

Tamaño de la muestra

Dra. Ana Berta Irineo Cabrales¹, Dr. Carlos A. Zambada Senties²

^{1,2}Grupo de Maestría y Doctorado en Ciencias Médicas

¿PORQUÉ NECESITAMOS CALCULAR EL TAMAÑO DE LA MUESTRA?

Para no cometer el Error Tipo II, al concluir que no hay diferencia entre los grupos, cuando en realidad si la hay. Esto se debe a que elegimos un pequeño tamaño de muestra. .

Error Tipo II: aceptar H_0 cuando H_0 es falsa (cuando debemos rechazarla-decir erróneamente que no hay diferencia, cuando en realidad si la hay).

También necesitamos conocer qué es el Error Tipo 1: rechazar H_0 cuando H_0 es verdadera (cuando debemos aceptada, decir erróneamente que si hay diferencia, cuando en realidad no la hay).

Definiciones y símbolos para el cálculo del tamaño muestral:

alfa: es el valor crítico de P, es el nivel de significancia. Es el valor crítico que escogimos para la probabilidad de que la H_0 sea verdadera. Si el valor P de nuestro ensayo es menor de alfa nosotros debemos rechazar H_0 y a la inversa, si $P > \alpha$ no debemos rechazarla. Por ejemplo, si nosotros establecimos un valor de alfa de 0.05, estamos diciendo que estamos dispuestos a aceptar un 5% de probabilidad de rechazar incorrectamente H_0 , es decir que no hay diferencia entre los grupos. Cuando decimos que hubo una asociación estadísticamente significativa nos referimos a que hay menos de 5% de probabilidades de estar cometiendo el Error Tipo 1.

beta: a la capacidad de una prueba de identificar correctamente que hay una diferencia entre los grupos en un ensayo se le denomina “poder” de la prueba. En otras palabras, es la capacidad de una prueba de rechazar hipótesis nula cuando efectivamente se debe rechazar.

Cuando en el estudio no se encuentra asociación (diferencia) y en verdad existe (celda c) se está cometiendo el Error Tipo II; este se representa con el valor de $\beta = 0.20$; esto significa que el investigador está dispuesto a aceptar 1 de 5 posibilidades de decir que no hay diferencia en las

intervenciones cuando en realidad si la hay. Su contraparte 1-beta ($1-0.20=0.80$ (80%)) es lo que se conoce como el poder de la muestra, es la sensibilidad de un estudio para probar una diferencia. El tipo de errores está en relación directa con el tamaño de la muestra; un tamaño de muestra muy grande (que nunca sucede) puede provocar que se presente el Error Tipo 1 y un tamaño de muestra pequeño (que casi siempre es el caso) puede provocar el Error Tipo II.

delta: es la magnitud de la diferencia entre los dos grupos que es clínicamente significativa. Es la magnitud de las diferencias que uno pretende probar. Se conoce también como el “tamaño del efecto”. En la práctica, la determinación del tamaño del efecto es a menudo arbitraria. Se obtiene de: $p2-p1$.

En ensayos clínicos controlados, si se mantienen constantes el poder y el nivel de significancia, el tamaño de la muestra (n) aumenta:

1-Si disminuye el tamaño del efecto (delta): a menor delta mayor tamaño muestral y a la inversa, a mayor delta menor tamaño muestral.

2-Si aumenta el porcentaje de fracasos en ambos tratamientos sin variaciones en el delta porque aumenta la p testada.

3-Si aumenta el poder) por ejemplo de 80% a 90% y/o el nivel de confianza de alfa, de 0.05 a 0.01.

4-Si se utiliza el valor de alfa media ($Z_{\alpha/2}$) (prueba de hipótesis de dos colas: $X1 \neq X2$) en lugar del valor de alfa (Z_{α}) (prueba de hipótesis de una cola: $X1 > X2$).

n : en un estudio de cohortes, el número de individuos expuestos y el mínimo de los no expuestos; en un estudio de casos y controles el número de casos y el mínimo de los controles; en un ensayo clínico controlado, el número de individuos en cada grupo, en el grupo de tratamiento control (supuestamente el peor tratamiento) y en el

grupo de tratamiento experimental (supuestamente el mejor tratamiento).

r: en un estudio de cohortes, la razón (división) entre el número de individuos no expuestos y expuestos; en un estudio de casos y controles, la razón entre el número de controles por caso; en un ensayo clínico controlado, la razón entre el número de individuos en el tratamiento experimental (mejor) y el tratamiento control (peor).

p2 : en un estudio de cohortes se refiere a la proporción de individuos expuestos que desarrollaron la enfermedad (celda a); en un estudio de casos y controles, la proporción de casos que estuvieron expuestos (celda a); en un ensayo clínico controlado, la proporción de individuos en el tratamiento habitual (peor) que no mejoraron (celda a).

p1 : en un estudio de cohortes, la proporción de individuos no expuestos que desarrollaron la enfermedad (celda c); en un estudio de casos y controles, la proporción de controles que estuvieron expuestos (celda b); en un ensayo clínico controlado, la proporción de individuos en el tratamiento experimental (mejor) que no mejoraron (celda c).

p testada: $p2+rp \frac{1}{1+r}$ = promedio ponderado de p2 y p1.

ENSAYO CLÍNICO CONTROLADO

La fórmula que se utiliza para una diferencia entre proporciones en los ensayos clínicos controlados, cohortes y casos y controles es la misma y es la siguiente:

$$n = \frac{(Za/2+Zb)^2 \times ptestada(1-ptestada)(r+1)}{(d)^2 r}$$

Los valores de Zb y Za/2 se refieren a valores críticos en la tabla de distribución normal.

Si el valor asignado a beta es 0.20, el valor en la tabla de distribución normal de Zb=0.84 (ver tabla Z)

El valor de Za/2 de un alfa de 0.05 es $\frac{0.05}{2} = 0.025 = 1.96$ (ver tabla Z)

Supongamos que se desea conocer la eficiencia del tratamiento A (experimental, supuestamente mejor) y lo vamos a comparar con el tratamiento B (habitual, estándar, supuestamente peor).

También supongamos que en un estudio previo se observó que el 70% de los paciente_ con el tratamiento A presentaron mejoría y el 50% mejoraron con el tratamiento B.

Cuadro No. 1

La mejor estrategia es construir una tabla de 2 x 2 y considerar 100 pacientes en el tratamiento habitual (B) y 100 pacientes en el tratamiento experimental (A)

TRATAMIENTO	NO MEJORÍA	MEJORÍA	TOTAL
TRATAMIENTO B HABITUAL ESTANDAR PEOR	p ² a 50% (0.50)	b 50% (0.50)	100
TRATAMIENTO A EXPERIMENTAL MEJOR	c p ¹ 30% (0.30)	d 70% (0.70)	100

Se establecen los valores de alfa=0.05 y de beta = 0.20

Si se considera que el 70% mejora con el tratamiento (A) experimental (mejor) y 50% con el tratamiento (B) habitual (peor), entonces:

p1 = es la frecuencia de no mejoría en el mejor tratamiento (A) (celda c) que es 1-0.7=0.3 (30%), esto es, si el 70% mejoró, el 30% no mejoró. p1=0.3

p2= es la frecuencia de no mejoría en el peor tratamiento (B) (celda a) que es 1-0.5=0.5 (50%), esto es, si el 50% mejoró, el 50% no mejoró. p2=0.5

Se procede a calcular p testada=

$$\frac{(p2+rp1)}{1+r} = \frac{(0.5)+1(0.3)}{1+1} = \frac{0.8}{2} = 0.4$$

Ptestada=0.4

delta=p2-p1 = 0.5-0.3=0.2

Se sustituye la fórmula

$$n = \frac{(7.849)^2 \times 0.4 \times (1-0.4) \times (1+1)}{(0.2)^2 \times 1} = \frac{3.76752}{0.04} = 94.18$$

n=94

Se necesitan 94 pacientes en cada grupo.

COHORTES

Ejemplo de cohortes

Se desea conocer si la exposición solar es un factor de riesgo para desarrollar melanoma

Cuadro No. 2

		MELANOMA		TOTAL
		SI	NO	
EXPOSICION SOLAR	SI	p ² a 62% (0.62)	b 38% (0.38)	100
	NO	c p ¹ 46% (0.46)	d 54% (0.54)	100

Se construye una tabla de 2 x 2 y se consideran 100 pacientes con exposición solar (definiendo previamente a qué se le llama exposición solar) y 100 pacientes sin exposición solar. Se siguen en el tiempo (de causa a efecto) y se registran tanto los que desarrollen como los que no desarrollen melanoma.

alfa=0.05 beta=0.20

p1=proporción de individuos SIN exposición solar que desarrollaron melanoma (0.18) (celda d).

p2=proporción de individuos CON exposición solar que desarrollaron melanoma (0.62)

$$p_{testada} = \frac{p_2 + r p_1}{1+r} = \frac{0.62 + 0.46}{2} = 0.54$$

$$\Delta = p_2 - p_1 = 0.62 - 0.46 = 0.16$$

$$n = \frac{7.849 (0.54) (0.46) (2)}{0.0256} = 152$$

n= 152 individuos expuestos y 152 no expuestos

CASOS Y CONTROLES

En un diseño de casos y controles:

p1 significa la proporción de CONTROLES (niños sanos, SIN leucemia) que estuvieron expuestos al factor de riesgo (tabaquismo paterno) (celda b)

p2 significa la proporción de CASOS (niños con leucemia) que estuvieron expuestos al factor de riesgo (tabaquismo paterno) (celda a).

Ejemplo

Se desea conocer si el tabaquismo paterno es un factor de riesgo para desarrollar leucemia en los hijos.

Supongamos que existe en la literatura una serie de casos que reporta que el 30% de los niños con leucemia tiene padres fumadores, entonces el valor de la celda "a" es de 0.30, esto es, porcentaje de niños con leucemia con padres fumadores (p2). Se consideran 100 niños con leucemia, por lo tanto son 70 en la celda "c".

Ahora, niños sin leucemia con padres fumadores: en una encuesta nacional se puede encontrar el porcentaje de hijos con padres fumadores en la población general, de esa cifra se puede obtener el valor de la celda b (p1), es decir, en los niños sin leucemia el 18% de los padres fuman; el valor de la celda d es 100-18=82 (0.82).

Se establecen los valores de alfa=0.05 y de beta=0.20

Se construye una tabla de 2 x 2

Cuadro no.3

		CASOS LEUCEMIA CONTROLES	
		SI	NO
TABAQUISMO PATERNO	SI	p2 30% (0.30)	b p1 18% (0.18)
	NO	c 70% (0.70)	d 82% (0.82)
TOTAL		100	100

Delta=p2-p1=0.30-0.18=0.12

$$p_{testada} = \frac{p_2 + r p_1}{1+r} = \frac{0.3 + 0.18}{2} = 0.24$$

n=199 casos y 199 controles como mínimo

$$n = \frac{7.849 (0.24) (0.76) (2)}{0.0144} = 198.84$$

Ejemplo no. 2

Se desea conocer si la infección cervicovaginal es un factor de riesgo de parto prematuro

Un estudio (Cota G y cols Rev Med IMSS 2001;39:289-293) reporta que de 48 pacientes con parto prematuro, 29 tuvieron infección cervicovaginal versus 14 de 48 pacientes de parto a término.

Calcule el tamaño de la muestra:

Ejemplo no. 3

		Parto prematuro casos	Parto término controles
INFECCIÓN CERVICO-VAGINAL	+	p2 29 60.4%	p1 14 29%
	-		

$$p = \frac{0.604 + 0.29}{2} = 0.447$$

$$\Delta = 0.604 - 0.29 = 0.314$$

$$d^2 = 0.09859$$

$$n = \frac{7.84 (0.447) (0.553) (2)}{1.09859} = 39$$

n= 39 casos Parto prematuro
39 controles Parto término

Se desea conocer si el uso de antibióticos (en infección de vías urinarias o de tracto respiratorio superior) es un factor de riesgo para candidiasis vulvovaginal sintomática. Existe en la literatura (Spinillo A et al Am J Obstet Gynecol 1999;180:14-17) un estudio con los siguientes resultados: la prevalencia del uso de antibióticos fue de 19.2% en pacientes con candidiasis comparado con 11.9% en pacientes sin candidiasis alfa= 0.05 beta= 0.20

		CANDIDIASIS CASOS	NO CANDIDIASIS CONTROLES
USO ANTIBIOTICO	+	P2 19.2%	P1 11.9%
	-		
		100	100

$$P = \frac{0.192 + 0.119}{2} = 0.1555 \quad d = 0.192 - 0.119 = 0.073$$

$$n = \frac{7.84 (0.1555) (0.8445) (2) + 2.059}{(0.073)^2} = \frac{2.059}{0.005329} = 38\%$$

n = 386 casos 386 controles

ESTUDIOS TRANSVERSALES (ENCUESTA PARA PREVALENCIA)

Estudios descriptivos:

En el estudio descriptivo tratamos de obtener una estimación de la proporción de la población que posee o presenta un determinado fenómeno o atributo de salud.

En los estudios descriptivos no se trata de probar una hipótesis. Sólo la precisión que deseamos obtener decidirá cuál ha de ser el tamaño muestral. En esta situación, para determinar el tamaño de la muestra se utiliza un método basado en el concepto de los intervalos de confianza.

El investigador ha de especificar un nivel de confianza para el intervalo. El nivel de confianza es la probabilidad de que el intervalo de confianza contenga a la verdadera proporción en la población.

Ejemplo no. 1

Se desea conocer la prevalencia de isquemia miocárdica silente (sin datos clínicos de la enfermedad) en pacientes con diabetes tipo 2.

La formula es:

$$n = \frac{(Z_{\alpha/2})^2 (p (1-p))}{d^2}$$

Z_{α/2}=1.96 si el valor de alfa=0.05, alfa media=0.025, se busca en la tabla Z el punto de 0.025, que es 1.96

p=es la prevalencia “esperada” (ya conocida en algunos estudios) de isquemia miocárdica silente en pacientes con diabetes mellitus tipo 2, que se ha reportado de 17% (referencia) diagnosticada con gamagrafía con ejercicio con talio.

d= es la distancia (o tolerancia): proximidad que se desea obtener en la estimación con respecto a la proporción que interesa, por ejemplo, dentro de 0.05. Es la diferencia entre el valor “esperado”

(prevalencia conocida) y el valor aceptable más .bajo. Supongamos que el investigador (cardiólogo) “acepta” arbitrariamente un 3% por arriba y por debajo de la prevalencia conocida de 17%, por lo tanto, la prevalencia más baja “aceptable” es 14%

d= prevalencia conocida (0.17) - prevalencia más baja aceptable (0.14): d=0.17-0.14=0.03

La prevalencia en el estudio podría variar, para ser confiable, de 14% a 20%.

Se sustituye la fórmula

$$n = \frac{(1.96)^2 (0.17 (1-0.17))}{(0.03)^2} = \frac{0.542049}{0.0009} = 602 \text{ pacientes diabéticos tipo 2}$$

Si sólo tenemos presupuesto para estudiar a 200 pacientes, se aplica el factor de corrección:

$$N = \frac{N1}{1 + \frac{N1}{n}} = \frac{602}{1 + \frac{602}{200}} = \frac{602}{4.01} = 150 \text{ pacientes diabéticos}$$

poblacion existente 200

Es condición que de los 200 pacientes existentes (disponibles) se deben elegir los 150 pacientes calculados al azar.

Ejemplo no.2

Se diseñó un estudio con objeto de estimar la incidencia acumulativa de embarazos ectópicos en EEUU en 1978, para poder compararla posteriormente con la obtenida en 1970.

Se examinó una muestra de expedientes clínicos. En 1970, la incidencia acumulativa de embarazos ectópicos fue de 4.5 x 1000 embarazos (0.0045) (.45%). Los investigadores deseaban alcanzar un nivel de confianza de 0.95 de que su cálculo estaba dentro de 0.0005 del valor real. El cálculo se basó en base a la siguiente información:

d=0.005

Z_{α/2}=1.96

p=0.0045, se usó la proporción calculada en 1970 como tentativa para la proporción de 1978.

n=68838

$$n = \frac{(Z_{\alpha/2})^2 (p (1-p))}{d^2} = \frac{3.8416 (0.0045) (0.9955)}{(0.0005)^2} = \frac{0.017209407}{0.0000025} = 68837.6$$

Si los investigadores hubieran decidido situarse dentro de 0.001 de la proporción real en lugar de 0.0005, el tamaño de la muestra hubiera sido:

n=17209

$$n = \frac{3.8416 (0.0045) (0.9955)}{(0.001)^2} = \frac{0.017209407}{0.000001} = 17209$$

n=17209

CÁLCULO DEL TAMAÑO DE LA MUESTRA PARA UN DISEÑO DE PRUEBAS DIAGNÓSTICAS

Ejemplo:

Se desea evaluar el uso de los niveles de complemento (CLs) para tamizaje de carcinoma hepatocelular, ya que la alfa-fetoproteína (AFP), determinada por radioinmunoanálisis, que es el estándar de oro para tamizaje de este padecimiento, es muy costosa.

Fórmula para sensibilidad:

$$n = \left[\frac{Z_{\alpha^1} \sqrt{2P^0(1-P^0)} + Z_{\beta} \sqrt{P^a(1-P^a) + P^b(1-P^b)}}{P^a - P^b} \right]^2$$

$$n = \left[\frac{1.65 \sqrt{2(0.8)(1-0.8)} + 1.28 \sqrt{0.90(1-0.90) + 0.70(1-0.70)}}{0.90 - 0.70} \right]^2$$

n = 66.5 (67 pacientes con hepatocarcinoma)

n = tamaño de la muestra

p^a = sensibilidad (proporción) del estándar de oro (AFP) (A) = 0.90

p^b = sensibilidad (proporción) de la prueba (CLs) a probar (B) = 0.70

p⁰ = valor ponderado de las dos proporciones (sensibilidades):

$$p^0 = \frac{p^a + p^b}{2} = \frac{0.90 + 0.70}{2} = 0.80$$

poder (1-beta) = 90% (0.90) beta = 10% (0.10):
Zbeta=1.28

a¹ = 0.05 : confianza 95%. Como los CLs serán excluidos como prueba útil solamente si su sensibilidad es significativamente menor (menos de 20%) que la AFP se llevará a cabo una prueba de significancia de una-cola: alfa (a¹)=0.05 con Z_{a¹}=1.65

CONCLUSIÓN

Para un 90% (poder) de nivel de certeza en detectar una sensibilidad diferente de 20% (sensibilidad de 70% Ó mayor de los CLs), se necesitan 67 pacientes con carcinoma hepatocelular comprobado (biopsia). Recuerde que sensibilidad es la probabilidad de que una PRUEBA sea POSITIVA (CLs positivos) en enfermos (carcinoma hepatocelular).

El investigador recomendaría el uso de CLs si hay 90% de certeza de que la diferencia en sensibilidad sea menor o igual a 20%.

Si además, queremos (casi nunca queremos)

comprobado estadísticamente se necesitan aplicar las dos pruebas (AFP y CLs) a los 67 pacientes con carcinoma hepatocelular

Usando los resultados de AFP y de CLs podemos probar si existe una real diferencia entre las proporciones de PRUEBAS POSITIVAS detectadas en cada método a través de una prueba Z o una X².

Si no se encuentra diferencia (en la X² o en la prueba Z) en la sensibilidad de las dos pruebas, podemos decir con un 90% (poder) de certeza que la sensibilidad de los CLs es aceptable.

Fórmula para especificidad:

$$n = \left[\frac{Z_{\alpha^1} \sqrt{2P^0(1-P^0)} + Z_{\beta} \sqrt{P^a(1-P^a) + P^b(1-P^b)}}{P^a - P^b} \right]^2$$

$$n = \left[\frac{1.65 \sqrt{2(0.85)(0.15)} + 1.28 \sqrt{0.95(0.05) + 0.75(0.25)}}{0.95 - 0.75} \right]^2$$

n = 53 Individuos sin hepatocarcinoma

poder = 0.90

a¹ = 0.05

p^a = 0.95 (especificidad)

p^b = 0.75 (especificidad)

$$p^0 = \frac{p^a + p^b}{2} = \frac{0.95 + 0.75}{2} = 0.85$$

Se requieren 53 pacientes libres de la enfermedad para probar una diferencia en especificidad de 20%. Se procede igual que en sensibilidad para la significancia estadística.

APÉNDICE

La distribución normal estándar

Proporciones bajo la curva normal entre la media y algún dato Z. Aparecen con cuatro cifras decimales, sin el punto, para ahorrar espacio.

Tablas en la siguiente página

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0159	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0739	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2133	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2518	2549
0.7	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3718	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4083	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
1.6	4452	4463	4474	4485	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4758	4762	4767

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.0	4773	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4898	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4980	4980	4981
2.9	4981	4982	4983	4984	4984	4984	4985	4985	4986	4986
3.0	4986.5	4987	4987	4988	4988	4988	4989	4989	4989	4990
3.1	4990.0	4991	4991	4991	4992	4992	4992	4992	4993	4993
3.2	4993.12 9									
3.3	4995.16 6									
3.4	4996.63 1									
3.5	4997.67 4									
3.6	4998.40 9									
3.7	4998.92 2									
3.8	4999.27 7									
3.9	4999.51 9									
4.0	4999.68 3									
4.5	4999.96 6									
5.0	4999.99 7133									