

Über dieses Buch

Dieses Buch bietet eine Einführung in die Statistik. Abgedeckt werden die Grundlagen der deskriptiven und induktiven Statistik, welche von mir in der Veranstaltung „Wirtschaftsstatistik“ der Bachelor-Studiengänge an der Fakultät Wirtschaftswissenschaften der HTWK Leipzig gelehrt werden. Im Vordergrund steht dabei die Anwendung statistischer Methoden. Die mathematischen Anforderungen sind auf ein Mindestmaß reduziert. Neben vielen Beispielen zur Anwendung und Verdeutlichung der statistischen Methoden werden zu jedem Kapitel Übungsaufgaben präsentiert. Aufgrund der knappen Zeit, die im Studium für Statistik zur Verfügung steht, ist eine Konzentration auf die essentiellen statistischen Konzepte notwendig. In Kap. 13 werden daher Hinweise für weiterführende Literatur gegeben.

Die Einführung in die Statistik wird in diesem Buch ergänzt mit einer Einführung in R. R ist eine mittlerweile sehr populäre, leistungsfähige und kostenfreie Statistiksoftware. Die Kombination der Einführung in die Statistik mit einem Grundkurs in R ermöglicht es, die statistischen Methoden direkt am Rechner anzuwenden und damit besser zu verstehen. Die Einführung in die Statistiksoftware R ist an die zuvor betrachteten statistischen Methoden gekoppelt und entspricht im Wesentlichen dem Statistik-Teil der Veranstaltung „Quantitative Methoden“ in den Master-Studiengängen an der Fakultät Wirtschaftswissenschaften der HTWK Leipzig. Jedes Kapitel im Buch ist daher in drei Teile gegliedert: (i) Statistische Methoden, (ii) Übungsaufgaben und (iii) Anwendung mit R. Dabei sind die Übungsaufgaben auch ohne Software lösbar.

In Kap. 14 (Anhang) sind Tabellen mit Verteilungen ausgewählter Zufallsvariablen zu finden. Ausführliche Lösungen zu den Übungsaufgaben, die im Buch verwendeten Datensätze und den R-Code für Abbildungen und Berechnungen in den Kapiteln gibt es im Internet unter <http://bsturm.htwk-leipzig.de/> oder auf Nachfrage per E-Mail (bodo.sturm@htwk-leipzig.de) beim Autor.

Dieses Buch wäre nicht entstanden ohne die großartige und unermüdliche Unterstützung meines Vaters, Martin Sturm, der die Darstellung inhaltlich und sprachlich deutlich verbessert hat. Vielen Dank für Deine Hilfe! An der Fehlerkorrektur und der Aufbereitung der Lösungen zu den Übungsaufgaben haben weiterhin Stefanie Burkhardt, Andrea Gauselmann, Natascha Götze, Mathis Kirchner und Philipp Radomski mitgewirkt. Auch ihnen gilt mein Dank. Alle verbleibenden Fehler gehen natürlich auf mein Konto.

Leipzig, im Juli 2016

Bodo Sturm

*IMAGINE a world without statistics.
Governments would fumble in the dark,
investors would waste money and
electorates would struggle to hold their political leaders to account.
The Economist, 25.2.2012*

1 Einführung

Zu Beginn dieser Einführung soll zunächst geklärt werden, was man unter Statistik versteht. Im Anschluss daran werden einige grundlegende Begriffe in der Statistik erläutert. Schließlich wird die Statistiksoftware R vorgestellt. Mit R wird im Rahmen dieses Buchs gearbeitet.

1.1 Was ist Statistik?

Der Begriff *Statistik* beschreibt grundsätzlich zwei unterschiedliche Aspekte. Erstens, in einer umgangssprachlichen Bedeutung, die Zusammenstellung von Daten, die bestimmte Bereiche der menschlichen Zivilisation beschreiben. Dies kann die Entwicklung der Bevölkerung oder des Bruttoinlandsprodukts eines Landes sein, die Umsatzstatistik eines Unternehmens oder das Wählerverhalten bei einer Abstimmung. In diesem Sinne ist auch das Churchill zugeschriebene Zitat *“I only believe in statistics that I doctored by myself“* („Ich glaube nur an Statistiken, die ich selbst gefälscht habe“) zu verstehen. Statistiken sind also mit dem Verdacht der Fälschung belastet – dies ist natürlich nicht Gegenstand dieses Buchs! Zweitens, die Gesamtheit der Methoden zur Beschaffung und Auswertung von statistischen Daten. Diese *statistische Methodenlehre* ist eine wissenschaftliche Disziplin und Gegenstand dieser Einführung. Die Auswertung unterteilt man im Allgemeinen in Analyse und Interpretation. Einer wissenschaftlichen Betrachtung sind vor allem die Analysemethoden zugänglich. Sie sind daher auch unser Schwerpunkt. Wesentliche Methoden wie Lagemaße, Streuungs- und Zusammenhangsmaße und lineare Regression werden in diesem Buch vorgestellt. Hierzu zählen auch Methoden zur Analyse von Stichproben wie Konfidenzintervalle und Hypothesentests. Die Antwort auf die Frage „Was ist Statistik?“ könnte man also folgendermaßen formulieren: Statistik beschreibt einerseits die Ergebnisse statistischer Arbeit und andererseits das methodische Vorgehen zur Beschaffung und Auswertung von Daten über Massenphänomene. Die Statistik als wissenschaftliche Disziplin stellt für letzteres das notwendige Instrumentarium bereit.

Statistik lässt sich unterscheiden in (i) *deskriptive* oder beschreibende Statistik und (ii) *induktive* oder schließende Statistik. Im Fall der deskriptiven Statistik werden die relevanten Daten der zu untersuchenden Grundgesamtheit gesammelt und ausgewertet. Alle Aussagen beziehen sich nur auf diese Daten: Hochrechnungen oder Verallgemeinerungen auf eine größere Datenmenge sind nicht zulässig. Im Fall der induktiven Statistik werden Daten nur von einem i.d.R. über einen Zufallsprozess ausgewählten Teil der Grundgesamtheit beschafft. Von dieser Stichprobe schließt man mit mathematischen Methoden wie der Wahrscheinlichkeitsrechnung und Grenzwertsätzen auf die Grundgesamtheit, z.B. bei Meinungsumfragen oder Materialprüfungen. Man spricht in diesem Fall auch von *Inferenz*.