

DESEMPENHO DO TESTE DA HERDABILIDADE UNIVARIADA MEDIANTE DELINEAMENTOS COM DIFERENTES ESTRUTURAS FAMILIARES

Alex de Oliveira RIBEIRO¹
Daniel Furtado FERREIRA¹
Júlia Maria Pavan SOLER²

- RESUMO: Estudos sobre a herdabilidade de caracteres quantitativos são essenciais para o melhor entendimento da contribuição genética a fatores associados a doenças, ao comportamento humano, dentre outros. Para essa finalidade, os delineamentos com famílias são bastante empregados, os quais possibilitam a captura de informações sobre a resposta de interesse em conjuntos específicos de parentes ou em famílias inteiras. Neste artigo, o efeito da variação da estrutura e do tamanho das famílias sobre o desempenho do teste de herdabilidade univariado proposto por Blangero *et al.* (2013) é avaliado por meio das taxas de erro tipo I e do poder do teste. Os resultados mostram que o teste é conservativo, independentemente do tipo de estrutura familiar utilizada, do tamanho das famílias e da relação de parentesco entre seus membros. Quanto ao poder, os resultados das simulações realizadas indicaram que delineamentos com irmandades extensas conduzem a um aumento no poder do teste de herdabilidade. Isto se deve a maior quantidade de relacionamentos diretos presentes nestes delineamentos, além do fato de que na decomposição espectral da matriz de parentesco este delineamento contribui com a maior proporção de autovalores menores que 1, os quais aumentam o valor da estatística de teste da razão de verossimilhanças (ELRT).
- PALAVRAS-CHAVE: Decomposição espectral; erro tipo I; herdabilidade; matriz de parentesco; poder.

¹Universidade Federal de Lavras - UFLA, Departamento de Estatística, CEP: 37.200-000, Lavras, MG, Brasil. E-mail: alex.ribeiro@ufla.br; danielff@ufla.br

²Universidade de São Paulo - USP, Instituto de Matemática e Estatística, CEP: 05508-090, São Paulo, SP, Brasil. E-mail: pavan@ime.usp.br

1 Introdução

O coeficiente de herdabilidade tem papel fundamental no entendimento da contribuição genética sobre caracteres quantitativos, envolvidos direta ou indiretamente na regulação de doenças, tanto em humanos como em animais e plantas. É definido como a proporção da variância total fenotípica resultante do efeito de fatores genéticos aditivos (VISSCHER *et al.*, 2008). A realização de inferências sobre a herdabilidade tem sido feita, principalmente, a partir de dados provenientes de delineamentos com indivíduos relacionados, como famílias, por exemplo, os quais possibilitam a captura de informações sobre a doença em conjuntos específicos de parentes (irmãos, pai-filho, meio-irmãos etc) ou em famílias inteiras (famílias nucleares ou famílias estendidas) (HAINES e PERICAK, 1998).

O modelo linear misto poligênico tem sido amplamente utilizado em estudos de herdabilidade no qual a estrutura familiar declarada pelo indivíduo é usada na definição das correlações entre as observações (ALMASY e BLANGERO, 1998, 2010; LANGE, 2002). A estrutura de covariância total Σ desse modelo, em geral, é particionada em um componente genético $2\Phi^{(j)}\sigma_g^2$ e outro ambiental $\mathbf{I}\sigma_e^2$, sendo $\Phi^{(j)}$ e \mathbf{I} as matrizes de parentesco (*kinship*) e identidade, respectivamente, cuja ordem é determinada pelo número de indivíduos da j -ésima família. Considerando a análise de um fenótipo específico, σ_g^2 e σ_e^2 representam as variâncias genética e ambiental, respectivamente. Nestes estudos, o foco está centrado na obtenção de estimativas dos seus componentes da variância σ_g^2 e σ_e^2 , pois através delas, é possível chegar a uma estimativa do coeficiente de herdabilidade (h^2) do fenótipo.

Ressalta-se que, devido à dependência entre os indivíduos de uma mesma família, a obtenção dessas estimativas requer considerável esforço computacional, especialmente, quando há muitos fenótipos e as famílias são muito extensas e desbalanceadas. Nesse sentido, Blangero *et al.* (2013), sob a formulação do modelo misto poligênico para um único fenótipo, propuseram o uso da decomposição espectral da matriz de covariâncias do modelo, simplificando assim a verossimilhança obtida. O resultado deste processo é um produto de normais univariadas, comparável à verossimilhança de dados não correlacionados. A simplificação proposta por esses autores possibilitou a obtenção de uma expressão analítica simples para testar o coeficiente de herdabilidade de uma única característica quantitativa baseada na esperança da razão de verossimilhanças, dada em função dos autovalores encontrados na decomposição espectral da matriz de correlação genética ($2\Phi^{(j)}$) e da estimativa de h^2 . Devido à simplicidade e facilidade de aplicação, esse método tem sido empregado por vários pesquisadores envolvidos com genética quantitativa (BRUCE *et al.*, 2017; CANTOR *et al.*, 2016; CLARK *et al.*, 2016, JUN *et al.*, 2018).

Para avaliar o desempenho do teste proposto, Blangero *et al.* (2013) usaram simulação bootstrap, considerando cenários com diferentes estruturas familiares, desde gêmeos monozigóticos a famílias extensas. Porém, esses autores padronizaram a estrutura familiar em cada simulação feita, ou seja, cada amostra gerada foi constituída por réplicas de uma mesma estrutura familiar. A acurácia do método foi

medida por meio do poder do teste para valores de herdabilidade variando dentro do seu espaço paramétrico. Seus resultados mostraram que delineamentos com famílias extensas são mais poderosos na região do espaço paramétrico próxima à hipótese nula, isto é, para herdabilidades pequenas. Entretanto, para valores de herdabilidade maiores que 0,47, delineamentos com gêmeos monozigóticos apresentaram maiores níveis de poder.

Estudos de herdabilidade têm sido conduzidos nas diferentes populações mundiais, envolvendo delineamentos com famílias constituídas por estruturas de relacionamento, número de indivíduos e gerações, diferentes em uma mesma amostra. Lin *et al.* (2005), por exemplo, estudaram a herdabilidade de fatores associados à Síndrome Metabólica com dados do *Northern Manhattan Family Study*, cuja amostra envolveu 803 pacientes de 89 famílias hispano-caribenhas com número de indivíduos por família variando de 3 a 53. Outro exemplo é o estudo realizado por Bosy-Westphal *et al.* (2008) em que foi avaliado o gasto de energia em repouso de 815 indivíduos participantes do *Kiel Obesity Prevention Study*. Neste estudo, o tamanho das famílias, bem como as relações de parentesco entre seus integrantes, se caracterizaram por forte desbalanceamento. No Brasil, o projeto Corações de Baependi (OLIVEIRA, 2008; ALVIM 2016) é pioneiro em avaliar muitas famílias extensas e desbalanceadas, tendo, inicialmente, como finalidade o estudo da herdabilidade de fenótipos associados a doenças cardiovasculares.

Quando o delineamento contempla famílias com tamanhos e estruturas de relacionamento diferentes, a decomposição espectral da matriz de covariâncias genética, proposta por Blangero *et al.* (2013), gera um conjunto de autovalores com maior variabilidade do que nos casos em que as famílias são balanceadas, isto é, fixadas em sua estrutura e tamanho. Como a estatística de teste da herdabilidade é função desses autovalores, variações na sua distribuição podem conduzir a oscilações nos seus níveis de poder.

Assim, neste artigo, pretende-se avaliar o efeito da variação da estrutura e do tamanho das famílias em uma mesma amostra, sobre o desempenho do teste de herdabilidade univariado proposto por Blangero *et al.* (2013). O procedimento adotado está dividido em três etapas, a saber: i) famílias de tamanhos e estrutura fixadas comparadas com famílias de tamanhos fixos e com estruturas de relacionamento diferentes; ii) famílias com tamanhos e estrutura fixados comparadas com famílias de estrutura e tamanhos variados; e, iii) famílias com tamanho fixo e estrutura variada comparadas com famílias de tamanho e estrutura variadas.

2 Método

2.1 Modelo linear misto poligênico

Para dados de famílias humanas (ou *pedigrees*), o modelo poligênico para um único fenótipo de interesse pode ser explicitado por:

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times q} \boldsymbol{\beta}_{q \times 1} + \mathbf{g} + \mathbf{e} \quad (1)$$

em que \mathbf{Y} , representa o vetor com os valores do fenótipo de interesse, com $n = \sum_{j=1}^F n_j$ e n_j o tamanho da j -ésima família ($j=1, \dots, F$), \mathbf{X} é a matriz do delineamento de posto completo, $\boldsymbol{\beta}$ é o vetor de parâmetros fixos e \mathbf{g} e \mathbf{e} são vetores de efeitos aleatórios associados aos componentes genético e ambiental (ou de erro), respectivamente.

Neste modelo, considera-se que os componentes genético e ambiental são não correlacionados, além de que $E(\mathbf{g}) = E(\mathbf{e}) = \mathbf{0}$, $Cov(\mathbf{g}) = \mathbf{K}\sigma_g^2$, $Cov(\mathbf{e}) = \mathbf{I}_n\sigma_e^2$, sendo $\mathbf{K} = 2\boldsymbol{\Phi}$ a matriz de parentesco (geralmente, uma matriz bloco diagonal com cada bloco construído a partir da informação das relações entre os indivíduos da família) e σ_g^2 e σ_e^2 os componentes da variância fenotípica, tal que, $Var(Y_{ij}) = \sigma_Y^2 = \sigma_g^2 + \sigma_e^2$, com $i = 1, 2, \dots, n_j$ e $j = 1, 2, \dots, F$. A matriz de covariâncias fenotípica $\boldsymbol{\Sigma}$ para o modelo poligênico (1) é:

$$Cov(\mathbf{Y}) = \boldsymbol{\Sigma} = \mathbf{K}\sigma_g^2 + \mathbf{I}_n\sigma_e^2$$

em que \mathbf{I}_n é a matriz identidade de ordem $n \times n$. Esta matriz de covariâncias pode ser reescrita em função do coeficiente de herdabilidade h^2 ($h^2 = \sigma_g^2/\sigma_Y^2$) e da matriz de parentesco, chegando-se à seguinte expressão:

$$Cov(\mathbf{Y}) = \boldsymbol{\Sigma} = \mathbf{K}[\sigma_Y^2 h^2] + \mathbf{I}_n[\sigma_Y^2(1 - h^2)]. \quad (2)$$

Assumindo, sem perda de generalidade, que $\sigma_Y^2 = 1$, tem-se:

$$Cov(\mathbf{Y}) = \boldsymbol{\Sigma} = \mathbf{K}h^2 + \mathbf{I}_n(1 - h^2). \quad (3)$$

Assim, pode-se observar que na expressão (3) a matriz de covariâncias fenotípica é particionada na soma ponderada de duas matrizes, sendo que a primeira expressa a variabilidade associada ao efeito genético e a segunda ao efeito ambiental ou do erro.

É interessante observar, que o modelo poligênico indicado em (1) pode ser expresso como um modelo linear misto clássico (SEARLE et al., 1992), fazendo $\mathbf{g} = \mathbf{Z}\boldsymbol{\gamma}$, tal que

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times q}\boldsymbol{\beta}_{q \times 1} + \mathbf{Z}_{n \times n}\boldsymbol{\gamma}_{n \times 1} + \mathbf{e}_{n \times 1} \quad (4)$$

com \mathbf{Z} uma matriz de delineamento de posto completo satisfazendo $\mathbf{Z}\mathbf{Z}^\top = 2\boldsymbol{\Phi}$, $\boldsymbol{\gamma}$ um vetor de efeitos aleatórios associados ao componente genético com $E(\boldsymbol{\gamma}) = \mathbf{0}$ e $Cov(\boldsymbol{\gamma}) = \mathbf{I}_n\sigma_g^2$. Hazelton and Gurrin (2003) apresentam uma formalização do modelo linear misto poligênico definindo a matriz \mathbf{Z} em termos de trios (parentesco entre pai, mãe e filho).

2.2 Estatística de teste baseada na esperança da razão de verossimilhanças - ELRT

Blangero *et al.* (2013) propuseram um teste, analiticamente simples, para a herdabilidade de uma característica ou fenótipo quantitativo. Em função das relações de dependência observadas entre os indivíduos de uma mesma família, esses autores realizaram uma transformação linear no vetor de resíduos do modelo (1), $\boldsymbol{\delta} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$, partindo da decomposição espectral da matriz de covariâncias ($\boldsymbol{\Sigma}$):

$$\boldsymbol{\Sigma} = \mathbf{S}\mathbf{D}_p\mathbf{S}' = \mathbf{S}[h^2\mathbf{D}_g + (1 - h^2)\mathbf{I}]\mathbf{S}' = \mathbf{S}[\mathbf{I} + h^2(\mathbf{D}_g - \mathbf{I})]\mathbf{S}', \quad (5)$$

em que \mathbf{S} é uma matriz ortogonal de autovetores e $\mathbf{D}_p = \text{diag}\{\lambda_{p_i}\}$ e $\mathbf{D}_g = \text{diag}\{\lambda_{g_i}\}$ são matrizes diagonais de autovalores fenotípicos e genético, respectivamente.

A partir dessa decomposição, a matriz de autovetores \mathbf{S} propicia condições para a condução da transformação linear no vetor de resíduos dependentes $\boldsymbol{\delta}$ do modelo, gerando um vetor $\boldsymbol{\tau} = \mathbf{S}'\boldsymbol{\delta}$, constituído por resíduos independentes. Considerando um fenótipo normalmente distribuído, cujo vetor de resíduos transformados $\boldsymbol{\tau}$ segue uma distribuição normal multivariada com vetor de média $\mathbf{S}'\boldsymbol{\mu}$, com $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, e matriz de covariância $\mathbf{S}'\boldsymbol{\Sigma}\mathbf{S}$, tem-se que da teoria multivariada padrão (STUART e ORD, 1987), a verossimilhança total pode ser fatorada como a soma entre a verossimilhança dos resíduos transformados e a verossimilhança da transformação, expressa pelo Jacobiano $J_{\boldsymbol{\delta} \rightarrow \boldsymbol{\tau}}$. Adotando-se esse procedimento, chega-se à função suporte, dada por:

$$\ln L(\boldsymbol{\beta}, h^2 \mid \mathbf{Y}, \mathbf{X}) = -\frac{1}{2} \sum_{i=1}^n \ln [1 + h^2(\lambda_{g_i} - 1)] - \frac{1}{2} \sum_{i=1}^n \frac{\tau_i^2}{1 + h^2(\lambda_{g_i} - 1)}. \quad (6)$$

É importante ressaltar que a decomposição espectral da matriz de parentesco \mathbf{K} é suficiente para a obtenção desta transformação, pois os autovetores de $\boldsymbol{\Sigma}$ são também autovetores de \mathbf{K} (THOMPSON e SHAW, 1990, 1992), e os autovalores de $\boldsymbol{\Sigma}$ podem ser expressos como combinação linear dos autovalores de \mathbf{K} .

Com a simplificação da verossimilhança, convertida em um produto de normais univariadas, similar ao caso de observações não correlacionadas, Blangero *et al.* (2013) chegaram à estatística de teste para a herdabilidade univariada. Neste processo, esses autores mostraram que a função (6) pode ser reduzida apenas ao primeiro somatório nela expresso, pois a esperança do segundo somatório conduz ao seu cancelamento em (6). Assim, após a retirada deste termo, a avaliação da função de verossimilhança resultante, sob as hipóteses nula ($h^2 = 0$) e alternativa ($h^2 > 0$) através da esperança da razão de verossimilhanças, conduz à estatística de teste (ELRT), expressa por:

$$ELRT[h^2 > 0] = - \sum_{i=1}^n \ln[1 + h^2(\lambda_{g_i} - 1)]. \quad (7)$$

Sob a hipótese nula ($h^2 = 0$), esta estatística é distribuída assintoticamente como uma mistura, em iguais proporções, de uma qui-quadrado com ponto de massa no zero e uma qui-quadrado com um grau de liberdade, $1/2\chi_0^2 + 1/2\chi_1^2$, (SELF e LIANG, 1987).

É importante destacar a contribuição dos autovalores λ_{g_i} na magnitude do resultado da estatística de teste (7). Autovalores menores que 1 contribuem positivamente para a ELRT, aumentando sua magnitude, enquanto autovalores maiores que 1 contribuem negativamente, reduzindo sua magnitude. Portanto, como esses autovalores provêm da decomposição espectral da matriz de parentesco (covariância genética), a estrutura familiar tem influência no resultado final desta estatística.

2.3 Estruturas familiares e simulação

A influência da estrutura familiar na acurácia do teste para a herdabilidade univariada foi avaliada, neste trabalho, por meio de simulação Monte Carlo. A variável, representando o fenótipo de interesse, foi gerada seguindo uma distribuição normal multivariada com vetor de média igual a $\mathbf{1}_n * 120$ e matriz de covariância formulada como em (2) com variância total igual a 25. Esses valores visaram retratar o comportamento de uma variável próxima aos parâmetros de uma pressão arterial sistólica, por exemplo. Para as herdabilidades (h^2) foram atribuídos os valores 0, 0,15, 0,20, 0,30, 0,40, 0,50 e 0,70, os quais possibilitaram a avaliação tanto das taxas de erro tipo I, quanto dos níveis de poder do teste.

Quanto às estruturas familiares definindo a matriz \mathbf{K} na expressão (2), foram consideradas três estruturas básicas, Família 1, 2 e 3, considerando desde pedigrees pequenos (Família 1 com 6 indivíduos), famílias nucleares de tamanho médio (Família 2 com 12 indivíduos) e pedigrees mais extensos constituídos por 56 indivíduos (Família 3). Os heredogramas dessas famílias encontram-se na Figura 1.

A partir da definição das estruturas familiares básicas, planejou-se os cenários para o estudo. Os três primeiros, denominados Estrutura 1, 2 e 3, avaliaram separadamente cada uma das estruturas básicas apresentadas na Figura (1), sendo a amostra total de cada cenário construída de forma balanceada e fixada, isto é, constituída por réplicas de cada estrutura. Para ampliar os cenários de avaliação do desempenho do teste, foi considerada ainda, uma quarta estrutura de dados balanceados, Estrutura 4, com núcleos familiares todos iguais (fixados) definidos de forma similar à Estrutura 1, porém compostos pelos pais e 8 filhos. Outros três cenários foram considerados, incluindo balanceamento no tamanho das famílias mas impondo estrutura familiar variável, isto é, incluindo diferentes graus de parentesco entre seus membros. Estes cenários foram denominados como Estrutura 5, 6 e 7, cada um, considerando famílias com o mesmo número de indivíduos como descrito na Figura (1), isto é, com 6, 12 e 56 membros, respectivamente. Assim, uma das famílias que compõem a amostra da estrutura 5, cujo tamanho tem por base a família F1, poderia ter, por exemplo, três gerações (pai, mãe, dois filhos sendo um deles casado e um neto) e não duas. Um último cenário, Estrutura 8, foi

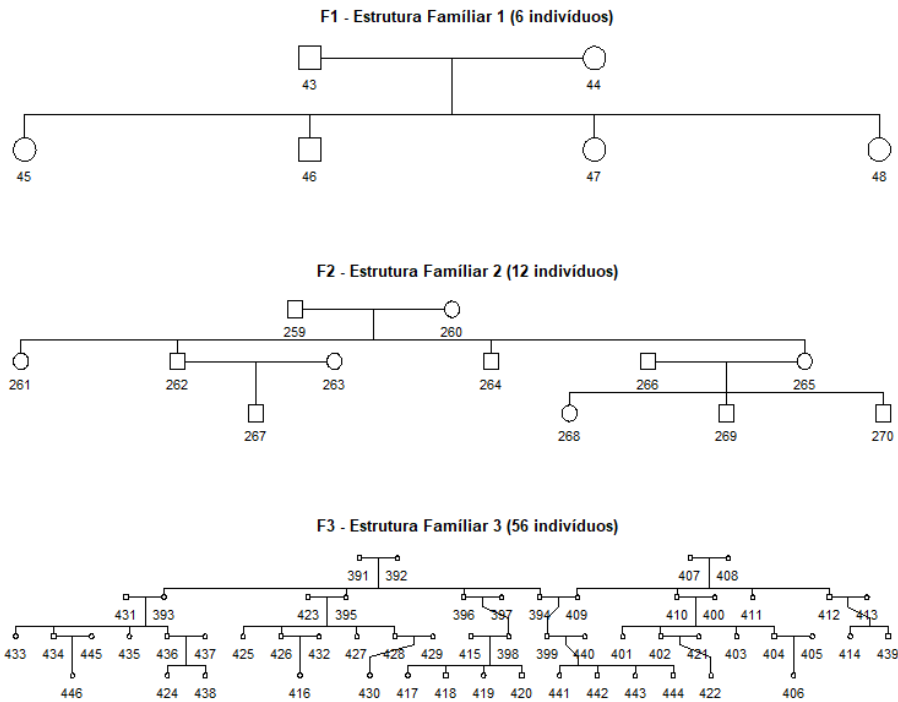


Figura 1 - Estruturas familiares simuladas para o estudo. Quadrados representam indivíduos do sexo masculino e círculos indivíduos do sexo feminino.

considerado sendo composto por uma amostra desbalanceada formada pelas três estruturas básicas da Figura (1), sendo a primeira replicada 15 vezes, a segunda 25 vezes e a terceira 2 vezes. A Tabela (1) descreve as características das amostras avaliadas.

O nível de significância adotado foi de 5% e cada uma das 8 estruturas familiares, combinadas com os parâmetros de herdabilidade considerados foi simulada 1000 vezes. As estimativas empíricas do erro tipo I e do poder do teste foram obtidas pela proporção de rejeições da hipótese nula em cada simulação. Todos os procedimentos envolvidos neste estudo foram implementados usando os recursos do *software* R (R CORE TEAM, 2018) e podem ser solicitados aos autores.

Tabela 1 - Descrição das amostras e estrutura familiar simulada

Cenários	Parentesco	Tamanho das Famílias	Número de Famílias	Amostra
Estrutura 1	F1 Fixado (4 irmãos)	6	84	504
Estrutura 2	F2 Fixado	12	42	504
Estrutura 3	F3 Fixado	56	9	504
Estrutura 4	F4 Fixado (8 irmãos)	10	50	500
Estrutura 5	F1 Variável	6	84	504
Estrutura 6	F2 Variável	12	42	504
Estrutura 7	F3 Variável	56	9	504
Estrutura 8	Misto Fixado	12*	42	502

* Média de indivíduos por família.

3 Resultados e discussões

Os resultados das simulações para os casos com parâmetro de herdabilidade igual a zero e sob a hipótese nula ($h^2 = 0$) estão apresentados na Tabela (2). Independente da estrutura familiar, observou-se que todas as taxas do erro tipo I foram inferiores ao nível de significância nominal considerado (5%), sugerindo um comportamento conservativo da estatística de teste ELRT. Este comportamento, observado também por Blangero *et al.* (2013), tem sido foco de alguns estudos que, considerando aplicações do modelo de componentes da variância na análise de dados de famílias, questionam a adequação da distribuição assintótica da ELRT como uma mistura 50:50 de Qui-quadrados, quando o parâmetro de interesse situa-se na borda do espaço paramétrico (CRAINICEANU *et al.*, 2005; CRAINICEANU e RUPPERT, 2004b; KUO, 1999). Esses autores mostram que para alguns casos a proporção de mistura pode não ser equilibrada como proposto por Self e Liang (1987), mas com desvios para um maior peso para a Qui-quadrado com zero graus de liberdade.

Tabela 2 - Erro tipo I estimado e médias da estatística de teste ELRT e da estimativa de herdabilidade para o fenótipo gerado

Estrutura Familiar	Erro Tipo I	ELRT Média	Estimativa média de h^2
1 Fixa (4 irmãos)	0,0020	0,0257	0,0005
2 Fixa	0,0010	0,0127	0,0002
3 Fixa	0,0000	0,0000	0,0000
4 Fixa (8 irmãos)	0,0020	0,0219	0,0004
5 Variável	0,0030	0,0252	0,0007
6 Variável	0,0020	0,0239	0,0005
7 Variável	0,0000	0,0012	0,0001
8 Mista (Fixa)	0,0010	0,0070	0,0002

Nota-se também, que famílias maiores, como a Estrutura 3, tendem a tornar o teste mais conservativo comparado à famílias menores (Estruturas 1 e 2), independentemente de replicações ou não das relações de parentesco. Os valores médios obtidos para a estatística de teste ELRT foram todos baixos, na ordem de 10^{-2} . Ainda, o modelo utilizado forneceu estimativas de h^2 bem próximas de zero, como já era esperado de ocorrer nesta avaliação.

Quanto ao poder do teste, os resultados para cada uma das herdabilidades paramétricas consideradas, encontram-se na Tabela (3). Apenas não foram descritos os resultados para a herdabilidade paramétrica igual a 0,70 pois, nesta situação, os níveis de poder obtidos convergiram para 1,0 em todos os cenários avaliados. Nesta tabela, observa-se que delineamentos de famílias nucleares constituídos por irmandades extensas, como foi o caso da Estrutura 4, implicam em maior poder para o teste baseado na ELRT, comparado com os demais delineamentos. Este resultado é mais evidente na região de menores valores do espaço paramétrico, cuja as herdabilidades situam-se entre 0,15 e 0,30. Em especial, para as herdabilidades de 0,15 e 0,20, este delineamento familiar apresentou taxas de poder equivalentes ao dobro das taxas obtidas nos demais cenários considerados neste estudo.

A região do espaço paramétrico com herdabilidades entre 0,20 e 0,40 merece bastante atenção devido ao fato de que grande parte dos estudos sobre herdabilidade de fenótipos quantitativos, como pressão sistólica, circunferência abdominal, níveis de colesterol, dentre outros, têm estimativas de herdabilidade de baixas a moderadas como estas. Como exemplo, Oliveira (2008), estudando a herdabilidade de fatores de risco para a Síndrome Metabólica, encontrou estimativas de herdabilidade, não ajustadas por covariáveis, iguais a 0,15 e 0,26 para a pressão arterial sistólica e circunferência abdominal, respectivamente. Quando estas estimativas foram ajustadas por sexo e idade dos indivíduos, os valores passaram para 0,26 e 0,40, respectivamente.

Outro ponto que merece destaque sobre os resultados da Tabela (3) refere-se ao fato de que os delineamentos balanceados e com estrutura fixada de relacionamentos (Estruturas 1, 2 e 3) apresentaram níveis de poder, em geral, maiores do que os correspondentes delineamentos com o mesmo balanceamento, mas cuja a estrutura de parentesco das famílias envolvidas foi variável (Estruturas 5, 6 e 7). Parte desse comportamento pode ser explicada pelo fato de que nas simulações das famílias com estrutura de parentesco variável, o número médio de gerações em cada família foi maior do que nas famílias cuja estrutura de parentesco foi fixada. Consequentemente, o número de relacionamentos diretos (entre pais e filhos e entre irmãos) em cada geração das famílias no delineamento balanceado e com estrutura familiar variável foi, em geral, menor. Além disso, comparando os delineamentos balanceados com famílias de estrutura variável, para herdabilidades moderadas (entre 0,30 e 0,40), o poder do delineamento definido pela Estrutura 7 foi maior que o da Estrutura 6 que, por sua vez, foi maior que o da Estrutura 5.

Tabela 3 - Poder estimado e média das ELRT sob as herdabilidades paramétricas 0,15, 0,20, 0,30, 0,40 e 0,50 nos cenários avaliados

Herdabilidade Paramétrica	Estrutura Familiar	Poder	ELRT Média
0,15	1 Fixa (4 irmãos)	0,164	2,616
	2 Fixa	0,161	2,466
	3 Fixa	0,212	3,508
	4 Fixa (8 irmãos)	0,355	5,788
	5 Variável	0,177	1,945
	6 Variável	0,142	2,061
	7 Variável	0,130	1,953
	8 Mista (Fixa)	0,180	2,856
0,20	1 Fixa (4 irmãos)	0,378	6,682
	2 Fixa	0,378	6,538
	3 Fixa	0,427	8,174
	4 Fixa (8 irmãos)	0,638	12,830
	5 Variável	0,349	4,347
	6 Variável	0,319	4,978
	7 Variável	0,337	5,638
	8 Mista (Fixa)	0,368	6,693
0,30	1 Fixa (4 irmãos)	0,826	19,900
	2 Fixa	0,820	19,510
	3 Fixa	0,855	22,390
	4 Fixa (8 irmãos)	0,934	29,640
	5 Variável	0,771	12,773
	6 Variável	0,784	15,590
	7 Variável	0,798	17,990
	8 Mista (Fixa)	0,843	20,360
0,40	1 Fixa (4 irmãos)	0,990	35,630
	2 Fixa	0,988	35,040
	3 Fixa	0,990	38,850
	4 Fixa (8 irmãos)	0,999	49,040
	5 Variável	0,951	23,370
	6 Variável	0,974	28,740
	7 Variável	0,980	32,870
	8 Mista (Fixa)	0,982	35,820
0,50	1 Fixa (4 irmãos)	1,000	53,800
	2 Fixa	1,000	52,980
	3 Fixa	1,000	57,900
	4 Fixa (8 irmãos)	1,000	71,650
	5 Variável	1,000	36,336
	6 Variável	0,999	44,100
	7 Variável	1,000	50,070
	8 Mista (Fixa)	1,000	53,860

Quanto à Estrutura 8 (desbalanceado), nota-se que o poder do teste foi maior que os observados na estrutura balanceada com relacionamentos variáveis (Estruturas 5, 6 e 7) e também maior que os observados na Estrutura 1 e 2 dos delineamentos balanceados fixados. Das avaliações consideradas, a Estrutura 8, somente não superou o poder obtido nos delineamentos das Estruturas 3 e 4, mas ficando com poder bem próximos a estes para herdabilidades maiores que 0,30. Aqui, cabe ressaltar que na maioria dos estudos reais que envolvem delineamentos com famílias, as estruturas de parentesco e o tamanho de cada família são variáveis (ALVIM, 2016; BOSY-WESTPHAL *et al.*, 2008; LIN *et al.*, 2005). Assim, o comportamento do teste observado nessa estrutura familiar tende a se aproximar melhor do que seria esperado em situações reais.

Deve-se considerar que os autovalores provenientes da decomposição espectral da matriz de parentesco (*kinship*) influenciam diretamente a magnitude da ELRT. Autovalores menores que 1,0, obtidos em maior proporção nos delineamentos com irmandades extensas, por exemplo, aumentam o valor da ELRT e, autovalores maiores que 1,0, diminuem o valor desta estatística (BLANGERO *et al.*, 2013).

A Figura 2(a) mostra as curvas de poder para as estruturas familiares balanceadas e com parentesco fixado. Pode-se observar que o teste tem melhor desempenho para delineamentos com irmandades extensas, como foi o caso da Estrutura 4, composta pelos pais e 8 filhos. Isto implica que para a estatística avaliada, o número de irmãos tem maior influência nos níveis de poder do teste, comparado ao tamanho da família e ao número de gerações, notadamente maiores na Estrutura 3. Vale ressaltar, que a maior contribuição desse tipo de delineamento decorre do maior número de relações de parentesco direto, isto é, entre pais e filhos e entre irmãos, cujo os coeficientes *kinship* são os maiores possíveis. Nesta mesma figura, o gráfico (b) mostra um comparativo entre as três estruturas balanceadas mas contendo famílias com graus de parentesco variável (Estruturas 5, 6 e 7). Neste caso, observa-se um desempenho muito similar entre as três curvas de poder. Já, no gráfico (c) foi feito um comparativo entre as estruturas balanceadas com famílias definidas por graus de parentesco fixado e variável, com até doze indivíduos por família (Estruturas 1, 2, 4, 5 e 6). Novamente, destaca-se a estrutura composta por irmandades mais extensas (Estrutura 4). No gráfico (d) cuja a comparação foi realizada entre a amostra com estrutura familiar desbalanceada (Estrutura 8) com as estruturas familiares balanceadas mais extensas (Estrutura 3 e 7), percebe-se claramente um comportamento do poder do teste da Estrutura 8 intermediário entre os outros dois, com o caso das estruturas fixadas (Estrutura 3) mostrando maior poder que o de estruturas variáveis (Estrutura 7). Ainda neste gráfico, nota-se poder equivalente entre as Estruturas 7 e 8 para herdabilidades maiores que 0,30.

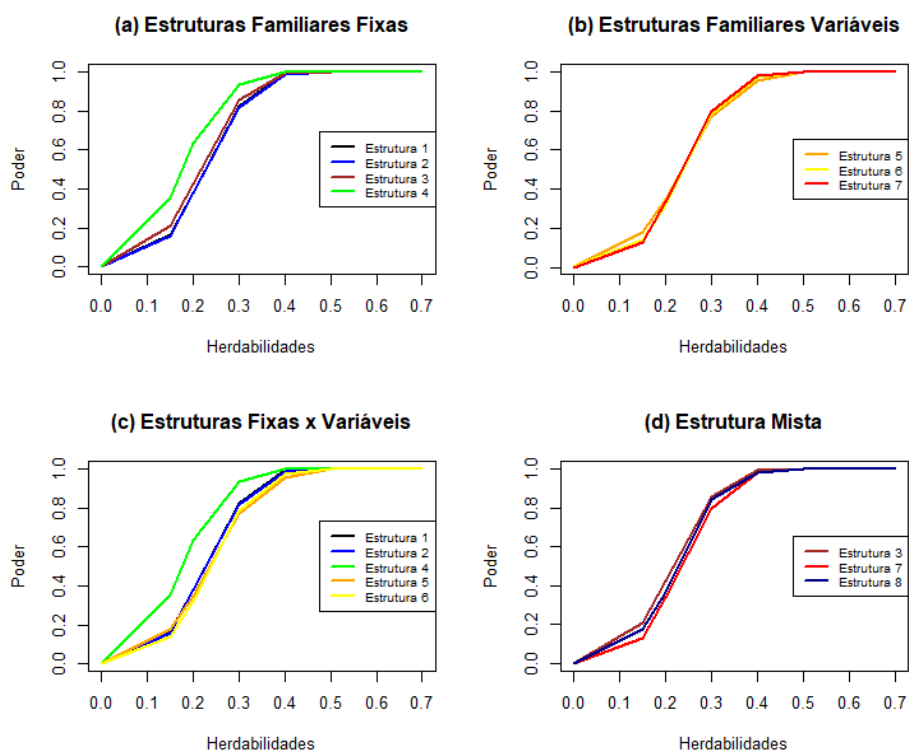


Figura 2 - Curvas do poder do teste para a herdabilidade univariada em função de diferentes estruturas familiares simuladas.

Na Figura (3) observa-se a distribuição dos autovalores, provenientes da decomposição espectral da matriz 2Φ , de cada estrutura familiar considerada no presente estudo. Nela pode-se observar claramente a maior concentração de autovalores menores que 1 na Estrutura 4, o que confere ao teste proposto por Blangero et al. (2013) maiores níveis de poder, como já discutido anteriormente.

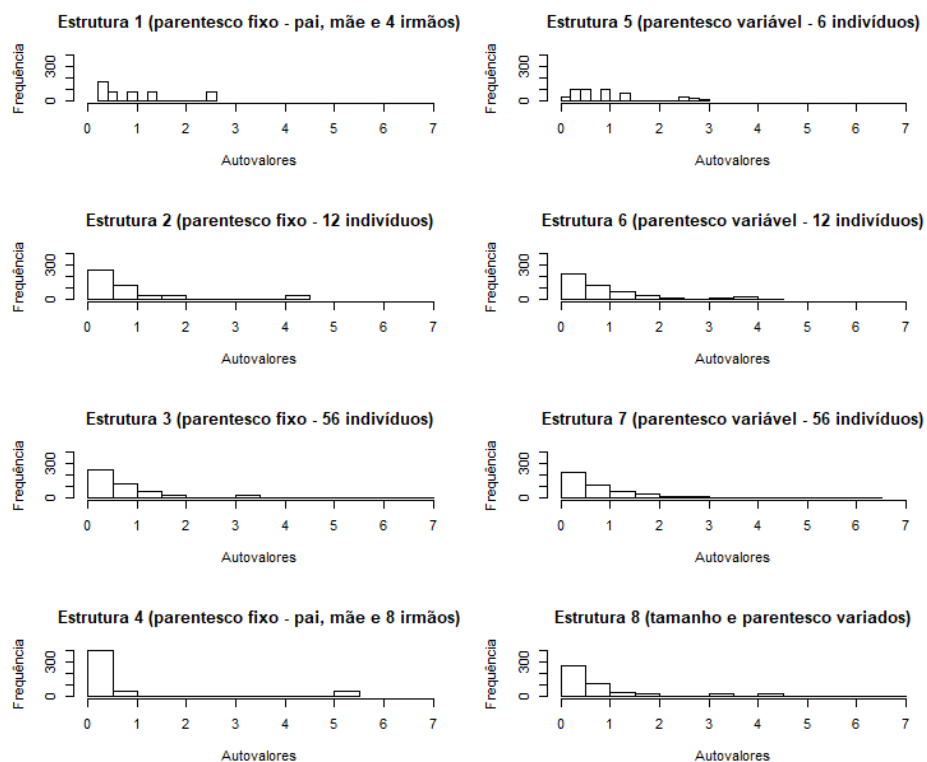


Figura 3 - Distribuição dos autovalores nas diferentes estruturas familiares simuladas.

Conclusões

Diante dos resultados obtidos é possível concluir que para a taxa de erro tipo I, o teste para a herdabilidade univariada proposto por Blangero *et al.* (2013) é conservativo, independentemente do tipo de estrutura familiar utilizada, do tamanho das famílias e da relação de parentesco entre os seus membros. Este resultado tem sido descrito por outros autores (BLANGERO *et al.*, 2012; CRAINICEANU *et al.*, 2005; CRAINICEANU e RUPPERT, 2004b; KUO, 1999) que indicam que as proporções de mistura associadas à distribuição assintótica da ELRT dependem da estrutura de covariância das observações por meio de seus autovalores.

Quanto ao poder do teste de herdabilidade, os resultados das simulações realizadas neste trabalho ilustram bem que delineamentos familiares constituídos por irmandades extensas (famílias com 8 irmãos ou mais) conduzem a um aumento no poder do teste de herdabilidade. Isto se deve aos relacionamentos diretos

presentes nestes delineamentos os quais implicam nos maiores valores do coeficiente *kinship*, além do fato de que na decomposição espectral da matriz de parentesco deste delineamento contribui com a maior proporção de autovalores menores que 1, os quais aumentam o valor da ELRT. Além disso, fixado o tamanho amostral total, que neste estudo foi de 500 indivíduos, comparado ao número de filhos, o tamanho das famílias e o número de gerações têm menor impacto no desempenho do teste.

Considerando a coleta de dados reais em populações naturais, principalmente em humanos, a estrutura desbalanceada (com famílias de diferentes tamanhos e graus de parentesco), similar à Estrutura 8, é a que mais ocorre. Neste caso, os resultados obtidos considerando um tamanho amostral total de 502 indivíduos, indicam que, para herdabilidades acima de 0,30 o poder do teste é alto. Para fenótipos com herdabilidade baixa (em torno de 0,20 ou menos) são requeridos elevados tamanhos amostrais para se obter alto poder, o que depende ainda da estrutura total do pedigree. Apesar do tamanho amostral do estudo ser um parâmetro fundamental em todas as áreas de Estatística, incluindo análises de poder de testes e taxas de erro tipo I, este não foi o objetivo do presente estudo e discussões sobre este tema podem ser encontrados em Raffa e Thompson (2016).

RIBEIRO, A. O.; FERREIRA, D. F.; SOLER, J. M. P. Performance of the test of univariate heritability on designs with different family structure. *Rev. Bras. Biom.*, Lavras, v.37, n.1, p.66-81, 2019.

■ **ABSTRACT:** *Studies on heritability of quantitative traits are essential for understanding of the genetic contribution of factors associated to diseases, human behavior, among others. For this purpose, designs with families are employed, which make it possible to capture information on the response of interest in specific sets of parents or whole families. In this article, the effect of variation in the structure and size of the families on the performance of the test of univariate heritability proposed by Blangero et al. (2013) is evaluated through error rates tipo I and the power of the test. The results shown that the test is conservative, independent of the type of family structure employed, the size of the families and the relationship of kinship between his members. Regarding the power, the results of simulations indicated that the designs with extensive brotherhood led to an increase of the power of the test of heritability. This is due to the greater amount of direct relationships present in these designs, besides the fact that in the spectral decomposition of the matrix of kinship this design contributes with a major proportion of the eigenvalues lower than 1, which increase the statistic of the test likelihood ratio (ELRT).*

■ **KEYWORDS:** *Spectral decomposition; type I error; heritability; kinship matrix; power.*

Referências

- ALMASY, L.; BLANGERO J. Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics*, v.62, p.1198-1211, 1998.
- ALMASY, L.; BLANGERO J. Variance component methods for analysis of complex phenotypes. *Cold Spring Harbor Protocols*, n.5, 2010.
- ALVIM, R. O. *Herdabilidade da velocidade de onda de pulso e associação do controle glicêmico e perfil lipídico com a rigidez arterial em uma população brasileira: "Projeto Corações de Baependi"*. 2016. 79p. Tese (Doutorado) - Faculdade de Medicina: Programa de Cardiologia. Universidade de São Paulo, São Paulo, 2016.
- BOSY-WESTPHAL, A. *et al.* Familial influences and obesity-associated metabolic risk factors contribute to the variation in resting energy expenditure: the Kiel Obesity Prevention Study. *The American Journal of Clinical Nutrition*. v.87, p.1695-1701, 2008.
- BLANGERO, J. *et al.* A kernel of truth: statistical advances in polygenic variance components models for complex human pedigrees. *Advances in Genetics*, v.81, n.1, p.1-31, 2013.
- BRUCE, M. D. H. A. *et al.* Potassium channel gene associations with joint processing speed and white matter impairments in schizophrenia. *Genes, Brain and Behavior*. v.16, p.515-521, 2017.
- CANTOR, R. M. *et al.* Genetic complexity at expression quantitative trait loci. *BMC Proceedings*. v.10(Suppl 7):37, p.85-89, 2016.
- CLARK, M. M. *et al.* Human birth weight and reproductive immunology: testing for interactions between maternal and offspring KIR and HLA-C genes. *Human Heredity*. v.81, p.181-193, 2016.
- CRAINICEANU, C. M. *et al.* Exact likelihood ratio tests for penalized splines. *Biometrika*, v.92, p.91-103, 2005.
- CRAINICEANU, C. M.; RUPPERT, D. Likelihood ratio tests in linear mixed models with one variance component. *Journal of Royal Statistical Society, Series B*, v.66, p.165-185, 2004b.
- HAINES, J. L.; PERICAK-VANCE, M. A. *Approaches to gene Mapping in Complex Human Disease*. New York: Wiley-Liss, 1998. 434p.
- HAZELTON, M. L.; GURRIN, L. C. A note on genetic variance components in mixed models. *Genetic Epidemiology*, v. 24, p.297-301, 2003.
- JUN, G. *et al.* Evaluating the contribution of rare variants to type 2 diabetes and related traits using pedigrees. *PNAS*, v.115, p.379-384, 2018.
- KUO, B. S. Asymptotics of ML estimator for regression models with a stochastic trend component. *Econometric Theory*, v.15, p.24-49, 1999.
- LANGE, K. *Mathematical and statistical methods for genetic analysis*. 2.ed. New York: Springer-Verlag, 2002.

- LIN, H. F. *et al.*. Heritability of the metabolic syndrome and its components in the Northern Manhattan Family Study. *Diabetologia*, v.48, p.2006-2008, 2005.
- OLIVEIRA, C. M. *Herdabilidade dos fatores envolvidos na síndrome metabólica.*, 2008. 100p. Tese (Doutorado) - Faculdade de Medicina: Programa de Cardiologia. Universidade de São Paulo, São Paulo, 2007.
- R CORE TEAM. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>, 2018.
- RAFFA, J. D.; THOMPSON, E. A. Power and effective study size in heritability studies. *Statistics in Biosciences*, v.8, p.264-283, 2016.
- SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. *Variance Components.* New York: John Wiley & Sons, 1992. 501p.
- SELF, S. G.; LIANG, K. Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, v.82, p.605-610, 1987.
- STUART, A.; ORD, J. K. *Kendall's advanced theory of statistics. 2. ed. Distribution theory*, v. 1. New York: Oxford University Press, 1987. 472p.
- THOMPSON, E. A.; SHAW, R. G. Pedigree analysis for quantitative traits: variance components without matrix inversion. *Biometrics*, v.46, p.399-413, 1990.
- THOMPSON, E. A.; SHAW, R. G. Estimating polygenic models for multivariate data on large pedigrees. *Genetics*, v. 131, p.971-978, 1992.
- VISSCHER, P. M.; ANDREW T.; NYHOLT, D. R. Genome-wide association studies of quantitative traits with related individuals: Little (power) lost but much to be gained. *European Journal of Human Genetics*, v.16, p.387-390, 2008.

Recebido em 08.12.2017.

Aprovado após revisão em 19.12.2018.