

Regresión Logística: Una Introducción Breve

Adaptado por Carlos a partir de Stoltzfus, J. C.

8 de abril de 2025

Resumen traducido

Las técnicas de regresión son versátiles en su aplicación a la investigación médica porque pueden medir asociaciones, predecir resultados y controlar los efectos de variables de confusión. Como una de estas técnicas, la regresión logística es una forma eficiente y poderosa de analizar el efecto de un grupo de variables independientes sobre un resultado binario cuantificando la contribución única de cada variable independiente. Utilizando componentes de regresión lineal reflejados en la escala logit, la regresión logística identifica iterativamente la combinación lineal más fuerte de variables con la mayor probabilidad de detectar el resultado observado.

Las consideraciones importantes al realizar una regresión logística incluyen la selección de variables independientes, asegurando que se cumplan los supuestos relevantes y eligiendo una estrategia adecuada de construcción del modelo. Para la selección de variables independientes, uno debe guiarse por factores como teoría aceptada, investigaciones empíricas previas, consideraciones clínicas y análisis estadísticos univariados, reconociendo las posibles variables de confusión que deben ser consideradas.

Los supuestos básicos que deben cumplirse para la regresión logística incluyen independencia de errores, linealidad en el logit para variables continuas, ausencia de multicolinealidad y falta de valores atípicos fuertemente influyentes. Adicionalmente, debe haber un número adecuado de eventos por variable independiente para evitar un modelo sobreajustado, con un mínimo comúnmente recomendado de “reglas prácticas” que van de 10 a 20 eventos por covariable.

Respecto a las estrategias de construcción de modelos, los tres tipos generales son: directa/estándar, secuencial/jerárquica y por pasos/estadística, cada uno con un énfasis y propósito diferente. Antes de llegar a conclusiones definitivas a partir de los resultados de cualquiera de estos métodos, se debe cuantificar formalmente la validez interna del modelo (es decir, su replicabilidad dentro del mismo conjunto de datos) y su validez externa (es decir, su generalizabilidad más allá de la muestra actual).

El ajuste general del modelo de regresión logística a los datos de muestra se evalúa utilizando varias medidas de bondad de ajuste, donde un mejor ajuste se caracteriza por una menor diferencia entre los valores observados y los valores predichos por el modelo. También se recomienda el uso de estadísticas de diagnóstico para evaluar aún más la adecuación del modelo. Finalmente, los resultados para las variables independientes suelen reportarse como razones de momios (odds ratios, ORs) con intervalos de confianza (IC) del 95 %.