Lectura 6

Poder estadístico

Poder, nivel de significación y tamaño de la muestra

->El poder corresponde a la probabilidad de no cometer un error de tipo II y que está muy relacionado al tamaño de la muestra

- El poder de la prueba aumenta mientras mayor es el tamaño del efecto (en este caso, la distancia entre el valor nulo y la media de la muestra).
- A medida que el tamaño del efecto disminuye (es decir, el estimador se acerca al valor nulo), el poder se aproxima al nivel de significación.
- Usar un valor de α más exigente (menor), manteniendo constante el tamaño de la muestra, hace que la curva de poder sea más baja para cualquier tamaño del efecto (lo que verifica la relación entre α y β).
- Usar una muestra más grande aumenta el poder de la prueba para cualquier tamaño del efecto distinto de 0.

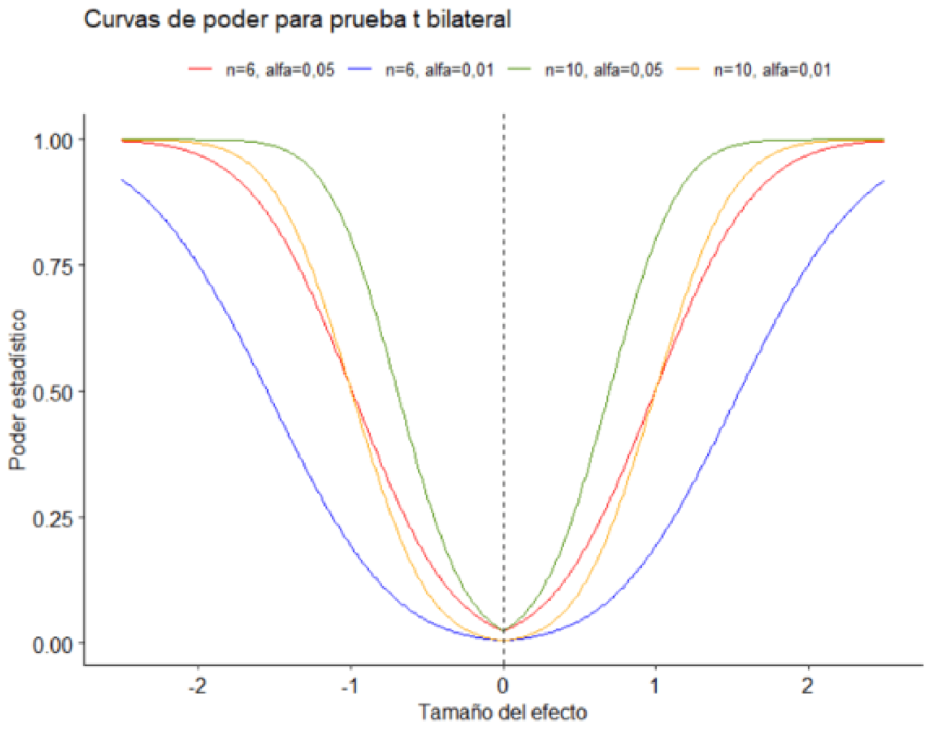

Figura 6.1: poder estadístico para prueba t bilateral.

Script 6.1: poder estadístico para prueba t bilateral.

```
library(ggpubr)
library(tidyverse)

# Generar un vector con un rango de valores para la efecto
# de medias.
efecto <- seq(-2.5, 2.5, 0.01)

# Calcular el poder para una prueba t bilareral, para cada tamaño
# del efecto, asumiendo una muestra con desviación estándar igual a 1.
# Se consideran 4 escenarios para calcular el poder:
# 1. Una muestra de tamaño 6 y nivel de significación 0.05.
```

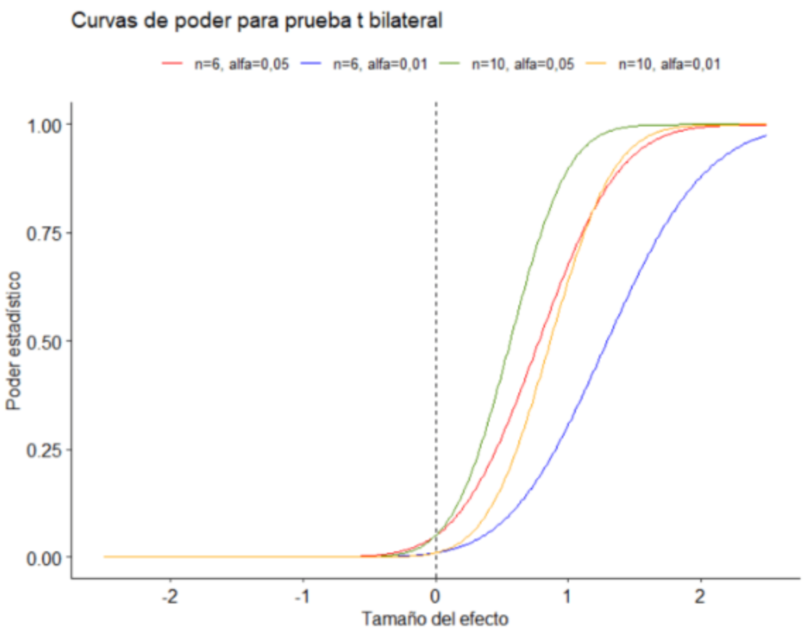

Figura 6.2: poder estadístico para prueba t unilateral.

```r
# 2. Una muestra de tamaño 6 y nivel de significación 0.01.
# 3. Una muestra de tamaño 10 y nivel de significación 0.05.
# 4. Una muestra de tamaño 10 y nivel de significación 0.01.
n_6_alfa_05 <- power.t.test(n = 6,
                      delta = efecto,
                      sd = 1,
                      sig.level = 0.05,
                      type = "one.sample",
                      alternative = "two.sided")$power

n_6_alfa_01 <- power.t.test(n = 6,
                      delta = efecto,
                      sd = 1,
                      sig.level = 0.01,
                      type = "one.sample",
                      alternative = "two.sided")$power

n_10_alfa_05 <- power.t.test(n = 10,
                       delta = efecto,
                       sd = 1,
                       sig.level = 0.05,
                       type = "one.sample",
                       alternative = "two.sided")$power

n_10_alfa_01 <- power.t.test(n = 10,
                             delta = efecto,
                             sd = 1,
                             sig.level = 0.01,
                             type = "one.sample",
                             alternative = "two.sided")$power
```

```r
# Construir matriz de datos en formato ancho.
datos <- data.frame(efecto, n_6_alfa_05, n_6_alfa_01,
                    n_10_alfa_05, n_10_alfa_01)

# Llevar a formato largo.
datos <- datos %>% pivot_longer(!"efecto",
                                names_to = "fuente",
                                values_to = "poder")

# Formatear fuente como variable categórica.
niveles <- c("n_6_alfa_05", "n_6_alfa_01", "n_10_alfa_05",
             "n_10_alfa_01")

etiquetas <- c("n=6, alfa=0,05", "n=6, alfa=0,01", "n=10, alfa=0,05",
               "n=10, alfa=0,01")

datos[["fuente"]] <- factor(datos[["fuente"]], levels = niveles,
                            labels = etiquetas)

# Graficar las curvas de poder.
g <- ggplot(datos, aes(efecto, poder,  colour = factor(fuente)))
g <- g + geom_line()
g <- g + labs(colour = "")
g <- g + ylab("Poder estadístico")
g <- g + xlab("Tamaño del efecto")

g <- g + scale_color_manual(values=c("red", "blue", "chartreuse4",
                                     "orange"))

g <- g + theme_pubr()
g <- g + ggtitle("Curvas de poder para prueba t bilateral")
g <- g + geom_vline(xintercept = 0, linetype = "dashed")

print(g)
```

Tamaño del efecto

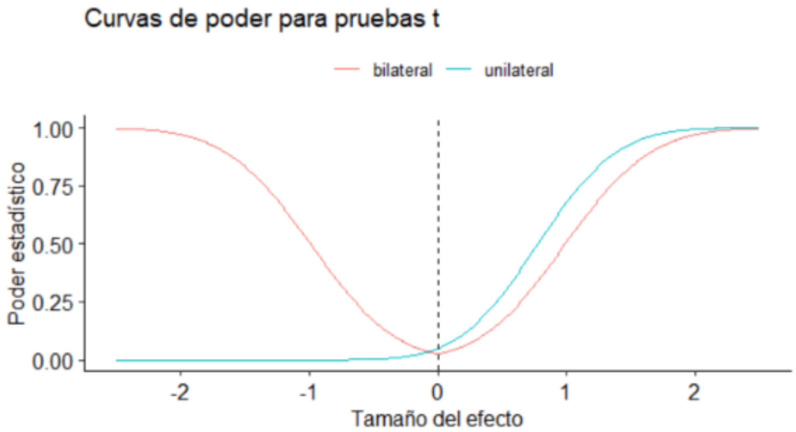

Figura 6.3: poder estadístico para pruebas t.

En términos generales, se considera que $d = 0,2$ es un efecto pequeño (imperceptible a simple vista), $d = 0,5$ es un efecto mediano (probablemente perceptible a simple vista) y $d = 0,8$, un efecto grande (definitivamente perceptible a simple vista).

En el caso de la prueba t de una muestra, la d de Cohen se calcula como muestra la ecuación 6.1, donde:

- $\overline{x}$: media muestral.
- μ_0: media teórica para el contraste (valor nulo).
- s: desviación estándar de la muestra con $n-1$ grados de libertad.

$$d = \frac{\overline{x} - \mu_0}{s} \tag{6.1}$$

Para la prueba t de diferencia de dos medias (también llamada prueba t para dos muestras independientes o, simplemente, prueba t independiente), si el tamaño de la muestra es mayor a 50 elementos, se calcula como muestra la ecuación 6.2, y para muestras pequeñas se aplica un factor de corrección, como indica la ecuación 6.3, donde:

- $\overline{x}_1$, $\overline{x}_2$: medias muestrales de cada grupo.
- n_1 y n_2 son los tamaños de ambas muestras.
- s_p: desviación estándar agrupada, dada por la ecuación 6.4[1].

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s_p} \tag{6.2}$$

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s_p} \cdot \frac{n_1 + n_2 - 3}{n_1 + n_2 - 2,25} \tag{6.3}$$

$$s_p = \sqrt{\frac{\sum(x - \overline{x}_1)^2 + \sum(x - \overline{x}_2)^2}{n_1 + n_2 - 2}} \tag{6.4}$$

En el caso de la variante de Welch para la prueba t independiente, la fórmula para el cálculo de la d de Cohen es ligeramente diferente, como puede apreciarse en la ecuación 6.5.

$$d = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}} \tag{6.5}$$

Por último, las ecuaciones 6.6 y 6.7 muestran cómo se calcula la d de Cohen en el caso de la prueba t con muestras pareadas grandes ($n > 50$) y pequeñas, respectivamente, donde D corresponde a las diferencias entre las observaciones pareadas.

$$d = \frac{\overline{x}_D}{s_D} \tag{6.6}$$

$$d = \frac{\overline{x}_D}{s_D} \cdot \frac{n_1 - 2}{n_1 - 1,25} \tag{6.7}$$

Poder, tamaño del efecto y tamaño de la muestra

-> **El poder** puede entenderse también como qué tan propensa es una prueba estadística para distinguir un efecto real de una simple casualidad y que podemos cuantificar este efecto.

-> Una gran **ventaja del poder estadístico** es que nos sirve para determinar el tamaño adecuado de la muestra para detectar un cierto tamaño del efecto.

Aumento

del poder estadístico a medida que el tamaño de la muestra aumenta (para un tamaño del efecto y nivel de significación fijos). En ella se aprecia que, a medida que el tamaño de la muestra crece, el poder estadístico también crece asintóticamente a 1, valor que equivale a tener la certeza de rechazar la hipótesis nula si esta es falsa.

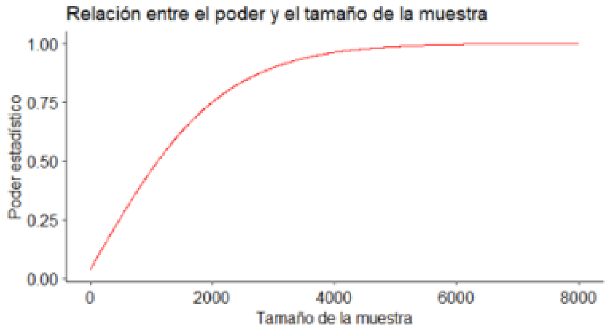

Figura 6.4: aumento del poder estadístico a medida que crece el tamaño de la muestra (manteniendo fijos el tamaño del efecto y el nivel de significación).

Script 6.2: aumento del poder estadístico a medida que crece el tamaño de la muestra.

```
library(ggpubr)

# Generar un vector con un rango para el tamaño de la muestra.
n <- seq(5, 8000, 5)

# Definir constantes
desv_est <- 6
alfa <- 0.05
tam_efecto <- 0.5

# Se calcula el poder con que se detecta el tamaño del efecto para
# cada tamaño de la muestra, asumiendo una prueba bilateral para
# una sola muestra.
poder <- power.t.test(n = n,
                      delta = tam_efecto,
                      sd = desv_est,
                      sig.level = alfa,
                      type = "two.sample",
                      alternative = "two.sided")$power

# Crear un data frame.
datos <- data.frame(n, poder)

# Graficar la curva de poder.
g <- ggplot(datos, aes(n, poder))
g <- g + geom_line(colour = "red")
g <- g + ylab("Poder estadístico")
g <- g + xlab("Tamaño de la muestra")
g <- g + theme_pubr()
g <- g + ggtitle("Relación entre el poder y el tamaño de la muestra")

print(g)
```

Cálculo teórico del poder

Lola Drones, estudiante de computación, ha diseñado dos nuevos algoritmos (A y B) que resuelven un mismo problema como parte de su trabajo de titulación. Lola desea saber si existe diferencia entre los tiempos de ejecución de ambos algoritmos. Para ello, ha decidido realizar una prueba t con muestras pareadas, con un nivel de significación $\alpha = 0,05$, usando para ello 36 instancias del problema de tamaño fijo que se ejecutan bajo iguales condiciones con cada algoritmo. Además, Lola ya sabe que la diferencia en el tiempo de ejecución sigue una distribución normal con desviación estándar $\sigma = 12$ milisegundos. Así, Lola ha formulado las siguientes hipótesis:

H_0: $\mu_{(A_i - B_i)} = 0$, es decir que la media de las diferencias en el tiempo de ejecución necesitado por los algoritmos A y B, para cada posible instancia i, es cero

H_A: $\mu_{(A_i - B_i)} \neq 0$

La figura 6.5 muestra cómo sería la distribución de la muestra (media de las diferencias en los tiempos de ejecución) si la hipótesis nula (H_0) fuese cierta, con las áreas correspondientes a la región de rechazo de H_0 coloreadas.

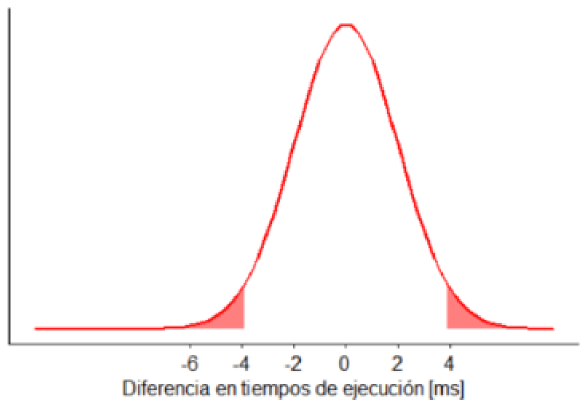

Figura 6.5: distribución de la diferencia de medias del tiempo de ejecución, señalando zonas de rechazo de la hipótesis nula.

Supongamos por un momento que, en realidad, el algoritmo B es en promedio 4 milisegundos más rápido que el algoritmo A. En este caso tendríamos que la media de las diferencias es de -4 [ms], correspondiente al tamaño del efecto. En este caso, su distribución sería como muestra la figura 6.6 (ver script 6.3) en color azul. Al superponer esta nueva curva a la que ya teníamos bajo el supuesto de que la hipótesis nula fuera verdadera, vemos que el área de la curva real que se situaría dentro de la región de rechazo de la curva teórica es aquella coloreada en azul. Esta área corresponde al poder de la prueba t, que en este caso es de $0,516$ de acuerdo al análisis teórico (ver script 6.1, líneas 77–86). Puesto que el poder corresponde a la probabilidad de **no** cometer un error de tipo II, de acuerdo al resultado obtenido se tiene que $\beta = 0,484$. ¡Lola no sería capaz de detectar una diferencia de -4 [ms] casi la mitad de las veces!

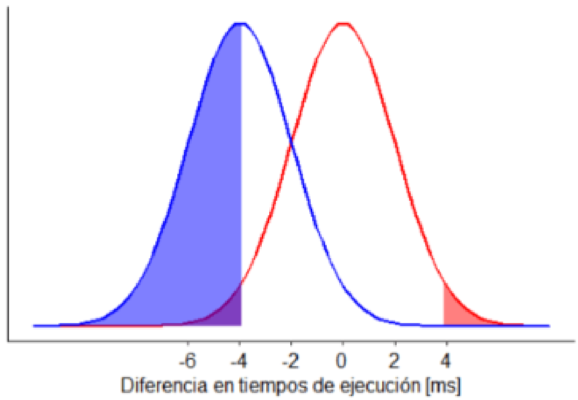

Figura 6.6: región de rechazo de la hipótesis nula en la distribución cuando el programa B es, en promedio, 4 milisegundos más rápido que el programa A.

Script 6.3: cálculo teórico del poder.

```
library(ggpubr)
library(pwr)

# Fijar valores conocidos.
sigma <- 12
alfa <- 0.05
n <- 36

# Calcular el error estándar.
SE <- sigma / sqrt(n)

# Gráficar la distribución muestral de la media de las diferencias si
```

```r
# la hipótesis nula fuera verdadera.
x <- seq(-6 * SE, 4 * SE, 0.01)
y <- dnorm(x, mean = media_nula, sd = SE)
g <- ggplot(data = data.frame(x, y), aes(x))

g <- g + stat_function(
  fun = dnorm,
  args = list(mean = media_nula, sd = SE),
  colour = "red", size = 1)

g <- g + ylab("")
g <- g + scale_y_continuous(breaks = NULL)
g <- g + scale_x_continuous(name = "Diferencia en tiempos de ejecución [ms]",
                            breaks = seq(-6, 4, 2))

g <- g + theme_pubr()

# Colorear la región de rechazo de la hipótesis nula.
media_nula <- 0
Z_critico <- qnorm(alfa/2, mean = media_nula, sd = SE, lower.tail = FALSE)
q_critico_inferior <- media_nula - Z_critico
q_critico_superior <- media_nula + Z_critico

g <- g + geom_area(data = subset(df, x < q_critico_inferior),
                   aes(y = y),
                   colour = "red",
                   fill = "red",
                   alpha = 0.5)

g <- g + geom_area(data = subset(df, x > q_critico_superior),
                   aes(y = y),
                   colour = "red",
                   fill = "red",
                   alpha = 0.5)

print(g)
```

```r
# Superponer la distribución muestral de la media de las diferencias
# si la la diferencia de medias fuera -4.
g <- g + stat_function(
  fun = dnorm,
  args = list(mean = media_efecto, sd = SE),
  colour = "blue", size = 1)

# Colorear la región de la nueva curva situada en la región de
# rechazo de la curva original.
x1 <- seq(-6 * SE, 4 * SE, 0.01)
y1 <- dnorm(x, mean = media_efecto, sd = SE)
g <- g + geom_area(data = subset(data.frame(x1, y1),
                                  x < q_critico_inferior),
                    aes(x = x1, y = y1),
                    colour = "blue",
                    fill = "blue",
                    alpha = 0.5)

g <- g + geom_area(data = subset(data.frame(x1, y1),
                                  x > q_critico_superior),
                    aes(x = x1, y = y1),
                    colour = "blue",
```

```r
                    fill = "blue",
                    alpha = 0.5)
print(g)

# Calcular el poder de acuerdo al análisis teórico.
poder <- pnorm(q_critico_inferior,
               mean = media_efecto,
               sd = SE,
               lower.tail = TRUE)
+ pnorm(q_critico_superior,
        mean = media_efecto,
        sd = SE,
        lower.tail =  FALSE)

cat("Poder = ", poder, "\n")

# Calcular la probabilidad de cometer un error tipo II.
beta <- 1 - poder_teorico
cat("Beta = ", beta, "\n")
```

agregar 6.5