

Análise de dados gemelares: uma aventura guiada para investigadores das Ciências do Desporto

CDD. 20.ed. 575.1
613.7

José António Ribeiro MAIA*
André SEABRA*
Rui Manoel Garganta da SILVA*
Duarte Luís de FREITAS**
Élvio Quintal Rúbio GOUVEIA*
Luciano BASSO***

*Faculdade de Desporto, Universidade do Porto - Portugal.
**Universidade da Madeira - Portugal.
***Escola de Educação Física e Esporte, Universidade de São Paulo.

Resumo

Este tutorial pretende apresentar, de forma sumária: 1) a importância dos estudos gemelares em Educação Física e Ciências do Desporto; 2) a estrutura básica deste tipo de delineamento; 3) as etapas da análise descritiva inicial; 4) os procedimentos de análise em grau crescente de complexidade - da análise de variância, à modelação de estruturas de covariância; 5) bem como a relevância da exploração da disocordância intra-par de gêmeos monozigóticos. Esta processologia, baseada num tutorial, recorrerá aos valores do fenótipo índice de atividade física nos tempos de lazer, com base numa amostra de 207 pares de gêmeos mono e dizigóticos. Todas as etapas da análise são comentadas e será interpretado o significado dos resultados, salientando o fato do fenótipo em causa ser explicado pelos fatores genéticos em cerca de 63%.

UNITERMOS: Gêmeos; Análise de variância; Modelos; Atividade física.

Introdução

Sir Francis Galton (1822-1911), brilhante cientista inglês, um dos progenitores da Biometria, foi o fundador da Genética do Comportamento (do inglês "Behavioral Genetics"), ao propôr o delineamento gemelar, para tentar interpretar os efeitos da famosa expressão cunhada pela sua mão - "nature versus nurture" (que poderíamos traduzir muito livremente por biologia versus cultura). Teríamos que esperar até 1971 para que este delineamento de pesquisa fosse adaptado pelo vasto território da pesquisa em Ciências do Desporto. De fato, nesse ano, Vassilis Klissouras publica um texto seminal no "Journal of Applied Physiology" sobre a importância dos fatores genéticos nos valores do consumo máximo de O_2 de gêmeos. Contudo, o maior destaque para o uso de gêmeos em investigação na Fisiologia do Exercício, Bioquímica do Exercício, Obesidade, Resposta ao Exercício, Controle Nutricional, entre outras temáticas, é devido ao trabalho magistral da equipe do Prof. Claude Bouchard [um vasto resumo da sua obra pode ser consultado nos diferentes capítulos do livro

editado por Bouchard, Malina e Perusse publicado em 1997 - "Genetics of Fitness and Physical Activity" (BOUCHARD, MALINA & PÉRUSSE, 1997)].

A investigação em Genética do Desporto é realizada com recurso ao modelo animal (normalmente ratos) e a humanos. Nas amostras de humanos, a primeira tarefa consiste em determinar a presença de agregação ou semelhança familiar num dado fenótipo (por exemplo: peso, consumo máximo de O_2 , percentagem de fibras musculares do tipo I, quantidade de massa isenta de gordura, etc). Normalmente a estatística mais adequada para estimar o grau de semelhança fenotípica é o coeficiente de correlação (uma estatística descritiva para determinar a associação entre membros de famílias num fenótipo qualquer). Os cálculos são feitos entre diferentes tipos de familiares (por exemplo, em famílias nucleares: pai-filho, mãe-filha, irmã-irmão, tio-sobrinha, avô-neta, etc; ou em gêmeos de diferente zigotia: pares de gêmeos monozigóticos do sexo masculino; pares de gêmeos dizigóticos do sexo masculino, etc). Se as correlações forem significativamente maiores do que

zero, numa segunda etapa calcula-se uma estatística designada por heritabilidade (h^2), que expressa o quanto da variação na característica em causa é devida a fatores genéticos. Esta estimativa é sempre corrigida para a presença de um conjunto diversificado de covariáveis¹, de forma a ter uma fenótipo cuja variância não seja nem amplificada nem reduzida na sua magnitude. Depois de verificado que a variação da característica em causa é governada, também, por fatores de natureza genética recorre-se, numa terceira etapa, a uma mistura de técnicas laboratoriais e estatísticas (estudos de “linkage” ou ligação fatorial) para tentar localizar regiões em diferentes cromossomas que se pensa conter genes (que se designam por candidatos) responsáveis pela semelhança verificada entre familiares. Um dos últimos passos nesta “caça de genes”, é o recurso a estudos de associação (normalmente de casos, os portadores de uma dada característica, em contraste com os controles, sujeitos que não possuem a característica em causa). Testa-se o valor de um gene candidato (mais concretamente dos seus alelos e genótipos possíveis) que se pensa condicionar os resultados obtidos no fenótipo em estudo (sobre esta sequência de etapas consultar, em língua Portuguesa - MAIA, 2007).

Gêmeos de diferente zigotia são universalmente utilizados em pesquisa no domínio da fisiologia (BOUCHARD, CHAGNON, THIBAUT, BOULAY, MARCOTTE, COTE & SIMONEAU, 1989; SIMONEAU, LORTIE, BOULAY, MARCOTTE, THIBAUT & BOUCHARD, 1986), crescimento somático (BEUNEN, THOMIS, MAES, LOOS, MALINA, CLAESSENS & VLIETINCK, 2000; HAUSPIE, BERGMAN, DIELICKI & SUSANNE, 1994), comportamentos de saúde (KAPRIO, PULKKINEN & ROSE, 2002), síndrome metabólica (EDWARDS, NEWMAN, MAYER, SELBY, KRAUSS & AUSTIN, 1997), somatótipo (PEETERS, THOMIS, CLAESSENS, LOSS, MAES, LYSSENS, VANDEN EYNDE, VLIETINCK & BEUNEN, 2003), atividade física (MAIA, THOMIS & BEUNEN, 2002; SAPAGE, MAIA, SEABRA, SILVA & LOPES, 2007) e aptidão física (MAIA, LOPES & MORAIS, 2001; WATANABE, MUTOH & YAMAMOTO, 2001), entre outras matérias. Recentemente tem sido atribuído a este tipo de delineamento um papel importante na identificação de genes responsáveis pela enorme variação populacional nas características complexas humanas (BOOMSMA, BUSJAHN & PELTONEN, 2002), ao ponto da Comunidade Económica Europeia ter financiado um vasto projeto de investigação designado “GenomEUtwin: a strategy to identify genetic influences on health and disease” (PELTONEN, 2003).

Estão disponíveis na literatura textos bem interessantes, de complexidade crescente, sobre o uso de informação gemelar, bem como a sua análise e

interpretação (BOUCHARD, PERUSSE & LEBLANC, 1990; NEALE & CARDON, 1992; NEALE & EAVES, 1995). A estes documentos juntam-se alguns livros fundamentais no vasto território da Genética do Comportamento de que destacamos os de Plomin e colaboradores (PLOMIN, DEFRIES, MCCLEARN & GUFFIN, 2008; PLOMIN, DEFRIES, CRAIG & MCGUFFIN, 2002).

Face ao interesse sempre renovado no recurso a gêmeos para pesquisar múltiplas facetas do comportamento e características de humanos, sobretudo no vasto espaço da Educação Física e das Ciências do Desporto, e porque não existe um texto, em Português, relativamente simples que explique algumas das etapas da análise da informação gemelar, o presente documento procura colmatar essa lacuna. Na sua generalidade, este trabalho pretende providenciar um conjunto de etapas de análise de dados quando se lida com informação gemelar. A sua estrutura, distinta da habitual, está dividida, sequencialmente, em dois domínios a partir de um conjunto de dados sobre um fenótipo relativamente complexo de medir - a atividade física. No primeiro trataremos de apresentar diferentes aspectos da análise de dados gemelares a partir de simples representações gráficas, processos de refinamento da variável dependente quando a amostra é relativamente reduzida, bem como cálculos distintos do coeficiente de correlação intraclasse e estimativas de heritabilidade a partir da Análise de Variância. Comentaremos e explicaremos os resultados e o seu significado. No segundo momento, a estrutura de análise será um pouco mais complexa a partir de Modelação de Estruturas de Covariância, em que são contrastados diferentes modelos para explicar o mais adequadamente possível as matrizes de covariância entre gêmeos das duas zigotias - monozigóticos (MZ) e dizigóticos (DZ). A partir deste ponto será adicionada uma nova frente de pesquisa promissora que lida com o problema, sempre desafiadora, da interpretação do significado da diferença intra-par nos gêmeos MZ.

Gostaríamos de salientar que este artigo localiza a sua atenção nas duas primeiras etapas da análise da Epidemiologia Genética aplicada às Ciências do Desporto e anteriormente referidas [sobre esta matéria ver, por exemplo, o capítulo 4 do livro de Bouchard (BOUCHARD, MALINA & PÉRUSSE, 1997)]. Como utilizaremos diferentes “softwares” estatísticos, de cada um deles, colaremos os “outputs” ao corpo do texto para melhor ilustrar o alcance dos resultados.

Antes de começarmos a nossa viagem, convém apresentar os pressupostos fundamentais da análise Genética Quantitativa quando se lida com gêmeos:

- Os pares de gêmeos são da mesma idade e partilham o mesmo ambiente familiar de origem (i.e. possuem o mesmo ambiente comum).

- Os gêmeos MZ partilham os mesmos fatores genéticos (os mesmos alelos em cada “locus”). Os gêmeos DZ partilham, em média, somente metade dos genes, tal como o que ocorre entre irmãos.

- Um ambiente semelhante afeta de igual modo os membros de cada par de gêmeos MZ e gêmeos DZ.

- A comparação da similaridade entre gêmeos MZ e gêmeos DZ relativamente a um fenótipo permite identificar as fontes de variação na população - genes, ambiente comumente partilhado e envolvimento único.

No modelo clássico, qualquer fenótipo é constituído por duas componentes aditivas - uma genética (G) e outra do ambiente (E). Quando se pensa em termos populacionais, no modelo clássico da Genética Quantitativa, a variância total (V_{TOT}) de qualquer traço fenotípico pode ser fraccionada em dois tipos de variância - a variância genética (V_G) e a variância do ambiente (V_E). Esta última pode ser ainda fraccionada em duas

outras fontes de variação - a que é devida ao ambiente comumente partilhado no seio da família (V_C) e a que é específica e única de cada par (V_{ESP}). Do contraste entre V_G e V_T é possível calcular um rácio designado de heritabilidade (h^2) que se refere à proporção de variância total que pode ser atribuída a efeitos genéticos ($h^2 = V_G / V_{TOT}$). De um modo semelhante podem calcular-se a contribuição dos fatores do ambiente comum ($c^2 = V_C / V_{TOT}$) e dos fatores específicos ($e^2 = V_{ESP} / V_{TOT}$). Nestas formulações do modelo clássico é imperioso que se cumpram três pressupostos considerados essenciais (BOUCHARD, MALINA & PÉRUSSE, 1997):

- Ausência de interação dos genótipos com o envolvimento (isto significa que diferentes genótipos reagem de modo similar aos mesmos fatores do envolvimento, traduzindo uma igualdade de suscetibilidade).

- Ausência de correlação entre genótipos e envolvimento (isto quer dizer que os vários genótipos são expostos a condições semelhantes do envolvimento).

- Ausência de interação gene-gene (i.e. não se verifica epistasia).

A amostra, determinação da zigotia e processo de medição da variável

Amostra

Os dados deste estudo provêm de uma amostra de gêmeos portugueses da Região Autónoma da Madeira (projeto de investigação GEAFAS) e do Continente (projeto de investigação sobre aspectos genéticos da atividade física, aptidão física associada à saúde, sobrepeso e obesidade - um

estudo em gêmeos dos seis aos 20 anos de idade - GEM). A distribuição da amostra está na TABELA 1.

A dimensão amostral total é composta por 414 sujeitos, o que prefaz 207 pares de gêmeos, cujas médias de idade são as seguintes: $MZ_{\delta} = 10,81 \pm 3,57$, $MZ_{\varphi} = 11,24 \pm 3,83$, $DZ_{\delta} = 10,24 \pm 2,98$, $DZ_{\varphi} = 10,43 \pm 3,40$, $DZ_{\delta\varphi} = 9,38 \pm 2,93$.

TABELA 1 - Linha de comandos, “output” do “software” STATA 10 e distribuição dos gêmeos (n=414 sujeitos que correspondem a 207 pares) pelas diferentes zigotias.

tabule zg zigotia, row column						
zg	zigotia					Total
	1	2	3	4	5	
1	86	82	0	0	0	168
	51,19	48,81	0,00	0,00	0,00	100,00
	100,00	100,00	0,00	0,00	0,00	40,58
2	0	0	60	72	114	246
	0,00	0,00	24,39	29,27	46,34	100,00
	0,00	0,00	100,00	100,00	100,00	59,42
Total	86	82	60	72	114	414
	20,77	19,81	14,49	17,39	27,54	100,00
	100,00	100,00	100,00	100,00	100,00	100,00

zg = gêmeos MZ(1) e DZ(2);
zigotia = MZ δ (1), MZ φ (2), DZ δ (3), DZ φ (4), DZ $\delta\varphi$ (5).

Determinação da zigotia

A determinação da zigotia (i.e., a identificação precisa de gêmeos MZ e DZ, à exceção dos gêmeos DZ de sexo oposto) pode ser efetuada através de métodos diretos (exame das membranas materno-fetais, marcadores sanguíneos e análise do DNA) e indiretos (apreciação somatoscópica da morfologia externa, dermatoglífia, questionários). Uma descrição detalhada destes métodos pode ser encontrado em qualquer manual de Genética Médica, ou em SAPAGE et al. (2007). No presente trabalho a informação detalhada sobre a zigotia foi obtida pela informação do seu DNA. A análise do DNA dos membros de cada par implicou a obtenção de uma amostra muito reduzida de sangue. Trata-se, pois, de um procedimento “algo invasivo”, para o qual foi obtido consentimento por parte dos pais, realizado por um elemento experiente. Também poderia ser feito através de saliva ou esfregaço bucal, mas o sangue foi o material preferido. A técnica de extração e amplificação do DNA por cadeia de reação da polimerase é bem conhecida e está suficientemente explicada em qualquer manual de genética médica. A extração do DNA a partir das amostras de sangue dos 414 indivíduos foi efetuada com um método baseado na utilização da resina CHELEX (LAREU, PHILLIPS, CARRACEDO, LINCOLN, SYNDERCOMBE COURT & THOMSON, 1994). Em todas as amostras de DNA, a análise de 17 STRs autossômicos (CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, PD, PE, TH01, TPO e VWA) e o “locus” da Amelogenina (determinação do sexo), foi efetuada por amplificação por PCR, utilizando os kits comerciais Powerplex 16 System (Promega Corporation) e Identifiler (AB Applied Biosystems), de acordo com as instruções dos fabricantes. A genotipagem foi efetuada em aparelhos ABI 310 Genetic Analyzer (AB Applied Biosystems), de acordo com as instruções do fabricante, por determinação do tamanho dos fragmentos de DNA e comparação com escalas alélicas fornecidos

com os kits comerciais. Nos casos onde os indivíduos apresentavam perfis genéticos de STRs idênticos, foi efetuado o cálculo de probabilidade de monozigotia, que segue, no fundamental, a metodologia de (Essen-Møller, 1939) e que se baseia nas frequências gênicas, neste caso determinadas na população residente no Centro e Norte de Portugal, bem como na Região Autónoma da Madeira.

Medição da variável

O questionário de BAECKE, BUREMA e FRIJTERS (1982) é composto por 16 itens de resposta tipo Lickert numa escala de 1 a 5 de ordem crescente de importância. Os itens procuram estimar a atividade física total (AFT) através de três índices: atividade física no tempo escolar/trabalho (IAFET); atividade no tempo dedicado ao desporto (IAD) e atividade física no tempo de recreação e lazer (IAFTL). O fenótipo que nos interessa estudar neste exemplo é o que se refere ao IAFTL, cujo processo de cálculo é o seguinte:

Índice de atividade física no tempo de lazer (IAFTL)

O índice de atividade física no tempo de lazer é determinado através das questões 13, 14, 15 e 16. Assim o cálculo deste índice resulta da seguinte fórmula:

$$- \text{IAFTL} = [(6 - I_{13}) + I_{14} + I_{15} + I_{16}] / 4]$$

em que I_i , $i = 13, 14, 15, 16$ são os itens do questionário.

Detalhes bem precisos deste questionário podem ser consultados em TEIXEIRA e SEABRA, MAIA, MENDONÇA, THOMIS, CASPERSEN e FULTON (2008). Este questionário tem sido utilizado com frequência em Portugal, providenciando informação relevante a que se adiciona a sua forte validade concorrente relativamente à água duplamente marcada e elevada fiabilidade dos respondentes (SAPAGE et al., 2007).

Estratégia de análise sequencial e interpretação dos resultados

Estudo exploratório da variável

A primeira etapa consistiu na análise exploratória dos resultados do IAFTL (nos gráficos esta variável é designada por “itlazer”) para identificar “outliers”, bem como verificar a normalidade da sua distribuição. A base original dos dados em Excel foi transferida

para o “software” STATA 10 para começar a análise. A primeira etapa consistiu, pois, em representar graficamente a variável (histograma) e verificar se haveria alguma transformação que fosse interessante de considerar. O STATA 10 produz graficamente o solicitado (ver FIGURA 1), ao mesmo tempo que apresenta informação numérica equivalente

(TABELA 2) com testes formais de aderência à normalidade, bem como o respectivo valor de prova.

A variável tem uma distribuição normal (ver FIGURA 2), ainda que fosse possível uma

transformação com recurso à raiz quadrada dos valores, conforme resultados da TABELA 2. Contudo, não há qualquer necessidade desta transformação.

. gladder itlazer

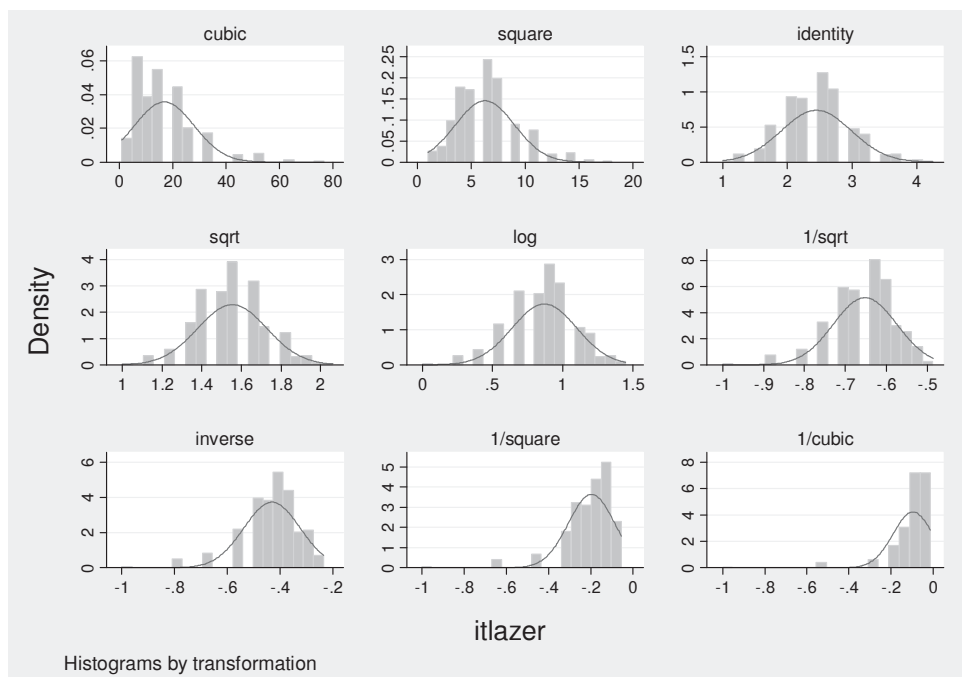


FIGURA 1 - Distribuição da variável (“itlazer” = IAFTL) e possíveis transformações para a normalizar (linha de comandos e “output” do STATA 10).

TABELA 2 - Resultados informativos sobre as diferentes possibilidades de transformação da variável para a normalizar (linha de comandos e “output” do STATA 10).

Transformation	formula	chi2(2)	P(chi2)
cubic	itlazer^3	-	0,000
square	itlazer^2	45,92	0,000
identify	itlazer	4,33	0,115
square root	sqrt(itlazer)	1,61	0,446
log	log(itlazer)	17,16	0,000
1/(square root)	1/sqrt(itlazer)	47,44	0,000
inverse	1/itlazer	-	0,000
1/square	1/(itlazer^2)	-	0,000
1/cubic	1/(itlazer^3)	-	0,000

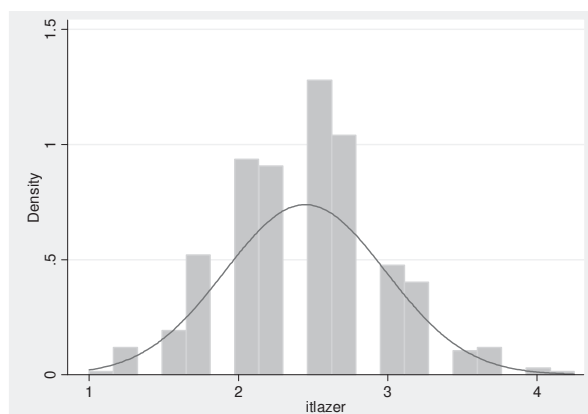


FIGURA 2 - Distribuição dos valores do “itlazer” e sua aproximação à normal (“output” do STATA 10).

Refinamento do fenótipo

Uma vez que temos gêmeos de diferentes idades, sexos e zigotia ($MZ\delta$, $MZ\eta$, $DZ\delta$, $DZ\eta$, $DZ\delta\eta$), e porque a dimensão amostral não é muito elevada (somente 207 pares de gêmeos), foi decidido trabalhar com resíduos “studentized” (em português são referidos como estudentizados) da regressão depois de removidos os efeitos das seguintes covariáveis: idade, idade², sexo, interação idade*sexo, e idade²*sexo. O fenótipo assim obtido é o reflexo mais preciso do “itlazer” depois de controlado para efeitos distintos da idade e sexo. A opção pelos resíduos “studentized” deve-se ao fato de evitar a inflação da variância sempre que um resíduo for elevado (DUPONT, 2002; STEVENS, 2002), bem como serve para verificar a assunção da correção do modelo em termos de distribuição de resíduos (i.e., violação da linearidade e homocedasticidade).

A TABELA 3 mostra os resultados dos preditores mais relevantes do modelo de regressão “stepwise”,

sendo que esta opção foi escolhida pelo fato de não termos um modelo preciso acerca da melhor combinação aditiva de preditores para calcular os resíduos. Os modelos de regressão múltipla com opção “stepwise” têm sido utilizados na literatura da especialidade quando se lida com informação gemelar ou familiar (BOUCHARD, MALINA & PÉRUSSE, 1997). É importante salientar a reduzida variância explicada pelos preditores considerados no modelo. Tal como referimos anteriormente, foram consideradas a idade, sexo, idade*sexo e idade²*sexo. Destas só entraram no modelo final o sexo, idade*sexo e idade²*sexo. A variância que estas covariáveis removem da variância total do fenótipo “itlazer” é somente de 6,3% (ajustamento ao número de variáveis e dimensão amostral).

A FIGURA 3 mostra a distribuição deste fenótipo mais refinado (neste caso os resíduos estudentizados) salientando a sua homogeneidade e normalidade (ver FIGURA 4).

TABELA 3 - Resultados da regressão “stepwise” para “afinar” o fenótipo “itlazer” (linha de comandos e “output” do STATA 10).

```
. stepwise. pe (0,05): regress itlazer sexo idecm idascm idasex id2sex
begin with empty model
```

p = 0,0001 < 0,0500 adding sexo

p = 0,0463 < 0,0500 adding id2sex

p = 0,0015 < 0,0500 adding idasex

Source	SS	df	MS
Model	8,42888381	3	2,80962794
Residual	112,027034	410	0,27323666
Total	120,455918	413	0,29166081

Number of obs = 414
 F (3, 410) = 10, 28
 Prob > F = 0,0000
 R-squared = 0,0700
 Adj R-squared = 0,0632
 Root MSE = 0,5227

A primeira parte do "output" refere-se às variáveis a incluir no modelo; a segunda trata dos valores da ANOVA às fontes de variação - explicada pelo modelo e residual; a terceira parte contém os parâmetros da regressão e respectivas estatísticas mais importantes.

itlazer	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sexo	0,214221	0,159276	1,34	0,179	-0,098878 0,527322
id2sex	0,004392	0,001258	3,49	0,001	0,001918 0,006865
idasex	-0,092252	0,028780	-3,21	0,001	-0,148828 -0,035677
_cons	2,775812	0,081693	33,98	0,000	2,615221 2,936403

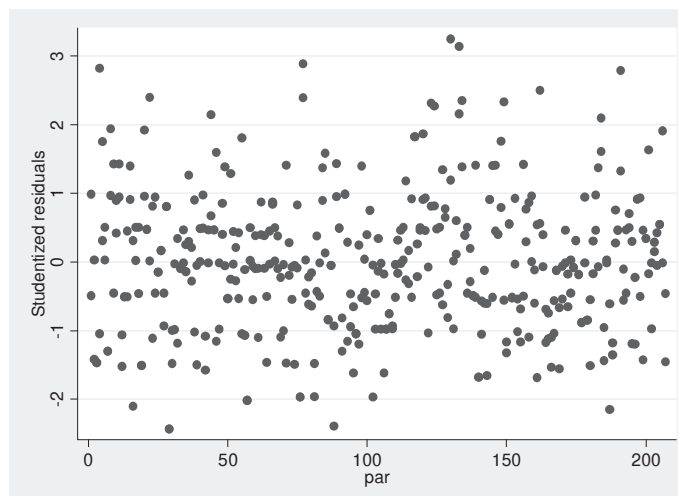


FIGURA 3 - Distribuição dos resíduos estudantizados (“output” do STATA 10).

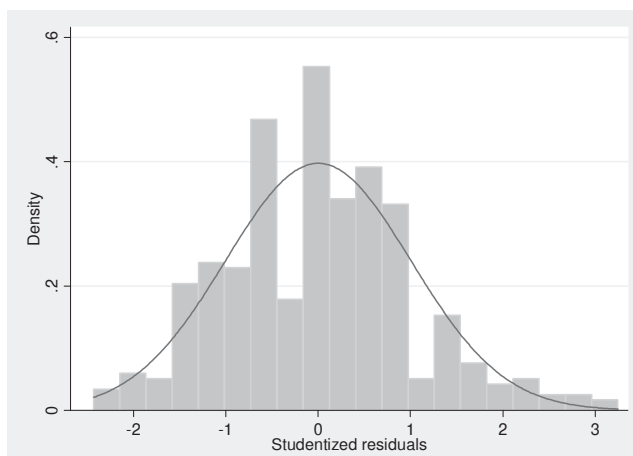


FIGURA 4 - Distribuição dos resíduos e aproximação à normal (“output” do STATA 10).

Representação gráfica das semelhanças intra-par e respectiva análise de homogeneidade intra zigotia

O passo seguinte foi representar, graficamente, a similaridade intra-par dos gêmeos MZ e DZ (ver FIGURA 5). É evidente a menor amplitude de diferenças entre gêmeos MZ do que DZ, o que sugere a presença de efeitos genéticos a governar a variação encontrada neste fenótipo.

De seguida foi necessário calcular a correlação intraclasse separadamente para gêmeos MZ e DZ. Esta estatística providencia informação acerca da homogeneidade de pares de valores. Ainda que seja possível utilizar a correlação interclasse ou de Pearson, convém lembrar que não estamos a lidar com duas variáveis (x e y), mas somente de uma (x) medida em dois sujeitos aparentados, e daqui o uso recorrente da correlação intraclasse (r_{xx}) como a mais adequada. Importa salientar, também, que é

com base na correlação intraclasse que se contrastam variâncias intra-par e entre pares, o que não ocorre na correlação interclasse.

É de esperar, de acordo com o modelo genérico sugerido pela formulação da Genética Quantitativa (FALCONER, 1989), que a correlação intraclasse seja maior nos gêmeos MZ. Se assim for, temos fortes indícios que fatores genéticos aditivos governam a variação presente no fenótipo IAFL. Este passo foi realizado no STATA 10 no módulo “loneway” que permite calcular aspectos relevantes da variação intra-par e entre pares a partir do modelo da análise de variância de efeitos aleatórios. Os resultados mais importantes estão na TABELA 4. Nos gêmeos MZ o valor é de $0,65 \pm 0,06$ (IC95% = 0,53, 0,78), enquanto que nos gêmeos DZ o valor é de $0,35 \pm 0,07$ (IC95% = 0,19, 0,51). É relevante salientar o menor valor da variância intra-par nos gêmeos MZ (0,34), do que nos gêmeos DZ (0,67), praticamente ½ da variância.

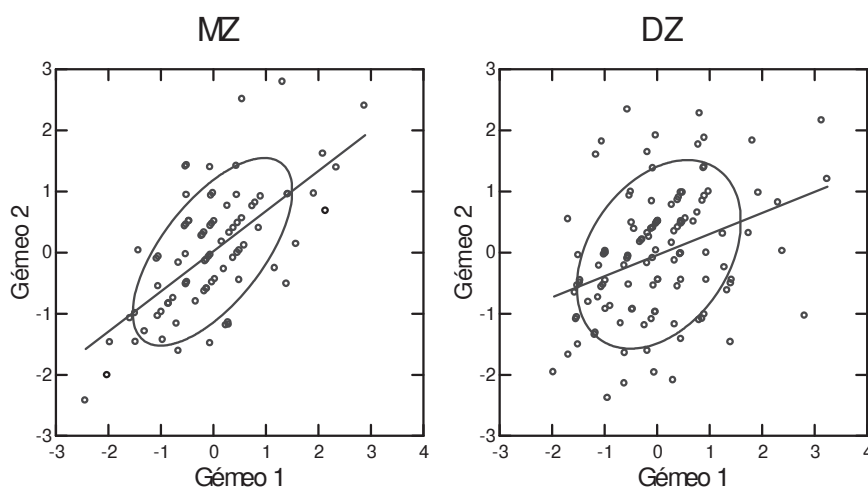


FIGURA 5 - Diagramas de dispersão intra-par nos gêmeos MZ e DZ (“output” do Systat 12).

TABELA 4 - Resultados da opção “loneway” do STATA 10 para calcular a correlação intraclasse, bem como os respectivos intervalos de confiança a 95% (linha de comandos e “output” do STATA).

```
. loneway ilazerst par if zg=1
```

One-way Analysis of variance for ilazerst: Studentized residuals

Number of obs = 168

R-squared = 0,8252

Source	SS	df	MS	F	Prob > F
Between par	135,5308	83	1,6329012	4,78	0,0000
Within par	28,716362	84	0,3418614		
Total	164,24716	167			

```
. loneway ilazerst par if zg=2
```

One-way Analysis of variance for ilazerst: Studentized residuals

Number of obs = 246

R-squared = 0,6733

Source	SS	df	MS	F	Prob > F
Between par	169,65752	122	1,3906354	2,08	0,0000
Within par	82,316595	123	0,6692406		
Total	251,97412	245	1,0284658		

Interclass correlation

Asy.
S.E.

[95% conf. Interval]

0,35021

0,07927

0,19485

0,50558

O STATA 10 tem implementado uma forma bem expedita de testar diferenças entre estas correlações através de um procedimento de re-amostragem (método Bootstrap). A partir de 50 replicações é estimado o erro-padrão das diferenças, que permite calcular um valor

z que é formalmente testado (a hipótese nula é que a diferença entre correlações intraclasse seja zero). Os principais resultados estão na TABELA 5. A diferença observada é significativamente maior que zero (CCIMZ-CCIDZ = 0,30 ± 0,09, p = 0,001; IC95% = 0,12, 0,48).

TABELA 5 - “Output” do STATA 10 relativo ao uso do processo de re-amostragem Bootstrap para testar as diferenças entre correlações intraclasse.

```
Bootstrap replications (50)
```

```
----- 1 ----- 2 ----- 3 ----- 4 ----- 5
```

```
..... 50
```

```
Bootstrap results
```

Number of obs = 414

Replications = 50

```
command: iccdiff
```

```
diff: r(iccdiff)
```

	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]	
diff	0,303556	0,091671	3,31	0,001	0,123884	0,483229

Estimação da heritabilidade a partir da Análise de Variância

Face aos resultados anteriores (gráficos e numéricos) sugerindo a forte presença de efeitos genéticos aditivos a explicarem a variação populacional no IAFTL, a etapa seguinte é estimar a magnitude desse valor. A estatística mais utilizada é a heritabilidade (h^2) que, tal como referimos anteriormente, se refere à fração da variância total que é variância genética. Esta parte da análise foi realizada no “software” TWINAN90 (WILLIAMS, CHRISTIAN & NORTON JUNIOR, 1992) especialmente concebido para lidar com informação gemelar. Os principais resultados estão na TABELA 6, que passaremos a apresentar com base em informação providenciada por WILLIAMS, CHRISTIAN e NORTON JUNIOR (1992) e CHRISTIAN e WILLIAMS (2000):

- A primeira parte do “output” refere-se a informação descritiva do fenótipo em causa: estatísticas descritivas básicas e teste estatístico formal à diferença de médias entre gêmeos MZ e DZ que não deve ser estatisticamente significativa (em cumprimento com suposições do modelo biométrico geral para dados gêmeos).

- A segunda trata de testar, também, um dos pressupostos do modelo e que diz respeito à ausência de

diferenças entre as variâncias totais do fenótipo nos gêmeos MZ e DZ, que também é respeitado nesta análise.

- Em terceira parte trata de formas distintas de calcular variância genética. Desta parte do “output” há que reter duas peças importantes de informação: os valores dos coeficientes de correlação intraclasse dos gêmeos MZ e DZ, bem como do teste formal à diferença entre variâncias genéticas de gêmeos MZ e DZ. No primeiro caso, os valores da correlação intraclasse vão no sentido certo, i.e., a correlação intraclasse dos gêmeos MZ é maior do que a dos DZ. No segundo caso, temos que a diferença em termos absolutos das variâncias genéticas dos gêmeos MZ e DZ é estatisticamente significativa. Tal fato traduz a maior variância genética dos gêmeos MZ.

- Finalmente, na quarta parte, temos diferentes estimativas de heritabilidade. A mais importante é a que se constrói com base na seguinte fórmula: $h^2 = 2(r_{MZ} - r_{DZ})$, tal como salientado por FALCONER (1989). Claro que r_{MZ} é a correlação intraclasse entre gêmeos MZ e r_{DZ} entre gêmeos DZ. Sessenta e cinco por cento da variação total dos valores do IAFTL são devidos a diferenças genéticas entre os sujeitos; dito de uma forma mais sumária: 65% da variação total do IAFTL é devida a fatores genéticos.

TABELA 6 - Principais resultados da análise dos efeitos genéticos no fenótipo IAFL (“output” do TWINAN90).

TWINAN90 TWIN ANALYSIS PROGRAM, VERSION 1.0				
FOR THE VARIABLE: iaftl				
NUMBER OF TWIN PAIRS: MZ = 84 DZ = 123				
ASSESSMENT OF MODEL ASSUMPTIONS, DESCRIPTIVE STATISTICS				
	Individuals		Twin absolute differences	
	MZ	DZ	MZ	DZ
Mean	- 0,954E-03	0,5262E-01	0,6133	0,7692
Variance	0,9808	0,8382	0,3056	0,5347
Skew; SE	0,32; 0,19	0,65; 0,16		
Ex Kurt; SE	0,42; 0,38	0,68; 0,31		
Test for difference in MZ and DZ means :				
MZ - DZ =	- 0,5357E-01	T" = - 0,45	DF = 155,6	P = 0,654
ANALYSIS OF VARIANCE				
	MZ	DZ	MZ	DZ
Among mean squares :	1,628	83	1,115	122
Within mean squares :	0,3409	84	0,5631	123
Sum of mean squares :	1,969	116,3	1,679	220,5

Continua

TABELA 6 - Principais resultados da análise dos efeitos genéticos no fenótipo IAFL ("output" do TWINAN90 - continuação).

----- TEST OF TWIN MODEL: -----					
Sum of mean squares ratio : F" = 0,85 Approx P = 0,313					
THE EQUAL VARIANCE HYPOTHESIS IS NOT REJECTED.					
ESTIMATES OF GENETIC VARIANCE					
		Estimate	Variance	F-ratio	Prob
Within pair	(WP):	0,444533	0,311048E-01	1,65	0,007
Among component	(AC):	0,735189	0,360936	1,50	0,005
Intraclass correlations: MZ = 0,654; P = 0,000 DZ = 0,329; P = 0,000					
Average abs difference test : T" = 1,97 DF = 203,5 P = 0,25					
HERITABILITY ESTIMATES					
		Formula	Estimate	Probability	
Within pair	:	2*(WDZ-WMZ) / ((SMZ + SDZ) / 4)	0,487	SEE PROB(WP)	
Among component	:	(2*AC) / ((SMZ + SDZ) / 4)	0,806	SEE PROB(AC)	
Intraclass correlations:		2*(RMZ-RDZ)	0,650	0,001	

Análise das estimativas das diferentes componentes de variância a partir de contraste de modelos

Esta parte da análise apresenta uma forma distinta, e um pouco mais complexa, de tratar a informação gêmealar.

Os gêmeos MZ são "cópias" um do outro, i.e., partilham os mesmos alelos idênticos por descendência e o mesmo envolvimento comum; os gêmeos DZ partilham, em média, metade dos seus alelos idênticos por descendência e o mesmo envolvimento comum. Decorre daqui que as correlações esperadas em cada zigotia sejam (NEALE & CARDON, 1992):

$$r_{MZ} = h^2 + c^2.$$

$$r_{DZ} = \frac{1}{2} h^2 + c^2.$$

Se subtrairmos os dois valores de correlação, obteremos

$$r_{MZ} - r_{DZ} = (h^2 + c^2) - (\frac{1}{2}h^2 + c^2)$$

$$r_{MZ} - r_{DZ} = h^2 + c^2 - \frac{1}{2}h^2 - c^2$$

$$r_{MZ} - r_{DZ} = \frac{1}{2}h^2$$

$$h^2 = 2(r_{MZ} - r_{DZ}).$$

Dado que a variância total é fixada em 1, e que é composta de forma aditiva pela variância genética (h^2), do ambiente comum (c^2) e do ambiente único (e^2), obtém-se:

$$h^2 + c^2 + e^2 = 1.$$

De acordo com regras simples de álgebra obtém-se uma estimativa da variância comumente partilhada no seio da família de natureza não-genética (c^2), i.e. efeitos ambientais comuns:

$$c^2 = 1 - h^2 - e^2$$

$$c^2 = 1 - 2r_{MZ} + 2r_{DZ} - 1 - r_{MZ}$$

$$c^2 = 2r_{DZ} - r_{MZ}.$$

Do mesmo modo, é fácil obter uma estimativa da variância que reflecte aspectos do ambiente único dos membros de cada par e que contém, também, aspectos de erros de medição (e_2):

$$e_2 = 1 - h^2 - c^2$$

$$e^2 = 1 - 2r_{MZ} + 2r_{DZ} - c^2$$

$$e^2 = 1 - 2r_{MZ} + 2r_{DZ} - 2r_{DZ} + 2r_{MZ}$$

$$e^2 = 1 - r_{MZ}.$$

Com base nestas regras simples de cálculo, é possível estimar as quantidades mais relevantes na explicação da variação observada no IAFTL (relembramos que o IAFTL se refere aos resíduos estudantizados). Ora a modelação de estruturas de covariância obriga à estimação iterativa, pelo método de máxima verosimilhança, das três fontes de variação total. O algoritmo implementado

no “software” TWINAN90 permite estimar, em simultâneo, as três fontes de variância, ao mesmo tempo que testa, sequencialmente, diversos modelos hierárquicos para explicar, “o melhor possível”, as matrizes de covariância entre gêmeos MZ e DZ. Este “melhor possível” é sempre baseado em diferenças de razões de verosimilhança (detalhes técnicos deste tipo de estimação e testagem sequencial de modelos pode ser encontrado em NEALE e CARDON (1992).

A vantagem deste tipo de abordagem reside no fato de se testarem hipóteses alternativas distintas para explicarem a extensão de variância e covariâncias entre os pares de gêmeos MZ e DZ. De fato, temos muito sinteticamente que organizar as diferentes fontes de explicação em domínios genéticos (aditivos e/ou dominância) e do envolvimento (comum e único). Daqui o recurso a modelos, mais adiante designados por ADE, ACE, CE, AE e E.

Interpretemos os resultados mais relevantes apresentados na TABELA 7:

- O “software” apresenta resultados para quatro modelos de partição da variância total. Um primeiro modelo (modelo ADE) com fatores genéticos aditivos (A), de dominância (D) e variância residual (E), também designada por envolvimento único dos gêmeos. Para cada parâmetro considerado é produzida a respectiva estimativa (\pm erro-padrão), a estimativa da proporção da variância total e o seu erro-padrão. A qualidade do modelo é apreciada a partir de duas peças importantes de informação: 1) a magnitude e sinal do parâmetro em causa, bem como; 2) o valor da qualidade de ajustamento do modelo em causa e cujo resultado é dado pelo valor da razão de verosimilhança (LR). Neste caso, $LR = 1,108$, com 3 graus de liberdade e um valor de prova excelente, $p = 0,778$ (relembremos os leitores que ao contrário dos testes clássicos de hipóteses em que se pretende rejeitar a hipótese nula, aqui é exactamente o contrário - a hipótese nula é a do autor, i.e., o modelo em causa é um excelente descritor da complexidade

das matrizes de variância-covariância entre pares de gêmeos MZ e DZ). O problema do Modelo ADE é que tem um valor negativo numa estimativa de variância, a de dominância, o que improvável.

- O segundo modelo, ACE, com os mesmos parâmetros do anterior, tem contudo uma outra configuração, uma vez que a variância total é fracionada em termos de fatores genéticos aditivos (A), de envolvimento comum (C) e único ou residual (E). A medida de ajustamento global é exactamente igual ao do modelo anterior. As estimativas de variância das três componentes contêm os respectivos erros-padrão, sugerindo que o valor da componente C não é estatisticamente significativo (basta ler com atenção a estimativa de C e o seu respectivo erro-padrão para se perceber que se construísse um intervalo de confiança o valor zero estaria contido).

- O modelo seguinte é o modelo AE, mais simples que os anteriores, uma vez que procura explicar toda a variação dos níveis do IAFTL através de duas componentes de variância - uma genética (A) e outra residual (E). Os valores dos parâmetros são elevados, os erros-padrão são baixos, e o resultado do LR é muito baixo (o que é excelente).

- O último modelo a ser considerado, CE, é um que só contém componentes de variância de natureza ambiental - comum (C) e única (E). Apesar dos valores dos parâmetros, o resultado do teste de LR é elevado, e o valor de prova é marginal ($p = 0,051$).

- Finalmente, o “software” produz uma comparação entre os diferentes modelos hierárquicos (i.e., de complexidade crescente) testando a sua qualidade. O melhor modelo (o mais parcimonioso), com menor valor de LR, e com parâmetros interpretáveis do ponto de vista estatístico e substantivo é o modelo AE. Temos assim que fatores genéticos aditivos (A) explicam cerca de 63% da variância total dos valores do IAFTL. O remanescente, i.e., 37% da variância deve-se a fatores ambientais únicos de cada membro do par.

TABELA 7 - Principais resultados da análise de diferentes modelos relativos ao fenótipo IAFL (partes mais importantes do "output" do TWINAN90).

TWINAN90 TWIN ANALYSIS PROGRAM, VERSION 1.0
TWINAN90 PATH ANALYSIS OUTPUT, page 1
FOR THE VARIABLE: iafl
NUMBER OF TWIN PAIRS: MZ = 84 DZ = 123

FOR THE COVARIANCE MATRICES

MZ =	0,9760	0,6412	DZ =	0,8620	0,2758
	0,6412	0,9970		0,2758	0,8189

ADE MODEL MAXIMUM LIKELIHOOD (ML) ESTIMATES

	Estimate	S. E.	Proportion	S. E.
Additive genetic variance	0,6741	0,2943	0,756	0,313
Dominance genetic variance	- 0,1176	0,2913	- 0,132	0,325
Error variance	0,3356	0,5124E-01	0,376	0,061

Goodness-of-fit likelihood ratio statistic = 1,108 ON 3 DF , P = 0,778

ACE MODEL (ML) ESTIMATES

	Estimate	S. E.	Proportion	S. E.
Additive genetic variance	0,4977	0,1729	0,558	0,189
Common environmental variance	0,5878E-01	0,1456	0,066	0,163
Error variance	0,3356	0,512E-01	0,376	0,061

Goodness-of-fit likelihood ratio statistic = 1,108 ON 3 DF , P = 0,778

AE MODEL (ML) ESTIMATES

	Estimate	S. E.	Proportion	S. E.
Additive genetic variance	0,5591	0,7844E-01	0,629	0,058
Error variance	0,3294	0,4715E-01	0,371	0,058

Goodness-of-fit likelihood ratio statistic = 1,256 ON 4 DF , P = 0,869

CE MODEL (ML) ESTIMATES

	Estimate	S. E.	Proportion	S. E.
Common environmental variance	0,4237	0,6945E-01	0,471	0,054
Error variance	0,4758	0,4700E-01	0,529	0,054

Goodness-of-fit likelihood ratio statistic = 9,427 ON 4 DF , P = 0,51

Model E vs model AE likelihood ratio statistic = 59,601 ON 1 DF , P = 0,000
Model E vs model CE likelihood ratio statistic = 51,430 ON 1 DF , P = 0,000
Model AE vs model ACE likelihood ratio statistic = 0,148 ON 1 DF , P = 0,700
Model CE vs model ACE likelihood ratio statistic = 8,319 ON 1 DF , P = 0,004
Model AE vs model ADE likelihood ratio statistic = 0,143 ON 1 DF , P = 0,700

Exploração do significado das diferenças intra-par

Uma avenida interessante de pesquisa gemelar tem centrado a sua atenção na exploração do significado das diferenças intra-par apenas nos gêmeos MZ, procurando entender a sua discordância. Um dos enfoques mais atuais é o que se refere ao domínio da Epigenética [os interessados nesta matéria podem consultar, por exemplo, PETRONIS (2006) ou o número especial da revista “Sciences et Avenir: Hors-Série (L'ENIGME, 2006)” dedicado inteiramente ao enigma dos gêmeos “verdadeiros” ou MZ]. Há muitas outras avenidas de pesquisa que percorrem aspectos metodológicos importantes (CARLIN, GURRIN, STERNE, MORLEY & DWYER, 2005; WESSEL, SCHORK, TIWARI & SCHORK, 2007), desenvolvimento embrionário (GOYEN, VEDDOVI & LUI, 2003), risco perinatal (RIESE, 2001), síndrome metabólica (GEHI, LAMPERT,

VELEDAR, LEE, GOLDBERG, JONES, MURRAH, ASHRAF & VACCARINO, 2008) e esquizofrenia (AHVENINEN, JÄÄSKELÄINEN, OSIPOVA, HUTTNNEN, ILMONIEMI, KAPRIO, LÖNNQUIST, MANNIEN, PAKARINEN, NÄÄTÄNEN & CANON, 2006), entre muitas outras.

Contudo, a nossa atenção neste texto é essencialmente metodológica. Neste sentido, exploraremos, muito sumariamente, algumas formas de representar este desafio de análise e interpretação. Duas das formas mais expeditas de estudar as diferenças intra-pares são as seguintes:

- A primeira forma exploratória é representar (histograma) as diferenças intra-par dos gêmeos no fenótipo em causa (resíduos estudantizados do IAFTL) para se ter uma ideia da sua distribuição, tal como representado na FIGURA 6 que mostra uma distribuição normal, mas com discordância (i.e., diferenças) à esquerda (valores menores do que -1) e à direita (valores maiores que +1) da distribuição.

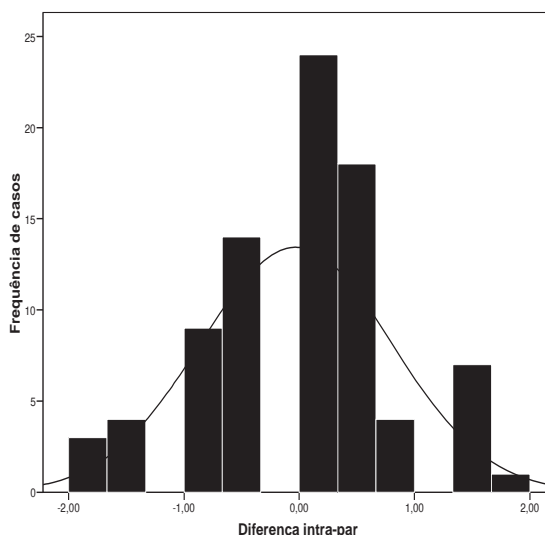
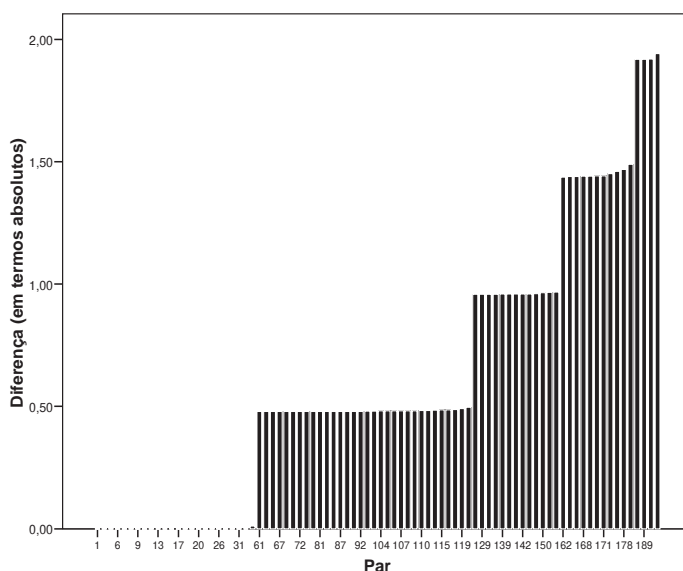


FIGURA 6 - Distribuição das diferenças intra-par nos gêmeos MZ (“output” do SPSS 15).

Outra forma, que não exploraremos, seria representar, separadamente, as diferenças intra-par nos gêmeos MZ masculinos e nos femininos, mas neste caso o fenótipo deveria estar ajustado para a covariável idade e/ou idade². Claro que tudo seria muito mais fácil se tivéssemos centenas de pares de gêmeos MZ masculinos e femininos da mesma idade cronológica, por exemplo todos terem 12 anos.

- Na segunda forma exploratória calcula-se a diferença em termos absolutos intra-par e faz-se a representação gráfica por par (FIGURA 7), ordenando a distribuição dos resultados. É evidente que há pares de gêmeos com diferenças nulas (têm exatamente o mesmo valor no IAFTL), enquanto que há outros, com diferenças “substanciais”, e que se situam acima de 1 (i.e., a diferença em termos absolutos intra-par é igual, ou superior, a uma unidade de desvio-padrão).



A numeração dos pares (valores ordenados) refere-se à numeração existente no ficheiro original do SPSS e não ao número de pares analisados.

FIGURA 7 - Distribuição da magnitude das diferenças intra-par nos gêmeos MZ (“output” do SPSS 15).

O passo seguinte seria dividir a distribuição dos resultados da FIGURA 7 em diferentes categorias (por exemplo a partir do percentil 90 da distribuição das diferenças, i.e., os mais discordantes) e estudar o seu “impacto” num outro conjunto de variáveis que se considera relevantes [ver o exemplo didático e bem esclarecedor no risco de asma (DUFFY, MITCHELL & MARTIN, 1998), ou nas influências ambientais nos momentos iniciais do desenvolvimento comportamental de humanos (ASBURY, DUNN, PIKE & PLOMIN, 2003)].

Por exemplo, no projecto de pesquisa GEAFAS (citado na amostra) será explorada, também, esta alternativa de análise nos valores da síndrome metabólica (valores contínuos dos cinco indicadores e frequências de risco); já no projeto GEM (também citado na amostra do presente estudo) será pesquisada a influência da discordância gêmear nos padrões de atividade física, na obesidade e nos níveis de aptidão física associados à saúde.

Conclusões e sugestões

As conclusões deste tutorial são as seguintes: 1) é bem evidente a necessidade de maior pesquisa com delineamentos gêmeares, de natureza descritiva e/ou experimental, no domínio da Educação Física e Ciências do Desporto nos países lusófonos; 2) foram apresentadas diferentes etapas da pesquisa gêmear, começando com aspectos de análise descritiva, gráfica e numérica, até ao plano da análise de variância; 3) é relevante considerar modelos alternativos para o fracionamento da variância total com base em modelação de estruturas de covariância; 4) é importante perspectivar a análise intra-par de gêmeos discordantes, pela sua potencialidade interpretativa, nos domínios da epigenética, e/ou nos domínios dos fatores de natureza sócio-cultural e educacional; 5) finalmente para reforçar a noção que o fenótipo “itlazer” tem uma forte componente genética, cerca de 63% da variância total.

As sugestões que apresentamos situam-se em dois planos: 1) no metodológico, e 2) no substantivo. Vejamos a primeira. Há necessidade de delinear pesquisa gêmear que considere não somente o teste a modelos mais complexos com informação univariada e/ou multivariada com dados transversais ou longitudinais tomando como amostra as cinco categorias de sexo nos gêmeos MZ e DZ; do mesmo modo se reclama mais investigação com o delineamento de “co-twin”, i.e., com gêmeos MZ discordantes, e pesquisa de natureza experimental (i.e., estudos de intervenção) com gêmeos MZ, e/ou MZ e DZ. A segunda avenida exige, num outro plano de problemas, que se efetuem no âmbito de convergência entre Genética Quantitativa e Genética Molecular com estudos de “Linkage” ou de Associação com genes candidatos. Qualquer que seja a avenida, os desafios a ultrapassar serão

bem interessantes, e as perguntas que encontrarão resposta colocarão a nova área da Genética do Desporto nos países lusófonos num plano de destaque bem merecido neste início de século.

Abstract

Analysis of twin data: a guided tour for researchers in Sport Sciences

This tutorial has the following purposes: 1) to summarize the importance of twin studies in Physical Education and Sport Sciences' research; 2) to show the basic structure of twin designs; 3) the levels of descriptive analysis; 4) the steps of growing complexity in data analysis – from analysis of variance to structural equation modeling; 5) as well as the analysis of discordant intra-pair differences of monozygotic twin data. The steps of this long list of procedures will be illustrated in this tutorial using the phenotype leisure time physical activity with data from 207 monozygotic and dizygotic twin pairs. Commentaries and explanations will be provided in all steps of the analysis, highlighting the fact that the used phenotype is genetically accounted for 63% of the total variation.

UNITERMS: Twins; Analysis of variance; Models; Physical activity.

Notas

1. Por covariável entende-se uma variável que pode condicionar, reduzindo ou amplificando, os valores de heritabilidade. Exemplos de covariáveis neste tipo de estudos são, por exemplo, a idade, o sexo, o estatuto sócio-econômico, e respectivas interações.

Luciano Basso - Bolsa de estágio de doutorado Capes - Proc. no. 4434-07-2*.

Projeto financiado pela Fundação para a Ciência e Tecnologia, POCTI/DES/62499/2004.

Referências

- AHVENINEN, J.; JÄÄKELÄINEN, I.P.; OSIPOVA, D.; HUTTINEN, M.O.; ILMONIEMI, R.J.; KAPRIO, J.; LÖNNWUIST, J.; MANNINEN, M.; PAKARINEN, S.; NÄÄTÄNEN, R.; CANON, T.D. Inherited auditory-cortical dysfunction in twin pairs discordant for schizophrenia. **Biological Psychiatry**, New York, v.60, n.6, p.612-20, 2006.
- ASBURY, K.; DUNN, J.F.; PIKE, A.; PLOMIN, R. Nonshared environmental influences on individual differences in early behavioral development: a monozygotic twin differences study. **Child Development**, Malden, v.74, n.3, p.933-43, 2003.
- BAECKE, J.A.H.; BUREMA, J.; FRIJTERS, J.E. A short questionnaire for the measurement of habitual physical-activity in epidemiological-studies. **American Journal of Clinical Nutrition**, Bethesda, v.36, n.5, p.936-42, 1982.
- BEUNEN, G.; THOMIS, M.; MAES, H.H.; LOOS, R.; MALINA, R.M.; CLAESSENS, A.L.; VLIETINCK, R. Genetic variance of adolescent growth in stature. **Annals of Human Biology**, London, v.27, n.2, p.173-86, 2000.
- BOOMSMA, D.; BUSJAHN, A.; PELTONEN, L. Classical twin studies and beyond. **Nature Reviews Genetics**, London, v.3, n.11, p.872-82, 2002.
- BOUCHARD, C.; CHAGNON, M.; THIBAUT, M.C.; BOULAY, R.R.; MARCOTTE, M.; COTE, C.; SIMONEAU, J.A. Muscle genetic-variants and relationship with performance and trainability. **Medicine and Science in Sports and Exercise**, Madison, v.21, n.1, p.71-7, 1989.
- BOUCHARD, C.; MALINA, R.M.; PÉRUSSE, L. **Genetics of fitness and physical performance**. Champaign: Human Kinetics, 1997.
- BOUCHARD, C.; PÉRUSSE, L.; LEBLANC, C. Using Mz twins in experimental research to test for the presence of a genotype-environment interaction effect. **Acta Geneticae Medicae et Gemellologiae**, Rome, v.39, n.1, p.85-9, 1990.

- CARLIN, J.B.; GURRIN, L.C.; STERNE, J.A.; MORLEY, R.; DWYER, T. Regression models for twin studies: a critical review. **International Journal of Epidemiology**, London, v.34, n.5, p.1089-99, 2005.
- CHRISTIAN, J.C.; WILLIAMS, C.J. Comparison of analysis of variance and likelihood models of twin data analysis. In: SPECTOR, T.D. et al (Eds.). **Advances in twin and sib-pair analysis**. Oxford: Oxford University Press, 2000. p.103-18.
- DUFFY, D.L.; MITCHELL, C.A.; MARTIN, N.G. Genetic and environmental risk factors for asthma: a cotwin-control study. **American Journal of Respiratory and Critical Care Medicine**, New York, v.157, n.3, p.840-5, 1998.
- DUPONT, W.D. **Statistical modeling for biomedical researchers: a simple introduction to the analysis of complex data**. Cambridge: Cambridge University Press, 2002.
- EDWARDS, K.L.; NEWMAN, B.; MAYER, E.; SELBY, J.V.; KRAUS, R.M.; AUSTIN, M.A. Heritability of factors of the insulin resistance syndrome in women twins. **Genetic Epidemiology**, New York, v.14, n.3, p.241-53, 1997.
- ESSEN-MØLLER, E. Die beweiskraft der ahnlichkeit im vaterschaftsnachweis: theoretische grundlagen. **Mitt. Anthropol. Ges.**, Wien, v.68, p.9-53, 1939.
- FALCONER, D.S. **Introduction to quantitative genetics**. Harlow: Longman Scientific & Technical, 1989.
- GEHI, A.; LAMPERT, R.; VELEDAR, E.; LEE, F.; GOLDBERG, J.; JONES, L.; MURRAH, N.; ASHRAF, A.; VACCARINO, V. A twin study of metabolic syndrome and autonomic tone. **Journal of Cardiovascular Electrophysiology**, Malden, v.34, p.1-7, 2008.
- GOYEN, T.-A.; VEDDOVI, M.; LUI, K. Development outcome of discordant premature twins at 3 years. **Early Human Development**, Amsterdam, v.73, p.27-37, 2003.
- HAUSPIE, R.C.; BERGMAN, P.; DIELICKI, T.; SUSANNE, C. Genetic variance in the pattern of the growth curve for height: a longitudinal analysis of male twins. **Annals of Human Biology**, London, v.21, n.4, p.347-62, 1994.
- KAPRIO, J.; PULKKINEN, L.; ROSE, R.J. Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. **Twin Research**, Bowen Hills, v.5, n.5, p.366-71, 2002.
- LAREU, M.V.; PHILLIPS, C.P.; CARRACEDO, A.; LINCOLN, P.J.; SYNDERCOMBE COURT, D.; THOMSON, J.A. Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. **Forensic Science International**, Limerick, v.66, n.1, p.41-52, 1994.
- L'ENIGME des vrais jumeaux. **Sciences et Avenir**, Paris, v.149, 2006. Hors-série.
- MAIA, J. Genética e desporto. In: BENTO, J.; CONSTANTINO, J. (Eds.). **Em defesa do desporto: mutações e valores em conflito**. Coimbra: Almedina, 2007. p.147-80.
- MAIA, J.; LOPES, V.P.; MORAIS, F.P. **Atividade física e aptidão física associada à saúde: um estudo de epidemiologia genética em gémeos e suas famílias realizado no arquipélago dos Açores**. Porto: FCDEF-UP/DREFD, 2001.
- MAIA, J.A.R.; THOMIS, M.; BEUNEN, G. Genetic factors in physical activity levels: a twin study. **American Journal of Preventive Medicine**, Amsterdam, v.23, n.2, p.87-91, 2002.
- NEALE, M.; EAVES, L. Research design and methods to study the genetic epidemiology of obesity. In: ALLISON, D.B.; PI-SUNIER, F.X. (Eds.). **Obesity treatment**. New York: Plenum Press, 1995. p.225-38.
- NEALE, M.C.; CARDON, L.R. **Methodology for genetic studies of twins and families**. Dordrecht: Kluwer Academic, 1992.
- PEETERS, M.W.; THOMIS, M.A.; CLAESSENS, A.L.; LOSS, R.J.; MAES, H.H.; LYSSENS, R.; VANDEN EYNDE, B.; VLIETINCK, R.; BEUNEN, G. Heritability of somatotype components from early adolescence into young adulthood: a multivariate analysis on a longitudinal twin study. **Annals of Human Biology**, London, v.30, n.4, p.402-18, 2003.
- PELTONEN, L. GenomEUtwin: a strategy to identify genetic influences on health and disease. **Twin Research**, Bowen Hills, v.6, n.5, p.354-60, 2003.
- PETRONIS, A. Le théoricien de l'épigénétique. **Sciences et Avenir**, Paris, v.149, 2006. Hors-série.
- PLOMIN, R.; DEFRIES, J.C.; CRAIG, I.W.; MCGUFFIN, P. **Behavioral genetics in the postgenomic era**. Washington: American Psychological Association, 2002.
- PLOMIN, R.; DEFRIES, J.C.; MCCLEARN, G.E.; MCGUFFIN, P. **Behavioral genetics**. 5.ed. New York: Worth, 2008.
- RIESE, M.L. Discordant and nondiscordant twins: comparative multimethod risk assessment in the neonatal period. **Journal of Developmental and Behavioral Pediatrics**, Hagerstown, v.22, n.2, p.102-12, 2001.
- SAPAGE, I.L.L.; MAIA, J.R.; SEABRA, A.; SILVA, R.G.; LOPES, V.P. **Fatores genéticos e ambientais nos níveis e padrões de atividade física: um estudo em gémeos**. Porto: FCDEF, 2007.
- SIMONEAU, J.A.; LORTIE, G.; BOULAY, M.R.; MARCOTTE, M.; THIBAUT, M.C.; BOUCHARD, C. Inheritance of human skeletal-muscle and anaerobic capacity adaptation to high-intensity intermittent training. **International Journal of Sports Medicine**, Stuttgart, v.7, n.3, p.167-71, 1986.

- STEVENS, J.P. **Applied multivariate statistics for the social sciences**. 4th. ed. Mahwah: Lawrence Erlbaum, 2002.
- TEIXEIRA E SEABRA, A.F.; MAIA, J.A.; MENDONÇA, D.M.; THOMIS, M.; CASPERSEN, C.T.; FULTON, J.E. Age and sex differences in physical activity of Portuguese adolescents. **Medicine and Science in Sports Exercise**, Madison, v.40, n.1, p.65-70, 2008.
- WATANABE, T.; MUTOH, Y.; YAMAMOTO, Y. Genetic variance in age-related changes in running performance and growth during adolescence: a longitudinal twin study. **American Journal of Human Biology**, New York, v.13, n.1, p.71-80, 2001.
- WESSEL, J.; SCHORK, A.J.; TIWARI, H.K.; SCHORK, N.J. Powerful designs for genetic association studies that consider twins and sibling pairs with discordant genotypes. **Genetic Epidemiology**, New York, v.31, n.7, p.789-96, 2007.
- WILLIAMS, C.J.; CHRISTIAN, J.C.; NORTON JUNIOR, J.A. Twinan90: a Fortran program for conducting anova-based and likelihood-based analyses of twin data. **Computer Methods and Programs in Biomedicine**, Amsterdam, v.38, n.2-3, p.167-76, 1992.

ENDEREÇO

José António Ribeiro Maia
Lab. Cineantropometria e Gabinete de Estatística Aplicada
Faculdade de Desporto
Universidade do Porto
R. Dr. Plácido Costa, 91
4200-450 - Porto - PORTUGAL
e-mail: jmaia@fade.up.pt

Recebido para publicação: 08/04/2011

Aceito: 13/05/2011