Tamaño muestral y potencia

Diseñamos un experimento para el cual el análisis estadístico adecuado es uno de estos contrastes *t* de los que hemos hablado anteriormente. ¿Cuántos individuos necesitaremos para cada muestra? La respuesta a esta pregunta dependerá de la “calidad” deseada del test de comparación de medias y de la propia variabilidad de los datos. La calidad del test depende de los siguientes factores:

a) Nivel de significación utilizado. Lo estándar es utilizar $\alpha = 0.05$. Para niveles más pequeños el contraste es mejor y esto revierte en un aumento del tamaño de la muestra.
b) Potencia requerida. Lo más frecuente es utilizar un valor de 0.8 o 0.9. Un aumento de la potencia también representa un aumento del tamaño de la muestra.
c) La mínima diferencia entre las medias que deseemos detectar. Se trata de una medida de la precisión o capacidad de separación del contraste. Cuanto menor sea la mínima diferencia que queremos detectar mayor será el tamaño de la muestra.

La variabilidad de los datos sólo puede ser determinada a través de un conocimiento previo de los mismos, mediante una experiencia piloto, la lectura de un informe en el que se describa una experiencia similar o un artículo, entre otros. A continuación estudiaremos cómo calcular el tamaño muestral necesario en función de todos los factores anteriormente mencionados para el test *t* para muestras independientes suponiendo las varianzas poblacionales iguales, así como para el test *t* para datos apareados.

Respecto al test de la *t* para muestras independientes, lo primero que cabe preguntarse es si los dos grupos o muestras deben tener el mismo tamaño o no. La respuesta es que fijado el número total de observaciones, la elección del mismo tamaño muestral para cada grupo maximiza la potencia del contraste. Así pues supondremos que emplearemos el mismo número de observaciones para cada grupo. La fórmula que permite obtener el tamaño muestral es la siguiente:

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

En esta expresión, N es el tamaño muestral de cada grupo, σ es el valor estimado de la desviación estándar poblacional, Δ es la mínima diferencia entre las medias que deseamos detectar, y z_α, z_β son valores constantes que dependen de si el contraste es a una o dos colas, del nivel de significación y de la potencia deseados.

Los valores z_α y z_β se obtienen a partir de una tabla de percentiles de la distribución normal estándar. Si el nivel de significación que queremos utilizar es de 0.05 (5%) y el test es a dos colas, entonces $\alpha = 0.05$ y $z_\alpha = 1.96$. Si la potencia que queremos obtener para nuestro test es de 0.9 (90%), entonces $\beta = 0.10$ y $z_\beta = 1.28$.

En la el Cuadro 24-1, se observan los tamaños muestrales adecuados en función de los diversos factores mencionados. Los valores de “d” que aparecen en la tabla son los de Δ/σ de la fórmula.

Ejemplo 1: estamos ensayando un nuevo fármaco y queremos ver si éste afecta al nivel de colesterol en sangre. Queremos ser capaces de detectar una diferencia entre las medias de 10 unidades ($\Delta = 10$). Por un experimento previo sabemos que la dispersión de los datos ha sido de unas 14 unidades ($\sigma = 14$). Supongamos que no tenemos ninguna información previa del resultado del experimento, de modo que haremos un test a dos colas, es decir, contrastaremos H_0: $\mu_1 = \mu_2$ frente a H_1: $\mu_1 \neq \mu_2$. Si trabajamos con un nivel de significación del 5% y una potencia del 90% tendremos.

$$N = 2\frac{(1.96 + 1.28)^2 14^2}{10^2} = 41.15 \approx 42$$

Así pues tomamos dos grupos de 42 individuos cada uno (el grupo control y el grupo tratamiento) a fin de realizar el experimento. Nótese como al aplicar la fórmula redondeamos el resultado por exceso a fin de obtener un contraste lo mejor posible.

En el ejemplo anterior teníamos una estimación de σ, fruto de un experimento previo. Esto significa que, por ejemplo, disponíamos de una serie de observaciones sobre el nivel de colesterol en sangre de individuos no sometidos al tratamiento, y hemos estimado σ calculando la desviación estándar muestral S de los mismos. También hemos podido obtener este dato a partir de un informe o experimento realizado por otros autores. Finalmente, tal vez hemos realizado un ensayo piloto con un tamaño muestral mucho más reducido. En ese caso, disponemos de dos estimaciones de σ al calcular la desviación estándar S_1 del grupo control y la desviación estándar S_2 del grupo tratamiento. A partir de estos dos valores podemos calcular una mejor estimación de σ, utilizando la denominada desviación estándar conjunta S_p, que se calcula de la siguiente manera:

$$S_p = \sqrt{\frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}}$$

Aquí n_1 y n_2 indican los tamaños muestrales de cada uno de los grupos, y S_1, S_2 sus respectivas desviaciones estándar. En el siguiente ejemplo veremos alguno de estos cálculos.

Ejemplo 2: deseamos probar un nuevo pienso de engorde de animales. Para ello realizamos una experiencia piloto, tomando 5 animales y suministrándoles el pienso ordinario, y a 4 animales dándoles el pienso experimental. Tras suministrarles los piensos durante un cierto período de tiempo registramos los incrementos de peso para cada animal. La composición del nuevo pienso nos hace pensar que como mínimo es tan eficaz como el ordinario. Por tanto tiene sentido plantear un contraste a una cola, en el que H_1: $\mu_1 < \mu_2$, siendo μ_1 el incremento de peso medio para los animales que toman el pienso ordinario y μ_2 el incremento de peso medio para los que toman el pienso experimental. Tras realizar una prueba T de Student, no detectamos diferencias significativas entre los pesos medios, siendo el nivel de significación empleado del 5%. Fruto del experimento, hemos calculado $S_1 = 4.7$ y $S_2 = 3.8$. Nos planteamos la siguiente pregunta: ¿cuál tendría que ser el tamaño muestral para que al repetir la experiencia seamos capaces de detectar una diferencia entre los incrementos medios de 3 unidades, realizando un contraste con una potencia del 90%? Para este caso tenemos que $\Delta = 3$, $\alpha = 0.05$, la potencia es de 0.90 y por lo tanto $\beta = 0.10$. Por otro lado calculamos:

$$S_p = \sqrt{\frac{(5-1)4.7^2 + (4-1)3.8^2}{5+4-2}} = 4.34$$

Así pues, $d = 3 / 4.34 \approx 0.7$, y observando en el Cuadro 24-1 obtenemos un tamaño muestral de N = 35 animales para cada grupo.

A veces nos podemos encontrar en la situación de que no disponemos de experiencias previas para poder estimar directamente σ. Sin embargo, si conocemos en líneas generales el rango de variación de los datos también podemos estimar su valor. Por ejemplo, supongamos que queremos estimar σ para unos datos referentes a temperaturas corporales de pacientes. Intuimos que la mayoría de individuos sanos tienen una temperatura corporal que oscila entre 36.5 y 37.5 grados, es decir, 37 ± 0.5. Por otro lado sabemos que para datos distribuidos normalmente, aproximadamente el 95% de las observaciones estarán comprendidas en el intervalo $\mu \pm 2\sigma$. Así pues, identificando esta “mayoría” con el 95% y 2σ con 0.5, obtenemos una estimación $\sigma \approx 0.25$. Lógicamente estos razonamientos son muy arriesgados pero a veces no podremos hacer nada mejor.

Cuando el diseño experimental requiera de medidas apareadas (antes y después, padres e hijos, etc...), el test *t* de Student de comparación de medias apropiado se efectúa de otra manera diferente al referido anteriormente. Por ello también el cálculo del tamaño muestral se hace de manera diferente, correspondiendo a la expresión:

$$N = \frac{(z_\alpha + z_\beta)^2 {\sigma_{dif}}^2}{\Delta^2}$$

Ahora el valor N nos indica el número de sujetos o parejas de datos necesarios para el experimento. Observemos la semejanza con la fórmula para el contraste para muestras independientes, aunque ahora σ_{dif} representa la desviación estándar de las diferencias entre los dos grupos de datos. Aparte de esto, dado que la fórmula es la misma excepto por el 2 que multiplicaba la expresión, se puede utilizar el mismo Cuadro 24-1, pero dividiendo el resultado entre dos. Para ilustrar esto vamos a ver un ejemplo.

> **Ejemplo 3**: En el ejemplo 1 en el que se medían el número de pulsaciones por minuto en 10 pacientes antes y después de tomar un fármaco, realizábamos un contraste a dos colas y concluíamos que existían diferencias significativas entre las medias. En una experiencia similar, ¿qué número de pacientes hay que tomar para ser capaces de detectar una diferencia de 1 latido por minuto? En este caso $\Delta = 1$ y podemos tomar $\sigma_{dif} \approx 1.767$, que es el valor muestral calculado en el experimento previo por el SPSS. Tenemos que $d = \Delta/\sigma_{dif} \approx 0.6$, si decidimos trabajar con $\alpha = 0.05$ y $\beta = 0.10$ consultando la tab. 24.1 obtenemos N = 58. Finalmente para calcular el número de sujetos se divide este número por dos, obteniendo un total de 29 individuos.

Observemos cómo el Cuadro 24-1 también nos indica la potencia aproximada de un contraste que ya se ha llevado a cabo. Así pues para el test del ejemplo anterior detectábamos una diferencia entre las medias de 3.7. Dado que $\sigma_{dif} \approx 1.767$ tenemos que $d = \Delta/\sigma_{dif} \approx 2.1$. Si hemos trabajado a dos colas con un nivel de significación de $\alpha = 0.01$ y un tamaño muestral de N = 10, buscando el valor de β en la tab. 24.1, encontramos que se corresponde con $\beta = 0.05$, es decir, una potencia del 95%.

Todos estos cálculos de tamaños muestrales se refieren a contrastes adecuados para observaciones distribuidas normalmente. En la siguiente sección hablaremos de las llamadas comparaciones no paramétricas para distribuciones no normales.

	α (1 cola) = 0.05				**0.025**				**0.005**			
	α (2 colas) = 0.10				**0.05**				**0.01**			
d	**β = 0.20**	**0.15**	**0.10**	**0.5**	**0.20**	**0.15**	**0.10**	**0.05**	**0.20**	**0.15**	**0.10**	**0.05**
0.10	1237	1438	1713	2165	1570	1795	2102	2599	2337	2609	2977	3563
0.20	309	359	428	541	393	449	526	650	584	652	744	891
0.25	198	230	274	346	251	287	336	416	374	417	476	570
0.30	137	160	190	241	174	199	234	289	260	290	331	396
0.40	77	90	107	135	98	112	131	162	146	163	186	223
0.50	49	58	69	87	63	72	84	104	93	104	119	143
0.60	34	40	48	60	44	50	58	72	65	72	83	99
0.70	27	31	35	44	32	37	43	53	48	53	61	73
0.75	24	28	30	38	30	32	37	46	42	46	53	63
0.80	21	24	29	34	27	30	33	41	37	41	47	56
0.90	17	20	23	29	21	24	28	32	31	32	37	44
1.0	14	16	19	23	18	20	23	28	25	28	30	36
1.1	12	14	16	19	15	17	19	23	21	24	27	31
1.2	11	12	14	17	13	14	17	20	18	20	23	27
1.3	9	11	12	15	11	13	14	17	16	17	20	23
1.4	8	10	11	13	10	11	13	15	14	15	17	20
1.5	7	9	10	12	9	10	11	14	12	14	15	18
1.6	7	8	9	10	8	9	10	12	11	12	14	16
1.7	6	7	8	9	7	8	9	11	10	11	12	14
1.8	6	6	7	9	7	8	8	10	9	10	11	13
1.9	6	6	7	8	6	7	8	9	8	9	10	12
2.0	6	6	6	7	6	6	7	8	8	9	9	11

2.1	6	6	6	7	6	6	7	8	7	8	9	10
2.2	6	6	6	6	6	6	6	7	7	7	8	9
2.3	6	6	6	6	6	6	6	7	6	7	8	9
2.4	6	6	6	6	6	6	6	7	6	7	7	8
2.5	6	6	6	6	6	6	6	6	6	6	7	8
3.0	2	6	6	6	6	6	6	6	6	6	6	6
3.5	2	2	2	6	2	2	6	6	6	6	6	6
4.0	2	2	2	2	2	2	2	6	2	6	6	6

Cuadro 24-1. Tamaños de muestra necesarios para el test de la *t* para datos independientes (tomada de Norman y Streiner, 1996)

Tomado de Piedrafita J, Puig P. Análisis estadístico, diseño experimental e interpretación de los resultados. http://minnie.uab.es/~veteri/00009/bloque4.html