

Tamaño del efecto

Effect size

Mario Luis Iovaldi*

La siguiente es una respuesta del chat GPT de Microsoft Bing a la pregunta: ¿Qué es el tamaño del efecto?

El tamaño del efecto es una medida de la fuerza de un fenómeno en estadística. Es una estadística descriptiva que transmite la magnitud estimada de una relación sin hacer ninguna declaración acerca de si la relación aparente en los datos reflejan una verdadera relación en la población¹. El tamaño del efecto se refiere a una forma de cuantificar la magnitud de la diferencia entre dos grupos².

Existen tres formas de medir el tamaño del efecto, según el tipo de análisis que esté realizando: 1. Diferencia de medias estandarizada 2. Coeficiente de correlación 3. Odds ratio².

Las citas 3 y 4 están sugeridas, sin embargo, no referenciadas en la respuesta.

El tema de la inteligencia artificial está muy de moda en la literatura médica y va a ser un motivo de discusión en los ámbitos científicos. Sin embargo, mi idea es dejar esto para otra ocasión y aprovechar la oportunidad para introducir el concepto y la importancia del tamaño del efecto (*effect size*) en los estudios de investigación clínica en contraste con el informe del valor p, que es una medida de la probabilidad de los resultados se deban al azar y solamente si el estudio está correctamente diseñado en cuanto a las condiciones de aplicabilidad.

De la respuesta del chat, resalto el concepto de la magnitud de la diferencia entre dos grupos, porque un valor bajo de p puede coincidir con un tamaño del efecto sin relevancia clínica.

Existen distintas medidas del tamaño del efecto para comparar 2 medias; la más conocida es la d de Cohen, otras son la g de Hedges y la delta de Glass³.

Dejo el riesgo relativo, relación de probabilidades (*odds ratio*), prueba de correlación, etc., para más adelante porque exceden el objetivo de este editorial y prefiero introducir conceptos lentamente.

La fórmula de d es $\text{media 1} - \text{media 2} / \text{desviación estándar del total de la muestra (sd pooled)}$ ⁴. La interpretación es 0,2 efecto pequeño, 0,5 efecto

mediano y 0,8 o más, efecto grande². Si el resultado fuera negativo, aunque matemáticamente correcto, no se usa para facilitar su interpretación. Es conocida también como diferencia de medias estandarizada y se denomina así porque las unidades del resultado son el número de desviaciones estándares en que se diferencian las medias. Es la moneda corriente en los metanálisis.

Un ejemplo simple en R⁵⁻⁸

Solamente los datos relevantes

```
mean (data1) # media grupo 1
[1] 41.21997
mean (data2) # media grupo 2
[1] 61.2
sd_pooled (data1, data2) # desviación estándar del
total de la muestra
[1] 27.63153
```

```
t. test (data2, data1) # Prueba (test) t de Student
de las 2 muestras sin significación estadística
```

```
t = 1.6169, df = 17.724, p-value = 0.1236 # Welch
Two Sample t-test
```

Cálculo de la d de Cohen según fórmula

```
mean (data2) - mean (data1) / sd_pooled (data1,
data2)
[1] 0.7230879
```

Cálculo de la d de Cohen con R, library (effect size)

```
cohens_d (data2, data1)
Cohen's d | 95% CI
-----|-----
0.72 | [0.37, 1.92]
```

- Estimated using pooled SD.

```
interpret_cohens_d(0.72)
[1] "medium"
(Rules: cohen1988)
```

* Director de la Revista Argentina de Cirugía.

Acá se ve un tamaño del efecto moderado y un valor p no significativo. El resultado puede darse a la inversa: un tamaño del efecto bajo y un valor p significativo.

Para proporciones hay otras estimaciones como la h de Cohen, con la misma interpretación, V de Cramer, etc. Otro ejemplo en R, *library* (pwr)^{5,9} con una línea de *script*

```
library (pwr)
```

```
ES.h (0.5,0.4) ### proporciones 0.5 y 0.4 en cada
```

muestra (50 y 40%) – h de Cohen

```
[1] 0.2013579 # el tamaño del efecto es bajo
```

De manera que, el hallazgo de un valor p estadísticamente significativo no es para festejar como si hubiéramos hallado un “Aleph” en el sótano¹⁰, sino para interpretarlo con cautela y profundizar el análisis de nuestros datos con otras herramientas estadísticas. El valor p aislado está sobrevaluado y conviene incluir el cálculo del tamaño del efecto en los estudios de investigación clínica.

ENGLISH VERSION

The following is an answer from Microsoft Bing GPT chat to the question: what is effect size?

“Effect size is a measure of the strength of a phenomenon in statistics. It is a descriptive statistic that conveys the estimated magnitude of a relationship without making any statement about whether the apparent relationship in the data reflects a true relationship in the population¹. Effect size refers to a way of quantifying the magnitude of difference between two groups².

There are three ways to measure effect size, depending on the type of analysis one is performing: 1. Standardized mean difference 2. Correlation coefficient 3. Odds ratio².”

Citations 3 and 4 are suggested but not referenced in the answer.

Artificial intelligence is fashionable in the medical literature and will become a topic of discussion in scientific settings. However, my idea is to leave artificial intelligence for another occasion and take this opportunity to introduce the concept and importance of the effect size in clinical research studies in contrast to reporting the p-value, which is a measure of the probability that the results are due to chance and only if the study is properly designed in terms of applicability.

From the chat response, I highlight the concept of the magnitude of the difference between two groups because a low p-value may coincide with an effect size without clinical relevance.

There are different measures for effect size to compare 2 means, The most common effect sizes are Cohen’s d, Hedges’ g and Glass’ delta³.

I leave relative risk, odds ratios, correlation coefficients, etc., for another occasion because they are beyond the scope of this editorial and I prefer to introduce concepts slowly.

d formula is $\frac{\text{mean 1} - \text{mean 2}}{\text{pooled standard deviation (SD}_{\text{pooled}})}$. A d of 0.2 is considered a small

effect size, a d of 0.5 is considered a medium effect size, and a d of 0.8 or larger is considered a large effect size². If the result is negative, although it is mathematically correct, it is not used to facilitate its interpretation. It is also known as standardized mean difference because the units of the result is the number of standard deviations in which the means differ. Standardized mean difference is a common measure in meta-analyses.

A simple example in R5,6,7,8

With only relevant data

```
mean(data1) # mean group 1
```

```
[1] 41.21997
```

```
mean(data2) # mean group 2
```

```
[1] 61.2
```

```
sd_pooled(data1, data2) # pooled standard deviation
```

```
[1] 27.63153
```

```
t.test(data2, data1) # Student’s t-test of the 2 samples without statistical significance
```

```
t = 1.6169, df = 17.724, p-value = 0.1236 # Welch Two Sample t-test
```

Calculation of Cohen’s d according to the formula

```
mean(data2) - mean(data1))/ sd_pooled (data1,data2)
```

```
[1] 0.7230879
```

Calculation of Cohen’s on R, library (effectsize)

```
cohens_d(data2, data1)
```

```
Cohen’s d | 95% CI
```

```
-----
```

```
0.72 | [0.37, 1.92]
```

- Estimated using pooled SD.

```
interpret_cohens_d(0.72)
[1] "medium"
(Rules: cohen1988)
```

Here, the effect size is moderate and the p-value is not significant. The result may be the other way around. Low effect size and significant p-value

There are other estimations for proportions, as Cohen's *h*, with the same interpretation, Cramer's *V*, etc. Another example in R, library (*pwr*)^{5,9} with one line of script

library (*pwr*)

```
ES.h (0.5,0.4) ### proportions 0.5 and 0.4 in each
sample (50 and 40%) – Cohen's h
[1] 0.2013579 # the effect size is low
```

Thus, finding a statistically significant p-value is not to celebrate as if we had found an "Aleph" in the basement¹⁰, but rather to interpret it with caution and perform a deeper analysis of our data with other statistical tools. An isolated p-value is overrated, and the effect size should be estimated in clinical research studies.

Referencias bibliográficas /References

1. Tamaño del efecto - Wikipedia, la enciclopedia libre
2. <https://statologos.com/tamano-del-efecto/> o Tamaño del efecto: qué es y por qué es importante (<https://statologos.com>) para evitar anuncios
3. ¿Por qué reportar el tamaño del efecto? - Comunicar. Escuela de Autores (<https://www.grupocomunicar.com/wp/escuela-de-autores/por-que-reportar-el-tamano-del-efecto/>)
4. ¿Qué es el tamaño del efecto? | Estadísticas (physiotutors.com) <https://www.physiotutors.com/es/wiki/effect-size/>
5. R Core Team (2023). *_R: A Language and Environment for Statistical Computing_*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.
6. Ben-Shachar M, Lüdtke D, Makowski D (2020). 'effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*. 2020;5(56): 2815. doi: 10.21105/joss.02815
7. <https://www.statology.org/pooled-standard-deviation-in-r/>
8. <https://www.statology.org/cohens-d-in-r/>
9. Champely S (2020). *_pwr: Basic Functions for Power Analysis_*. R package version 1.3-0, <<https://CRAN.R-project.org/package=pwr>>.
10. Borges JL. *El Aleph*. Buenos Aires: Losada; 1949.