



Lo que no se dijo...



Regla del $n > 30$

Como hemos visto, esta regla tiene varias suposiciones:

- La población es infinita (o muy grande)
- Los datos fueron obtenidos con muestreo aleatorio simple
- La población tiene una distribución “razonable”

Sin embargo, estas **condiciones suelen olvidarse**, al igual que para muchas otras reglas y métodos clásicos.

Peor, debemos recordar que los métodos clásicos de inferencia se basan en **resultados asintóticos...**

--- Hesterberg, T. (2008). It's Time To Retire the " $n \geq 30$ " rule.
Proceedings of the American Statistical Association
Inferencia y Modelos Estadísticos 2



Regla del $n > 30$

Como el teorema del límite central y la ley de los grandes números...

- la distribución de un estimador **tiende** a una distribución normal (o t) a medida que el tamaño de la muestra **crece a infinito**...

Estas **aproximaciones teóricas** fueron desarrolladas por Ronald Fisher, Jerzy Neyman and Egon Pearson entre 1920 y 1935 aproximadamente.

Mucho antes de que existieran computadores.

¿Se justifica seguir con esta regla?

--- Hesterberg, T. (2008). It's Time To Retire the " $n \geq 30$ " rule.
Proceedings of the American Statistical Association



Regla del $n > 30$

Pareciera que no. Hesterberg (2008), por ejemplo, mostró que:

- existen alternativas basadas en **remuestreo** que hoy son viables y fáciles de utilizar
- **diagnósticos** con remuestreo son más efectivos que la regla clásica
- se logran **inferencias más precisas**, aunque a veces se necesitan muchas repeticiones (>60.000) y no mejores métodos se necesitan para muestras muy

--- Hesterberg, T. (2008). It's Time To Retire the " $n \geq 30$ " rule.
Proceedings of the American Statistical Association
Inferencia y Modelos Estadísticos 4

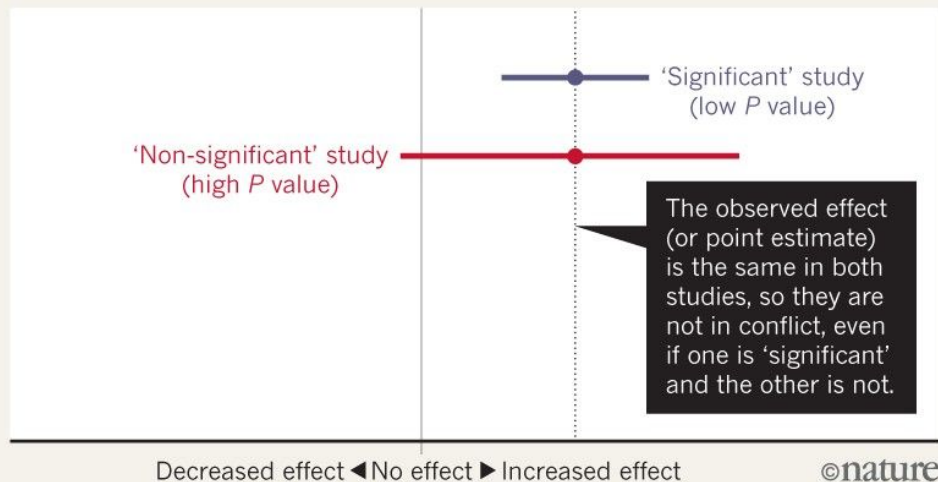


Críticas al p-valor

Centrarse en p-valores ha sido y todavía es **controversial**.

BEWARE FALSE CONCLUSIONS

Studies currently dubbed 'statistically significant' and 'statistically non-significant' need not be contradictory, and such designations might cause genuine effects to be dismissed.



“After 4 decades of severe criticism, the ritual of null hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists.”

--- Cohen, J. (1994). The Earth is round ($p < .05$). American psychologist, 49(12).

--- Amrhein, V., Greenland, S., & McShane, B. (2019). Retire statistical significance. NATURE, 567.



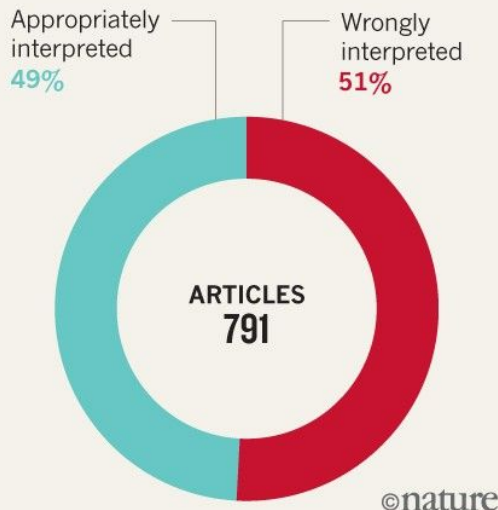
Críticas al p-valor

Nunca se debe concluir que “**no hay diferencia** o **asociación**” solamente porque un p-valor es menor que un umbral como .05.

WRONG INTERPRETATIONS

An analysis of 791 articles across 5 journals* found that around half mistakenly assume non-significance means no effect.

*Data taken from: P. Schatz *et al.* *Arch. Clin. Neuropsychol.* **20**, 1053–1059 (2005); F. Fidler *et al. Conserv. Biol.* **20**, 1539–1544 (2006); R. Hoekstra *et al. Psychon. Bull. Rev.* **13**, 1033–1037 (2006); F. Bernardi *et al. Eur. Sociol. Rev.* **33**, 1–15 (2017).



Sin embargo, se han encontrado **cientos de artículos científicos** cometiendo estos errores de interpretación en diferentes estudios y en **diferentes áreas** del saber.

--- Amrhein, V., Greenland, S., & McShane, B. (2019). Retire statistical significance. *NATURE*, 567.



Críticas al p-valor

Esto ha llevado a **desperdiciar** muchos recursos innecesariamente.

Se recomiendan las **3R** (¡no el pisco!):

1. **Randomize** data production (muestras aleatorias, grupos aleatorios)
2. **Repeat** by simulation (hasta ver qué resultados son típicos y cuáles no)
3. **Reject** any model that puts your data in its tail (contradice los datos)

--- Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum?
Technology innovations in statistics education, 1(1).



Críticas a la hipótesis nula

Pero esta interpretación errada tiene raíz en el **uso de hipótesis nulas**.

De acuerdo a Cohen (1994), las pruebas de hipótesis **no nos dicen** lo que queremos saber:

- **queremos**: dados estos datos, ¿cuál es la probabilidad que H_0 es verdadera?
- **nos dice**: Si H_0 es verdadera, ¿con qué probabilidad veríamos estos datos o datos más extremos?

Y sabemos desde hace mucho que **$P(D|H_0) \neq P(H_0|D)$** .

--- Cohen, J. (1994). The Earth is round ($p < .05$). American psychologist, 49(12).



Críticas a la hipótesis nula

Veamos un ejemplo:

La incidencia de esquizofrenia en adultos es $\approx 2\%$. Se estima que un test puede detectar $\approx 95\%$ de los casos positivos (sensibilidad) y $\approx 97\%$ de los negativos (especificidad).

Así: H_0 : el caso es normal, H_1 : el caso es esquizofrenia, T^+ : el test resulta positivo para esquizofrenia (dato).

Luego: $P(T^+|H_0) < .05$, por lo que se puede rechazar H_0 en favor de H_1 (con 95% confianza).

Pero, aunque a muchos le parezca así o quieran que sea así, esto **no nos dice** cuál es la probabilidad de observar un caso normal a pesar de tener un examen positivo: $P(H_0|T^+)$.

--- Cohen, J. (1994). The Earth is round ($p < .05$). American psychologist, 49(12).



Críticas a la hipótesis nula

Veamos un ejemplo:

Usando **estadística Bayesiana**, se puede estimar que $P(H_0|T^+) = .607$!

Puede que sea más fácil ver esta probabilidad en una tabla con 1.000 datos:

	Normal	Esquiz.	Total
T^-	949	1	950
T^+	30	20	50
Total	979	21	1.000

Esto ocurre los casos de esquizofrenia son poco frecuentes, $P(H_0) = .98$.-

--- Cohen, J. (1994). The Earth is round ($p < .05$). American psychologist, 49(12).



Críticas a la hipótesis nula

También se ha criticado porque muchas veces no tiene sentido.

Una verdadera H_0 : no hay diferencia, igual probabilidad, no hay correlación;
siempre es falsa.

“It is foolish to ask ‘Are the effects of A and B different?’ They are always different for some decimal place” (Tukey, 1991)

--- Tukey, J. W. (1991). The Philosophy of Multiple Comparisons. Statistical Science, 6(1)..



Críticas más antiguas

Desde Berkson (1938) que hay personas que cuestionan el uso de la prueba de hipótesis en la ciencia...

Some Difficulties of Interpretation Encountered in the Application of the Chi-Square Test

Joseph Berkson

Journal of the American Statistical Association, Vol. 33, No. 203. (Sep., 1938), pp. 526-536.

Journal of the American Statistical Association is currently published by American Statistical Association.





Críticas más antiguas

La principal tiene que ver con la idea de **probabilidad \approx frecuencia**.

La teoría se formuló con monedas, dados, cartas y bolas de colores... que pueden generar **variables aleatorias** asociadas a **distribuciones de probabilidad**.

Luego, como vimos con la interpretación de los intervalos de confianza:

Si un parámetro tiene un **valor fijo**,
¿por qué se trata como una variable aleatoria?



¿Alternativas?

Varias, la más popular es la **estadística Bayesiana**. Las 3R como alternativa a medio camino.

Pero no vamos a entrar en eso este semestre.

Última reflexión:

Question: Why, then, is the t-test the centerpiece of the introductory statistics curriculum? Answer: The t-test is what scientists and social scientists use most often.

Question: Why does everyone use the t-test? Answer: Because it's the centerpiece of the introductory statistics curriculum.

--- Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum?.
Technology innovations in statistics education, 1(1).



¿Dudas?
¿Comentarios?