

Uma nota didáctica breve no uso esclarecido de procedimentos estatísticos em análise de dados repetidos no tempo. Um estudo guiado para investigadores das Ciências do Desporto

José A. R. Maia¹
Rui M. Garganta¹
André Seabra¹
Vitor P. Lopes²
António Prista³
Duarte Freitas⁴

¹ Faculdade de Ciências do Desporto e de Educação Física
Universidade do Porto, Portugal

² Instituto Politécnico de Bragança, Portugal

³ Faculdade de Ciências da Educação Física e Desporto
Universidade Pedagógica, Moçambique

⁴ Secção Autónoma de Educação Física e Desporto
Universidade da Madeira, Portugal

RESUMO

Este texto pretende ser um auxiliar didáctico no uso esclarecido de procedimentos estatísticos relativos à análise longitudinal de dados. Servir-nos-emos de um exemplo ilustrativo de complexidade crescente para introduzir a estrutura de um delineamento de grupo único e da essência do ensaio de hipóteses estatísticas. De seguida apresentamos os principais resultados do uso de teste t e da análise de variância de medidas repetidas. Os resultados são interpretados de modo formal e substantivo, introduzindo um pensamento alternativo à estrutura binária do resultado do ensaio de hipóteses. O recurso a procedimentos gráficos é fortemente explorado. Finalmente, recorremos à modelação hierárquica para salientar a sua riqueza e flexibilidade interpretativa no estudo de dados longitudinais.

Palavras-chave: dados longitudinais, teste t, análise de variância, exploração gráfica, modelação hierárquica.

ABSTRACT

A didactical note on the use of statistical procedures for longitudinal data. A guide for researchers in Sport Sciences

This report aims at being a didactical tool to instruct in the use of statistical procedures related to the analysis of longitudinal data sets. We shall use an example of growing complexity, so that we may introduce the basic ideas of a single group design and the essence of hypothesis testing. The main results of the t test and analysis of variance shall be analyzed. These results are formally and substantively interpreted, and an alternative way of thinking is introduced, as a contrast to the binary results of the hypothesis testing. We shall focus on graphic displays of data. Finally, hierarchical modeling is introduced, mainly its richness and flexibility in terms of interpretation of the data structure at hand.

Key Words: longitudinal data, t test, analysis of variance, graphic displays, hierarchical modeling.

INTRODUÇÃO

O propósito essencial desta nota breve não é o de substituir capítulos de um qualquer manual de Estatística, tão pouco instruir no uso exclusivo dos procedimentos do t teste e da análise de variância (ANOVA)¹. Pretende, isso sim, explorar diferentes alternativas de interrogar um conjunto de dados repetidos no tempo e, deste modo, ser um auxiliar didáctico para um pesquisador iniciante que nem sempre é capaz de se libertar do estereótipo de só se concentrar na magnitude do valor da prova (p), que nem sempre sabe o que significa, mas que espera que seja sempre inferior ao mágico número de 5%. Isto, apesar de na maior parte das vezes não ter reflectido seriamente porque é que tal valor é fixado, “desde longa data”, em 5%.

Está pois lançado o propósito desta nota breve. A sua estratégia é bem simples, e percorrerá as seguintes etapas:

Apresentaremos, em primeiro lugar, aspectos essenciais de um delineamento de grupo único com observações repetidas no tempo. Por questões de natureza gráfica, e para não abordar procedimentos mais complexo², limitaremos o número de observações temporais a três.

De seguida discutiremos, do ponto de vista formal e substantivo, o problema do ensaio de hipóteses e o nível do erro que se está disposto a “correr” para rejeitar uma hipótese nula verdadeira.

Os dados para os dois pontos do tempo serão descritos e situados num contexto bem conhecido - resultados numéricos da aplicação de um programa de treino para melhorar a força explosiva dos músculos extensores do joelho.

Em quarto lugar apresentaremos os resultados da aplicação do t teste, onde abordaremos duas interpretações alternativas e bem mais interessantes - a do intervalo de confiança para a média das diferenças e a magnitude do efeito.

Passaremos à exploração gráfica dos resultados centrado a nossa atenção no comportamento ordenado das diferenças e seu significado.

Um novo aspecto do comportamento dos dados nos dois pontos do tempo será explorado a partir da noção de estabilidade das mudanças intraindividuais nas diferenças interindividuais. Será utilizado, aqui, o coeficiente de correlação intraclasse.

Apresentaremos de seguida a vantagem em ter mais do que dois pontos no tempo. O modelo estatístico utilizado será a ANOVA de medidas repetidas. Serão lançadas e discutidas várias possibilidades para ensaio de hipóteses. Discutiremos o problema nuclear da simetria composta ou esfericidade e o recurso alternativo a uma análise univariada ou multivariada.

Abordaremos, tal como no ponto anterior, interpretações alternativas ao teste formal de hipóteses e que são os intervalos de confiança e o eta quadrado (η^2). O comportamento dos resultados será explorado graficamente a partir dos diagramas de extremos e quartis (*Box plot*) e do *high-low-close* dos dados ordenados. A sua importância e interpretação serão salientadas.

Veremos o interesse em pesquisar aspectos da tendência dos resultados e uma opção bem interessante para o teste de *post-hoc* ou testes *a posteriori*.

Trataremos, também, do problema da estabilidade ou *tracking* das mudanças intraindividuais no seio do grupo de observações, recorrendo ao coeficiente de correlação intraclasse.

Finalmente apresentaremos, ainda que de modo muito breve, diferentes formulações ou modelos que impõem comportamentos distintos aos dados. Para tal recorreremos à modelação hierárquica ou multinível (MHMN), um modelo estatístico altamente versátil e flexível, e bem mais interessante do que a ANOVA.

Os procedimentos gráficos e de análise quantitativa estão implementados nos *softwares* que utilizaremos, e que são o SPSS 12 e o SYSTAT 10. Porções relevantes dos *outputs* serão “coladas” no texto para facilitar a sua apresentação e interpretação.

DELINEAMENTO DE GRUPO ÚNICO

ASPECTOS ESSENCIAIS.

O plano conceptual e operativo de uma qualquer pesquisa reclama do investigador, para além do enquadramento teórico das inquietações que emergem dos problemas a investigar, uma atenção redobrada à validade das variáveis utilizadas e fiabilidade da informação, i.e., ao controlo apertado que faz de todo o processo de avaliação e medição. A validade de generalização dos seus resultados depende, sempre, do modo como reflectiu e solucionou as questões centrais da validade interna e externa da sua pesquisa (3).

O delineamento de grupo único pode assumir várias formas - desde a estrutura mais simples da pesquisa quasi-experimental de pré-pós, até ao estudo mais sofisticado de *cross-lagged panel* (5, 8, 21).

O delineamento de grupo único, que procura marcar numa estrutura longitudinal aspectos da história natural de um evento, exige um controlo apertado de todo o processo de medição, por forma a garantir uma elevada qualidade da informação. Este controlo implica a execução de estudos-piloto prévios que informem sobre a precisão de todo o protocolado de avaliação, bem como de uma estrutura designada de *reliability in field*, única forma de controlar, para efeitos de aprendizagem nas medições, e estimar a magnitude da variância erro implícita em todos os momentos de medição. Estes aspectos remetem-nos, necessariamente, para a frequência de amostragem das medições. Quando lidamos com o lato universo das aptidões físico-motoras, registos semestrais ou anuais são mais do que suficientes para evidenciar verdadeira mudança intraindividual. Contudo, por questões de natureza operativa, pode acontecer que por motivos de monitorização do treino das aptidões haja a necessidade de impor uma outra estrutura na temporalidade das medições, que pode ser uma frequência mensal (20). Aqui enfrentamos, necessariamente, aspectos de variabilidade intraindividual que podem não representar uma verdadeira mudança e uma quantidade associada à verdadeira mudança intraindividual. Há pois que delinear a pesquisa no sentido de recorrer a procedimentos estatísticos suficientemente flexíveis para separar aquilo que é variabilidade intraindividual e aquilo que é mudança intraindividual (11, 15, 16, 17).

O PROBLEMA DO ENSAIO DE HIPÓTESES

Um dos problemas centrais de toda a investigação de natureza empírica prende-se com o entendimento claro da estrutura conceptual e operativa do ensaio de hipóteses, tal como é entendida no quadro formal da estatística inferencial.

Convém que fique claro, desde já, que uma coisa é o conteúdo semântico da estrutura de uma qualquer hipótese substantiva avançada por um investigador e outra coisa é a sua transposição para o domínio da formulação “simples” em termos estatísticos. É importante que se diga, e de modo inequívoco, que

a(s) hipótese(s) avançada(s) pelo investigador raramente é (são) colocada(s) à prova em termos estatísticos. Dito por outras palavras, a hipótese do investigador nunca é formalmente testada. Aquilo que se ensaia é, tão-somente, uma outra hipótese contrária à do investigador, e que este pretende rejeitar, que se designa por hipótese nula. O “caricato” da situação é o seguinte: o investigador assume que a hipótese nula é verdadeira (!), e no entanto pretende rejeitá-la em favor de uma alternativa (a sua hipótese) que nunca testa formalmente, e que portanto aceita se se verificar que os dados fornecem evidência substancial contra a hipótese nula.

No ensaio de hipóteses lida-se com dois conceitos intimamente associados e relativamente aos quais nem sempre se entende a sua diferença e significado. São eles o nível de significância (α) e o valor da prova (p). Para ilustrar a confusão que se estabelece entre a estrutura operativa do ensaio de hipóteses, o α e o p , vamos servir-nos de um exemplo muito bem relatado (5), e que é o seguinte:

Suponha que um dado programa de treino é capaz de alterar o desempenho numa dada tarefa. O nível de significância foi previamente fixado em 5%.

Depois de comparar as médias dos grupos experimentais e de controlo (digamos com 10 sujeitos cada um), os resultados obtidos com o teste t independente foram os seguintes: $t=2.7$; graus de liberdade=18; $p=0.001$. Marque agora com verdadeiro ou falso as seguintes possibilidades de conclusão: 1ª A hipótese nula, que estabelecia que não havia diferenças entre populações, é rejeitada completamente;

2ª Encontrou-se a probabilidade da hipótese nula ser verdadeira;

3ª Provou-se, de modo absoluto, a hipótese experimental;

4ª Pode deduzir-se a probabilidade da hipótese experimental ser verdadeira;

5ª Sabe-se, se se rejeitar a hipótese nula, que existe a probabilidade em cometer uma decisão errada;

6ª Tem-se uma experiência fiável no sentido em que, se fosse hipoteticamente repetida um grande número de vezes, se obteria um resultado significativo em 99% das ocasiões.

De facto, as seis possibilidades de resposta estão todas erradas. Se quisermos interpretar o resultado

da experiência com base no valor da prova, a sua correcta interpretação seria a seguinte: há uma probabilidade de 1 em 1000 em obter os dados observados (ou um conjunto de dados que representem um desvio mais extremo da hipótese nula) se a hipótese nula for verdadeira.

Claramente que aquilo que o autor pretende evidenciar (fazendo eco das posições de um número cada vez maior de estatísticos e metodólogos) é, tão-somente, a informação extremamente limitada do valor da prova (p) face à natureza binária da decisão associada ao teste de hipóteses. A sua sugestão, muito mais esclarecedora em termos interpretativos, é recorrer à informação proveniente dos intervalos de confiança, da magnitude do efeito, ou da noção de variância explicada. É este percurso que iremos apresentar com base num exemplo bem simples.

OS DADOS DO ESTUDO

Os dados desta ilustração provêm de um estudo realizado num conjunto de voleibolistas seniores masculinos. Tratou-se de uma pesquisa que pretendia mapear a “história natural” da resposta ao treino dos atletas que foram submetidos a um programa específico de pliometria³, no sentido de melhorar substancialmente o seu desempenho explosivo no salto vertical.

A avaliação da impulsão vertical foi realizada na plataforma de Bosco, e os resultados obtidos referem-se ao salto com contra-movimento. Foi realizado um reteste com uma semana de intervalo numa sub-amostra de 10 atletas para estimar a fiabilidade do seu desempenho, e o valor obtido do coeficiente de correlação intraclasse foi de 0.87.

Os resultados dos 20 atletas são pois os seguintes (obtidos em 3 pontos do tempo espaçados, cada um, de 1 mês).

Quadro 1: Valores de impulsão vertical (salto com contra-movimento) dos 20 atletas nos 3 pontos do tempo (valores em cm).

Atletas	t1 (momento 1)	t2 (momento 2)	t3 (momento 3)
1	44,0	48,5	51,5
2	37,5	46,4	49,1
3	41,9	43,7	47,4
4	45,9	49,7	53,0
5	46,5	48,2	50,3
6	40,7	43,6	46,5
7	40,9	43,2	46,3
8	40,1	42,6	47,4
9	40,3	44,5	46,9
10	40,5	43,7	47,5
11	42,7	45,8	49,2
12	41,3	44,3	44,3
13	42,6	45,2	47,4
14	40,7	41,5	41,9
15	40,7	41,2	43,1
16	41,0	43,6	44,1
17	40,6	44,0	44,6
18	41,4	44,1	44,9
19	41,8	41,1	41,2
20	40,2	42,7	45,6

Teste t

Se partirmos do princípio, e é sempre este o ponto de partida desta pesquisa, que o programa foi bem delineado, é de esperar a sua eficácia que se traduzirá, naturalmente, nos incrementos dos valores de força explosiva dos atletas já no 2º momento de avaliação (i.e., em t2). É pois esta a hipótese que o investigador gostaria de testar, mas que não consegue realizar de acordo com o formalismo estatístico dos ensaios de hipóteses.

Medidas descritivas e teste formal

No Quadro 2 temos os resultados das medidas descritivas mais importantes. De t1 para t2 verificou-se um incremento no valor das médias, de 41.57 cm passou para 44.38 cm. Regista-se, também, um ligeiro aumento na variação interindividual da performance no salto vertical no segundo momento de avaliação (de 2.04 para 2.36 cm).

Quadro 2: Medidas descritivas do salto vertical de t1 (MOM1) para t2 (MOM2), output obtido no SPSS 12.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	MOM1	41,5650	20	2,03787	,45568
	MOM2	44,3800	20	2,35721	,52709

Os principais resultados do teste t estão no Quadro 3. O valor da média das diferenças é de 2.82 cm (incremento médio de t1 para t2). Dado que $t(19) = -6.655$, $p < 0.001$, rejeita-se a hipótese nula ($H_0: \mu_d = 0$). Se a hipótese nula fosse verdadeira, a probabilidade de se verificar uma média de diferenças na impulsão vertical da magnitude de 2.82 cm por mero acaso seria menor do que 1 vez em 1000. Dado que a probabilidade é extremamente baixa, há

que rejeitar a hipótese nula e aceitar a alternativa, que é a do pesquisador e que não é submetida a qualquer teste formal. Esta só é aceite caso se rejeite a que lhe é oposta, a nula.

De um modo estatístico, trata-se aqui de testar a hipótese do programa não ter tido qualquer eficácia, o que é à partida uma insensatez, dado ser contrária àquilo que está nos planos do investigador. Mas é assim mesmo. Não há qualquer engano formal.

Uma vez que se aceita a hipótese alternativa, a que sugere que há ganhos significativamente diferentes de zero, estamos a suportar, indirectamente, a eficácia do programa na melhoria da força explosiva dos membros inferiores dos atletas. E nada mais há a extrair do teste formal de hipóteses, i.e., aquilo que o teste formal dá é tão-somente um contributo numa decisão dicotómica e que é efectuada em termos probabilísticos.

Quadro 3: Resultados da aplicação do teste t (output do SPSS 12).

Paired Samples Test									
		Paired Differences			95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	MOM1 - MOM2	-2,81500	1,89161	,42298	-3,70030	-1,92970	-6,655	19	,000

Interpretação alternativa

Uma interpretação alternativa ao resultado formal do teste de hipóteses, que só informa sobre a rejeição ou não da hipótese nula, é centrar a nossa atenção na elucidação da “qualidade” das mudanças ocorridas. Vejamos, pois, o resultado do intervalo de confiança a 95% para a média das diferenças, e que é de 1.93 a 3.70. Isto significa que em termos populacionais, e se submetêssemos os atletas seniores de voleibol a um programa pliométrico igual ao do presente estudo, a média dos ganhos situar-se-ia, com uma confiança de 95%, entre 2.0 e 4.0 cm, aproximadamente. A questão que aqui se levanta é justamente saber, se os resultados que se encontram neste intervalo são

relevantes em termos de expressão de força explosiva para se submeter atletas a um programa desta natureza. Se pensarmos, também, em termos da magnitude do efeito em termos relativos $[(\text{Média}_{t2} - \text{Média}_{t1}) / \text{Média}_{t1}] \times 100$, os ganhos de força explosiva só reflectem a “história natural” desta mudança em cerca de 7%. Uma vez mais poderíamos perguntar se um programa bem delineado e que implicou 4 treinos semanais durante 4 semanas só consegue induzir alterações percentuais numa magnitude de 7%. A representação gráfica relativa aos ganhos individuais expressos em termos percentuais está na figura seguinte.

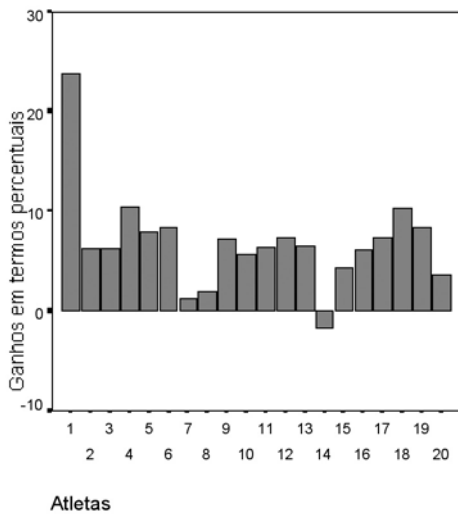


Figura 1: Ganhos individuais de t1 para t2 expressos em termos percentuais.

Uma forma gráfica interessante e alternativa da anterior no olhar das alterações que ocorrem de t1 para t2 é a que representa, de modo ordenado, os resultados numéricos das diferenças de valores de cada atleta (Figura 2).

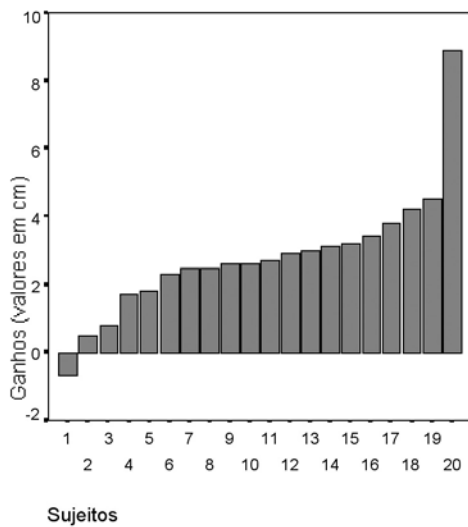


Figura 2: Representação gráfica dos ganhos em termos absolutos, ordenados pelo valor mais baixo (output obtido no SPSS 12).

Estamos diante de uma representação que salienta alguma heterogeneidade individual na resposta ao treino e que pode ter explicações variadas de natureza neuro-muscular. Há 1 atleta que diminui o seu valor de impulsão vertical, há quem ganhe de 1 a 2 cm, de 3 a 4 cm, ou mais de 8 cm. A questão a explorar é saber a que se deve a circunstância de haver sujeitos com “resistência” à mudança nos seus valores de força explosiva, enquanto que há outros de forte sensibilidade ao treino pliométrico.

Estabilidade

Os atletas não têm todos o mesmo valor de partida (em t1), dado que o desvio-padrão é de 2.04 cm. É pois provável que a sua resposta ao treino seja distinta (conforme vimos anteriormente). Nesta circunstância, é importante verificar se os ganhos de t1 para t2 evidenciam estabilidade, i.e., se não se identificam cruzamentos nas trajetórias interindividuais (ver Figura 3).

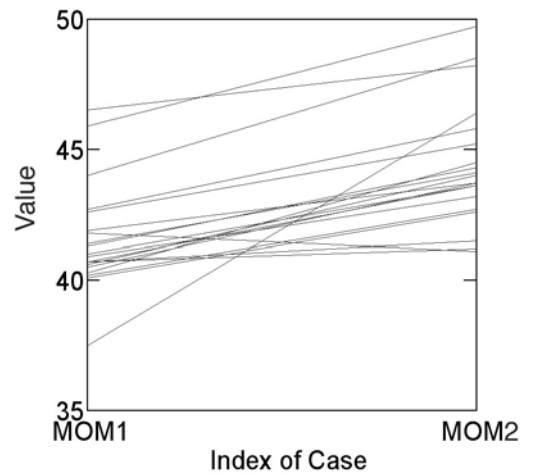


Figura 3: Trajetórias dos atletas de t1 (MOM1) para t2 (MOM2), gráfico obtido no SYSTAT 10.

De facto, há cruzamentos nas trajetórias dos atletas. O coeficiente de correlação intraclassa (obtido no SPSS 12) é de 0.774 (ler na linha *average measures*), com um intervalo de confiança de 0.429 a 0.911 (ver porção do *output* no Quadro 4), indicador

de alguma instabilidade nos ganhos e que reflecte a ausência de forte manutenção da posição relativa dos sujeitos no seio do grupo.

Quadro 4: Estatísticas relativas ao coeficiente de correlação intraclasse (average measures).

Intraclass Correlation Coefficient							
	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig.
Single Measures	,631 ^b	,273	,836	4,427	19,0	19	,001
Average Measures	,774 ^c	,429	,911	4,427	19,0	19	,001

Two-way mixed effects model where people effects are random and measures effects are fixed.

- Type C intraclass correlation coefficients using a consistency definition—the between-measure variance is excluded from the denominator variance.
- The estimator is the same, whether the interaction effect is present or not.
- This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

ANOVA

Dois pontos no tempo definem, exclusivamente, uma recta. Contudo, não é pacífico, no domínio das variáveis de interesse no território das Ciências do Desporto, que os fenómenos a observar tenham comportamentos que possam ser descritos exclusivamente por rectas. Com mais pontos de registo temporal é possível identificar outros comportamentos nos dados, os quais podem ser de natureza não-linear. Ainda que a maioria dos autores sugira que 4 pontos são o mínimo necessário para se ter uma noção mais adequada do comportamento de uma variável, neste exemplo só consideraremos 3 pontos, tal como foi explicado no início do texto.

Aquilo que aqui se pretende verificar é, uma vez mais, a hipótese substantiva do pesquisador. Dado que o programa foi bem delineado e os atletas responderam com empenho e sem qualquer falta ao treino, é de esperar que a sua resposta, ainda que evidencie alguma variabilidade, seja traduzida em incrementos distintos nas médias nos três pontos do tempo. A haver mudanças na força explosiva dos membros inferiores, a única causa plausível é a que se refere à aplicação do programa. No caso vertente, é formulada, em termos estatísticos, a seguinte hipótese: $H_0: \mu_1 = \mu_2 = \mu_3$ (contra a alternativa, $H_1: \mu_i \neq \mu_j$, i.e., pelo menos duas médias em dois pontos do tempo são significativamente diferentes). Uma vez

mais enfrentamos a situação da ausência de qualquer teste formal à hipótese do investigador que refere, justamente, que o programa foi eficaz, e que portanto há diferenças significativas nos valores médios de força explosiva em pelo menos dois pontos do tempo. Como não há um teste formal àquilo que o investigador pretende evidenciar, há que testar a validade de H_0 , e esperar que seja rejeitada! E bem mais difícil de entender é a assunção de que é verdadeira!

Medidas descritivas e teste formal

No Quadro 5 estão descritos os resultados numéricos das estatísticas elementares para os três pontos no tempo. Verifica-se um incremento nas médias, que é mais ou menos equivalente entre pontos adjacentes no tempo. Constata-se, também, que a variabilidade do desempenho aumenta no tempo, já que se verifica um ligeiro incremento nos valores dos desvios-padrão.

Quadro 5: Medidas descritivas da força explosiva nos três pontos do tempo (output do SPSS 12).

Descriptive Statistics			
	Mean	Std. Deviation	N
MOM1	41,5650	2,03787	20
MOM2	44,3800	2,35721	20
MOM3	46,6100	3,05664	20

Independentemente de se assumir que os registos repetidos no tempo são independentes e provêm de uma distribuição normal multivariada, é importante assumir, também (ainda que não seja uma condição necessária), simetria composta para a matriz de covariância das medidas repetidas. Trata-se aqui de assumir que as variâncias dos resultados nos 3 pontos do tempo são iguais e que as covariâncias entre momen-

tos são também iguais. Uma face da simetria composta refere-se, exclusivamente, à noção de esfericidade ou circularidade, que trata da noção de que as variâncias nos três pontos do tempo não devem ser significativamente diferentes entre si. O teste formal da ANOVA de medidas repetidas assenta nestes pressupostos. Uma forma de verificar o pressuposto da esfericidade é recorrer ao teste W de Mauchly (ver Quadro 6).

Quadro 6: Resultados do teste W de Mauchly ao pressuposto da esfericidade (output do SPSS 12).

Mauchly's Test of Sphericity^b

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
tempo	,480	13,193	2	,001	,658	,688	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

- a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
- b.

Design: Intercept
 Within Subjects Design: tempo

Dado que o valor da prova do teste é de 0.001, é rejeitada a hipótese da esfericidade. A extensão desta violação é reflectida num parâmetro designado de epsilon. Quando se verifica que a matriz de covariância das medidas repetidas não viola o pressuposto de esfericidade, o parâmetro epsilon é igual a 1. O pior resultado da violação é dado por $1/(k-1)$, em que k é igual ao número de medidas repetidas. No nosso caso a estimativa de epsilon é igual a 0.50 e corresponde ao limite inferior do epsilon, conforme é visível no Quadro 6. Greenhouse-Geisser e Huynh-Feldt propõem fórmulas alternativas para o cálculo deste parâmetro e daqui que o seu resultado seja ligeiramente diferente. Como foi violado o pressuposto da esfericidade há que recorrer a uma correcção do teste F da ANOVA com base nas sugestões de Greenhouse-Geisser ou Huynh-Feldt. Estas correcções modificam os valores dos graus de liberdade conforme pode ser consultado no Quadro 7. Ainda que as correcções de Greenhouse-Geisser e Huynh-

Feldt sejam ligeiramente diferentes, uma (Greenhouse-Geisser) mais conservadora do que a outra (Huynh-Feldt), o valor de epsilon, tradutor da magnitude da violação da esfericidade, não é superior a 0.75⁴ e o valor da prova é sempre significativo em qualquer das situações, é sugerida a opção pela correcção de Greenhouse-Geisser. A conclusão é pois a rejeição da hipótese nula, dado que $F(1.316,25.008)=60.898$, $p<0.001$. Neste caso aceita-se a hipótese alternativa que é a que o investigador enunciou, mas que não foi verificada directamente pelo teste estatístico formal.

Quadro 7: Resultados da ANOVA com e sem violação do pressuposto da esfericidade (output do SPSS 12).

Tests of Within-Subjects Effects

Measure: MFASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
tempo	Sphericity Assumed	255,661	2	127,830	60,898	,000	,762
	Greenhouse-Geisser	255,661	1,316	194,240	60,898	,000	,762
	Huynh-Feldt	255,661	1,376	185,866	60,898	,000	,762
	Lower-bound	255,661	1,000	255,661	60,898	,000	,762
Error(tempo)	Sphericity Assumed	79,766	38	2,099			
	Greenhouse-Geisser	79,766	25,008	3,190			
	Huynh-Feldt	79,766	26,135	3,052			
	Lower-bound	79,766	19,000	4,198			

Uma atitude diversa das explicações anteriores é recorrer a um teste multivariado (MANOVA - Análise de Variância Multivariada) que não precisa de assumir esfericidade para se obterem resultados correctos. De facto, e tal como refere por exemplo (5), trata-se de um pressuposto “demasiado forte” para ser verificado numa qualquer condição em que se tenha medidas repetidas no tempo, sobretudo devido à natureza quasi-simplex⁵ de uma tal matriz de covariância. O teste multivariado é uma opção mais “sólida”. Contudo, tem várias alternativas. Uma das mais utilizadas é o Λ de Wilks=0.139, que tem uma distribuição amostral muito complexa. É usualmente aproximada para uma distribuição de χ^2 de Bartlett ou F de Rao, cujos resultados são $F(2,18)=34.981$, $p<0.001$. Tanto o teste univariado (ANOVA) como o multivariado (MANOVA) referem-se, exclusivamente, ao teste formal acerca da validade de H_0 . A ser verdadeira, a probabilidade de ocorrência de um valor tão extremo quanto o obtido na razão F é muito baixa ($p<0.001$) e, neste caso, rejeita-se a hipótese nula, tal como tínhamos referido para o caso do teste univariado com o ajustamento de G-G.

Quadro 8: Resultados da MANOVA (output do SPSS 12).

Multivariate Tests^b

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
tempo	Pillai's Trace	,795	34,981 ^a	2,000	18,000	,000	,795
	Wilks' Lambda	,205	34,981 ^a	2,000	18,000	,000	,795
	Hotelling's Trace	3,887	34,981 ^a	2,000	18,000	,000	,795
	Roy's Largest Root	3,887	34,981 ^a	2,000	18,000	,000	,795

a. Exact statistic

b.

Design: Intercept

Within Subjects Design: tempo

O problema a enfrentar agora é o seguinte: se as médias nos três pontos do tempo não são iguais, onde é que se situam as diferenças? Há várias alternativas para testes *post-hoc*⁶. Contudo, é importante salientar, ainda que se trate de um delineamento de natureza pré-experimental num único grupo, que o teste *a posteriori* utilizado deverá ser acompanhado de algum pensamento por parte do pesquisador. Dado que ele é a pessoa que conhece bem o proble-

ma e aquilo que está “em jogo”, deve ter uma atitude bem esclarecida sobre esta matéria. Só assim se compreende a sua opção e justificação. Ora, na ausência de uma qualquer explicação que suporte a escolha de um procedimento de *post-hoc* ou teste *a posteriori*, as sugestões que apresentaremos aqui representam pensamentos alternativos. A primeira, contrasta os três momentos entre si (t1-t2; t1-t3; t2-t3) e é efectuada com base no teste de Bonferroni⁷ e os resultados estão no Quadro 9.

Pairwise Comparisons

Measure: MEASURE_1

(I) tempo	(J) tempo	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
1	2	-2,815*	,423	,000	-3,925	-1,705
	3	-5,045*	,595	,000	-6,606	-3,484
2	1	2,815*	,423	,000	1,705	3,925
	3	-2,230*	,312	,000	-3,048	-1,412
3	1	5,045*	,595	,000	3,484	6,606
	2	2,230*	,312	,000	1,412	3,048

Quadro 9: Resultados do teste de post-hoc com ajustamento de Bonferroni (output do SPSS 12).

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Bonferroni.

É mais do que evidente dos resultados que se verificam diferenças significativas ($p < 0.001$) entre todos os pontos do tempo. De t1 para t2 verifica-se um incremento médio de 2.815 cm, de t1 para t3 de 5.045 cm, e de t2 para t3 de 2.230 cm.

Uma outra forma de analisar os resultados é realizar somente testes para valores adjacentes no tempo, t1-t2 e t2-t3. Os resultados estão no quadro seguinte e mostram-se todos significativos.

Quadro 10: Resultados de testes de post-hoc para valores médios adjacentes no tempo (output do SYSTAT 10)

Univariate F Tests

Source	SS	df	MS	F	P
1	158.484	1	158.484	44.292	0.000
Error	67.985	19	3.578		
2	99.458	1	99.458	51.181	0.000
Error	36.922	19	1.943		

Tendência

Um aspecto importante da análise de medidas repetidas refere-se à possibilidade de identificar a tendência do comportamento temporal da informação. Isto é, de verificar se em função do tempo o comportamento dos valores da força explosiva é de natureza linear ou não linear. Como só temos três pontos no tempo, a análise quantitativa da tendência não pode ultrapassar um polinómio de grau 2. O Quadro 11 refere, justamente, duas possibilidades: linear e quadrática. Ora, a que é estatisticamente significativa é a tendência linear [$F(1,19) = 71.968$, $p < 0.001$], conforme veremos na ilustração gráfica dos perfis individuais dos atletas.

Quadro 11: Teste à tendência dos resultados (output do SPSS 12).

Tests of Within-Subjects Contrasts							
Measure: MEASURE_1							
Source	tempo	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
tempo	Linear	254,520	1	254,520	71,968	,000	,791
	Quadratic	1,141	1	1,141	1,724	,205	,083
Error(tempo)	Linear	67,195	19	3,537			
	Quadratic	12,571	19	,662			

Interpretação alternativa

Já tivemos oportunidade de apresentar alguma argumentação para não nos atermos somente ao valor da prova que reflecte a necessidade exclusiva de uma decisão dicotómica. Bem pelo contrário, aquilo que se exige é largura de visão quando se explora um dado conjunto de dados, por forma a extrair dele toda a informação que contém. Daqui que se recorra, uma vez mais, aos intervalos de confiança e à noção de variabilidade explicada.

No caso dos intervalos de confiança chamamos a atenção do leitor para o Quadro 9. Nas duas últimas colunas temos resultados numéricos relativos a intervalos de confiança a 95% para as diferenças entre médias. Assim, o intervalo do momento 2 menos o momento 1 = 1.705 a 3.925; do momento 3 para o momento 2 = 1.412 a 3.048, e do momento 3 para o momento 1 = 3.484 a 6.606. Se centrarmos a nossa atenção exclusivamente neste último, verificamos que em termos populacionais a mudança esperada nos valores da força explosiva se situam, aproximadamente, entre 4 e 7 cm. Uma vez mais a questão a colocar situa-se ao nível da “qualidade” do intervalo destes ganhos, i.e., se reflectem aquilo que é esperado pelo treinador e se reflectem um ganho substantivo em termos de desempenho.

O eta quadrado parcial (η^2), é sobretudo uma sobrestimativa do actual valor da magnitude do efeito (SPSS Advanced Models, pag. 327). Refere-se à proporção de variabilidade total atribuível ao factor tempo. No caso do delineamento sobre as alterações da força explosiva, 79.1% (ver Quadro 11) da variabilidade total observada tem que ser imputada ao factor tempo, que não é mais do que o efeito do programa. Chamamos a atenção que esta interpretação não deve ser confundida com a que foi utilizada no teste t, por se tratar de conceitos e estatísticas diferentes.

Representações gráficas

O diagrama de extremos e quartis (*Box plot*) é uma representação gráfica extremamente interessante, dado salientar aspectos nucleares da uma qualquer distribuição de valores numéricos, sobretudo nos seus aspectos de medidas centrais e de dispersão, bem como a eventual presença de resultados que ultrapassam os valores adjacente inferior e superior, e que podem ser candidatos a *outliers*. Para além de salientar alguma assimetria à direita em t1 e t2, observa-se (a) ainda a presença de algumas observações discrepantes nos 3 momentos, (b) o incremento contínuo da variabilidade dos resultados nos 3 pontos, (c) bem como o aumento do valor da mediana.

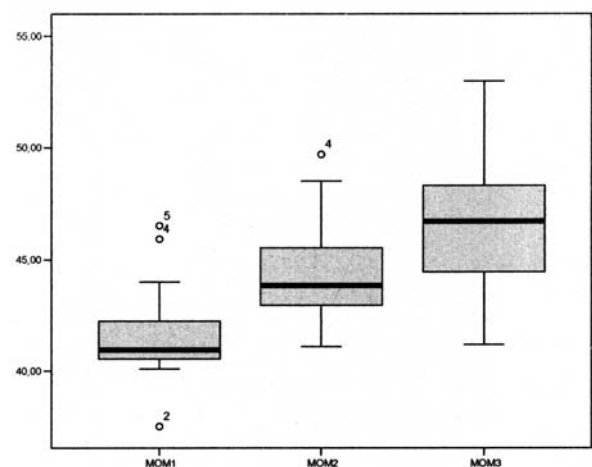


Figura 4: Diagrama de extremos e quartis.

Este diagrama de *high-low-close* (*high*=t3; *close*=t2; *low*=t1) representado na Figura 5, depois de ordenar os valores por t1, procura ilustrar vários pontos: (a) a heterogeneidade dos valores de partida; (b) as diferenças nos ganhos que nem sempre parecem estar condicionados aos valores iniciais; (c) a posição relativa dos ganhos em t2 (MOM2) relativamente a t1 (MOM1) e t3 (MOM3). É pois uma representação gráfica bem elucidativa do comportamento díspar na resposta ao treino, por exemplo, dos atletas 1, 8, 18 e 20.

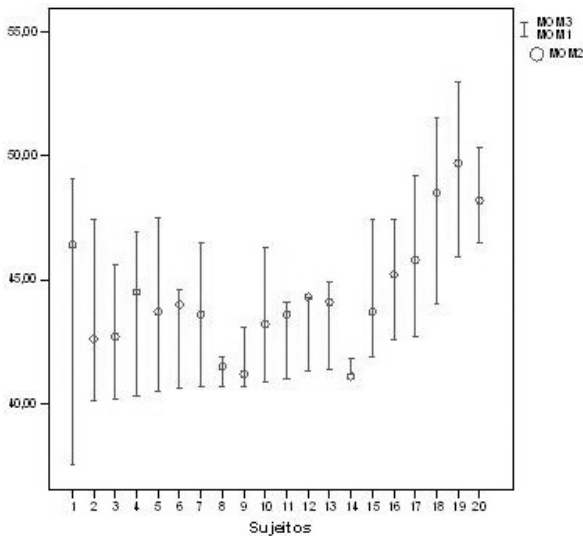


Figura 5: Diagrama de high-low-close.

Estabilidade ou tracking

A análise da estabilidade ou *tracking* das mudanças intraindividuais nas diferenças que ocorrem entre sujeitos requer agora um olhar bem mais amplo do que nos resultados do teste t.

O primeiro passo consiste, justamente, em observar o traçado das trajetórias dos sujeitos nos 3 pontos do tempo, o que se encontra ilustrado na figura seguinte.

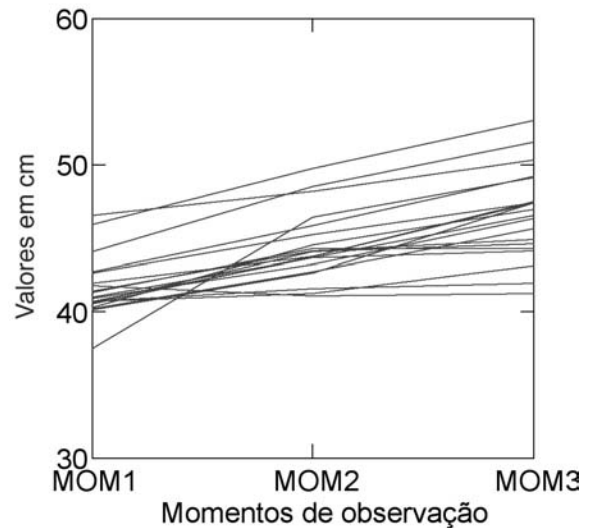


Figura 6: Spaghetti plot das trajetórias individuais.

Exige-se, agora, uma análise mais detalhada e ampla dos resultados que se encontram no quadro seguinte: - Na matriz de covariância os valores que mais nos interessam são os que se encontram na diagonal principal (variâncias dos resultados em cada ponto do tempo). Ora, é notório que a variância vai aumentando no tempo, produzindo um efeito de “abertura de leque” nas trajetórias dos sujeitos (confirmado pelos traçados da figura anterior). Isto significa que vai aumentando a heterogeneidade dos valores da impulsão vertical.

- Uma forma expedita de identificar o *tracking* é verificar o comportamento dos valores de correlação de Pearson entre os três momentos. É evidente que a correlação entre t1 e t3 é moderada ($r=0.516$). Contudo, acima do valor requerido para um traço ou característica evidenciar *tracking* ($r \geq 0.50$). É altamente estável o comportamento da força explosiva de t2 para t3 ($r=0.899$), ao contrário do que acontece de t1 para t2 ($r=0.638$). É provável que tal estabilidade esteja associada a efeitos de aprendizagem do próprio processo de avaliação e/ou resposta neuromuscular ao treino.

- A estabilidade global ou *tracking* é dada pelo valor do coeficiente de correlação intraclassa que é, neste momento, de 0.859 (IC95%=0.702; 0.940), sugerindo a manutenção moderada a elevada da posição relativa de cada sujeito no seio do grupo.

Quadro 12: Resultados (do SPSS e SYSTAT) relativos à estabilidade ou tracking.

Inter-Item Covariance Matrix

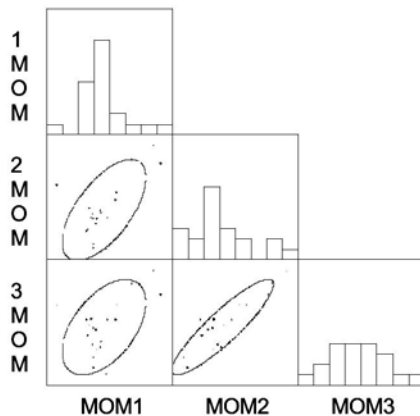
	MOM1	MOM2	MOM3
MOM1	4,153	3,066	3,211
MOM2	3,066	5,556	6,478
MOM3	3,211	6,478	9,343

The covariance matrix is calculated and used in the analysis.

Inter-Item Correlation Matrix

	MOM1	MOM2	MOM3
MOM1	1,000	,638	,516
MOM2	,638	1,000	,899
MOM3	,516	,899	1,000

The covariance matrix is calculated and used in the analysis.



Intraclass Correlation Coefficient

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single measures	,669	,440	,839	7,076	19,0	38	,000
Average measures	,859 ^c	,702	,940	7,076	19,0	38	,000

Two-way mixed effects model where people effects are random and measures effects are fixed.

^a Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

^b The estimator is the same, whether the interaction effect is present or not.

^c This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Modelação hierárquica ou multinível (MHMN)

Não pretendemos terminar este ensaio breve sem sugerir que, com base no pensamento, metodologia e estratégia altamente versáteis e flexíveis da MHMN⁸ é possível testar, para o mesmo conjunto de dados longitudinais, um corpo variado de hipóteses-modelos hierarquicamente embricados (*nested within*), por forma a identificar aquele que se afigura mais plausível. Esta versatilidade na formulação de modelos *nested within* não é possível de realizar com base na ANOVA ou MANOVA. Formularemos quatro modelos⁹ para descrever¹⁰ o comportamento da força explosiva dos membros inferiores.

1º É possível que a variância observada nos valores de partida dos sujeitos (em t1) seja irrelevante, sem qualquer significância estatística. Trata-se de variabilidade biológica sem qualquer significado. Do mesmo modo assumimos que não se verifica qualquer divergência substancial nas trajectórias dos atletas no que respeita aos seus ganhos de força. Neste caso uma única equação de regressão para todo o grupo seria a explicação mais plausível.

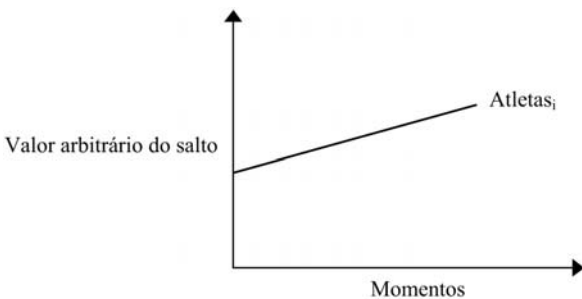


Figura 7: Comportamento dos valores dos atletas em função do tempo para o 1º modelo.

2º Uma outra forma de olhar para o comportamento dos dados seria pensar que ainda que não haja variação significativa no momento inicial (um ponto único serve para descrever os resultados dos atletas em t1), é bem possível que as suas trajectórias sejam diferentes, i.e., que haja variância significativa nos declives das trajectórias dos diferentes atletas, resultado inequívoco das diferenças na sua resposta ao treino e que pode ser de natureza biológica.

Graficamente representaríamos tal possibilidade como na figura 8.

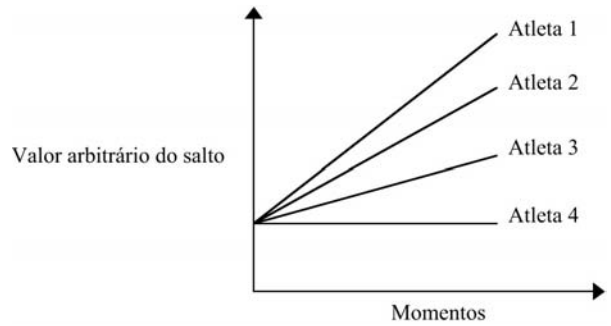


Figura 8: Comportamento dos valores dos atletas em função do tempo para o 2º modelo.

3º É possível pensar, ainda, num formato bem mais flexível para o comportamento das trajectórias inter-individuais. Não só os valores de partida dos sujeitos são diferentes entre si, i.e., há variância significativa em t1 (hipótese contrária às anteriores), como se constata que não existe qualquer diferença nas trajectórias dos sujeitos (a sua resposta ao treino é a mesma, independentemente do seu valor de partida). Trata-se, pois, de especificar rectas paralelas para todos os atletas. Os ganhos de força explosiva no tempo são descritos pelo mesmo declive, aquilo que se designa por estabilidade paralela e que pode representar-se assim (ver Figura 9).

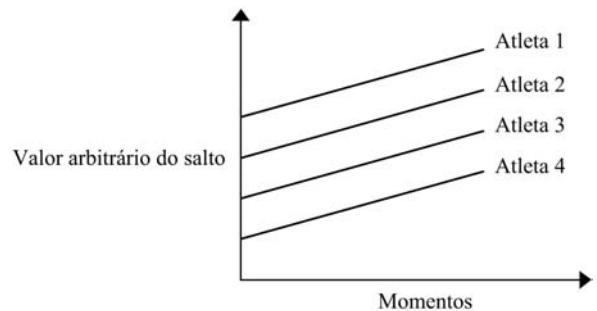


Figura 9: Comportamento dos valores dos atletas em função do tempo para o 3º modelo.

4º Finalmente, o modelo mais flexível, do qual todos os outros não são mais de que aspectos parcelares, *nested within*, é o que sugere que não só existe uma variância substancial no momento inicial do estudo (os valores de força explosiva são o resultado da variabilidade biológica e do carácter “residual” da magnitude do destreino¹¹), como também se verifica uma forte heterogeneidade na resposta ao treino e que é visível no comportamento diferenciado dos valores de força explosiva. Portanto, teríamos trajetórias interindividuais bem diferentes (Figura 10).



Figura 10: Comportamento dos valores dos atletas em função do tempo para o 4º modelo.

A forma mais adequada de testar a plausibilidade destes 4 modelos é pensar em termos da sua expansão no que respeita ao número de parâmetros a estimar. O primeiro modelo é o mais simples, e os outros mais complexos. Uma estatística adequada para determinar a qualidade de ajustamento dos dados a uma dada hipótese é a razão do logaritmo da verosimilhança (RV), do inglês *log likelihood ratio*. Com base no valor desta razão é possível testar a plausibilidade interpretativa de dois modelos, contrastando o mais simples com um outro mais complexo e multiplicando o seu valor por -2, tal que a estatística $G = -2 * (RV_{\text{modelo mais restritivo}} - RV_{\text{modelo mais expansivo}})$, e esperando que o resultado, que tem uma distribuição de χ^2 com número de graus de liberdade igual à diferença dos parâmetros em cada modelo seja significativo (i.e., $p < 0.05$). Os resultados obtidos nos 4 modelos foram os seguintes (Quadro 13).

Quadro 13: Valores da razão do logaritmo de verosimilhança para os 4 modelos testados.

Modelos	RV (Log da função de Verosimilhança)	G
Modelo1	-139.149	
Modelo2	-129.479	19.340 [M ₁ -M ₂]
Modelo3	-125.694	-
Modelo4	-117.274	24.41 [M ₂ -M ₄]; 16.84 [M ₃ -M ₄]

Importa salientar que os modelos 2 e 3 não são *nested within*, mas sim competitivos ou alternativos, dado terem o mesmo número de parâmetros. Qualquer deles seria uma hipótese viável dado terem uma RV praticamente idêntica. Do contraste dos vários modelos e da qualidade do seu ajustamento, fácil se torna verificar que o modelo mais plausível é pois o que postula que não só os valores de partida de força explosiva são diferentes, como também a trajetória dos ganhos evidencia uma variabilidade substancial. Interpretemos, agora, as partes mais importantes do *output* (*Mixed Regression* executada no SYSTAT 10).

Quadro 14: Resultados mais importantes do output da MHMN.

Perform 100 EM iterations
2 random terms
0 fixed terms

Numbers of observations

Level 2 observations = 20
Level 1 observations = 60

* Final Results - Maximum Marginal Likelihood (MML) Estimates *

Empirical Bayes estimates:

ID	Observations	INTERCEPT	MOMENTO
1	3	44.104	3.717
2	3	39.555	4.937
3	3	41.635	2.701
4	3	45.558	3.689
5	3	45.652	2.386
6	3	40.902	2.754
7	3	40.924	2.597
8	3	40.191	3.286
9	3	40.887	3.071
10	3	40.751	3.217
11	3	42.640	3.188
12	3	41.599	1.704
13	3	42.504	2.501
14	3	40.552	0.895
15	3	40.400	1.358
16	3	41.221	1.710
17	3	41.058	2.053
18	3	41.572	1.901
19	3	41.165	0.235
20	3	40.380	2.548

EM Iterations = 100
Fisher Iterations = 2
Total Iterations = 102
Log Likelihood = -117.274

Variable	Estimate	Standardized		p-value
		Error	Z	
INTERCEPT	41.663	0.428	97.398	0.000
MOMENTO	2.523	0.290	8.704	0.000

Residual variance:

0.686 0.217 3.162 0.001

Random-effect variance & covariance term(s):

Estimate	1	2
	INTERCEPT	MOMENTO
1 INTERCEPT	3.088	
2 MOMENTO	0.025	1.337

Standardized Error	1	2
	INTERCEPT	MOMENTO
1 INTERCEPT	1.171	
2 MOMENTO	0.569	0.542

Z	1	2
	INTERCEPT	MOMENTO
1 INTERCEPT	2.637	
2 MOMENTO	0.044	2.466

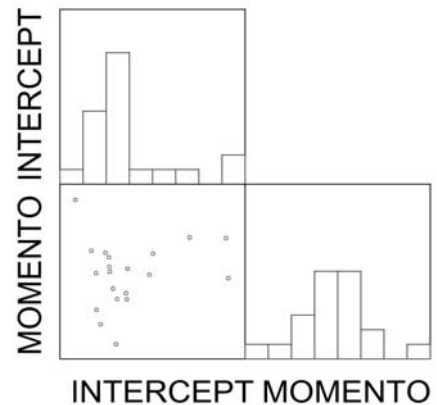
p-value	1	2
	INTERCEPT	MOMENTO
1 INTERCEPT	0.004	
2 MOMENTO	0.965	0.007

Note: p-values are 2-tailed except for those associated with variances, which are 1-tailed.

Random-effect covariances expressed as correlations	1	2
	INTERCEPT	MOMENTO
1 INTERCEPT	1.000	
2 MOMENTO	0.012	1.000

Correlation of the MML estimates of the fixed terms	1	2
	INTERCEPT	MOMENTO
1 INTERCEPT	1.000	
2 MOMENTO	-0.128	1.000

Empirical Bayes Estimate:



- É evidente a presença de diferenças interindividuais nos valores de partida (ordenada), bem como nas suas trajectórias de modificação da capacidade de salto (declive), tal como é referido nas estimativas individuais apresentadas (Empirical Bayes Estimates).
- O valor na ordenada (em t1) é de 41.663 cm (erro-padrão=0.428), com um resultado da estatística $z=97.398$, $p<0.001$. O resultado numérico 41.663 é uma estimativa do ponto médio da performance do grupo no início do estudo, dado que t1 foi codificado com o valor 0. Com base neste resultado seria muito fácil construir um intervalo de confiança com 95% para aquele valor ($41.663 \pm 1.96 \times 0.428$).
- O resultado numérico do declive é de 2.523 (erro-padrão=0.290), $z=8.704$, $p<0.001$. Trata-se de um valor que confirma a presença de uma trajectória ascendente nos valores de força explosiva dos atletas. Também aqui seria fácil construir um intervalo de confiança com 95% para o declive ($2.523 \pm 1.96 \times 0.290$).
- A verificação da forte heterogeneidade nos valores de partida dos atletas é dada pela estimativa de variância no valor da ordenada e que é de 3.088 ± 1.171 cm ($z=2.637$, $p=0.004$); o mesmo ocorre para os valores dos declives das trajectórias interindividuais, e que é de 1.337 ± 0.542 ($z=2.466$, $p=0.007$).
- Dado que se verifica uma correlação irrelevante ($r=0.012$) e não significativa ($p>0.05$) entre os valores na ordenada e os declives (ver diagrama de dispersão entre *intercept* e *trial*), estamos diante da circunstância de não haver qualquer relação entre os valores de partida e os ganhos de força.
- É importante salientar que há variância residual igual a 0.686 (significativa, $p=0.001$) que exige algum esforço suplementar de modelação e que conduziria, necessariamente, para preditores dos sujeitos que originaram as suas diferenças nas trajectórias de desenvolvimento da força explosiva.
- Em resumo, estamos diante de um conjunto de ideias e resultados bem diferentes e porventura mais interessantes e elucidativos do que o simples teste formal à diferença de médias.

Em conclusão, esperamos que este pequeno exercício e viagem guiada a um conjunto simples de dados tenha sido esclarecedor ao ponto de alertar o leitor

para uma exploração mais atenta e a uma interpretação mais esclarecida da sua informação.

É importante ter sempre presente a noção de que há que “conversar” com a informação disponível e extrair dela não só sumários relevantes de natureza diversa, como também representações gráficas interessantes e auto-explanatórias do comportamento dos dados.

NOTAS

- ¹ Chamamos a atenção do leitor para dois textos magistrais no domínio da associação estreita entre argumentação, delineamento de pesquisa e análise de dados (1, 18).
- ² Uma leitura atenta e cuidada dos textos “A primer in longitudinal data analysis” (23), “Longitudinal data analysis: designs, models and methods” (2), “Modeling longitudinal and multilevel data” (12), “New methods for the analysis of change” (4) será da maior utilidade para o leitor mais exigente e desejoso de mergulhar profundamente neste domínio fascinante.
- ³ Na sua essência, o programa era constituído por exercícios de pliometria de profundidade.
- ⁴ Sobre a matéria da opção pelas correcções de Greenhouse-Geisser ou Huynh-Feldt sugerimos, por exemplo, o estudo dos textos (6, 7, 23). Uma leitura bem esclarecedora e extremamente didáctica desta matéria pode ser efectuada no texto (26), “Análisis de varianza com medidas repetidas”. Uma aplicação ao domínio das Ciências do Desporto pode ser consultada no capítulo 10 do manual “Statistics in Kinesiology” (25).
- ⁵ Sobre este assunto consultar (10). Uma aplicação desta noção ao estudo da estabilidade dos valores da actividade física é encontrada em (16).
- ⁶ Sobre esta matéria consultar, por exemplo, o manual do SPSS 12 (23).
- ⁷ Para evitar a inflação do erro do tipo 1 pelas múltiplas comparações a efectuar, o alfa nominal (global) de 5% é dividido por três. Nesta situação o alfa de cada comparação individual passa para 0.017 (23) para explorar alternativas no uso dos testes de *post-hoc*.
- ⁸ Este é um brevíssimo apontamento relativo ao uso da MHMN cujas potencialidades na análise de dados longitudinais são extremamente elevadas e muito ricas de pensamento alternativo e mais esclarecedor daquilo que a informação contém. Uma estratégia alternativa à da MHMN é a da Modelação de Estruturas de Covariância (13).
- ⁹ Para se identificar com aspectos básicos da MHMN consultar, em língua portuguesa (16), ou os textos nucleares (9, 19).
- ¹⁰ Neste exemplo só se possui informação sobre uma variável de “saída” - a impulsão vertical. É evidente que se estivessemos na posse de outra informação dos sujeitos e do “ambiente” que variasse, ou não, em função do tempo, poderíamos modelar este conjunto multivariado e encontrar a importância diversa dos factores que explicam as mudanças ocorridas no tempo.
- ¹¹ O programa de treino teve lugar logo no início da época desportiva.

CORRESPONDÊNCIA

José António Ribeiro Maia

Laboratório de Cineantropometria
e Análise de Dados
Faculdade de Ciências do Desporto
e de Educação Física
Universidade do Porto
Rua Dr. Plácido Costa, 91
4200-450 Porto
Portugal
jmaia@fedef.up.pt

BIBLIOGRAFIA

1. Abelson RP (1995). *Statistics as a principled argument*. Mahawah: Lawrence Erlbaum Associates, Publishers.
2. Bijleveld CCJH, van der Kamp LJT, Mooijaart A, van der Kloot WA, van der Leeden R, van der Burg E (1998). *Longitudinal data analysis. Designs, models and methods*. Thousand Oaks: Sage Publications.
3. Campbell DT, Stanley JC (1963). *Experimental and quasi-experimental designs for research*. Skokie: Rand McNally.
4. Collins, L.M, Sayer, A.G. (Eds). (2001). *New methods for the analysis of change*. Washington: American Psychological Association.
5. Cuttance P, Ecob R (1987). *Structural equation modeling by example. Applications in educational, sociological, and behavioral research*. Cambridge: Cambridge University Press.
6. Everitt BS (1996). *Making sense of statistics in psychology. A second-level course*. Oxford: Oxford University Press.
7. Field, A (2000). *Discovering statistics using SPSS for windows. Advanced techniques for the beginner*. Thousand Oaks: Sage Publications.
8. Finkel SE (1995). *Causal analysis with panel data*. Sage University paper series on Quantitative Applications in the Social Sciences. Thousand Oaks: Sage Publications, 7-105.
9. Hox J (2002). *Multilevel analysis. Techniques and applications*. Mahwah : Lawrence Erlbaum Associates, Publishers.
10. Joreskog KG (1970). Estimating and testing of simplex models. *British Journal of Mathematical and Statistical Psychology* 23:121-145.
11. Joreskog KG, Sorbom D (1976). Statistical models and methods for analysis of longitudinal data. In DJ Aigner, AS Goldberger (Eds.) *Latent Variables in Socioeconomic Models*. Amsterdam: North Holland, 285-325.
12. Little, TD, Schnabel KU, Baumert J (2000). *Modeling longitudinal and multilevel data. Practical issues, applied approaches, and specific examples*. Mahwah: Lawrence Erlbaum Associates, Publishers.
13. Maia JAR, Beunen G, Lafevre J, Claessens A (1998). Testes empíricos a formulações desenvolvimentalistas: um estudo centrado na modelação da estrutura de covariância. *Revista Paulista de Educação Física* 12(2):160-180.
14. Maia JAR, Beunen G, Lafevre J, Claessens A, Renson R, Vanreusel B (2003). Modeling stability and change in strength development: a study in adolescent boys. *American Journal of Human Biology* 4:579-591.
15. Maia JAR, Lopes VP, da Silva RG, Seabra A (2002). A importância do estudo do *tracking* (estabilidade e previsão) em delineamentos longitudinais: um estudo aplicado à epidemiologia da actividade física e à performance desportivo-motora. *Revista Portuguesa de Ciências do Desporto* 2(4):41-56.
16. Maia JAR, Lopes VP, da Silva RG, Seabra A, Ferreira JV, Cardoso MV (2003). Modelação hierárquica ou multinível. Uma metodologia estatística e um instrumento útil de pensamento na investigação em Ciências do Desporto. *Revista Portuguesa de Ciências do Desporto*. 3 (1):92-107.
17. Marsh HW, Grayson D (1994). Longitudinal factor analysis: common, time-specific, item-specific and residual-error components of variance. *Structural Equation Modeling* 1:116-145.
18. Maxwell SE, Delaney HD (2000). *Designing experiments and analyzing data. A model comparison perspective*. Mahawah:

- Lawrence Erlbaum Associates, Publishers.
19. Raudenbush SW, Bryk AS (2002). *Hierarchical linear models. Applications and data analysis methods*. 2nd edition. Thousand Oaks: Sage Publications.
 20. Silva PJMM (2003). *Estudo da dinâmica da performance em andebolistas júniores*. Tese de Mestrado. Faculdade de Ciências do Desporto e de Educação Física, Universidade do Porto. Portugal.
 21. Spector PE (1981). *Research Designs*. Sage University Paper series on Quantitative Applications in the Social Sciences, series nº 35. Berverly Hills: Sage Publications.
 22. SPSS Advanced Models. Chicago: SPSS Inc.
 23. Stevens JP (2002). *Applied multivariate statistics for the social sciences*. 4th edition. Mahawah: Lawrence Erlbaum Associates, Publishers.
 24. Taris TW (2000). *A primer in longitudinal data analysis*. Thousand Oaks. Sage Publications.
 25. Vincent WJ (1999) *Statistics in Kinesiology*. 2nd edition. Champaign: Human Kinetics.
 26. Ximénez C, Martín RS (2000). *Análisis de varianza com medidas repetidas*. Madrid: Editorial la Muralla S.A.