



Tema 25: Pruebas de equivalencia (parte 1)

*Absence of evidence
is not evidence of absence.
Douglas G. Altman; J. Martin Bland¹*

□ Peter B. Mandeville

Una prueba de diferencia, que se emplea habitualmente en ensayos clínicos, tiene una hipótesis de nulidad (H_0), que sostiene que dos cantidades son iguales. La hipótesis alterna (H_a), afirma que las dos cantidades son diferentes. Si los datos proporcionan suficiente evidencia de que la H_a es verdadera, se rechazará la H_0 y se aceptará la H_a .¹⁻⁷

El objetivo de una prueba de equivalencia es comprobar que dos cantidades son iguales, es decir, demostrar la H_0 de una prueba de diferencia. Una prueba de diferencia es inadecuada para este propósito, porque nunca se puede aceptar la H_0 de un ensayo.¹⁻⁷

Para probar la equivalencia de dos tratamientos, la H_a debe ser que los tratamientos son iguales. Este hecho requiere una inversión de las hipótesis de una prueba de diferencia. Específicamente, una prueba de equivalencia requiere la especificación de ninguna diferencia en la H_a y una diferencia en la H_0 . Además, una prueba de equivalencia contiene un parámetro adicional delta (δ), que es la diferencia

máxima permitida para que un tratamiento experimental sea considerado equivalente a un tratamiento estándar.²⁻⁷

Las hipótesis de equivalencias bilaterales son:

$$H_0: \theta \neq \delta$$

$$H_a: \theta = \delta$$

en las que θ es la diferencia o la razón de las medias para el tratamiento experimental y el tratamiento estándar. Si θ es menor que δ , se puede aceptar la hipótesis alterna de que los dos tratamientos son iguales con un $100*(1-\alpha)\%$ nivel de confianza. Las hipótesis de equivalencia típicamente son unilaterales:

$$H_0: \theta \geq \delta$$

$$H_a: \theta < \delta$$

También, a veces, un criterio de equivalencia no simétrica es apropiado, y se utilizan dos valores de δ .²⁻⁷

Error tipo I (α) en el contexto de una prueba de diferencia es la probabilidad de concluir que un tratamiento experi-

mental es diferente a un tratamiento estándar cuando no existe ninguna diferencia, es decir, rechazar H_0 cuando es cierto. Error tipo II (β) es la probabilidad de que el tratamiento experimental no sea diferente a un tratamiento estándar, cuando en realidad hay una diferencia, es decir, no rechazar H_0 cuando es falsa.²⁻⁷

Para que una prueba de equivalencia indique que un tratamiento experimental no es diferente de un tratamiento estándar cuando no lo es, no rechazar H_0 cuando es falsa, es equivalente al error tipo II de una prueba de diferencia. Por lo tanto, no solamente las H_0 y H_a están invertidas, sino las definiciones de los errores. El error tipo I de una prueba de equivalencia es el error tipo II de una prueba de diferencia. Sin embargo, se elige la probabilidad de error tipo I (α) para una prueba de equivalencia de la forma habitual.²⁻⁷

Los tratamientos estándar y experimental deberían tener distribuciones similares para utilizar análisis paramétrico. Al contrario, debe efectuar la prueba de equivalencia con el procedimiento Mann-Whitney.^{7, 8}

Como ejemplo se utilizan los datos de un ensayo comparativo de moxonodín [un agente de bloqueo de los receptores alfa] y captopril [un inhibidor de la enzima de la angiotensina] en el tratamiento antihipertensivo de pacientes que sufren de depresión mayor. Específicamente, la variable de respuesta es la reducción de la presión arterial diastólica [mm Hg] observada en pacientes que sufren de depresión mayor después de cuatro semanas de tratamiento diario con 0.2-0.4 mg moxonodín (M) o 25-30 mg captopril (C). El objetivo del ensayo fue establecer la equivalencia y se eligió un criterio de equivalencia asimétrico para la diferencia estandarizada de las medias especificando $(\delta_1, \delta_2)=(-0.50, 1.00)$.⁷

```
> M <- c(10.3,11.3,2,-6.1,6.2,6.8,3.7,-3.3,-3.6,-
3.5,13.7,12.6)
> M
[1] 10.3 11.3 2.0 -6.1 6.2 6.8 3.7 -3.3 -3.6 -3.5 13.7 12.6
> C <- c(3.3,17.7,6.7,11.1,-5.8,6.9,5.8,3.0,6.0,3.5,18.7,9.6)
> C
```

```
[1] 3.3 17.7 6.7 11.1 -5.8 6.9 5.8 3.0 6.0 3.5 18.7 9.6
```

Para efectuar el análisis paramétrico de dos muestras independientes o pares de observaciones en R,⁹ se puede utilizar la función *tost* (*two one-sided testing*) en el paquete *equivalence*.⁸

```
> tost(M,C,epsilon=1)
```

```
$mean.diff
```

```
[1] -3.033
```

```
$se.diff
```

```
[1] 2.793
```

```
$alpha
```

```
[1] 0.05
```

```
$ci.diff
```

```
[1] -7.830 1.763
```

```
attr("conf.level")
```

```
[1] 0.9
```

```
$df
```

```
df
```

```
21.92
```

```
$epsilon
```

```
[1] 1
```

```
$result
```

```
[1] "not rejected"
```

```
$p.value
```

```
[1] 0.7629
```

```
$check.me
```

```
NULL
```



En este caso no se rechaza la H_0 . No existe evidencia que las medias de los tratamientos sean equivalentes.

Hay una idea general, falsa, de que un ensayo de equivalencia requiere una muestra más grande que un ensayo de diferencia. La veracidad de esta suposición depende del tamaño de la diferencia (δ) y del criterio de equivalencia (δ) que uno está dispuesto a permitir. La fórmula para la determinación del tamaño mínimo de la muestra con siendo el tamaño de la muestra por cada grupo es:²⁻⁶

$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2 (Vs + Ve)}{(Ms - Me - \delta)^2}$$

donde Vs y Ve son las varianzas, y Ms y Me son las medias de los tratamientos estándar y experimental. Se pueden efectuar los cálculos con el código siguiente:

```
samsiz <- function(x1,x2,delta,alpha=0.05,beta=0.2){
  m1 <- mean(x1); m2 <- mean(x2)
  v1 <- var(x1); v2 <- var(x2)
  cat("mean1:",m1,"\\n"); cat("mean2:",m2,"\\n")
  cat("varianza1:",v1,"\\n"); cat("varianza2:",v2,"\\n")
  n <- (qnorm(1-alpha/2)+qnorm(1-beta))^2 * (v1+v2)/
  (m1-m2-delta)^2
  cat("n:",n,"\\n")
}

> samsiz(M,C,.1)
mean1: 4.175
mean2: 7.208333
varianza1: 49.70386
varianza2: 43.88629
n: 74.8212
```

Se necesitará un mínimo de 75 repeticiones por tratamiento o un total de 150 unidades experimentales. Las recomendaciones para reportar los resultados de una prueba de equivalencia están dadas por Piaggio, Elbourne, Altman, Pocock y Evans.¹⁰

Referencias

1. Douglas G. Altman y J. Martin Bland. 1995. Absence of evidence is not evidence of absence. BMJ volume 311, 19 August 1995, p. 485.
2. Pamela Atherton-Skaff y Jeff Sloan. 1998. Design and Analysis of Equivalence Trials via the SAS System. Proceedings of the Twenty-Third Annual SAS Users Group International. Paper 219. March 22-25, 1998, Nashville, TN, USA.
3. William C. Blackwelder. 1982. "Proving the Null Hypothesis" in Clinical Trials. Controlled Clinical Trials 3:345-353.
4. Dieter Hauschke, Volker Steinijans y Iris Pigeot. 2007. Bioequivalence Studies in Drug Development: Methods and Applications. Statistics in Practice. John Wiley & Sons, Ltd., Chichester, UK.
5. John M. Hoenig y Dennis M. Heisey. 2001. The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis. The American Statistician. February 2001. Vol 55, No. 1, pp. 19-24.
6. Scott Patterson y Byron Jones. 2006. Bioequivalence and Statistics in Clinical Pharmacology. Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton, FL, USA.
7. Stefan Wellek. 2003. Testing Statistical Hypotheses of Equivalence. Chapman & Hall/CRC, Boca Raton, FL, USA.
8. Andrew Robinson A. Robinson@ms.unimelb.edu.au . 2010. equivalence: Provides tests and graphics for assessing tests of equivalence. R package version 0.5.6. <http://CRAN.R-project.org/package=equivalence> .
9. R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
10. Gilda Piaggio, Diana R. Elbourne, Douglas G. Altman, Pocock, S.J., y Evans, S.J.W. (2006). Reporting of Noninferiority and Equivalence Randomized Trials: An Extension of the Consort Statement. JAMA 2006;295(10):1152-1160.