

# El tamaño de la muestra cuando hay más de un grupo

En el tema anterior vimos cómo estimar el tamaño de muestra mínimo que nos permite llevar a cabo una investigación, con un poder adecuado, en el caso que ésta se produzca con un grupo único (p. ej., estimar la glucemia basal media o la prevalencia de individuos con antecedentes de episodio cardiovascular de la población de referencia de nuestra farmacia).

Este tipo de investigaciones tiene lugar en las fases preliminares de muchas líneas de investigación, pero con mayor frecuencia nos hallaremos ante un estudio que precise dos grupos de pacientes. En este caso, el cálculo del tamaño de la muestra se efectúa de una manera algo diferente.

## Investigaciones con dos grupos

La implementación de una intervención farmacéutica tiene como objetivo influir en la variable resultado, con el fin de mejorar su valor, sea éste el que sea. Hay que tener en cuenta que en este caso el mero cambio de la variable no implica que la exposición del paciente a esta intervención haya sido la causa de la variación final del resultado. Hay diversas causas que pueden provocarlo, de modo que será imprescindible contar con un grupo de idénticas características que el de intervención, el grupo control, donde la única diferencia sea la exposición a esta intervención farmacéutica.

El objetivo es obvio: analizar posteriormente si hay una variación estadísticamente significativa entre la variable resultado del grupo de intervención y la correspondiente al grupo control.

Del mismo modo que en el tema anterior, la estimación del tamaño de la muestra se analizará en función del tipo de variable, es decir, que se trate de una media aritmética o de un porcentaje.

## El resultado se expresa como una media

Es un caso muy habitual en la investigación sobre resultados de la intervención farmacéutica implementada. Pensemos que se desea iniciar una determinada acción farmacéutica en pacientes diabéticos tipo 2, después de detectar que la glucemia media de los pacientes de nuestra farmacia es elevada. Esta acción, obviamente, se evaluará para estimar su efectividad. Para ello, se procederá inicialmente a una prueba piloto. Al concluir esta prueba, las reducciones en la glucosa basal de cada paciente tratado son las que se muestran en la tabla 1.

Con Excel se calculará el valor de la media aritmética ( $m$ ) y la desviación estándar ( $\sigma$ ) de los grupos control (c) e intervención (i):

- $m_c = 21,25$  mg/dl.
- $m_i = 25,86$  mg/dl.
- $\sigma_c = 10,89$  mg/dl.
- $\sigma_i = 9,81$  mg/dl.

La diferencia entre las dos variables resultado ( $m_i$  y  $m_c$ ) es muy probable que no sea significativa por su exiguo tamaño; no obstante, a partir de esta diferencia se puede estimar el tamaño mínimo de cada grupo, ya

**Tabla 1. Reducciones individuales de la glucosa basal correspondientes a cada paciente donde se ha implementado, como prueba piloto, la intervención farmacéutica que se quiere evaluar**

ID PAC	GB-GC	ID PAC	GB-GC	ID PAC	GB-GI	ID-PAC	GB-GI
1	25	5	32	1	26	5	33
2	15	6	18	2	29	6	28
3	19	7	11	3	39	7	11
4	9	8	41	4	15		

GB: glucosa basal; GC: grupo control; GI: grupo de intervención; ID pac: número de identificación del paciente.

que aquélla deberá ser, al menos, dos veces superior al error estándar, de modo que:

$$(m_i - m_c) = 2 \cdot \sqrt{[(\sigma_i^2 + \sigma_c^2)/n]}$$

de donde se despeja n:

$$n = [4 \cdot (\sigma_i^2 + \sigma_c^2)] / (m_i - m_c)^2$$

de donde se obtiene el valor de 41 sujetos en cada grupo.

Es posible que, dadas las características de nuestra farmacia, no se pueda obtener el número de individuos calculado para el grupo de intervención, aunque sí para los del grupo control. En ese caso, la estimación del tamaño de cada grupo se atenderá a los requerimientos que hagamos previamente. Supongamos, entonces, que nuestras estimaciones indiquen que el grupo de intervención tendría sólo una parte de personas (a) del grupo control, por ejemplo, la mitad (a: 1/2). En este caso  $n_c/a = n_i$  ( $n_c/2$ ), por lo que  $n_c = a \cdot n_i$  ( $:2 \cdot n_i$ ). Así, la diferencia entre las variables resultados de cada grupo sería de:

$$(m_i - m_c) = 2 \cdot \sqrt{[(\sigma_c^2/a \cdot n_i) + (\sigma_i^2/n_i)]}$$

de donde n:

$$n_i = [4 \cdot (a \cdot \sigma_i^2 + \sigma_c^2)] / [a \cdot (m_i - m_c)^2]$$

de modo que el tamaño de la muestra es de 30 personas para el grupo de intervención y de 60 ( $:30 \cdot 2$ ) para el control.

### El resultado se expresa como un porcentaje

Es un caso algo diferente que el anterior, si bien el razonamiento es similar. Para comprenderlo mejor, supongamos que después de una prueba piloto de una intervención dirigida a controlar el nivel de A1C en diabéticos tipo 2 se han obtenido los datos de la tabla 2, donde constan los valores de A1C obtenidos.

Los datos que se pueden obtener a partir de los valores hallados son los siguientes:

#### Grupo control

- Número de pacientes totales: 15.
- Número de pacientes controlados: 5.
- Porcentaje de pacientes controlados: 0,33 ( $:5/15$ ).

**Tabla 2. Valores de la A1C (%) obtenidos en la prueba piloto, en los grupos control y de intervención\***

GRUPO CONTROL			GRUPO DE INTERVENCIÓN		
ID PAC	A1C (%)	CONTROLADO	ID PAC	A1C (%)	CONTROLADO
1	7,3	0	1	6,8	1
2	7,9	0	2	7,6	0
3	7,4	0	3	8,1	0
4	6,8	1	4	6,9	1
5	6,7	1	5	6,4	1
6	8,3	0	6	7,6	0
7	9,2	0	7	7,9	0
8	7,5	0	8	6,2	1
9	6,8	1	9	6,4	1
10	8,2	0	10	7,6	0
11	6,4	1	11	7,1	0
12	7,2	0	12	6,8	1
13	9,2	0			
14	7,3	0			
15	6,2	1			

ID pac: número de identificación del paciente.

\*En la columna «Controlado», 1 es el código dado a «sí» y 0 el dado a «no».

#### Grupo de intervención

- Número de pacientes totales: 12.
- Número de pacientes controlados: 6.
- Porcentaje de pacientes controlados: 0,50 ( $:6/12$ ).

En el caso de que el valor de la variable resultado sea una media, la diferencia entre ambos grupos debe ser al menos el doble que el error estándar para que aquélla sea estadísticamente significativa. De esta forma, ha de cumplirse que:

$$(m_i - m_c) = \sqrt{[p \cdot (1-p) \cdot (1/n_i + 1/n_c)]}$$

de modo que n:

$$n = (8p - 8p^2) / (p_i - p_c)^2$$

siendo p el porcentaje global de resultados (tomando en consideración los dos grupos simultáneamente):

$$p = (x_i + x_c) / (n_i + n_c)$$

donde  $x_i$  y  $x_c$  son el número de casos positivos de cada grupo (intervención y control) en el ejemplo del

tema, el número de pacientes que han alcanzado un grado de control en función de su porcentaje de A1C y  $n_c$  y  $n_i$  se corresponden con el tamaño de cada grupo en la prueba piloto. Así,  $p$  toma el valor de:

$$p = (6+5)/(15+12) = 0,41$$

con lo que  $n$ :

$n = (8 \cdot 0,41 - 8 \cdot 0,41^2) / (0,50 - 0,33)^2$   
de lo que se deduce que  $n = 70$  pacientes en cada grupo de estudio.

Como en el apartado anterior, puede ocurrir que dispongamos de menor número de pacientes disponibles para el grupo de intervención que para el control (o viceversa). En tal caso, el tamaño de la muestra se calculará para cada grupo separadamente.

Entonces, si se desea que  $n_c = n_i \cdot a$ , el cálculo se efectuará de la forma siguiente: si el grupo de intervención se estima que será la mitad del de control ( $n_i = n_c / 2$ ;  $n_c = n_i \cdot 2$ ), el tamaño de cada grupo se calcula mediante:

$$n_i = [(1+a) \cdot (4p - 4p^2)] / [a \cdot (p_i - p)^2]$$
$$n_i = [(1+2) \cdot (4 \cdot 0,41 - 4 \cdot 0,41^2)] / [2 \cdot (0,50 - 0,33)^2]$$

de modo que  $n_i = 53$ . Así, el valor de  $n_c$  será de 106 pacientes ( $: 53 \cdot 2$ ). ■