

# Buchbesprechung: Regression Methods in Biostatistics

• Rainer Muehe<sup>1</sup>

## Bibliographische Angaben

E. Vittinghoff, D.V. Glidden, S.C. Shiboski, C.E. McCulloch:

Regression Methods in Biostatistics

Linear, Logistic, Survival, and Repeated Measures Models

Springer-Verlag Inc., New York, 2005, 340 Seiten, ISBN 0-387-20275-7

## Rezension

Regressionsanalysen sind ein wichtiges Auswertungstool geworden bei der Analyse von Daten im Bereich der Lebenswissenschaften. Dabei wird im Wesentlichen von der Datensituation ausgegangen, dass man eine einzelne Zielgröße durch mehrere Einflussgrößen beschreiben kann. Mögliche Fragestellungen können dabei sein: (1) Vorhersage der Zielgröße auf Basis der Einflussgrößen, (2) Untersuchung des Effektes einer Einflussgröße adjustiert um den Effekt der weiteren Variablen im Modell und (3) Untersuchung des gemeinsamen Effektes mehrerer Einflussgrößen. Die jeweilige Fragestellung bedingt im Modellierungsprozess teilweise ein anderes Herangehen, was im Buch angesprochen und anhand von Beispielen beschrieben wird.

Die Autoren haben sich in dem Buch vorgenommen, die wichtigsten Regressionsmodelle für vier verschiedene Anwendungssituationen vorzustellen und anhand vieler Beispiele die Durchführung der Modellierung und speziell auch die Interpretation der Ergebnisse dem Leser näher zu bringen. Dazu gehören für stetige Zielgrößen das Lineare Regressionsmodell, für dichotome Zielgrößen das Logistische Regressionsmodell, für Überlebenszeitdaten das Cox-Proportional-Hazard Modell und für Messwiederholungen Modelle für Repeated Measurements (u.a. GEE und Mixed Model).

Das Buch ist von der Intention sehr anwendungsbezogen geschrieben und wendet sich hauptsächlich an Forscher in den Lebenswissenschaften, die diese Modelle in ihren Auswertungen benötigen und damit umgehen müssen sowie Statistikstudenten, die die

Anwendung der Methoden in „real life“ Situationen interessiert. Deshalb sind im Text nur die absolut notwendigen Formeln ohne Herleitungen, Beweise o. ä. angegeben, um sich auf die Modellierungsaspekte und Interpretation der Ergebnisse konzentrieren zu können. Darüber hinaus werden die Beispiele anhand einer häufig genutzten Statistiksoftware (Aufrufe und Output) dargestellt und interpretiert. Die Autoren haben sich hierbei für Stata entschieden, was heute zu einem der meist genutzten Statistiksoftwarepakete in der medizinischen Forschung und Verwandten Gebieten wie der Epidemiologie zählt.

Jedes der 11 Kapitel in dem Buch endet mit weitergehenden kurz kommentierten Literaturhinweisen, Übungsaufgaben und der Angabe der Lernziele. Die im Text und in den Übungsaufgaben benutzten Daten und Stata-Programme können von der Internetseite <http://www.biostat.ucsf.edu/vgsm/> herunter geladen werden, Musterlösungen für die Übungsaufgaben zur Selbstkontrolle werden allerdings nicht angeboten. Das ist sicher ein Manko bei der sonst so konsequent durchgezogenen didaktischen Aufbereitung des Lehrstoffes in dem Buch.

Vorausgesetzt für das Verständnis des Buches werden grundlegende Kenntnisse in Medizinstatistik wie t-Tests und Varianzanalyse, Korrelationen, einfache lineare Regression, Kontingenztafeln, Kaplan-Meier-Analysen und Logrank-Test. Trotzdem werden diese im Kapitel 3 des Buches auf etwa 40 Seiten wiederholt, so dass das Buch auch für Regressions-Anfänger geeignet ist.

Die wesentlichen Kapitel im Buch sind das Kapitel 4 mit der Beschreibung des Linearen Regressionsmodells, Kapitel 5 über Variablenselektionsaspekte, Kapitel 6 über die Logistische Regression, Kapitel 7 über Survival Analysis und die Kapitel 8 und 9 über „Repeated Measures Analysis“ und „Generalized Linear Models“. Darüber hinaus wird im Kapitel 10 etwas über die Schätzung von Effekten in komplexen Surveys angeschnitten und im abschließenden Kapitel 11 wird die Planung und Auswertung einer Studie mit Regressionsanalyse zusammengefasst und weitere praktische Aspekte wie Studienplan, Datenmanagement,

<sup>1</sup> Institut für Biometrie, Universität Ulm, Ulm, Deutschland

Softwareauswahl, Datenschutz und Beratung durch Biometriker angesprochen. Gerade dies letzte Kapitel macht das Buch zu einem wertvollen Ratgeber für interessierte Leser, die vor dem Beginn einer Forschungsaktivität stehen und nicht nur die entsprechenden statistischen Methoden kennen lernen wollen.

Das wichtigste an dem Buch sind die vielen Hinweise auf praktische Modellierungsaspekte, die oft in Lehrbüchern nur am Rande oder gar nicht beschrieben werden. So gibt es z.B. im Kapitel 4 über das Lineare Regressionsmodell neben der Beschreibung des Modells und der wichtigsten Kenngrößen wie den Koeffizienten, Tests und  $R^2$  umfangreiche Hinweise zur Untersuchung von Confounding, Wechselwirkungen, Modellvoraussetzungen und Modell-Fit. Darüber hinaus werden im folgenden Kapitel die Durchführung und Probleme der Variablenselektion intensiv bespro-

chen bis hin zur Kollinearität, Anzahl Variablen im Modell und Alternativen zu Stepwise-Verfahren. Diese Aspekte (Confounding, Kollinearität, Modellprüfung, Wechselwirkung etc.) finden sich dann auch jeweils kürzer in den anderen Modellbeschreibungen wieder.

Gerade diese letztgenannten Aspekte in Verbindung mit dem praxisnahen und beispielhaften Ansatz machen dies Buch sehr lesenswert, nicht nur für die von den Autoren oben genannten Lesergruppen. Als Hintergrundlektüre für die Beratungstätigkeit an einer Universität kann ich das Buch dringend empfehlen.

---

### Korrespondenzadresse:

• Dr. biol. hum. Rainer Muche, Institut für Biometrie, Universität Ulm, Schwabstr. 13, 89075 Ulm, Tel: ++49-731-502-6903, Fax: ++49-731-502-6902 [rainer.muche@uni-ulm.de](mailto:rainer.muche@uni-ulm.de)