

Regressão Stepwise aplicada ao estudo do Absenteísmo e sua relação com a Fadiga Crônica

Anderson Cristiano Neisse ¹, Fernando Luiz Pereira de Oliveira ², Anderson Castro Soares de Oliveira ³, Frederico Rodrigues Borges da Cruz ⁴

Introdução

A Síndrome da Fadiga Crônica (CFS), ou Encefalomielite miálgica, tem sido comum nas práticas clínicas das últimas décadas após sua primeira definição de caso por Holmes et al. (1988). Várias revisões da definição de caso têm sido feitas desde então, mas a de Fukuda et al. (1994) permanece a mais utilizada até hoje, definindo como caso de CFS aquele que tenha presença de pelo menos 4, dos 8 sintomas: Problemas de concentração ou memória, dor de garganta, gânglios linfáticos sensíveis, dor muscular, dor articular, dor de cabeça, sono inefetivo e mal-estar pós-esforço. Dentre fatores relacionados com a CFS, estudos sugerem: sono insatisfatório/irregular, estresse psicológico, colesterol, nutrientes, infecções e outros. Segundo Kennedy et al. (2010), a qualidade de vida de crianças com CFS se compara à de crianças com diabetes mellitus tipo 1 ou asma.

Trabalhadores em turnos são naturalmente mais suscetíveis ao desenvolvimento da CFS devido aos hábitos inusuais de sono e descanso. Segundo Costa (2010) o trabalho em turnos noturnos é muito estudado por perturbar o ciclo de sono, resultando em desregulação dos ritmos circadianos biológicos de humanos, seres naturalmente diurnos. Em condições de trabalho de risco o desenvolvimento da CFS pode aumentar a chance de acidentes fatais, como o trabalho em turnos alternados da indústria de mineração que apresenta naturalmente os fatores de sono irregular/insatisfatório e estresse psicológico. Déficit de sono e alteração persistente do ritmo circadiano pode levar a CFS, neuroticismo, ansiedade crônica e/ou depressão, aumentando risco de absenteísmo e necessidade de medicamentos psicotrópicos (Cole et al., 1990; Colquhoun et al., 1996; Nakata et al., 2004).

De acordo com Murphy et al. (2011), modelagem preditiva pode ser uma ferramenta efetiva na prevenção da CFS. No intuito de buscar evidências que permitam auxiliar na prevenção da CFS e procurar fatores relacionados, este trabalho estudou a relação entre absenteísmo, 8 variáveis antropométricas e 11 bioquímicas por meio de regressão Logística de forma a estabelecer relação com os fatores relacionados à CFS.

Metodologia

Um estudo transversal foi realizado em 2012 com trabalhadores em turno de uma mina localizada nos Inconfidentes, região de Minas Geras, Brazil. Tratam-se de operadores de caminhão em turnos de 6 horas seguidas por descanso de 12 horas. Os dados coletados consistem de 22 variáveis: 8 antropométricas, 11 bioquímicas, sexo, idade e a variável que indica se houve absenteísmo durante o ano de 2012. Esta última sendo a variável

¹Universidade Federal de Viçosa (UFV). e-mail: a.neisse@gmail.com

²Universidade Federal de Ouro Preto (UFOP)

³Universidade Federal de Mato Grosso (UFMT)

⁴Universidade Federal de Minas Gerais (UFMG).

dependente do estudo de forma a identificar relações entre absenteísmo e síndrome de fadiga crônica (CFS). Após a análise descritiva, e análise de correlação foi feita a análise de dados faltantes. Variáveis com mais de 15% de dados faltantes foram removidas uma vez que proporções de dados faltantes muito altas não são confiáveis para imputação de dados. Devido ao fator que levou a dados faltantes foi a perda de dados pelas coletoras, foi assumido padrão MAR para os dados faltantes. Foi feita padronização locação-escala nas variáveis e então os dados faltantes foram imputados com o método dos K vizinhos mais próximos (KNN). Foram gerados dois bancos de dados: um contendo somente os casos sem dados faltantes (*Complete-Cases*) e outro contendo o banco de dados com imputações (*KNN-Imputed*) com o intuito de comparar desempenhos. Por fim, 70% de cada um dos bancos de dados foi considerado como treinamento.

Aos dois bancos de dados de treinamento foram ajustadas regressões logísticas Stepwise (bidirecional), Lasso, Ridge e Elastic-Net. Tanto a escolha de modelo do Stepwise quanto a procura de parâmetros das demais regressões foram feitas com validação cruzada *repeated 10-fold* com 10 repetições, utilizando como medida a área abaixo da curva ROC (AUC). A qualidade de ajuste dos modelos foi testada por meio do teste Hosmer-Lemeshow. Foram obtidas medidas de Sensibilidade, Especificidade e AUC para cada modelo, assim como suas distribuições empíricas estimadas por bootstrap não-paramétrico ($r = 1000$) aplicado nos dados de teste. A partir das distribuições empíricas, foram obtidas as medidas médias e seus intervalos de confiança de bootstrap com 95% de confiança. Utilizando os intervalos de confiança foi comparado: (i) Performance entre bancos de dados imputados e de casos completos; (ii) Performance entre modelos Stepwise, Ridge, Lasso e Elastic Net. Finalmente, o melhor modelo teve seus efeitos explorados por meio de intervalos de confiança de bootstrap de 95% obtidos nos dados de teste.

Resultados e Discussão

Os testes de Hosmer-Lemeshow indicaram bom ajuste de todos os modelos. Não houve diferença de performance nos dados de teste entre os dois bancos de dados para nenhum modelo. Também não houve diferença de performance entre os modelos de acordo com os intervalos de confiança. Os intervalos de AUC incluem o valor de 0,5, indicando falta de poder discriminativo para a variável de absenteísmo. Os métodos de regularização e seleção de variáveis não trouxeram melhoria na performance. Este fato somado ao fato de não haver diferença entre os modelos levou à decisão pelo modelo Stepwise para exploração dos efeitos. Os coeficientes acompanhados de suas estimativas e intervalos de bootstrap são mostrados na Tabela 1, estimativas acompanhadas de “.”, “*” ou “**” são significativas a 10%, 5% e 1% respectivamente.

Tabela 1: Coeficientes de regressão e estimativas Bootstrap ($r = 1.000$).

	Estimativas	Média Boot.	IC Boot.
Intercep.	-1,3266**	-1,3610	(-1,6330; -1,1166)
AvDBP	-0,1935	-0,1881	(-0,4710; 0,0868)
LDL	0,5362.	0,5680	(-0,0316; 1,1882)
Trig.	0,4911**	0,5028	(0,1760; 0,8216)
Col.	-0,7375*	-0,7740	(-1,4359; -0,1252)
Sódio	-0,4939*	-0,5464	(-1,1451; -0,0911)

Deve-se notar que os dados foram padronizados e que os coeficientes estão na escala de log-odds. Os resultados da Tabela 1 correspondem ao banco de dados imputado. A Figura 1 mostra os gráficos de violino (densidades estimadas espelhadas) em conjunto com os intervalos de confiança de 95% obtidos por Bootstrap ($r = 1.000$) para os coeficientes.

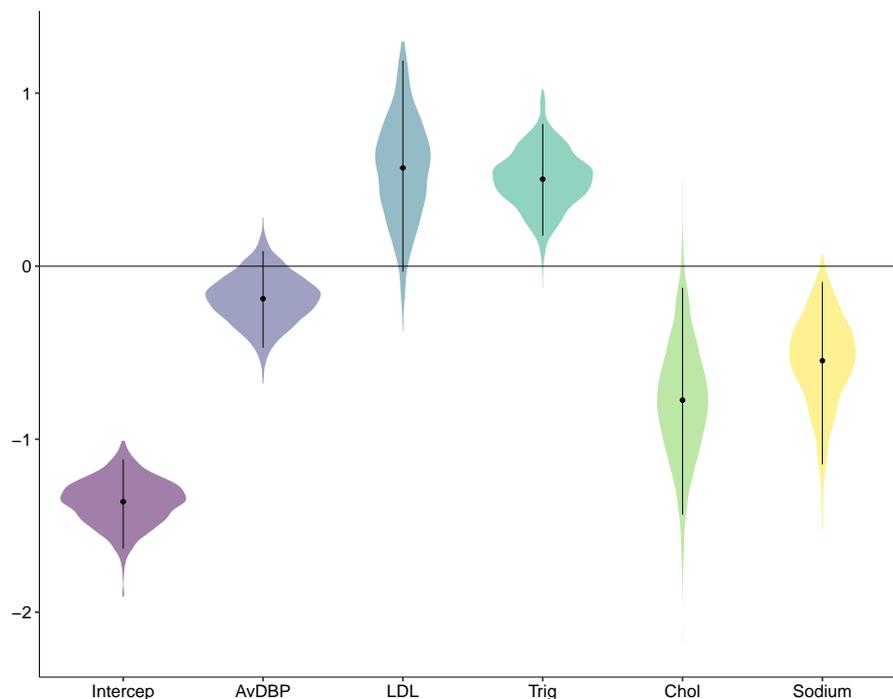


Figura 1: Violinos (Densidade espelhada) e Intervalos de confiança Bootstrap ($r=1.000$) 95% para os coeficientes da regressão Stepwise.

Os resultados sugerem efeito significativo em Triglicerídeos, Colesterol Total e Sódio. Um indivíduo que está um desvio padrão acima da média de Triglicerídeos terá um aumento de 0,49411 no log-odds de absenteísmo. Estudos sugerem que indivíduos com CFS apresentam níveis elevados de triglicerídeos (Tomic et al., 2010; De Lorenzo et al., 1997). Quanto ao colesterol total é sugerido decréscimo de log-odds para indivíduos acima da média, resultado similar aos obtidos por De Lorenzo et al. (1997). Por fim, sódio também apresenta decréscimo no log-odds, o que concorda com estudos que indicam relação entre hipotensão neuralmente mediada com a CFS (Rowe e Calkins, 1998), condição esta que é tratada com consumo supervisionado de sódio (Comhaire, 2018).

Conclusões

Resultados indicam falta de poder discriminatório dos modelos ajustados aos dados. Este fato pode resultar do ruído elevado inerente à natureza da variável explicativa. Além disso, todos os modelos se mostraram estatisticamente equivalentes de acordo com as medidas de performance e concordaram quanto aos efeitos de variáveis. Apesar da falta de discriminação, a exploração dos efeitos do modelo Stepwise indicou efeitos significativos de Triglicerídeos, Colesterol Total e Sódio para a descrição da chance de absenteísmo. Os resultados também sugerem importância do Colesterol LDL como um fator que tem correlação com a chance de absenteísmo.

Agradecimentos

Este estudo foi financiado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Universidade Federal de Ouro Preto (UFOP), Universidade Federal de Viçosa e o projeto Prevenção da Fadiga.

Referencias Bibliográficas

- Cole, R. J.; Loving, R. T.; Kripke, D. F. (1990). Psychiatric aspects of shiftwork. *Occupational medicine (Philadelphia, Pa.)*, 5(2), 301-314.
- Colquhoun, W. P., Costa, G., Folkard, S., and Knauth, P. (1996). Shiftwork. Problems and solutions.
- Comhaire, F. (2018). Treating patients suffering from myalgic encephalopathy/chronic fatigue syndrome (ME/CFS) with sodium dichloroacetate: An open-label, proof-of-principle pilot trial. *Medical hypotheses*, 114, 45-48.
- Costa, G. (2010). Shift work and health: current problems and preventive actions. *Safety and health at Work*, 1(2), 112-123.
- De Lorenzo, F., Xiao, H., Mukherjee, M., Harcup, J., Suleiman, S., Kadziola, Z., and Kakkar, V. V. (1998). Chronic fatigue syndrome: physical and cardiovascular deconditioning. *QJM: monthly journal of the Association of Physicians*, 91(7), 475-481.
- Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., and Komaroff, A. (1994). The chronic fatigue syndrome: a comprehensive approach to its definition and study. *Annals of internal medicine*, 121(12), 953-959.
- Holmes, G. P., Kaplan, J. E., Gantz, N. M., Komaroff, A. L., Schonberger, L. B., Straus, S. E., ... and Tosato, G. (1988). Chronic fatigue syndrome: a working case definition. *Annals of internal medicine*, 108(3), 387-389.
- Kennedy, G., Underwood, C., and Belch, J. J. F. (2010). Physical and functional impact of chronic fatigue syndrome/myalgic encephalomyelitis in childhood. *Pediatrics*, 125(6), e1324-e1330.
- Murphy, S. M., Castro, H. K., and Sylvia, M. (2011). Predictive modeling in practice: improving the participant identification process for care management programs using condition-specific cut points. *Population health management*, 14(4), 205-210.
- Nakata, A., Haratani, T., Takahashi, M., Kawakami, N., Arito, H., Kobayashi, F., ... and Araki, S. (2004). Association of sickness absence with poor sleep and depressive symptoms in shift workers. *Chronobiology International*, 21(6), 899-912.

R CORE TEAM. *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Rowe, P., and Calkins, H. (1998). Neurally mediated hypotension and chronic fatigue syndrome. *The American journal of medicine*, 105(3), 15S-21S.

Tomic, S., Brkic, S., Maric, D., and Mikic, A. N. (2012). Lipid and protein oxidation in female patients with chronic fatigue syndrome. *Archives of medical science: AMS*, 8(5), 886.