



Modelos de regressão linear e logística: quando utilizá-los e como interpretá-los?

Horacio Matias Castro^{1,2} , Juliana Carvalho Ferreira^{1,3} 

CENÁRIO PRÁTICO

Uma análise secundária⁽¹⁾ do estudo denominado “*Integrating Palliative and Critical Care*”, ensaio randomizado de *cluster*, foi realizada para explorar as diferenças no recebimento de elementos de cuidados paliativos entre pacientes que morreram na UTI com doença pulmonar intersticial (DPI) ou DPOC em comparação com aqueles que morreram de câncer. Os autores utilizaram dois métodos de análise de regressão múltipla: regressão linear para estimar o impacto da DPOC e da DPI, em comparação com o do câncer, no tempo de internação na UTI, e regressão logística para avaliar os efeitos da DPOC e da DPI sobre a presença ou ausência de elementos de cuidados paliativos. Todos os modelos de regressão foram ajustados para fatores de confusão (idade, sexo, status de minoria, escolaridade, entre outros) da associação entre o diagnóstico do paciente e os desfechos em cuidados paliativos.

INTRODUÇÃO

As regressões linear e logística são métodos estatísticos amplamente utilizados para avaliar a associação entre variáveis em pesquisa médica. Esses métodos estimam se existe associação entre a variável independente (também chamada de preditor, exposição ou fator de risco) e a variável dependente (desfecho).⁽²⁾

A associação entre duas variáveis é avaliada por meio da análise de regressão simples. No entanto, em muitos cenários clínicos, mais de uma variável independente pode estar associada ao desfecho, podendo haver necessidade de controle de variáveis de confusão. Quando mais de duas variáveis independentes estão associadas ao desfecho, utiliza-se a análise de regressão múltipla. A análise de regressão múltipla avalia o efeito independente de cada variável sobre o desfecho, ajustando para o efeito das outras variáveis incluídas no mesmo modelo de regressão.

QUANDO UTILIZAR REGRESSÃO LINEAR OU LOGÍSTICA?

O fator determinante do tipo de análise de regressão a ser utilizado é a natureza da variável de desfecho. A regressão linear é utilizada para variáveis de desfecho contínuas (dias de hospitalização ou VEF₁, por exemplo), e a regressão logística é utilizada para variáveis de desfecho categóricas, como óbito. As variáveis independentes podem ser contínuas, categóricas ou uma combinação das duas.

Em nosso exemplo, os autores queriam saber se havia relação entre câncer, DPOC e DPI (doença de base, a variável independente) e dois diferentes desfechos. Um desfecho era contínuo (tempo de internação na UTI), e o outro, categórico (presença ou ausência de elementos de cuidados paliativos). Portanto, foram construídos dois modelos: um modelo linear para examinar a associação entre a doença de base (doença pulmonar crônica ou câncer) e o tempo de internação na UTI, e uma análise de regressão logística para examinar a associação entre a doença de base e o recebimento de elementos de cuidados paliativos.

COMO INTERPRETAR OS RESULTADOS DA ANÁLISE DE REGRESSÃO?

Os modelos de regressão são realizados em pacotes estatísticos, e os resultados de saída (*output*) incluem vários parâmetros, que podem ser complexos de interpretar. Os clínicos que estão aprendendo os fundamentos dos modelos de regressão devem se concentrar nos parâmetros-chave apresentados no Quadro 1.

Em nosso exemplo, a doença de base — DPOC, DPI ou câncer (categoria de referência) — é a variável independente, e o tempo de internação na UTI e o recebimento de elementos de cuidados paliativos são os desfechos de interesse. Além disso, os modelos de regressão também incluíram outras variáveis independentes consideradas potenciais fatores de confusão, como idade, sexo e status de minoria. No modelo de regressão linear, o tempo de internação na UTI para os pacientes com DPI foi maior do que para aqueles com câncer ($\beta = 2,75$; IC95%: 0,52-4,98; $p = 0,016$), o que significa que, em média, ter DPI aumentou o tempo de internação na UTI em 2,75 dias em comparação com o tempo de internação na UTI entre os pacientes com câncer. No modelo de regressão logística, os autores constataram que a probabilidade de haver qualquer documentação da avaliação da dor nas últimas 24 h de vida era menor para os pacientes com DPI do que para aqueles com câncer (OR = 0,43; IC95%: 0,19-0,97; $p = 0,042$), o que significa que ter DPI diminui em mais da metade as chances de documentação da avaliação da dor.

PONTOS-CHAVE

- Regressão linear e regressão logística são métodos estatísticos importantes para testar relações entre variáveis e quantificar a direção e a força da associação.

1. Methods in Epidemiologic, Clinical, and Operations Research-MECOR-program, American Thoracic Society/Asociación Latinoamericana del Tórax, Montevideo, Uruguay.

2. Pulmonary Medicine Department, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

3. Divisão de Pneumologia, Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo (SP) Brasil.

Quadro 1. Parâmetros mais importantes em análises de regressão e suas interpretações.

Parâmetros	Regressão linear	Regressão logística
Direção e força da associação entre a variável independente e a variável dependente (desfecho)	Coefficiente beta: Descreve a mudança média (esperada) na variável de desfecho para cada mudança de uma unidade na variável independente para variáveis contínuas, ou a mudança média na variável de desfecho para uma categoria da variável independente em comparação com a categoria de referência para variáveis categóricas.	OR: A OR para uma variável independente contínua é interpretada como a mudança nas chances de ocorrência do desfecho para cada aumento de uma unidade na variável independente. A OR para variáveis independentes categóricas é interpretada como o aumento ou diminuição nas chances entre duas categorias (homens vs. mulheres, por exemplo). OR = 1: sem associação; OR > 1: associação positiva ou fator de risco; e OR < 1: associação negativa ou fator de proteção
Exemplo (para variável independente contínua)	O aumento esperado no VEF ₁ para cada aumento de um centímetro na estatura	O aumento esperado nas chances de óbito para cada aumento de um ano na idade entre pacientes com sepse
Exemplo (para variável independente categórica)	O aumento esperado no VEF ₁ para homens em comparação com mulheres da mesma estatura e idade	O aumento esperado nas chances de óbito para homens em comparação com mulheres entre pacientes com COVID-19
Precisão da estimativa	O IC95% do coeficiente beta	O IC95% da OR
Significância estatística	O valor de p (significativo quando < 0,05)	O valor de p (significativo quando < 0,05)

- A regressão linear é utilizada com desfechos contínuos, e a regressão logística é utilizada com desfechos categóricos.
- Esses procedimentos requerem experiência na construção de modelos de regressão e, geralmente, a assistência de um bioestatístico.

REFERÊNCIAS

1. Brown CE, Engelberg RA, Nielsen EL, Curtis JR. Palliative Care for Patients Dying in the Intensive Care Unit with Chronic Lung Disease Compared with Metastatic Cancer. *Ann Am Thorac Soc.* 2016;13(5):684-689. <https://doi.org/10.1513/AnnalsATS.201510-667OC>
2. Bzovsky S, Phillips MR, Guymer RH, Wykoff CC, Thabane L, Bhandari M, et al. The clinician's guide to interpreting a regression analysis. *Eye (Lond).* 2022;36(9):1715-1717. <https://doi.org/10.1038/s41433-022-01949-z>