

Rezension zu: Siegmund, F. (2020). Statistik in der Archäologie: eine anwendungsorientierte Einführung auf Basis freier Software. Norderstedt: BoD. Softcover, 408 Seiten mit 125 z. T. farbigen Abb., dazu Beispiel- und Übungsdateien. ISBN: 978-3-75282-570-1. PrePrint kostenfrei erhältlich unter <https://frank-siegmund.de/statistik> [13.7.2020]).

Sophie C. Schmidt

Das hier besprochene Buch ist das erste deutsche Lehrbuch für statistische Methoden in der Archäologie seit dem bekannten „Ihm“ von 1978 (IHM, 1978) und damit eine seit langem überfällige Neuauflage des Themas. Die Rezensentin (Rez.) unterrichtet selbst seit 2016 quantitative Methoden in der Archäologie und ihre Umsetzung mit der Programmiersprache R. Sie hat das Fehlen eines neuen Lehrwerks schon seit längerem bedauert.

Siegmund baut seine Einführung auf „Basis freier Software“ auf und leistet damit einen wichtigen und begrüßenswerten Beitrag zur Zugänglichkeit von statistischem Anwendungswissen und der Verbreitung offener und freier Software, nicht zuletzt R. Dies wird dadurch verstärkt, dass der Pre-Print des Buchs kostenlos zusammen mit Übungsdateien und -skripten herunterladbar ist. Es ist hier nur zu bedauern, dass das Buch, Code und Daten nicht unter einer DOI oder einer anderen persistenten URL abgelegt wurden, was eine Zitation der Online-Fassung über die Lebensdauer der persönlichen Homepage des Autors erlauben würde. Dies und eine zitierbare Versionierung der Überarbeitungen wäre z. B. mithilfe von Zenodo, dem Open Science Framework oder ähnlichen Repositorien möglich.

Das siebenseitige, stark gegliederte Inhaltsverzeichnis deutet auf die Art und Weise hin, wie der Autor das Buch gelesen wissen möchte: Als Beihilfe zum Selbststudium, für das gezielt Kapitel herausgesucht und nachgelesen werden können. Dabei geht der Autor „von Null aus“: Es handelt sich um ein Grundlagenwerk, für das keinerlei Vorkenntnisse benötigt werden, das dann aber auch nicht zu komplexeren Themen, z. B. multivariaten Analysen, vordringt. Dies ist eine der Stärken des Buchs, da es so einen nachvollziehbaren Schwerpunkt legt und damit sowohl für alle Anfänger geeignet ist als auch von etwas Fortgeschritteneren als Nachschlagewerk genutzt werden kann. Der Stil des Buches orientiert sich an dem Ursprung des Werkes als Vorlesungsskript, ist verständlich und reich an Beispielen. Dem Text ist die langjährige Erfahrung des Autors als Dozent anzumer-

ken, der an vielen Stellen auf mögliche Fehlerquellen eingeht und auch einfache Prozesse sehr kleinschrittig erklärt. An einigen Stellen treibt der eigentlich angenehme Erklärfluss der Vorlesung jedoch leider strukturell absonderliche Blüten: So bespricht der Autor z. B. unter „17.1 Empfehlungen zu statistischen Grafiken“ acht Empfehlungen unterschiedlicher Länge, deren letzte in detaillierten Hinweisen zu Histogrammen mündet, welche zu einem Unterkapitel „17.1.1 Schätzung der empirischen Dichtefunktion“ überleiten. Nach diesem Ausflug folgt „17.2 Beispiel-Grafiken“, ein Kapitel ohne eine einzige Beispiel-Grafik, in dem sich eigentlich weitere Empfehlungen zum Thema Lesbarkeit und Farbschemata verbergen, die gut in das vorhergehende Empfehlungs-Kapitel gepasst hätten. Daran schließt sich Kapitel 17.3 zu Balken- und Kreisdiagrammen an, welches von einem Kapitel „17.4 Histogramm“ gefolgt wird. Ein winziger Hinweis bekommt ein eigenes Unterkapitel „17.4.2.1 Hinweis: Grafiken nachbearbeiten mit InkScape“, der nicht unbedingt zu den Histogrammen gehört, genausowenig wie Kapitel „17.4.4 Datenreihen im Balkendiagramm vergleichen“, da Balkendiagramme keine Histogramme sind. Eine redaktionelle Überarbeitung der Struktur der Unterkapitel hätte der Nachvollziehbarkeit des Buches geholfen und es Nutzern erleichtert, bestimmte Informationen im Inhaltsverzeichnis zu finden.

Siegmund beginnt das Buch mit Erläuterungen über Statistik, Archäologie und den Vorzügen freier Software und widerlegt im Kapitel 6 die Skepsis gegenüber statistischen Methoden. Es folgt „Ein kurzer Rückblick in die Forschungsgeschichte: die Ko-Evolution von angewandter Statistik und Archäologie“. Die Rez. teilt hier die Einschätzung zu den 2010er Jahren als in der Methodenentwicklung „stagnierend“ nicht ganz. So zeigen z. B. theoretische Ansätze von Nakoinz (NAKOINZ, 2012; 2013; POPA & NAKOINZ, 2016) und die (Weiter-)Entwicklung von geostatistischen Methoden (CASPARI & JENDRYKE, 2017; GILLINGS ET AL., 2020), dass durchaus methodisch Neues in den archäologischen Werkzeugkasten integriert wird. Auch die Einschätzungen zum Thema Datenmanagement greifen nach Ansicht der Rez. etwas zu kurz, da es hier nicht nur darum geht „big data“ zu erzeugen, sondern nach einer ersten Phase der Digitalisierung nun das Entstehen eines „dark age“ der Wissenschaft zu verhindern (z. B. JEFFREY, 2012; SCHMIDT, 2018; siehe auch <https://www.seadda.eu/> [13.7.2020]). Nichtsdestotrotz ist dieser Ausflug in die Forschungsgeschichte ein spannendes Kapitel und die ausführliche Darstellung der Entwicklung der Rechenkapazitäten in der Archäologie für die Verfas-

serin, sowie sicherlich die meisten Leser, die diese Entwicklung nicht miterleben konnten, ein faszinierender forschungsgeschichtlicher Rückblick.

Das Einführungskapitel „*Erster Schritt: persönliche Bedürfnisse und Möglichkeiten klären*“ ist ein sinnvoller Einschub, der in jedem statistischen Lehrwerk seinen Platz finden sollte. Siegmund rät mit einem Hinweis auf Zeitbeschränkungen für Abschlussarbeiten davon ab, R als Skriptsprache zu lernen und stellt in Kapitel 14 die Programme PAST, JASP und JAMOVI vor, allesamt freie, Graphic-User-Interfaces (GUIs) nutzende Software. Dabei geht der Autor sehr ausführlich vor und erläutert alles von der Installation bis zu nützlichen Shortcuts. Das kann sehr hilfreich sein, wird aber, wie der Autor selbst feststellt, durch die zügige Weiterentwicklung der besprochenen Software-Pakete relativ schnell fehlerbehaftet sein. Durch den Vertriebsweg des Buches als *Book on Demand* und die zusätzliche Online-Publikation besteht jedoch die Möglichkeit, regelmäßig Überarbeitungen zu veröffentlichen, weswegen die oben angesprochene Versionierung mit DOIs von Vorteil wäre. Ein eigenes Kapitel erhält die Einführung in R als Arbeitsumgebung, in der auch unterschiedliche GUIs vorgestellt werden. Siegmund arbeitet größtenteils mit R Commander, ermöglicht so aber seinen Lesern, selbständig eine Auswahl zu treffen.

Das Abwägen von Aufwand und Nutzen des Programmieren-Lernens ist legitim und Lernende sollten natürlich über Lernkurven informiert werden. Nach Erachten der Rez. stellt der Autor jedoch die Vorteile des Programmierens in Betracht auf Reproduzierbarkeit (siehe z. B. SCHMIDT & MARWICK, 2020) und der Automatisierung sich wiederholender Aufgaben zu wenig heraus. Obwohl er immer wieder Code-Beispiele bringt, zeigen diese manchmal redundante Arbeitsabläufe, die anhand einer for-Schleife leicht zu automatisieren gewesen wären (z. B. S. 320 f. oder S. 346 f.). Insbesondere in Teil IV, in dem Anwendungsbeispiele vorgeführt werden, wäre eine Demonstration des Mehrwertes von Programmierung gut aufgehoben gewesen, ohne diejenigen, die lieber mit einer GUI arbeiten, zu überfordern.

Ab Kapitel 9 beginnt „*Teil II: Werkzeuge und Grundlagen kennenlernen und auswählen*“. Siegmund geht von hier bis Kapitel 13 auf verschiedene Themen rund um die Datenaufnahme und -redaktion ein. Sehr hilfreich ist Kapitel 10, welches mit wertvollen Hinweisen und „*Praxistipps*“ Anfängern v. a. anhand von LibreOffice Calc (also an einem freien und offenen Äquivalent zu Microsoft Excel) Datenmanagement beibringt. Im Kapitel 11, „*Skalenniveau*“, unterläuft dem Autor ein Flüchtig-

keitsfehler, wo er das Prägejahr einer Münze als verhältnisskaliertes Merkmal aufführt (wobei er es später aber auf S. 158 richtig als intervallskaliertes Merkmal behandelt). Kapitel 12 mit Anleitungen zur Datenverwaltung, einfachen Umformungen und Datenprüfverfahren ist eine gelungene Zusammenfassung diverser Fehlerquellen, u. a. zum Beobachterfehler (*Observer Error*). Ein „*pet peeve*“ der Rez. ist jedoch die Kodierung von Variablen, bei der Siegmund regelhaft (und anscheinend in anthropologischer Forschungstradition) nominale Merkmale mit Zahlen kodiert. Dies führt jedes Mal, wenn diese Daten in ein Statistikprogramm wie R eingeladen werden, dazu, dass das Programm diese Kategorien als „*Zahlen*“ (numeric/integer) einordnet, wie auch auf S. 155 und S. 233 in der Praxis beschrieben. Unnötige Umkodierungen kann man sich sparen, wenn die Kodierung von Nominal- und Ordinaldaten mit Text/Buchstaben vorgenommen wird.

Teil III mit den Kapiteln 16–24, ist der „Kern“ des statistischen Lehrbuchs. Hier beginnt Siegmund mit der Erstellung von Grafiken, die er ausführlich in Funktionsweise und Erstellungsmöglichkeiten erläutert. Abgesehen von strukturellen Uneinheitlichkeiten werden die Grundlagen gut beschrieben. Anschließend widmet sich Siegmund Häufigkeiten und Prozentzahlen. Hier geht er einführend auf die Zuverlässigkeit von Prozentangaben ein und stellt Konfidenzintervalle und deren Berechnungen vor, um in die schließende Statistik einzuführen. Wie auch in anderen Lehrwerken üblich, nutzt Siegmund den Chi-Quadrat-Test zur Erläuterung allgemeiner statistischer Verfahren und Begriffe wie Irrtumswahrscheinlichkeit, kritischer Wert, Nullhypothese etc. Diese Begriffe finden sich auch in einem Stichwortverzeichnis am Ende des Buchs, das die Handhabung als Nachschlagewerk deutlich erleichtert. Das Kapitel 19 zur „*Analyse von Messwerten*“ mit der Erläuterung von Zentral- und Streuungsmaßen, Kapitel 20 zu Normalverteilung und Kapitel 21 zu Testverfahren auf Mittelwertsunterschiede sind klar durchstrukturiert und gut verständlich. Die mathematisch bewanderten Leser des Buchs mag es verwundern, dass Siegmund ohne eine einzige „richtige“ mathematische Formel auskommt. Stattdessen nutzt er beschreibende Worte anstelle mathematischer Symbolik, was Anfängern sicherlich das Verständnis erleichtert. Erfahrenere Nutzer werden hier eine Tiefe der Erklärungen vermissen, sie sind jedoch nicht die erste Zielgruppe des Autors.

Es folgt ein großes Kapitel zu Korrelation und Regression (Kapitel 23). Hier scheint Siegmund

auf S. 267 Scheinkorrelation mit Autokorrelation zu verwechseln. Zur (richtigen) Scheinkorrelation folgt auf S. 276 ein kurzer Absatz unter Kapitel „23.3.4 Praktisches Vorgehen und Ausblick“, der nach Erachten der Rez. mehr Bedeutung hätte erhalten sollen. Auch sind die Voraussetzungen für eine Regression eher knapp gehalten, es fehlt das Stichwort Homoskedastizität (das in einem Anwendungsbeispiel S. 323 jedoch für die multiple lineare Regression erläutert wird) und der Hinweis, dass die Normalverteilung der Residuen eine der Voraussetzungen ist und nicht ein Hinweis darauf, dass andere Voraussetzungen nicht erfüllt sind (missverständliche Formulierung S. 277). Insgesamt werden die unterschiedlichen Regressionsverfahren jedoch angemessen und verständlich erläutert, gespickt mit wertvollen Hinweisen wie „Vorsicht: mathematisch passender heißt nicht zwingend sachlich passender“ (S. 279) bei nicht-linearen Regressionen.

Abschließend werden in Teil IV sechs Anwendungsbeispiele dargelegt, und zwar von der Fragestellung über die Beschreibung der Daten und der Auswahl der geeigneten Methoden bis zur Interpretation. Dies ist für Lernende eine wertvolle Handreichung, die die Übertragung der erlernten Methoden auf eigene Fragestellungen erleichtern wird. Die kommentierte Literaturliste am Ende ist eine Fundgrube an Büchern und Webseiten, in denen nicht nur Anfänger weiterführende Hilfe finden werden.

Dem Autor ist zu gratulieren, dass er das anspruchsvolle Vorhaben eines Statistiklehrbuches für die Archäologie in Eigenregie umgesetzt hat. Das Buch ist Anfängern zu empfehlen, die anhand archäologischer Beispiele und Anwendungsfälle in die statistische Methodik eingeführt werden wollen. Die einfache Sprache wird ihnen den Einstieg erleichtern und insbesondere die ausführlichen Tipps zum Datenmanagement sind wertvoll, da diese häufig in anderen Lehrwerken fehlen. Wer sich später tiefergehend mit statistischen Methoden beschäftigen möchte, dem wird auch im vorliegenden Buch empfohlen, sich von einer GUI zu lösen und das Programmieren zu erlernen, sowie tiefer in die mathematischen Themen einzudringen. Dass das Buch im Selbstverlag erschienen ist, sieht man ihm leider an einigen wenigen Stellen an, an denen Unsauberkeiten im Satz bestehen (fehlende Referenzen, unnötige Zeilenumbrüche), die Struktur nicht ganz nachvollziehbar ist und kleine Verwechslungen passieren, die vor einer Veröffentlichung hätten ausgemerzt werden sollen.

Literatur

Caspari, G. & Jendryke (2017). Archsphere – A cluster algorithm for archaeological applications. *Journal of Archaeological Science: Reports*, 14, 181–188. <https://doi.org/10.1016/j.jasrep.2017.05.052>.

Gillings, M., Hacigüzeller, P. & Lock, G. (Hrsg.) (2020). *Archaeological Spatial Analysis: A Methodological Guide*. London: Routledge. <https://doi.org/10.4324/9781351243858>.

Ihm, P. (1978). *Statistik in der Archäologie*. Köln: Rheinland-Verlag.

Jeffrey, S. (2012). A new Digital Dark Age? Collaborative web tools, social media and long-term preservation. *World Archaeology*, 44(4), 553–570. <https://doi.org/10.1080/00438243.2012.737579>.

Nakoinz, O. (2012). Datierungskodierung und chronologische Inferenz – Techniken zum Umgang mit unscharfen chronologischen Informationen. *Prähistorische Zeitschrift*, 87(1), 189–207. <https://doi.org/10.1515/pz-2012-0010>.

Nakoinz, O. (2013). *Archäologische Kulturgeographie der ältereisenzeitlichen Zentralorte Südwestdeutschlands* (Universitätsforschungen zur prähistorischen Archäologie 224). Bonn: Habelt.

Popa, C. & Nakoinz, O. (27.3.2016). A plead for non-mathematical reasoning in computational archaeology. *Rethinking the use of computational methods in archaeology*. Vortrag bei der CAA-Tagung 2016 in Oslo: <https://doi.org/10.13140/RG.2.1.5101.8642>.

Schmidt, S. C. (25.7.2018). Long term storage of things digital. *Archaeoinformatics*. <https://web.archive.org/web/20200709132650/http://archaeoinformatics.net/long-term-storage-of-things-digital/> [20.7.2020].

Schmidt, S. C. & Marwick, B. (2020). Tool-Driven Revolutions in Archaeological Science. *Journal of Computer Applications in Archaeology*, 3(1), 18–32. <https://doi.org/10.5334/jcaa.29>.

Sophie C. Schmidt, M.A.
Archäoinformatik, Archäologisches Institut
Universität zu Köln
Albertus-Magnus-Platz 1
50923 Köln
s.c.schmidt@uni-koeln.de

<https://orcid.org/0000-0003-4696-2101>

