
PROCEDIMIENTOS Y FÓRMULAS ESTADÍSTICAS PARA MEDIR LA INFLUENCIA HERENCIA/ AMBIENTE SOBRE CARACTERES MÉTRICOS EN POBLACIONES HUMANAS

MAX CONTASTI y EDGAR MATHEUS

RESUMEN

El objetivo de este trabajo consiste en el estudio y el análisis, desde un punto de vista estadístico, de los diversos procedimientos y fórmulas propuestos para calcular en las poblaciones humanas, la influencia de la herencia/ambiente en algunas variables fenotípicas de tipo continuo que se puedan medir con una precisión aceptable, tales como estatura, peso, presión arterial e inteligencia. Se presenta el análisis de las limitaciones más relevantes en la formulación estadística considerada y cómo, según los casos, se han corregido las

limitaciones observadas y se han generado nuevos procedimientos de cálculo para minimizar esas limitaciones estadísticas y metodológicas. Desde un punto de vista epistémico y metodológico, hasta 1974 las investigaciones estuvieron orientadas únicamente por el enfoque genético, que calcula la influencia de la herencia a través del uso de datos obtenidos de gemelos o de otros miembros del grupo familiar. Posteriormente el ambiente comenzó a ser considerado como una variable adicional para la formulación estadística en el análisis.

El objetivo de este trabajo consiste en estudiar y analizar, desde un punto de vista estadístico, diferentes procedimientos y fórmulas que han sido propuestas para calcular, en poblaciones humanas, la influencia herencia/ambiente (heredabilidad/ambientalidad) sobre cierto tipo de caracteres poligénicos, que al constituirse en variables resultan empíricamente medibles con aceptable precisión. Tales caracteres son, por ejemplo, la estatura, el peso, la presión arterial y, en menor grado debido a las dificultades de su medición, la inteligencia y variables temperamentales de la personalidad. Este análisis permitirá mostrar cuáles han sido las principales limitaciones de las fórmulas propuestas y cómo, según los casos, en cada uno de los desarrollos sucesivos se ha tratado de superar tales limitaciones. Será destacado el desplazamiento del foco epistémico y metodológico en las fórmulas y procedimientos, al pasar de diseños y mediciones orientadas genéticamente sobre gemelos y otros parentescos hacia observaciones dirigidas directamente a la medición ambiental.

Las primeras investigaciones sobre la influencia estadística de la herencia son las de Galton (1883), donde se comparan y se establecen correlaciones estadísticas entre características de los padres con hijos y nietos, manteniéndose como problemas de difícil tratamiento las diferencias de temporalidad y ambiente entre las distintas generaciones. En ese sentido, surgió la posibilidad de estudiar gemelos como un camino para determinar estadísticamente la relación herencia/ambiente. El desarrollo estadístico posterior en este campo de investigación ha sido influido por el diseño de estudios con plantas y animales. En el caso de las poblaciones humanas existen limitaciones, especialmente de carácter ético, para conformar los diseños experimentales, inclusive en la llamada experimentación de campo, donde se procura, por una parte, la homogeneización y, por la otra, la asignación aleatoria de los ambientes, procedimientos que constituyen una práctica común en la genética cuantitativa aplicada a plantas y animales, pero que no se facilita en poblaciones humanas.

Por estas razones se optó por comparar las características de los dos tipos de gemelos, univitelinos y bivitelinos, que al ser criados juntos permiten un control estadístico del ambiente (Holzinger, 1929). Posteriormente, Jensen (1967) extendió y generalizó el procedimiento, al incluir la posibilidad de la comparación de los dos tipos de gemelos con la de niños adoptados criados juntos. En los últimos años, Rao *et al.* (1974) estudiaron de manera conjunta y mediante comparaciones múltiples y simultáneas varios tipos de parentesco, tales como hermanos, primos, padres y tíos. Los procedimientos estadísticos que utilizan gemelos y otros parentescos colaterales, iniciados por Rao *et al.* (1974), siguen vigentes hasta ahora. MacCallum y Austin (2000) hacen un recuento sobre la construcción y uso del modelaje de ecuaciones estructurales (SEM, del inglés *structural equation modelling*) en la investigación de la influencia herencia/ambiente. En todos estos trabajos se continúa utilizando como dato básico de entrada y bajo diferentes procedimientos de cálculo (doble entrada/

PALABRAS CLAVE / Ambiente / Estadística / Herencia / Metodología /

Recibido: 28/02/2008. Modificado: 18/06/2009. Aceptado: 20/06/2009.

Max Contasti. Doctor en Filosofía y Letras, Universidad Complutense de Madrid, España. Doctor en Ciencias, Universidad Central de Venezuela. Profesor, Universidad Nacional Abierta, Venezuela. Dirección Postal: Circunvalación del Sol, Edificio Tarabay, Apartamento 3-A, Santa Paula Caracas, 1061. e-mail: contasti1938@gmail.com

Edgar Matheus. Ingeniero Civil, Universidad Los Andes, Venezuela. Maestría en Ciencias Políticas, Universidad Simón Bolívar, Venezuela. Ingeniero Consultor, Consorcio Decoyne. e-mail: ematheusb@gmail.com

correlación intraclase) el coeficiente de correlación de Pearson.

Históricamente, todos los intentos basados en el parentesco se habían centrado en el cálculo de la heredabilidad, pero a partir de Rao *et al.* (1974) se comenzó a tomar en consideración el ambiente, aunque solo fuese como un factor de ajuste, conformándose así índices que trataban de medir la similitud o discrepancia ambiental entre los dos miembros del par. Loehlin (1989), en las conclusiones del trabajo titulado "Partición de las contribuciones genética y ambiental sobre el desarrollo conductual", expresa que todavía el tema ambiental no ha sido estudiado en profundidad y que se está bien orientado en las investigaciones, aunque no se han logrado respuestas definitivas. Concluye que será en el lado de la exploración ambiental, más que en el lado genético, donde, en el futuro, los estudios darán sus mayores frutos. Plomin y Rende (1991) afirman a su vez que resulta extraño, tras décadas de ambientalismo, que el factor limitante en este esfuerzo de la determinación de la influencia herencia/ambiente haya sido la necesidad de mejores mediciones del ambiente, y que en particular son pocas las mediciones que se focalizan en el ambiente del niño.

Contasti (1996) introdujo, en su llamada "Fórmula Metodológica", una propuesta para la consideración del ambiente. Esta fórmula tiene la misma estructura estadística de un caso particular de Jensen (1967), diferenciándose porque en lugar de pares de niños adoptados relacionados por crianza, en esa fórmula se utilizarían muestras aleatorias de sujetos individuales, de la misma edad y sexo, más extensas y representativas, que presentan únicamente una medición que indicaría similitud del ambiente de la familia, similitud que sería lo equivalente a la crianza común del par de niños adoptados.

Definición Estadística de Heredabilidad. Aspectos Metodológicos

Se asume que, en relación con un carácter poligénico cualquiera, influyen fenoménicamente la herencia y el ambiente. Tratándose de caracteres continuos y mensurables empíricamente, se encuentra que entre las fuentes de variación, además de los generados por la herencia y el ambiente, habrá que agregar el residuo aleatorio correspondiente, que se genera por la confiabilidad de las mediciones.

De manera ideal, puede postularse que un carácter fenotípico medido por una variable t , se conforma como la suma de tres componentes: el primero, la influencia de la variable herencia $t(H)$, para simplificar señalado como (h) ; el segundo, la influencia de la variable ambiente $t(A)$, para simplificar

señalado como (a) y en tercer lugar el residuo aleatorio (ϵ) , de manera tal que

$$t = h + a + \epsilon$$

En principio, h y a no son variables observables y medidas empíricamente, por tanto se trata de constructos; por ello el residuo aleatorio ϵ es generado únicamente por la medición de la variable t . Cuando se utilizan pares de gemelos o de niños adoptados criados juntos, se considera como un supuesto básico que los dos ambientes correspondientes a cada uno de los dos sujetos que conforman el par sean absolutamente idénticos ($A_1 = A_2$); consecuentemente, generan el mismo efecto: $t(A_1) = t(A_2)$, será denominado ambiente común compartido y su efecto, para simplificar, será señalado como (c) . Si este supuesto no se cumpliera, implícitamente se estaría realizando una medición diferencial del ambiente; entonces, además del residuo aleatorio ϵ correspondiente a la medición de la variable t , también será generado otro tipo de residuo aleatorio correspondiente a las diferencias de los dos efectos ambientales. Este residuo (α) es llamado ambiente intangible (Falconer, 1960). Bajo estas condiciones, se puede expresar como el modelo biométrico de Prescott y Kendler (1996):

$$t = h + c + \alpha + \epsilon$$

Para la estimación de la heredabilidad, utilizando un enfoque genético, las muestras de gemelos comúnmente utilizados han sido las siguientes:

-En primer lugar se han tomado gemelos criados juntos; en este caso no existiría la posibilidad de ningún tipo de sesgo de carácter estadístico ya que biológicamente el par comparte la misma carga genética e idealmente son criados juntos por sus padres naturales en ambientes sociales y educativos idénticos. Para simplificar, la correlación r_{jj} simboliza el tomar dos mediciones de la misma variable (t) , una para cada gemelo que conforma el par. En este caso, cada par comparte la herencia y, supuestamente, el ambiente común compartido. Quedaría el residuo aleatorio, cuya correlación, por su propia condición aleatoria, es nula tanto con la herencia como con el ambiente. Idealmente, al nacer, si se pudiese medir la variable t con una confiabilidad perfecta, el valor de esta correlación r_{jj} debería ser igual a la unidad, ya que solo en el vientre materno podría existir un ambiente idéntico, pero posteriormente, por diversas circunstancias subjetivas e individuales, aun para gemelos univitelinos, esta correlación, cuando se estima a lo largo de distintas edades, tenderá a disminuir y será ligeramente inferior a la unidad. Esto indicaría la existencia de una diferenciación en la variable ambiente, ya que los gemelos, aunque puedan pasar juntos veinticuatro horas diarias, que sería un indicador de un ambiente común

compartido, se van diferenciando en alimentación, salud, gustos y otros aspectos que influyen en la conformación del ambiente intangible definido anteriormente. Este ambiente intangible tendrá efectos aleatorios en conjunto con el residuo aleatorio producto de la confiabilidad de la medición de la variable t (Loehlin, 1989; Plomin y Rende, 1991).

-En segundo lugar, se presenta una situación que en condiciones ideales podría servir para estimar de manera directa la influencia de la variable herencia. Sería el caso de gemelos univitelinos pero criados separadamente. En este caso, para que pudiera estimarse la influencia neta de la variable herencia, sin sesgo de carácter estadístico, habría que aleatorizar los valores nominales de la variable ambiente; ello implica asignar aleatoriamente a los dos gemelos de cada par a los futuros ambientes de crianza, lo cual no será posible porque no se trata de un experimento, tal como se realiza en animales y plantas, donde las variables pueden controlarse y aleatorizarse. En este caso precisamente se tiene que, por una parte, no es frecuente que existan gemelos univitelinos que se críen separadamente y cuando excepcionalmente estas circunstancias se han dado, podemos suponer que existirá una cierta posibilidad de que algunas de las familias adoptantes tengan algún parentesco entre ellas, que vivan en lugares cercanos tales que los gemelos compartan muchas actividades sociales comunes, como asistir a las mismas escuelas. La correlación r_{ss} bajo esta circunstancia presentaría un cierto sesgo estadístico, sobreestimándose el parámetro correspondiente a la influencia de la herencia, ya que el par puede compartir algún componente del ambiente, cuyo efecto le será acreditado estadísticamente a la influencia de la herencia. Adicionalmente, siendo las muestras muy pequeñas, las restricciones en el recorrido y variabilidad de la variable ambiente vulneran la representatividad de la estimación.

-El tercer caso se genera al tomar gemelos bivitelinos del mismo sexo, criados juntos por sus padres naturales. Cada par de gemelos puede compartir de manera común 0, 1, 2, ..., hasta 23 pares de cromosomas, calculándose su probabilidad según una distribución binomial. Las probabilidades de que un par de gemelos compartan exactamente los 23 pares de cromosomas (caso equivalente a gemelos univitelino), o que no compartan ninguno, es muy pequeña. Por ser una distribución simétrica, la mayor probabilidad estará al compartir 11 o 12 cromosomas, siendo 11,50 la esperanza matemática de la distribución. Tomando entonces un gran número de gemelos bivitelinos, el promedio, en el conjunto de los pares será $\sim 11,50$. Existirá adicionalmente el residuo aleatorio que se combinará en sus efectos con la herencia no compartida. Esta correlación se simboliza como r_{ff} .

El cuarto caso, distinto y extremo, será el de niños del mismo sexo y edad sin ningún tipo de parentesco, ambos adoptados y criados juntos desde su nacimiento por la misma familia. Difieren por tanto solo en el ambiente intrauterino. En este caso no existe mayor problema de aleatoriedad de la variable herencia, ya que la gran variabilidad genotípica propia de los seres humanos asegura esta condición. El problema sería en cuanto al ambiente, ya que los dos niños, de la misma edad y sexo, deben ser ambos adoptados, porque un par hijo/adoptado, por un posible trato afectivo diferencial, no asegurará la homogeneidad ambiental. En las condiciones antes mencionadas, las muestras que pueden obtenerse son tan singulares que no solo será difícil conformarlas, sino que por su singularidad no contemplarían todos los tipos de ambientes posibles; es decir, es una situación poco representativa y, además, puede generarse un sesgo estadístico muy difícil de detectar o de prever. El par, supuestamente, comparte un mismo ambiente común compartido (c), y la variable que vendrá aleatorizada será la herencia, que por tanto se combinaría en sus efectos con el residuo aleatorio. Se simboliza esta correlación como r_{pp} .

Para evaluar en términos estadísticos los diferentes coeficientes de correlación utilizados (r_{jj} , r_{ss} , r_{ff} , r_{pp}), en todos los casos se dará la influencia de los cuatro componentes que hemos denominado herencia (h), ambiente común compartido (c), ambiente intangible (α) y residuo aleatorio (ϵ). Por otra parte y según el caso, será posible separar dos partes entre los sumandos: una que será igual y/o equivalente (E) para los dos componentes del par (primer grupo entre paréntesis) y otra que es distinta y aleatoria (segundo grupo entre paréntesis).

De tal forma, la variable t vendrá dada por:

$t = (h + c) + (\alpha + \epsilon)$ para gemelos univitelinos criados juntos,

$t = (h) + (c + \alpha + \epsilon)$ para gemelos univitelinos criados separados,

$t = (0,5h + c) + (0,5h + \alpha + \epsilon)$ para gemelos bivitelinos criados juntos, y

$t = (c) + (h + \alpha + \epsilon)$ para niños adoptados criados juntos.

Bajo esas condiciones se puede llegar a fórmulas de correlación similares a las desarrolladas en la teoría de confiabilidad. Allí queda demostrado (Nunnally, 1967) que toda correlación del tipo r_{ij} donde x_i , x_j tienen una parte común equivalente (E) y otra aleatorizada, representará directamente una proporción entre varianzas. Específicamente, la proporción de la varianza total de x que es explicada por la parte común equivalente (E). Este desarrollo ha sido confirmado por Jensen (1971).

La definición estadística de heredabilidad se hará explícita evaluando las correlaciones entre gemelos y adoptados, de la manera siguiente:

$r_{jj} = (\sigma_h^2 + \sigma_c^2) / \sigma_t^2$ para gemelos univitelinos criados juntos en ambiente común compartido.

$r_{ss} = \sigma_h^2 / \sigma_t^2 = h^2$ (heredabilidad), para gemelos univitelinos criados separadamente en ambientes diferentes supuestamente aleatorizados. La heredabilidad (h^2) se define entonces como la proporción de varianza de la variable t que es explicada por las variaciones de la variable herencia σ_h^2 / σ_t^2 .

$r_{ff} = (0,5\sigma_h^2 + \sigma_c^2) / \sigma_t^2$ para gemelos bivitelinos criados juntos en ambiente común compartido

$r_{pp} = \sigma_c^2 / \sigma_t^2 = c^2$ para pares adoptados criados juntos en ambiente común compartido.

Nótese que r_{pp} no puede definir operacionalmente la ambientalidad (a^2), pues solo se toma en consideración el ambiente común compartido. La ambientalidad (a^2) será definida como la proporción de varianza de la variable t que es explicada por las variaciones de la variable ambiente.

$$a^2 = (\sigma_c^2 / \sigma_t^2) + (\sigma_\alpha^2 / \sigma_t^2) \quad (\text{Ambientalidad})$$

Si, idealmente, t pudiese ser medido con una confiabilidad perfecta y no existiese el ambiente intangible, entonces se tendría que $h^2 + c^2 = 1$. Como en una situación empírica ello no es posible, ni es posible tener ambientes absolutamente idénticos, entonces $h^2 + c^2 < 1$.

Al calcular r_{pp} , la no existencia de ambientes absolutamente idénticos ($A_i \neq A_j$) crea un error aleatorio generado por el ambiente intangible en la variable ambiente (A), que influirá subestimando el valor correspondiente al ambiente común compartido (c) y complementariamente sobreestimando el valor de la heredabilidad (h^2). Por otra parte, al calcular r_{ss} en gemelos univitelinos criados separadamente, la falta de aleatoriedad del ambiente del par puede generar una correlación que estadísticamente se acredita a la herencia, por tanto, esa relación influirá sobreestimando el valor h^2 . Nótese entonces que históricamente, en cualquiera de los casos, se habrá presentado siempre una sobreestimación de la heredabilidad.

Fórmulas y Procedimientos Propuestos

Para estudiar la influencia de la herencia en un determinado carácter fenotípico, Holzinger (1929) utilizó las diferencias en ese carácter entre los dos gemelos criados juntos que forman el par. Al comparar el promedio de las diferencias entre los gemelos univitelinos ($I_i - I_j$) con el promedio de las diferencias entre los gemelos bivitelinos ($F_i - F_j$), es de esperarse que tomando los valores absolutos, el promedio correspondiente a los

gemelos bivitelinos ($F_i - F_j$) resulte mayor que el promedio correspondiente a los gemelos univitelinos ($I_i - I_j$). La expresión estadística del procedimiento se presenta en términos de la varianza de las diferencias y sería

$$h^2 = (\text{Var}(F_i - F_j) - \text{Var}(I_i - I_j)) / \text{Var}(F_i - F_j) \quad (1)$$

Utilizando ciertos supuestos que se consideran aceptables cuando se trata de muestras grandes o poblaciones generales, esta fórmula puede ser expresada en términos de los correspondientes coeficientes de correlación de Pearson.

$$h^2 = (r_{jj} - r_{ff}) / (1,00 - r_{ff}) \quad (2)$$

Al evaluar la Ec. 2 de Holzinger (1929), encontramos que si $r_{jj} = 1$, entonces, $h^2 = 1$, independientemente de cuál sea el valor r_{ff} . Esto constituye una limitación de la ecuación presentada. Al revisar los cálculos en muestras pequeñas según los dos procedimientos propuestos, dados los supuestos utilizados para pasar de una ecuación a la otra, en general se encuentran resultados distintos. Por esta razón, al utilizar la Ec. 2, que se expresa por el coeficiente de correlación de Pearson, se procede a duplicar los datos de cada par de gemelos, considerando la duplicación en el orden inverso. Este procedimiento de cálculo del coeficiente de correlación de Pearson fue denominado históricamente correlación de doble entrada.

Falconer (1960) utiliza la fórmula correspondiente a la correlación intraclase, que se expresa como una relación entre dos tipos de varianzas. En el análisis de varianza, la varianza de una variable (t) medida en diferentes grupos se puede descomponer en dos partes: la varianza entre grupos, $\text{Var}(e)$ o varianza ponderada entre las medias de los grupos, y la varianza intra-grupos, $\text{Var}(i)$ o promedio de las varianzas internas dentro de los grupos. La fórmula general conocida del coeficiente de correlación intraclase (I) será

$$I = \text{Var}(e) / (\text{Var}(e) + \text{Var}(i))$$

Las fórmulas utilizadas, una calculada por el procedimiento de doble entrada en pares, y la otra como estimado de la correlación intraclase, generan idénticos resultados (Taylor, 1980). Como lo expresa Loehlin (1989), las diferentes investigaciones sobre heredabilidad han venido utilizando las dos versiones de cálculo, dependiendo del programa electrónico que se utilice. Saudino *et al.* (1995), en uno de sus últimos estudios con gemelos, sobre la influencia herencia/ambiente en aspectos temperamentales, afirma un tanto reiterativamente que se utiliza la correlación intraclase usando un procedimiento de doble entrada. El coeficiente de correlación intraclase (I), por ser conformado mediante una relación entre varianzas, sería equivalente a un coeficiente de determinación (r^2) en su estructura algebraica. De la misma

manera, un coeficiente de correlación de doble entrada, r_{ij} y/o r_{ff} , y tal como lo afirmó Jensen (1971), expresa directamente y sin necesidad de tomar su valor al cuadrado una relación entre varianzas.

Falconer (1960) trabajó sobre diseños experimentales en animales, especialmente en ratones y en la mosca *Drosophila*. Aunque no presenta una expresión algebraica particular como fórmula de heredabilidad, al referirse a poblaciones humanas expresa que la diferencia entre los coeficientes de correlación de los gemelos univitelinos y de los gemelos bivitelinos podría tomarse en casos ideales como estimación del doble de la heredabilidad.

En nuestra notación, lo expresado por Falconer (1960) sería formalizado como

$$h^2 = (r_{ij} - r_{ff}) / (0,50)$$

o, equivalentemente, como

$$h^2 = (r_{ij} - r_{ff}) / (1 - 0,50)$$

Nótese que el valor de 0,50 surge como consecuencia de una expresión simétrica entre el numerador y el denominador. En el numerador la diferencia se da entre dos expresiones empíricamente calculadas (r_{ij} y r_{ff}); en el denominador, en cambio, la diferencia se da entre dos valores estimados probabilísticamente y se refiere a la diferencia entre la carga genética compartida por los dos gemelos univitelinos (100%) y el promedio de la carga genética compartida por los dos gemelos bivitelinos, que como esperanza matemática se aproxima al 50%, valor que se expresa operacionalmente por medio de la proporción de pares de cromosomas comunes en los dos gemelos. De manera exacta, precisa e idéntica en los gemelos univitelinos, y de manera probabilística y promedio en los gemelos bivitelinos.

Jensen (1967), a partir de la expresión de Falconer (1960), propuso para el cálculo de la heredabilidad una generalización que contempla cualquier par colateral relacionado por parentesco o por crianza, teniéndose en un extremo a los gemelos univitelinos que comparten 100% de su carga genética (G) y en el otro extremo a los pares de niños adoptados del mismo sexo que, idealmente, si sus padres no tienen antepasados comunes, tienen una carga genética compartida que se aproxima a cero (0%).

Jensen (1967) generalizó esta fórmula, proponiendo:

$$h^2 = (r_{xx} - r_{yy}) / (G_{xx} - G_{yy})$$

con $r_{xx} > r_{yy}$ y $G_{xx} > G_{yy}$

donde r_{xx} : correlación entre el par del parentesco xx criados juntos, r_{yy} : correlación entre el par del parentesco yy criados juntos, G_{xx} : proporción promedio de carga genética compartida por el conjunto de pares xx, y G_{yy} : proporción promedio de carga genética com-

partida por el conjunto de pares yy.

De esa manera se obtienen las siguientes heredabilidades:

$h^2 = (r_{ij} - r_{ff}) / (1,00 - 0,50)$ (Expresión de Falconer), utilizando gemelos univitelinos y gemelos bivitelinos,

$h^2 = (r_{ff} - r_{pp}) / (0,50 - 0,00)$ utilizando gemelos bivitelinos y pares adoptados, y

$h^2 = (r_{ij} - r_{pp}) / (1,00 - 0,00)$ utilizando gemelos univitelinos y pares adoptados.

Las críticas más persistentes a la aplicación de estas fórmulas se refieren a las proporciones promedios de cargas genéticas compartidas (Taylor, 1980) y al supuesto de que la influencia del ambiente común compartido sea igual para los dos elementos del par en todos los tipos de parentesco (Loehlin, 1989). Para un conjunto de gemelos bivitelinos, aunque en promedio comparten 50% de la carga genética, en situaciones empíricas se dan factores que incrementan este valor, tal como las afinidades electivas entre los padres, como por ejemplo, matrimonios consanguíneos. Adicionalmente, los dos niños deben ser del mismo sexo. De manera similar para pares adoptados, aunque teóricamente sería nula la proporción promedio de carga genética compartida, en una situación empírica real el valor se incrementaría por la escogencia de los dos niños a ser adoptados y la posibilidad de que exista consanguinidad remota entre ellos. Igualmente, ambos niños deben ser de igual sexo.

A partir de Rao *et al.* (1974) todos los estudios donde fue usada la comparación de pares relacionados por crianza o parentesco se ubican dentro del método general conocido como modelaje de ecuaciones estructurales (SEM). De manera particular, para la estimación de las correlaciones utilizan la técnica conocida como análisis de ruta (MacCallum y Austin, 2000). Por cada par de gemelos, parentescos o adoptados criados juntos en ambiente común compartido, se genera mediante la técnica del análisis de ruta la siguiente ecuación básica como modelo teórico:

$$r_{ij} = (c_i c_j) \times rA_i A_j + (h_i h_j) \times rG_i G_j + (c_i h_i) \times rA_i G_i + (c_j h_j) \times rA_j G_j$$

Donde c representa el efecto sobre la variable fenotípica del ambiente común compartido, h representa el efecto sobre la variable fenotípica de la herencia y rAA , rGG y rAG representan constructos teóricos, no observables y/o medibles. Se asignan valores de acuerdo a la teoría probabilística de la herencia y a la condición aleatoria de la muestra que implica independencia entre las variables.

Cada ecuación expresa el valor teórico del coeficiente de correlación de Pearson, conformado a partir de rutas y correlaciones que son de carácter teórico y/o hipotético, y que se estructuran entre las dos varia-

bles observadas (t_i , t_j) por una parte, y un conjunto de variables no observadas o latentes, por la otra. Así, $c_i = c_j$; $h_i = h_j$; $r_{GG} = 1,00$ en gemelos univitelinos; $r_{GG} \sim 0,50$ en promedio, para gemelos bivitelinos y hermanos; $r_{GG} \sim 0,25$ en promedio, para primos; $r_{GG} \sim 0,00$ en promedio, para pares de adoptados; $r_{AG} = 0,00$ por correlación herencia/ambiente $rA_i G_i = rA_j G_j$; y $r_{AA} = 1,00$ por definición de ambiente común compartido

La manera como se constituye el programa de modelaje de ecuaciones estructurales es que se tienen valores observados del coeficiente de correlación, uno para cada parentesco y/o adoptados, y en cada ecuación se estima r_{GG} según la teoría genética. Se tiene un sistema sobredeterminado si el número de coeficientes de correlación empíricamente observados es mayor que el número de incógnitas.

La manera como funcionan los procedimientos de estimación de las incógnitas para estas situaciones, en las cuales existe un mayor número de valores empíricamente calculados que de incógnitas, es mediante procedimientos iterativos. Utilizando programas de computación, se van asignando a las incógnitas ciertos valores inicialmente arbitrarios, que luego se van cambiando y ajustando, hasta que los valores teóricos de correlación, resultantes dentro del modelo en comparación con los valores empíricamente observados de esa misma correlación, en los gemelos, parentescos y adoptados, se ajusten lo más posible entre sí. Se utiliza un criterio estadístico de ajuste, siendo el más usual el conocido como la técnica de los mínimos cuadrados (Rao *et al.*, 1974; Eaves, 1975).

La evaluación de la bondad del ajuste logrado se expresará mediante el cálculo del estadístico χ^2 , con un total de grados de libertad igual a la diferencia entre el número de valores empíricamente observados y el número de incógnitas a ser estimadas. El valor asociado de probabilidad (p) de la prueba χ^2 expresaría el mínimo nivel de significación (probabilidad de error del primer tipo), con el cual se puede rechazar la hipótesis de nulidad. En caso de no ser estadísticamente significativa esa diferencia, se podría inferir que los valores de las incógnitas pueden ser considerados estimaciones consistentes de los parámetros/constructos correspondientes. Ello no implica, sin embargo, que se pueda afirmar que los valores encontrados son los correctos y/o los únicos posibles (Bentler y Bonett, 1980). Esta observación puede considerarse como la principal limitación de este procedimiento.

El Enfoque Ambiental

Desde el ángulo metodológico de la consideración y medición ambiental, Contasti (1996) propone una fórmula metodológica cuya estructura estadística es simi-

lar a un caso particular de la fórmula general de Jensen (1967). El caso particular de las fórmulas propuestas por Jensen sería aquel en el cual se compara a los gemelos univitelinos criados juntos con pares de adoptados del mismo sexo y edad, también criados juntos desde su nacimiento por la misma familia. La fórmula de Jensen es $h^2 = (r_{jj} - r_{pp}) / (G_{jj} - G_{pp})$. Asumiendo que para los gemelos univitelinos $G_{jj} = 1$ y que para los pares adoptados $G_{pp} = 0$, la fórmula se transforma en $h^2 = r_{jj} - r_{pp}$.

Contasti (1996) propuso utilizar, en lugar del coeficiente r_{pp} , una correlación intraclase equivalente, en la cual se sustituye la medición de la variable fenotípica (t) en pares adoptados criados juntos, por una medición nominal del ambiente. Esta medición debe indicar ambientes similares y/o equivalentes en las familias de un grupo numeroso de niños y/o sujetos de la misma edad y sexo. Al obtener una caracterización nominal del ambiente para una muestra aleatoria, los resultados serán sometidos a un procesamiento estadístico pertinente según el procedimiento establecido, para obtener una estimación equivalente a r_{pp} . La fórmula metodológica propone categorizar nominalmente el ambiente, de tal forma que para todos los sujetos que pertenezcan a una determinada categoría nominal se asume que comparten un ambiente similar, que se corresponde con el ambiente común compartido del par de adoptados criados juntos. Para todos los sujetos 1,2,..., j,..., n, que pertenecen a la categoría nominal i, se asume: $A_{i1} = A_{i2} = \dots = A_{ij} = \dots = A_{in}$, y más específicamente en relación con el efecto de ese ambiente A_j , en la variable t, consecuentemente se supone que

$$t(A_{i1}) = t(A_{i2}) = \dots = t(A_{ij}) = \dots = t(A_{in})$$

Contasti (1996) describe un procedimiento específico para el cálculo de un coeficiente de correlación intraclase. Para aplicar este procedimiento, será necesario que se transforme la variable categorial de ambiente (A) en una variable (s) de nivel métrico. Esta manera de calcular un coeficiente de correlación intraclase será diferente al procedimiento clásico y tradicional como una relación entre dos varianzas. La ventaja del procedimiento, hacia el futuro, será que hará más expedito el cálculo de un coeficiente de correlación intraclase de carácter múltiple, es decir, con dos o más variables independientes resultará de elevar al cuadrado a un coeficiente de correlación lineal de Pearson r_{ts} entre la variable fenotípica de interés (t) y una variable (s) a ser construida. Esta variable s representará una transformación métrica de las categorías nominales; para construirla se tomarán como valores asignados de ella, para todos los sujetos j que pertenecen a una categoría nominal i, precisamente el promedio (t^*) de los valores de la variable t de todos los sujetos j que pertenecen a esa categoría nominal i. Por tanto, $S_{ij} = t_i^*$ para $i=1 \dots i \dots k$. En general t_i^* distinto t_j^* .

Nótese que mientras el coeficiente de correlación intraclase I_{ts} , al igual que el coeficiente de doble entrada r_{pp} , tiene la misma estructura de un coeficiente de determinación (r^2); en cambio r_{ts} , como coeficiente de correlación lineal de Pearson entre dos variables diferentes, debe ser elevado al cuadrado para que adquiera la misma estructura de un coeficiente de correlación intraclase y que por tanto signifique una proporción de la varianza explicada. Para entender mediante una aproximación conceptual el porqué el valor r_{ts}^2 representa un coeficiente de correlación intraclase, vale analizar el comportamiento de la variable construida (s) en dos situaciones límites: 1) cuando el coeficiente de correlación intraclase sea igual a cero, y 2) cuando el coeficiente de correlación intraclase sea igual a uno. Es conocido que r_{ts}^2 representa la proporción de la varianza de la variable dependiente t que es explicada por las fluctuaciones de la variable independiente s. Entonces

$$I = V_e / V_{tot} = V_e / (V_e + V_i)$$

donde V_e : varianza entre las medias (t^*) de las k categorías nominales i; V_i : promedio ponderado de las varianzas internas (V_i en las variable t en las k categorías nominales i).

Caso 1: $I_{ts} = 0$. Si $V_i = V_{tot}$, entonces $V_e = 0$ e $I_{ts} = 0$. Este caso se dará cuando la distribución de la variable dependiente (t) es idéntica para todas y cada una de las k categorías nominales i. La varianza interna (V_i) en cada una de las k categorías nominales i, será igual a la varianza total (V_{tot}) de la distribución de la variable dependiente (t) en el colectivo, razón por la cual, para cualquier valor individual t, en cualquiera de las k categorías nominales i, le correspondería en la variable s siempre un valor $s = t^*$, que será el mismo en todas las k categorías nominales i. Para todo ij se tendrá $t_i^* = t_j^*$. Nótese entonces que $r_{it}^* = 0$, $r_{ts} = 0$ y $r_{ts}^2 = 0$.

Caso 2: $I_{ts} = 1$. Si $V_i = 0$, entonces $I_{ts} = 1$. Este caso se dará cuando las varianzas internas V_i , en todas las k categorías nominales i son iguales a cero. Por ello, en cualquiera de las k categorías nominales i para todos los sujetos j de esa categoría nominal i, se tendrá un mismo un valor constante de la variable t. Este valor será distinto para las k categorías nominales i, y en cada caso será idénticamente igual al promedio de t^* , ($s = t^*$), para todo valor s dentro de cada una de las k categorías nominales i. Nótese entonces que $r_{it}^* = 1$, $r_{ts} = 1$, $r_{ts}^2 = 1$.

El procedimiento de cálculo de la correlación r_{ts} , tal como se hace explícito, constituye una fórmula metodológica que para ser operacionalizada pertinentemente exige que se construyan las categorías nominales, de tal forma que, en cuanto al ambiente, sean lo más perfectamente homogéneas. Esta con-

dición, en las correlaciones donde se utilizan pares relacionados por crianza, se logra cuando los dos niños adoptados son criados juntos desde su nacimiento. Por ello, el problema instrumental para operacionalizar la fórmula metodológica de Contasti consiste en cómo construir esas categorías nominales de manera homogénea. Contasti (1996), para presentar un ejemplo práctico, donde se mostrase principalmente el uso y funcionamiento estadístico de la fórmula, propone una medición nominal que implicaría, más que el concepto de ambiente común compartido entre adoptados, el concepto de ambiente equivalente de las familias, en el sentido que para establecer la similitud del ambiente (categorías nominales) utiliza datos obtenidos con un instrumento para la medición del nivel socioeconómico de las familias, denominado según el nombre de su autor como escala Graffar (1956), que para Venezuela fue adaptada por Méndez Castellano (1982). Dado que esta escala no fue diseñada con el objetivo de evaluar similitud en el ambiente de las familias, ello constituye una limitación de la aplicación del procedimiento en cuanto a la precisión de la estimación de los valores numéricos particulares de los parámetros, pero ello no le resta validez estadística a la fórmula metodológica propuesta.

Conclusiones y Recomendaciones

Desde los estudios de Galton (1883) hasta las formulaciones de Jensen (1971) el problema fue enfocado hacia el cálculo de la heredabilidad, utilizando gemelos, parentescos y niños adoptados. Esta tendencia ha continuado (MacCallum y Austin, 2000), siendo por lo tanto el enfoque predominante hasta el presente, cuya mayor limitación es la sobrestimación sistemática del coeficiente de heredabilidad obtenido.

A partir de Rao *et al.* (1974) se toma en consideración la medición del ambiente, aunque solo fuese como un factor de ajuste para compensar estadísticamente diferencias ambientales entre los miembros del par. Nótese que al tener como elemento básico la utilización de pares se mantiene el enfoque genético inicial en la medición de la heredabilidad. Con la fórmula metodológica de Contasti (1996) se cambia el enfoque y se pasa a tratar de medir directamente el ambiente, posibilidad que ya había sido prevista y promocionada por Loehlin (1989) y por Plomin y Rende (1991). Debe notarse como una limitación de carácter instrumental, que la operacionalización propuesta por Contasti como ejemplo del uso de su fórmula metodológica, implica una medición muy cruda e indirecta del ambiente familiar mediante factores de carácter socioeconómico que constituyen la escala Graffar.

En la propuesta para el cálculo de la correlación r_{ts} de Contasti, a diferencia del uso de r_{pp} , que está signada por la

homogeneización del ambiente, mediante la consideración de los pares de niños adoptados criados juntos, al categorizar el ambiente en una escala nominal se permite que, de manera progresiva y perfectible, por una parte, se pueda mejorar continuamente la medición nominal del ambiente y, por la otra, que pueda ser utilizada en poblaciones más amplias y sin restricciones, excepto la aleatoriedad de la muestra escogida.

Lógicamente, el uso de mediciones crudas y poco elaboradas en relación al ambiente estará generando una subestimación de la ambientalidad y, en consecuencia, una sobrestimación de la heredabilidad, pero a su vez se posibilita que, sistemáticamente, en el futuro, al incorporar e integrar mediciones múltiples, diversas y más precisas del ambiente, este sesgo se pueda reducir continuamente. En este sentido será necesario prever y diseñar un coeficiente de correlación intraclase de carácter múltiple, que integre dos o más mediciones o dimensiones del ambiente.

Desde una perspectiva epistémica, se puede reiterar (Contasti, 1996) que históricamente, al utilizar en los estudios de heredabilidad la crianza en común como indicador de similitud en el ambiente de gemelos, otros parentescos y en

pares de niños adoptados, este problema se ha venido trabajando con métodos observacionales propios de una ciencia natural básica. Al incorporar una medición nominal del ambiente se están utilizando métodos constructivos y operacionales, propios de una ciencia artefactual tecnológica.

REFERENCIAS

- Bentler PM, Bonett DG (1980) Significance test and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88: 588-606.
- Contasti M (1996) Fórmula Metodológica Herencia/Ambiente para variables biológicas y sociales. *Informe Inv. Educ.* X: 51-56.
- Eaves LJ (1975) Testing models for variation in intelligence. *Heredity* 34: 132-136.
- Falconer DS (1960) *Introduction to Quantitative Genetics*. Ronald. Nueva York, EEUU. 365 pp.
- Galton F (1883) *Inquiries into the Human Faculty and its Development*. MacMillan. Londres, RU. 387 pp.
- Graffar M (1956) Une methode de classification sociales d'echantillons de population. *Courrier VI*: 445-459.
- Holzinger KJ (1929) The relative effect of nurture and nature influences on twin differences. *J. Educ. Psychol.* 20: 241-248.
- Jensen AR (1967) Estimation of the limits of heritability of traits by comparison of monozygotic and

dizygotic twins. *Proc. Nat. Acad. Sci.* 58: 149-156.

- Jensen AR (1971) A note on why genetic correlations are not squared. *Psychol. Bull.* 75: 23-24.
- Loehlin JC (1989) Partitioning environmental and genetic contributions to behavioral development. *Am. Psychol.* 44: 1285-1292.
- MacCallum RC, Austin JT (2000) Applications of structural equation modelling in psychological research. *Annu. Rev. Psychol.* 51: 201-226.
- Méndez Castellano H (1982) *Método Graffar Modificado para Venezuela*. Fundacredesa. Caracas, Venezuela. 56 pp.
- Nunnally J (1967) *Psychometric Theory*. McGraw-Hill. Nueva York, EEUU. 640 pp.
- Plomin R, Rende R (1991) Human behavioral genetics. *Annu. Rev. Psychol.* 42: 161-190.
- Prescott CA, Kendler KS (1996) Longitudinal stability and change in alcohol consumption among female twins: contributions of genetics. *Dev. Psychopathol.* 8: 849-866.
- Rao DC, Morton NE, Yee S (1974) Analysis of family resemblance II. A linear model for familial correlation. *Am. J. Human Genet.* 26: 331-359.
- Saudino KF, Mcguire S, Reiss D, Hetherington EM, Plomin R (1995) Parent ratings of EAS temperaments in twins, full siblings, half siblings, and step siblings. *J. Pers. Soc. Psychol.* 68: 723-733.
- Taylor HF (1980) *The IQ Game: A Methodological Inquiry into the Heredity-Environment Controversy*. Rutgers University Press. New Brunswick, NJ, EEUU. 276 pp.

STATISTICAL FORMULAS AND PROCEDURES TO MEASURE IN HUMAN POPULATIONS THE INFLUENCE OF HEREDITY/ENVIRONMENT IN SOME CONTINUOUS METRICAL VARIABLES

Max Contasti and Edgar Matheus

SUMMARY

The object of this work consists of the study and the analysis, from a statistical viewpoint, of different procedures and formulas used to calculate in human populations, the influence of inheritance/environment in some continuous phenotypic variables that can be measured with acceptable precision, such as stature, weight, arterial pressure, and intelligence. The document presents the analysis of the most important statistical and methodological limitations on the formulations considered and how, according to

the cases, these limitations have been corrected and other new procedures of calculation have been generated to minimize them. From an epistemic and methodological point of view, up to 1974 research was oriented by the genetic approach that calculates the influence of the inheritance by means of the use of data on twins or other members of the family group. After that, the environment began to be considered like an additional variable for the statistical formulation in the analysis.

PROCEDIMENTOS E FÓRMULAS ESTATÍSTICAS PARA MEDIR A INFLUÊNCIA HERANÇA/AMBIENTE SOBRE CARACTERES MÉTRICOS EM POPULAÇÕES HUMANAS

Max Contasti e Edgar Matheus

RESUMO

O objetivo deste trabalho consiste no estudo e a análise, desde o ponto de vista estatístico, dos diversos procedimentos e fórmulas propostos para calcular, nas populações humanas, a influência da herança/ambiente em algumas variáveis fenotípicas de tipo contínuo que se possam medir com uma precisão aceitável, tais como estatura, peso, pressão arterial e inteligência. Apresenta-se a análise das limitações mais relevantes na formulação estatística considerada e como, segundo os casos, tem sido corrigidas as limitações observadas e gerado novos procedimentos

de cálculo para minimizar essas limitações estatísticas e metodológicas. Desde o ponto de vista epistêmico e metodológico, até 1974 as investigações estiveram orientadas unicamente pelo enfoque genético, que calcula a influência da herança através do uso de dados obtidos de gêmeos ou de outros membros do grupo familiar. Posteriormente o ambiente começou a ser considerado como uma variável adicional para a formulação estatística na análise.