Práctica 2 - Modelos de distribución de probabilidad

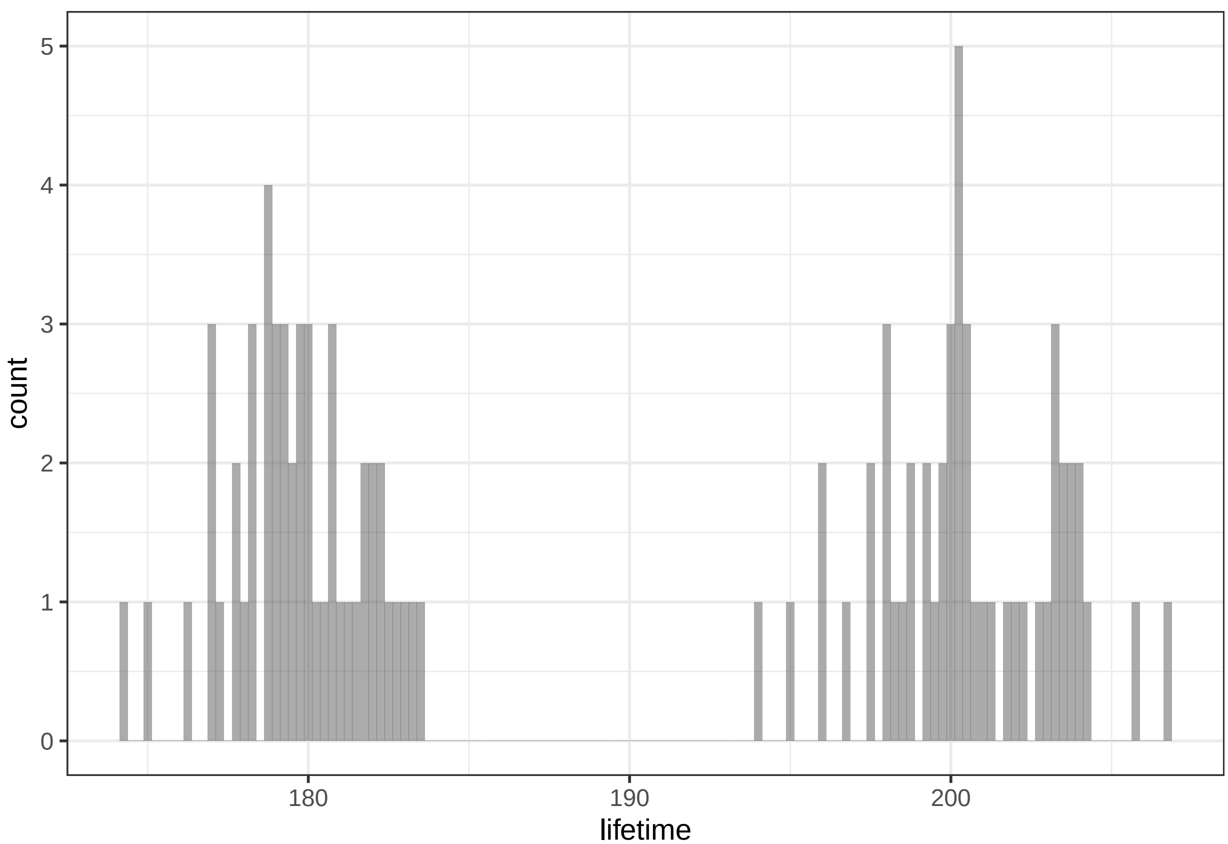
Alejandro Perela Posada, Andrea Condado Gómez y Miguel Calvo Pérez

9/11/2021

```
library(PASWR2); library(ggplot2);library(dplyr);library(scales);library(nortest)
datos <- BATTERY
```

Actividad 1

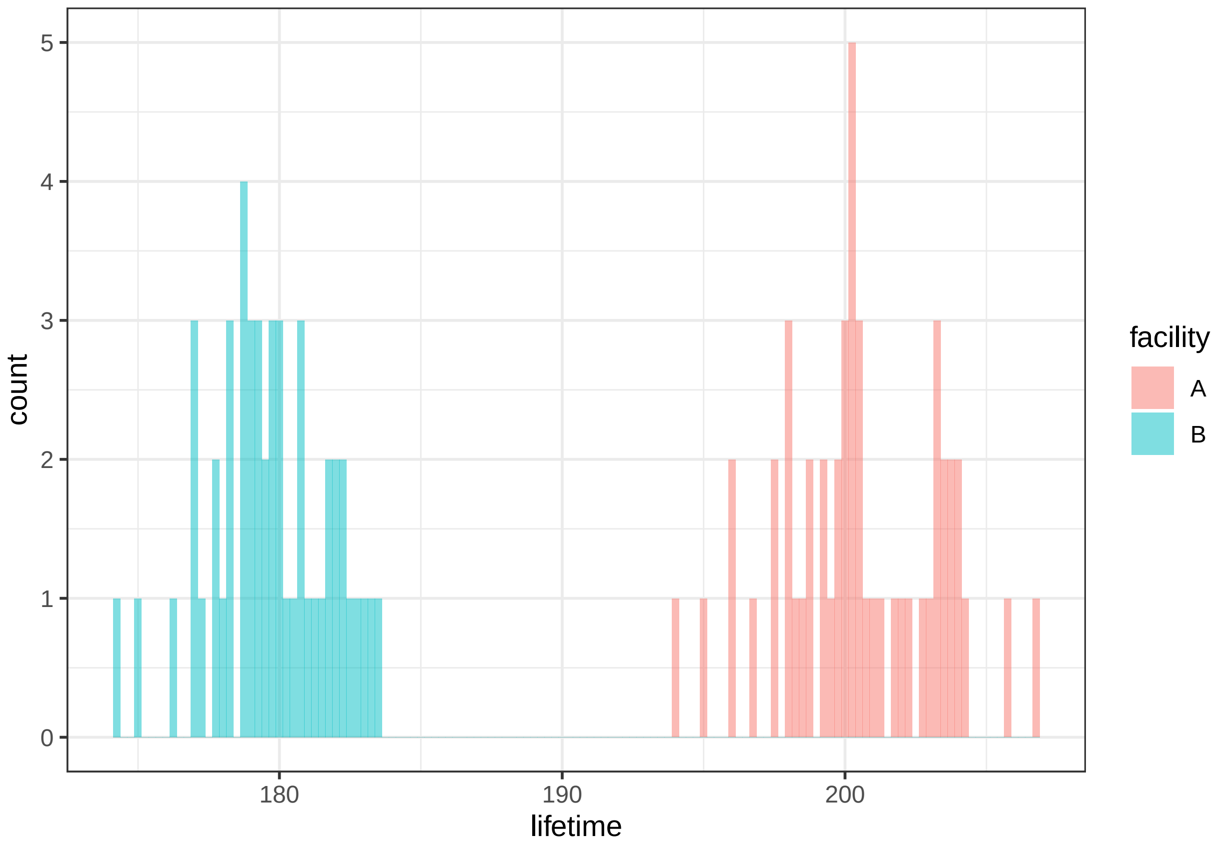
- **Realiza un histograma de todas las filas de la variable lifetime y comprueba que efectivamente nos interesa separar los datos.**



- **Crea dos conjuntos de datos diferentes para los dos tipos de baterías, por ejemplo datosA y datosB.**

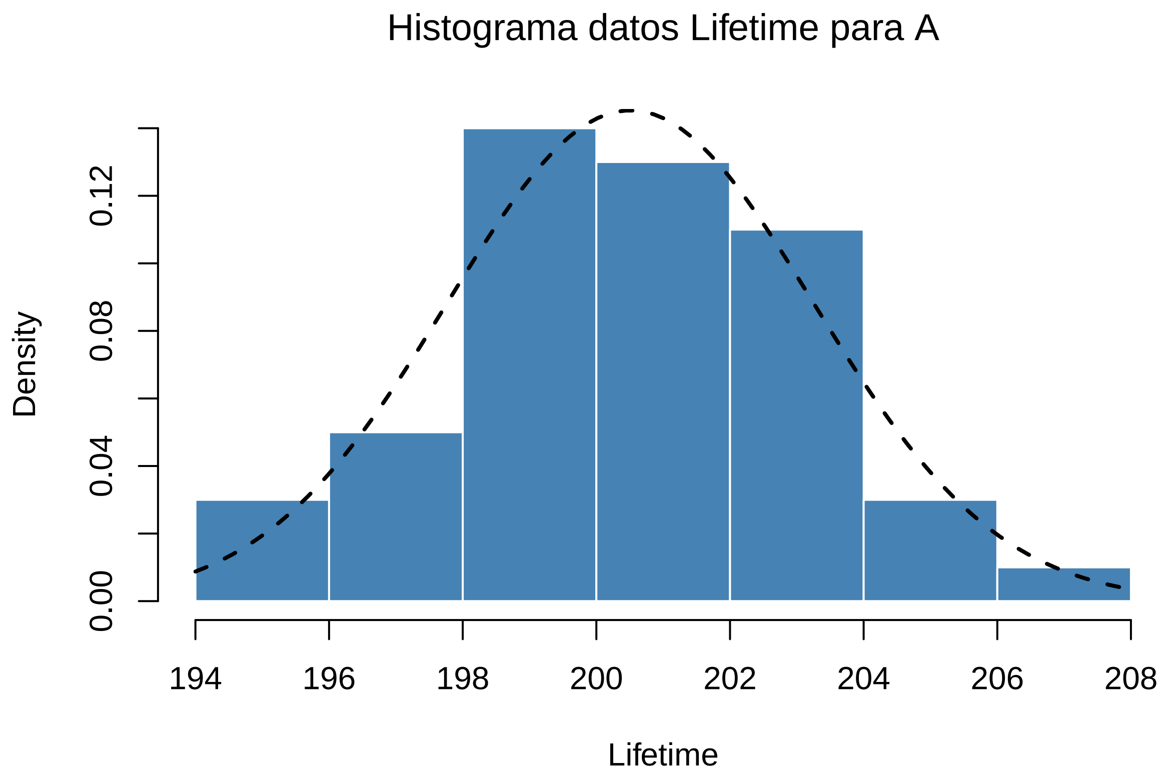
```
datosA <- datos %>% filter(facility == "A")
datosB <- datos %>% filter(facility == "B")
```

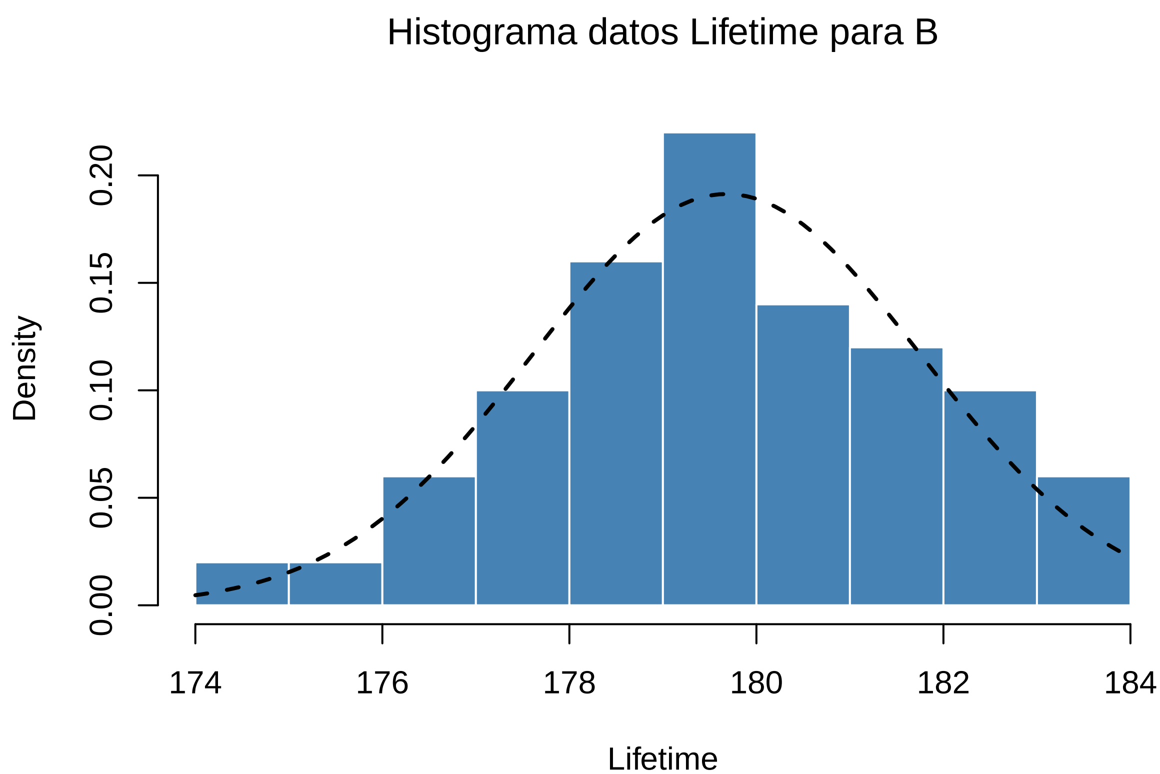
- **Realiza ahora un histograma de cada uno de los tipos y comenta si te parece que los datos siguen una distribución normal.**

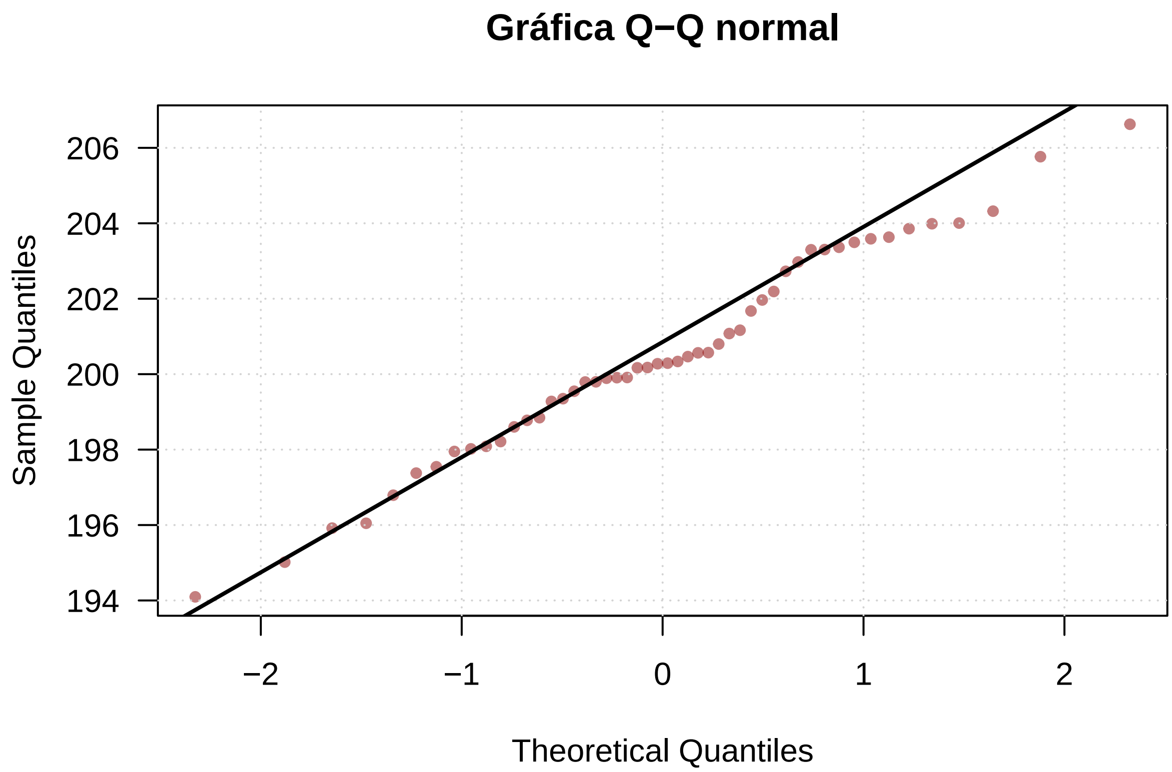


Se observan dos distribuciones que son compatibles con la distribución normal.

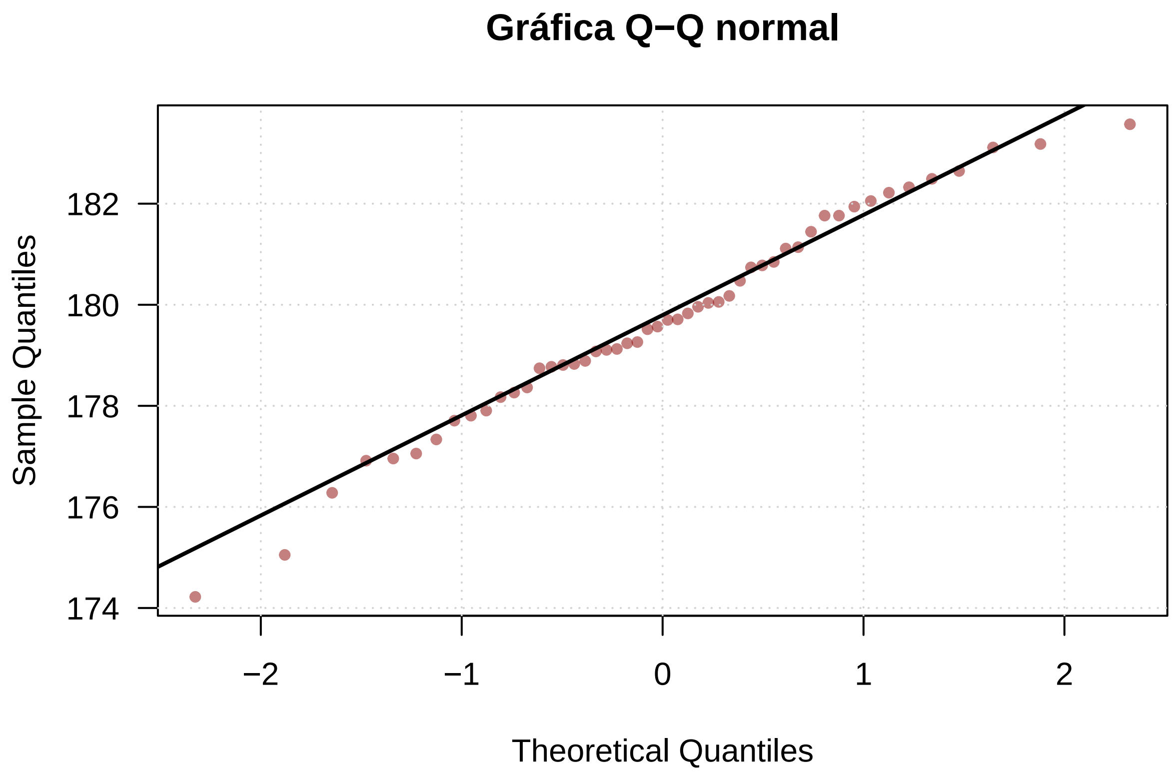
- **Confirma tus conclusiones con alguna/s de las herramientas vistas en clase (test de normalidad, gráfico Quantil-Quantil, tests de normalidad,...)**







Gráfica Q−Q normal
Sample Quantiles
194
196
198
200
202
204
206
−2
−1
0
1
2
Theoretical Quantiles



El método del Gráfico Q-Q también nos indica que ambas distribuciones siguen una distribución normal, dado que la distribución de la variable es muy similar a la distribución de comparación, siendo sobre todo en el centro una línea recta.

Comparando visulamente con una distribución normal que toma como inputs la media y desviación estándar de cada uno de los tipos de batería, parece que ambos tipos siguen una distribución normal.

```
##
##  Shapiro-Wilk normality test
##
## data:  datosA$lifetime
## W = 0.9848, p-value = 0.7632
```

```
##
##  Shapiro-Wilk normality test
##
## data:  datosB$lifetime
## W = 0.98395, p-value = 0.7256
```

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

```
##
```

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

Al evaluar a través de pruebas de normalidad, se encuentra que con el test de Shapiro-Wilk se obtiene que

$$W = 0.9848, p = 0.7632$$

en la muestra de baterías A y

$$W = 0.9835, p = 0.7256$$

en la muestra de baterías B. Por lo tanto, podemos concluir que para un nivel de significancia de 0.05 no hay evidencias estadisticas para rechazar la hipótesis de normalidad en ambas muestras. Se concluye lo mismo al aplicar el test de normalidad de Anderson-Darling, obteniendo en la muestra de baterías A

$$A = 0.35332, p = 0.4511$$

y

$$A = 0.19478, p = 0.8873$$

en la muestra de baterías B.

Actividad 2

Ahora que sabemos que nuestros datos siguen aproximadamente una distribución normal, tendríamos que estimar sus parámetros y . A partir de ahí, podemos realizar cálculo de probabilidades de la normal.

• Realiza una estimación puntual de la media y la desviación típica de la población de cada tipo de baterías.

$$\overline{X}_A = 200.5087; \sigma_A = 2.745777$$
$$\overline{X}_B = 179.6805; \sigma_B = 2.0849777$$

• Calcula la probabilidad de que una batería tomada al azar del tipo A dure más de 210 horas.

X: Batería de tipo A

$$X \sim N(200.5086628, 2.7457769)$$

```
A_mas_210 = pnorm(q = 210, mean = mean_lifetimeA, sd = sd_lifetimeA, lower.tail = FALSE)
```

$$P[X > 210] = 0.0002734129$$

• Calcula la probabilidad de que una batería tomada al azar del tipo B dure menos de 175 horas.

X: Batería de tipo B

$$X \sim N(179.6805239, 2.0849773)$$

```r
B_menos_175 = pnorm(q = 175, mean = mean_lifetimeB, sd = sd_lifetimeB, lower.tail = TRUE)
```

$$P[X < 175] = 0.01238792$$

- **Encuentra cuál es la duración máxima del 3% de las pilas del tipo B que duran menos (ayuda: esto es equivalente a encontrar el cuantil 0.03 de la distribución)**

```r
qnorm(p = .03,
      mean = mean_lifetimeB,
      sd = sd_lifetimeB,
      lower.tail = TRUE,
      log.p = FALSE)
```

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

$$p_{0.03} = 175.7591$$

Actividad 3

Vamos a centrarnos ahora en las baterías de tipo B. Supongamos que una duración por debajo de 175 horas no es aceptable para el usuario de la batería. En la actividad anterior hemos calculado la probabilidad p que esto suceda. Entonces, si tomamos una batería del tipo B al azar y comprobamos si dura menos de 175 de horas, estamos realizando un experimento de Bernoulli con probabilidad p.

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

X: Numero de baterias defectuosas en 1 lote de 10 baterias con probabilidad de ser defectuosas 0.0124

$$X \sim Bin(10, 0.0124)$$

```r
pbinom(q = 0, size = 10, prob = B_menos_175)
```

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

$$P[X = 0] = 0.8828033$$

- **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.)**

X:Numero de baterias no defectuosas hasta la primera bateria defectuosa con probabilidad de ser defectuosa 0.0124

$$X \sim Ge(P[X < 175])$$

```
dgeom(5, prob = B_menos_175)
```

```
## [1] 0.01163939
```

$$P(X = 5) = 0.01163939$$

- **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)**

X: Numero de elementos defectuosos obtenidos en un muestreo sin reemplazo de tamaño 5(n) de un conjunto con 20(N) baterias de las que 3(D) son defectuosas

$$X \sim HG(20; 3; 5)$$

```
1- dhyper(x = 0, m = 3, k = 5, n = 20-3)
```

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

$$P(X \geq 1) = 1 - P[X < 1] = 1 - P[X = 0] = 0.6008772$$

Actividad 4

Seguimos con las baterías de tipo B, pero en vez de hacer experimentos de Bernoulli queremos estudiar el número de baterías defectuosas fabricadas cada día. Supongamos que se fabrican 1000 baterías cada día. Entonces, cada día en promedio se estarán produciendo aproximadamente 1000 × p baterías, y el número de baterías defectuosas por día sigue una distribución de Poisson. Tomemos 12 como ese promedio de baterías defectuosas cada día. (ayuda: repasa qué modelo de distribución modeliza estos recuentos de eventos raros con una tasa media por unidad de tiempo)

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

X:Numero de baterias defectuosas por día

$$X \sim Poiss(12)$$

```r
ppois(q = 20, lambda = 12, lower.tail = FALSE)
```

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

$$P[X > 20] = 1 - P[X \leq 20] = 0.01159774$$

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

```r
ppois(q = 0, lambda = 12)
```

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

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

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

$$P[X = 0] = 6.144212 \cdot 10^{-6}$$

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

Sigue tambien una distribucion de Poisson. La distribucion Poisson tiene la propiedad aditiva o reproductiva, esto es, la suma de variables aleatorias independientes con distribucion Poisson de parametro λ_i tambien tiene una distribucion Poisson de parametro λ (suma de los diferente parametros $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_k$).

Fuente:https://www.cartagena99.com/recursos/alumnos/apuntes/2014%20-T1%20GradoADE_EstadisticaEmpresarialII_Tema1___Resumen_v1.pdf

X: numero de baterias defectuosas por dia

Y: numero de baterias defectuosas por semana

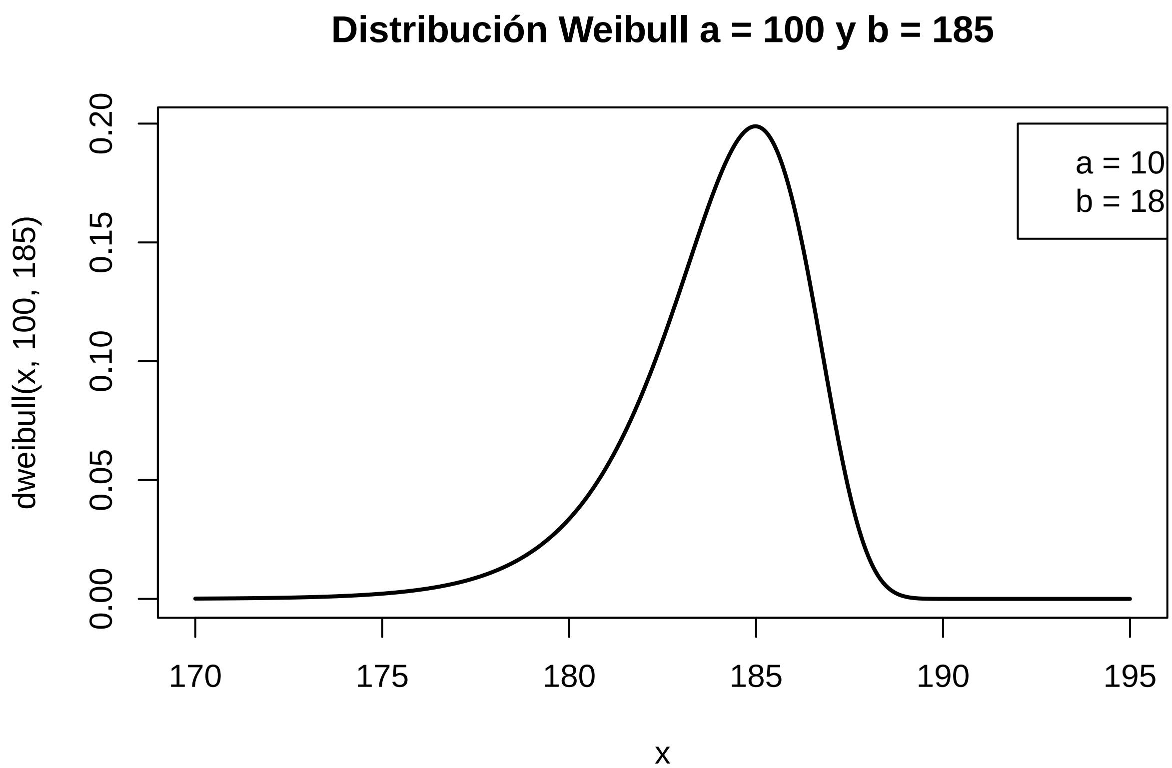
Aditiva: $Y = \sum_{j=1}^{m} X_j, X_j \sim Poiss(\lambda_j) \Longrightarrow Y \sim Poiss(\sum_{j=1}^{m} \lambda_j)$

Entonces, $m = 5$ por los días de la semana (lunes a viernes): $Y \sim Poiss(\sum_{j=1}^{5} \lambda_j)$

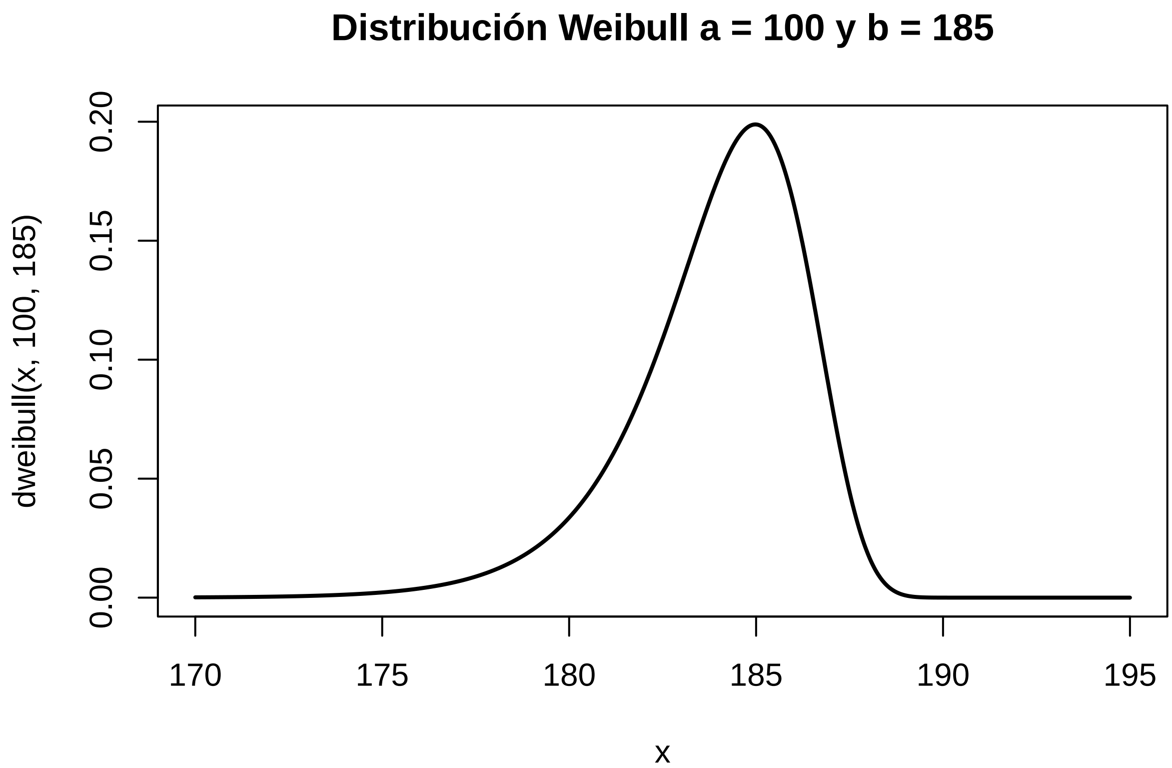
Actividad 5

El departamento de I+D de la empresa que fabrica las baterías tipo B está investigando nuevos materiales y métodos para mejorar la vida útil de las baterías. En particular, quieren llegar a diseñar una batería cuya duración siga una distribución de Weibull con parámetros a = 100 y b = 185.

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




```
weibull_vector = curve(dweibull(x, 100, 185),
                       from = 170,
                       to = 195,
                       n = 5000,
                       lwd = 2,
                       main = "Distribución Weibull a = 100 y b = 185")
```



- **Con este nuevo proceso, ¿se mejora realmente la duración media de las baterías?(ayuda: puedes usar los datos simulados o la expresión de la esperanza de una Weibull)**

$$\overline{X}_B = 179.6805$$

$$X \sim We(100, 185)$$

$$\overline{X}_{WeB} = 183.9177$$

```
esp_weibull = mean(rweibull(5000, shape = 100, scale = 185))
```

Efectivamente hay diferencia entre la media de los datos y la esperanza de la distribución Weibull, aunque no se detecta inicialmente unas diferenecias destacables.

- **Los ingenieros no lo tienen muy claro (parece que la diferencia no es tanta en promedio y los nuevos materiales son costosos). Para demostrarles que merece la pena, calcula la proporción de baterías defectuosas que producirá probablemente el nuevo proceso y compárala con el anterior (la p que calculamos en la actividad 2)**

$$P[X < 175] = 0.01238792$$

```r
B_menos_175_weibull = pweibull(175, 100, scale = 185)
B_menos_175 / B_menos_175_weibull
```

```
## [1] 3.215173
```

$$P[X < 175]_{We} = 0.003852956$$

$$Ratio = \frac{P[X < 175]}{P[X < 175]_{We}} = \frac{0.01238792}{0.003852956} = 3.215173$$

Relativo al nuevo proceso, la distribución Weibull muestra una probabilidad más de 3 veces inferior a la probabilidad de tener baterías defectuosas original.