



## Tema 12:

# Tamaño muestral

Jordi Cortés

José Antonio González, Héctor Rufino, Laura Riba y Erik Cobo

Enero 2014



## Tamaño muestral

### Índice

Presentación .....	3
1. Estimación de un parámetro.....	4
1.1. Estimación de una media .....	4
1.2. Estimación de una probabilidad .....	6
1.3. Garantías de observar un evento con baja probabilidad .....	8
2. Comparación de 2 medias .....	9
2.1. Datos independientes .....	10
2.2. Datos apareados .....	16
2.3. Análisis del cambio .....	18
2.4. Estudios de equivalencia y no-inferioridad.....	19
2.5. Precisión de la estimación del efecto .....	22
2.6. Tamaño desigual en ambos grupos * .....	23
3. Comparación de probabilidades.....	26
4. Comparación del tiempo hasta un evento .....	27
4.1. Participantes reclutados en un mismo instante .....	27
4.2. Participantes reclutados a lo largo del tiempo .....	30
5. Consejos prácticos.....	32
Soluciones a los ejercicios .....	34

\* Indica tema más avanzado que conviene mirar pero no es crucial para los ejercicios.

---

## Presentación

El tamaño importa: sabemos ya calcular la precisión de un estimador en unos datos aleatorios. En este tema, estudiamos el tamaño necesario para una precisión determinada. Y para un estudio pivote, de decisión, cómo garantizar la potencia deseada.

Por progresividad en la exposición, la primera parte no aplica a los ensayos clínicos sino a estudios con un solo grupo.

**Contribuciones:** Basado en transparencias de José Antonio González, Jordi Cortés preparó una primera versión que ha sido revisada por Héctor Rufino, Laura Riba y Erik Cobo.

# 1. Estimación de un parámetro

Cuando el objetivo es estimar un parámetro poblacional (p.ej, una media), la (im)precisión puede definirse como la amplitud (A) de dicho intervalo, la semi-amplitud ( $e = A/2$ ) del intervalo o el error estándar del estimador (EE). En este apartado se mostrará cómo obtener el tamaño muestral para la estimación de una media o de una probabilidad a partir de  $e$  por coherencia con la medida de (im)precisión usada por R.

## 1.1. Estimación de una media

Para el cálculo del tamaño muestral es necesario fijar de antemano unos parámetros que dependen del tipo de estudio. En el caso de la estimación de una media, estos parámetros son la desviación típica de la respuesta ( $\sigma$ ), la confianza ( $1-\alpha$ ) y la semi-amplitud ( $e$ ) del intervalo.

### Fórmula

El tamaño muestral para la **estimación de una media** es:

$$n = \left( \frac{z_{\alpha/2} \cdot \sigma}{e} \right)^2$$

donde  $z_{\alpha/2}$  es el cuantil de la distribución Normal estándar correspondiente ( $z_{\alpha/2} = 1.96$  en el caso de un IC<sub>95%</sub>).

En esta fórmula, la confianza  $1-\alpha$  y la semi-amplitud  $e$  del intervalo son dos parámetros que escoge el investigador, si bien es tradicional fijar  $1-\alpha=0.95=95\%$ . Por contra, la desviación típica  $\sigma$  de la variable de interés viene dada por el fenómeno en estudio y su valor debe provenir de la literatura existente o de la experiencia profesional de cada uno. La ambición del investigador determina el nivel de confianza y la precisión, pero la desviación típica es una premisa que depende del conocimiento previo.

**Nota técnica:** Se vio que la amplitud del intervalo de confianza depende del error típico del estimador y del nivel de confianza. La forma de limitar el grado de incertidumbre o amplitud de este intervalo sin disminuir el nivel de confianza  $1-\alpha$  es disminuir el error típico de la estimación del parámetro a través de modificar el tamaño muestral. En el caso de la estimación de la media poblacional o esperanza matemática, el intervalo de confianza, asumiendo  $\sigma$  conocida, es:

$$IC_{1-\alpha}(\mu) = \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Por tanto, para una semi-amplitud del intervalo de confianza que valga  $e$ , se tiene:

$$e = \frac{\text{Lím. superior IC} - \text{Lím. inferior IC}}{2} = \frac{\left(\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) - \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)}{2} = \frac{2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}{2} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Aislando la  $n$  en esta expresión se obtiene la fórmula anterior.

**Ejemplo 1.1:** Para estudiar la imagen de los diferentes políticos, se pide a los encuestados que los evalúen en una escala (continua) de 0 a 10 puntos. Si se acepta que la desviación típica de esta variable es de 1.5 puntos, ¿cuántos casos se necesitan para que la semi-amplitud del intervalo de confianza ( $e$ ) al 95% de la media poblacional sea de 0.05 puntos?

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{e}\right)^2 = \left(\frac{1.96 \cdot 1.5}{0.05}\right)^2 \approx 3457.44 \rightarrow 3458 \text{ casos}$$

**Nota técnica:** En los cálculos de tamaño muestral el número obtenido siempre se redondea al alza para alcanzar el objetivo especificado.



**Ejercicio 1.1**

Dado que 3458 casos son demasiados, se rebaja la ambición de conocimiento desde una semi-amplitud de 0.05 puntos hasta 0.25 puntos. ¿Cuál es ahora el tamaño necesario?

En R, existen diversos paquetes para el cálculo del tamaño muestral. El paquete *samplingbook* contiene funciones aplicables a la estimación de un parámetro; por ejemplo la función *sample.size.mean* realiza el cálculo para la estimación de una media. [Recuerde que lo primero es realizar la instalación y la carga del paquete].



**Ejemplo de R**

```
# Instalación y carga de 'samplingbook'
> install.packages('samplingbook')
> library(samplingbook)

# Aplicación al Ejemplo 1.1. (e es la semi-amplitud, S es
# la sigma (σ) y level es la confianza)
> sample.size.mean(e=0.05, S=1.5, level = 0.95)
Sample size needed: 3458
```



### Ejercicio 1.2

En un estudio se pretende estimar la PAS de un grupo de pacientes con una determinada patología. Se asume una  $\sigma = 10$  mmHg. Si se desea amplitud (A) de 5 mmHg para el intervalo de confianza del 95%, ¿Cuál es el tamaño necesario? Haga los cálculos con R.

### Ejercicio 1.3

Después de realizar el estudio del ejercicio anterior con el número de pacientes calculado, resulta que el  $IC_{95\%}$  es más ancho que los 5 mmHg pretendidos en el diseño del estudio. ¿Qué ha ocurrido?

## 1.2. Estimación de una probabilidad

En el caso de una respuesta dicotómica, únicamente se deben especificar 2 parámetros: la confianza y la semi-amplitud del intervalo.

### Fórmula

El tamaño muestral para la **estimación de una probabilidad** es:

$$n = \left( \frac{z_{\alpha/2}}{2 \cdot e} \right)^2$$

En el caso particular de un  $IC_{95\%}$ , la fórmula será  $n \approx 1/e^2$  (ya que  $1.96 \approx 2$ )

**Nota:** La amplitud del intervalo debe expresarse en tanto por uno (no en porcentaje).

**Nota técnica:** El intervalo de confianza, en la situación de máxima incertidumbre, es:

$$IC_{95\%}(\pi) = P \pm z_{\alpha/2} \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}}$$

Por tanto, la amplitud del intervalo A vale:

$$A = 2 \cdot z_{\alpha/2} \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}}$$

Aislando la  $n$  se llega a la fórmula descrita.

**Ejemplo 1.2:** Para conocer el porcentaje de votos de un partido político, con una semi-amplitud total del intervalo de confianza al 95% igual a 0.5% (Amplitud del 1%) ¿Cuántos casos se necesitan?

$$n = \left(\frac{z_{\alpha/2}}{2e}\right)^2 = \left(\frac{1.96}{2 \cdot 0.005}\right)^2 \approx 38416 \text{ casos}$$



#### Ejercicio 1.4

Decididamente 38416 son demasiados casos, por lo que una vez más se rebaja la ambición de conocimiento desde una semi-amplitud de 0.5 puntos (0.5%=0.005) hasta 2.5 puntos (2.5%=0.025). Calcule el nuevo tamaño muestral.

En R, la instrucción *sample.size.prop* del paquete *samplingbook* realiza el cálculo del tamaño necesario para el intervalo de confianza de una probabilidad. [Instálese y cárguese el paquete si no se hizo previamente]



#### Ejemplo de R

```
# Aplicación al Ejemplo 1.2 (e es la semi-amplitud del IC y
# level es la confianza)
> sample.size.prop(e=0.005, level = 0.95)
Sample size needed: 38415
```

**Nota:** El cálculo “a mano” mediante la fórmula da un resultado de 38416 casos, mientras que el resultado con la función de R es de 38415 casos. El motivo es el redondeo del valor de  $z_{\alpha/2}$  que realizamos cuando resolvemos el cálculo a mano. Usamos 1.96 mientras que R usa el valor exacto (1.959964). Siempre que sea posible es recomendable usar el software.



#### Ejercicio 1.5

Se desea conocer la prevalencia de diabetes en la población mayor de 18 años con una incertidumbre del 2% ( $e = 0.02$ ) y una confianza del 90%. ¿Cuál es el tamaño necesario? Haga los cálculos con R.

**Nota:** Por simplicidad, es habitual considerar infinita a la población objetivo, ya que hacerlo así garantiza los riesgos estadísticos —que en realidad serían ligeramente menores: p.e., la [corrección](#) necesaria para una muestra muy grande de 10000 casos, de una población pequeña de 100000 indica que la amplitud real del IC se reduciría en menos de 0.05.

### 1.3. Garantías de observar un evento con baja probabilidad

Un objetivo secundario de un ensayo clínico es tener una probabilidad razonable de observar eventos no deseados.

**Lectura:** No, mejor, escucha: [la vida te da sorpresas, sorpresas te da la vida](#), o [Life is what happens to you while you're busy making other plans](#).

El reto es que los efectos no deseados se pueden manifestar en infinidad de formas. Además, algunos de estos efectos, posiblemente los más graves, pueden presentarse con muy baja frecuencia, por lo que necesitarían tamaños muestrales inmensos para tener una probabilidad razonable de observarlos.

Podemos recurrir a las distribuciones de probabilidad estudiadas en el capítulo 5 para estudiar, para una muestra de  $n$  casos, la probabilidad de observar eventos raros.

**Ejemplo 1.3:** Si suponemos que cierta reacción adversa grave se presenta en 1 de cada 200 pacientes, la probabilidad de observar por lo menos un caso en una muestra de 100 pacientes vale 8.98% [con R, `pbinom (q=1, size=100, prob=1/200, lower.tail = FALSE)`].

**Ejemplo 1.4:** Si deseamos tener una probabilidad del 90% de observar por lo menos un caso con un evento no deseado que tenga una frecuencia tan baja como 1 cada mil, necesitamos observar 2303 casos. El cálculo de este valor no es directo y requiere de un tanteo usando la distribución de Poisson.

$$X \sim \text{Poisson}(n \cdot \lambda)$$

[ $n$  es el número de pacientes necesarios y  $\lambda$  es la tasa de aparición del evento ( $\lambda=1/1000$ )]

Lo que se desea es que  $P(X \geq 1) \geq 0.9$ . Con R, se observa que la primera  $n$  que cumple este propósito es 2303:

```
> n = 2303
> ppois(q=0, lambda=n*1/1000, lower.tail = FALSE)
[1] 0.9000415
```

Nótese que se especifica  $q=0$  (y no  $q=1$ ) porque en el cálculo de colas superiores se calcula la  $P(X > q)$ , que es equivalente a la  $P(X \geq q+1)$ . Pruebe con  $n$ 's más bajas para verificar que no se llega a la probabilidad deseada.



Aunque el estudio de estos efectos se beneficia del análisis globalizador de los diferentes ensayos disponibles, en la práctica conviene una buena recogida de información observacional una vez autorizada la intervención.

**Lectura:** La [ICH-E1A](#) da recomendaciones sobre el número de casos necesarios para estudiar la seguridad de tratamientos crónicos. Pretende detectar aquellos fenómenos cuya incidencia supera el 1% a los 3 meses, pero no pretende ‘caracterizar’ acontecimientos adversos por debajo del 1 por mil. Sugiere entre 300 y 600 casos seguidos y tratados durante seis meses y 100 casos durante un año. También alerta sobre el necesario rigor científico de la comparación con los no tratados.



### Ejercicio 1.6

Un fenómeno tiene una incidencia de un 1% a los 3 meses. Calcule la probabilidad de observar al menos un evento en los siguientes casos usando la distribución de Poisson:

- a) Con 300 casos seguidos durante 3 meses
- b) Con 600 casos seguidos durante 3 meses
- c) Con 300 casos seguidos durante 6 meses
- d) Con 600 casos seguidos durante 6 meses
- e) Con 100 casos seguidos durante 1 año

Haga los cálculos usando la función *ppois* con R.

## 2. Comparación de 2 medias

En respuesta continua, el análisis más común consiste en la comparación de la media entre dos grupos de tratamiento.

Un Ensayo Clínico Aleatorizado (ECA) pivote implica una decisión posterior. El entorno de decisión de Neyman-Pearson permite limitar ambos riesgos de error.

**Ejemplo 2.1:** Una agencia de regulación que autoriza productos sanitarios desea un mecanismo de decisión que le garantice que: 1) sólo un 2.5% de las intervenciones no eficaces alcanzan el mercado; pero 2) que sí lo hagan un 90% de las que tienen cierto efecto positivo (al que llamarán Delta:  $\Delta$ ).

## 2.1. Datos independientes

La metodología de decisión de Neyman-Pearson permite considerar simultáneamente los riesgos  $\alpha$ ,  $\beta$  y la magnitud  $\Delta$  que se desea establecer.

En la comparación de dos medias, se está interesado en tomar una decisión entre dos valores concretos, por ejemplo 0 y  $\Delta$ , que se sitúan en las hipótesis nula y alternativa:

$$\begin{cases} H_0: \mu_A - \mu_B = 0 \\ H_1: \mu_A - \mu_B = \Delta \end{cases}$$

**Ejemplo 2.2:** Puede imaginarse, por ejemplo, que cierto tratamiento A tenga interés sanitario y comercial si, respecto a la versión clásica B,  $\Delta$  representa aquella diferencia que hace rentable el desarrollo y la sustitución de B por A. Como es habitual en la prueba de diferencias, el valor de la hipótesis nula indica la absoluta igualdad entre ambos.

Supóngase que se conoce el grado de dispersión ( $\sigma$ ) existente entre los resultados en varios pacientes sometidos al mismo tratamiento. Supóngase también que se ha decidido que los riesgos de adoptar decisiones erróneas sean exactamente  $\alpha$  (bilateral) y  $\beta$  (unilateral). Para determinar el número “n” de pacientes necesario en cada grupo, por simplicidad, se considera la situación (de máxima eficiencia) en la que se dispone de exactamente el mismo número de casos en ambos grupos:  $n_A = n_B = n$ .

### Fórmula

El tamaño muestral en cada grupo para la **comparación de 2 medias** es:

$$n = \frac{2 \cdot \sigma^2 \cdot (z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$

**Nota técnica:** en esta situación, dado que la varianza de la diferencia de las medias en muestras independientes (asumiendo iguales las  $n$  y las  $\sigma$  bajo cada tratamiento) es:

$$V(\bar{y}_1 - \bar{y}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \frac{2\sigma^2}{n}$$

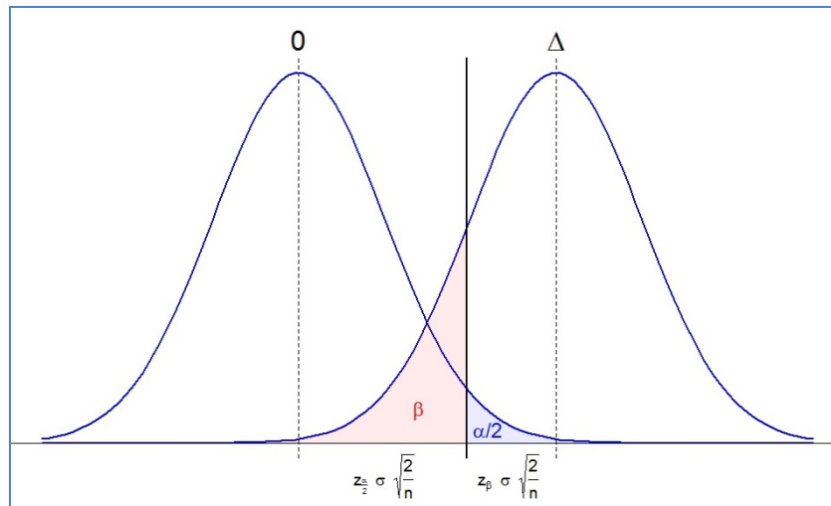
Si la distribución de  $y$  es Normal o el número de casos es razonablemente grande, la distribución de esta diferencia de medias será Normal con esta varianza y centrada en 0, bajo  $H_0$ , o en  $\Delta$ , bajo  $H_1$ .

En la Figura 2.1 puede verse que la distancia entre los centros de ambas distribuciones es:

$$\Delta = Z_{\alpha/2} \cdot \sigma \sqrt{\frac{2}{n}} + Z_{\beta} \cdot \sigma \cdot \sqrt{\frac{2}{n}}$$

Si ahora, se especifican los riesgos  $\alpha$  (probabilidad de actuar acorde a  $H_1:\Delta$  siendo cierta  $H_0:0$ ) y  $\beta$  (probabilidad de actuar acorde a  $H_0:0$  siendo cierta  $H_1:\Delta$ ) que se está dispuesto a tolerar, ya se puede conocer el tamaño muestral  $n$  en cada muestra:

$$n = \frac{2 \cdot \sigma^2 \cdot (Z_{\alpha/2} + Z_{\beta})^2}{\Delta^2}$$



**Figura 2.1** El tamaño muestral  $n$  requerido es aquel que permite que el solapamiento de las distribuciones del estadístico bajo  $H_0$  y bajo  $H_1$  proporcione los valores  $\alpha$  y  $\beta$  especificados. Nótese que, si  $n$  aumenta, las distribuciones se hacen más “puntiagudas”, disminuye el solapamiento y, por tanto, los riesgos  $\alpha$  y  $\beta$ . La línea continua vertical marca el valor a partir del cual se tomará una u otra decisión.



**Recuerde**

El cálculo del tamaño muestral depende de:

- los riesgos  $\alpha$  y  $\beta$  que esté dispuesto a aceptar: cuanto menores, mayor tamaño.
- la dispersión  $\sigma$  del fenómeno estudiado: cuanto mayor, mayor tamaño.
- la magnitud  $\Delta$  de la diferencia que se desea demostrar: cuanto menor, mayor tamaño.



**Ejercicio 2.1**

¿Qué significa  $\Delta$  ? ¿La diferencia “ideal” que se quiere demostrar? ¿La diferencia “real” que se cree que se puede demostrar?

**Ejercicio 2.2**

En un ensayo clínico de cuyo éxito depende la autorización comercial de un fármaco, ¿qué consecuencias se derivan de los riesgos  $\alpha$  y  $\beta$  de cometer errores de 1ª y 2ª especie? ¿Qué implicaciones tienen para el usuario y el patrocinador?

**Nota técnica:** Tradicionalmente se define un riesgo  $\alpha$  bilateral, pero la decisión de autorizar la intervención sólo se toma en el lado superior, por lo que, por coherencia, algunos metodólogos definen un  $\alpha = 0.025$  unilateral en lugar de un  $\alpha = 0.05$  bilateral. Mire [este número](#) de J. Biopharm. Stat. donde se le dedican varios artículos a este asunto (p. 133-170).



### Recuerde

La **potencia** de un estudio para establecer una alternativa de interés es el complementario del riesgo  $\beta$ .

La **potencia** es la probabilidad de hallar diferencias entre dos tratamientos que realmente son diferentes.

**Ejemplo 2.3:** un riesgo  $\beta = 0.20$  implica una potencia de  $0.8 = 80\%$ .

**Ejemplo 2.4:** ¿Qué tamaño muestral sería necesario para detectar una diferencia en la altura media de hombres y mujeres de 10 cm? Sea  $\sigma = 8$  cm y los riesgos habituales ( $\alpha = 0.05$  ;  $\beta = 0.20$ ).

$$n = \frac{2 \cdot 8^2 \cdot (1.96 + 0.84)^2}{10^2} = 10.04$$

Se necesitan 11 casos por grupo (aunque con 10, prácticamente se alcanzaría el objetivo).

Note que también se puede hacer la pregunta a la inversa, es decir, preguntar por la diferencia que se podría detectar dado un determinado tamaño muestral.

**Ejemplo 2.4 (cont.):** ¿Qué diferencia en la altura media de hombres y mujeres puede detectar suponiendo que puede reclutar 40 pacientes en total?

Aislando  $\Delta$  de la fórmula anterior se obtiene:

$$n = \frac{2 \cdot \sigma^2 \cdot (z_{\alpha/2} + z_{\beta})^2}{\Delta^2} \quad \rightarrow \quad \Delta = \sqrt{\frac{2 \cdot \sigma^2 \cdot (z_{\alpha/2} + z_{\beta})^2}{n}}$$

Por lo tanto, suponiendo que puede reclutar 20 pacientes por grupo (40 en total):

$$\Delta = \sqrt{\frac{2 \cdot 8^2 \cdot (1.96 + 0.84)^2}{20}} = 7.084$$

Con 20 pacientes por grupo se puede detectar una diferencia de 7.084 cm en la altura media de hombres y mujeres.

Nota: El valor exacto de  $z_{\beta} = z_{0.8}$  es 0.8416212, por lo que al usar el valor redondeado tanto de  $z_{\beta}$  como de  $z_{\alpha/2}$  los resultados obtenidos pueden variar un poco de los cálculos exactos. Por ejemplo, el valor de  $\Delta$  usando todos los decimales es 7.088.

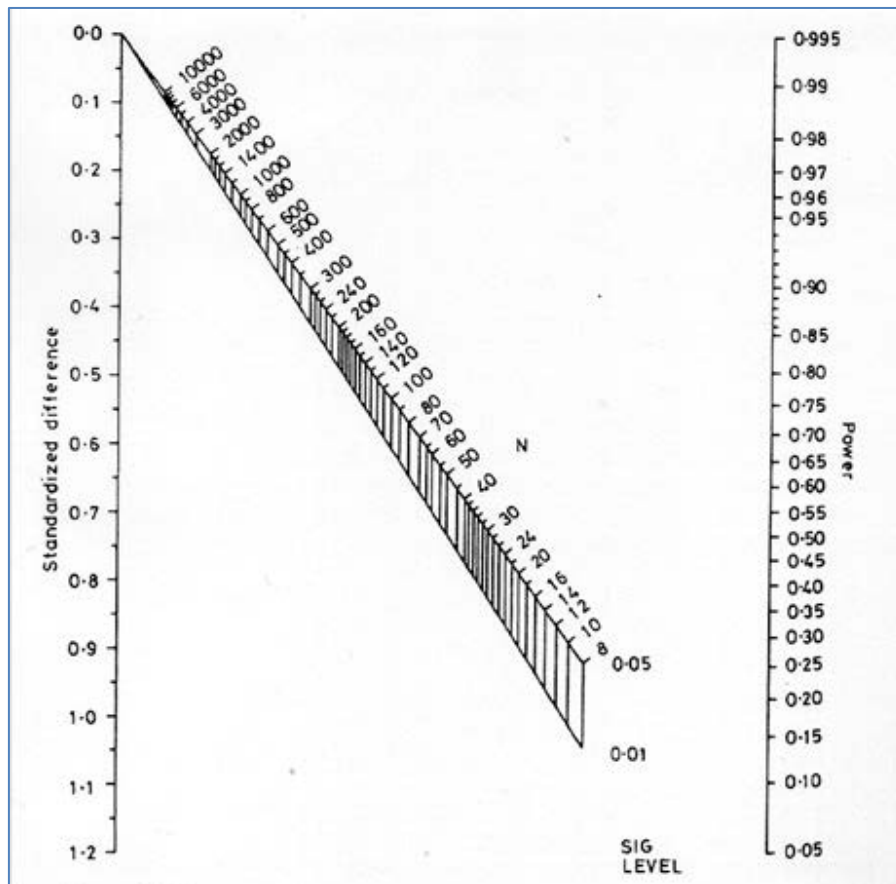
Se puede definir la diferencia tipificada o estandarizada  $\Delta_s$  como la razón entre la diferencia que se quiere detectar y la desviación típica (diferencia esperada entre dos observaciones):

$$\Delta_s = \frac{\Delta}{\sigma}$$

Así, esta diferencia tipificada representa el efecto relativo a la dispersión natural de los casos.

**Ejemplo 2.5:** Si se deseara aumentar la altura en 4 centímetros y la desviación típica se ha dicho que son 8 cm, el efecto tipificado sería del 50%.

Para un cálculo orientativo preliminar, se puede usar el gráfico de Douglas Altman (Figura 2.2), donde N representa el tamaño total considerando ambos grupos ( $N = 2 \cdot n$ ). Una la diferencia estandarizada y la potencia deseadas de los ejes verticales izquierdo y derecho con una línea. El punto de corte de dicha línea con la línea de  $\alpha=0.05$  (o  $\alpha=0.01$ ), le indicará el tamaño requerido.



**Figura 2.2** Nomograma para el cálculo del tamaño muestral y potencia (Altman, 1982)



### Ejercicio 2.3

¿Cuántos casos se necesitan si  $\Delta = 5u$ ,  $\sigma = 8u$ ,  $\alpha = 0.05$  bilateral y  $\beta = 0.20$ ? Haga el cálculo con la fórmula y usando el nomograma.

### Ejercicio 2.4

Calcule el tamaño necesario para un caso real propio.

El paquete *TrialSize* de R contiene funciones de cálculo del tamaño muestral en ensayos clínicos. Para la comparación de dos medias independientes se usa la instrucción *TwoSampleMean.Equality*



### Ejemplo de R

```
# Instalación y carga de 'TrialSize'
> install.packages('TrialSize')
> library(TrialSize)

# Aplicación al Ejemplo 2.4 (k es el cociente entre los
# tamaños de ambos grupos. k=1 implica grupos iguales)
> n <- TwoSampleMean.Equality(alpha=0.05, beta=0.20, sigma=8,
margin=10, k=1)
> n
[1] 10.04657
```

### Ejercicio 2.5

Sea  $\sigma = 10u$ ,  $\alpha = 0.05$  bilateral y  $\Delta = 5u$ . Calcule los casos necesarios por grupo para potencias de 10%, 20%...90%. Dibuje un gráfico con R en el que se relacione la potencia con el tamaño muestral.



### Recuerde

Mayor tamaño muestral implica mayor potencia.

La instrucción *power.t.test* realiza el proceso inverso, dada una  $n$ , calcula la potencia en el caso de una comparación de medias. [De hecho, puede calcular cualquier parámetro especificando el resto. Véase la ayuda con *?power.t.test*]



### Ejemplo de R

```
# Potencia con n=15, Δ=15, σ=15 y α=0.05
> power.t.test(n=15, delta=15, sd=15, sig.level=0.05)
Two-sample t test power calculation
```

```
n = 15
delta = 15
sd = 15
sig.level = 0.05
power = 0.752921
alternative = two.sided
NOTE: n is number in *each* group
```

**Nota:** Esta función retorna una lista con varios objetos. Añada *\$pow* al final de la instrucción para obtener únicamente la potencia.



### Ejercicio 2.6

Sea  $\sigma = 10u$ ,  $\alpha = 0.05$  bilateral y  $N = 100$ . Dibuje con R un gráfico en el que se relacione la potencia con el efecto tipificado para  $\Delta = 1, 2, \dots, 9$ .



### Recuerde

Mayor efecto  $\Delta$  en estudio implica mayor potencia.



### Ejercicio 2.7

Suponga que los investigadores han sido algo conservadores y han especificado un tamaño del efecto  $\Delta$  algo menor que el real. Discuta cómo afecta este hecho a la potencia del estudio.



### Recuerde

Si el efecto real fuera mayor que el  $\Delta$  empleado en la fórmula, la potencia real sería mayor.



### Ejercicio 2.8

Sea  $\sigma=10u$ ,  $\alpha=0.05$  bilateral y  $\beta=0.80$ . Halle los casos que necesita para  $\Delta$  desde 1 hasta 9. Dibuje un gráfico con R en el que se relacione los casos necesarios con el efecto tipificado.

### Recuerde

Mayor efecto  $\Delta$  en estudio implica menor tamaño muestral.

## 2.2. Datos apareados

Los diseños con datos apareados consideran diferentes fuentes de variación. Recuerde que, como cada unidad proporciona información sobre la diferencia del efecto de ambos tratamientos en comparación, se definía una nueva variable (D), diferencia entre la respuesta observada en ambas alternativas:

$$D_i = Y_{iA} - Y_{iB}$$

¿Qué tiene que ver la varianza de esta nueva variable,  $\sigma_D^2$ , con la varianza  $\sigma^2$  que se utiliza en los datos independientes? Un sencillo modelo descompone  $\sigma^2$  (la varianza total) en dos componentes: **entre**-individuos  $\sigma_E^2$  o verdaderas diferencias entre los casos e **intra**-individuos  $\sigma_I^2$  o discordancias entre dos medidas del mismo individuo:

$$\sigma^2 = \sigma_E^2 + \sigma_I^2$$

En datos apareados, se puede utilizar la misma fórmula que en los datos independientes pero teniendo en cuenta que el error se refiere a la variabilidad **intra**-sujeto  $\sigma_I^2$  y la  $n$  resultante de la fórmula es la  $N$  total, ya que cada observación aporta los dos valores.

### Fórmula

El número de **observaciones necesarias** para la **comparación de 2 medias apareadas** es:

$$N = \frac{2 \cdot \sigma_I^2 \cdot (z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$



### Recuerde

Misma fórmula pero: la varianza ahora es intra-sujetos y la 'N' resultante es la total.

**Nota:** veremos que, si una tercera variable define el apareamiento (por ejemplo, el orden o lugar de administración), conviene dividir la  $N$  total en 2 subgrupos, uno para cada orden de la tercera variable (por ejemplo, administrar en orden AB y BA, o administrar alternado los lugares).



### Ejercicio 2.9

¿Cuántos casos totales se necesitan en un diseño paralelo y en uno apareado si  $\Delta = 5u$ ,  $\sigma_E^2 = (9u)^2$ ,  $\sigma_I^2 = (4u)^2$ ,  $\alpha = 0.05$  bilateral y  $\beta = 0.20$ ? Haga los cálculos con las fórmulas.



**Ejercicio 2.10**

¿Cuántos casos totales se necesitan en un diseño paralelo y en uno apareado si  $\Delta = 5u$ ,  $\sigma_E^2 = 50 u^2$ ,  $\sigma_I^2 = 50 u^2$ ,  $\alpha = 0.05$  bilateral y  $\beta = 0.20$ ? Haga los cálculos en R.

**Ejercicio 2.11**

Invente valores de  $\sigma_I^2$  y  $\sigma_E^2$  para una variable que le sea familiar.

En ocasiones, se dispone de la varianza total  $\sigma^2$  y de la correlación entre ambas observaciones en los datos apareados. Se obtiene la varianza intra-sujetos mediante:

**Fórmula**

$$\sigma_I^2 = \sigma^2 (1 - \rho)$$

**Ejemplo 2.6.** Se quiere comprobar la eficacia de un nuevo fármaco sobre la presión arterial sistólica (PAS) frente el tratamiento habitual. Se realiza un diseño apareado. La correlación esperada entre la presión inicial y final es 0.8 y la varianza total es  $34(\text{mmHg})^2$ .

Se obtiene la siguiente varianza intra-sujetos:

$$\sigma_I^2 = 34 \cdot (1 - 0.8) = 6.8 (\text{mmHg})^2$$

**Ejercicio 2.12**

Calcule la varianza intra-sujetos de un diseño apareado que presenta una varianza total de  $12u^2$ . Haga los cálculos para los siguientes valores de  $\rho$ : 0.2 , 0.5 y 0.8. ¿Qué efecto tiene una mayor correlación entre ambas observaciones sobre la varianza intra-sujetos?

Como la varianza intra-sujetos será menor cuanto mayor sea la correlación o similitud entre ambas respuestas, el beneficio de este diseño es mayor cuanto más *apareados* están los datos.

Otras veces se dispone de la variancia de la variable diferencia  $\sigma_D^2$  entre ambas medidas. Se obtiene la varianza intra-sujetos mediante:

**Fórmula**

$$\sigma_I^2 = \sigma_D^2 / 2$$

**Ejemplo 2.7.** En un estudio con datos apareados en el que la varianza de la variable diferencia es  $16u^2$ , la varianza intra-sujetos es de  $8u^2$ .



**Ejercicio 2.13**

¿Cuál es la varianza intra-casos en un diseño apareado en el que se conoce que la desviación típica de la variable diferencia es  $12u$ ?

**2.3. Análisis del cambio**

Una combinación de las dos situaciones anteriores estudia el efecto comparando (entre los grupos de tratamiento) el cambio en una variable desde el momento basal hasta el final del seguimiento. En este caso, dentro de cada grupo, disponemos de datos apareados (medidas basal y final) que después se compararan entre grupos como datos independientes.

En un ensayo aleatorizado, como las basales vienen de la misma población, comparar el cambio estima el mismo efecto que comparar las respuestas finales: ambos son insesgados. El objetivo de usar el cambio como variable principal es disminuir la varianza de la respuesta, y en consecuencia, aumentar la potencia del estudio. El análisis del "cambio" es más eficiente si la correlación entre la variable basal y final es mayor que 0.5.

**Fórmula**

La varianza de la nueva variable cambio ( $\sigma_C^2$ ) será:

$$\sigma_C^2 = 2 \cdot (1 - \rho) \cdot \sigma^2$$

**Nota:** se debe asumir que ni el paso del tiempo ni el tratamiento afectan a la dispersión, es decir, homoscedasticidad o misma variabilidad entre tratamientos y entre tiempos.

**Nota:** En este caso, también se deberá echar mano de la literatura para encontrar estimaciones de la correlación basal-final.

**Nota:** Fíjese que para valores de correlación ( $\rho$ ) inferiores a 0.5, la varianza del cambio resulta superior a la de la respuesta final y por tanto, no aporta ventajas.

**Fórmula**

El **tamaño** necesario en la comparación de 2 medias del **cambio** es:

$$n = \frac{2 \cdot \sigma_C^2 \cdot (z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$

**Ejemplo 2.8:** Se desea comparar dos tratamientos para eliminar la placa dental. Se planea un estudio a 6 meses donde se estudiará el cambio en la presencia de dicha placa a través de un índice estándar de medida. La correlación esperada entre la cantidad de placa inicial y final es 0.75 y la desviación típica de la respuesta final es 0.2. Si se pretende detectar una diferencia de 0.1 con una potencia del 80% y un riesgo  $\alpha=0.05$  bilateral, ¿cuál es el tamaño muestral necesario?

$$\sigma_C^2 = 2 \cdot (1 - 0.75) \cdot 0.2^2 = 0.02$$

$$n = \frac{2 \cdot 0.02 \cdot (1.96 + 0.84)^2}{0.1^2} = 31.36 \rightarrow 32 \text{ pacientes por grupo}$$



#### Ejercicio 2.14

Compruebe que este número hubiese sido exactamente el doble en caso de usar el índice a los 6 meses como respuesta y explique las razones.

**Nota técnica:** Un análisis basado en el modelado estadístico (ANCOVA) es sistemáticamente más eficiente que el análisis de la variable final o el análisis del cambio. Se estudia en el curso “observacionales”.

### 2.4. Estudios de equivalencia y no-inferioridad

Se ha visto que demostrar equivalencia requiere poder afirmar que  $\varepsilon_1 < \mu_1 - \mu_2 < \varepsilon_2$ , o bien, si hacemos ambos límites simétricos, que  $|\mu_1 - \mu_2| < \varepsilon$ . El contraste de hipótesis para establecer equivalencia es pues:

$$\begin{cases} H_0: \mu_A - \mu_B = \varepsilon \\ H_1: \mu_A - \mu_B = 0 \end{cases}$$



#### Ejercicio 2.15

En los estudios de diferencias,  $\Delta$  representaba cierta diferencia de interés, con relevancia clínica. ¿Qué significa  $\varepsilon$  en los estudios de equivalencia y no-inferioridad?



#### Recuerde

$\Delta$  representa el efecto relevante en estudios de diferencias y  $\varepsilon$  el efecto irrelevante en los de equivalencia y no inferioridad.

El margen  $\varepsilon$  de irrelevancia debe ser menor que el  $\Delta$  de relevancia.

**Ejemplo 2.9:** Si un hipotensor A comparado con placebo ha demostrado bajar la PA, pongamos, 15mmHg, y deseamos demostrar que nuestro hipotensor B es equivalente a A, parece sensato proponer que  $\varepsilon$  sea la mitad o la tercera parte de 15mmHg. EMEA, la agencia europea de regulación, ofrece [consideraciones](#) más detalladas.

**Nota:** Podemos distinguir  $\Delta$  y  $\varepsilon$  al verlos consecutivamente. Muchas referencias monográficas sobre estudios de equivalencia o el mismo R utilizan  $\Delta$  para ambos conceptos –lo que facilita el uso de la misma fórmula, pero dificulta entender la principal razón de un mayor tamaño muestral. Una vez más, el cálculo se hace igual que antes con las salvedades siguientes:

- El efecto  $\Delta$  del planteamiento clásico, que hace relevante una diferencia, se sustituye por el efecto  $\varepsilon$ , que la hace irrelevante. Como  $\varepsilon \ll \Delta$ , las  $n$ 's serán muy mayores.
- El planteamiento bajo la hipótesis nula es unilateral. Es decir, se cambia  $z_{\alpha/2}$  por  $z_{\alpha}$  y las  $n$ 's serán (algo) menores.
- El planteamiento bajo la hipótesis alternativa es bilateral en equivalencia (cambia  $z_{\beta}$  por  $z_{\beta/2}$ ) y unilateral en no inferioridad —dónde sólo interesa un lado.

### Fórmula

El tamaño muestral para la **comparación de medias**

- en un **estudio de equivalencia** es  $n = \frac{2\sigma^2(z_{\alpha} + z_{\beta/2})^2}{\varepsilon^2}$
- en un **estudio de no-inferioridad** es  $n = \frac{2\sigma^2(z_{\alpha} + z_{\beta})^2}{\varepsilon^2}$

**Ejemplo 2.10:** Se desea estudiar un nuevo hipotensor, más seguro y barato. Se considera suficiente con demostrar que el clásico no le supera en 5 mmHg (no-inferioridad). ¿Cuántos casos se necesitan si  $\sigma = 15$  mmHg,  $\alpha=0.025$  unilateral,  $\beta=0.2$ ?

$$\begin{cases} H_0: \mu_A - \mu_B = 5 \\ H_1: \mu_A - \mu_B = 0 \end{cases}$$

$$n = \frac{2 \cdot 15^2 \cdot (1.96 + 0.84)^2}{5^2} \approx 141.28 \rightarrow 142 \text{ casos por grupo}$$



### Ejercicio 2.16

¿Cuántos casos se necesitan si  $\varepsilon=5$  mmHg,  $\sigma=15$  mmHg,  $\alpha=0.05$  unilateral,  $\beta=0.2$  en un estudio de no-inferioridad?

**Ejercicio 2.17**

En estos estudios, ¿qué riesgo(s)  $\alpha$  y/o  $\beta$  pueden perjudicar al usuario y al patrocinador?

**Recuerde**

La potencia en estudios de equivalencia (o no-inferioridad) es la probabilidad de establecer que dos tratamientos son equivalentes (o uno no-inferior a otro) cuando realmente es así.

Ahora bien, se puede desear establecer equivalencia a pesar de que se crea que los dos tratamientos no son absolutamente idénticos: aunque tengan diferencias muy pequeñas entre ellos, menores que la irrelevancia  $\varepsilon$ , podrían ser alternativas terapéuticas.

**Lectura:** “when the power of a non-inferiority trial is assessed at a zero difference, then the sample size needed to achieve that power will be underestimated if the effect of the investigational product is less than that of the active control” ([ICH E9](#), point 3.5).

**Recuerde**

Es conveniente disponer de cierto margen de seguridad, por si los productos no fueran absolutamente idénticos.

Para ello, en las fórmulas habituales se pone, en lugar del límite de no-equivalencia  $\varepsilon$ , la diferencia  $\delta$  entre este límite de no-equivalencia y el margen de seguridad MS que se desea cubrir:  $\delta = \varepsilon - MS$  donde  $\delta$  es el valor a poner en la fórmula:

$$n = \frac{2\sigma^2(z_\alpha + z_\beta)^2}{\delta^2} \quad \text{ó} \quad n = \frac{2\sigma^2(z_\alpha + z_{\beta/2})^2}{\delta^2}$$

**Ejemplo 2.11** (continuación del Ejemplo 2.10) Ahora bien, se sospecha que este nuevo hipotensor más seguro y barato puede no ser absolutamente idéntico al clásico y se desea seguir teniendo la misma probabilidad de demostrar su eficacia incluso en el caso de que el clásico le superara en 1 mmHg. En resumen, ¿cuántos casos se necesitan si  $\varepsilon=5$  mmHg, MS= 1mmHg,  $\sigma=15$  mmHg,  $\alpha=0.05$  unilateral,  $\beta=0.2$  (es decir, potencia del 80%)?

$$\delta = \Delta - MS = 5 - 1 = 4$$

$$n \approx \frac{2 \cdot 15^2 (1.645 + 0.84)^2}{4^2} = 173.68 \rightarrow 174 \text{ casos por grupo}$$



### Ejercicio 2.18

¿Cuántos casos se necesitan si aumentamos el margen hasta 2 mmHg [ $MS=2$  mmHg,  $\varepsilon=5$  mmHg,  $\sigma=15$  mmHg,  $\alpha=0.05$  unilateral,  $\beta=0.2$ ?

En R, las funciones *TwoSampleMean.Equivalence* y *TwoSampleMean.NIS* realizan el cálculo del tamaño muestral en comparación de medias en estudios de equivalencia y de no-inferioridad respectivamente.



### Ejemplo de R

```
# Aplicación al Ejemplo 2.11 (k=1 implica grupos iguales, delta se
refiere a  $\varepsilon$ )
> n <- TwoSampleMean.NIS (alpha=0.025, beta=0.2, sigma=15, k=1,
delta=5, margin=0)
> n
[1] 141.2798
```



### Ejercicio 2.19

Para probar la no-inferioridad de un nuevo fármaco (B) para la degeneración macular asociada a la edad, que es menos costoso que el de referencia (A), se decide diseñar un ensayo con variable respuesta el cambio (antes/después del tratamiento) en el número de letras que el paciente es capaz de identificar a una cierta distancia. Calcule con R el tamaño de la muestra por grupo con los parámetros siguientes:

$MS=0$  letras,  $\varepsilon=5$  letras,  $\sigma=15$  letras,  $\alpha=0.004$  unilateral,  $\beta=0.1$ .

## 2.5. Precisión de la estimación del efecto

Hemos visto el cálculo del tamaño muestral que se usa habitualmente, basado en el enfoque de decisión de Neyman y Pearson para escoger entre dos alternativas. Sin embargo, el objetivo de un estudio podría ser estimar el efecto de una intervención con una determinada precisión. Estos estudios son análogos a los de estimación de un parámetro poblacional, pero en este caso, este parámetro es el efecto de una intervención concreta —definido como la diferencia entre las respuestas en los 2 grupos. Ahora ya no es necesario especificar la potencia sino la semi-amplitud (e) del intervalo requerida.

**Fórmula**

El tamaño muestral necesario para una **comparación de medias** para una **precisión determinada del efecto** de la intervención es:

$$n = \frac{2 \cdot \sigma^2 \cdot z_{\alpha/2}^2}{e^2}$$

**Ejemplo 2.12.** Se está diseñando un ensayo clínico para probar la eficacia de un nuevo fármaco antidiabético (B) respecto a uno de referencia (A). Se supone que la desviación estándar de la respuesta es 14 mg/dl. Se desea conocer el efecto con una confianza del 95% y un error (e) de 3 mg/dl

$$n = \frac{2 \cdot 14^2 \cdot 1.96^2}{3^2} = 167.32 \rightarrow 168 \text{ casos por grupo}$$



**Ejercicio 2.20**

Se desea comparar dos tratamientos para reducir el nivel de colesterol en sangre. Se estima una variabilidad en la respuesta de 8 mg/l. Se desea obtener un IC con un 99% de confianza y una amplitud no mayor de 5 mg/l (e=2.5). ¿Cuántos casos son necesarios?

**2.6. Tamaño desigual en ambos grupos \***

Razones éticas, logísticas o económicas pueden aconsejar asignar a los pacientes en lugar del equilibrio habitual 1 a 1, con razones 2 a 1, 3 a 1, etc.

Ya se ha comentado que, en general, la situación de máxima eficiencia es aquella en que ambos grupos están balanceados. El hecho de que no lo estén comporta una pérdida de potencia. Por ejemplo, la potencia para el caso particular de un efecto  $\Delta = \sigma$  es:

$$Potencia = 1 - \beta = \Phi \left( \sqrt{\frac{n_A \cdot n_B}{N}} - z_{\alpha/2} \right)$$

donde  $\Phi$  es la función de distribución normal estándar (En R, *pnorm*).

**Ejemplo 2.13.** En un estudio de comparación de medias con  $\alpha=0.05$  y  $\Delta = \sigma$ , el tamaño global es  $N=40$ . La potencia resultante dependiendo del equilibrio entre los grupos será:

$$n_A = 20 \text{ y } n_B = 20 \rightarrow Potencia = \Phi \left( \sqrt{\frac{20 \cdot 20}{40}} - 1.96 \right) = \Phi(1.20) = 0.89$$

$$n_A = 15 \text{ y } n_B = 25 \rightarrow \text{Potencia} = \Phi\left(\sqrt{\frac{15 \cdot 25}{40}} - 1.96\right) = \Phi(1.10) = 0.86$$

$$n_A = 10 \text{ y } n_B = 30 \rightarrow \text{Potencia} = \Phi\left(\sqrt{\frac{10 \cdot 30}{40}} - 1.96\right) = \Phi(0.78) = 0.78$$

Con un tamaño global de  $N=40$ , pasar de tener grupos balanceados a tener un desequilibrio de 3 a 1 supone una pérdida de potencia del 11%.



**Recuerde**

La potencia máxima en una comparación de medias se obtiene con los grupos balanceados.



**Recuerde**

La potencia de un estudio se define a priori.

En capítulos posteriores, se verá cómo controlar el equilibrio entre los grupos en un ensayo a través de asignación por bloques.

Las notas siguientes explican dos razones para desequilibrar los grupos: diferentes costes de las intervenciones y diferentes dispersiones de la variable respuesta (heteroscedasticidad) entre grupos.

**Nota:** El coste de los tratamientos no siempre es comparable en un ensayo. Si el objetivo no es maximizar la eficiencia sino minimizar los costes, entonces se introducirá un desequilibrio entre los grupos que cumpla con este objetivo.

**Fórmula**

El cociente entre los tamaños entre grupos según los costes de ambos tratamientos ( $c_A$  y  $c_B$ ) es:

$$\frac{n_A}{n_B} = \sqrt{\frac{c_B}{c_A}}$$

**Ejemplo 2.14.** En un estudio donde el tratamiento nuevo (B) es 4 veces más costoso que el convencional (A), el ratio de asignación debe ser de 2:1.

$$\frac{n_A}{n_B} = \sqrt{\frac{c_B}{c_A}} = \sqrt{4} = 2$$



**Ejercicio 2.21**

El tratamiento B es 2 veces más costoso que el tratamiento A. Si se han asignado 20 pacientes al tratamiento B, ¿Cuántos pacientes deben asignarse al tratamiento A si se consideran los costes? ¿Qué pérdida de potencia supone este desequilibrio si  $\alpha=0.05$ ?

**Nota:** La variabilidad que presentan ambos tratamientos puede diferir. En este caso, es conveniente obtener más casos de aquel tratamiento que presente mayor variabilidad para lograr la máxima eficiencia.

**Fórmula**

El cociente entre los tamaños entre grupos considerando las desviaciones de la respuesta en ambos tratamientos ( $\sigma_A$  y  $\sigma_B$ ) es:

$$\frac{n_A}{n_B} = \frac{\sigma_A}{\sigma_B}$$

**Ejemplo 2.15.** En un estudio donde el tratamiento nuevo (B) presenta la mitad de variabilidad que el convencional (A), el ratio de asignación debe ser 1:2.

$$\frac{n_A}{n_B} = \frac{\sigma_A}{\sigma_B} = \frac{1}{2}$$

**Nota:** Nótese que no se han detallado las fórmulas teóricas para llevar a cabo los cálculos con desequilibrios. Sin embargo el parámetro  $k$  que contienen las funciones del paquete *TrialSize* permite hacer los cálculos considerando este aspecto.

**Ejemplo 2.16 (cont).** Se quiere detectar una diferencia de 10 unidades entre el tratamiento nuevo (B) y el convencional (A), siendo  $\sigma=8$  y los riesgos habituales ( $\alpha = 0.05$  ;  $\beta=0.20$ ). Si se quiere que el ratio de asignación sea 1:2, el tamaño muestral es:

```
> n = TwoSampleMean.Equality(alpha=0.05, beta=0.20, sigma=8,
margin=10, k=1/2)
> n
[1] 7.534925
```

Por lo que en el grupo de control habría 8 pacientes y en el grupo del nuevo tratamiento habría 16 pacientes.

En cambio, si se quisiera que el ratio de asignación fuera 2:1, el tamaño muestral sería:

```
> n = TwoSampleMean.Equality(alpha=0.05, beta=0.20, sigma=8,
margin=10, k=2)
> n
[1] 15.06985
```

En el grupo de control se reclutarían 16 pacientes y en el del nuevo tratamiento 8 pacientes. Note que la n que proporciona la función de R siempre es el tamaño muestral del grupo de control.

### 3. Comparación de probabilidades

La teoría que subyace detrás del cálculo en una comparación de probabilidades es análoga a la de comparación de medias ya que la distribución del estadístico sigue, como ya vimos, una distribución Normal. Sin embargo, la fórmula es un poco más compleja. Sea el contraste:

$$\begin{cases} H_0: \pi_A = \pi_B \\ H_1: \pi_A = p_A \text{ y } \pi_B = p_B \end{cases}$$

#### Fórmula

El tamaño muestral en cada grupo para la **comparación de 2 probabilidades** es:

$$n = \left( \frac{z_{\alpha/2} \cdot \sqrt{2p(1-p)} + z_{\beta} \cdot \sqrt{p_A(1-p_A) + p_B(1-p_B)}}{p_A - p_B} \right)^2$$

donde  $p = \frac{p_A + p_B}{2}$

**Ejemplo 3.1:** El porcentaje de pacientes que tras sufrir un ictus isquémico se encuentran en un estado de gravedad leve ( $mRS \leq 1$ ) al cabo de 3 meses es un 20% con el tratamiento convencional (A). Se desea probar un nuevo fármaco (B) en un ECA que incremente esta proporción hasta el 30%. Con un riesgo alfa del 5% y una potencia del 80%, ¿cuál es el tamaño necesario por grupo?

$$n = \left( \frac{1.96 \cdot \sqrt{2 \cdot 0.25(1-0.25)} + 0.84 \cdot \sqrt{0.2(1-0.2) + 0.3(1-0.3)}}{0.2 - 0.3} \right)^2 = 292.82$$

Se necesitan 293 casos por grupo.

Con R, se puede emplear la función *TwoSampleProportion.Equality* del paquete *TrialSize* para comparar proporciones.

**Ejemplo de R**

```
# Aplicación al Ejemplo 3.1 (k=1 implica grupos iguales)
> TwoSampleProportion.Equality(alpha=0.05,beta=0.2,p1=0.2,
                               p2=0.3, k=1, delta=0.1)
[1] 290.4086
```

**Nota:** El resultado difiere ligeramente del ejemplo por una corrección por continuidad que realiza R.

**Ejercicio 3.1 [Basado en un ejemplo del libro de [Julious](#)]**

En un estudio se quiere comparar las eficacias del midazolán oral con la del diazepam rectal en las emergencias por convulsiones en niños. Se sabe que el éxito del midazolán está en torno al 55% y se quiere saber si el diazepam incrementa la probabilidad de éxito hasta el 75%. Con un riesgo  $\alpha$  del 5% y una potencia del 90% ¿Que tamaño se requiere por grupo?

**Nota:** Una fórmula alternativa a la propuesta para la comparación de proporciones es:

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2}{2(\text{asen}(\sqrt{\pi_A}) - \text{asen}(\sqrt{\pi_B}))^2}$$

## 4. Comparación del tiempo hasta un evento

Este apartado hace referencia a los estudios donde la variable de interés es el tiempo que transcurre hasta un evento, como los de supervivencia. En este tipo de estudios primero se calcula el número de eventos E que se necesita observar y luego se obtiene el número de casos y el tiempo que deben ser seguidos para poder observar esos eventos E.

Existen diversos métodos dependiendo de las premisas y el reclutamiento.

### 4.1. Participantes reclutados en un mismo instante

Un posible tipo de estudio más sencillo es aquel en que todos los participantes se incorporan al unísono. Se trata de poblaciones “secuestradas” en un ámbito determinado (trabajadores de un centro, ancianos en una residencia, alumnos de un colegio...).

En el cálculo del tamaño muestral, se puede tener la premisa de que las tasas ( $\lambda_A$ ,  $\lambda_B$ ) de aparición del evento en ambos grupos son constantes a lo largo del tiempo o se puede relajar asumiendo únicamente que su cociente, llamado Hazard Rate Ratio ( $\text{HRR} = \lambda_A/\lambda_B$ ) es constante.

### Fórmula

El número de **eventos totales** para la comparación de 2 **tiempos hasta un evento** en el caso de **reclutamiento instantáneo** es:

$$\text{Tasas constantes} \rightarrow E = \frac{4 \cdot (z_{1-\alpha/2} + z_{1-\beta})^2}{[\ln(\text{HRR})]^2}$$

$$\text{HRR constante} \rightarrow E = \frac{(\text{HRR} + 1)^2 \cdot (z_{1-\alpha/2} + z_{1-\beta})^2}{(\text{HRR} - 1)^2}$$

Y el **tamaño muestral total** (en ambas situaciones) es:

$$N = \frac{2E}{2 - \pi_A - \pi_B}$$

donde  $\pi_A$  y  $\pi_B$  son la proporción estimada de casos donde NO se presentará el evento durante el estudio.

**Nota:** Tasas constantes implica HRR constante. El recíproco no es cierto.

**Nota:** En supervivencia, el diseño más eficiente no es aquel tal que  $N_A=N_B$  sino aquel en que  $E_A=E_B$

**Nota:** Dado que la premisa de tasas constantes en ambos grupos es más restrictiva, está fórmula proporcionará tamaños menores (más información a priori comporta menos necesidad de información a posteriori).



### Recuerde

La premisa puede ser o bien que la tasa de aparición de eventos es constante a lo largo del tiempo, o bien que su cociente (HRR) es constante.

**Ejemplo 4.1:** En una universidad, se decide hacer un estudio para comparar el tiempo de adherencia a dos dietas saludables (el evento en este caso es el abandono de la dieta). Para ello, se recluta personal laboral de la universidad que iniciará una de las dos dietas asignada aleatoriamente al inicio del curso escolar. El tiempo de seguimiento es de 9 meses (=0.75 años) y se estiman unas proporciones de personas que acabarán el curso con la dieta de  $\pi_A = 0.55$  y  $\pi_B = 0.65$ . Asumiendo un HRR constante de 0.72, con una potencia del 90% y un riesgo  $\alpha$  del 5%, estime el tamaño muestral necesario.

El número de eventos totales necesarios (personas que abandonarían la dieta) son:

$$E = \frac{(0.72 + 1)^2 \cdot (1.96 + 1.28)^2}{(0.72 - 1)^2} = \frac{2.96 \cdot 10.51}{0.078} = 396.49$$

Ahora, ya se puede calcular la N:

$$N = \frac{2 \cdot 396.49}{2 - 0.55 - 0.65} = 991.24$$

Por tanto, se requieren 496 participantes por grupo (la mitad de 991.24 redondeado al alza)

**Nota técnica:** Fíjese que el tiempo de seguimiento no ha sido utilizado en los cálculos. De hecho, este dato proporciona el HRR a partir de  $\pi_A$  y  $\pi_B$ .

$$\left. \begin{aligned} \lambda_A &= -\frac{\ln(\pi_A)}{T} \\ \lambda_B &= -\frac{\ln(\pi_B)}{T} \end{aligned} \right\} \rightarrow HRR = \frac{\lambda_B}{\lambda_A}$$

[Véase el apartado de la distribución exponencial en el capítulo 6. El HRR se ha introducido en capítulos anteriores y se ampliará su explicación más adelante]



### Ejercicio 4.1

Repita el Ejemplo 4.1 pero asumiendo que las tasas son constantes.

- A) A priori, el tamaño que se obtendrá ¿será mayor o menor que el del ejemplo?
- B) Dadas las proporciones, ¿cuáles son las tasas? Use las fórmulas de la nota técnica.
- C) Finalmente, ¿cuál es el tamaño por grupo requerido?



### Ejercicio 4.2

En una residencia se desea comparar el tiempo hasta un evento cardiovascular entre dos grupos de pacientes mayores de 70 años asignados aleatoriamente a dos programas de mantenimiento deportivo (A: Clásico y B: Nuevo). El estudio está pensado a 3 años y se espera obtener una proporción de eventos al cabo de estos 3 años de 0.2 en el clásico y 0.15 en el nuevo. Asuma tasas constantes y riesgos  $\alpha$  y  $\beta$  de 0.05 y 0.2, respectivamente.

La función `ssizeCT.default` del paquete `powerSurvEpi` calcula el tamaño en el caso de HRR constante.



### Ejemplo de R

```
# Aplicación al Ejemplo 4.1
> ssizeCT.default(power=0.9,k=1, pE=0.45, pC=0.35, RR=0.72, alpha=0.05)
  nE  nC
496 496
```

## 4.2. Participantes reclutados a lo largo del tiempo

Es la situación más habitual en los ensayos clínicos: los pacientes van entrando en el estudio a lo largo del tiempo. Los parámetros necesarios para el cálculo son el tiempo de reclutamiento (TR), el tiempo que dura el estudio (T), los riesgos  $\alpha$  y  $\beta$  y las tasas de ocurrencia del evento en ambos grupos (en este caso deben asumirse constantes).

### Fórmula

El tamaño muestral **total** para la comparación de 2 **tiempos hasta un evento** es:

$$N = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 (\Phi(\lambda_A) + \Phi(\lambda_B))}{(\lambda_A - \lambda_B)^2}$$

donde:

$$\Phi(\lambda) = \frac{\lambda^3 \cdot TR}{\lambda \cdot TR - (e^{-\lambda(T-TR)} - e^{-\lambda T})}$$

$\lambda$  es la tasa [eventos/paciente x unidad de tiempo] esperada.

**Nota:** El tiempo de duración del estudio (T) se refiere al de obtención de datos e incluye el de reclutamiento TR y el de seguimiento de los pacientes, que puede ser fijo o variable. T debe ser estrictamente superior a TR —de lo contrario, los pacientes reclutados el último segundo serían seguidos sólo 1 segundo.

**Nota:** La unidad de tiempo en que se especifique la tasa ( $\lambda$ ) debe ser la misma en la que se especifiquen TR y de duración del estudio T. Por ejemplo, si ambos se especifican en años, la  $\lambda$  se deberá especificar en proporción de eventos al año.

### Recuerde

La premisa es que la tasa de aparición de eventos es constante a lo largo del tiempo.

**Ejemplo 4.2:** Se quiere diseñar un estudio para comparar la eficacia de dos tratamientos respecto al tiempo hasta la progresión de la enfermedad en un determinado tipo de cáncer. Las tasas esperadas son  $\lambda_A = 50\%$  (referencia) y  $\lambda_B = 35\%$  (nuevo tratamiento). Se fija un tiempo de reclutamiento de 1.5 años y un tiempo de seguimiento máximo de 2 años. Asumiendo tasas constantes, con una potencia del 90% y un riesgo  $\alpha$  del 5%, estime el tamaño muestral necesario.

Se empieza calculando las  $\Phi$ 's:

$$\Phi(\lambda_A) = \frac{0.5^3 \cdot 1.5}{0.5 \cdot 1.5 - (e^{-0.5(2-1.5)} - e^{-0.5 \cdot 2})} = \frac{0.1875}{0.75 - 0.41} = 0.553$$

$$\Phi(\lambda_B) = \frac{0.35^3 \cdot 1.5}{0.35 \cdot 1.5 - (e^{-0.35(2-1.5)} - e^{-0.35 \cdot 2})} = \frac{0.064}{0.525 - 0.34} = 0.353$$

Ahora, ya se puede calcular la N:

$$N = \frac{2(1.96 + 1.28)^2(0.553 + 0.353)}{(0.5 + 0.35)^2} = 846.27$$

Por tanto, se requieren 424 participantes por grupo (la mitad de 846.27 redondeado al alza)

Con R, la función *TwoSampleSurvival.Equality* realiza el cálculo del tamaño en este supuesto. El parámetro *gamma* de la función con un valor positivo suficientemente bajo (sin llegar a ser cero) asume que el reclutamiento es uniforme a lo largo del tiempo (supuesto razonable en la mayoría de estudios). Esta función retorna 3 valores:  $\Phi(\lambda_A)$ ,  $\Phi(\lambda_B)$  y la *n* en cada grupo.



#### Ejemplo de R

```
# Aplicación al Ejemplo 4.2
> TwoSampleSurvival.Equality(alpha=0.05,beta=0.1,lam1=0.50,
lam2=0.35,k=1, ttotal=2, taccrual=1.5,gamma=0.000001)
[1] 0.552969
[1] 0.3531164
[1] 423.1388
```



#### Ejercicio 4.3

Se quiere diseñar un ensayo clínico que compare el tiempo hasta la muerte entre dos tratamientos en pacientes con cancer de ovario en estadios avanzados. El tratamiento A consistirá en quimioterapia y el tratamiento B en quimioterapia + cirugía de citorreducción. Calcule con R el tamaño necesario por grupo con los siguiente parámetros:

Potencia = 80%,  $\alpha=0.05$ ,  $\lambda_A = 0.40$ ,  $\lambda_B = 0.30$ , T=6 años, TR =5 años

## 5. Consejos prácticos

Hemos visto que el cálculo muestral precisa valores que reflejan la ambición de la investigación (confianza, precisión, delta, etc.) y conocimientos previos sobre las variables en estudio (sigma, tasa en el grupo control, etc.). Estos últimos forman parte de las premisas del estudio y conviene ser prudente al fijar sus valores.

**Nota técnica:** estos parámetros reciben el calificativo de molestos o estorbo (“nuisance”) ya que son necesarios para especificar el modelo o realizar los cálculos, pero no son el objetivo principal del estudio.

Todas estas premisas necesarias para el cálculo del tamaño muestral hacen que no se pueda considerar una ciencia exacta. Nosotros aconsejamos un proceso iterativo conjunto entre investigadores y metodólogos.

El **primer paso** es mirar qué han hecho otros investigadores: ¿cuáles eran sus objetivos, tipo de diseño, variable principal, análisis estadístico, tamaño empleado, ...?.

El **segundo paso** es diseñar el borrador del propio estudio con valores aproximados de los parámetros necesarios ( $\Delta$ ,  $\sigma^2$ ): ¿Es razonable el número resultante? ¿Incluiremos este número de casos en un plazo y con un coste razonable? ¿Tendrán capacidad de convicción los resultados?

El **tercer paso** consiste en comparar este diseño con otros alternativos y repetir el segundo paso, hasta optar por un diseño concreto.

Una vez decidido un diseño, el **cuarto paso** consiste en ajustar por las pérdidas de información. Aunque el informe final debe incluir todos los casos reclutados, eso no significa que todos ellos aporten la información deseada. Si la previsión es tener una proporción  $r$  de casos no informativos, conviene re-ajustar el tamaño calculado ( $n$ ) a uno nuevo ( $n'$ ) que tiene en cuenta estos casos no informativos:

$$n' = \frac{1}{1-r} \cdot n$$

Finalmente, el **quinto paso** consiste en documentar el cálculo del tamaño muestral, lo que requiere justificar los valores empleados en las fórmulas y usar tablas publicadas o programas validados para el cálculo definitivo.

Especialmente aconsejables son, en castellano, el programa [ene](#) de Llorenç Badiella; y en Inglés, los libros: (1) [Sample Size Tables for Clinical Studies](#) (Machin & Campbell); (2) [Sample Sizes for Clinical Trials](#) (Julious); y (3) [Fundamentals of Clinical Trials](#) (Friedman) para tiempo hasta el evento.

**Historieta:** según [Montgomery](#), el mejor momento para diseñar un estudio es cuando termina.



		Fórmula	R
Estimación de un parámetro	Media	$n = \left(\frac{2 \cdot z_{\alpha/2} \cdot \sigma}{A}\right)^2$	<i>sample.size.mean</i>
	Probabilidad	$n = \left(\frac{z_{\alpha/2}}{A}\right)^2$	<i>sample.size.prop</i>
Comparación de medias	Independientes	$n = \frac{2 \cdot \sigma^2 \cdot (z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$	<i>TwoSampleMean.Equality</i>
	Apareadas	$N = \frac{2 \cdot \sigma_I^2 \cdot (z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$	-
	Análisis del cambio	$n = \frac{2 \cdot \sigma_c^2 \cdot (z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$	-
	Equivalencia	$n = \frac{2\sigma^2(z_{\alpha} + z_{\beta/2})^2}{\Delta^2}$	<i>TwoSampleMean.Equivalence</i>
	No-Inferioridad	$n = \frac{2\sigma^2(z_{\alpha} + z_{\beta})^2}{\Delta^2}$	<i>TwoSampleMean.NIS</i>
	Precisión	$n = \frac{8 \cdot \sigma^2 \cdot z_{\alpha/2}^2}{A^2}$	-
Comparación de probabilidades	Independientes	$n = \left(\frac{z_{\alpha/2}\sqrt{2p(1-p)}}{p_A - p_B} + \frac{z_{\beta}\sqrt{p_A(1-p_A) + p_B(1-p_B)}}{p_A - p_B}\right)^2$	<i>TwoSampleProportion.Equality</i>
Comparación de tiempos	Reclutamiento instantáneo	$N = \frac{2E}{2 - \pi_A - \pi_B}$	<i>ssizeCT.default</i> (para HRR constantes)
	Reclutamiento durante un periodo	$N = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2(\Phi(\lambda_A) + \Phi(\lambda_B))}{(\lambda_A - \lambda_B)^2}$	<i>TwoSampleSurvival.Equality</i>

**Tabla 5.1.** Tabla resumen de las fórmulas vistas en este capítulo

**Nomenclatura.**  $\sigma^2$ : varianza total,  $\sigma_I^2$ : varianza intra-sujetos,  $\sigma_c^2$ : varianza de la variable cambio,  $\Delta$ : diferencia que se quiere detectar,  $\rho$ : correlación entre ambas observaciones,  $A$ : amplitud del intervalo,  $E$ : número de eventos que se necesita observar,  $\lambda$ : tasa de aparición del evento y  $\pi$ : proporción estimada de casos donde NO se presentará el evento durante el estudio.

## Soluciones a los ejercicios

1.1  $n \approx (Z_{\alpha/2} \cdot \sigma / e)^2 = (1.96 \cdot 1.5 / 0.25)^2 \approx 138.2976 \rightarrow 139$  casos

1.2 `> sample.size.mean(e=2.5, S=10, level = 0.95)`  
Sample size needed: 62

1.3 Se infra-estimó la variabilidad de la respuesta en el cálculo del tamaño.

1.4  $n \approx (Z_{\alpha/2} / 2 \cdot e)^2 = (1.96 / 2 \cdot 0.025)^2 \approx 1536.64 \rightarrow 1537$  casos

1.5 `> sample.size.prop(e=0.02, level = 0.90)`  
Sample size needed: 1691

1.6 a) `> ppois(q=0, lambda=300*1/1000, lower.tail = FALSE)` [1] 0.2591818 (25.9%)  
b) `> ppois(q=0, lambda=600*1/1000, lower.tail = FALSE)` [1] 0.4511884 (45.1%)  
c) `> ppois(q=0, lambda=300*2/1000, lower.tail = FALSE)` [1] 0.4511884 (45.1%)  
d) `> ppois(q=0, lambda=600*2/1000, lower.tail = FALSE)` [1] 0.6988058 (69.9%)  
e) `> ppois(q=0, lambda=100*4/1000, lower.tail = FALSE)` [1] 0.32968 (33.0%)

2.1  $\Delta$  es el valor de la diferencia entre los tratamientos para el que se desea tener una probabilidad  $1-\beta$  de demostrar que los tratamientos son diferentes. Conviene que coincida con la eficacia real y también con la ideal.

2.2 El riesgo  $\alpha$  es la probabilidad de que un tratamiento no eficaz ( $H_0$ ) se declare eficaz y se ponga en el mercado ( $A_1$ ). El riesgo  $\beta$  es la probabilidad de que un tratamiento eficaz ( $H_1$ ) se declare no eficaz y no se lleve al mercado ( $A_0$ ). Ambos repercuten negativamente en el usuario, a quien representa la administración, y en el patrocinador. Al usuario porque puede estar pagando por un producto que no es eficaz ( $\alpha$ ) o porque no se puede beneficiar de uno que lo es ( $\beta$ ). Al patrocinador, porque no comercializa un producto eficaz ( $\beta$ ) o porque pierde energías en uno que no lo es ( $\alpha$ ).

2.3  $n = [2 \cdot 8^2 (1.96 + 0.84)^2] / 5^2 \approx 40.14 \rightarrow 41$  casos por grupo (debe redondearse al alza).

El nomograma ofrece un resultado similar (doble ya que su  $N$  es la total =  $2n$ )

2.4 Use la fórmula y compruebe que coincide su cálculo con el nomograma.

2.5 La instrucción `seq (inicio,fin,separación)` genera todas las potencias para las que queremos hacer el cálculo.

Recuerde que las  $\beta$ 's son el complementario de las potencias.

```
> potencias <- seq(0.1,0.9,0.1)
> betas <- 1-potencias
> betas
[1] 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1
```

Con la función `TwoSampleMean.Equality` se realiza el cálculo de todos los tamaños a la vez. La instrucción `ceiling` redondea al alza.

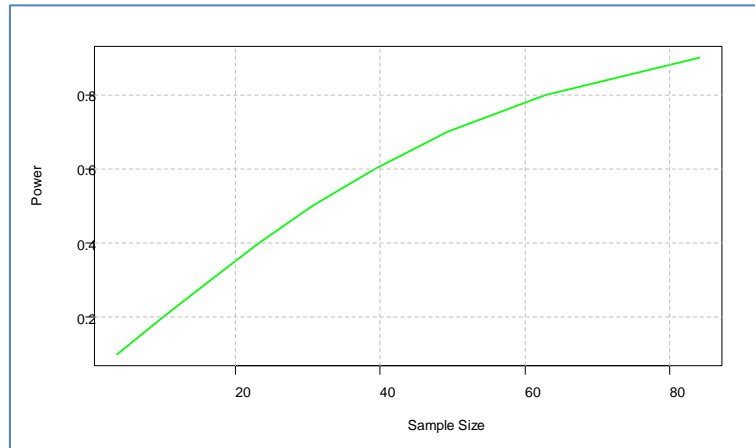
```
> n <- TwoSampleMean.Equality(alpha=0.05,beta=betas,sigma=10,margin=5,k=1)
> ceiling(n)
[1] 4 11 17 24 31 40 50 63 85
```

La función `plot` con el parámetro `type="l"` (une los puntos con líneas) realiza el gráfico:

```
> plot(n,potencias,type="l")
```

Puede mejorar el gráfico añadiendo más parámetros (*?par*) y poniendo una rejilla con abline (*?abline*).

```
> plot(n,potencias,type="l",lwd=2,col="green",las=1,xlab="Sample Size",
      ylab="Power")
> abline(v=seq(20,80,20),h=seq(0.2,0.8,0.2),lty=2,col="grey")
```



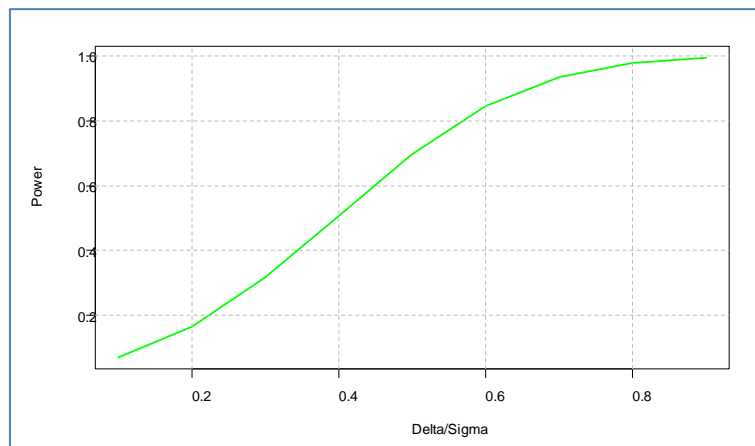
Puede verse como, para un efecto  $\Delta$  que representa el 50% de la desviación típica  $\sigma$ , se necesitan, para la potencia usual del 80%, algo más de 60 casos por grupo. Nótese en la representación gráfica como crece la potencia a medida que aumenta el número de casos.

## 2.6 El siguiente código permite hacer el gráfico

```
# Deltas
> deltas <- 1:9
> deltas
[1] 1 2 3 4 5 6 7 8 9

# Potencias
> potencias <- power.t.test(n=50, delta=deltas, sd=10, sig.level=0.05)$pow
> round(potencias,2)
[1] 0.07 0.17 0.32 0.51 0.70 0.84 0.93 0.98 0.99

# Gráfico
> efecto_tipificado <- deltas/10
> plot(efecto_tipificado,potencias,type="l",lwd=2,col="green",las=1,
      xlab="Delta/Sigma",ylab="Power")
> abline(v=seq(0.2,0.8,0.2),h=seq(0.2,1,0.2),lty=2,col="grey")
```



Puede verse que con 50 casos por grupo se tiene una potencia algo superior al 80% para un efecto que represente el 60% de la dispersión entre los casos estudiados. Nótese también que, si el efecto se acerca al 100% de la dispersión entre los casos, con un diseño de 100 casos por grupo se tiene una potencia que se acerca al 100%, es decir, que, de ser cierto este efecto, la probabilidad de que el resultado del estudio sea significativo se acerca al 100%.

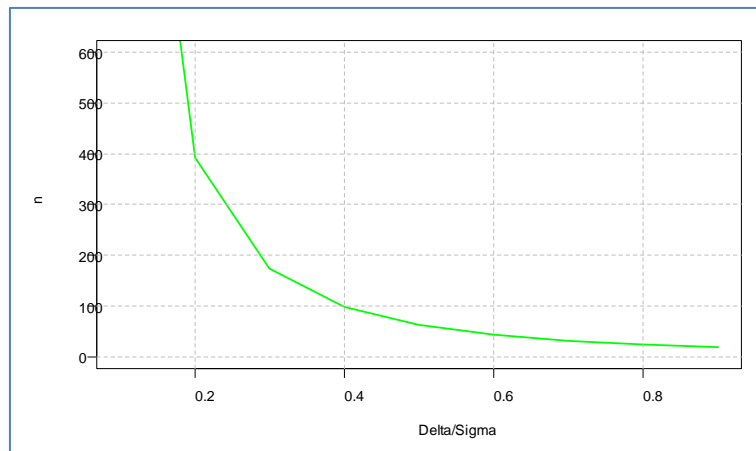
2.7 La potencia del estudio será mayor que la estipulada.

2.8 El siguiente código permite hacer el gráfico:

```
# Deltas
> deltas <- 1:9
> deltas
[1] 1 2 3 4 5 6 7 8 9

# Tamaños
> n <- TwoSampleMean.Equality(alpha=0.05,beta=0.8,sigma=10,margin=deltas,k=1)
> ceiling(n)
[1] 251 63 28 16 11 7 6 4 4

# Gráfico
> efecto_tipificado <- deltas/10
> plot(efecto_tipificado, n, type="l", lwd=2, col="green", las=1,
       xlab="Delta/Sigma", ylab="n", ylim=c(0,600))
> abline(v=seq(0.2,0.8,0.2),h=seq(0,600,100),lty=2,col="grey")
```



Igual que antes, puede verse como, para un efecto  $\Delta$  que representa el 40% de la desviación típica  $\sigma$ , se necesitan aproximadamente 100 casos por grupo. Nótese como va disminuyendo el número de casos necesarios a medida que aumenta la magnitud del efecto que se desea establecer y que el decremento de casos es progresivamente menos acusado.

2.9  $n = [2 \cdot (9^2 + 4^2)(1.96 + 0.84)^2] / 5^2 \approx 60.84 \rightarrow 61$  casos por grupo.

$N = [2 \cdot (4^2)(1.96 + 0.84)^2] / 5^2 \approx 10.04 \rightarrow 11$  casos en total, que dividido por 2 supone 6 casos.

2.10 El siguiente código realiza los dos cálculos:

```
# Independientes
> n <- TwoSampleMean.Equality(alpha=0.05,beta=0.2,sigma=10,margin=5, k=1)
> ceiling(n)
[1] 63
```

```
# Apareadas
> N <- TwoSampleMean.Equality(alpha=0.05, beta=0.2, sigma=sqrt(50), margin=5, k=1)
> ceiling(N/2)
[1] 16
```

**2.11** Compruebe los valores con las referencias de la literatura.

**2.12**  $\rho = 0.2 \rightarrow \sigma_I^2 = 12 \cdot (1 - 0.2) = 9.6 u^2$

$\rho = 0.5 \rightarrow \sigma_I^2 = 12 \cdot (1 - 0.5) = 6 u^2$

$\rho = 0.8 \rightarrow \sigma_I^2 = 12 \cdot (1 - 0.8) = 2.4 u^2$

Cuanto mayor es la correlación entre ambas observaciones, menor es la varianza intra-sujetos.

**2.13**  $\sigma_D^2 = 12^2 = 144 \rightarrow \sigma_I^2 = \frac{144}{2} = 72 u^2$

**2.14**  $n = \frac{2 \cdot 0.2 \cdot (1.96 + 0.84)^2}{0.1^2} = 62.72 \rightarrow 63$  pacientes por grupo.

Utilizando el cambio como variable principal disminuye la varianza de la respuesta, y en consecuencia, aumenta la potencia del estudio. Por lo tanto, el tamaño muestral necesario es menor.

**2.15** Mientras que en los estudios de diferencias,  $\Delta$  representa la diferencia a partir de la cual se empieza a considerar relevantes a las diferencias entre los tratamientos, en los de equivalencia y no-inferioridad, deber representar un valor lo suficientemente pequeño como para que la diferencia entre los dos tratamientos sea irrelevante.

**2.16**  $n \approx [2 \cdot 15^2 (1.645 + 0.84)^2] / 5^2 \approx 111.29 \rightarrow 112$  casos por grupo

**2.17** El riesgo  $\alpha$  es la probabilidad de que dos tratamientos no equivalentes ( $H_0$ ) se declaren equivalentes ( $A_1$ ). El riesgo  $\beta$  es la probabilidad de que un tratamiento equivalente ( $H_1$ ) se declare que no lo es ( $A_0$ ). Igual que antes, ambos deben preocupar a ambos, pero las razones se invierten. El error tipo I del riesgo  $\alpha$  implica sustituir un fármaco por otro cuando no son equivalentes, las consecuencias dependen de la dirección de la no equivalencia y del objetivo del estudio (eficacia o seguridad). Por su parte, el error tipo II del riesgo beta, implica no sustituirlo cuando en realidad sí que son equivalentes. Las consecuencias, como antes, dependen de la situación, aunque la habitual es de tipo económico, ya que no se autoriza un genérico más barato.

**2.18**  $n \approx [2 \cdot 15^2 (1.645 + 0.84)^2] / 3^2 \approx 309.13 \rightarrow 310$  casos por grupo

```
2.19 > n <- TwoSampleMean.NIS (alpha=0.004, beta=0.10, sigma=15,
                                k=1, delta=5, margin=0)
> n
[1] 278.5208
```

Se necesitan 279 pacientes por grupo

**2.20**  $n \approx [2 \cdot 8^2 \cdot 2.58^2] / (2.5)^2 \approx 136.32 \rightarrow 137$  casos por grupo

**2.21**  $n_A \approx \sqrt{(c_B/c_A) \cdot n_B} = \sqrt{2 \cdot 20} = 28.28 \rightarrow 28$  casos asignados al tratamiento A

Potencia con equilibrio  $\rightarrow 1 - \beta = \Phi(\sqrt{(24 \cdot 24 / 48)} - 1.96) = \Phi(1.50) = 0.933$

Potencia sin equilibrio  $\rightarrow 1 - \beta = \Phi(\sqrt{(20 \cdot 28/48)} - 1.96) = \Phi(1.50) = 0.927$

Supone una pérdida de potencia del 0.6%

```
3.1> TwoSampleProportion.Equality(alpha=0.05,beta=0.10,p1=0.55,p2=0.75,  
k=1,delta=0.2)
```

```
[1] 114.2682
```

115 casos por grupo

4.1 A) El tamaño ha de ser ligeramente menor, ya que la premisa es más restrictiva. B)  $\lambda_A = 0.797$ ;  $\lambda_B = 0.574$

C)  $E = 4 \cdot (1.96 + 1.28)^2 / [\ln(0.72)]^2 = 391.3 \rightarrow N = 391.3 / (2 - 0.55 - 0.65) = 978.4 \rightarrow n = 440$  por grupo

4.2  $\pi_A = 0.8$ ;  $\pi_B = 0.85 \rightarrow \lambda_A = 0.074$ ;  $\lambda_B = 0.054 \rightarrow HRR = 0.728$

$E = 4 \cdot (1.96 + 0.84)^2 / [\ln(0.728)]^2 = 312.4 \rightarrow E = 312.4 / (2 - 0.80 - 0.85) = 892.5 \rightarrow 447$  casos por grupo

```
4.3> TwoSampleSurvival.Equality(alpha=0.05,beta=0.2,lam1=0.4,lam2=0.30, k=1,  
ttotal=6, taccrual=5, gamma=0.000001)
```

```
[1] 0.2252889
```

```
[1] 0.1460279
```

```
[1] 291.4421
```

Se requieren 292 pacientes por grupo