Logistic Regression: A Brief Primer

(Stoltzfus, 2011)

Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Academic Emergency Medicine*, *18*(10), 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

As técnicas de regressão são versáteis na sua aplicação à investigação médica porque podem medir as associações, prever resultados, e controlar para efeitos variáveis confusos. Como uma dessas técnicas, a regressão logística é uma forma eficiente e poderosa de analisar o efeito de um grupo de variáveis independentes sobre um resultado binário, quantificando a contribuição única de cada variável independente. Utilizando componentes da regressão linear reflectidos na escala logítica, a regressão logística identifica iterativamente a combinação linear mais forte de variáveis com a maior probabilidade de detectar o resultado observado.

Considerações importantes na condução da regressão logística incluem a selecção de variáveis independentes, a garantia de que os pressupostos relevantes são cumpridos, e a escolha de uma estratégia adequada de construção de modelos. Para a selecção de variáveis independentes, deve-se orientar por factores tais como teoria aceite, investigações empíricas anteriores, considerações clínicas, e análises estatísticas univariadas, com reconhecimento de potenciais variáveis confusas que devem ser contabilizadas. Os pressupostos básicos que devem ser cumpridos para a regressão logística incluem independência de erros, linearidade no logit para variáveis contínuas, ausência de multicolinearidade, e ausência de outliers fortemente influentes. Além disso, deve haver um número adequado de eventos por variável independente para evitar um modelo de sobreajustamento, com "regras de polegar" mínimas geralmente recomendadas, variando de 10 a 20 eventos por covariada. Relativamente às estratégias de construção de modelos, os três tipos gerais são direct⁄ standard, sequencial ⁄ hierárquico, e passo a passo ⁄ estatístico, tendo cada um deles uma ênfase e um objetivo diferentes. Antes de se chegar a conclusões definitivas a partir dos resultados de qualquer destes métodos, deve-se quantificar formalmente a validade interna do modelo (ou seja, a replicabilidade dentro do mesmo conjunto de dados) e a validade externa (ou seja, a generalizabilidade para além da amostra atual). A adequação global do modelo de regressão logística resultante aos dados da amostra é avaliada utilizando várias medidas de adequação, com uma melhor adequação caracterizada por uma menor diferença entre os valores observados e os valores previstos no modelo. Recomenda-se também a utilização de estatísticas de diagnóstico para avaliar melhor a adequação do modelo. Finalmente, os resultados para variáveis independentes são normalmente reportados como odds ratios (ORs) com intervalos de confiança de 95% (CIs).