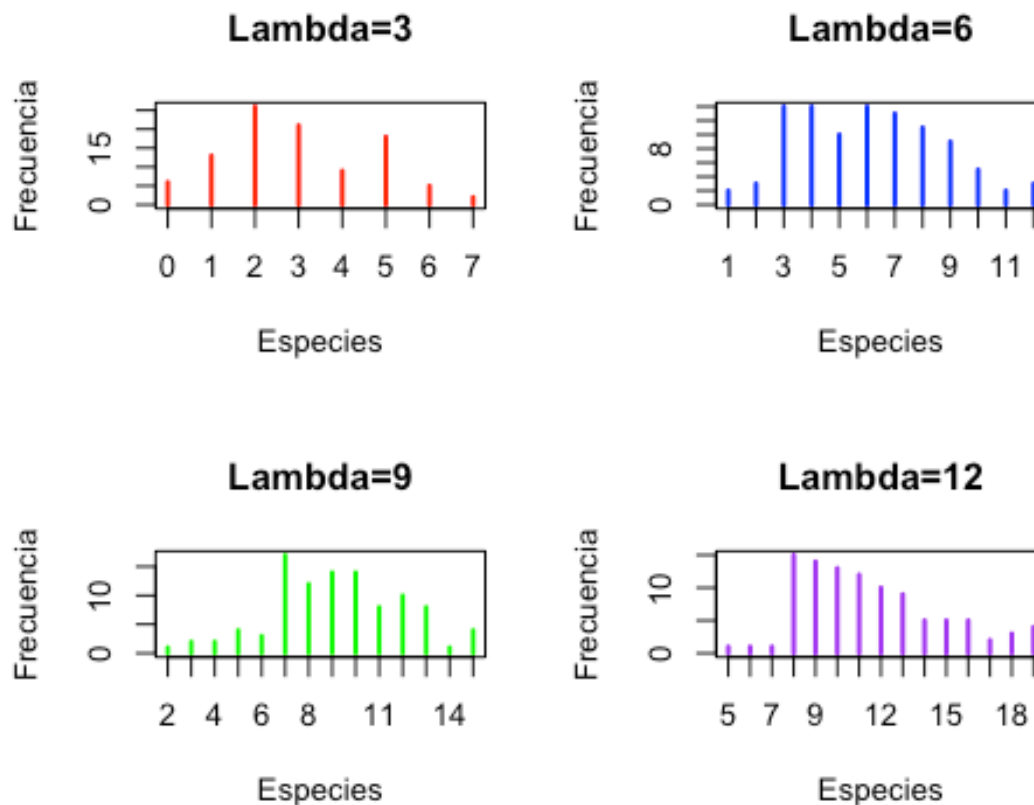
Distribución Binomial

```r
x02<-rbinom(100,50,0.2)
x05<-rbinom(100,50,0.5)
x09<-rbinom(100,50,0.9)
x10<-rbinom(100,50,1)
data02<-table(x02)
data05 <- table(x05)
data09 <- table(x09)
data10 <- table(x10)
par(mfrow=c(2,2))
plot(data02, ylab="Frecuencia", xlab="Ensayos", col="red", main="p=0.2")
plot(data05, ylab="Frecuencia", xlab="Ensayos", col="blue",main="p=0.5")
plot(data09, ylab="Frecuencia", xlab="Ensayos", col="green",main="p=0.9")
plot(data10, ylab="Frecuencia", xlab="Ensayos", col="purple",main="p=1")
```

p=0.2

Frecuencia

Ensayos

p=0.5

Frecuencia

Ensayos

p=0.9

Frecuencia

Ensayos

p=1

Frecuencia

Ensayos