

Einführung

Durch die Entwicklung umfassender statistischer Berichtssysteme im vergangenen Jahrhundert – zum Beispiel in der Volkswirtschaftslehre: Entwicklung der Volkswirtschaftlichen Gesamtrechnungen – erlebte das Arbeiten mit (Massen-)Daten, das heißt die empirische Forschung, einen ungeheuren Auftrieb. Ermöglicht wurde dies durch gewaltige Sprünge der Hardware (gerade in Bezug auf Prozessorleistung und Speicherkapazitäten). In diesem Zusammenhang sind auch die statistischen Softwarehilfestellungen ausgebaut worden.

Eines der bekanntesten Statistikprogramme ist SPSS®. Die Abkürzung SPSS kommt von »Statistical Package for the Social Sciences« (zeitweise stand SPSS im Übrigen auch mal für »Superior Performing Software System«). Zwischen 2009 und 2010 hieß SPSS kurzzeitig PAW = Predictive Analysis SoftWare.

Das inzwischen von IBM® angebotene SPSS wurde von seinen Funktionen her ständig erweitert beziehungsweise verbessert. Dies gilt nicht zuletzt für die Benutzerfreundlichkeit des Programms. Ich habe anfangs der 1990er-Jahre die empirischen (massenstatistischen) Auswertungen im Rahmen meiner Doktorarbeit noch mit einer frühen SPSS-Version auf einem Großrechner – im »Vor-Windows-Zeitalter« – bewerkstelligt (SPSS^X 2.0). Hierbei musste ich höllisch auf Programmierfallen aufpassen. Heutzutage brauchen Sie zur Nutzung von SPSS im Grunde genommen keine (größeren) Programmierkenntnisse mehr, um mittels der SPSS-Bedienoberfläche in einer Geschwindigkeit, die noch vor 25 Jahren undenkbar schien, komplizierte statistische Berechnungen durchzuführen.

Dieser Gewinn an Benutzerfreundlichkeit ist aber zugleich eine Gefahr, verleitet dies doch mitunter zum »Quick-and-dirty-Arbeiten«, soll heißen: zu Auswertungen gewissermaßen »auf Knopfdruck«, ohne dass die Auswertungsergebnisse inhaltlich immer besonders sinnvoll wären. Mit meinem Buch möchte ich helfen, solche »Nonsense-Auswertungen« zu vermeiden. Außerdem möchte ich Hilfestellungen anbieten, wie die mithilfe von SPSS gewonnenen Ergebnisse inhaltlich-sinnvoll interpretiert werden können. Hierbei habe ich mich bemüht, die Funktionalität von SPSS so umfassend wie möglich darzustellen, damit Sie aus diesem Buch den größtmöglichen Nutzen ziehen.

Das Buch bezieht sich auf die SPSS-Version 24.0 mit möglichst vielen Programmpaketen. Trotz des Umfangs dieses Buchs musste eine Auswahl aus den vielfältigen SPSS-Programmelementen getroffen werden. Das hat dazu geführt, dass Erweiterungen in Form bestimmter SPSS-Zusatzmodule – wie zum Beispiel AMOS (= Analysis of Moment Structures) zur Schätzung von Strukturgleichungsmodellen – im Folgenden eher nicht behandelt werden.

Ich würde mich sehr freuen, wenn ich bei Ihnen durch das Buch zumindest ein bisschen Begeisterung auch für solche statistischen Zusammenhänge erwecken würde, die Ihnen am Anfang vielleicht ziemlich kompliziert und eher langweilig erschienen sind.

Über dieses Buch

Sie leben in einer Zeit der Informationsflut: Tagtäglich werden Sie mit vielen Informationen konfrontiert, das heißt mit Daten aus Politik, Wirtschaft, Kultur und so weiter, aber auch aus dem eigenen Alltagserleben. Um in diesem »Wirrarr« den Überblick zu behalten, gilt es, diese Informationen zu sortieren, das heißt: zu strukturieren, um daraus wesentliche Schlussfolgerungen ziehen zu können, das heißt: um diese Informationen interpretieren und für Ihre Zwecke sinnvoll nutzen zu können. Bei all diesen Dingen ist die Statistik ein unentbehrliches Mittel zum Zweck. Aufgrund der Masse an verfügbaren Daten brauchen Sie aus meiner Sicht leistungsfähige Statistiksoftware. Eine solche ist unzweifelhaft das Statistikprogramm SPSS, um das es in diesem Buch geht. Dabei knüpft dieses Buch an das erfolgreiche, schon bestehende Wiley-Buch »SPSS 24 für Dummies« von Felix Brosius an, indem es über die dort vorrangig präsentierten SPSS-»Basics« hinaus gerade auch für spezielle Statistikfragestellungen SPSS-Lösungswege anbietet. Dies betrifft in erster Linie die umfangreiche Behandlung der statistischen Verfahren Regressionsanalyse und Testtheorie, aber auch der Automatisierung von SPSS-Befehlsfolgen. Das lesenswerte Buch von Felix Brosius wird nicht benötigt, um dieses Buch hier zu verstehen.

Dem Vorgenannten Folge leistend, werde ich im Buch Folgendes mit Ihnen besprechen:

- ✓ Es werden zunächst Grundlagen der statistischen Datenverarbeitung vermittelt, natürlich mit speziellem Blick auf SPSS.
- ✓ Danach geht es um das Beschreiben und die Darstellung statistischer Sachverhalte im Rahmen der sogenannten deskriptiven (= beschreibenden) Statistik.
- ✓ Demgegenüber beschäftigt sich die sogenannte induktive (= schließende) Statistik mit dem Schließen aus Stichprobenbefunden auf die jeweilige statistische Gesamtheit, wobei wahrscheinlichkeitstheoretische Grundzusammenhänge bestehen.
- ✓ Um Ihnen den breiten Anwendungsbereich von SPSS beispielhaft zu verdeutlichen und Ihnen das Erstellen eigener SPSS-Programme zu erleichtern, biete ich Ihnen außerdem auch noch – gewissermaßen als Muster – einige spezielle, selbst geschriebene SPSS-Programme für bestimmte Zusammenhänge (zum Beispiel für die Messung sozialer Ungleichheit) an. Dabei wird ausführlich auf die Automatisierung von SPSS-Befehlsfolgen – vor allem in Form sogenannter Makros – eingegangen.

Sie müssen im Übrigen die Kapitel in dem Buch nicht in einer bestimmten Reihenfolge lesen. Wenn Sie an einem Sachverhalt besonders interessiert sind, können Sie an der entsprechenden Textstelle unmittelbar nachschlagen. In diesem Sinne ist das Buch auch eher als Nachschlagewerk und weniger als Lehrbuch konzipiert.

Törichte Annahmen über meine Leser

Dieses Buch bietet sich vor allem für Studierende, wissenschaftliche Mitarbeiter und dergleichen an Hochschulen (Universitäten, Fachhochschulen, Berufsakademien), aber auch in (Politik-)Beratungs- und Forschungsinstitutionen an. Darüber hinaus ist es aber durchaus auch für Praktiker geeignet, die mit nicht immer ganz einfachen empirischen Fragestellungen

konfrontiert sind. Hierbei denke ich an Praktiker aus Industrie, Handel, Verwaltungen und Behörden, aus dem Gesundheitswesen oder aus dem Medienbereich. Auch diese Personengruppen, so ist meine sicherlich nicht unbegründete Hoffnung, sollen durch dieses Buch mit den für Ihre Tätigkeiten wichtigen statistischen Zusatzkenntnissen ausgestattet werden.

Sie, liebe Leserinnen und Leser, müssen dabei gar nicht so viel Vorwissen mitbringen. Es ist im Grunde genommen lediglich ein gewisses statistisches Grundverständnis vonnöten. Zur Wiederholung derartiger Grundlagen sei auf mein Statistikbuch »Statistik für Wirtschafts- und Sozialwissenschaftler für Dummies«, Weinheim 2015: Wiley, verwiesen. Allerdings bin ich überzeugt, dass Sie auch ohne nennenswerte Statistikkenntnisse nachfolgend gut zu recht kommen werden, weil ich mich bemüht habe, an den einzelnen Textstellen kurz zum Verständnis der entsprechenden Zusammenhänge die jeweils notwendigen Statistikausführungen zu machen.

Wichtige Konventionen in diesem Buch

Fachbegriffe werden im Text **fett** hervorgehoben. Zum Teil werden weniger bedeutsame Begriffe beziehungsweise Formulierungen auch *kursiv* hervorgehoben. Die entsprechenden Begriffe finden sich im Übrigen auch im Sachwortverzeichnis, wodurch ein schnelles Auffinden zentraler Begriffe im Text ermöglicht wird.

SPSS-Menüelemente werden in Kapitälchen (das heißt: in kleinen Großbuchstaben) dargestellt, zum Beispiel: DATEI|SPEICHERN. Das heißt: Die übergeordnete Menüebene findet sich vor dem senkrechten Strich, die nachgeordnete Menüebene danach. Ebenfalls großgeschrieben sind im Text SPSS-Syntaxbefehle, zum Beispiel: COMPUTE. Demgegenüber werden Variablenbezeichnungen (zum Beispiel: *variable1*) kleingeschrieben und zudem kursiv dargestellt. Ebenfalls in kursiver Schrift dargestellt, aber großgeschrieben sind im Folgenden Verzeichnisangaben (zum Beispiel: C:/SPSS-BUCH/DATEIEN) und Dateiangaben (zum Beispiel: 123.SAV).

Wie dieses Buch aufgebaut ist

Das vorliegende Buch setzt sich aus fünf Teilen zusammen. Zunächst einmal werden Sie in Teil I mit eher allgemeinen Hilfestellungen der statistischen Datenverarbeitung versorgt, ehe Sie in Teil II die Anwendungsmächtigkeit von SPSS in Bezug auf deskriptive Statistikfragen kennenlernen. Ergänzend kommt dann in Teil III der SPSS-Anwendungsbezug hinsichtlich der induktiven Statistik hinzu. In Teil IV werden unter anderem einige spezielle, selbst erstellte SPSS-Programme diskutiert. Zuletzt folgt, wie bei jedem Dummies-Buch, der Top-Ten-Teil. Letzterer bietet Ihnen Kurzübersichten über SPSS in Form von To-do-Listen, Tipps und so weiter.

Teil I: Grundlagen der statistischen Datenverarbeitung und von SPSS

Hier geht es nach einigen Vorbemerkungen zur Bedeutung von Daten und ihrer Skalierung um die grundlegende Benutzerführung/-steuerung in SPSS. Das heißt: Es werden das Einlesen und die Verarbeitung von Daten mit SPSS angesprochen. Dabei spielen verschiedene Datei- und Variablenoperationen eine Rolle.

Teil II: Deskriptive SPSS-Statistiken

Gegenstand dieses Teils sind beschreibende (deskriptive) SPSS-Statistikoperationen. Nachdem hierbei auf ein- wie mehrdimensionale Häufigkeitsverteilungen inklusive ihrer grafischen Darstellungsmöglichkeiten eingegangen worden ist, werden statistische Zusammenhänge zwischen zwei beziehungsweise noch mehr Variablen untersucht. Dies geschieht unter Bezugnahme auf die Korrelations- und die Regressionsmethodik. Betrachtungen unter anderem zur Cluster- und zur Zeitreihenanalyse schließen Teil II ab.

Teil III: Induktive SPSS-Statistiken

In Teil III geht es im Wesentlichen um das statistische Testen. Dabei werden sowohl Tests auf (Mittelwert-)Unterschiede zwischen Variablen als auch Tests auf bestimmte Verteilungsformen der interessierenden Variablen präsentiert.

Teil IV: SPSS-Syntaxprogrammen für Fortgeschrittene

Zunächst werden in Teil IV Ausführungen zu der Einbindung von Elementen aus den Programmen R und Python in SPSS gemacht. Danach wird auf die Automatisierung von SPSS-Syntaxbefehlen – vor allem in Form von Makros – Bezug genommen. Nach weiteren allgemeinen (zum Teil wiederholenden) Anmerkungen zur Struktur von SPSS-Syntaxprogrammen wird beispielhaft eingegangen auf SPSS-Syntaxprogramme bezüglich der Armuts-/Reichtumsmessung beziehungsweise allgemein der Ungleichheitsmessung.

Teil V: Top-Ten-Teil

Hinzu kommt noch der Top-Ten-Teil, in dem die Highlights von SPSS gelistet angegeben sind. Hierbei wird auf die zehn wichtigsten SPSS-Befehle eingegangen, und es werden die zehn wichtigsten SPSS-Tipps gegeben.

Symbole in diesem Buch

In diesem Buch wird eine Reihe von Symbolen genutzt. Diese Symbole stelle ich Ihnen im Folgenden kurz vor.



Statistik verleitet leicht zu Fehlinterpretationen. Sie kennen sicherlich die Frage: »Wie lügt man am besten mit Statistik?«, oder Sie kennen den Spruch: »Traue keiner Statistik, die du nicht selbst gefälscht hast.« Beim »Achtung-Icon« geht es entsprechend vor allem darum, mit den dort von mir gegebenen Informationen zu vermeiden, dass Sie in entsprechende Statistikfallen tapen.



Manchmal ist das »Achtung-Icon« noch zu vorsichtig angelegt. Dies gilt etwa dann, wenn Sie durch Unachtsamkeit Gefahr laufen, Ihre kompletten Daten zu vernichten. In solchen »Katastrophen«-Fällen ist eine besonders nachhaltige Warnung angebracht.



Um Sie nicht nur zur Vorsicht zu mahnen (mittels der beiden eben genannten Symbole), sondern um Ihnen auch Tipps zu geben, wie Sie bestimmte Fallstricke mit SPSS umgehen beziehungsweise wie Sie möglichst effizient mit SPSS arbeiten, werden Ihnen durch das »Tipp-Icon« hoffentlich hilfreiche Anmerkungen seitens meiner Wenigkeit gegeben.



Viele Dinge lernen sich leichter, wenn sie mit Beispielen unterlegt werden. Dies veranschaulicht an den entsprechenden Stellen die dortigen Zusammenhänge. Hierzu dienen im Buch viele mehr oder weniger kleine SPSS-Beispieldatensätze. Sie sind im Internet unter <http://www.wiley-vch.de/ISBN978-3-527-71367-7> abrufbar. Damit das »Beispiel-Symbol« angesichts der vielen Beispiele im Buch nicht inflationär eingesetzt wird, finden Sie es nur an besonders prägnanten Stellen, wo komplexe Zusammenhänge unbedingt einer zusätzlichen beispielhaften, eher speziellen Erläuterung bedürfen.



Unter diesem Symbol werden wichtige statistische Sachverhalte beziehungsweise Kennziffern wie zum Beispiel Mittelwerte erläutert, damit die zugehörigen SPSS-Operationen besser nachvollzogen werden können. Wem das aber zu viel »theoretischer Ballast« ist, der kann über diese theoretischen Hintergrundinformationen hinweglesen, ohne dass seine »Ausbildung« zum SPSS-Experten dadurch grundsätzlich infrage gestellt wäre.

Wie geht es weiter im Buch?

Es folgt Teil I. In diesem Teil werden die Grundlagen der statistischen Datenverarbeitung mit SPSS vermittelt. Zunächst wird dort der Aufbau von SPSS-Datendateien beschrieben, ehe das Zusammenlegen und die Bearbeitung von Datendateien skizziert werden. In diesem Zusammenhang wird auch auf spezielle SPSS-Syntaxbefehle eingegangen.