2018, Vol. 18, No 3.
ISSN 1667-4545
Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología - Universidad Nacional de Córdoba

Otras formas de entender la *d* de Cohen

Other ways of understanding Cohen's *d*

José Ventura-León * [1]

1 - Universidad Privada del Norte, Lima, Perú.

Recibido: 27/02/2018 **Revisado:** 18/05/2018 **Aceptado:** 28/05/2018

Resumen

La *d* de Cohen (*d*) es una medida del tamaño del efecto bastante utilizada y su reporte es una condición necesaria para los análisis estadísticos. No obstante, los investigadores reportan que la diferencia entre dos distribuciones es pequeña ($d > .20$). Sin embargo, la interpretación de ese coeficiente no es clara en estudios de psicología. En ese sentido, es necesario convertir la *d* en una medida de probabilidad y de esa forma facilitar la interpretación de las distribuciones que son objeto de comparación. Dentro de las medidas más frecuentes se encuentran: U3 de Cohen, el coeficiente de superposición (OVL), la probabilidad de superioridad (PS) y el número necesario para tratar (NNT) que pueden ser considerados como medidas alternativas de la magnitud de una diferencia. Para tales fines se proporcionan códigos en R que los investigadores pueden usar fácilmente, además de una tabla que evidencia las modificaciones de las medidas alternativas ante el incremento del tamaño del efecto.

Palabras clave: d *de Cohen, tamaño del efecto, medidas alternativas, probabilidad de superioridad, Coeficiente de superposición*

Abstract

Cohen's *d* (*d*) is quite a used measure of the size of the effect and its report is compulsory necessary in statistical analyzes. Nevertheless, researchers report that the difference between two distributions is small ($d > .20$). However, the interpretation of this coefficient is not clear in psychology studies. In this sense, it is necessary to convert the *d* into a probability measure to facilitate the interpretation of the distributions that are object of comparison. Among the most frequent measures are: Cohen's U3, the superposition coefficient (OVL), the probability of superiority (PS) and the number needed to treat (NNT), which can be considered as alternative measures of the magnitude of a difference. For such purposes, R codes that can be easily used by the researchers are provided, as well as a table showing the modifications of the alternative measures before the increase in the size of the effect.

Key words: *Cohen's* d*, effect size, alternative measures, probability of superiority, superposition coefficients*

*** Correspondencia a:** José Ventura-León. Av. Tingo María 1122, Breña, Lima. E-mail: jose.ventura@upn.pe

Cómo citar este artículo: Ventura-León, J. (2018). Otras formas de entender la *d* de Cohen. *Revista Evaluar, 18*(3), 73-78. Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Introducción

El *tamaño del efecto* (TE) puede ser entendido como "el grado en que el fenómeno está presente en la población" (Cohen, 1988, p. 9) y es clasificado en dos grandes grupos de familia: (a) familia *r*, que implica las medidas de asociación y (b) familia *d*, que se refiere a las medidas de diferencias de grupos (Ellis, 2010). De acuerdo con revisiones sistemáticas, estos coeficientes aparecen entre el 9 % y el 12 % en revistas psicológicas (Castillo, 2014; García, Ortega, & De la Fuente, 2008) y se incluyen entre las recomendaciones para el análisis de datos (American Psychology Association [APA], 2010). Tanto ha sido su impacto que el artículo A Power Primer, publicado por Cohen en 1992, cuenta con 6839 citas, lo que lo convierte en el artículo más citado en la historia de la psicología (Ho & Hartley, 2016).

Pese a la existencia de numerosos coeficientes acerca del TE, la *d* continúa siendo el más utilizado para examinar la diferencia de medias estandarizadas entre dos grupos independientes. Matemáticamente se expresa como $d = y\ (y'1 - y'2)' / S_p$, donde $y'1$ e $y'2$ son las medias de las poblaciones y S_p es la desviación estándar combinada; de ese modo, el coeficiente resultante puede ser considerado pequeño ($d > 0.20$), mediano ($d > 0.50$) o grande ($d > 0.80$; Cohen, 1988). No obstante, estos valores fueron establecidos de forma arbitraria, y el propio Cohen (1988) advertía que "esta es una operación cargada de muchos peligros" (p. 12). El principal riesgo es que estos valores sean tomados como reglas rígidas que conviertan el tamaño del efecto en una especie de *p*-valor < .05 (Gurnsey, 2017). Por esto, existe un consenso creciente de que estos puntos de corte pueden ser engañosos y deben evitarse (Baguley, 2009), porque no pueden realizarse independientemente del contexto de investigación y de los conceptos asociados a este como la potencia, significación y tamaño de la muestra (Gurnsey, 2017). Dado que el TE representa el grado en que la hipótesis nula es falsa, un TE alto implica un incremento en la potencia estadística (Cohen, 1988, 1992), y cuanto más pequeña sea la significancia, menor será esta potencia (Cárdenas-Castro & Arancibia, 2014).

En ese contexto, es importante transformar la *d* en una medida de probabilidad para facilitar la interpretación de la magnitud de la diferencia entre dos medias a partir de medidas alternativas. En este sentido, Magnusson (2014) expone los siguientes coeficientes, que pueden ser utilizados si se satisfacen los supuestos de normalidad, igualdad de varianza y tamaños de muestra equivalentes (Cohen, 1988): U3 de Cohen, el coeficiente de superposición (OVL), la probabilidad de superioridad (PS) y el número necesario para tratar (NNT), cuya utilización sólo es posible si se satisfacen los supuestos de normalidad, igualdad de varianza y tamaños de muestra equivalentes (Cohen, 1988).

Previo a la explicación de las medidas alternativas, se presenta la Tabla 1 que recoge a modo de resumen todas las fórmulas y comandos de R del presente artículo. Si bien los comandos de la U3 y OVL pueden resultar innecesarios, son útiles para buscar un valor preciso basado en una *d* específica (por ejemplo, $d = 0.623$). Si el usuario decide no utilizarlos, puede recurrir a la Tabla 2 en la que se presentan los diferentes valores de las medidas alternativas para cada incremento de una *d* previamente establecida. Asimismo, se advierte que la PS de la Tabla 2 asume igualdad muestral y el NNT está basado en una tasa de evento de grupo control de .20. Si se desea cambiar las condiciones, será necesario recurrir a las fórmulas y comandos de la Tabla 1.

Tabla 1

Coeficiente, fórmulas y códigos en R de las diferentes medidas alternativas.

Coeficiente	Fórmula	Códigos en R
U3	$U_3 = \Phi(\delta)$	d<-.69 U3<-pnorm(d) U3
OVL	$OVL = 2\Phi\left(\frac{-\lvert\delta\rvert}{2}\right)$	d<-.69 OVL<-pnorm(-(d)/(2))*2 OVL
PS	Grupos Iguales $PS = \Phi\left(\frac{d}{\sqrt{2}}\right)$	d<-.69 CL1<-pnorm(d/sqrt (2)) CL1
	Grupos Desiguales $PS = \Phi(d\sqrt{\frac{p_1 s_1^2 + p_2^2 s_2^2}{s_1^2 + s_2^2}})$	p1<-.40 p2<-.60 DE1<-(4.55)^2 DE2<-(9.11)^2 <-pnorm(d*(sqrt(p1*DE1)+(p2*DE2)/(DE1+DE2))) Pse
NNT*	$NNT = \frac{1}{\Phi(d - \Psi(CER)) - CER}$	d<-.69 CER<-.2 NNT<-1/(pnorm (d + qnorm (CER)) - CER) NNT

Nota. Es la función de distribución acumulativa para la distribución normal y es la *d* de Cohen de la población; s^2 corresponde a la desviación estándar y *p* es la proporción de sujetos; es la inversa; CER es la tasa de eventos del grupo control. * = Los códigos en R son de Magnusson (2014).

U3 de Cohen

El TE puede ser expresado en términos de superposición de una distribución, y en este caso se denomina medida U (Cohen, 1988). La medida U3 resulta de la multiplicación de la *d* por una función de distribución acumulativa para la distribución normal (véase Tabla 1), lo que convierte a la *d* en la probabilidad de que ese valor esté por debajo o encima de la media de la segunda distribución.

Supóngase que un grupo de estudiantes de psicología que asisten a clase puntualmente supera a los estudiantes que llegan tarde; siendo él TE *d* = 0.30, la U es aproximadamente de 62. Esto significa que, si un estudiante del grupo de los puntuales dejara de ser puntual y pasara a formar parte del grupo de los impuntuales, entonces pasaría del percentil 50 al 62.

Coeficiente de Superposición (OVL). El OVL es un coeficiente que revela el porcentaje en que las distribuciones de ambos grupos se superponen (Reiser & Faraggi, 1999), y de esa forma permite probar la similitud o cercanía entre las dos distribuciones (Al-Saleh & Samawi, 2007). Este coeficiente resulta de multiplicar dos veces la función de la distribución acumulativa para la distribución normal por el negativo de la mitad de la *d* (véase Tabla 1).

Supóngase que un grupo de niños con TDAH (trastorno por déficit de atención con hiperactividad) es superado por otro grupo de niños sin TDAH en lectura (*d* = 0.75); para este TE, la OVL es aproximadamente de .71. Esto significa que el 71% de las puntuaciones de lectura de niños con y sin TDAH se superponen.

Probabilidad de Superioridad (PS). La PS es un coeficiente que indica la probabilidad de que un individuo extraído al azar de un grupo sea superior en su puntaje con respecto a una variable, comparado con un individuo extraído al azar del otro grupo (McGraw & Wong, 1992). No obstante, es importante advertir que existen diferencias en las ecuaciones según se trate con grupos iguales o desiguales. En el caso de la igualdad de grupos, el coeficiente resulta de la multiplicación de la función de distribución acumulativa por la d entre la raíz cuadrada de dos y, en el caso de grupos desiguales, se considera la proporción de sujetos de cada grupo con sus respectivas desviaciones estándar (véase Tabla 1).

Supóngase que un psicólogo observa que la diferencia entre un grupo de 20 personas diagnosticadas con depresión (p1 = .54 [proporción de sujetos del grupo 1]; DE = 4.55) y un grupo de 17 personas sin depresión (p2 = .46 [proporción de sujetos del grupo 2]; DE = 9.11) es moderada con una d = 0.55; la PS es aproximadamente .98. Esto indica que hay una probabilidad del 98 % de que una persona escogida al azar del grupo con depresión obtenga una puntuación mayor que una persona escogida al azar del grupo sin depresión.

Número necesario para tratar (NNT). El NNT es un cálculo que permite conocer la cantidad de pacientes necesarios para que un grupo experimental tenga un éxito más (o uno menos) en comparación con el grupo de control (Furukawa & Leucht, 2011). Para calcular este coeficiente es necesario contar con el CER (por sus siglas en inglés), que es el porcentaje de veces que ocurre el evento en el grupo de control. En caso de desconocerlo, se puede establecer un valor por defecto de .20 (en dicho caso los resultados deben tomarse con precaución). Además, el NNT refleja la importancia clínica relacionada con la reducción del riesgo absoluto (McGough & Faraone, 2009). Pese a ello, hay investigadores que indican que existe una pérdida de información en la conversión (Scholten, de Beurs, & Bouter, 1999).

Supóngase que un psicólogo desea conocer la efectividad de un programa de intervención en agresión para niños; por tanto, cuenta con un grupo de control de 20 niños que no reciben el programa y un grupo experimental en el que se efectúan sesiones de intervención. La magnitud de la diferencia entre los grupos es grande (*d* = 0.95) y, por ende, su NNT es 2.91. Esto indica que será necesario tratar a 3 pacientes con el programa de intervención para obtener una mejor respuesta que la del grupo de control.

Tabla 2
Valores para todas las medidas alternativas según diferentes tamaños de la *d* de Cohen.

d	U3	OVL	PS	NNT
0.00	50.00	100.00	50.00	Inf.
0.10	53.98	96.01	52.82	34.30
0.20	57.93	92.03	55.62	16.51

0.30	61.79	88.08	58.40	10.63
0.40	65.54	84.15	61.14	7.73
0.50	69.15	80.26	63.82	6.01
0.60	72.57	76.42	66.43	4.89
0.70	75.80	72.63	68.97	4.10
0.80	78.81	68.92	71.42	3.53
0.90	81.59	65.27	73.77	3.09
1.00	84.13	61.71	76.02	2.76
1.10	86.43	58.23	78.17	2.49
1.20	88.49	54.85	80.19	2.27
1.30	90.32	51.57	82.10	2.10
1.40	91.92	48.39	83.89	1.95
1.50	93.32	45.33	85.56	1.84
1.60	94.52	42.37	87.11	1.74
1.70	95.54	39.53	88.53	1.65
1.80	96.41	36.81	89.85	1.58
1.90	97.13	34.21	91.04	1.53
2.00	97.72	31.73	92.14	1.48
2.20	98.61	27.13	94.01	1.40
2.40	99.18	23.01	95.52	1.35
2.60	99.53	19.36	96.70	1.31
2.80	99.74	16.15	97.61	1.29
3.00	99.87	13.36	98.31	1.27
3.20	99.93	10.96	98.82	1.26
3.40	99.97	8.91	99.19	1.26
3.60	99.98	7.19	99.45	1.25
3.80	99.99	5.74	99.64	1.25
4.00	100.00	4.55	99.77	1.25

Nota. d = d de Cohen; el NNT es calculado según un CER de .20; el PS es calculado para muestral igual. Estos valores fueron creados mediante códigos ad hoc en el programa de acceso libre R. Inf. = un valor infinito

En conclusión, transformar la *d* en valores de probabilidad resulta de mucha utilidad, porque facilita la interpretación de la magnitud de la diferencia y el investigador puede conocer mejor lo que significa un coeficiente *d*. Aunque cabe advertir que la relevancia clínica no puede solo deducirse de la interpretación de *d* de Cohen (Scholten et al., 1999), resulta útil explorar e interpretar de forma más sencilla las distribuciones de los dos grupos a partir de los coeficientes aquí brindados. Además, el hecho de brindar códigos en R facilitará los cálculos de investigadores inexpertos en estadística, que podrán reproducir los códigos, modificando el valor de la *d* y otros índices.

Referencias

Al-Saleh, M. F., & Samawi, H. M. (2007). Interference on overlapping coefficients in two exponential populations. *Journal of Modern Applied Statistical Methods, 6*(2), 503-516. doi: 10.22237/jmasm/1193890440

American Psychological Association (2010). *Publication Manual of the American Psychological Association*. Washington, DC: American Psychological Association.

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100*(3), 603-617. doi: 10.1348/000712608X377117

Castillo, R. (2014). *Reporte del tamaño del efecto en los artículos de tres revistas de psicología peruanas en los años 2008 al 2012* (Tesis de pregrado). Universidad Mayor de San Marcos. Lima, Perú.

Cárdenas-Castro, J. M., & Arancibia, H. (2014). Potencia estadística y cálculo del tamaño del efecto en G* Power: Complementos a las pruebas de significación estadística y su aplicación en psicología. *Salud & Sociedad, 5*(2), 210-244. doi: 10.22199/s07187475.2014.0002.00006

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2a ed.). Hillsdale, NJ: Erlbaum. doi: 10.4324/9780203771587

Cohen, J. (1992). Cosas que he aprendido (hasta ahora). *Anales de Psicología, 8*(1-2), 3-17. Recuperado de http://revistas.um.es/analesps

Ellis, P. D. (2010). *The essential guide to effect sizes. Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, UK: Cambridge University. doi: 10.1017/cbo9780511761676

Furukawa, T. A., & Leucht, S. (2011). How to obtain NNT from Cohen's d: Comparison of two methods. *PloS One, 6*(4), 1-5. doi: 10.1371/journal.pone.0019070

García, J., Ortega, E., & De la Fuente, L. (2008). Tamaño del efecto en las revistas de Psicología indizadas en Redalyc. *Informes Psicológicos, 10*(11), 173-188.

Gurnsey, R. (2017). *Statistics for research in Psychology: A modern approach using estimation*. Thousand Oaks, CA: SAGE.

Ho, Y. S., & Hartley, J. (2016). Classic articles in psychology in the Science Citation Index Expanded: A bibliometric analysis. *British Journal of Psychology, 107*(4), 768-780. doi: 10.1111/bjop.12163

Magnusson, K. (3 de febrero de 2014). *Interpreting Cohen's* d *effect size: An interactive visualization*. Recuperado de http://rpsychologist.com/d3/cohend

McGough, J. J., & Faraone, S. V. (2009). Estimating the size of treatment effects: Moving beyond P values. *Psychiatry (Edgmont), 6*(10), 21-29. Recuperado de https://www.ncbi.nlm.nih.gov/pmc/journals/901

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*(2), 361-365. doi: 10.1037/0033-2909.111.2.361

Reiser, B., & Faraggi, D. (1999). Confidence intervals for the overlapping coefficient: The normal equal variance case. *Journal of the Royal Statistical Society: Series D (The Statistician), 48*(3), 413-418. doi: 10.1111/1467-9884.00199

Scholten, R. J., de Beurs, E., & Bouter, L. M. (1999). From effect size into number needed to treat. *The Lancet, 354*(9178), 598. doi: 10.1016/s0140-6736(05)77952-6