

Comparación de los tamaños del efecto y sus intervalos de confianza: Guía para las pruebas de equivalencias en la investigación educativa

1. INTRODUCCIÓN

La prueba de equivalencia (PE) es una técnica utilizada para comparar tamaños del efecto o medias de dos estudios diferentes a fin de determinar si serían estadísticamente equivalentes.

OBJETIVO enseñar cómo realizar las pruebas de equivalencia aplicadas a la investigación educativa, haciendo hincapié en las recomendaciones encaminadas a mejorar la calidad de la investigación.

¿Cómo funciona la PE?



H01: Difference between ES or mean original study and ES or mean replicated study \leq Lower Equivalence Limit (ΔL). This implies $p \geq \alpha$.

HA1: Difference between ES or a mean original study and ES or a mean replicated study $>$ Lower Equivalence Limit (ΔL): $p < \alpha$.

H02: Difference between ES or a mean original study and ES or a mean replicated study \geq Upper Equivalence Limit (ΔU): $p \geq \alpha$.

HA2: Difference between ES or a mean original study and ES or a mean replicated study $<$ Upper Equivalence Limit (ΔU): $p < \alpha$.

TRADUCCIÓN: Área de no-rechazo // Área de rechazo // Área de no-equivalencia // Área de equivalencia

Diferencia entre el TE o la media de un estudio original y el TE o la media de un estudio replicado

Límite de equivalencia inferior // Límite de equivalencia superior

Dos pruebas unilaterales

HA1 y HA2: dos hipótesis alternativas

H01 y H02: dos hipótesis nulas

ES: Tamaño del efecto

ΔL: Límite de equivalencia inferior

ΔU: Límite de equivalencia superior

Valor de p: probabilidad de observar el valor de un estadístico de prueba o menor en una distribución, dada una hipótesis nula verdadera

α: Nivel alpha

2. EJEMPLOS

1. Una única prueba de muestra tras un pre-test y un post-test

Límites de datos y equivalencias tomados de De Muth (2019). Objetivo: comparar los hallazgos (puntuaciones del post-test) con un valor objetivo (Media de la población = 100 puntos).

Este ejemplo ha contextualizado la PE en la investigación educativa con un diseño experimental. El problema se resolvió...

- mediante cálculos manuales utilizando valores críticos de la tabla de distribución de t .
- usando una calculadora virtual para estimar los valores de p a partir de los valores de t y SPSS 24 para Prueba de Significancia Estadística para la Hipótesis nula.

Hipótesis:

$$H_{01} : \bar{x} - \mu \leq \Delta L$$

$$H_{A1} : \bar{x} - \mu > \Delta L$$

$$H_{02} : \bar{x} - \mu \geq \Delta U$$

$$H_{A2} : \bar{x} - \mu < \Delta U$$

$\bar{x} - \mu =$ observed average - Target average (the Target average can be found in the literature or other reliable source).

μ Valor objetivo

\bar{x} Media de la muestra

Conclusión: las hipótesis nulas se rechazaron porque los valores de $p < \alpha$. Existía una equivalencia estadísticamente significativa con un intervalo de confianza del 90% entre las medias de la muestra y la población. La diferencia de medias con un IC (Intervalo de Confianza) del 95% [-0,855, 0,555], que representaba a una población, se situó entre ΔL (-1,5) and ΔU (+1,5), así que se puede considerar estadísticamente equivalente para el objetivo de 100 puntos.

2. Dos pruebas de grupo independientes tras un pre-test y un post-test

Datos extraídos de De Muth (2019). Escenario: el investigador A intentaba replicar un estudio realizado por el investigador B con el fin de observar si un tratamiento para mejorar la lectura tenía la misma relevancia en ambos estudios. Pregunta de investigación: ¿Existe una equivalencia significativa entre estas dos medias?

Ejemplo utilizando:

- SPSS 24 y cálculos manuales con valores críticos de la tabla de distribución de t y calculadoras virtuales.
- Código R creado por Lakens (2017) para las pruebas de equivalencia.

Hipótesis:

$$H_{01} : \mu_A - \mu_B \leq \Delta L$$

$$H_{A1} : \mu_A - \mu_B > \Delta L$$

$$H_{02} : \mu_A - \mu_B \geq \Delta U$$

$$H_{A2} : \mu_A - \mu_B < \Delta U$$

Si se rechazan las nulas; la equivalencia queda demostrada.

Conclusión: para un objetivo dado de 100, el límite inferior se fijó en -2 y el superior en +2. La hipótesis nula (H01) no se rechazó a causa del IC [-2,4638, 0,3638], que superaba el rango de |2|. Además, la ratio de t no servía para rechazar la hipótesis para el límite inferior $t(9) = 1,22$, $p = ,25$. Así pues, no se logró mostrar equivalencia.

3. CONCLUSIONES

La PE puede ser una herramienta útil para comparar medias y efectos dentro de ciertos límites que esperamos que pueda adquirir una relevancia práctica a la hora de dar sentido a los hallazgos.

REFERENCIAS:

De Muth, J. E. (2019). *Practical Statistics for Pharmaceutical Analysis with Minitab Applications*.

Springer. <https://doi.org/10.1007/978-3-030-33989-0>

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355-362. <https://doi.org/10.1177/1948550617697177>