El tamaño del efecto (*d de Cohen*)

El hallazgo de efectos estadísticamente significativos (cuando se rechaza la Hipótesis Nula) pueden ser irrelevantes cuando son de baja magnitud, lo que puede ocurrir cuando las muestras son bastante grandes. Por ello se dice que las pruebas de significación estadística son insuficientes en situaciones prácticas, donde la magnitud del efecto observado es fundamental. Los procedimientos estadísticos de tamaño del efecto tienen como finalidad fundamental la cuantificación de la relevancia del efecto obtenido. Dicho de otra forma, se trata de establecer si efectos estadísticamente significativos son relevantes en el campo de aplicación de la investigación.

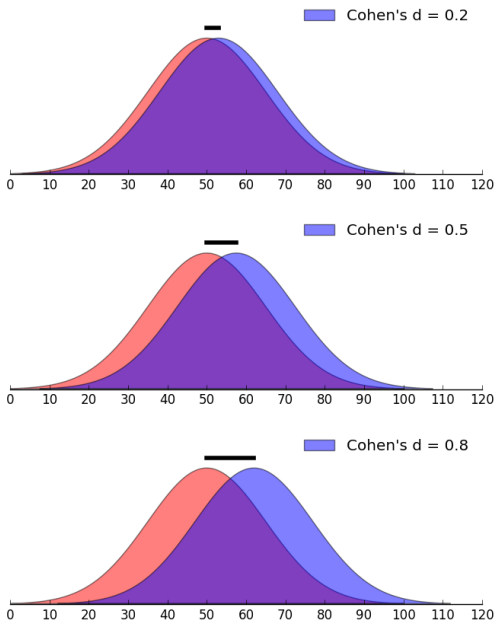
La *d de Cohen* es una medida del tamaño del efecto como diferencia de medias estandarizada. Es decir, nos informa de cuántas desviaciones típicas de diferencia hay entre los resultados de los dos grupos que se comparan (grupo experimental y grupo de control, o el mismo grupo antes y después de la intervención).

Para una práctica eficiente resulta útil conocer el tamaño del efecto, una medida estadística que cuantifica la relación entre dos variables, o la diferencia entre dos grupos. La *d de Cohen* es una medida muy difundida en la que el tamaño del efecto se calcula restando la media obtenida por el grupo experimental menos la media del grupo de control y dividiendo el resultado entre la desviación típica de la población a la que pertenecen ambos grupos.

$$d = \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sqrt{\frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}}}$$

Se puede considerar que la *d* de Cohen representa el número de desviaciones típicas que separan a dos grupos. Por ejemplo, una *d* de 0,5 indica que la diferencia entre el grupo experimental y el grupo de control en la variable evaluada es de media desviación típica. Cohen dio algunas referencias para interpretar la magnitud de los tamaños del efecto:

- *d* = 0,20: tamaño del efecto pequeño.
- *d* = 0,50: tamaño del efecto mediano.
- *d* = 0,80: tamaño del efecto grande.



Ejemplo

Supongamos que un profesor elabora un material que si bien consume muchas horas de estudio, pretende ayudar a sus alumnos a mejorar el rendimiento académico.

Lógicamente, nuestro profesor parte de la hipótesis de que las notas serán superiores si los alumnos utilizan el material al que nos referimos. Para comprobar su hipótesis, el profesor utiliza dos muestras aleatorias de 900 alumnos cada una, obteniendo una nota media igual a 5,5 en el grupo que ha utilizado el nuevo material (Grupo 1), y una nota media igual a 5 en el grupo que no ha utilizado el nuevo material (Grupo 2). Vamos a suponer que conocemos las varianzas poblacionales y que ambas valen 12,5. Plantemos un contraste unilateral derecho cuyas hipótesis son:

$$H_0: \mu_1 - \mu_2 \leq 0$$
$$H_1: \mu_1 - \mu_2 > 0$$

Siendo el valor del estadístico de contraste:

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{5{,}5 - 5}{\sqrt{\frac{12{,}5}{900} + \frac{12{,}5}{900}}} = 3$$

Acudiendo a la tabla de curva normal comprobamos que el nivel crítico p vale: 0,0013.Por lo tanto los resultados son significativos, superando con creces un nivel de confianza del 99%.

Ahora bien, el incremento de la nota (es decir, el tamaño del efecto) es muy pequeño, por lo que, teniendo en cuenta que el nuevo material consume muchas horas de estudio, ¿merece la pena emplear dicho material? Si tomamos una decisión teniendo en cuenta únicamente que hemos obtenido resultados significativos, concluiríamos que sí merece la pena, pero si valoramos el tamaño del efecto (un incremento de tan sólo medio punto), probablemente concluiríamos que no merece la pena utilizar el nuevo material. Por otro lado, es obvio que los resultados son significativos porque el tamaño de las muestras es muy grande, porque, por ejemplo, con muestras $n_1 = n_2 = 100$, el mismo efecto (medio punto de mejora en el rendimiento académico) no habría resultado significativo.

Es importante, por lo tanto, desarrollar medidas que cuantifiquen cuál es el tamaño del efecto, medidas que, por otro lado, han de ser independientes del tamaño muestral.

Aunque hay una amplia variedad de fórmulas, vamos a ilustrar el proceso básico de obtención del efecto con el denominado índice "*d de Cohen*". Básicamente este índice no es más que la estandarización de una diferencia en dos medias, una que sería la del grupo al que se ha aplicado un determinado tratamiento, y otra, la del llamado grupo control.

Práctica

Se realiza un experimento por el que se trata de estudiar si la verbalización del proceso facilita la realización de tareas manuales complejas. Se seleccionan aleatoriamente 60 sujetos y se asignan 30 a cada uno de dos grupos: el experimental, en el cual los sujetos verbalizan la tarea, y el de control, en el que los sujetos realizan la tarea en silencio. Como variable dependiente se registra el tiempo en segundos que se requiere para completar la tarea, y los valores promedio por grupos son los siguientes:

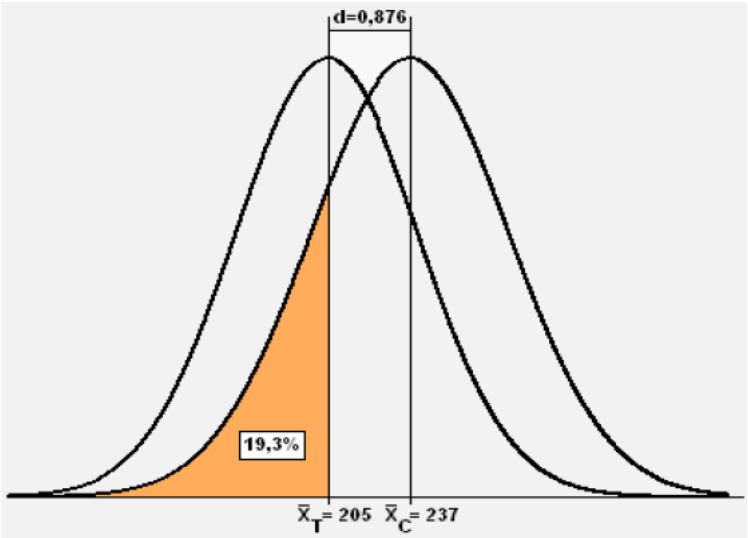
	$\bar{Y}$	$\hat{S}$
Grupo Experimental	205	35
Grupo Control	237	38

Cuantificar la mejora en rapidez que se produce al verbalizar la tarea:

$$d = \frac{|\bar{Y}_{Tratamiento} - \bar{Y}_{Control}|}{\sqrt{\frac{(n_1-1)\hat{S}_1^2+(n_2-1)\hat{S}_2^2}{n_1+n_2-2}}} = \frac{|205-237|}{\sqrt{\frac{29\cdot 35^2+29\cdot 38^2}{n_1+n_2-2}}} = 0{,}876 \approx 0{,}88$$

0,88 es la distancia estandarizada entre las medias de los dos grupos, y su probabilidad asociada, Z_{PROB}= *d de Cohen*, es 0,8106, lo que indica que el 81,06% de los sujetos del grupo experimental tardan menos tiempo que el promedio de los sujetos que no verbalizan. Sólo un 18,94% de los niños que no verbalizan tardan menos tiempo que el promedio de los que sí lo hacen.

En la Figura se observa la situación del grupo de control respecto del experimental:



Magnitud del efecto del grupo experimental respecto del de control