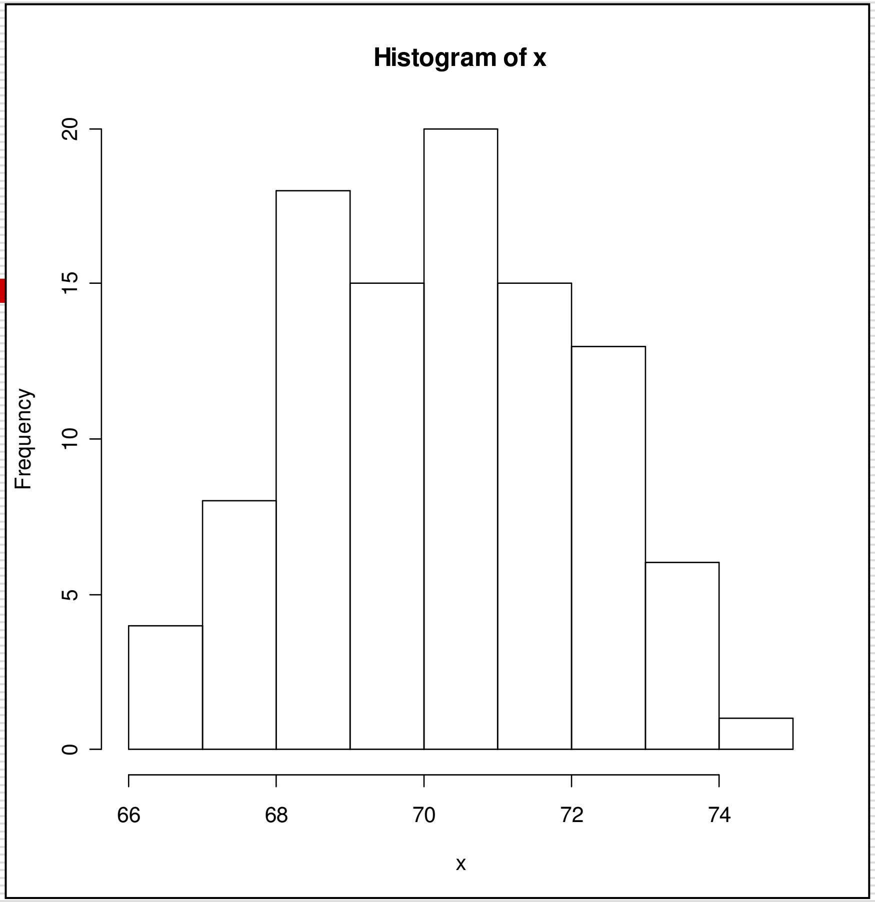
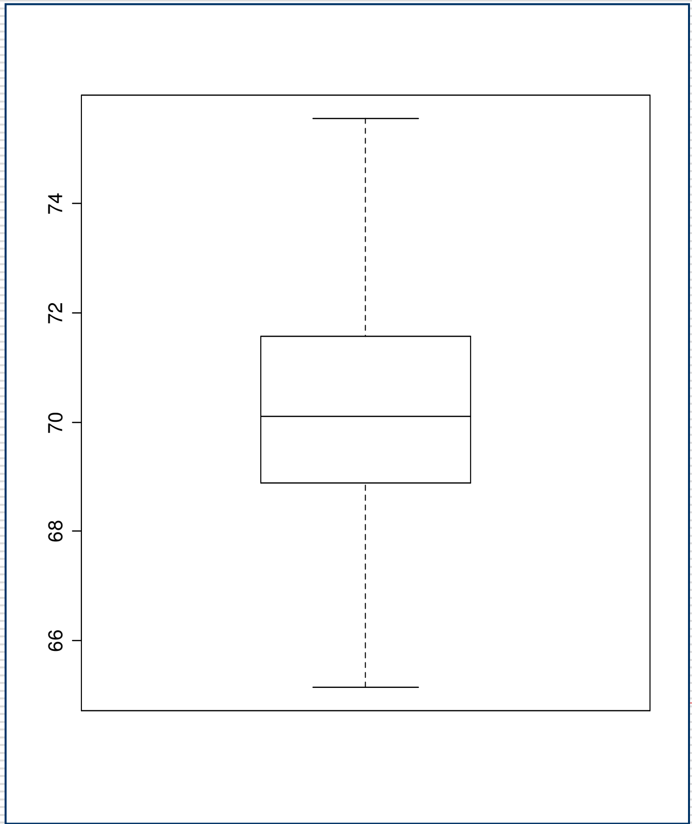
Cálculos con distribución normal en **R**

Obtención de una muestra de una distribución normal

- Muestra de tamaño 100 de una distribución N(70,2)

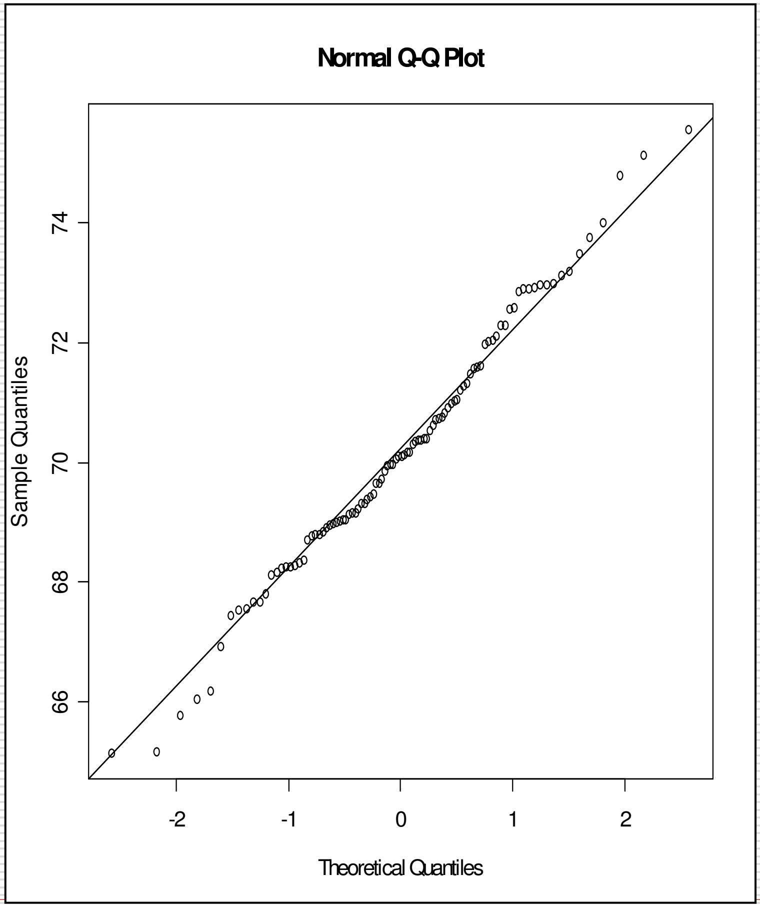





```
# Generar la muestra
x <- rnorm(100,70,2)
# Histograma
hist(x)
# Box Plot
boxplot(x)
```

Gráfica Q-Q

- Para obtener un gráfico Q-Q utilizamos la función
  - **`qqline(x)`**



Función de distribución y cuantiles

- Cálculo de la función de distribución
  - Untilizar la función **pnorm(x,μ,σ)**
  - Ejemplo para una N(70,2)

```
> pnorm(75,70,2)
[1] 0.9937903
> pnorm(72,70,2)
[1] 0.8413447
> pnorm(69,70,2)
[1] 0.3085375
>
```

$$\rightarrow \quad P(X \leq 69) = 0.3085$$

Cuantiles

- Cálculo de quantiles de una distribución normal
  - Untilizar la función **qnorm(q,μ,σ)**
  - Ejemplo para una N(70,2)

```
> qnorm(0.95,0,1)
[1] 1.644854
>
qnorm(0.95,70,2)
[1] 73.2897
>
```

$\rightarrow$ $P(Z \leq 1.644854) = 0.95$

$\rightarrow$ $P(X \leq 73.2897) = 0.95$

Cálculo del intervalo de normalidad

- Definimos la función

```
IntNorm <- function(m=0,s=1,conf=0.95)
{         alfa=1-conf
          z <- qnorm(1-alfa/2,0,1)
          c(m-z*s,m+z*s)
}
```

```
> IntNorm()
[1] -1.959964  1.959964
> IntNorm(m=70,s=2)
[1] 66.08007 73.91993
> IntNorm(m=70,s=2,conf=0.90)
[1] 66.71029 73.28971
>
```

Probabilidad de un intervalo

- Definimos la función

```
PIntervalo <- function(a,b,m=0,s=1)
{
        pnorm(b,m,s)-pnorm(a,m,s)
}
```

```
> PIntervalo(-1.96,1.96)
[1] 0.9500042
> PIntervalo(70,72,m=70,s=2)
[1] 0.3413447
> PIntervalo(70,72,m=70,s=1)
[1] 0.4772499
>
```

Distribución de la media de n observaciones

- Si X es una distribución N(μ,σ), entonces la media aritmética de n observaciones se distribuye:

$$\overline{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- La varianza de la media de n observaciones disminuye en función de n. Las media de muestra grandes son más parecidas que las media de muestras pequeñas.

Distribución de la media de n observaciones

- Consideremos una N(100,3) y muestras de tamaño 25.
- La media muestral se distribuirá como una N(100,3/5).
- Veamos que esto es cierto generando 300 muestras de tamaño 25 de una N(100,3) y calculando sus medias
- Primero definimos las funciones:

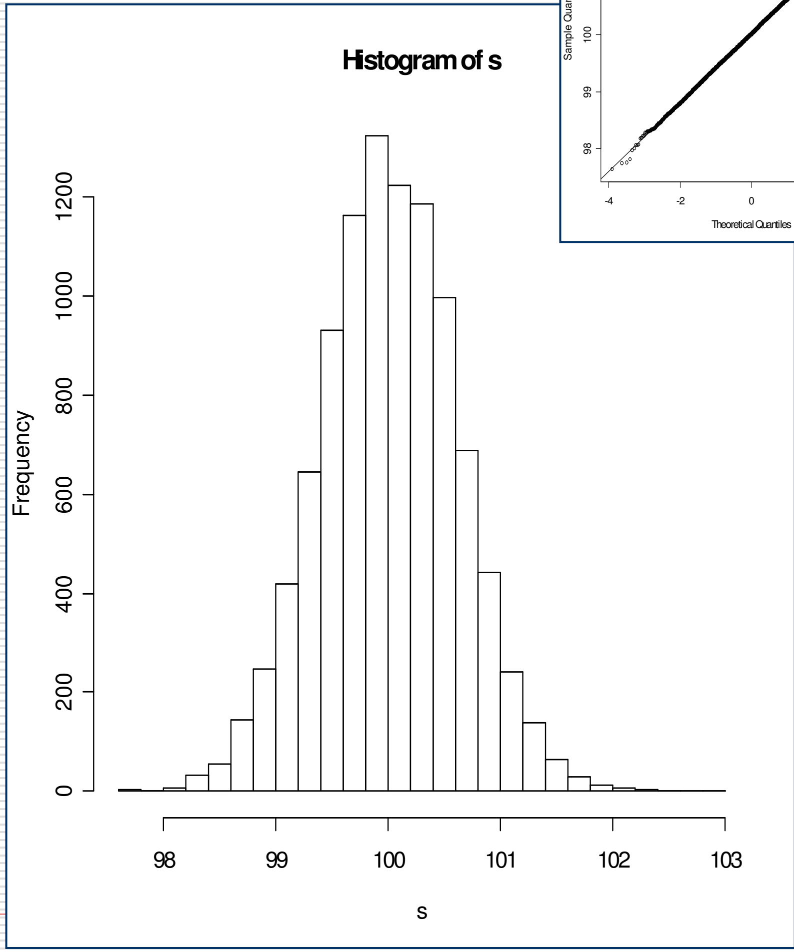
```
sample <- function(n,m,s)
{
        x <- rnorm(n,m,s)
        mean(x)
}

MeansSample <- function(nsamples,n,m,s)
{
        x <- c(1:nsamples)
        for (i in 1:nsamples) x[i]<-sample(n,m,s)
        x
}
```

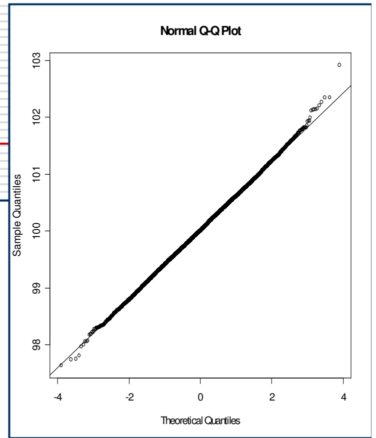
Distribución de la media de n observaciones

- Consideremos una N(100,3) y muestras de tamaño 25.
- La media muestral se distribuirá como una N(100,3/5), es decir N(100, 0.6).
- Veamos que esto es cierto generando 10000 muestras de tamaño 25 de una N(100,3) y calculando sus medias

```
> s <- MeansSample(10000,25,100,3)
> hist(s,nclass=30)
```

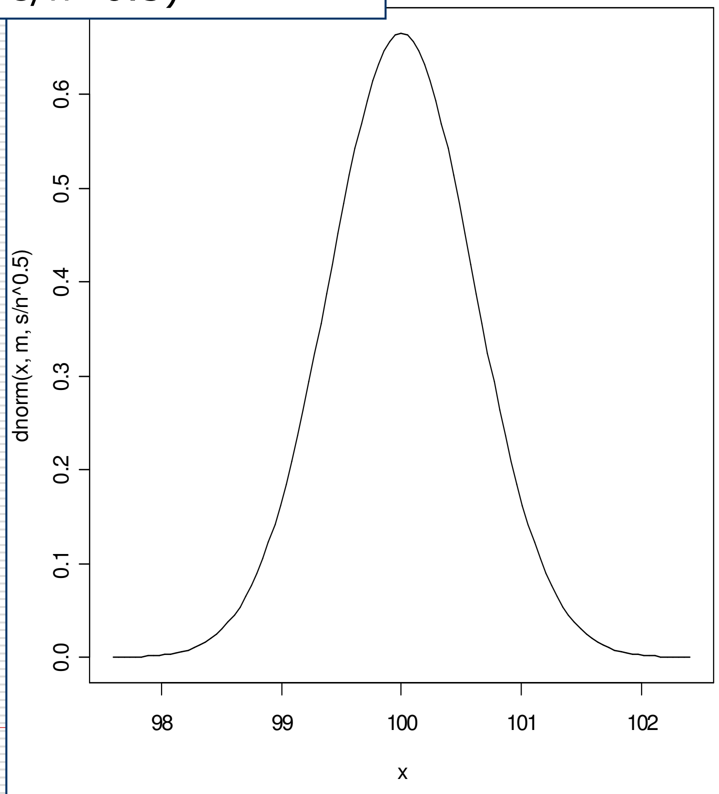
```
> c(mean(s),var(s)^0.5)
[1] 100.010012   0.607991
>
```





Dibujar una curva normal
Comportamiento de las medias de muestras de tamaño 25

```
> m<-100
> s<-3
> n<-25
> curve(dnorm(x,m,s/n^0.5),m-4*s/n^0.5,m+4*s/n^0.5)
```



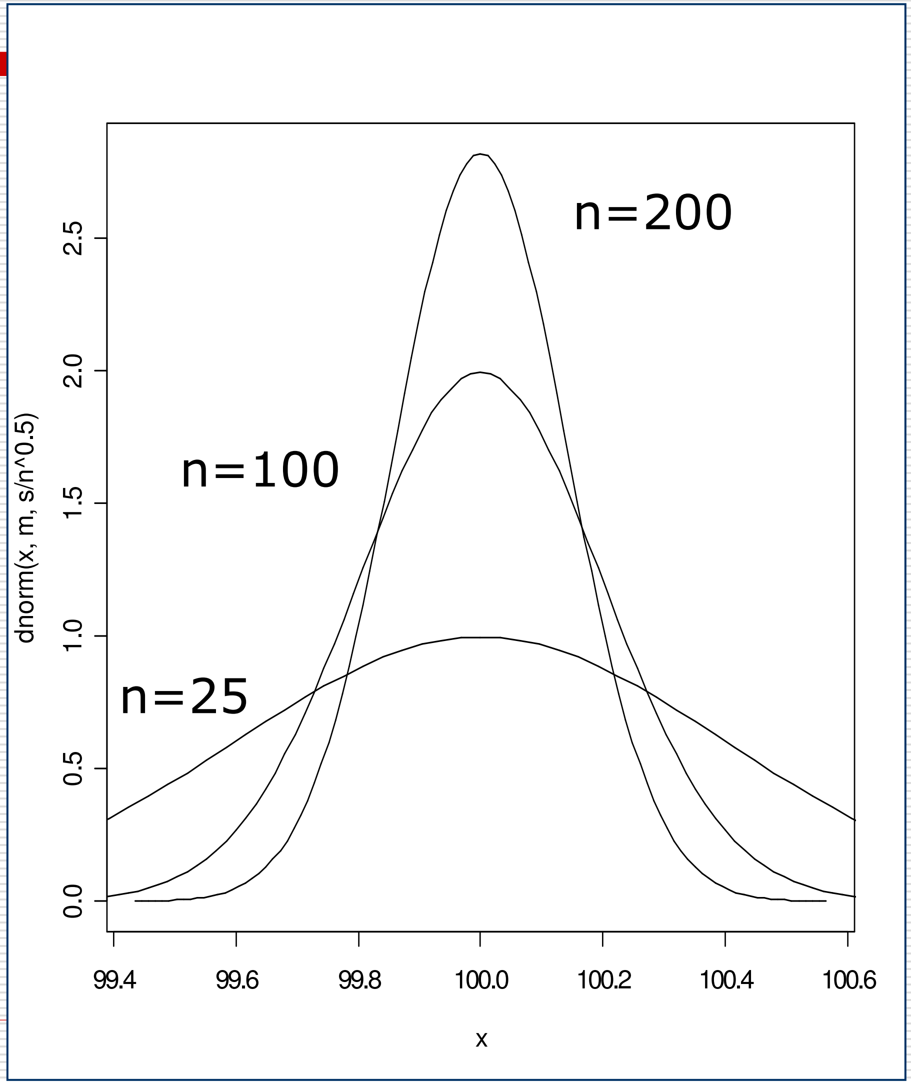
Comportamiento de las medias muestrales

- Dibujaremos la función de densidad de la media de muestras de distintos tamaños de una N(100,2)
- Las instrucciones son:

```
m <- 100
s <- 2

n <- 200
curve(dnorm(x,m,s/n^0.5),m-4*s/n^0.5,m+4*s/n^0.5)

n <- 100
curve(dnorm(x,m,s/n^0.5),m-4*s/n^0.5,m+4*s/n^0.5,add=TRUE)

n <- 25
curve(dnorm(x,m,s/n^0.5),m-4*s/n^0.5,m+4*s/n^0.5,add=TRUE)
```

Comportamiento de las medias muestrales

- Dibujaremos la función de densidad de la media de muestras de distintos tamaños de una N(100,2)



Obtención de muestras de dos poblaciones

- Consideremos la producción de dos bioreactores (gramos por litro). Un primer proceso tiene una producción N(5,0.5), mientras que un segundo proceso tiene una producción N(5.5, 0.4)
  - Se pide obtener una muestra aleatoria de 25 observaciones de cada proceso en fase estacionaria.
  - Comparar los resultados mediante Box Plots
  - Calcular la media de producción de cada proceso

Obtención de muestras de dos poblaciones

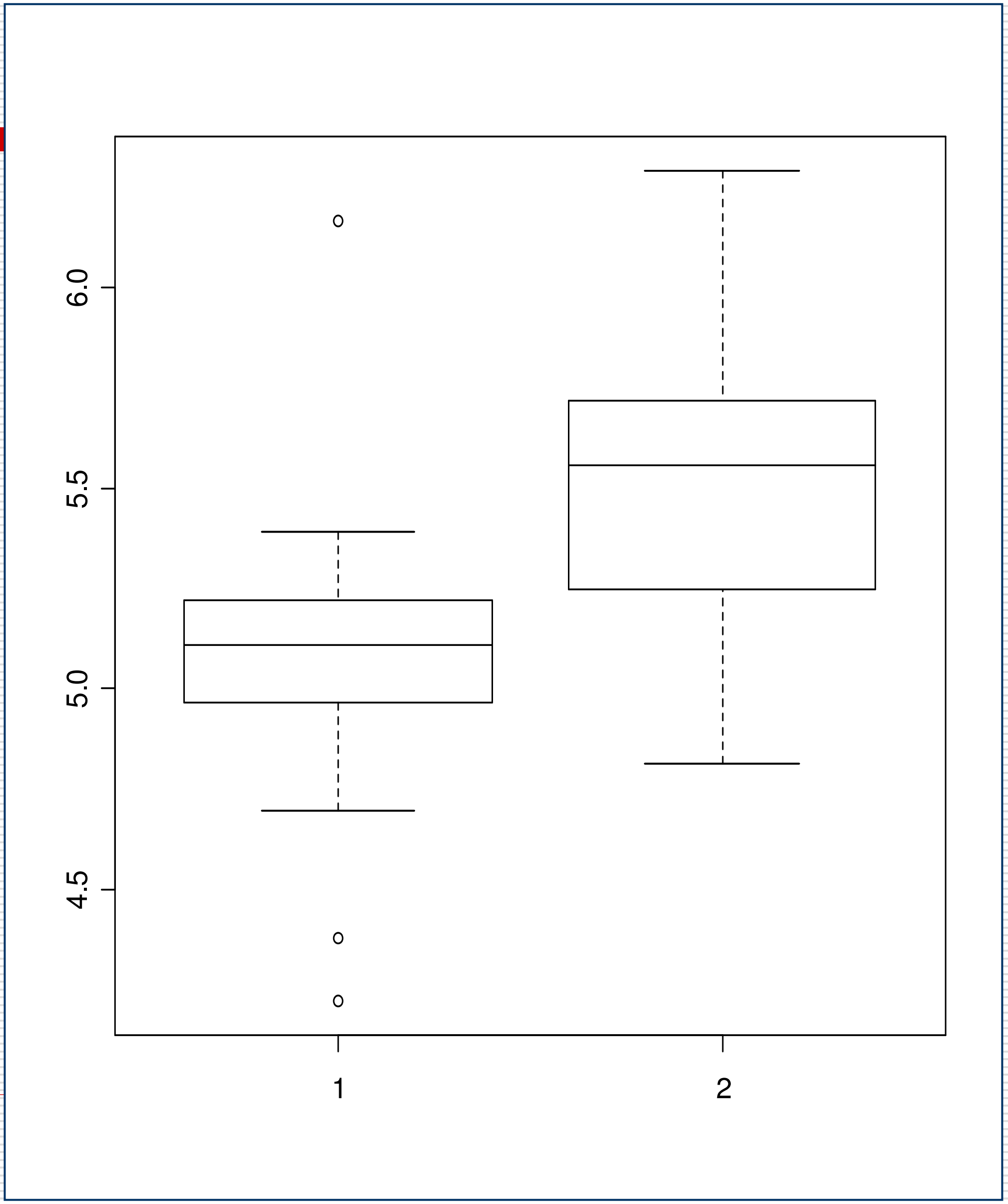
- Obtención de las muestras
  - Creamos cada grupo con su etiqueta (1:Grupo 1, 2:Grupo 2)
  - Creamos un **`data.frame`** para cada grupo
  - Juntamos los grupos en un **`data.frame`**

```
grup <- rep(1,25)
x <- rnorm(25,5,0.5)
d1 <- data.frame(grup,x)

grup <- rep(2,25)
x <- rnorm(25,5.5,0.4)
d2 <- data.frame(grup,x)

data <- rbind(d1,d2)
```

Obtención de muestras de dos poblaciones

- Box plot por grupos

```
boxplot(x~grup,data=data)
```



Cálculo de medias para cada grupo

- El **`data.frame`** que contiene los datos se llama **`data`**. Las variables son **`grup`** y **`x`**.
  - Podemos seleccionar los casos correspondientes un cierto valor de **`x`** mediante la instrucción

```
>  data[data$grup==1,]
```

  - Por lo tanto, las medias de cada grupo pueden obtenerse haciendo:

```
> mean(data[data$grup==1,]$x)
[1] 5.08195
> mean(data[data$grup==2,]$x)
[1] 5.484446
>
```