

Tejero-González, C.M.; Castro-Morera, M. y Balsalobre-Fernández, C. (2012). Importancia del tamaño del efecto. Una ejemplificación estadística con medidas de condición física / The importance of effect size: a statistical example using physical condition measurements. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 12 (48) pp. 715-727 [Http://cdeporte.rediris.es/revista/revista48/artimportancia318.htm](http://cdeporte.rediris.es/revista/revista48/artimportancia318.htm)

ORIGINAL

IMPORTANCIA DEL TAMAÑO DEL EFECTO. UNA EJEMPLIFICACIÓN ESTADÍSTICA CON MEDIDAS DE CONDICIÓN FÍSICA

THE IMPORTANCE OF EFFECT SIZE: A STATISTICAL EXAMPLE USING PHYSICAL CONDITION MEASUREMENTS

Tejero-González, C.M.¹; Castro-Morera, M.² y Balsalobre-Fernández, C.³

1 Profesor Contratado Doctor. Departamento de Educación Física, Deporte y Motricidad Humana de la Universidad Autónoma de Madrid (Spain). Web personal: www.uam.es/carlos.tejero; Correo electrónico: carlos.tejero@uam.es

2 Profesora Titular. Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universidad Complutense de Madrid (Spain). Correo electrónico: maria.castro@edu.ucm.es

3 Personal de Investigación en Formación del Departamento de Educación Física, Deporte y Motricidad Humana de la Universidad Autónoma de Madrid (Spain). Correo electrónico: carlos.balsalobre@uam.es

Código UNESCO / UNESCO Code: 5802.06. Análisis, Realización de Modelos y Planificación Estadística / Analysis, Modelling and Statistical Planning.

Clasificación del Consejo de Europa / European Council classification: 17 Otras: Estadística aplicada a las ciencias de la actividad física / Others: Statistics applied to physical activity sciences.

Recibido 8 de noviembre de 2010 **Received** November 8, 2010

Aceptado 20 de febrero de 2012 **Accepted** February 20, 2012

RESUMEN

Tanto en las ciencias de la actividad física y del deporte como en otras áreas de conocimiento científico es habitual investigar con diseños que requieren comparar grupos, concluyendo sobre la existencia o no de diferencias estadísticamente significativas. Sin embargo, no siempre se informa sobre la magnitud de las diferencias encontradas. Este trabajo subraya la importancia de acompañar la significación estadística con valores que informen sobre el tamaño del efecto. Con este propósito y en aras de una argumentación didáctica, se contrasta la capacidad física de chicos y chicas adolescentes,

ejemplificando estadísticamente la estimación del *parámetro delta* con la *g de Hedges ajustada*.

PALABRAS CLAVE: tamaño del efecto, *g* de Hedges, parámetro delta, significación estadística, α y β , comparación grupos, condición física.

ABSTRACT

It is common in the field of physical activity and sports science, as well as in other scientific disciplines, to use designs that require groups to be compared in order to determine the existence of statistically significant differences. However, information regarding the magnitude of any differences found is not always provided. This work highlights the importance of combining statistical significance with values that provide information regarding the effect size. With this in mind, and in order to provide a more didactic discussion, herein we compare the physical abilities of adolescent boys and girls and estimate the *delta parameter* statistically using the *corrected Hedges' g parameter*.

KEYWORDS: effect size, Hedges' *g*, delta parameter, statistical significance, α and β , group comparison, physical condition.

INTRODUCCIÓN

En ciencias de la actividad física y del deporte, al igual que ocurre en otras áreas de conocimiento científico, es habitual encontrar informes que comparan resultados de grupos con el objeto de inferir si existen o no diferencias significativas entre ellos en una determinada característica o variable. Así, las diferencias estadísticamente significativas informan sobre la probabilidad (p) de que los resultados observados en la variable de respuesta o dependiente no se hayan producido por azar, sino por la acción o influencia de la variable independiente.

En este sentido, la significación estadística es la verosimilitud de que la diferencia entre los dos grupos pueda ser un accidente del muestreo. Dicho con otras palabras, mide la probabilidad de que la diferencia observada sea del mismo tamaño que la que se hubiera obtenido por azar, incluso en el caso de que no hubiera diferencias entre los dos grupos. Sin embargo, existen problemas con el uso de pruebas de significación, ya que el valor p es resultado de dos cuestiones: del tamaño de las diferencias y del tamaño de la muestra. Se podrían obtener resultados significativos tanto si las diferencias entre los grupos son muy grandes, aunque la muestra fuera pequeña, como si la muestra fuera muy grande, aunque el tamaño de la diferencia fuera pequeño.

Al respecto, bien a la hora de diseñar el estudio o en la interpretación de resultados, los investigadores deben vigilar dos posibles errores: el *tipo I* y el *tipo II*. El error tipo I, también denominado *falso positivo*, se produce cuando se

rechaza una hipótesis nula que en realidad es verdadera; es decir, cuando el investigador infiere que hay diferencia estadísticamente significativa cuando en realidad no la hay. Por su parte, el error tipo II, también llamado *falso negativo*, se produce cuando se acepta una hipótesis nula que en realidad es errónea, de tal forma que el investigador interpreta que no hay diferencia estadísticamente significativa cuando en realidad sí existe. Para controlar los errores tipo I y tipo II existen dos indicadores: α (*alfa*) y β (*beta*), respectivamente.

El más conocido es el *nivel alfa* o *significación estadística*, que indica el nivel de riesgo tipo I que asume el investigador. La comunidad científica ha establecido dos estándares de alfa: $\alpha=0,05$, para hacer estimaciones con un margen máximo de error tipo I del 5% (nivel de confianza del 95%), o $\alpha=0,01$, para un margen de error del 1% (nivel de confianza del 99%).

También es relevante el *nivel beta*, que hace referencia al riesgo que asume el investigador de que se produzca un error tipo II o falso negativo. En relación con beta, la comunidad científica establece dos estándares: $\beta=0,10$, cuando se quiere garantizar un margen máximo de error tipo II del 10% (confianza o *potencia* de contraste del 90%), y $\beta=0,20$, para asegurar un margen de error del 20% (confianza o *potencia* de contraste del 80%).

Pues bien, una de las posibles estrategias para optimizar los niveles de confianza α y β es aumentar el tamaño de la muestra (Cañadas, Borges, Sánchez y San Luis, 2000). Sin embargo, ni α ni β informan adecuadamente de la magnitud o importancia de las diferencias, en el caso de que las hubiera. Y es precisamente al hilo de esta cuestión donde el investigador debe contar con otro tipo de indicador: el *tamaño del efecto* o *magnitud de la diferencia* (Fan, 2001; Frías, Pascual y García, 2000; Monterde, Pascual y Frías, 2000; Thompson, 2006; Thomas y Nelson, 2007; Valera y Sánchez, 1997).

Al respecto, el tamaño del efecto indica la eficacia cuantificada entre los distintos niveles de la variable independiente, complementado así la información ofrecida por la probabilidad de ocurrencia de la hipótesis nula, pues además de confirmar la existencia de diferencias también informa de la magnitud de éstas.

Para ilustrar la aportación conceptual de la magnitud del efecto citaremos la investigación de Dowson (2000), quien estudió el efecto de la variable *momento de estudio* —con dos niveles: mañana y tarde—, sobre la variable *aprendizaje*. La investigación se realizó con una muestra incidental de 38 sujetos que fueron distribuidos aleatoriamente en los grupos diurno y vespertino, presentándoles el mismo estímulo de aprendizaje. Para tal fin, se midió la comprensión del texto a través del número de respuestas correctas (nivel máximo de la escala 20). La puntuaciones medias fueron de 15,2 para el grupo de mañana y de 17,9 para el grupo de tarde. Llegados aquí, el debate aflora a propósito de dos cuestiones sustantivas: ¿son suficientemente grandes las diferencias entre ambos grupos?, ¿se puede concluir que se aprende más

por la tarde que por la mañana? Además de la significación estadística, una forma de solucionar este tipo de problemas es utilizar el tamaño del efecto. Si no hubiera superposición entre las distribuciones de los dos grupos, estaríamos ante una importante diferencia. En el otro extremo, si la superposición fuera mayor, la diferencia entre los grupos sería menos importante.

Con el objeto de representar gráficamente lo anteriormente expuesto, en Figura 1 se muestran dos situaciones en las que la importancia de la diferencia varía en función de la superposición de las distribuciones. A la izquierda las diferencias son muy grandes y significativas, a la derecha la diferencia existente es menos relevante.

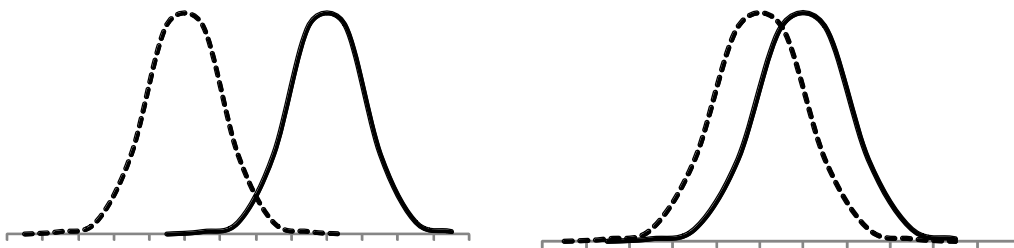


FIGURA 1. Ilustración de la magnitud del efecto: superposición de distribuciones.

Existen diversos procedimientos de estimación del tamaño del efecto; por ejemplo, sin el ánimo de ser exhaustivos: el coeficiente de determinación, eta cuadrado, omega cuadrado, Phi, etc. (Rosnow, Rosenthal y Rubin, 2000; Sink y Stroh, 2006; Trusty, Thompson y Petrocelli, 2004; Vacha y Thompson, 2004). Si bien, este trabajo, con el fin de ejemplificar qué es el tamaño del efecto y que el lector comprenda su utilidad y relevancia, renuncia a revisar las distintas estimaciones del tamaño del efecto y se centrará en la *diferencia estandarizada de medias* o *parámetro delta* (en adelante δ) obtenida mediante la *g de Hedges ajustada* (en adelante, g_{ajust}), siguiendo para ello las directrices de Ledesma, Macbeth y Cortada de Kohan (2008). Decisión que los autores de este trabajo adoptan en virtud de lo que a su juicio son tres criterios favorables a la g_{ajust} : (1) estimación precisa e insesgada, (2) simplicidad de cálculo y (3) fácil interpretación del resultado.

Para obtener la g_{ajust} primero ha de procederse al cálculo de g y después a su ajuste. El parámetro g se obtiene mediante:

$$g = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}}}$$

Siendo, \bar{X}_1 la media aritmética del grupo 1, \bar{X}_2 la media aritmética del grupo 2, n_1 el tamaño muestral del grupo 1, S_1^2 la varianza de las puntuaciones

del grupo 1, n_2 el tamaño muestral del grupo 2 y S_2^2 la varianza de las puntuaciones del grupo 2.

Posteriormente, se ajusta g de la siguiente manera:

$$g_{ajust} = g \left[1 - \frac{3}{4gl - 1} \right]$$

Donde,

$$gl = n_1 + n_2 - 2$$

En definitiva, la g_{ajust} estima la diferencia entre las medias de los grupos y estandariza dicha diferencia dividiéndola entre la desviación típica unificada de los dos grupos, con lo que el procedimiento aporta un parámetro tipificado (puntuación z), al que finalmente se le elimina el sesgo derivado del tamaño muestral. Así, este parámetro expresa un valor tipificado que en última instancia es de gran utilidad ya que permite inferir mediante la tabla de la curva normal el porcentaje de casos que un grupo está por debajo del promedio del otro grupo. Como contrapartida, es necesario el cumplimiento de los supuestos de normalidad y homocedasticidad, especialmente con tamaños muestrales pequeños (por ejemplo, menos de 30 observaciones por grupo) (Pardo y San Martín, 2004).

En virtud de lo expuesto hasta el momento, este artículo tiene como objetivo evidenciar la importancia de acompañar la significación estadística de valores que informen del efecto o magnitud de las diferencias. Con este fin, se desarrolla un estudio de comparación de grupos que se ha diseñado y llevado a cabo expresamente para la ocasión, no previamente publicado.

MÉTODO

Participantes

La muestra está formada por 271 participantes de los que el 53% son hombres (grupo 1; $n_1= 142$), con edades comprendidas entre 12 y 18 años ($M=14,44$, $DT=1,52$), y el 48% son mujeres (grupo 2; $n_2= 129$), también con edades comprendidas entre 12 y 18 años ($M=14,46$; $DT= 1,52$). Los participantes fueron seleccionados mediante muestreo no aleatorio incidental, por motivos de facilidad de acceso. La participación fue voluntaria, autorizada y no recompensada.

Diseño, objetivo y variables

Diseño *ex post facto* retrospectivo. El objetivo del estudio es analizar si hombres y mujeres adolescentes difieren en sus niveles de condición física,

para lo que se midieron tres variables dependientes: fuerza, velocidad y flexibilidad.

Hipótesis

Primera. Los hombres tienen más fuerza que las mujeres.

$$1^a-H_0: \bar{X}_{\text{fuerza hombres}} \leq \bar{X}_{\text{fuerza mujeres}}$$

$$1^a-H_1: \bar{X}_{\text{fuerza hombres}} > \bar{X}_{\text{fuerza mujeres}}$$

Segunda. Los hombres son más veloces que las mujeres.

$$2^a-H_0: \bar{X}_{\text{velocidad hombres}} \leq \bar{X}_{\text{velocidad mujeres}}$$

$$2^a-H_1: \bar{X}_{\text{velocidad hombres}} > \bar{X}_{\text{velocidad mujeres}}$$

Tercera. Los hombres son menos flexibles que las mujeres.

$$3^a-H_0: \bar{X}_{\text{flexibilidad hombres}} \geq \bar{X}_{\text{flexibilidad mujeres}}$$

$$3^a-H_1: \bar{X}_{\text{flexibilidad hombres}} < \bar{X}_{\text{flexibilidad mujeres}}$$

Procedimiento

La toma de datos se llevó a cabo por licenciados en ciencias de la actividad física y del deporte en un instituto de enseñanza secundaria de la Comunidad Autónoma de Madrid (España), durante diversas clases de educación física. El Consejo Escolar del centro educativo, como órgano colegiado superior con representantes electos de alumnos, padres o tutores y profesorado, autorizó el desarrollo del estudio.

La variable *fuerza* se midió con el test de *Lanzamiento de balón medicinal* (Legido, Segovia y Ballesteros, 1995), utilizando para ello un balón de 2 kg y operativizando la fuerza en metros de lanzamiento. Para la medición de la variable *velocidad* los participantes corrieron una distancia de 50 m (Rosandich, 1999), utilizando el tiempo en segundos como expresión de velocidad. Asimismo, la variable *flexibilidad* se valoró mediante flexión de tronco hacia delante desde posición sentada (Prueba de *Sit and Reach*) (Eurofit, 1993), utilizando un banco de flexibilidad con escala graduada en centímetros donde el valor 0 se situó en la planta de los pies.

Una vez recogidos y analizados los datos, todos los participantes recibieron un informe individual sobre su nivel de condición física.

Análisis de los datos

Se procedió con estadística inferencial de contraste de grupos, estableciendo desde un primer momento niveles de confianza α y β iguales a

0,05 y 0,10, respectivamente. Los análisis se llevaron a cabo con ayuda de la aplicación informática *IBM SPSS Statistics 18*.

RESULTADOS

Supuestos de normalidad y homocedasticidad

Tanto para la comparación de grupos mediante *prueba t* de Student como para la estimación del tamaño del efecto es necesario el cumplimiento del *supuesto de normalidad* de las variables, especialmente cuando los grupos a comparar son pequeños, que no es el caso de este estudio.

A tenor de los resultados de la prueba de Kolmogorov-Smirnov sin corrección de significación Lilliefors, no se puede rechazar el supuesto de normalidad en las distribuciones de fuerza, velocidad y flexibilidad de las mujeres ($p=,20$, $p=,93$ y $p=,40$, respectivamente) ni de la fuerza y la flexibilidad de los hombres ($p=,74$ y $p=,38$, respectivamente). Si bien, debe rechazarse la hipótesis de normalidad en el caso de los hombres para la variable velocidad ($p=,03$); lo que no es determinante para el objetivo de este estudio, teniendo en cuenta el tamaño numeroso del grupo afectado y sabiendo que los datos del resto de variables son compatibles con la hipótesis de normalidad de distribución.

En cuanto al *supuesto de homocedasticidad* o igualdad de varianzas, la *prueba de Levene* confirma que hombres y mujeres tienen la misma dispersión en las variables velocidad ($F=,199$; $p=,65$) y flexibilidad ($F=,060$; $p=,80$), pero no en la variable fuerza ($F=53,4$; $p<,001$). No empero, el incumplimiento de este supuesto no es determinante para estimar el tamaño del efecto ya que sólo se produce violación en una de las tres variables y, además, los grupos tienen tamaños grandes y no desequilibrados (como se apuntó más arriba al hablar de la configuración de la muestra, 142 hombres y 129 mujeres).

Comparación de medias

Mediante *prueba t de Student para dos muestras independientes* comparamos los dos grupos en las variables dependientes, obteniendo probabilidades alfa que permiten rechazar las tres hipótesis nulas planteadas. Es decir, es plausible inferir con un nivel de confianza del 99% que es falso que los hombres tengan igual o menor fuerza que las mujeres ($t=11,87$; $gl=213,13$; $p<,001$), que los hombres sean igual o menos veloces que las mujeres ($t=10,57$; $gl=269$; $p<,001$), o que los hombres tengan igual o mayor flexibilidad que las mujeres ($t=-5,61$; $gl=269$; $p<,001$).

Evidencia empírica que permite considerar como provisionalmente verdadero que los hombres son más fuertes y veloces que las mujeres y éstas más flexibles que aquéllos (Tabla 1).

TABLA 1. Prueba T para comparación de medias independientes

| | Sexo | M | DT | p |
|---|---------|------|------|--------|
| Fuerza: metros de lanzamiento con balón de 2kg. | Hombres | 7,33 | 1,85 | < ,001 |
| | Mujeres | 5,23 | 0,94 | |
| Velocidad: segundos en 50m. | Hombres | 8,29 | 1,06 | < ,001 |
| | Mujeres | 9,60 | 0,96 | |
| Flexibilidad: centímetros en flexión de tronco sentados | Hombres | 1,85 | 7,8 | < ,001 |
| | Mujeres | 7,26 | 7,9 | |

Tamaño del efecto

Procediendo a la estimación del tamaño del efecto mediante g_{ajust} , desarrollando la formulación descrita en la introducción, se obtienen las siguientes magnitudes *delta*: $\delta_{fuerza} = 1,39$, $\delta_{velocidad} = 1,27$ y $\delta_{flexibilidad} = 0,68$.

Estadísticos que corresponden a las representaciones gráficas ilustradas en Figura 2.

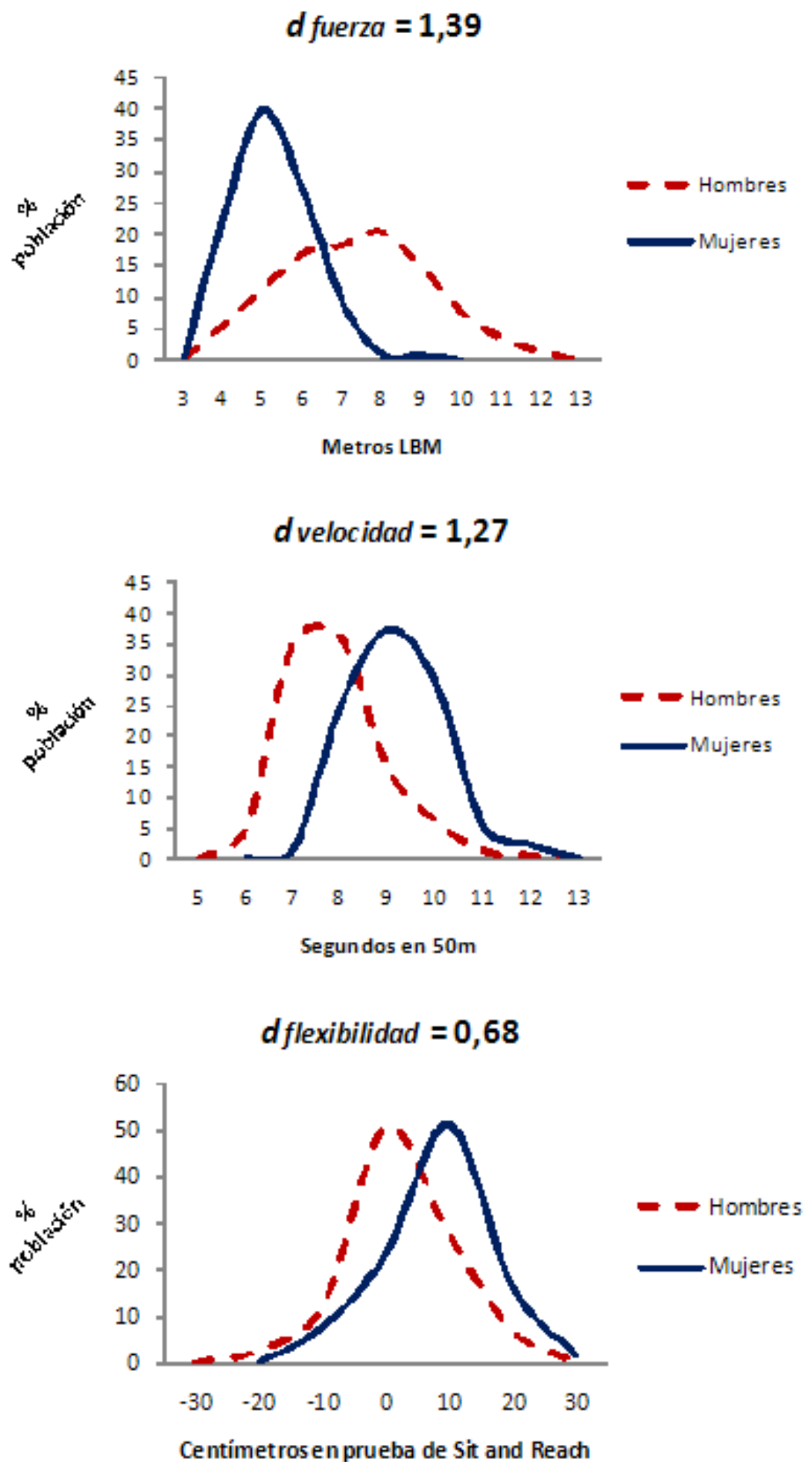


FIGURA 2. Tam años de efecto y representaciones gráficas

Posteriormente, utilizando la *Tabla de distribución normal estandarizada*, analizando la probabilidad que acumula cada una de las diferencias tipificadas (véase por ejemplo Vincent, 2005) se infieren tres nuevos resultados: el 91% de las mujeres tiene una fuerza igual o inferior al promedio de los hombres (puntuación $z=1,39$), el 89% de las mujeres tiene igual o menor velocidad que el promedio de los hombres (puntuación $z=1,27$), y el 75% de los hombres tiene una flexibilidad igual o inferior al promedio de las mujeres (puntuación $z=0,68$) (Tabla 2).

TABLA 2. Distribución normal estandarizada

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |

DISCUSIÓN Y CONCLUSIONES

Este trabajo, valiéndose de dos grupos estadísticamente comparables tanto en cuanto presentan tamaños similares y mismas distribuciones de edad, ha evidenciado que durante la adolescencia hombres y mujeres son distintos en sus capacidades físicas de fuerza, velocidad y flexibilidad. Para ello se ha procedido a contrastar los dos grupos con *Prueba t de Student para muestras independientes* en las tres variables dependientes, observándose la misma significación estadística en todas las ocasiones: $p < 0,001$. Es decir, la probabilidad de que se produzca un error por *falso positivo* es inferior al 1% en las tres variables.

Hasta el momento, sólo se puede concluir que hombres y mujeres son distintos en las variables mencionadas y que nivel de confianza al hacer esta afirmación es superior al 99%. También se sabe que los hombres son

significativamente más fuertes y veloces que las mujeres, y que éstas son estadísticamente más flexibles, pero no sabemos cuánto más un grupo sobre el otro. Así, no se sabe cuál es la importancia o magnitud de las diferencias encontradas. Para esta última cuestión, se ha estimado la *diferencia estandarizada de medias* o *parámetro delta*, obtenida mediante g_{ajust} , encontrando los siguientes valores: $\delta_{fuerza}=1,39$, $\delta_{velocidad}=1,27$, $\delta_{flexibilidad}=0,68$. Datos que a su vez permiten inferir, de acuerdo a la Tabla de distribución normal, que sólo el 9% de las mujeres tiene más fuerza que el promedio de los hombres, que sólo el 11% de las mujeres es más veloz que el promedio de los hombres y que sólo el 25% de los hombres tiene más flexibilidad que el promedio de las mujeres.

En consecuencia, los valores δ obtenidos, independientemente de la escala de medida, puesto que las variables analizadas se operativizaron en metros, segundos y centímetros, indican que la diferencia de mayor magnitud entre los grupos se sitúa en la variable fuerza, seguida de la velocidad, y finalmente la flexibilidad. Información que pudiera ser relevante a efectos teóricos y prácticos, y de la que no informa la significación estadística.

Por tanto, como se ha podido ver en los resultados expuestos, y en sintonía con lo defendido por otros autores (Fernández-Cano y Fernández-Guerrero, 2009; Lustig y Trauser, 2004; Rhea, 2004; Smith y Honoré, 2008; Thompson, 1999), son muchas las virtudes del tamaño del efecto. Se destacan las siguientes:

1. el tamaño del efecto es relativamente fácil de calcular;
2. el tamaño del efecto, como expresión de desviación típica o de varianza explicada, proporciona una interpretación intuitiva de los resultados y de la magnitud de las diferencias, pudiendo considerarse un indicador de significación práctica y no siendo incompatible con la significación estadística;
3. el tamaño del efecto permite comparar diferencias entre variables con distinta métrica por ser adimensional;
4. el tamaño del efecto es un estadístico que facilita la acumulación de conocimiento, pues permite la comparación meta-analítica de los resultados de distintas investigaciones que tienen un mismo objeto de estudio; y
5. el tamaño del efecto mantiene mayor independencia que la significación estadística, con respecto a la influencia del tamaño de la muestra.

En definitiva, la intención de este trabajo ha sido evidenciar mediante argumentos y ejemplificación estadística la conveniencia de acompañar las probabilidades de significación estadística de otros valores que informen del

tamaño del efecto o magnitud de las diferencias. De no ser así, a juicio de los autores de este trabajo, el informe manifestará limitaciones en la presentación de las conclusiones.

REFERENCIAS BIBLIOGRÁFICAS

- Cañadas, I., Borges, A., Sánchez, A. y San Luis, C. (2000). Estudio de la potencia de los contrastes de medias con dos y tres grupos con tamaño del efecto pequeño y en condiciones de normalidad y homocedasticidad, *Psicothema*, 12(2), 114-116.
- Dowson V. (2000). Time of day effects in school-children's immediate and delayed recall of meaningful material. *TERSE Report* (en <http://www.cem.dur.ac.uk/ebeuk/research/terse/library.htm>).
- Eurofit (1993). *Eurofit Tests of Physical Fitness*, 2nd Edition. Strasbourg.
- Fan, X. (2001). Statistical significance and effect size in education research: two sides of a coin. *The Journal of Educational Research*, 94(5), 275-282.
- Fernández-Cano, A. y Fernández-Guerrero, I. (2009). *Crítica y alternativas a la significación estadística en el contraste de hipótesis*, Madrid: La Muralla.
- Frías, M. D., Pascual, J. y García, J. F. (2000). Tamaño del efecto del tratamiento y significación estadística. *Psicothema*, 12(2), 236-240.
- Ledesma, R., Macbeth, G. y Cortada de Kohan, N. (2008). Tamaño del efecto: revisión teórica y aplicaciones con el sistema estadístico ViSta, *Revista Latinoamericana de Psicología*, 40(3), 425-439.
- Legido, J. C., Segovia, J. C. y Ballesteros, J. M. (1995). *Valoración de la condición física por medio de test*. Madrid: Ediciones pedagógicas.
- Lustig, D. y Trauser, D. (2004). Effect size and rehabilitation research. *Journal of Rehabilitation*, 70(4), 3.
- Monterde, H., Pascual, J. y Frías, M. D. (2000). Errores de interpretación de los métodos estadísticos: importancia y recomendaciones, *Psicothema*, 18(4), 848-856.
- Pardo, A. y San Martín, R. (1994). *Análisis de datos en psicología II*, Madrid: Pirámide.
- Rhea, M. (2004). Determining the magnitude of treatment effects in strength training research through the use of the effect size. *Journal of Strength & Conditioning Research*, 18(4), 918-920.
- Rosandich, T. P. (1999). International Physical Fitness Test. *The Sport Journal*, 2(1).
- Rosnow, R. L., Rosenthal, R. y Rubin, D. B. (2000). Contrasts and correlation in effect-size estimation. *Psychology Science*, 11(6), 446-453.
- Sink, A. y Stroh, H. R. (2006). Practical significance: the use of effect sizes in school counseling research. *Professional School Counseling Journal*, 8(1), 115-120.
- Smith, M. L. y Honoré, H. H. (2008). Effect size reporting in current health education literature. *American Journal of Health Studies*, 23(3), 130-135.
- Thomas, J. R. y Nelson, J. K. (2007). *Métodos de investigación en actividad física*. Badalona: Editorial Paidotribo.

- Thompson, B. (1999). Why “encouraging” effect size reporting is not working: The etiology of researcher resistance to changing practices. *The Journal of Psychology*, 133(2), 133-140.
- Thompson, B. (2006). Research synthesis: effect sizes. En J. Green, G. Gamilli y P. B. Elmore (Eds.). *Handbook of complementary methods in education research* (pp. 583-603). Washington, DC: American Educational Research Association.
- Trusty, J., Thompson, B. y Petrocelli, J. V. (2004). Practical guide for reporting effect size in quantitative research in the Journal of Counselling and Development. *Journal of Counselling & Development*, 82, 107-110.
- Vacha, T. y Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473-481.
- Valera, A. y Sánchez, J. (1997). Pruebas de significación y magnitud del efecto: Reflexiones y propuestas, *Anales de Psicología*, 13(1), 85-90.
- Vincet, W. J. (2005). *Statistics in Kinesiology* (3rd ed.). Human Kinetics.

Número de citas totales / Total references: 23 (100%)

Número de citas propias de la revista / Journal's own references: 0 (0 %)