

# Errores estadísticos y potencia: desafíos en la investigación actual

## Resumen

El presente artículo examina la naturaleza, los mecanismos y las consecuencias de los errores de tipo I ( $\alpha$ ) y tipo II ( $\beta$ ) dentro del marco de las pruebas de hipótesis nula, analizando la transición histórica desde los modelos de Fisher hacia la síntesis de Neyman-Pearson. Se aborda la relación inversa entre ambos errores y cómo factores como el tamaño de la muestra, la variabilidad de los datos y el tamaño del efecto determinan la potencia estadística. Mediante el uso del lenguaje de programación R y la librería ggplot2, se desarrollan simulaciones visuales para ilustrar la superposición de distribuciones y la generación de curvas de potencia acordes a los estándares de la investigación cuantitativa. Se concluye que el equilibrio entre falsos positivos y falsos negativos es una decisión ética y metodológica, que influye directamente en la actual crisis de replicación científica.

## Abstract

This article examines the nature, mechanisms, and consequences of type I ( $\alpha$ ) and type II ( $\beta$ ) errors within the framework of null hypothesis testing, analyzing the historical transition from Fisher's models to the Neyman-Pearson synthesis. The inverse relationship between the two errors is addressed and how factors such as sample size, data variability, and effect size determine statistical power. Utilizing the R programming language and the ggplot2 library, visual simulations are developed to illustrate the overlapping of distributions and the generation of power curves according to the standards of quantitative research. It is concluded that the balance between false positives and false negatives is an ethical and methodological decision, which directly influences the current crisis of scientific replication.

## Introducción

La inferencia estadística es una herramienta indispensable en la investigación, pues permite extrapolar los hallazgos observados en una muestra para formular conclusiones aplicables a una población más amplia (Casella & Berger, 2002). No obs-

Julio César Robles Venzor<sup>1</sup>.  
Elvia Rosario Chávez Tellez<sup>1</sup>.  
Miriam Gabriela Flores Granados<sup>1</sup>.

- (1) Facultad de Ciencias Químicas/  
Universidad Autónoma de Chihuahua

Recibido: 11 de mayo de 2026

Aceptado: 13 de junio de 2026

tante, al analizar datos bajo condiciones de incertidumbre, este proceso conlleva márgenes de error inherentes. Un componente central de este análisis es la prueba de hipótesis, cuyo propósito es evaluar si la evidencia experimental tiene el peso suficiente para rechazar una premisa base, conocida como hipótesis nula ( $H_0$ ). Puesto que estas decisiones se fundamentan en modelos probabilísticos, el investigador asume dos riesgos metodológicos: reportar un efecto inexistente u omitir la detección de un efecto real (Neyman & Pearson, 1933). En la literatura cuantitativa, estas situaciones se definen formalmente como el error tipo I ( $\alpha$ ) y el error tipo II ( $\beta$ ), respectivamente (García-Garzón et al., 2018).

El diseño estadístico de experimentos vivió un cambio de paradigma a principios del siglo XX gracias a las contribuciones de Sir Ronald A. Fisher, evolucionando de la mera acumulación de datos a un método sistemático de inferencia. Según Pantoja (2023), su obra de 1935, *The Design of Experiments*, demostró que la validez de las conclusiones no depende exclusivamente de un volumen masivo de información, sino de la pertinencia del método empleado. Fisher evidenció que las deficiencias analíticas frecuentemente se originan en un planteamiento experimental inadecuado. Por consiguiente, propuso un marco metodológico riguroso que permite evaluar múltiples variables de forma simultánea, minimizando los sesgos y garantizando que los hallazgos cuenten con un soporte estadístico robusto (Amrhein et al., 2019; Ato et al., 2013).

Posteriormente, la síntesis teórica de Neyman y Pearson redefinió la estadística empírica, trasladándola de una simple evaluación de evidencia hacia una metodología formal para la toma de decisiones. Este avance introdujo la hipótesis alternativa ( $H_a$ ) y estableció el concepto de regiones de rechazo en función de la probabilidad de error a largo plazo (Gelman & Loken, 2014; Cumming, 2014). Dicho esquema resulta crítico en disciplinas aplicadas como la investigación biomédica, donde incurrir en un error tipo I podría derivar en la aprobación de tratamientos ineficaces, mientras que un error tipo II significaría descartar terapias potencialmente benéficas (Ioannidis, 2018; Gómez-Gómez et al., 2013).

Durante la última década, se ha evidenciado que el manejo inadecuado de la potencia estadística —entendida como el complemento del error tipo II— representa uno de los factores principales detrás de la denominada “crisis de replicación” (Domínguez-Lara, 2018; Ioannidis, 2018). Las investigaciones con baja potencia no solamente presentan dificultades para detectar efectos reales, sino que tienden a sobrestimar la magni-

tud de los resultados que resultan significativos, un sesgo metodológico documentado en la literatura como la “maldición del ganador” (Button et al., 2013; Ioannidis, 2018).

## **Materiales y Métodos**

Para la evaluación del comportamiento de las decisiones estadísticas, se empleó una metodología basada en la simulación computacional de distribuciones muestrales utilizando el lenguaje de programación R, el cuál es un sistema de código abierto, con amplia versatilidad para manejar datos y el ecosistema de paquetes tidyverse y pwr, que son pequeños programas capaces de realizar análisis estadísticos de alto nivel, pero fácilmente programables a las necesidades de cualquier investigador (Champely, 2020; R-bloggers, 2016; Wickham et al., 2019). Con este software se simularon datos para obtener un conjunto de valores con los cuales realizar un diseño experimental (Anexo 1) mediante el que se pretendió demostrar la relación que guardan los valores que representan la distribución (parámetros de contraste) con el tamaño de la muestra y la capacidad de probar un resultado esperado (hipótesis alternativa) válido, sustentado en los valores estadísticos, mientras se rechaza el resultado donde no existió un cambio (hipótesis nula). Por lo tanto, es fundamental para todo aquel que realice un análisis estadístico con la finalidad de discernir entre dos resultados, saber interpretar el valor de error que tiene esta prueba y diferenciarlo de las dos posibilidades: error tipo I, que es rechazar una hipótesis nula que es verdadera, o error tipo II, es decir, aceptar una hipótesis nula que es falsa.

Dado que la simulación requería modelar escenarios teóricos bajo la hipótesis nula ( $H_0$ ) y la hipótesis alternativa ( $H_a$ ), se asumieron distribuciones normales, o datos en forma de campana de Gauss, para el estadístico de prueba (Casella & Berger, 2002). Sin embargo, dado que es imposible tomar un infinito de valores, se selecciona una muestra pequeña, que incluya valores que describan de manera fidedigna toda la curva, es decir, una muestra significativa con la cual se va a trabajar, y de esta, es imprescindible considerar el nivel de significancia ( $\alpha$ ), que es la certeza de que hubo cambios por una causa específica, y no por el simple azar. Durante el desarrollo de esta simulación se mantuvo dicha significancia ( $\alpha$ ) en un nivel constante de 0.05, reconocido internacionalmente como valor estándar. Para evidenciar el cambio en los valores de error respecto al tamaño de la muestra, se varió el tamaño de la muestra ( $n$ ), la diferencia entre el valor esperado y el hipotetizado (Diferencia de medias) y la variabilidad de los datos entre sí mismos, ex-

presada mediante la desviación estándar ( $\sigma$ ).

En la primera etapa de la modelación, se estructuraron dos funciones de densidad de probabilidad para representar los contrastes de hipótesis: una distribución normal estándar bajo  $H_0$  definida con una media  $\mu_0 = 0$ , y una distribución normal desplazada bajo  $H_a$  con una media  $\mu_1 = 2$ . En ambos modelos teóricos se asumió una desviación estándar conocida y constante  $\sigma = 1$ . El dominio de la variable continua  $x$  se discretizó en un vector uniforme de 1,000 observaciones dentro del intervalo  $[-4, 6]$  para asegurar la resolución y suavidad de las curvas de densidad resultantes. El criterio de decisión estadística se estableció fijando un nivel de significancia nominal  $\alpha = 0.05$ . A partir de este parámetro, se determinó el límite exacto de la región de rechazo para una prueba de una cola calculando el valor crítico teórico ( $Z_{\text{crit}}$ ) mediante la función de cuantiles de la distribución normal acumulada bajo  $H_0$ , correspondiente al percentil 95% ( $1 - \alpha$ ).

La integración y visualización de estas regiones de decisión se realizó mediante el paquete `ggplot2` (Wickham, 2016), construyendo un modelo gráfico de alta fidelidad con capas geométricas de líneas para las funciones de densidad y capas de área para delimitar las probabilidades de error. El error tipo I ( $\alpha$ ) se sombreó bajo la curva de  $H_0$  para el sector donde  $x \geq Z_{\text{crit}}$ , reflejando el riesgo de rechazar incorrectamente una hipótesis nula verdadera. Por su parte, el error tipo II ( $\beta$ ) se estimó integrando el área bajo la curva de  $H_a$  a la izquierda del valor crítico determinado por la regla de decisión (Gelman & Carlin, 2014), visualizándose en el intervalo  $x \leq Z_{\text{crit}}$  para proyectar la probabilidad de no rechazar la hipótesis nula siendo esta falsa. El umbral de separación entre ambas regiones se marcó explícitamente mediante un vector vertical discontinuo en la coordenada de dicho valor crítico.

En la segunda etapa, se evaluó cómo interactúan en la práctica las variaciones paramétricas sobre la sensibilidad del contraste mediante un análisis de potencia computacional. Para ello, se analizó el comportamiento de la potencia estadística frente al tamaño de la muestra considerando tres niveles de diferencia de medias ( $\delta = 3, 5, 7$ ) y tres niveles de variabilidad basados en la desviación estándar ( $\sigma = 8, 10, 12$ ). Los tamaños de muestra ( $n$ ) se fijaron en un rango desde las 5 hasta las 100 observaciones, incrementándose de manera gradual en intervalos de 5 en 5. Se construyó una matriz con todas las combinaciones posibles de estos factores para ejecutar un algoritmo automatizado que implementó la función de la prueba  $t$  de Student para una sola muestra sobre cada escenario, realizando cálculos exactos

fundamentados en las propiedades de esta distribución (Ferrando & Lorenzo-Seva, 2014). En cada interacción, la magnitud del efecto se estandarizó bajo el criterio de la  $d$  de Cohen, calculada dinámicamente como el cociente de la diferencia de medias entre la desviación estándar ( $d=\delta/\sigma$ ).

Este procedimiento matemático se configuró manteniendo el nivel de significancia estático en  $\alpha=0.05$ , obteniéndose la potencia estadística resultante ( $1-\beta$ ) entendida como la probabilidad matemática de detectar un efecto real cuando este verdaderamente existe. Finalmente, las curvas de potencia simuladas se unificaron gráficamente con ggplot2. En la representación, se dispuso el tamaño de la muestra ( $n$ ) en el eje horizontal y la potencia calculada ( $1-\beta$ ) en el eje vertical. Para asegurar una lectura intuitiva de la interacción multifactorial, las diferencias de medias ( $\delta$ ) se codificaron mediante una escala de colores, mientras que los niveles de variabilidad ( $\sigma$ ) se asignaron a distintos estilos de línea. Como elemento de validación metodológica, se incorporó una asíntota horizontal en el valor de 0.80, la cual funge como el umbral normativo convencionalmente aceptado en la investigación científica para dictaminar si un estudio posee un tamaño de muestra confiable y con la sensibilidad adecuada.

## Resultados y discusiones

Para obtener la potencia de la prueba respecto al tamaño de muestra, se generaron diferentes escenarios manteniendo los valores de  $\sigma$  y  $\delta$  en tres niveles arbitrarios fijos (8, 10 y 12 para  $\sigma$ ; 3, 5 y 7 para  $\delta$ ) y un nivel de significancia ( $\alpha$ ) constante de 0.05, iterando en los valores de  $n$  entre 5 y 100, cuyas dispersiones se graficaron y se contrastan entre sí para determinar que si se busca una potencia mínima de 0.8, es necesario maximizar el valor  $\delta$  mientras se minimiza  $\sigma$ , de manera que  $n$  sea mínimo, y en contraparte, si se tiene un  $\delta$  pequeño y una desviación estándar grande, será necesario aumentar el valor de  $n$  para llegar a la meta de potencia mínima de 0.8. Este juego matemático conlleva que se debe contemplar desde el diseño del experimento, el tamaño de  $n$  y su consecuente aumento en repeticiones que se refleja en mayor costo en tiempo y dinero. Estos resultados concuerdan con lo obtenido por Cohen (1988), quien menciona que si se busca una potencia fuerte (0.80) en un experimento, se debe aumentar el tamaño de la muestra o asegurarse de que el efecto buscado sea lo suficientemente grande para sobresalir por encima del error.

De manera arbitraria, al inicio del experimento se define el valor de  $\alpha$  (Error tipo I), siendo generalmente de 0.05, que es

Decisión del Investigador	$H_0$ es Realmente Verdadera	$H_0$ es Realmente Falsa ( $H_a$ es Cierta)
No Rechazar $H_0$	Decisión Correcta ( $1 - \alpha$ )	Error tipo II ( $\beta$ )
Rechazar $H_0$	Error tipo I ( $\alpha$ )	Decisión Correcta ( $1 - \beta$ ) (Potencia)

el umbral comúnmente establecido para tomar una decisión respecto a la hipótesis nula, ya sea rechazarla o no rechazarla. Y el valor de  $\beta$  (Error tipo II) regularmente ignorado, indica la probabilidad de no rechazar la hipóte-

**Cuadro 1. Relación entre realidad poblacional y conclusión del estudio.**

sis nula cuando realmente el falsa (Benjamin et al., 2018). Numéricamente, los valores de  $\beta$  para indicar una buena potencia de la prueba son de máximo 0.2, ya que la potencia de prueba es  $1 - \beta$  (Cuadro 1).

Utilizando ggplot2, se visualiza cómo el valor crítico segmenta las regiones de error en la gráfica 1. El área bajo  $H_0$  a la derecha representa  $\alpha$ , y el área bajo  $H_a$  a la izquierda representa  $\beta$  (Halsey et al., 2015; Monterrey, 2012)

Siguiendo la lógica de la Gráfica 1, se graficaron curvas representadas en la Gráfica 2.

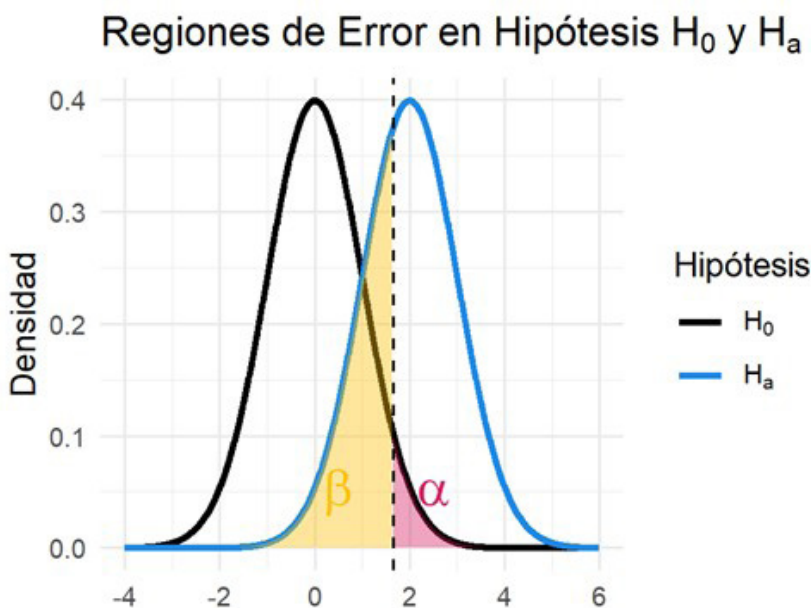
En esta, se observa que la potencia aumenta con el tamaño de la muestra y el tamaño del efecto, y disminuye con la varianza (Cohen, 1988; Wasserstein & Lazar, 2016).

En la estadística tradicional, existe una fijación por el valor  $p$  que, por su facilidad de uso, en ocasiones causa que se olvide lo que realmente significan los datos. El problema inherente a esto es que, si existe una muestra de gran tamaño, hasta la diferencia más insignificante puede parecer estadísticamente significativa, lo cual es una posible fuente de error. Por eso es de vital importancia tomar en cuenta el tamaño del efecto, tal cual lo establece Cohen en 1988.

**Gráfica 1. Visualización del solapamiento de distribuciones.**

Como mencionan Gutiérrez y De La Vara (2008), un buen experimento no solo trata de reducir el error, sino de tener la sensibilidad necesaria para detectar cambios que realmente impacten a la realidad de la muestra. Cohen en 1988 diseñó este parámetro con la finalidad de estandarizar esas diferencias y permitir comparar estudios distintos sin importar sus unidades de medida.

Según los resultados obtenidos en este experimento, que son concordantes con Cohen



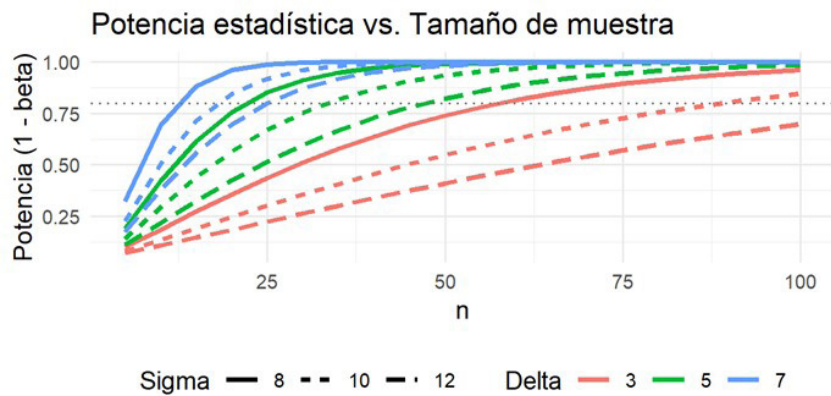
(1988), la relación entre este valor y la potencia estadística es directa, puesto que, si el efecto es pequeño, se necesitará una muestra mucho más grande para tener seguridad de no cometer errores tipo I o II (Lakens et al., 2018). Por tanto, limitar una decisión estadística a solo el rechazo o no rechazo de una hipótesis es una visión sobre simplificada, pues es necesario usar los umbrales de Cohen (0.2 para efectos pequeños, 0.5 para medianos y 0.8 para grandes) para alejarse de esa decisión binaria y plantear una imagen real de qué tan relevante es el hallazgo estadístico (Cuadro 2).

Determinar la ponderación entre el error tipo I y el error tipo II trasciende el cálculo probabilístico, constituyendo una decisión basada en las implicaciones prácticas de cada fallo inferencial. En el análisis experimental, existe un balance asimétrico: en ciertos contextos resulta más perjudicial reportar un falso positivo, mientras que, en otros, la omisión de un efecto verdadero conlleva consecuencias más severas. Esta dualidad exige que el diseño de la investigación se aborde como una justificación metodológica rigurosa, más allá de la adopción mecánica de criterios estandarizados.

Actualmente, existe un debate considerable en la literatura sobre la pertinencia de endurecer los criterios de evidencia, particularmente mediante la reducción del nivel de significancia tradicional de 0.05 a 0.005. El propósito central de esta medida es incrementar la reproducibilidad y la robustez de los hallazgos empíricos. No obstante, las proyecciones demuestran que esta modificación metodológica conlleva exigencias logísticas sustanciales. Para preservar la capacidad de identificar fenómenos reales manteniendo una potencia estadística del 80%, sería indispensable incrementar el tamaño de la muestra en aproximadamente un 70%.

**Conclusiones**

El análisis inferencial evidencia la relación inversamente proporcional que existe entre las probabilidades de incurrir en errores de tipo I y tipo II. Bajo condiciones estándar, la minimización de un riesgo



**Gráfica 2. Potencia estadística contra tamaño de muestra.**

**Cuadro 2. Magnitudes del efecto d de Cohen**  
Fuente: Cohen, 1988.

Magnitud	d de Cohen	Interpretación
Pequeño	0.20	Difícil de detectar sin muestras grandes.
Mediano	0.50	Detectable en gráficos de dispersión.
Grande	0.80	Diferencia sustancial y obvia.

conlleva inexorablemente el incremento del otro. La principal estrategia metodológica para atenuar ambos márgenes de error de manera simultánea radica en la optimización del diseño experimental, fundamentalmente a través del incremento del tamaño muestral y la reducción del error de medición.

En este sentido, la estimación a priori del tamaño de la muestra, fundamentada en la magnitud del efecto esperado, constituye un requerimiento indispensable para garantizar la solidez empírica de los hallazgos. Lejos de representar un mero formalismo técnico, este procedimiento asegura que la investigación alcance la potencia estadística necesaria para detectar efectos científicamente relevantes, mitigando la incertidumbre inherente al contraste de hipótesis.

En definitiva, el rigor metodológico exige trascender la aplicación mecánica de pruebas de significancia. La evaluación consciente de las consecuencias derivadas de los falsos positivos y los falsos negativos transforma el análisis cuantitativo en un proceso de toma de decisiones fundamentado. Esta perspectiva crítica garantiza que las inferencias extraídas sean válidas, reproducibles y contribuyan de manera efectiva al avance del conocimiento científico.

## Referencias Bibliográficas

- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305-307. <https://doi.org/10.1038/d41586-019-00857-9>
- Ato, M., López-García, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. <https://dx.doi.org/10.6018/analesps.29.3.178511>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Johnson, V. E. (2018). Re-define statistical significance. *Nature Human Behaviour*, 2(1), 6-10. <https://doi.org/10.1038/s41562-017-0189-z>
- Burbano Pantoja, V. M. A., Valdivieso Miranda, M. A., & Burbano Valdivieso, Á. S. (2023). La estadística no paramétrica en los libros de texto universitarios: situaciones problema para promover la investigación científica desde el aula. Editorial UPTC. <https://doi.org/10.19053/9789586607308>
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365-376. <https://doi.org/10.1038/nrn3475>
- Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury/Thomson Learning. ISBN: 0534243126.

- Champely, S. (2020). pwr: Basic functions for power analysis (Versión 1.3-0) [Software de computadora]. Comprehensive R Archive Network. <https://CRAN.R-project.org/package=pwr>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29. <https://doi.org/10.1177/0956797613504966>
- Domínguez-Lara, S. A. (2018). Magnitud del efecto, una guía rápida. *Educación Médica*, 19(4), 251-254. <https://doi.org/10.1016/j.edumed.2017.07.002>
- Ferrando, P. J., & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: algunas consideraciones prácticas y revisadas. *Anales de Psicología*, 30(3), 1151-1169. <https://dx.doi.org/10.6018/analesps.30.3.199991>
- García-Garzón, E., Lecuona, O. y Carbajal, G. V. (2018). Estudios de replicación, pre-registros y ciencia abierta en Psicología. *Apuntes de Psicología*, 36(1-2), 75-83. <https://doi.org/10.55414/ap.v36i1-2.713>
- Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, 9(6), 641-651. <https://doi.org/10.1177/1745691614551642>
- Gelman, A., & Loken, E. (2014). The statistical crisis in science: data-dependent analysis—a “garden of forking paths”—explains why many statistically significant comparisons don't hold up. *American Scientist*, 102(6), 460-466. <https://doi.org/10.1511/2014.111.460>
- Gómez-Gómez, M., Danglot-Banck, C., & Vega-Franco, L. (2013). Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. *Revista Mexicana de Pediatría*, 80(2), 91-99. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=8084>
- Gutiérrez Pulido, H., & de la Vara Salazar, R. (2008). *Análisis y diseño de experimentos* (2ª ed.). México, D.F.: McGraw-Hill/Interamericana Editores. ISBN: 9789701065266.
- Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, 12(3), 179-185. <https://doi.org/10.1038/nmeth.3288>
- Ioannidis, J. P. A. (2018). The proposal to lower P value thresholds to .005. *JAMA*, 319(14), 1429-1430. <https://doi.org/10.1001/jama.2018.1536>
- Lakens, D., Adolfs, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., ... & Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, 2(3), 168-171. <https://doi.org/10.1038/s41562-018-0311-x>
- Monterrey, P. (2012). ¿Criterio mágico para resolver cualquier problema o leyenda urbana? *Universitas Scientiarum*, 17(2), 20–25. <https://dialnet.unirioja.es/descarga/articulo/210715.pdf>
- Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions*

of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 231(694-706), 289–337. <https://doi.org/10.1098/rsta.1933.0009>

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129-133. <https://doi.org/10.1080/00031305.2016.1154108>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer-Verlag. <https://doi.org/10.1007/978-3-319->

## Anexo 1

Código de R utilizado:

```
library(ggplot2)
library(dplyr)
library(pwr)
# Parámetros técnicos
mu0 <- 0; mu1 <- 2; sd_val <- 1; alpha_val <- 0.05
x <- seq(-4, 6, length.out = 1000)
df <- data.frame(x, y0 = dnorm(x, mu0, sd_val), y1 = dnorm(x, mu1, sd_val))
crit_val <- qnorm(1 - alpha_val, mu0, sd_val)
# Gráfico de distribuciones y errores
ggplot(df, aes(x = x)) +
  geom_line(aes(y = y0, color = "H0"), size = 1) +
  geom_line(aes(y = y1, color = "H1"), size = 1) +
  geom_area(data = filter(df, x >= crit_val), aes(y = y0), fill = "#D81B60", alpha = 0.4)
+
  geom_area(data = filter(df, x <= crit_val), aes(y = y1), fill = "#FFC107", alpha = 0.4)
+
  geom_vline(xintercept = crit_val, linetype = "dashed") +
  annotate("text", x = 2.5, y = 0.05, label = expression(alpha), color = "#D81B60", size
= 6) +
  annotate("text", x = 0.5, y = 0.05, label = expression(beta), color = "#FFC107", size
= 6) +
  scale_color_manual(
    values = c("H0" = "black", "H1" = "#1E88E5"),
    labels = c(expression(H[0]), expression(H[a]))
  ) +
  labs(
    title = expression(paste("Regiones de Error en Hipótesis ", H[0], " y ", H[a])),
    x = "",
    y = "Densidad",
    color = "Hipótesis"
  ) +
  theme_minimal()
#Potencia Estadística vs. Tamaño de Muestra
library(pwr)
library(ggplot2)
# Simulación multivariante
n_range <- seq(5, 100, by = 5)
deltas <- c(3, 5, 7); sigmas <- c(8, 10, 12)
params <- expand.grid(n = n_range, delta = deltas, sigma = sigmas)
params$power <- mapply(function(n, d, s) pwr.t.test(n=n, d=d/s, sig.level=0.05,
type="one.sample")$power, params$n, params$delta, params$sigma)
# Gráfica final
ggplot(params, aes(x = n, y = power, color = factor(delta), linetype = factor(sigma))) +
  geom_line(size = 1) +
  geom_hline(yintercept = 0.8, linetype = "dotted", color = "grey50") +
  labs(title = "Potencia estadística vs. Tamaño de muestra",
    x = "n", y = "Potencia (1 - beta)", color = "Delta", linetype = "Sigma") +
  theme_minimal() + theme(legend.position = "bottom")
```

