PracticaProbabilidad

Práctica 2 - Modelos de distribución de probabilidad

Ejercicio 1

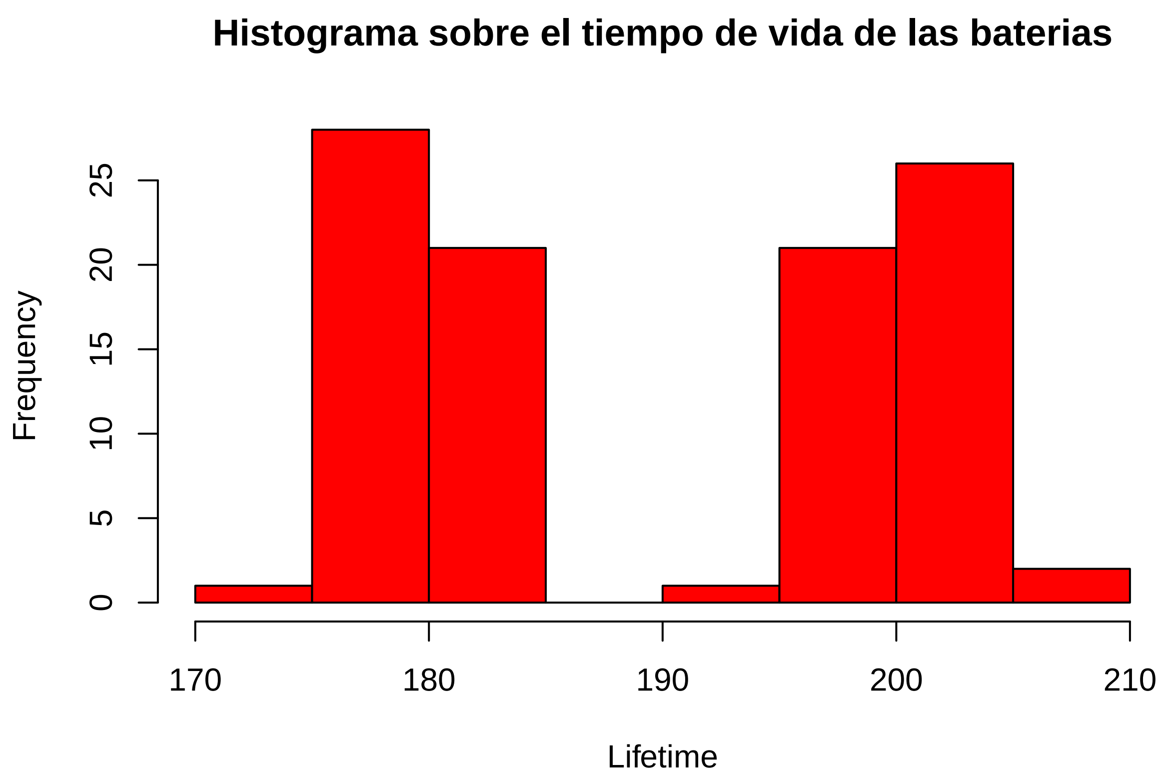
```r
library(scales)
library(PASWR2)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```r
library(normtest)
library(nortest)
datos <- BATTERY
```

Histograma Relizamos un histograma sobre la variable “Lifetime”.

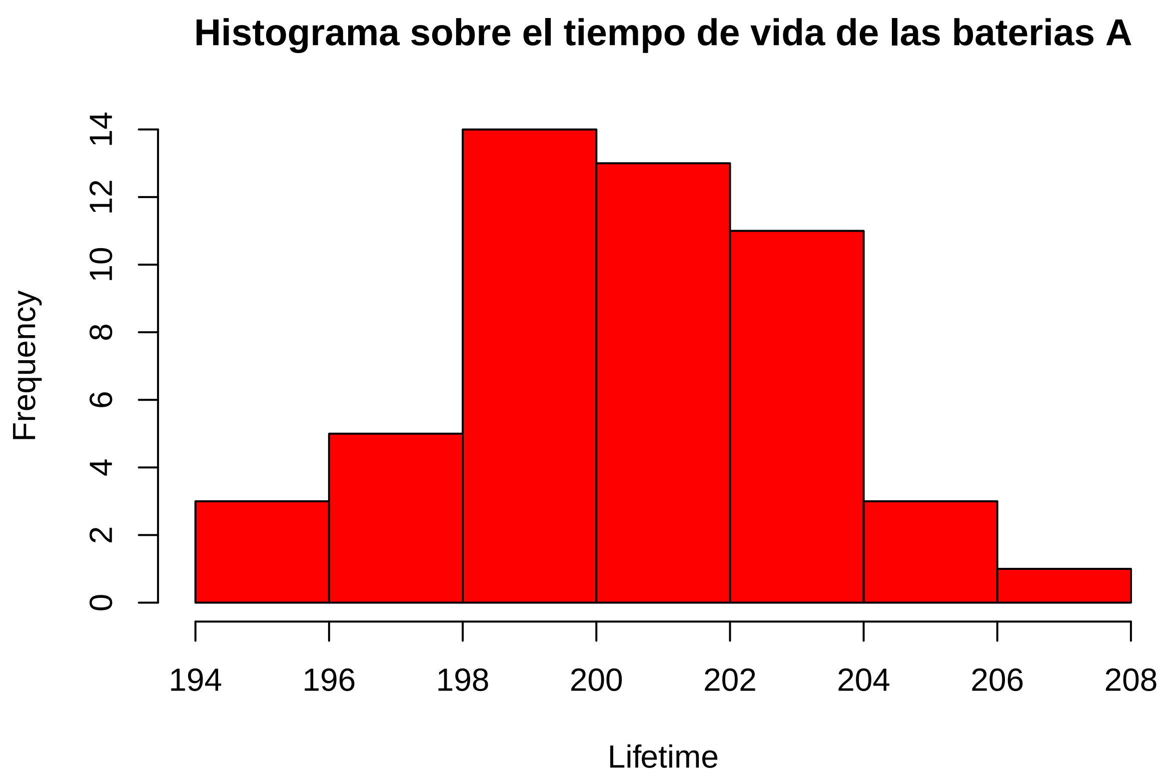


Conjuntos de datos Separamos los datos en función del campo “Facility”.

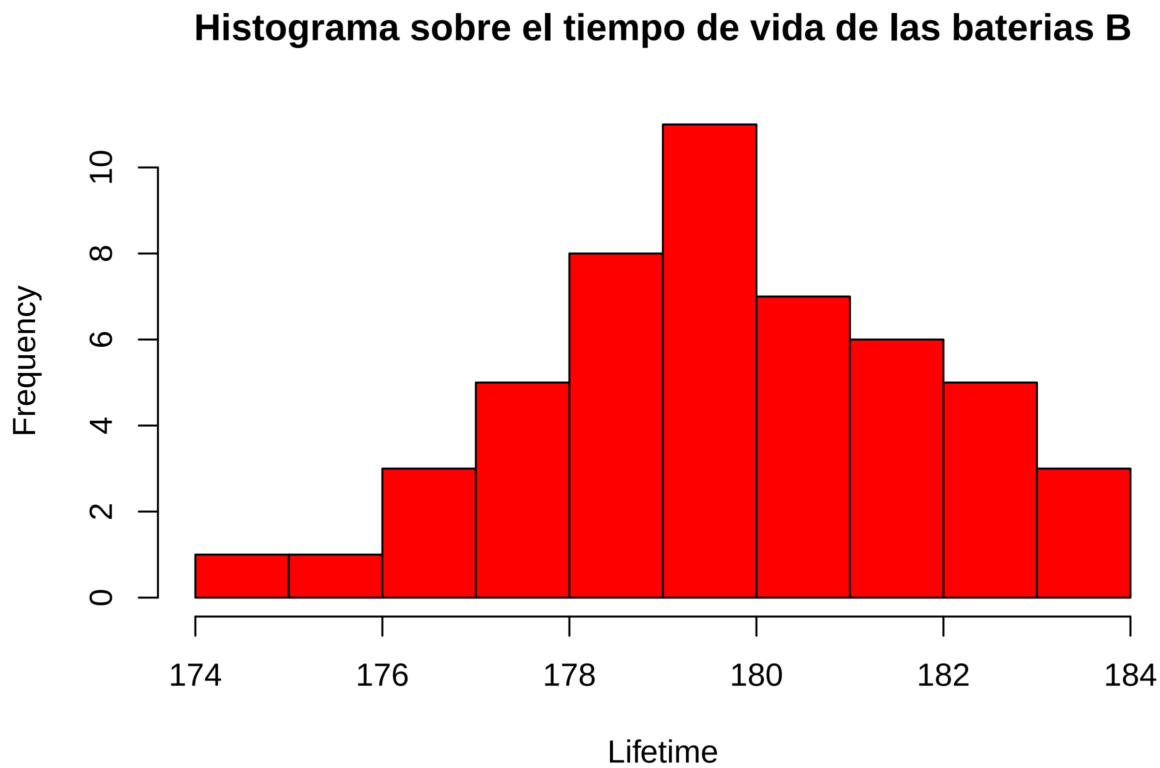
```r
datos_a <- datos[which(datos$facility=='A'),]
datos_b <- datos[which(datos$facility=='B'),]
datos_a1 <- datos$lifetime[datos$facility=='A']
datos_b1 <- datos$lifetime[datos$facility=='B']
```

Histogramas independientes Procedemos a crear un histograma para cada conjunto de datos y observar sus resultados.

```r
hist(datos_a$lifetime, col='red', main="Histograma sobre el tiempo de vida de las baterias A",
     xlab="Lifetime")
```




```r
hist(datos_b$lifetime, col='red', main="Histograma sobre el tiempo de vida de las baterias B",
     xlab="Lifetime")
```



Como podemos observar, a simple vista ya se puede apreciar que ambos histogramas siguen una distribución normal, pero realizaremos test de normalidad.

```r
ad.test(datos_a1)
```

```
##
##  Anderson-Darling normality test
##
## data:  datos_a1
## A = 0.35332, p-value = 0.4511
```

```r
jb.norm.test(datos_a1)
```

```
##
##  Jarque-Bera test for normality
##
## data:  datos_a1
## JB = 0.24858, p-value = 0.8765
```

```r
hegazy1.norm.test(datos_a1, nrepl=20000)
```

```
##
##  Hegazy-Green test for normality
##
## data:  datos_a1
## T = 0.10812, p-value = 0.5154
```

```r
ad.test(datos_b1)
```

```
##
##  Anderson-Darling normality test
##
## data:  datos_b1
## A = 0.19478, p-value = 0.8873
```

```r
jb.norm.test(datos_b1)
```

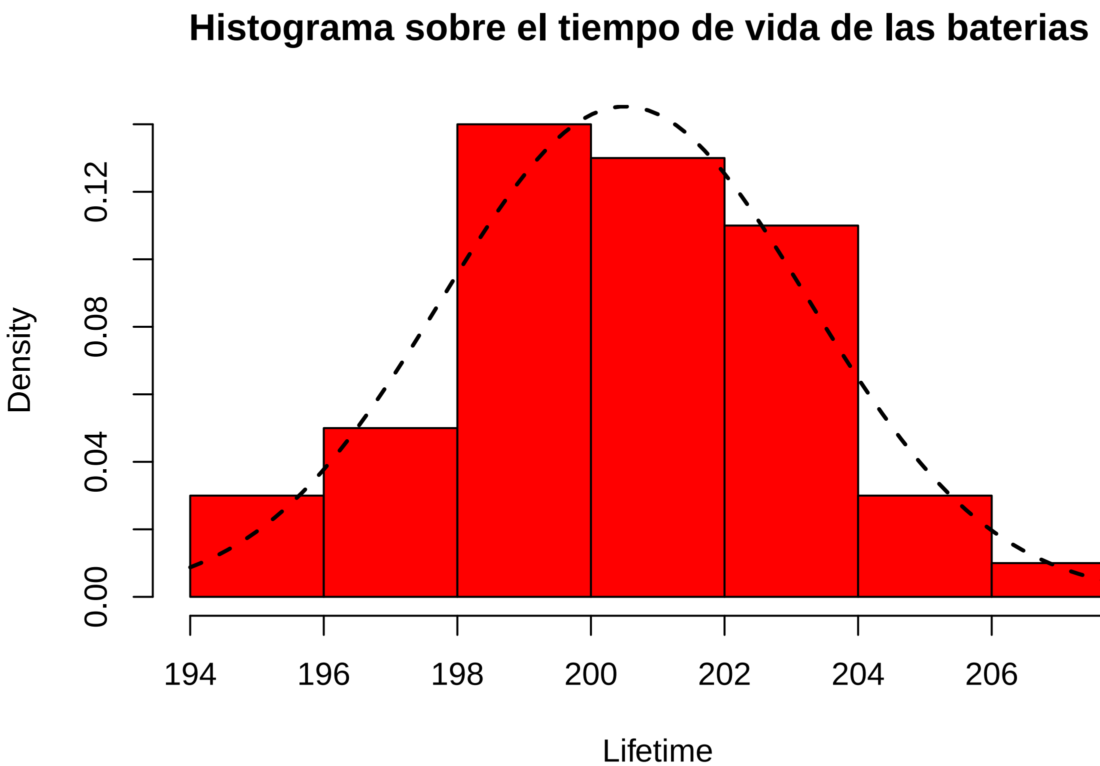
```
##
##  Jarque-Bera test for normality
##
## data:  datos_b1
## JB = 0.68004, p-value = 0.6605
```

```r
hegazy1.norm.test(datos_b1, nrepl=20000)
```

```
##
##  Hegazy-Green test for normality
##
## data:  datos_b1
## T = 0.086234, p-value = 0.8782
```

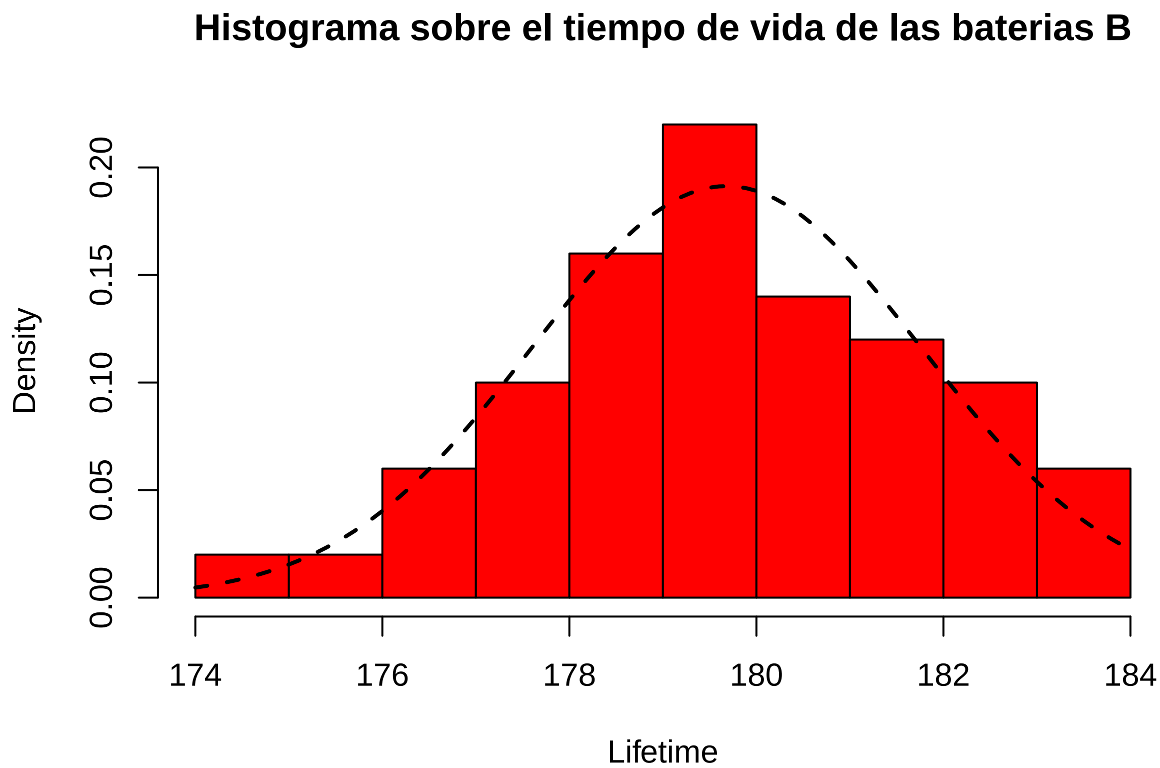
Podemos observar que los p-value son menor que la significancia, de modo que aceptamos la normalidad de los datos.

```r
hist(datos_a$lifetime, col='red', main="Histograma sobre el tiempo de vida de las baterias A",
     xlab="Lifetime", freq=FALSE)
curve(dnorm(x,mean(datos_a$lifetime),sd(datos_a$lifetime)), add = TRUE, lwd = 2, lty = 2)
```



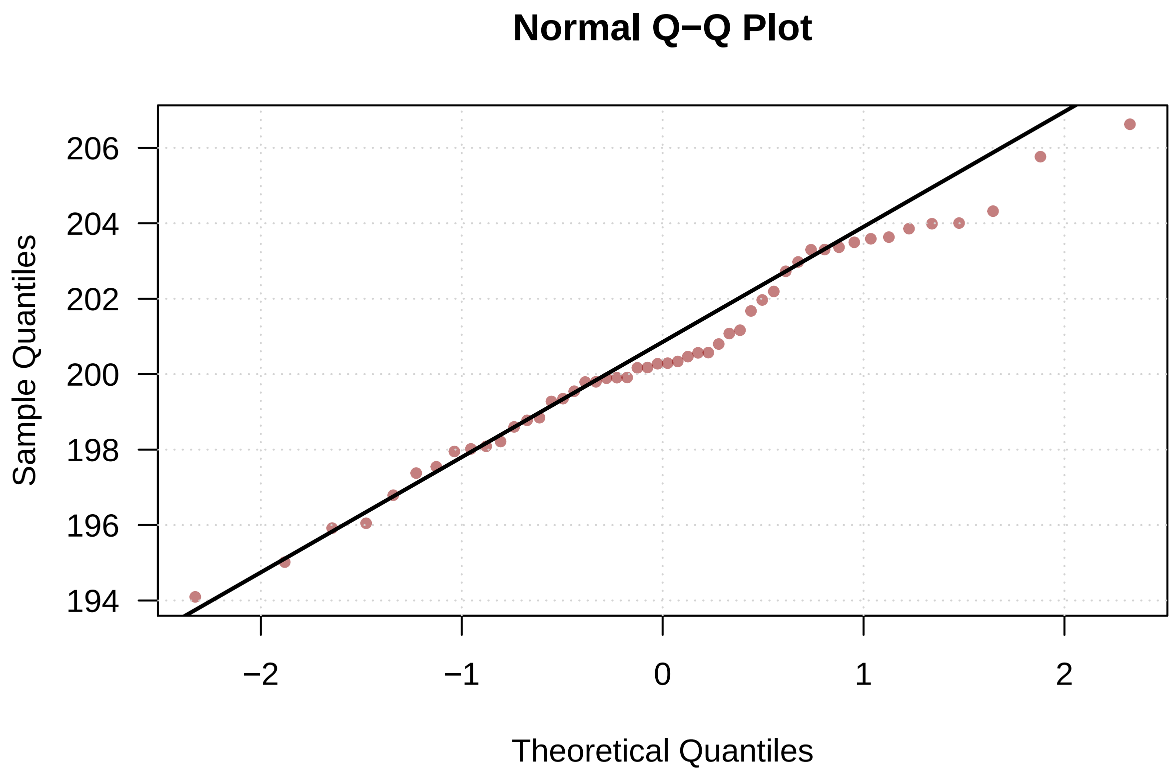
Análisis de histogramas

```
hist(datos_b$lifetime, col='red', main="Histograma sobre el tiempo de vida de las baterias B",
     xlab="Lifetime", freq=FALSE)
curve(dnorm(x,mean(datos_b$lifetime),sd(datos_b$lifetime)), add = TRUE, lwd = 2, lty = 2)
```

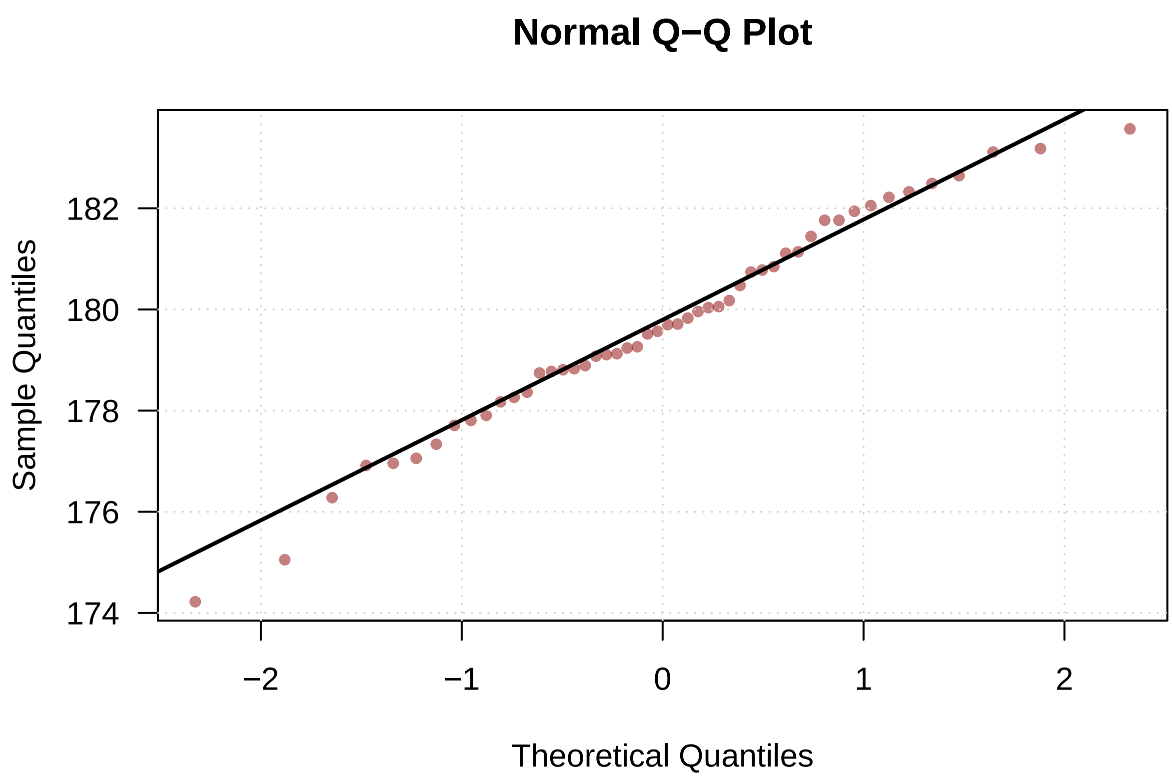


Gráficos quantil-quantil Representando los gráficos quantil-quantil, también podemos ver como sigue una distribución normal.

```
qqnorm(datos_a$lifetime, pch = 20, col =alpha("red4", 0.5),las = 1)
grid()
qqline(datos_a$lifetime, lwd = 2)
```




```r
qqnorm(datos_b$lifetime, pch = 20, col =alpha("red4", 0.5),las = 1)
grid()
qqline(datos_b$lifetime, lwd = 2)
```



Ejercicio 2

Estimación puntual de media y desviación típica. Calculamos la media y desviación típica en ambos conjuntos de datos.

```
mean_a = mean(datos_a$lifetime)
sd_a = sd(datos_a$lifetime)
mean_a
```

```
## [1] 200.5087
```

```
sd_a
```

```
## [1] 2.745777
```

```
mean_b = mean(datos_b$lifetime)
sd_b = sd(datos_b$lifetime)
mean_b
```

```
## [1] 179.6805
```

```
sd_b
```

```
## [1] 2.084977
```

```
pnorm(210,mean=mean_a,sd=sd_a,lower.tail=FALSE)
```

Probabilidad de que una bateria del tipo A dure más de 210 horas.

```
## [1] 0.0002734129
```

```r
pnorm(210,mean=mean_a,sd=sd_a,lower.tail=TRUE)
```

```
## [1] 0.9997266
```

```r
prob_a2_3<-pnorm(175,mean=mean_b,sd=sd_b,lower.tail=TRUE)
prob_a2_3
```

Probabilidad de una batería del tipo B dure menos de 175 horas.

```
## [1] 0.01238792
```

```r
qnorm(0.03,mean=mean_b,sd=sd_b,lower.tail=TRUE)
```

Cuantil 0.03

```
## [1] 175.7591
```

Actividad 3

```r
pbinom(q = 0, size = 10, prob = prob_a2_3)
```

Calcula la probabilidad de que en un lote de 10 baterías, no haya ninguna defectuosa (ayuda: distribución binomial).

```
## [1] 0.8828033
```

```r
pgeom(q = 4, prob = prob_a2_3, lower.tail = FALSE)
```

Imagina que las baterías se fabrican en serie e independientemente. ¿Cuál es la probabilidad de que la batería producida en quinto lugar sea la primera defectuosa? (ayuda: distribución geométrica.)

```
## [1] 0.9395761
```

```r
1-dhyper(x=0,m=3,n=17,k=5)
```

Supongamos que en una caja de 20 baterías van 3 defectuosas. ¿Cuál es la probabilidad de que al tomar una muestra sin reposición de 5 baterías al menos una sea defectuosa? (ayuda: distribución hipergeométrica)

```
## [1] 0.6008772
```

Actividad 4

¿Cuál es la probabilidad de que un día se produzcan más de 20 baterías defectuosas?

```r
1 - ppois(lambda = 12, q = 20)
```

```
## [1] 0.01159774
```

¿Cuál es la probabilidad de que un día no salga ninguna batería defectuosa de la fábrica?

```r
dpois(0,12)
```

```
## [1] 6.144212e-06
```

La fábrica funciona de lunes a viernes. ¿Qué distribución sigue el número de baterías defectuosas porsemana? Justifica qué propiedad se aplica.

Podría seguir una distribución continua exponencial, ya que esta solo se basa en la media en la que ocurre un evento, y en este caso se puede calcular esa media a partir de la media de que ocurra cada día (no varía)

Actividad 5

Realiza una simulación de la producción semanal de baterías (recuerda: 5 días de produccción, a 1000baterías por día). Guarda los datos en un vector.

```r
a=100
b=185
set.seed(1)
simulation = rweibull(5000, shape=a, scale=b)
```

Con este nuevo proceso, ¿se mejora realmente la duración media de las baterías?

```r
esperanza = b * gamma(1+1/a)
```

Si que se mejoran con respecto al ejercicio 2, ya que tenían una esperanda de 179.68 y ahora 183.95, aunque no es muy notable la mejora

Los ingenieros no lo tienen muy claro (parece que la diferencia no es tanta en promedio y los nuevosmateriales son costosos). Para demostrarles que merece la pena, calcula la proporción de bateríasdefectuosas que producirá probablemente el nuevo proceso y compárala con el anterior (lapquecalculamos en la actividad 2).

```r
mean = mean(simulation)
sd = sd(simulation)
pnorm(174,mean=mean,sd=sd,lower.tail=TRUE)
```

```
## [1] 1.62445e-05
```

Ahora sale una probabilidad de 1.62 por diez a la menos cinco, y antes salía 0.03