

<https://doi.org/10.5232/ricyde2021.06407>

**El grado de la evidencia de las hipótesis estadísticas mediante el factor Bayes
en las ciencias del deporte**

**The degree of evidence for statistical hypotheses using the Bayes factor
in sports science**

Cristian Antony Ramos-Vera
Universidad Cesar Vallejo. Perú

Carta al editor

Señor editor,

En el volumen 16 de *RICYDE. Revista Internacional de Ciencias del Deporte*, se publicó una importante investigación que reporta la existencia de una correlación negativa y estadísticamente significativa entre la frecuencia cardíaca media (FCM) y la diferencia de medias en porcentaje del tiempo de agotamiento en la prueba aeróbica (Dif. % Tagot), en 15 corredores de montaña, cuyos datos fueron analizados mediante las pruebas de significancia estadística ($p < 0,05$) mediante el coeficiente de correlación de Pearson (Landart, Cámara, Urdampilleta, Santos-Concejero, Gomez, y Yanci, 2020). Otro estudio más reciente en la presente revista reportó una diferencia significativa mediante la prueba *t* de student, el cual presentó un mayor nivel en la medida física de la capacidad cardiorrespiratoria en el post-test en contraste al pre-test en 14 alumnos de educación secundaria (Polo-Recuero, Moreno-Barrio y Ordóñez-Dios, 2020), que evidenció un aumento del efecto de intervención del estudio.

La presente carta tiene como objetivo presentar dos ejemplos de reanálisis bayesiano (Ly, Raj, Etz, Gronau y Wagenmakers, 2018) a partir de los valores estadísticos, el coeficiente de correlación ($r = -0,78$), el valor *t* (-3,422) y los datos muestrales (15 y 14) de ambos estudios respectivamente.

La estadística bayesiana permite contrastar dos hipótesis opuestas mediante probabilidades a priori y posteriori dado los datos, cuya utilidad es más idónea en la investigación en las ciencias del deporte a diferencia de las pruebas de significancia estadística que se basan en el rendimiento promedio hipotético al realizar una serie de repeticiones idénticas del estudio (Marsman y Wagenmakers, 2017). El factor Bayes (FB) es recomendable en las investigaciones de las ciencias del deporte que tengan como objetivo confirmar la hipótesis alterna dado los datos (p. ej., determinar un mayor efecto de intervención pre y post-test; Polo-Recuero y col., 2020), debido a que permite precisar el nivel de certeza de esta hipótesis más allá de las limitaciones del poder estadístico, el uso de las pruebas post hoc o la intención del muestreo a diferencia de las pruebas de significancia (Marsman y Wagenmakers, 2017).

Palabras clave: replicación estadística; análisis bayesiano; significancia estadística.

Key words: statistical replication; Bayesian analysis; statistical significance.

Correspondencia/correspondence: Cristian Antony Ramos Vera
Área de investigación. Facultad de Ciencias de la Salud. Universidad Cesar Vallejo. Perú
Email: cristonny_777@hotmail.com

La utilización única de las hipótesis de significancia para reportar conclusiones es controversial, pues solo permite inferir el rechazo de la hipótesis nula, mientras se asume la hipótesis alterna sin confirmar su validez. Esto ha generado un mal uso de los valores de significancia e interpretaciones erróneas (Greenland y col., 2016). Cuyos sesgos de interpretación más comunes son “la falacia de la probabilidad inversa”, que es la falsa creencia que el valor p se refiere a la probabilidad de tener una hipótesis nula (H_0) verdadera, otra es “la falacia del tamaño del efecto” que vincula la significación estadística con la magnitud del efecto. Así, los p valores pequeños se interpretan como efectos grandes, sin embargo, no informan la magnitud de un efecto (Greenland y col., 2016).

Los hallazgos significativos con pequeños datos muestrales presentan mayormente valores estadísticamente significativos con efectos inflados, los cuales refieren una mayor incertidumbre de precisión del efecto verdadero que se evidencia en los intervalos de confianza muy amplios (Abt y col., 2020; Bridges, 2019). El beneficio de la inferencia bayesiana como análisis o reanálisis permite confirmar cuan probable es el verdadero tamaño de efecto significativo en contraste a la nulidad o viceversa (Ly y col., 2018), y una correcta estimación e interpretación de los intervalos de credibilidad dado los datos a diferencia de los intervalos frecuentistas. A su vez permite contrastar la fuerza probatoria de las hipótesis direccionadas de correlación y comparación entre grupos (p. ej., correlación positiva; Ramos-Vera, 2020).

El uso de la *Magnitude Based Inference* (MBI) también ha sido considerado como una alternativa más robusta en contraste a las hipótesis de significancia en el ámbito deportivo, el cual se ha referido que presenta características del modelo bayesiano, sin embargo, no tiene el rigor estadístico de la inferencia bayesiana (Welsh y Knight, 2015). El método MBI no se escapa del marco de los p valores, pues tiende a una mayor prevalencia de índices de error tipo I (falsos positivos), esta es incluso mayor que la estimación frecuentista, a su vez también está condicionado por la potencia estadística y el tamaño muestral (Welsh y Knight, 2015). Dichas cuestiones no refuerzan la consideración del MBI como una mejora metodológica para la inferencia estadística en relación a las hipótesis de significancia a diferencia del modelo bayesiano en el ámbito de las ciencias del deporte (Harrison y col., 2020; Sainani, Lohse, Jones y Vickers, 2019).

Desde el enfoque bayesiano, el FB es el método alternativo para evaluar la probabilidad dado los datos de las hipótesis estadísticas, el cual estima el grado de evidencia de la hipótesis alterna en relación, a la hipótesis nula, a partir del esquema de clasificación de Jeffreys (1961): débil, moderado, fuerte y muy fuerte (tabla 1). Esto brinda información adicional más allá de la interpretación dicotómica del rechazo o aceptación de la hipótesis nula e intervalos de credibilidad mas precisos de las estimaciones estadísticas.

Tabla 1. Valores de interpretación cuantificable del factor Bayes

>30	Muy fuerte	Hipótesis alternativa
10-30	Fuerte	Hipótesis alternativa
3,1-10	Moderado	Hipótesis alternativa
1,1-3	Débil	Hipótesis alternativa
1	0	No evidencia
0,3-0,99	Débil	Hipótesis nula
0,3-0,1	Moderado	Hipótesis nula
0,1-0,03	Fuerte	Hipótesis nula
<0,03	Muy fuerte	Hipótesis nula

Nota: Creación propia según Jeffreys (1961)

El FB permite inferir dos interpretaciones: FB_{10} (a favor de la hipótesis alternativa) y FB_{01} (a favor de la hipótesis nula) y el intervalo de credibilidad al 95% (Goss-Sampson, 2020). Los resultados, obtenidos mediante el factor Bayes son $FB_{10} = 17,499$ y $FB_{01} = 0,057$ e IC95% [0,254 a 0,879] en el estudio de deportistas de montaña y $FB_{10} = 10,894$ y $FB_{01} = 0,092$ e IC95% [-1,435 a -0,199] en la investigación con adolescentes. Estos hallazgos respaldaron con una fuerza probatoria concluyente (fuerte) a favor de las hipótesis alternas de ambas investigaciones. También, se reportaron los parámetros del factor Bayes máximo ($\max FB_{10} = 17,52$; $\max FB_{01} = 10,91$), cuyas estimaciones similares afianzan la estabilidad de los resultados de la inferencia bayesiana.

Muchos estudios en ciencias del deporte utilizan datos muestrales de poblaciones particulares de interés que incluyen menos de 250 participantes (p. ej., Landart y col., 2020; Polo-Recuero y col., 2020) que favorecen un mayor error aleatorio y evidencian estimaciones imprecisas con sesgos al replicar tales hallazgos, y promueven una mayor prevalencia en reportar falsos positivos o errores de tipo I (Ramos-Vera, en prensa; Schönbrodt y Perugini, 2013).

Además, no hay un consenso claro de las pautas de interpretación de los tamaños de efectos (TE), pues varían entre las diversas áreas y subdisciplinas científicas debido al tipo de investigación, las medidas específicas utilizadas y las poblaciones de interés (Bridges, 2019; Kinney, Eakman y Graham, 2020).

Un estudio reciente determinó directrices de interpretación del TE en el área de rehabilitación, en base a 3381 tamaños de efecto extraídos de 99 investigaciones metaanalíticas de tratamientos de rehabilitación (Kinney et al., 2020), esta investigación precisó los criterios de interpretación de TE en: efecto pequeño (0,08 a 0,15), efecto mediano (0,19 a 0,36), y efecto grande (0,41 a 0,67), los cuales difieren de los criterios de Cohen (1988).

En la investigación de Landart et al. (2020) se consideró las pautas del TE según Cohen (1988) para reforzar las pruebas de significancia, sin embargo, es más idóneo considerar criterios de interpretación en base a artículos sistemáticos de TE más recientes del área deportiva de interés (Harrison y col., 2020; Correll, Mellinger, McClelland y Judd, 2020). Las directrices de Cohen (1988) son controversiales en la investigación deportiva dado que están orientadas a otro ámbito científico (ciencias del comportamiento). Además, no toman en cuenta la incertidumbre, pues hay un mayor nivel de error aleatorio cuando se usan pequeños datos muestrales (Bridges, 2019).

Asimismo, en la investigación de las ciencias del deporte no hay un estándar interpretativo fijo del TE para cada subdisciplina particular, por lo tanto, el uso del FB es un gran aporte metodológico inclusivo para las diferentes áreas de la investigación deportiva (Abt y col., 2020; Correll y col., 2020). Cuyo uso como replicación estadística de hallazgos significativos que reporten una evidencia concluyente o superior ($FB_{10} > 10$) permita reforzar la credibilidad práctica (p. ej., la implicancia de un mayor efecto de una intervención de una prueba física mediante pre-test y post-test; Polo-Recuero et al., 2020).

La inclusión del FB como replicación estadística ha sido utilizada en un estudio reciente de Kelter (2020), el cual reevaluó varios artículos previos, uno de los cuales estuvo compuesto por 800 pacientes y reportó el análisis de varianza ANOVA para evaluar la hipótesis de diferencia del promedio de las frecuencias cardíacas según género en cuatro grupos: dos de control y dos de tratamiento. A su vez Kelter (2020) reportó la inferencia bayesiana de otra investigación anterior que aplicó el análisis de regresión lineal entre el número promedio de pasos y su efecto en el índice de masa corporal en 100 universitarios.

La conversión del TE y otras medidas estadísticas que se basan en las hipótesis de significancia (d , f , η^2 , OR, χ^2 , z) a un coeficiente de correlación (r) permite afianzar futuros análisis y reanálisis bayesianos. Se recomienda realizar estos hallazgos mediante el paquete estadístico gratuito JASP (Goss-Sampson, 2020), este programa (similar al SPSS) es de fácil acceso y comprensible para cualquier investigador o profesional de las ciencias del deporte. Cabe precisar que las estimaciones de conversión son fáciles de realizar mediante la calculadora online de Lenhard y Lenhard (2016). Adicionalmente, es primordial para reforzar las investigaciones cuantitativas sistemáticas al permitir la inclusión de mayor número de artículos con diversos valores estadísticos convertibles. La interpretación del FB en estos estudios, fortalece una mayor credibilidad en las conclusiones meta-analíticas con intervalos bayesianos más precisos.

Otro de los beneficios del FB es su utilidad en la planificación de la investigación para detectar un mínimo tamaño muestral que estime la probabilidad bayesiana de obtener falsos positivos, $FB_{10} < 1$, e incluir datos muestrales que afiancen el poder estadístico de futuros estudios de las ciencias del deporte y favorezcan su replicación (Abt y col., 2020).

En conclusión, el factor Bayes es un gran aporte metodológico que presenta una implicancia práctica en la toma de decisiones importantes a partir de la confirmación de resultados que sean eficazmente concluyentes, cuya aplicación es inclusiva para las diversas subdisciplinas de la investigación en las ciencias del deporte.

Referencias

- Abt, G.; Boreham, C.; Davison, G.; Jackson, R.; Nevill, A.; Wallace, E., & Williams, M. (2020). Power, precision, and sample size estimation in sport and exercise science research. *Journal of Sports Sciences*, 38(17), 1933-1935. <https://doi.org/10.1080/02640414.2020.1776002>
- Brydges, C. R. (2019). Effect size guidelines, sample size calculations, and statistical power in gerontology. *Innovation in Aging*, 3(4), igz036. <https://doi.org/10.1093/geroni/igz036>.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates
- Correll, J.; Mellinger, C.; McClelland, G. H., & Judd, C. M. (2020). Avoid Cohen's 'Small', 'Medium', and 'Large' for Power Analysis. *Trends in Cognitive Sciences*, 24(3), 200-207. <https://doi.org/10.1016/j.tics.2019.12.009>
- Goss-Sampson, M. A. (2020). *Bayesian Inference in JASP: A Guide for Students*. University of Amsterdam: JASP team. <http://dx.doi.org/10.17605/OSF.IO/CKNXM>.
- Greenland, S.; Senn, S. J.; Rothman, K. J.; Carlin, J. B.; Poole, C.; Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*, 31(4), 337-350. <https://doi.org/10.1007/s10654-016-0149-3>
- Harrison, A. J.; McErlain-Naylor, S. A.; Bradshaw, E. J.; Dai, B.; Nunome, H.; Hughes, G. T., ... & Fong, D. T. (2020). Recommendations for statistical analysis involving null hypothesis significance testing. *Sports Biomechanics*, 19(5), 561-568. <https://doi.org/10.1080/14763141.2020.1782555>
- Jeffreys, H. (1961). *Theory of probability* (3ra ed.). Oxford, England: Oxford University Press.

- Landart, A.; Cámara, J.; Urdampilleta, A.; Santos-Concejero, J.; Gomez, J., y Yanci, J. (2020). Análisis de la fatiga neuromuscular y cardiovascular tras disputar una maratón de montaña. *RICYDE. Revista Internacional de Ciencias del Deporte*, 16(59), 43-56. <https://doi.org/10.5232/ricyde2020.05904>
- Lenhard, W., & Lenhard, A. (2016). Calculation of Effect Sizes. Disponible en: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. <https://doi.org/10.13140/RG.2.2.17823.92329>
- Ly, A.; Raj, A.; Etz, A.; Gronau, Q. F., & Wagenmakers, E.J. (2018). Bayesian reanalyses from summary statistics: a guide for academic consumers. *Advances in Methods and Practices in Psychological Science*, 1(3), 367-374. Recuperado de: <https://journals.sagepub.com/doi/full/10.1177/2515245918779348>
- Kelter, R. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: a non-technical introduction to Bayesian inference with JASP. *BMC Medical Research Methodology*, 20, 1-12. <https://doi.org/10.1186/s12874-020-00980-6>
- Kinney, A. R.; Eakman, A. M., & Graham, J. E. (2020). Novel Effect Size Interpretation Guidelines and an Evaluation of Statistical Power in Rehabilitation Research. *Archives of Physical Medicine and Rehabilitation*, 101(12), 2219-2226. <https://doi.org/10.1016/j.apmr.2020.02.017>
- Marsman, M., & Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, 14(5), 545-555. <https://doi.org/10.1080/17405629.2016.1259614>
- Polo-Recuero, B.; Moreno-Barrio, A., y Ordóñez-Dios, A. (2020). Lecciones activas: estrategia para aumentar la actividad física de los escolares durante la jornada lectiva. *RICYDE. Revista Internacional de Ciencias del Deporte*, 62(16), 342-357. <https://doi.org/10.5232/ricyde2020.06201>
- Ramos-Vera, C. A. (2020). Una inferencia adicional mediante el factor Bayes en análisis correlacionales. *Revista Investigación en Educación Médica*, 9(36), 103-104. <https://doi.org/10.22201/fm.20075057e.2020.36.20249>
- Ramos-Vera C. A. (En prensa). Replicación bayesiana: cuán probable es la hipótesis nula e hipótesis alterna. *Educación Médica*. <https://doi.org/10.1016/j.edumed.2020.09.014>
- Sainani, K. L.; Lohse, K. R.; Jones, P. R., & Vickers, A. (2019). Magnitude-based Inference is not Bayesian and is not a valid method of inference. *Scandinavian Journal of Medicine & Science in Sports*, 29(9), 1428-1436. <https://doi.org/10.1111/sms.13491>
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609-612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Welsh, A. H., & Knight, E. J. (2015). "Magnitude-based inference": a statistical review. *Medicine and science in sports and exercise*, 47(4), 874-884. <https://doi.org/10.1249/MSS.0000000000000451>