

## ¿Conviene la nota media simple en los exámenes conjuntos de enseñanza secundaria?

JOSE FERNANDEZ HUERTA

Quando los hombres, como legisladores y examinadores, pretendemos simplificar la calificación de los escolares, buscamos una sola nota como síntesis del conocimiento del alumno. El primer obstáculo con el que tropezamos es con la división del curso escolar en asignaturas. El segundo, advertido solamente por aquellos que tienen una preocupación pedagógica más honda, consiste en la parcelación de cada asignatura en cuanto a conocimientos, comprensión, actitudes, etcétera. Mas la simplicidad presiona sobre nosotros de tal modo que nos empuja a recaer en la nota única como medio resolutorio para la decisión uniforme sobre los examinandos.

Es innegable que, conseguida una nota como compendio de los saberes del alumno, podemos arbitraria o rigurosamente determinar el punto de partida del cual se logrará una calificación determinada. El procedimiento a que estamos acostumbrados y que por ello surge con fresca espontaneidad es el de la media aritmética.

¿Qué es o representa la media aritmética? La media aritmética es ahora un parámetro estadístico que se obtiene al dividir la suma de todas las puntuaciones por el número de sujetos examinados. No interesa en este momento todo el aparato científico que supone la admisión de la media; pero sí conviene saber que en términos de probabilidad media aritmética y esperanza matemática coinciden. También se admite que la media representa el centro de gravedad de las puntuaciones consideradas en su cantidad y es el valor central que convierte en mínimo el segundo momento de una distribución estadística. El hecho de que todos los valores intervengan con su cuantía hace que la media aritmética se vea muy afectada por las discrepancias de los extremos.

Un sencillo ejemplo nos puede hacer patente la influencia de los extremos. Si los sujetos A y B han obtenido, respectivamente, las siguientes puntuaciones: A (5-5-5-6-6) y B (1-6-6-7-7), la media aritmética  $27:5 = 5,40$  es la misma para los dos; pero nadie aseguraría que ambas puntuaciones son equivalentes. Todos los jueces, excepto el primero, han concedido un punto más a B que a A. Se comprende que el examinando B es mejor que el A, pero... el primer juez ha rebajado excesivamente la nota de B y su decisión ha pesado con exceso sobre el valor medio. Igual podría haber acontecido por el extremo superior.

Admitimos que el procedimiento operatorio es elemental, ya que la suma de todas las notas divididas

por cinco se hace elementalmente, y, simplificaría aún más si se duplicase el total y se dividiese por 10. Es un criterio económico el que completa al estadístico; pero ¿no podríamos encontrar otra medida de tendencia central que fuese de tanta o más garantía que la media?

Conviene detenernos unos instantes. Pensemos en los jueces. Para aquellos examinadores que consideran las puntuaciones subjetivas (si nos referimos a éstas es por la única razón de ser las que tanto de hecho como legalmente son las predominantes en España) como números invariables, lo que contraría todas las investigaciones realizadas hasta la fecha, es innegable que la media aritmética constituye el mejor valor central cuando las puntuaciones se distribuyen simétricamente. Pero, para quienes las puntuaciones subjetivas son evaluaciones de gradación sin precisión numérica, la media aritmética pierde virtualidades por no interesar el número sino el orden. También la media pierde validez si las puntuaciones se distribuyen asimétricamente.

El nuevo valor central que puede sustituir en este segundo caso a la media aritmética es el valor mediano o mediana. ¿Qué nos señala la mediana? La mediana es un valor posicional que nos indica aquella puntuación que deja tantos valores por debajo como por encima. Ordenadas todas las puntuaciones de mayor a menor, la mediana es el valor que ocupa el centro de la serie. También la mediana posee propiedades estadísticas. Estas propiedades son la posicional y la de convertir en mínimo las diferencias del primer momento absoluto.

Si operamos con dos, tres o cinco jueces, la mediana aparece como más equilibrada que la media. Viene a representar en el campo de la ordenación de puntuaciones lo que la posición ecléctica en el doctrinal: no se deja arrastrar por los extremos. Dejan de preocupar los excesos de un juez, ya que su valor ocupará el extremo. Si truncamos la puntuación de los cinco jueces por separación de la del que puntuó con menor nota, ¿qué nos interesa la cuantía de los puntos que tal juez otorgó? Igual mínimo será en ese caso B con 0, 1, 2, 3, 4 ó 5 puntos. La mediana no sufre alteración aunque sí la sufriese la media.

En los ejemplos A y B anteriores las medianas serán 5 y 6, respectivamente, lo que confirma aquella tendencia que advertimos de un punto más al escolar B. Ofrece la ventaja, ya aparente, de eliminar el influjo pernicioso o provechoso de una de las notas.

Si proponemos dos nuevos ejemplos C y D (1-7-7-7-) y D (5-5-5-5-9-) y estudiamos sus dos medidas de tendencia central, advertimos que las dos medias coinciden, mientras que las medianas serían 7 y 5, respectivamente.

Cuando analizamos las puntuaciones comprobamos una clara tendencia para el sujeto C de obtener siete puntos y para el D de obtener cinco. Pero en el primero uno de los jueces le otorgó solamente un punto, es decir, le rebajó la nota por razones imprevisibles; mientras que en el segundo otro, o el mismo juez, se la incrementó exageradamente. Estas anomalías temporales desaparecen al utilizar la mediana, gracias a la cual concedemos que el escolar C supera al D, lo que no ocurriría con la media.

Abogamos, en principio, por la mediana en lugar de la nota media cuando se utilizan puntuaciones subjetivas de escaso número de jueces. La nota es más equilibrada y, lógicamente, más comprensible.

¿Sería suficiente esta sustitución de medidas de tendencia central para resolver las dificultades que aparezcan? El legislador, en fecha muy reciente (4 de junio de 1954), dice: "Sin embargo, si el presidente juzgara excesiva, en algún caso, la discrepancia entre la puntuación dada por los distintos vocales el ejercicio de un alumno, podrá someterlo al juicio del tribunal en pleno."

Con esta determinación se establece un criterio superior al de la media de los miembros del tribunal restringido: la media de los cinco miembros. Pero este nuevo criterio sólo entrará en vigor cuando el presidente juzgue excesiva la discrepancia. Luego su finalidad no consiste en proporcionar un arma técnica, sino en evitar el exceso de benevolencia o de dureza con *este* alumno precisamente. *Supuesto* un nivel equivalente de juicio y de aptitud examinadora en los jueces, la divergencia *exagerada*, en algún caso, no puede ser originada más que por prejuicios o factores independientes al mismo examen.

No obstante, la orden ministerial introduce un nuevo valor: la discrepancia. Media, discrepancia y nueva media (aunque esto último no se indique), parece ser el camino a seguir para alcanzar la nota definitiva. Así, aparecen ahora dos focos en la calificación: media y discrepancia. Esto último lo podríamos expresar de otro modo si proponemos en lugar de discrepancia (que parece ceñirse más al caso de tribunal restringido) el término variabilidad.

Para nosotros, media o mediana y desviación típica o error probable serán los aspectos que consideremos al sustituir por términos científicos los primeramente expuestos. Otra ventaja ofrece la variabilidad: la posibilidad de eliminar el supuesto que hemos subrayado: el de nivel equivalente de juicio y aptitud examinadora. La ciencia nos enseña la irrealidad de esta equivalencia. La experiencia también.

Como vía introductoria para clarificar el procedimiento apoyado solamente en media o mediana y varianza o error probable, ofrecemos un ejemplo a base de cinco supuestos. Estos supuestos, cuya coincidencia es improbable se presente en algún tribunal, representan las calificaciones atribuidas u otorgadas por cinco jueces (E, F, G, H, I), a diez alumnos innumerados.

E	4	4	5	5	5	5	5	6	6
F	2	2	3	3	3	3	3	4	4
G	6	6	7	7	7	7	7	8	8
H	2	3	3	4	5	5	6	7	8
I	1	2	3	4	5	5	6	7	8

Las medias respectivas son: 5-3-7-5-5 y la media general de todas las puntuaciones es 5. Las medianas serían idénticas a las medias y la general también.

Un pequeño análisis de las notas nos dirá que mientras para los tres primeros jueces los alumnos poseían una información similar, para los dos últimos existían grandes variabilidades. Es cierto que entre E, F y G se manifiesta clara discrepancia, ya que mientras F parece exigir mucho a sus alumnos, G se presenta altamente benévolo. El juez E aparece como el más ponderado.

Los valores o notas dadas por E, F y G, son comparables entre sí y cualquiera medida de tendencia central que empleemos a partir de las tres notas nos proporcionará la solución esperada; pero ¿son equiparables E, F y G con H e I?

Anticipémoslo. No son equiparables. No lo son porque las puntuaciones concedidas por H e I presionan más sobre la media del conjunto de las cinco notas que cada una de las otras tres. De los primeros jueces podría afirmarse poca aptitud discriminativa durante los exámenes debida ora al tiempo de examen, ora al tipo de preguntas, ora a la rigidez de criterio o a cualquier otra causa. No obstante la diferente elevación de las notas en dichos tres jueces, estamos radicalmente incapacitados para determinar cuál es el mejor. Los tres son científicamente iguales. Sólo podremos determinar la superioridad de uno de ellos cuando utilicemos otros criterios que faciliten tal caracterización.

Pero las notas de H e I es probable que no se puedan comparar con las tres antedichas. En el campo de los juicios o notas se puede admitir que los jueces juzguen *lo mismo* cuando sus medias y variabilidad son científicamente iguales (en variabilidad incluimos tanto la varianza como la forma y las covarianzas). Esa igualdad estadística no presupone la eliminación de la dispersión de muestral de notas, que existirá, sino que las diferencias entre los valores indicados sean explicables por la misma fluctuación muestral. Las diferencias no han de ser significativas.

Dados estos supuestos básicos, ¿han juzgado lo mismo el juez E y el juez I? Nuestra primera impresión es negativa. ¿Quién afirmarí que han juzgado lo mismo si mientras la variabilidad del primero alcanza un rango total de tres (4, 5 y 6), la del último abarca nueve (1, 2, ..., 8 y 9)? Los elementos de juicio parecen diferentes.

Ahora, quizá, se nos pretenda preguntar: ¿acaso el juez I es mejor que el juez E porque ha matizado más las notas? Nuestra respuesta es dilatoria. No cabe responder. No cabe responder porque el problema enfocado es distinto. La variabilidad de los juicios no resuelve acerca de la exactitud del juez. Podría ocurrir que el juez con rango inferior haya juzgado con más precisión que el juez con rango superior. Lo único que en verdad podemos concluir es acerca de la diversidad discriminativa; pero los elementos en que se apoya dicha diversidad son demasiado opacos.

Por otra parte, es de todos conocida la inconvenien-

cia de sumar u operar con magnitudes heterogéneas. Si el estudio de las distribuciones de notas nos las presenta como faltas de homogeneidad, nunca deben sumarse, al igual que nadie totaliza en su heterogeneidad dos clases distintas de objetos, como no sea elevándolos el género superior común a todos. Así, nadie suma mesas y sillas, como no sean bajo el concepto de muebles u otro superior. Y el total de ambos podrá sumarse al de libros si lo elevamos a otro término superior: objetos. En ambos casos, hemos transformado la heterogeneidad en homogeneidad; homogeneidad en muebles u objetos.

Esto quiere decir, en nuestro caso, que dada una evidente heterogeneidad de notas, manifiesta en la diversidad significativa de medias y varianzas, sólo podremos aunar las puntuaciones para conseguir la nota media cuando hayamos homogeneizado las puntuaciones, cuando las hayamos igualado en un género superior. De acuerdo con lo dicho, hemos de arbitrar un procedimiento que nos permita homogeneizar las puntuaciones.

Admitida la hipótesis de normalidad en las puntuaciones de los escolares, se considerará científicamente lograda la homogeneidad cuando la media, varianza, forma y covarianzas son las mismas. Para simplificar el procedimiento nos referimos solamente a la igualdad de medias y varianzas. Para la forma, véase la *Revista Española de Pedagogía*, núm. 44, octubre-diciembre, 1953, págs. 517-527.

Se concede que dos distribuciones corresponden a la misma población—en nuestro caso se admitiría la discriminación equivalente de los jueces—cuando su definición coincide. La definición económica se verifica a base de la igualdad práctica de media y desviación típica. Podría utilizarse el error probable por la indudable ventaja en su distribución: 75, 90 y 98 por 100 más comprensible que la de la desviación típica. Por comodidad presentaremos solamente el procedimiento más utilizado.

Si regresamos al último ejemplo de los cinco profesores supuestos, advertimos una gran facilidad para igualar las medias de todos mediante la adición o sustracción de cierta cantidad. Con sumar dos a cada puntuación de B y restar dos a cada una de C, obtendríamos la deseada igualdad de medias, ya que  $\bar{X}_a = 5$ ;  $\bar{X}_b = 3$ ;  $\bar{X}_c = 7$ ;  $\bar{X}_d = 5$ ;  $\bar{X}_e = 5$ , y  $\bar{X}_b = (\sum X_b) : N$ .  $\bar{X}_b + 2 = [\sum (X_b + 2)] : N$ .

Pero el proceso iniciado de este modo es inconveniente, ya que es muy probable que al igualar las discrepancias o variabilidades las medias obtenidas sufran alteración, que debería ser revisada para conseguir la igualdad. Lo aconsejable es igualar primero, con cierta aproximación, las variabilidades típicas y después las medias por la vía que hemos citado. En el caso de que concedamos *a priori* índices de ponderación, se procederá como luego indicaremos.

En el ejemplo base, las desviaciones típicas encontradas en los diversos jueces son las siguientes: E = 0,67; F = 0,67; G = 0,67; H = 2,05; I = 2,58. Sin ulterior demostración admitimos que dado el número de alumnos las desviaciones típicas son diferenciables.

Para lograr la igualación es suficiente multiplicar por un número que transforme en valores idénticos todos los obtenidos; así, si elegimos como tamaño de

desviación típica básico el del juez H, nos sería suficiente multiplicar todas las puntuaciones de E, F, y G por tres y la de I por 0,80. Aunque pudiera aproximarse más el factor no es necesario, dada la variabilidad admisible en el análisis de varianzas.

Hechas las multiplicaciones pertinentes, obtenemos los siguientes conjuntos de puntuaciones: E (12-12-15-15-15-15-18-18); F (6-6-9-9-9-9-12-12); G (18-18-21-21-21-21-24-24); H (2-3-3-4-5-5-6-7-7-8) y I (0,8-1,6-2,4-3,2-4,4-4,8-5,6-6,4-7,2)

Completaremos la labor si las reducimos a la misma media, que, para mayor sencillez, haremos sea la general: 5. Con restar o sumar un constante obtendremos las siguientes puntuaciones definitivas, con las que ya podremos componer la nota final.

E (2 - 2 - 5 - 5 - 5 - 5 - 5 - 8 - 8 -)  
 F (2 - 2 - 5 - 5 - 5 - 5 - 5 - 8 - 8 -)  
 G (2 - 2 - 5 - 5 - 5 - 5 - 5 - 8 - 8 -)  
 H (2 - 3 - 3 - 4 - 5 - 5 - 6 - 7 - 7 - 8 -)  
 I (1,8 - 2,6 - 3,4 - 4,2 - 5 - 5 - 5,8 - 6,6 - 7,4 - 8,2).

El valor central elegible como representativo de la nota alcanzada por cada sujeto debería ser el que correspondiese al obtenible de este último cuadro.

Si analizamos someramente los resultados, advertimos la igualdad de los jueces E, F y G, cuando se les ha suprimido su bondad o dureza y se los ha pesado de acuerdo con su aptitud discriminativa. Además, se puede apreciar un mayor parecido en el conjunto de puntuaciones. El procedimiento está logrado.

El haber expuesto este procedimiento no debe llevar al pensamiento ilusorio de innecesidad del mismo, puesto que en nuestro caso teórico hemos propuesto los alumnos de modo ordenado. Lo que normalmente ocurrirá al juzgar los escolares es que se manifieste clara variabilidad dentro de los valores asignables a cada examinando por los jueces. El procedimiento, en principio, está logrado siempre que sean idénticas media, o mediana y desviación típica o error probable y la forma de la distribución; pero sería más perfecto si las intercorrelaciones posibles entre los valores dados se mantienen dentro de los límites de igualdad. Renunciamos a exponer este procedimiento, que, más científico, sería excesivamente laborioso.

Tampoco debe suponerse que sea éste el único procedimiento: escalas sigmas, T y C, junto a ecuaciones de regresión son otras tantas vías para lograr una verdadera nota media. También pueden ser sometidas a crítica como el procedimiento mentado; pero constituyen la base tanto para una sola materia, desde perspectiva única, como para varias asignaturas o una sola con diferentes escores.

¿En qué nos apoyaremos cuando intentemos obtener un verdadero valor medio de asignaturas diferentes o de aspectos claramente diferenciables de una sola materia? El proceder lógico es el mismo aunque se complique algo la técnica. En cuanto hablamos de diferentes materias surge la cuestión de la "importancia" de dichas materias. ¿Es legítimo, se preguntarán algunos, que la materia A influya sobre la media lo mismo que la materia B? A esta pregunta se responde con dos soluciones: 1. Todas las materias admitidas a examen son de la misma importancia. 2. Entre estas

materias las más importantes son... B y D, la una con mayor importancia que la otra.

En el primer caso, cuando todas las materias son igualmente importantes, el proceso a seguir es exactamente el que hemos presentado respecto de una. Consiste en igualar desviaciones y medidas de tendencias centrales. Igualados dichos parámetros se consigue, después de las transformaciones, averiguar la puntuación que corresponde a cada alumno en las asignaturas. La media o mediana para cada examinando producirá la nota final.

En el segundo caso, es necesario conceder índices de ponderación. Así, las materias A y C valen como dos, B como cuatro, D como cinco y E como tres. Pronto se advierte que un diez o un cero en la materia D tiene más importancia para la nota final que un diez o un cero en las materias A o C.

Aunque se pueden señalar varios procedimientos para resolver la nueva complejidad, bosquejaremos uno que, dado lo ya expuesto, nos parece más sencillo.

Si queremos mantener la misma media general como símbolo de la puntuación aprobatoria, el índice de ponderación influirá directamente sobre la variabilidad. La relación de ponderación parece actuar como constante que multiplica a todos los valores, luego la desviación típica de dichos valores queda multiplicada. Así, si en el ejemplo de los cinco jueces E, F, G, H e I hubiéramos pensado que los jueces eran materias ligadas por los índices de ponderación E y  $F = 2$ , G y H  $= 3$  e I  $= 5$ , las desviaciones típicas parece debían ser: 0,67 — 0,67 — 1,00 — 3,07 —

6,45. Estos resultados son el producto de multiplicar cada puntuación por su índice dividido entre dos. Pero si ahora utilizásemos el criterio anterior de igualamiento de las varianzas nuestra meta sería la contraria de la deseada, ya que habríamos conseguido que, por la multiplicación realizada, creciese la desviación típica, y, como se ha podido observar, operamos con pesos equivalentes a números e inversos de la variabilidad.

El procedimiento que seguiremos será contrario. Dividiremos las desviaciones típicas entre la ponderación concedida. En nuestro caso obtendremos los siguientes valores: 0,67 — 0,67 — 0,44 — 1,37 — 1,03. El procedimiento operatorio posterior es como el que ya hemos descrito. Igualaremos las desviaciones típicas y luego las medias. Para igualar las desviaciones típicas, dejaremos constantes las dos primeras asignaturas, multiplicaremos por 1,5 las puntuaciones de G, por 0,5 las de H y por 0,65 ó 0,7 las de I. Las medias quedarán equiparadas al restar 5,5 a los resultados de G y sumar 2,5 a H y 1,75 ó 1,50 a I. La puntuación de F se iguala sumando 2.

De este modo habríamos homogeneizado las puntuaciones bajo la condición de índice ponderativo y la media final representaría lo que se habría proyectado al ponderar la importancia de las diversas asignaturas.

Lo que nos permite concluir acerca de la inconveniencia de utilizar la media aritmética simple o bruta para alcanzar una sola nota como expresión simplista de los saberes del alumno.

Cualquier procedimiento indicado en este artículo satisfaría con más rigor la obtención de la nota única.