

EL PODER ESTADÍSTICO. DIFERENCIAS OBSERVADAS CUANDO SE CAMBIA EL ALFA ESTABLECIDO EN UN ESTUDIO DE INVESTIGACIÓN

Mylord Reyes Tosta, EdD

Resumen

Se utilizaron dos ejemplos para una prueba de hipótesis pareada, en donde se calculó el poder estadístico utilizando un alfa de 5% y un alfa de 1% con los mismos datos. En el primer ejemplo, utilizando un alfa de 5% con una muestra de 48 sujetos y con una diferencia entre las medias de 40.69, se obtuvo un valor $p = 0.000$ y un poder estadístico de 1.000 en una prueba de hipótesis pareada. En el mismo ejemplo, utilizando los mismos datos, se calculó el poder estadístico utilizando un alfa de 1% y el resultado fue el mismo. Es decir, se obtuvo un poder estadístico de 1.000 en la prueba de hipótesis pareada. En el segundo ejemplo, utilizando un alfa de 5% con una muestra de 30 sujetos y con una diferencia entre las medias de 2.33, se obtuvo un valor $p = 0.012$ y un poder estadístico de .738 en una prueba de hipótesis pareada. En el mismo ejemplo, utilizando los mismos datos, se calculó el poder estadístico utilizando un alfa de 1% y el resultado fue de .483 en la prueba de hipótesis. Es decir, se obtuvo un poder estadístico menor en la prueba de hipótesis pareada. El análisis y las conclusiones de estos resultados se presentan en este artículo.

Palabras claves: poder estadístico, prueba de hipótesis, tamaño del efecto, alfa, beta

Abstract

Two examples were used for paired hypothesis test, where the statistical power was calculated using an alpha of 5% and an alpha of 1% with the same data. In the first example, using an alpha of 5% with a sample of 48 subjects and with a difference between the means of 40.69, yielded a $p\text{-value} = 0.000$ and a power of 1.000 in the paired hypothesis test. In the same example, using the same data, the statistical power was calculated using an alpha of 1% and the result was the same. That is, it was obtained a power of 1.000 in the paired hypothesis test. In the second example, using an alpha of 5% with a sample of 30 subjects and with a difference between the means of 2.33, it was obtained a $p\text{-value} = 0.012$ and a power of .738 in the paired hypothesis test. In the same example, using the same data, the statistical power was calculated using an alpha of 1% and the result was of .483 in the hypothesis test. That is, it was obtained a lower statistical power in the paired hypothesis test. The analysis and conclusions from these results are presented in this paper.

Keywords: statistical power, hypothesis test, size of effect, alpha, beta

INTRODUCCIÓN

El poder de una prueba estadística es la probabilidad de que la prueba va a rechazar la hipótesis nula cuando la hipótesis nula es falsa. Es decir, la probabilidad de no cometer un error de tipo II o tomar una decisión de falsos negativos (Cohen, 1988). Es decir, representa la capacidad de una prueba para detectar como estadísticamente significativas diferencias o asociaciones de una magnitud determinada (Díaz & Fernández, 2003). Cuando el poder aumenta, las posibilidades de un error de tipo II disminuyen. La probabilidad de que ocurra un error de tipo II se refiere a la tasa de falsos negativos (β). Por lo tanto, la potencia es igual a $1 - \beta$, donde β (beta) es el error tipo II que también se conoce como la sensibilidad.

Un análisis de poder se puede utilizar para calcular el tamaño mínimo de muestra requerida. También, puede utilizarse para calcular el efecto mínimo de un tamaño de muestra dado. Además, el concepto de poder se utiliza para hacer comparaciones entre diferentes procedimientos de análisis estadísticos. Por ejemplo, entre una prueba paramétrica y una prueba no paramétrica con las mismas hipótesis (Cohen, 1988).

Factores que influyen en el poder estadístico de cualquier estudio

Los factores que influyen en el poder estadístico de un estudio van a depender de cada situación que es particular de un estudio dado. Los cuatro factores siguientes siempre influyen en el poder estadístico de una prueba.



1. El tamaño de la muestra usado para detectar el efecto. Este determina la cantidad de error de muestreo inherente al resultado de la prueba. Es difícil detectar un efecto en muestras pequeñas. Aumentando el tamaño de la muestra, se puede obtener un poder más alto.
2. La magnitud del efecto de interés en la población. Esta puede ser cuantificada en términos del tamaño del efecto. Donde hay un poder mayor, hay un efecto mayor.
3. El nivel de significancia estadística utilizado en la prueba. Un nivel de significancia estadística es una declaración de lo improbable que puede ser un resultado, si la hipótesis nula es verdadera, para ser considerada significativa.

En otras palabras, cuánto se está dispuesto para tomar el riesgo de asumir una conclusión equivocada. Los criterios más utilizados son las probabilidades de 0.05 (5%, 1 en 20), 0.01 (1%, 1 en 100) y 0.001 (0.1%, 1 en 1000). Si el criterio es 0.05, la probabilidad de obtener el efecto observado cuando la hipótesis nula es verdadera, debe ser inferior a 0.05 y así sucesivamente. Si se utiliza un nivel de significancia de 5%, conocido como alfa o la probabilidad de cometer el error de tipo I, significa que se tiene un nivel de confiabilidad de 95%. Una manera fácil de aumentar la potencia de una prueba, es realizar una prueba menos conservadora al utilizar un nivel de significancia mayor. Esto aumenta la probabilidad de rechazar la hipótesis nula. Es decir, la obtención de un resultado estadísticamente significativo cuando la hipótesis nula es falsa. De esta forma, se reduce el riesgo de cometer un error tipo II. Pero también aumenta el riesgo de obtener un resultado estadísticamente significativo rechazando la hipótesis nula cuando la hipótesis nula es verdadera. En este caso, se aumenta el riesgo de cometer un error de tipo I (Cohen, 1988).

4. La variabilidad de la respuesta o desviación estándar del estudio. Así, cuanto mayor sea la variabilidad en la respuesta, más difícil será detectar diferencias entre los grupos que se comparan y menor será el poder estadístico del estudio. Se recomienda estudiar grupos que sean equivalentes.

Interpretación de los resultados

El análisis adecuado del poder estadístico de una investigación, que es en definitiva la capacidad que tiene el estudio para encontrar diferencias si es que realmente las hay, es un paso fundamental tanto en la fase de diseño como en la interpretación y discusión de sus resultados. A la hora del diseño, por lo tanto, debe establecerse la magnitud mínima de la diferencia o asociación que se considere de relevancia, así como el poder estadístico que se desea para el estudio y, de acuerdo con ello, calcular el tamaño de la muestra necesario (Gordon, Finch, Nothnagel & Ott, 2002). Tanto si los hallazgos son estadísticamente significativos como si no lo son, la estimación de intervalos de confianza pueden también facilitar la interpretación de los resultados en términos de magnitud y relevancia,

proporcionándonos una idea de la precisión con la que se ha efectuado la estimación, de la magnitud y de la dirección del efecto. De este modo, los intervalos de confianza nos permiten tener una idea acerca del poder estadístico de un estudio y, por tanto, de la credibilidad de la ausencia de hallazgos significativos (Díaz & Fernández, 2003). Se pueden considerar los siguientes puntos para interpretar los resultados del poder estadístico:

- a) De acuerdo a Myoung (2003), el estándar adecuado de poder por la mayoría de los investigadores es 0.80
- b) 0.20 = error tipo II
- c) Un poder estadístico ≥ 0.80 indica las probabilidades de decir que hay una relación, diferencia o ganancia. Son las probabilidades que confirman nuestra teoría correctamente.
- d) Un poder estadístico de 0.80 indica que 80 de cada 100 veces, cuando hay un efecto, vamos a decir que lo hay (Myoung, 2003).
- e) Si el poder estadístico es mayor de 0.80, el poder es más dominante.

La Tabla 1 muestra una ilustración de las conclusiones correctas o incorrectas a las que se puede llegar, dependiendo del rechazo o aceptación de la hipótesis nula. En donde α = probabilidad de cometer un error de tipo I y β = probabilidad de cometer un error de tipo II.

Tabla 1. Prueba estadística de contraste de hipótesis

Posibles conclusiones		
Escenarios	Resultado de la prueba	
	Existe asociación o diferencia H0 falsa	Asociación o diferencia significativa. Se rechaza la hipótesis nula
No error (1- β) Poder estadístico		Error de tipo II β Beta
No existe asociación o diferencia H0 verdadera	Error de tipo I α Nivel de significancia	No error (1- α) Nivel de confiabilidad

Con el objetivo de analizar el comportamiento del poder estadístico cuando cambia el alfa establecido en un estudio, se presentan dos ejemplos. En el primer ejemplo, se cuenta con los resultados de una preprueba y una posprueba realizadas a un grupo de 48 sujetos. Se asume que son las notas obtenidas por ellos en una escala del 1 al 100. La Tabla 2 muestra una media para la preprueba de 37.19 con una desviación estándar de 17.43. Mientras que la posprueba, tiene una media de 77.88 con una desviación estándar de 12.26 para una diferencia entre las medias de 40.69 en este primer ejemplo. Claramente se puede notar que los resultados de la posprueba superan a los resultados de la preprueba. Para la prueba de hipótesis pareada se utilizó un alfa de 5% y se obtuvo un $p = .000$ que se muestra en la Tabla 3. Con el valor obtenido ($p = .000$) se puede concluir que sí existe una diferencia estadísticamente significativa entre los resultados de la preprueba y la posprueba. Por lo tanto, se rechazó la hipótesis nula. Con este resultado, se procedió a calcular el poder estadístico de esta prueba que se muestra en la Tabla 4. Se obtuvo un poder estadístico de 1.000, lo cual significa una potencia estadística máxima. Es decir, que se puede afirmar con un 100% de seguridad que verdaderamente existe una diferencia estadísticamente significativa entre las medias. Para todos los cálculos estadísticos se utilizó el programa SPSS versión 20.

Se utilizaron los datos del primer ejemplo para calcular el poder estadístico utilizando un alfa de 1%. Como se puede observar en las Tablas 5 y 6, tanto el valor de p en la prueba de hipótesis pareada, como el valor del poder estadístico, son exactamente los mismos que cuando se utilizó un alfa de 5%.

Tabla 2. Media y desviación estándar de la preprueba y la posprueba

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pretest	37.19	48	17.432	2.516
	Posttest	77.88	48	12.261	1.770

Tabla 3. Prueba de hipótesis pareada con un alfa de 5%

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest - Posttest	-40.688	20.133	2.906	-46.533	-34.842	-14.002	47	.000

Tabla 4. Poder estadístico de la prueba de hipótesis pareada con un alfa de 5%

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	factor1	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power ^a
factor1	Linear	39731.344	1	39731.344	196.046	.000	196.046	1.000
Error (factor1)	Linear	9525.156	47	202.663				

a. Computed using alpha = .05

Tabla 5. Prueba de hipótesis pareada con un alfa de 1%

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	99% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest - Posttest	-40.688	20.133	2.906	-48.489	-32.886	-14.002	47	.000

Tabla 6. Poder estadístico de la prueba de hipótesis pareada con un alfa de 1%**Tests of Within-Subjects Contrasts**

Measure: MEASURE_1

Source	factor1	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power ^a
factor1	Linear	39731.344	1	39731.344	196.046	.000	196.046	1.000
Error (factor1)	Linear	9525.156	47	202.663				

a. Computed using alpha = .01

En el segundo ejemplo, se cuenta con los resultados de una preprueba y una posprueba realizadas a un grupo de 30 sujetos. Se asume que son las notas obtenidas por ellos en una escala del 1 al 100. La Tabla 7 muestra una media para la preprueba de 62.20 con una desviación estándar de 19.04. Mientras que la posprueba tiene una media de 64.53 con una desviación estándar de 16.88 para una diferencia entre las medias de 2.33, en este segundo ejemplo. Claramente se puede notar que los resultados de la posprueba superan a los resultados de la preprueba. Para la prueba de hipótesis pareada se utilizó un alfa de 5% y se obtuvo un $p = .012$ que se muestra en la Tabla 8. Con el valor obtenido ($p = .012$) se puede concluir que sí existe una diferencia estadísticamente significativa entre los resultados de la preprueba y la posprueba. Por lo tanto, se rechazó la hipótesis nula. Con este resultado, se procedió a calcular el poder estadístico de esta prueba que se muestra en la Tabla 9. Se obtuvo un poder estadístico de .738, lo cual significa una potencia estadística débil. Es decir, que solamente se puede afirmar con un 74% de seguridad que verdaderamente existe una diferencia estadísticamente significativa entre las medias.

Se utilizaron estos mismos datos del segundo ejemplo para calcular el poder estadístico utilizando un alfa de 1%. Como se puede observar en la Tabla 10, el resultado en la prueba de hipótesis pareada con un alfa de 1% fue de $p = .012$. Este es igual al obtenido con un alfa del 5%. Mientras que el poder estadístico utilizando un alfa de 1%, como se muestra en la Tabla 11, fue de .483. Este valor es totalmente diferente al obtenido con un alfa de 5% (.738). Este resultado en el poder estadístico indica que

solamente se puede afirmar con un 48% de seguridad que sí existe una diferencia estadísticamente significativa entre las medias de la preprueba y de la posprueba.

A la luz de los hallazgos obtenidos, en el primer ejemplo no hubo diferencia alguna en el poder estadístico cuando se utilizó un alfa del 5% y un alfa del 1%. Mientras que en el segundo ejemplo, cuando se usó un alfa de 5% y se cambió por un alfa de 1%, el valor del poder estadístico de la prueba disminuyó dramáticamente. Este fenómeno radica principalmente en dos factores. El tamaño del efecto es mayor en el primer ejemplo porque la diferencia entre las medias fue mayor (40.69) comparado con la diferencia del segundo ejemplo que fue de 2.33. En la Tabla 12, se muestra el tamaño del efecto observado para el primer ejemplo que dio como resultado .964, lo que representa un tamaño de efecto grande. Por último, la muestra en el primer ejemplo era mayor (48), comparado con la muestra del segundo ejemplo (30). Se concluye que, en el segundo ejemplo, el estudio no tiene el poder estadístico suficiente para asegurar que existen diferencias significativas entre la preprueba y la posprueba. Esto nos indica que en el estudio debe aumentarse el tamaño de la muestra. Finalmente, se puede señalar, también, que al reducir el nivel alfa o error tipo I, si no existe un tamaño del efecto significativo y un tamaño de muestra apropiado, el poder estadístico bajará irremediabilmente.

Tabla 7. Media y desviación estándar de la preprueba y la posprueba

		Paired Samples Statistics			
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pretest	62.20	30	19.038	3.476
	Posttest	64.53	30	16.882	3.082

Tabla 8. Prueba de hipótesis pareada con un alfa de 5%

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest - Posttest	-2.333	4.759	.869	-4.110	-.556	-2.686	29	.012

Tabla 9. Poder estadístico de la prueba de hipótesis pareada con un alfa de 5%

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	factor1	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power ^a
factor1	Linear	81.667	1	81.667	7.213	.012	7.213	.738
Error (factor1)	Linear	328.333	29	11.322				

a. Computed using alpha = .05

Tabla 10. Prueba de hipótesis pareada con un alfa de 1%

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	99% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest - Posttest	-2.333	4.759	.869	-4.728	.061	-2.686	29	.012

Tabla 11. Poder estadístico de la prueba de hipótesis pareada con un alfa de 1%**Tests of Within-Subjects Contrasts**

Measure: MEASURE_1

Source	factor1	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power ^a
factor1	Linear	81.667	1	81.667	7.213	.012	7.213	.483
Error (factor1)	Linear	328.333	29	11.322				

a. Computed using alpha = .01

Tabla 12. Tamaño del efecto observado para el primer ejemplo**Tests of Between-Subjects Effects**

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	317745.094	1	317745.094	1263.303	.000	.964
Error	11821.406	47	251.519			

REFERENCIAS

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). New Jersey: Lawrence Erlbaum.
- Díaz, P. & Fernández, P. (2003). Cálculo del poder estadístico de un estudio. *Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario-Universitario Juan Canalejo. A Coruña, España. Cad Aten Primaria*, 10, 59-63.
- Gordon, D., Finch, S.J., Nothnagel, M. & Ott, J. (2002). Power and sample size calculations for case-control genetic association tests when errors are present: Application to single nucleotide polymorphisms. *Hum Hered* 54, 22-23.
- Myoung, H. (2003). *Understanding the Statistical Power of a Test*. UITS *Center for Statistical and Mathematical Computing*. Indiana University.
- Statistical Package for the Social Sciences (SPSS). IBM SPSS Statistics 21.0. (August, 2012). *IBM Corporation*.

AUTORA

Mylord Reyes Tosta, EdD – Consultora. Correo electrónico: mylord_reyes21@hotmail.com

Copyright 2013 Non-Profit Evaluation & Resource Center, Inc.