

COMPARACIÓN ENTRE TAMAÑOS DE MUESTRAS DE MEDIAS PAREADAS CON ANÁLISIS DE PODER USANDO PAQUETES DEL SOFTWARE R Y EL SOFTWARE G*POWER

Comparison between paired sample sizes with power analysis using R software packages and G*power software

Patricio Fernando Badillo Badillo*, Rubén Antonio Pazmiño Maji

Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Grupo de Investigación Ciencia de Datos/Carrera de Estadística Informática, Riobamba, Ecuador.

*patricio.badillo@esPOCH.edu.ec

Resumen

La verificación de hipótesis requiere un previo planteamiento de factores como el tamaño de muestra y su fiabilidad de las pruebas estadísticas para abordar estudios experimentales, dado que se puede rechazar la presencia de un efecto derivado de un tratamiento, cuando en realidad no hay suficiente poder estadístico para arribar a dicha conclusión. El objetivo de este artículo es definir que es el poder de una prueba estadística, explicar su cálculo y determinar el nivel de aproximación de los tamaños de muestras de la prueba estadística de medias pareadas generados por tres paquetes del software R, identificar si existen similitudes en la generación de resultados con el software G*power; dando un grado de seguridad y confianza para las pruebas estadísticas en el análisis de poder.

Palabras clave: poder estadístico, tamaños de muestras, paquetes de R, G*power.

Abstract

Hypothesis verification requires a prior approach of factors such as the sample size and its reliability of statistical tests to address experimental studies, since the presence of an effect derived from a treatment can be rejected, when in reality there is not enough statistical power to arrive at that conclusion. The objective of this article is to define what is the power of a statistical test, explain its calculation and determine the level of approximation of the sample sizes of the statistical test of paired means generated by three R software packages, and identify if there are similarities in the generation of results with the G*power software; giving a degree of security and confidence for statistical tests in power analysis.

Key words: statistical power, sample sizes, R packages, G*power.

Fecha de recepción: 20-07-2019

Fecha de aceptación: 01-12-2019

I. INTRODUCCIÓN

La potencia de una prueba estadística o el poder estadístico representa la probabilidad de rechazar la hipótesis nula cuando es realmente falsa(1). Es decir, representa la capacidad de un test para detectar como estadísticamente significativas diferencias o asociaciones de una magnitud determinada(1). Jacob Cohen (1962,1990), fue pionero en el uso de esta técnica.

El poder de una prueba estadística es la probabilidad de que produzca resultados estadísticamente significativos(2). Cohen (1988 pag.16) (3) afirma que si se diseña un estudio para que tenga alto poder $1-\beta$ (sensibilidad o capacidad del estimador), para detectar alejamientos de la hipótesis nula (siendo [Efecto de la muestra]= Δ), y no se rechaza la hipótesis nula, entonces la conclusión de que el verdadero valor del parámetro se encuentra en Δ unidades del valor de la hipótesis

nula, es significativo al nivel β (probabilidad de que ocurra un error tipo II)(4).

El análisis de poder es muy importante cuando se está realizando un diseño de experimento, permitiendo determinar el tamaño de muestra requerido con parámetros establecidos como es el tamaño del efecto y el nivel de significancia. Si la probabilidad es baja, sería prudentemente de alterar sus parámetros o abandonar el experimento(5).

El análisis de poder para diseños de medidas repetidas de dos celdas es lógicamente simple. El concepto de prueba pareada se puede extender a comparaciones de más de dos grupos y hablaremos entonces de bloques de m elementos(6). Si conocemos el promedio de la puntuación de diferencia (Δ) y su desviación estándar (sd), el tamaño del efecto viene dado por $d = \Delta/sd$. La hipótesis nula y alterna viene dada de la siguiente forma:

$H_0: \mu_T = \mu_C$ (caso medias poblacionales iguales)

$H_a: \mu_T \neq \mu_C$ (caso de medias poblacionales distintas)

Para ello se plantea el estadístico t-test

$$t = \frac{\bar{\mu}_T - \mu_C}{\frac{\sigma}{\sqrt{n}}} \quad (1)$$

donde $\bar{\mu}_T$ es la media de la variable en el grupo de tratamiento, μ_C la media de la variable en el grupo de control, σ el error estándar y n la cantidad de casos en la muestra(7).

El poder depende del tamaño del efecto de la población, que por lo general se tiende a desconocer, llevando a utilizar la mejor estimación disponible de dicho tamaño. En esta investigación comprobaremos si los tamaños de las muestras generadas por diferentes paquetes del software R son iguales o son distintas que los tamaños de las muestras generada por el software G*power.

El software G*power es una herramienta para calcular análisis estadísticos de potencia para muchas pruebas t, pruebas F, pruebas χ^2 , pruebas z y algunas pruebas exactas(8). G*power también se puede utilizar para calcular los tamaños de los efectos y para mostrar gráficamente los resultados de los análisis de potencia(9). En nuestro caso

utilizamos el software R ya que posee distintos paquetes estadísticos que tienen funciones para realizar análisis de poder como son los paquetes: *pwr*, *powerSurvEpi*, *powerGWASinteraction*, *Simr*, *SampleSize*, *powerEQTL*, *WebPower*(10). No todos estos paquetes calculan el tamaño de la muestra de medias pareadas.

Para nuestro estudio se utilizó *pwr*, *WebPower* y *powerAnalysis* con los que se obtuvo diferentes tamaños de muestras de medias pareadas mediante diferentes combinaciones entre el tamaño de efecto, nivel de significancia y poder.

II. MATERIALES Y MÉTODOS

Se realizó una investigación experimental, de tipo descriptiva, para proporcionar elementos que permitan comparar diferentes parámetros expuestos en un análisis de poder y posibles alternativas para tamaños de muestras pareadas, las cuales son importantes en estrategia de búsqueda sobre metodología estadística de investigación.

La herramienta de soporte para este estudio fue el software R versión 3.5.3 y la utilización de tres paquetes *pwr*, *WebPower* y *powerAnalysis* los cuales son gratuitos y se encuentran en el CRAN y el software G*power versión 3.1.

El paquete *pwr* tiene funciones de análisis de poder a lo largo de las líneas de Cohen (1988) utilizan en particular las mismas notaciones para los tamaños de efecto, su versión que se utilizó es 1.2-2, con fecha de publicación 2018-03-03, "*pwr.t.test()*" calcule la potencia de las pruebas o determina los parámetros para obtener la potencia objetivo(11).

El paquete *WebPower* es una colección de herramientas para realizar tanto análisis de potencia estadística básica como avanzada, incluyendo correlación, proporción, prueba t, ANOVA de una vía, ANOVA de dos vías, regresión lineal, regresión logística, etc. La versión que se utilizó es 0.5.2, con fecha de publicación 2018-11-29, "*wp.t()*" puede evaluar la significación estadística de la diferencia entre la población media y un valor específico, la diferencia entre dos medias de

población independientes y emparejadas(12). El paquete *powerAnalysis* posee funciones básicas para el análisis de potencia y el cálculo del tamaño del efecto, La versión que se utilizó es 0.2.1, con fecha de publicación 2017-02-02, "*power.t()*" puede realizar cálculos de potencia para la prueba t(13).

III. RESULTADOS

Con la utilización de los paquetes *pwr*, *WebPower* y *powerAnalysis* del software R en la versión 3.5.3, se obtuvieron mediante una simulación matemática combinaciones entre el tamaño de efecto de 0.1 a 0.9, el nivel de significancia de 0.01 a 0.05 y un poder de 0.50 a 0.99 con tres repeticiones, dando un total de 6750 datos en cada paquete.

Paquete	Repeticion	Efecto	Alfa	Poder	Tamaño_muestra	
1	pwr	1	0.1	0.01	0.50	666.8077
2	pwr	2	0.1	0.01	0.50	666.8077
3	pwr	3	0.1	0.01	0.50	666.8077
4	pwr	1	0.1	0.01	0.51	679.7852
5	pwr	2	0.1	0.01	0.51	679.7852
6	pwr	3	0.1	0.01	0.51	679.7852
7	pwr	1	0.1	0.01	0.52	692.8967
8	pwr	2	0.1	0.01	0.52	692.8967
9	pwr	3	0.1	0.01	0.52	692.8967
10	pwr	1	0.1	0.01	0.53	706.1508
11	pwr	2	0.1	0.01	0.53	706.1508
12	pwr	3	0.1	0.01	0.53	706.1508
13	pwr	1	0.1	0.01	0.54	719.5566
14	pwr	2	0.1	0.01	0.54	719.5566
15	pwr	3	0.1	0.01	0.54	719.5566
16	pwr	1	0.1	0.01	0.55	733.1234
17	pwr	2	0.1	0.01	0.55	733.1234
18	pwr	3	0.1	0.01	0.55	733.1234

Tabla 1. Base de datos generada con los paquetes *pwr*, *Webpower* y *powerAnalysis* en R.

Se probaron los supuestos de normalidad(14), independencia(15) y homocedasticidad(16), con cada uno de los paquetes, con el fin de verificar si existen diferencias significativas en su uso.

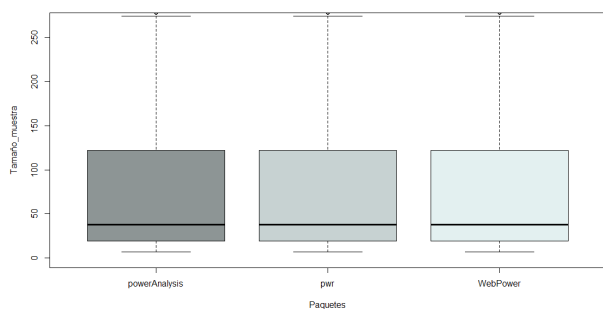


Figura 1. Representación de los tamaños de las muestras con respecto a cada paquete.

La asimetría de los datos generados(17) por los tres paquetes del software R se presenta a continuación:

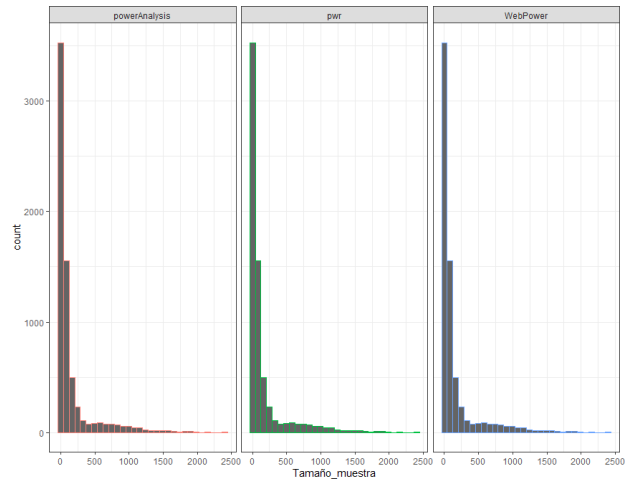


Figura 2. Asimetría de los datos de los tres paquetes de R

En la representación gráfica de los datos se observa que las muestras no siguen una distribución normal. Las tres muestras de los paquetes "*pwr*", "*Webpower*" y "*powerAnalysis*" presentan asimetría hacia la derecha. Con el uso de diferentes test para verificar si las muestras siguen una distribución normal se obtuvo que mediante el test de Anderson-Darling, Lilliefors y Jarque Bera(18), resultó un p-value inferior a $2.2e-16$ con lo que se pudo afirmar que las muestras del tamaño del efecto, nivel de significancia, poder y del tamaño de la muestra no seguían distribución normal. Mediante el test Chi-cuadrado para verificar si las muestras son independientes(19), se obtuvo que el paquete con respecto al tamaño de muestra resultó un p-value inferior a $2.2e-16$, por lo que se pudo afirmar que estas variables son dependientes; de igual forma el poder es dependiente del paquete que se utilizó. Mediante el test de Levene y Bartlett para verificar si las muestras siguen homocedasticidad(20), se obtuvo que el tamaño de muestra, efecto y poder resultó un p-value igual a 1, afirmando que poseen varianzas iguales.

Al hacer el análisis de nuestros datos se comprobó que no seguían una distribución normal, por lo cual, se decidió recurrir a pruebas de contraste no paramétrico, puesto que las pruebas no paramétricas, a pesar de ser igualmente rigurosas que las paramétricas, no permiten rechazar la hipótesis nula de nuestro estudio, se procedió a hacer una transformación de los datos(21) con fin de que los nuevos datos sigan una distribución

normal, a continuación, se aplica el test de Anderson Darlin para probar normalidad; se describen las técnicas usadas a continuación:

Transformación (tres paquetes de R)	p-value	Transformación (pwr y G*power)	p-value
Logaritmos	3.7e-24	Logaritmos	3.7e-24
Exponencial	3.7e-24	Exponencial	3.7e-24
Paquete de Johnson	3.7e-24	Paquete de Johnson	0.0001698978
Box-Cox	3.7e-24	Box-Cox	3.7e-24

Tabla 2. Transformaciones realizadas en los datos

Se determinó que los datos nuevos son estrictamente no normales ya que se rechazó la hipótesis nula resultando un su p-valor es inferior a 0.05; llevando a la conclusión que se aplicaría en el estudio pruebas no paramétricas(22). Mediante un análisis multifactorial Kruskal-Wallis por rango, es una alternativa no paramétrica ya que esta prueba se recomienda cuando las muestras no provienen de una distribución normal, con igual varianza y misma distribución en los grupos(23). Con el test de Kruskal-Wallis se obtuvo un valor de p es de 0.999 siendo mayor que el nivel de significancia de 0.05, concluyendo que no existen diferencias significativas entre los grupos de tratamiento, es decir los tamaños de muestras de medias pareadas son influenciados a ser iguales por los paquetes que se utilizó en R como son pwr, WebPower y powerAnalysis.

De la población de los datos generados del paquete "pwr" del software R se sacó una muestra de 363 datos mediante un muestreo aleatorio simple, los cuales fueron replicados en las mismas condiciones en el software G*power; se muestran a continuación:

Paquete	Efecto	Alfa	Poder	Tamaño_muestra
357 pwr	0.1	0.02	0.86	1163.245869
358 pwr	0.5	0.03	0.53	22.569335
359 pwr	0.9	0.04	0.50	7.447270
360 pwr	0.5	0.02	0.77	40.328738
361 pwr	0.9	0.02	0.76	14.164889
362 pwr	0.5	0.02	0.54	26.299033
363 pwr	0.3	0.02	0.53	66.806831
364 G*power	0.1	0.03	0.90	1194.000000
365 G*power	0.2	0.05	0.98	405.000000
366 G*power	0.5	0.02	0.76	40.000000
367 G*power	0.5	0.02	0.97	74.000000
368 G*power	0.1	0.01	0.78	1125.000000
369 G*power	0.8	0.04	0.75	14.000000

Tabla 3. Base de datos generada con el paquete "pwr" del software R y el software G*power

Se escogió aleatoriamente el paquete "pwr" del

software R, se sacó una muestra de 363 datos los cuales se replicaron en el software G*power para determinar si existen diferencias en los resultados que generan los dos softwares.

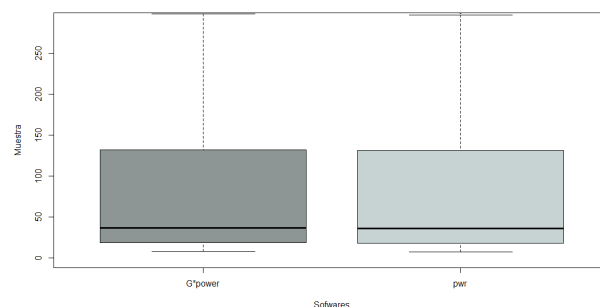


Figura 3. Representación de las muestras de G*power y pwr.

La asimetría de los datos generados en el paquete pwr del software R y el software G*power se presenta a continuación:

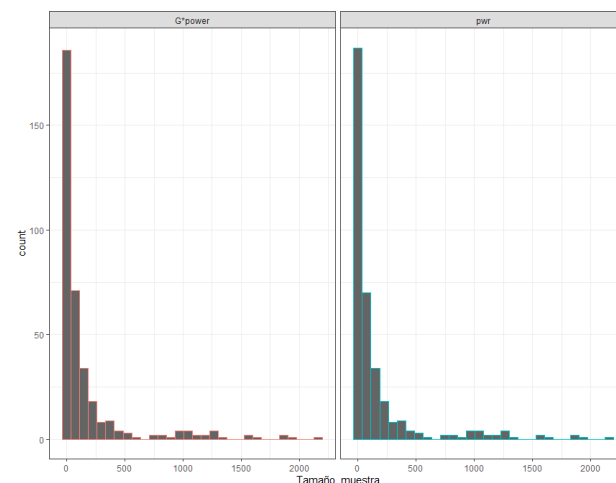


Figura 4. Asimetría de los datos del paquete pwr de R y G*power

En la representación gráfica de los datos se observa que la muestra no sigue una distribución normal, los datos del paquete "pwr" del software R y los datos del software G*power presentan asimetría positiva.

La comprobación de los supuestos mediante el test de Anderson-Darling, Lilliefors y Jarque Bera resultó un p-value inferior a 2.2e-16 con lo que se pudo afirmar que las muestras del tamaño del efecto, nivel de significancia, poder y el tamaño de la muestra no seguían una distribución normal. Con el test Chi-Cuadrado se obtuvo que el paquete con respecto al tamaño de muestra resultó un p-value de 5.292e-10 por lo que se pudo afirmar que son dependientes, de igual forma el poder es dependiente del paquete que se utilice. Con el test de Levene y Bartlett para verificar si las muestras siguen Homocedasticidad o varian-

zas iguales, se obtuvo que el tamaño de muestra, efecto y el poder un p-value de 0.9998 por lo que se pudo afirmar que poseen varianzas iguales.

Mediante el análisis del test de Kruskal-Wallis se obtuvo un p-value es de 0.7518 siendo mayor que el nivel de significancia de 0.05, concluyendo que no existen diferencias significativas entre los grupos de tratamiento, es decir la muestra de medias pareadas del paquete "pwr" del software R y en el software G*power tienden a ser iguales; mediante las comparaciones por pares utilizando la prueba de suma de rangos de Wilcoxon(24) dió un p-value de 0.75 de forma que corrobora con el test de Kruskal-Wallis.

IV. CONCLUSIONES

Según autores como Marco Perugini, Marcello Gallucci, Giulio Costantini(25) el análisis de potencia se pueden realizar en diferentes softwares como G*power o R los cuales están equipados para que sean exactos en sus resultados, pero también pueden ser comprados con softwares que brinden las mismas condiciones para el análisis de poder.

Con la utilización de tres paquetes del software R como son pwr, WebPower y powerAnalysis se obtuvieron varios tamaños de muestras de medias pareadas mediante combinaciones entre el

tamaño de efecto, nivel de significancia y poder. Se determinó que los grupos generados por el software fueron significativamente similares ya que no se rechazó la hipótesis de nula con el test de Kruskal-Wallis, se obtuvo un p-value de 0.999 siendo mayor que el nivel de significancia de 0.05, llevando a la conclusión que tanto los tamaños de muestras generados por los tres paquetes son iguales en un análisis de poder para medias pareadas. También se hizo una comparación con los datos que genera el software G*power, llevando a tomar de la población de los datos generados del paquete "pwr" del software R una muestra de 363 datos mediante un muestreo aleatorio simple, los que se los replicó a las mismas condiciones en software G*power, obteniendo como resultados de que el software "G*power" y el paquete de R "pwr" entregan valores similares para los tamaños de muestras utilizando la prueba de suma de rangos de Wilcoxon con un p-value de 0.75 como estimado.

V. AGRADECIMIENTOS

Mi sincero agradecimiento al Dr. Rubén Pazmiño docente de la Escuela Superior Politécnica de Chimborazo por impartir sus conocimientos para que se plasmen en realidad de forma aplicada en este trabajo siendo de utilidad para el sector académico y científico.

Referencias

1. Cálculo del poder estadístico de un estudio [Internet]. [citado 25 de noviembre de 2019]. Disponible en: https://www.fisterra.com/mbe/investiga/poder_estadistico/poder_estadistico.asp
2. Para C, Investigación LA, Fisher R. TAMAÑO DE EFECTO : 2007;
3. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edi. New York, New York, USA: LAWRENCE ERLBAUM ASSOCIATES; 1988. 579 p.
4. Gesto Giannattasio N. Análisis de Poder Estadístico y su Aplicación a Evaluaciones Experimentales. 2016 ene.
5. Perugini M, Gallucci M, Costantini G. A Practical Primer To Power Analysis for Simple Experimental Designs. Int Rev Soc Psychol. julio de 2018;31(1):20.
6. Pruebas pareadas. SHE. mayo de 2003;6.
7. Unknown P por. Prueba t de Student [Internet]. [citado 25 de noviembre de 2019]. Disponible en: <http://pruebatstudentf.blogspot.com/>
8. Kiel C. G * Power 3 : A flexible statistical power analysis program for the social , behavioral , and biomedical sciences. 2007;39(2):175-91.
9. (PDF) A short tutorial of GPower [Internet]. [citado 25 de noviembre de 2019]. Disponible en: https://www.researchgate.net/publication/49619426_A_short_tutorial_of_GPower
10. Blomberg S. Power Analysis using R. 2015.
11. pwr.pdf [Internet]. [citado 26 de noviembre de 2019]. Disponible en: <https://cran.r-project>.

org/web/packages/pwr/pwr.pdf

12. WebPower.pdf [Internet]. [citado 26 de noviembre de 2019]. Disponible en: <https://cran.r-project.org/web/packages/WebPower/WebPower.pdf>
13. powerAnalysis.pdf [Internet]. [citado 26 de noviembre de 2019]. Disponible en: <https://cran.r-project.org/web/packages/powerAnalysis/powerAnalysis.pdf>
14. Assumptions Part 1: Normality [Internet]. Discovering Statistics. 2012 [citado 26 de noviembre de 2019]. Disponible en: <https://www.discoveringstatistics.com/2012/08/06/assumptions-part-1-normality/>
15. Zhang S, Jalali D, Wuttke J, Mu K, Lam W, Ernst MD, et al. Empirically Revisiting the Test Independence Assumption.
16. Assumptions Part 2: Homogeneity of Variance/Homoscedasticity [Internet]. Discovering Statistics. 2012 [citado 26 de noviembre de 2019]. Disponible en: <https://www.discoveringstatistics.com/2012/09/13/assumptions-part-2-homogeneity-of-variancehomoscedasticity/>
17. Garson GD. Single User License . Do not copy or post . Single User License . Do not copy or post . 2012;1-52.
18. Moni M, Shuaib M. A Comparison of Power of Normality Tests: Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, Anderson-Darling and Jarque-Bera Tests. 2015.
19. Tests S. Chapter 10 Chi Square Tests.
20. Mu Z. Comparing the Statistical Tests for Homogeneity of Variances. En 2006.
21. Consejos para reconocer y transformar datos no normales - iSixSigma [Internet]. 2010 [citado 26 de noviembre de 2019]. Disponible en: <https://www.isixsigma.com/tools-templates/normality/tips-recognizing-and-transforming-non-normal-data/>
22. Cowan AR, Sergeant AMA. Interacting biases, non-normal return distributions and the performance of tests for long-horizon event studies. *J Bank Finance* [Internet]. 1 de abril de 2001 [citado 26 de noviembre de 2019];25(4):741-65. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0378426600000947>
23. anova_un_factor-lectura.pdf [Internet]. [citado 26 de noviembre de 2019]. Disponible en: https://previa.uclm.es/profesorado/mdsalvador/58109/teoria/anova_un_factor-lectura.pdf
24. tests noparametricos.pdf [Internet]. [citado 26 de noviembre de 2019]. Disponible en: https://www.ugr.es/~rruizb/cognosfera/sala_de_estudio/estadistica/tests%20noparametricos.PDF
25. Perugini M, Gallucci M, Costantini G. Salvaguardar el poder como protección contra estimaciones imprecisas del poder. *Perspect Psychol Sci* [Internet]. 1 de mayo de 2014 [citado 26 de noviembre de 2019];9(3):319-32. Disponible en: <https://doi.org/10.1177/1745691614528519>