

## Tamaño de muestra en grupos secuenciales

Felipe Peraza Garay,<sup>a,\*</sup>

<sup>a</sup>Centro de Investigación y Docencia en Ciencias de la Salud (CIDOCS), Universidad Autónoma de Sinaloa, Sinaloa, México.

Recibido Mayo 2011; aceptado Julio 2011

El cálculo del tamaño de muestra es un paso crucial en la planeación de un ensayo clínico. Y si se considera que un objetivo primordial de un ensayo clínico es el de determinar si un tratamiento beneficia o perjudica al paciente, o bien si un nuevo tratamiento es mejor o peor que otro establecido, es importante estar en la posibilidad de detener el experimento en una etapa temprana, por ejemplo, estableciendo cortes o análisis intermedios. La metodología para implementarla es en general más compleja que la de un diseño clásico e indudablemente deberá utilizarse algún paquete estadístico avanzado. En este artículo, se presenta breve una introducción a los métodos secuenciales para el caso de la comparación de dos medias. Se enfatiza en el uso del software estadístico R.

**Palabras clave:** tamaño de muestra, métodos secuenciales, R.

The sample size calculation is a crucial step in planning a clinical trial. And considering that a primary aim of a clinical trial is to determine whether a treatment benefit or harm the patient, or if a new treatment is better or worse than another set, it is important to be able to stop the experiment an early stage, for example by establishing intermediate analysis. The general methodology for their application is more complex than that of a classic design and certainly requires an advanced statistical package. In this paper we present a brief introduction to sequential methods for comparing two means. It emphasizes the use of statistical software R.

**Keywords:** sample size, sequential methods, R package.

En un clásico diseño cerrado, la duración del ensayo clínico es tan larga como el tiempo que dure el coleccionar el total de los pacientes determinados por el tamaño de muestra fijo  $n$ , calculado al inicio del estudio. Pero usualmente, los pacientes no entran al estudio de una vez, estos se van incorporando de manera gradual y se aleatorizan en el momento justo de entrar al estudio. Y si se considera que un objetivo primordial de un ensayo clínico es el de determinar si un tratamiento aporta beneficios o perjudica al paciente, o bien si un nuevo tratamiento es mejor o peor que otro establecido. Es claro la ganancia que se obtendrá para el paciente que el experimento se termine en el momento que se obtenga evidencia estadística a favor o en contra. Evidencia estadística que permita detener el experimento, significa diferencia suficientemente amplia en magnitud entre los tratamientos y nivel de significancia correcto.

Pero evidencia estadística conlleva a realizar pruebas de hipótesis antes de completar el tamaño  $n$  y al

hacerlas de manera repetida se infla el error Tipo I a niveles no aceptables. Para corregir esto se han desarrollado numerosos métodos para ajustar los niveles de significancia para mantener los niveles iniciales.<sup>1</sup>

Un tipo de métodos estadísticos que permiten este tipo de procedimientos se llaman *secuenciales* y se sugieren de cuatro a cinco análisis intermedios.<sup>2</sup>

En esta artículo de revisión analizaremos la comparación de dos tratamientos donde la variable respuesta es continua, tanto el método como la determinación del tamaño de muestra.

### 1. Método

Suponga dos tratamientos, A y B, y sean  $X_1, X_2, \dots, X_{n_1}$  e  $Y_1, Y_2, \dots, Y_{n_2}$  las respuestas numéricas continuas de dos muestras aleatorias de tamaño  $n_1$  y  $n_2$  de pacientes en el tratamiento A y B, respectivamente.

En un diseño cerrado, se determina el tamaño muestral de  $n_1$  pacientes del tratamiento A y  $n_2$  pacientes del tratamiento B, el experimento finaliza cuando se ingresan y se mide la respuesta de cada uno de los

\*Dr. Felipe Peraza Garay. **Correspondencia:** Eustaquio Buena No. 91 Col. Gabriel Leyva, Culiacán, Sinaloa, México. CP: 80030, Tel-Fax: (667) 7137978. E-mail: fperaza@uas.uasnet.mx

$n_1 + n_2$  pacientes. Entonces, se realiza la comparación estadística de los tratamientos.

En el caso en que la distribución normal es la adecuada y las varianzas de la respuesta es conocida, un estadístico  $Z$  para probar la hipótesis

$$H_0 : \mu_A = \mu_B,$$

de no diferencia entre los tratamientos es:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Para una prueba biliateral, se rechaza  $H_0$  si  $|z| > b$ . Es decir, si el valor calculado de  $Z$  cae fuera de la región de aceptación  $RA = (-b, b)$ .

El valor crítico  $b$  se determina de manera tal que la prueba tenga un nivel de significancia  $\alpha$ , especificado de antemano; en este caso  $b = z_{\alpha/2}$  el cuantil  $1 - \alpha/2$  de la distribución normal estándar.

En un diseño secuencial, se decide el número  $K$  de análisis intermedios y se determina el tamaño de muestra. En cada análisis se realiza una prueba de hipótesis, y se toma la decisión de continuar o detener el experimento. Consecuentemente se requieren  $K$  estadísticos de prueba:

Y se requieren  $K$  criterios de rechazo o regiones de aceptación  $RA_k = (-b_k, b_k)$ ; esto es, en el análisis  $k$ -ésimo se rechazará  $H_0 : \mu_A = \mu_B$  si  $|z_k| > b_k$ . El rechazo de  $H_0$  conducirá a detener el experimento, en otro caso continuará hasta el próximo análisis intermedio.

Los valores críticos  $b_k$  se eligen de manera tal que el nivel de significancia de la prueba se mantenga fijo en  $\alpha$  independientemente del número de análisis intermedios evaluados.

Un método para determinar los  $b_k$ , es utilizar una función de gasto alfa (*alpha spending function*) definida por DeMets, (1983).<sup>3</sup> Se denotan por  $\alpha(\tau)$ . Estas funciones satisfacen:  $\alpha(0) = 0$  y  $\alpha(1) = \alpha$  y

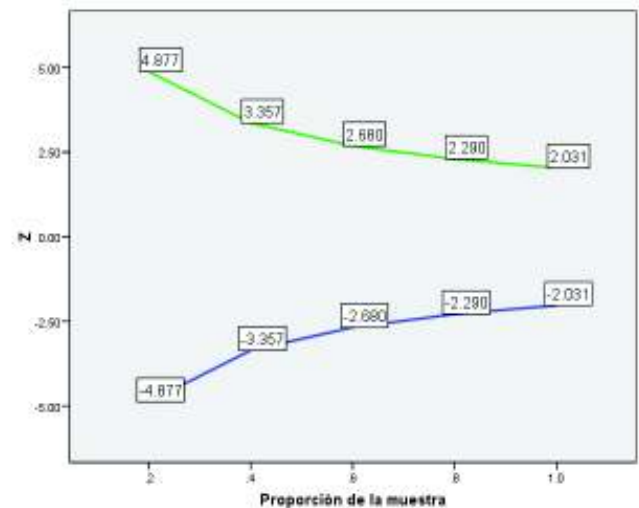
$$\Pr(|z_1| > b_1 \text{ o } |z_2| > b_2 \text{ o } \dots \text{ o } |z_k| > b_k) = \alpha(\tau).$$

La función de gasto alfa mas popular es la de O'Brien-Fleming.<sup>4</sup>

$$\alpha(\tau) = 2 - 2\Phi\left(\frac{z_{\alpha/2}}{\sqrt{\tau}}\right)$$

donde  $\tau$  representa la proporción de la muestra que se ha colectado, digamos  $\tau = \frac{k}{K}$  o la proporción de tiempo transcurrido con respecto al tiempo máximo del experimento, y  $\Phi(\cdot)$  la función de distribución normal acumulada. Un ejemplo de los valores críticos  $b_k$ , con  $\alpha = .05$ , y  $K = 5$  análisis intermedios calculado con esta función se observa en la Figura 1.

Fig. 1. Valores críticos calculados con la función de gasto de O'Brien-Fleming



Existen otras funciones para calcular los  $b_k$  como la función de gasto Pocock:

$$\alpha(\tau) = \alpha \ln(1 + (e - 1)\tau)$$

La función de gasto de O'Brien-Fleming es la más utilizada en ensayos clínicos, es conservadora en etapas tempranas del ensayo, al requerir una gran diferencia entre las medias para ser rechazada la hipótesis  $H_0$ .

Utilizando la notación de Jennison and Turnbull,<sup>5</sup> la prueba de O'Brien-Fleming se plantea de la siguiente manera: para el análisis intermedio  $k$ , se rechaza,

$$H_0 : \mu_1 = \mu_2, \text{ si } |Z_k| \geq C_B(K, \alpha) \sqrt{\frac{K}{k}}$$

donde

$$Z_k = \frac{\bar{X}_k - \bar{Y}_k}{\sqrt{\frac{\sigma_1^2}{n_{1,k}} + \frac{\sigma_2^2}{n_{2,k}}}}, \quad k = 1, 2, \dots, K$$

es una variable aleatoria con distribución normal estándar,  $K$  es el tamaño de muestra máximo,  $\tau = \frac{k}{K}$

es la proporción de la muestra, la constante  $C_B(K, \alpha)$  depende del tamaño de muestra máximo  $K$  y del nivel de significancia  $\alpha$ . Las varianzas  $\sigma_1^2, \sigma_2^2$  usualmente son desconocidas y se reemplazan por sus estimadores con los datos acumulados.<sup>2</sup> No existen fórmulas cerradas para calcular  $C_B(K, \alpha)$  pero se puede consultar las tablas en varios libros.<sup>1,2,4,5</sup>

Existe software comercial, por ejemplo SAS o S-Plus con módulos que cubren todos los aspectos de los diseños secuenciales, y otros que lo hacen de manera parcial como NCSS. Otra opción es utilizar "R" (<http://www.r-project.org/>), el cual es un software de uso libre, con numerosos procedimientos (paquetes) libres de bajar en internet para realizar análisis estadísticos. Consultar Mazumdar M and Bang H (2008) para una amplia referencia de software sobre este tema.<sup>1</sup>

En este trabajo se utilizó "R", y el paquete "ldbounds" descargado de la página <http://cran.r-project.org/web/packages/ldbounds/index.htm> y desarrollado por T. Charles Casper y Oscar A. Perez, del cual se ejecutaron la función "bounds" y "drift", como se describe en el manual.<sup>6</sup>

### Usando *ldbounds*

Mostraremos una aplicación de *ldbounds* a manera de ejemplo. Defina una sucesión de números que indiquen los momentos en que se realizará cada análisis intermedio, en proporciones de la unidad. Por ejemplo, si planea 5 análisis intermedios en cada 20% del tamaño de la muestra, la selección deberá ser: 0.2, 0.4, 0.6, 0.8, 1. No es necesario que sean de la misma longitud, por ejemplo: 0.2, 0.3, 0.5, 0.7, 1 es una elección válida. Ahora elija un nivel de significancia alfa, por ejemplo  $\alpha = 0.05$ . Sustituya en los primeros dos renglones del siguiente código, y ejecute dentro de R:

```
time<-{0.2,0.4,0.6,0.8,1}
alfa<-.05
bk<-bounds(time, iuse = c(1, 1), alpha = c(alfa/2,
alfa/2))
summary(bk)
```

La salida se muestra en el Cuadro 1; las primeras dos columnas muestran el número de análisis intermedios y la proporción de la muestra y la columna 3 y 4 los valores críticos  $-b_k$  y  $b_k$ .

**Cuadro 1.** Salida de *ldbounds*

Time	Lower	Upper	Exit pr.	Diff. pr.
1 0.2	-4.8769	4.8769	1.0777e-06	1.0777e-06
2 0.4	-3.3569	3.3569	7.8830e-04	7.8723e-04
3 0.6	-2.6803	2.6803	7.6161e-03	6.8278e-03
4 0.8	-2.2898	2.2898	2.4424e-02	1.6807e-02
5 1.0	-2.0310	2.0310	5.0000e-02	2.5576e-02

En el Cuadro 2 y 3, se muestran los valores críticos para niveles de significancia  $a = .05$  y diferente número de análisis intermedios.

## 2. Tamaño de muestra

Para calcular el tamaño de la muestra, además del nivel de significancia  $\alpha$  y el número de análisis intermedios  $K$ , se debe considerar la diferencia mínima  $\delta$ , y la potencia  $1 - \beta$ .

Siguiendo el método descrito en Jennison and Turnbull (2000) y en Julius (2010) primero se calcula el tamaño de muestra utilizando la fórmula para un diseño fijo,

$$n_{fijo} = \frac{(z_{\alpha/2} + z_{\beta})^2 (\sigma_1^2 + \sigma_2^2)}{\delta^2}$$

y luego se multiplica por una constante  $R_B(K, \alpha, \beta)$ , tabulada; para obtener el tamaño de muestra para este diseño:

$$n = n_{fijo} R_B(K, \alpha, \beta).$$

Finalmente, el resultado de dividir  $n/K$ , proporciona el tamaño de muestra en cada uno de los análisis intermedios.

### Usando *ldbounds*

Continuando con la aplicación en R, usaremos la función "drift" de *ldbounds*, drift estima el parámetro  $\theta$ , la deriva del proceso estocástico (movimiento Browniano) asociado al método. La deriva está relacionada a  $Z_k$  mediante:

$$\theta = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

En el Cuadro 2 y 3, se muestran las derivas para  $a = .05, 1 - \beta = .8, .9$  y diferente número de análisis intermedios, y se calcularon ejecutando:

```
time<- seq(1/5, 1, length = 5)
```

```
obf.bd<-bounds(time, iuse = c(1, 1), alpha =
c(0.025, 0.025))
drift.dr<-drift(obf.bd$lower.bounds,
obf.bd$upper.bounds, time,pow = 0.8)
summary(drift.dr)
```

Para determinar el tamaño de muestra elija el nivel de significancia  $\alpha$ , la potencia  $1-\beta$ , la mínima diferencia clínica  $\delta$ , y el número de total de análisis intermedios  $K$  que aquí supondremos igualmente espaciados. En el Cuadro 2 o 3, seleccione la deriva  $\theta$ . Entonces, el tamaño de muestra máximo por grupo con esa selección de parámetros, está dado por:

$$n = \theta^2 \frac{(\sigma_1^2 + \sigma_2^2)}{\delta^2}.$$

Entonces,  $n/K$ , es el tamaño de muestra en cada análisis intermedio.

Para grupos desbalanceados, con tamaño muestra  $n_1$  para el Grupo 1 y  $n_2 = rn_1$ , para el Grupo 2, use:

$$n_1 = \theta^2 \frac{(\sigma_1^2 + \sigma_2^2/r)}{\delta^2} \text{ y } n_2 = rn_1.$$

Es importante aclarar que no se requiere que los análisis sean igualmente espaciados.

**Una aplicación.** Se quiere determinar el tamaño de muestra para comparar de manera bilateral un tratamiento A y un tratamiento B. Se asume una desviación estándar  $\sigma = 8.1$ , diferencia  $\delta = 5$  y potencia de 0.8. Se planean  $K = 3$  análisis intermedios.

*Procedimiento:*

En el Cuadro 3, en  $K = 3$ , seleccione la deriva  $\theta = 2.8195$ . Sustituya los valores para determinar el tamaño máximo:

$$n_{\max} = \theta^2 \frac{(\sigma_1^2 + \sigma_2^2)}{\delta^2} = 2.8195^2 \frac{8.1^2 + 8.1^2}{5^2} = 41.726.$$

Así, el número máximo de pacientes es de 42 por grupo.

Como requiere tres análisis intermedios, el tamaño de muestra para el primer análisis es de:

$$n = n_{\max}/K = 42/3 = 14$$

pacientes por grupo.

Una vez completado los  $n_1 = n_2 = 14$  pacientes, calcule, la media para cada grupo,  $\bar{X}$ ,  $\bar{Y}$ , y sustituya:

$$Z_1 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

En el Cuadro 3, en  $K = 3$ , seleccione el valor crítico  $b_1 = 3.7103$ ; si el valor calculado de  $Z_1$  es mayor que  $b_1$  o menor que  $-b_1$  rechaze  $H_0$  y detenga el experimento.

Si no rechazó  $H_0$ , continúe el ensayo hasta que colecte  $n_1 = n_2 = 28$ , calcule la media para cada grupo,  $\bar{X}$ ,  $\bar{Y}$ , y sustituya:

$$Z_2 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Nuevamente en el Cuadro 3, en  $K = 3$ , seleccione el valor crítico  $b_1 = 2.5114$ ; si el valor calculado de  $Z_1$  es mayor que  $b_1$  o menor que  $-b_1$  rechaze  $H_0$  y detenga el experimento.

Si no rechaza  $H_0$ , continúe el ensayo hasta que colecte el número máximo de pacientes  $n_1 = n_2 = 42$ , calcule, la media para cada grupo,  $\bar{X}$ ,  $\bar{Y}$ , y sustituya:

$$Z_3 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Nuevamente en el Cuadro 3, en  $K = 3$ , seleccione el valor crítico  $b_1 = 1.9930$ ; si el valor calculado de  $Z_1$  es mayor que  $b_1$  o menor que  $-b_1$  rechaze  $H_0$ , en otro caso, concluya que no existe diferencia significativa entre los tratamientos.

### 3. Conclusión

En este trabajo se muestra solo el caso de comparación de medias y varianzas conocidas, en el cual se pueden realizar las generalizaciones usuales en pruebas de hipótesis. Existen pruebas para muestras pequeñas utilizado la prueba  $t$ , para proporciones, análisis de sobrevivencia y otras. Se sugiere consultar las referencias.

La complejidad de los nuevos métodos estadísticos obliga a utilizar software en ocasiones poco amigable para el investigador, tanto el software comercial pero sobretodo el libre. Pero, este último, con el uso cada vez más amplio de la comunicación vía internet, permite aplicarlos antes que aparezca una versión comercial.

**Cuadro 2.** Valores críticos y deriva.  $\alpha = .05, \beta = .9$ 

	k	Fracción	Inf	Sup	Deriva
K=2	1	0.5	-2.9626	2.9626	3.247271
	2	1.0	-1.9686	1.9686	
K=3	1	0.33	-3.7103	3.7103	3.260754
	2	0.66	-2.5114	2.5114	
	3	1.00	-1.9930	1.9930	
K=4	1	0.25	-4.3326	4.3326	3.271059
	2	0.50	-2.9631	2.9631	
	3	0.75	-2.3590	2.3590	
	4	1.00	-2.0141	2.0141	
K=5	1	0.2	-4.8769	4.8769	3.278782
	2	0.4	-3.3569	3.3569	
	3	0.6	-2.6803	2.6803	
	4	0.8	-2.2898	2.2898	
	5	1.0	-2.0310	2.0310	
K=6	1	0.16	-5.3666	5.3666	3.284674
	2	0.33	-3.7106	3.7106	
	3	0.50	-2.9697	2.9697	
	4	0.66	-2.5386	2.5386	
	5	0.83	-2.2522	2.2522	
	6	1.00	-2.0447	2.0447	

**Referencias**

1. Mazumdar M, Bang H. Sequential and Group Sequential Designs in Clinical Trials: Guidelines for Practitioners. Handbook of Statistics, Elsevier 2008: Vol. 27: pp 491-512.
2. Chow SC, Shao J, Wang H. . Sample Size Calculations in Clinical Research. Marcel Dekker. New York; 2003.
3. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983; Vol. 70, 659-663.
4. Julious SA. Sample sizes for clinical trials. Taylor and Francis Group; 2010.
5. Jennison C, Turnbull B. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CR; 2000.
6. Casper TC, Perez OA. An R Package for Group Sequential Boundaries Using Alpha Spending Functions. <http://cran.r-project.org/web/packages/lid-bounds/index.html>; 2006.

**Cuadro 3.** Valores críticos y deriva.  $\alpha = .05, \beta = .8$ 

	k	Fracción	Inf	Sup	Deriva
K=2	1	0.5	-2.9626	2.9626	2.806940
	2	1.0	-1.9686	1.9686	
K=3	1	0.33	-3.7103	3.7103	2.8195
	2	0.66	-2.5114	2.5114	
	3	1.00	-1.9930	1.9930	
K=4	1	0.25	-4.3326	4.3326	2.829
	2	0.50	-2.9631	2.9631	
	3	0.75	-2.3590	2.3590	
	4	1.00	-2.0141	2.0141	
K=5	1	0.2	-4.8769	4.8769	2.836052
	2	0.4	-3.3569	3.3569	
	3	0.6	-2.6803	2.6803	
	4	0.8	-2.2898	2.2898	
	5	1.0	-2.0310	2.0310	
K=6	1	0.16	-5.3666	5.3666	2.841409
	2	0.33	-3.7106	3.7106	
	3	0.50	-2.9697	2.9697	
	4	0.66	-2.5386	2.5386	
	5	0.83	-2.2522	2.2522	
	6	1.00	-2.0447	2.0447	