

# ESTRATIFICAÇÃO DA APTIDÃO AERÓBIA A PARTIR DA VARIABILIDADE DA FREQUÊNCIA DE REPOUSO PELA ANÁLISE DE COMPONENTES PRINCIPAIS E AGRUPAMENTO POR K-MEANS

W. Materko, R. Bartels, A.R.S. Carvalho, J. Nadal

Programa de Engenharia Biomédica (PEB)/COPPE/UFRJ, Rio de Janeiro, Brasil

e-mail:wollner.materko@gmail.com

**Resumo:** O objetivo do presente estudo foi propor um método para estratificação do grau de aptidão aeróbica a partir da variabilidade da frequência cardíaca em um grupo de indivíduos saudáveis com características físicas e antropométricas similares. Vinte voluntários praticantes de corrida em esteira participaram do presente estudo. Após coleta do tacograma com cardiografocímetero durante 10 min em repouso, um teste de esforço com incremento progressivo de carga foi realizado e o consumo máximo de oxigênio ( $\dot{V}O_{2\max}$ ) determinado. Os espectros de potência dos tacogramas de repouso foram estimados por periodogramas de Welch, após interpolação por splines cúbicas e reamostragem a 4 Hz. A análise de componentes principais (ACP) foi aplicada aos espectros de potência, seguida pelo método de agrupamento K-means, obtendo-se dois grupos, com 9 e 11 sujeitos. De acordo com o teste *t* de Student, o agrupamento com nove sujeitos apresentou valores de  $\dot{V}O_{2\max}$  significativamente maiores que o outro ( $48,3 \pm 5,1$  vs  $33,5 \pm 3,2$  ml.kg<sup>-1</sup>.min<sup>-1</sup>, respectivamente;  $p = 0,01$ ). Conclui-se que o método proposto parece ser capaz de estratificar o nível de aptidão aeróbica em voluntários saudáveis em repouso.

**Palavras-chave:** variabilidade da frequência cardíaca, análise de agrupamentos, análise de componentes principais.

**Abstract:** The aim of this study was stratifying the degree of aerobic fitness from the heart rate variability (HRV) in subjects with similar physical and anthropometric characteristics. Twenty treadmill runners were volunteers to the present study. After recording the resting tachogram with a cardiografocymeter for 10 min, a maximal cardiopulmonary incremental test were performed to measure the maximum oxygen uptake ( $\dot{V}O_{2\max}$ ). The power spectra of resting tachograms were estimated by Welch periodograms after cubic spline interpolation and resampling at 4 Hz. Principal component analysis (PCA) was applied to the power spectra, followed by clustering method K-means, obtaining two groups, with 9 and 11 subjects. According to the Student *t* test, the cluster with nine subjects presented  $\dot{V}O_{2\max}$  values significantly higher than the other ( $49.4 \pm 4.4$  vs

$37.4 \pm 3.2$  ml.kg<sup>-1</sup>.min<sup>-1</sup>, respectively;  $p = 0.01$ ). As a conclusion, the proposed method appears to be capable for stratifying the degree of aerobic fitness in resting healthy volunteers.

**Keywords:** heart rate variability, K-means clustering, principal component analysis.

## Introdução

O coração tem suas funções moduladas pelo sistema nervoso autônomo, que promove respostas adaptativas às demandas periféricas pelos ramos simpático e parassimpático, com aumento e diminuição da frequência cardíaca (FC), respectivamente [1,2]. A variabilidade da FC (VFC) tem sido estimada através da análise das variações dos intervalos entre ondas R consecutivas do eletrocardiograma. A diminuição da VFC está relacionada com um maior índice de morbidade e mortalidade cardiovascular [2], enquanto o aumento da VFC está relacionado a uma boa condição de saúde [3].

A estratificação da aptidão aeróbica é uma prática comum e apropriada nos programas de exercício físico, sendo estimada pelo teste cardiopulmonar máximo ( $\dot{V}O_{2\max}$ ). Entretanto, o alto custo e a inacessibilidade a tais equipamentos acabam por limitar sua aplicação, conduzindo à busca por técnicas menos sofisticadas e de menor custo para a estimativa da aptidão aeróbica. Embora alguns métodos venham sendo propostos e validados para amostras muito distintas, tais como sujeitos normais vs sujeitos com doenças cardiovasculares [1] ou sujeitos sedentários vs atletas de alto nível [4], os mesmos falham ao comparar indivíduos com características físicas e antropométricas similares.

Métodos de análise multivariada são essenciais para o processamento de grandes massas de dados. Entre eles, o método de análise de agrupamentos tem como objetivo organizar os dados em diferentes grupos a partir das suas características. Um dos métodos de agrupamento mais populares e eficiente é o *K-means*, que forma grupos a partir da distância a centróides, otimizando a função do erro quadrático [5]. Por outro lado, quando o objetivo é compactação de dados a análise de componentes principais (ACP) é usada para projetar dados para um subespaço de menor dimensão buscando preservar ao máximo a variância original dos

dados [6]. A ACP promove grande compactação dos dados, além da filtragem de ruídos com flutuações aleatórias.

Neste sentido, a análise de agrupamentos pode ser facilitada pelo pré-processamento com a ACP. Sendo assim, o objetivo do presente estudo foi propor um método para estratificação do grau de aptidão aeróbica a partir da VFC em um grupo de indivíduos saudáveis com características físicas e antropométricas similares. Para isso, utilizou-se um algoritmo de agrupamento *K-means*, após a utilização da ACP para compactação dos dados da análise espectral do tacograma de repouso, com intuito de separar os indivíduos da amostra em dois grupos de acordo com o nível de aptidão aeróbica.

## Materiais e métodos

### Voluntários

Participaram deste estudo 20 voluntários do sexo masculino de baixo risco, praticantes de corrida em esteira. Considerou-se como critérios de inclusão: não utilizar qualquer recurso ergogênico ou medicamento que atue no sistema cardiovascular; não apresentar histórico de doenças pulmonares ou cardiovasculares; e não apresentar lesão ostomioarticular prévia.

Os voluntários foram orientados para não ingerir bebida alcoólica na noite anterior a coleta dos dados, ter uma boa noite de sono, evitar qualquer fonte de estresse, não ingerir chá ou café e não praticar exercício físico, a fim de não interferir nos resultados.

A coleta e a análise de dados foram realizadas no Laboratório de Avaliação Física do Centro de Atenção à Saúde (HU/CAS) da Universidade Federal de Juiz de Fora. Os experimentos ocorreram no período entre 15 h e 17 h, a temperatura não excedeu 22° C e a umidade relativa manteve-se no máximo de 70%.

Os procedimentos experimentais tiveram início após o consentimento verbal e a assinatura do Termo de Consentimento Livre e Esclarecido, conforme o protocolo experimental aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Juiz de Fora (UFJF) com o parecer número 1230.276.2007.

### Avaliação Antropométrica

Esta consistiu das medidas de massa corporal e estatura, realizadas numa balança mecânica com estadiômetro (Filizola, Brasil), e tomada das medidas das sete dobras cutâneas, através de um compasso científico (Cescorf, Brasil). A partir destas medidas, estimou-se o percentual de gordura e a massa livre de gordura usando as equações de Jackson e Pollock [7] para a estimativa da densidade corporal em homens, respectivamente, combinada com a equação de Siri [8].

### Teste de incremento progressivo de carga

Realizou-se o teste cardiopulmonar máximo (TCPM) em cicloergômetro Ergo 167 (ERGOFIT, Alemanha), seguindo um protocolo incremental com potência inicial de 25 W e incrementos de 25 W/min, com manutenção de 50 a 60 rpm. A coleta e análise de

gases expirados foi realizada continuamente, pelo analisador metabólico VO<sub>2000</sub> (MedGraphics, EUA) calibrado no modo automático antes de cada teste. O consumo de oxigênio e demais variáveis foram amostradas em intervalos de 20 s. A partir do TCPM, identificou-se o  $\dot{V}O_{2\max}$  de cada voluntário.

### Protocolo de aquisição

Os voluntários foram posicionados na posição supina, sendo colocado no tórax, próximo ao processo xifóide do voluntário, uma cinta de transmissão para obtenção dos dados, e o cardiofrequencímetro modelo RS810i (Polar, Finlândia) no pulso esquerdo, que fez aquisição dos iRR de repouso durante o período de 10 min.

Obtido os tacogramas, os dados foram transmitidos para o software Polar Precision Performance SW (Polar, Finlândia), através do sensor infravermelho, onde foram filtrados (aceitos dados com correção de até 5% do sinal), e depois salvos em arquivo “.txt”. Todo o processamento de sinais foi realizado com programas escritos em Matlab versão 6.5 (The MathWorks, EUA).

### Processamento de Sinais

Para a análise no domínio da frequência do tacograma de repouso foi utilizado o software SinusCor (desenvolvido no Laboratório de Engenharia Pulmonar, PEB/COPPE/UFRJ). Os tacogramas do iRR foram reamostrados em intervalos constantes, com frequência de 4 Hz, pelo método de interpolação *spline cúbica*. Após a remoção da tendência linear, foi realizada a análise espectral das séries, utilizando-se o método do periodograma de Welch (transformada rápida de Fourier, 256 pontos, 50% de sobreposição e janela Hanning).

A matriz formada pelo conjunto de espectros de potências foi então utilizada para o cálculo da matriz de covariâncias, à qual foi aplicada a ACP [6]. A seguir, o método de agrupamento por *K-means*, fixado para o agrupamento de dois clusters, foi aplicado às componentes principais de acordo com um algoritmo proposto por Napoleon e Pavalakodi [9].

O centróide de cada cluster é definido a partir de:

$$m_k = \frac{1}{n_k} \sum_{i \in C_k} X_i \quad (1)$$

onde  $X_i$  é o vetor com os coeficientes das componentes principais, e  $n_k$  é o número de pontos em  $C_k$ .

Com isto o algoritmo busca minimizar o somatório do erro quadrático dado por:

$$J_k = \sum_{k=1}^k \sum_{i \in C_k} (x_i - m_k)^2 \quad (2)$$

Após a fixação do centro de agrupamento e alocação dos pontos no agrupamento mais próximo, o método de agrupamento por *K-means* recalcula os centros dos *clusters*, como a média dos pontos que ele representa, e

repete o procedimento de realocação até que os centroides não mais se movam.

Os dados são expressos como média  $\pm$  desvio padrão (SD) e coeficiente de variação (SD/média). Após a implementação do algoritmo de agrupamento e a identificação de dois grupos, os  $\dot{V}O_{2\max}$  dos indivíduos de cada grupo foram comparados com o teste *t de Student* com  $\alpha = 0,05$ . Todas as análises foram realizadas no Matlab versão 6.5 (The MathWorks, EUA).

## Resultados

Após a ACP, foram utilizados apenas dois componentes, os quais representam 86% da variância total das variáveis originais.

A partir da análise de dispersão dos escores dos dois componentes, o método *K-means* efetuou o agrupamento em dois clusters utilizando apenas o segundo componente (Figura 1): grupo A com 9 voluntários e grupo B com 11 voluntários.

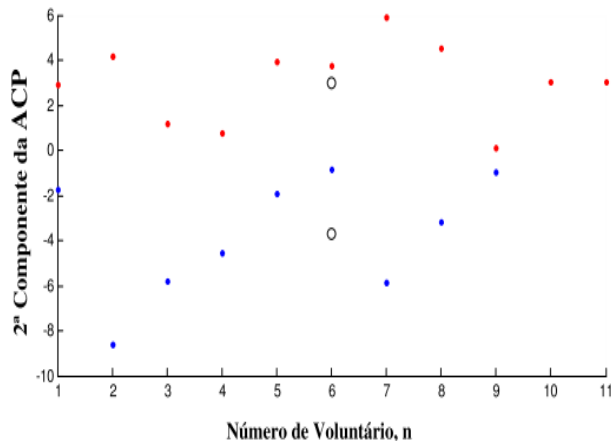


Figura 1. Resultado da análise de *K-means* via ACP: grupo A representado por pontos em vermelho, grupo B por pontos em azul. Os círculos correspondem aos centroides.

As características antropométricas e físicas dos 20 voluntários foram muito similares (Tabela 1), conforme os baixos valores de desvio padrão e coeficiente de variação de cada variável. Com isso, as características físicas e antropométricas do grupo A foram comparadas àquelas do grupo B, não se registrando diferenças significativas entre os mesmos (Tabela 2).

Tabela 1. Características antropométricas e físicas dos voluntários

Variáveis	Média $\pm$ SD	CV (%)
Idade (anos)	23,9 $\pm$ 4,2	17
Massa corporal (kg)	73,4 $\pm$ 9,5	10
Estatura (cm)	170,0 $\pm$ 10,0	5
Gordura relativa (%)	11,8 $\pm$ 2,4	20

SD = desvio padrão; CV = coeficiente de variação

Tabela 2. Características antropométricas e físicas dos voluntários para cada cluster

Variáveis	Grupo A	Grupo B
Idade (anos)	25,8 $\pm$ 4,2	26,4 $\pm$ 8,1
Massa corporal (kg)	79,0 $\pm$ 9,5	73,4 $\pm$ 9,5
Estatura (cm)	172,9 $\pm$ 5,7	175,0 $\pm$ 5,8
Gordura relativa (%)	12,4 $\pm$ 2,4	11,3 $\pm$ 2,4

Na análise comparativa do  $\dot{V}O_{2\max}$  dos dois agrupamentos, os indivíduos do grupo B ( $49,4 \pm 4,4 \text{ ml.kg}^{-1}.\text{min}^{-1}$ ) apresentaram valores significativamente maiores ( $p = 0,01$ ) que o grupo A ( $37,4 \pm 3,2 \text{ ml.kg}^{-1}.\text{min}^{-1}$ ). Tendo em vista este resultado, o método *K-means* foi reutilizado para o agrupamento dos sujeitos em função do segundo componente e do  $\dot{V}O_{2\max}$ , apenas para fins de ilustração (Figura 2).

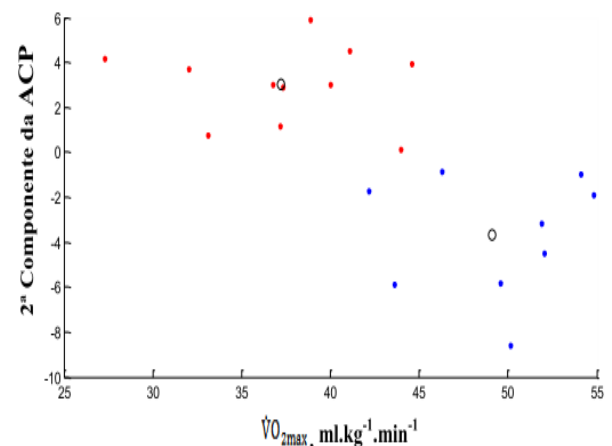


Figura 2. Resultado da análise de *K-means* via ACP: grupo A representado por pontos em vermelho, grupo B por pontos em azul. Os círculos correspondem aos centroides.

## Discussão

A ACP é um método análise multivariada de dados de que tem por finalidade básica a sua redução, com eliminação de variáveis correlacionadas e a escolha das formas mais representativas de dados a partir de combinações lineares das variáveis originais [4,6]. O *K-means*, por sua vez, é um método de análise de agrupamentos que organiza os dados em grupos, minimizando a variabilidade intragrupo e maximizando a variabilidade intergrupo [5]. A junção dessas duas ferramentas de processamento é potencialmente útil por facilitar a organização dos grupos pela segunda, através da redução de complexidade efetuada pela primeira. No presente caso, ao invés de analisar os agrupamentos utilizando todos os componentes espectrais, passou-se a analisá-los apenas no espaço dos dois componentes principais.

Embora o estudo tenha sido limitado à análise de apenas dois componentes principais, posteriormente

reduzidos a um pelo método de agrupamento, os grupos criados apresentaram diferenças significativas em termos de aptidão aeróbica.

Com vistas a identificar quais características espectrais dos tacogramas contribuíram para a diferenciação dos grupos, pode-se examinar o autovetor correspondente ao segundo componente principal. A distribuição dos fatores de carga desse componente com relação ao espectro original (Figura 3) é indicativa da contribuição relativa de cada componente de frequência [6] para este componente principal. Este autovetor apresenta maiores pesos na banda entre 0,09 e 0,22 Hz, os quais devem ter contribuído mais para a discriminação em termos de aptidão aeróbica. Esta banda combina efeitos das ações dos dois ramos do sistema nervoso autônomo, em particular porque o teste foi realizado com respiração espontânea, mas reflete predominantemente a modulação simpática [2]. Como os sujeitos com maior aptidão aeróbica têm *scores* negativos para esse componente, seus espectros situam-se abaixo da média do grupo amostral nessa banda de energias, estando sujeitos a um menor tônus simpático.

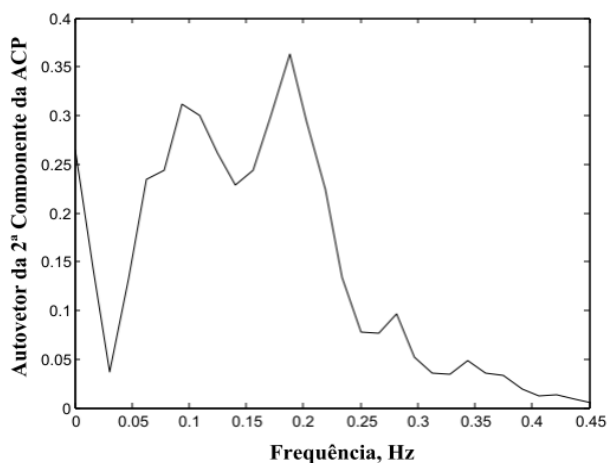


Figura 3. Autovetor do segundo componente principal obtido por ACP a partir dos espectros de potência do grupo amostral.

O resultado do presente trabalho é promissor pelo fato de que a amostra analisada é muito homogênea, dada por jovens praticantes de atividade física regular, com características antropométricas e físicas similares.

Novos estudos, envolvendo um número maior de sujeitos e amostras de sujeitos com outros níveis de condicionamento físico são necessários para a validação do método utilizado. Uma vez validado, este método constituiria uma ferramenta de utilidade prática para avaliação da aptidão aeróbica, ao combinar as vantagens de baixo custo e facilidade de aquisição das medidas, em particular pela aplicabilidade mesmo em uma amostra de características físicas e antropométricas homogêneas.

### Conclusão

O uso da análise de componentes principais aplicada aos espectros de potência dos tacogramas, associado ao

método de agrupamento *k-means* parece ser capaz de estratificar o nível de aptidão aeróbica em voluntários saudáveis em repouso, portanto sem a necessidade de aplicação de uma prova de esforço.

### Agradecimentos

Este projeto recebeu apoio das agências CAPES e CNPq.

### Referências

- [1] Rosenwinkel ET, Bloomfield DM, et al. Exercise and autonomic function in health and cardiovascular disease. *Clinical Cardiology*. 2001; 19(3):369-87.
- [2] Task Force of the European society of cardiology and the North American society of pacing and electrophysiology, Heart rate variability-standards of measurement, physiological interpretation and clinical use. *European Heart Journal*. 1996; 93(5):1043-65.
- [3] Kawaguchi LYA, Nascimento ACP, Lima MS, Frigo L, Paula Jr. AR, Tierra-Criollo CJ, Lopes-Martins AB. Caracterização da variabilidade da frequência cardíaca e sensibilidade do barorreflexo em indivíduos sedentários e atletas do sexo masculino. *Revista Brasileira de Medicina do Esporte*. 2007; 13(4):207-12.
- [4] Maroco M, Nadal, J, Barbosa PR. The effect of an aerobic training program on the electrical remodeling of the heart: High-frequency components of the signal-averaged electrocardiogram are predictors of the maximal aerobic power. *Brazilian Journal of Medical and Biological Research*. 2007; 40(2): 199-208.
- [5] Hartigan J, Wang M. A K-means clustering algorithm. *Applied Statistics*. 1979, 28:100-8.
- [6] Muniz AM, Nadal J, Lyons KE, Pahwa R, Liu W. Long-term evaluation of gait initiation in six Parkinson's disease patients with bilateral subthalamic stimulation. *Gait Posture*. 2012; 35(3):452-7.
- [7] Jackson AS, Pollock, M. Generalized equations for predicting body density of men, *The British Journal of Nutrition*. 1978; 40:497-504.
- [8] Siri, WE. Body composition from fluid spaces and density: analysis of methods, apud: Brozek, J and Henschel. *Techniques for measuring body composition*. Washington National Academic of Science, 1961.
- [9] Napoleon D, Pavalakodi S. A new method for dimensionality reduction using K-Means clustering algorithm for high dimensional data set. *International Journal of Computer Applications*. 2011; 13(7):41-6.