

Book Reviews

Heinz GROHMANN, Walter KRÄMER, und Almut STEGER (Herausgeber) (2011). **Statistik in Deutschland; 100 Jahre Deutsche Statistische Gesellschaft**. Springer Verlag, Heidelberg, Dordrecht, London, New York, 262 Seiten, gebunden, ISBN 978-3-642-15634-2 (€ 46.21)

Die Deutsche Statistische Gesellschaft (DStatG) wurde im Juli 1911 gegründet. Aus dem Anlass dieses Jubiläums erschien dieser Sammelband, der die hundertjährige Geschichte dieser wissenschaftlichen Vereinigung nachzeichnet.

Die 17 Beiträge sind drei großen Abschnitten zugeordnet: Der erste Teil ist der DStatG im Wandel der Zeit gewidmet und damit gleichzeitig der Entwicklung der Statistik in Deutschland selbst. Der Schwerpunkt der Darstellung liegt deutlich bei der angewandten, bei der amtlichen Statistik. Im zweiten Teil werden die wichtigsten Arbeitsgebiete der Gesellschaft vorgestellt. Der dritte Teil bringt zwei prominente Beispiele, bei denen sich die DStatG aktiv in der politischen Auseinandersetzung engagiert hat: Der Beitrag von Stäglin beschäftigt sich mit den intensiven Bemühungen der DStatG nach der Wiedervereinigung, als es um die Neuausrichtung der Statistik in den neuen Bundesländern, um die Rolle des Faches Statistik an den Universitäten, die Integration der Statistiker in die Gesellschaft und nicht zuletzt um die Sicherung der Datenbestände der ehemaligen DDR ging. Der Artikel von Grohmann beleuchtet den wichtigen Beitrag der DStatG zur Durchführung der Volkszählung 1987 und zur Entwicklung des Mikrozensus und zeichnet den Paradigmenwechsel zu einem registergestützten Zensus nach. Recht ausführliche Literaturhinweise ergänzen die meisten der Beiträge. 13 sehr informative Anhänge runden den Band ab.

Wie es im Vorwort formuliert ist, wendet sich das Buch an “Freunde der Statistik”. Für Freundinnen und Freunde der Statistik in Österreich ist der Sammelband aus zwei Gründen lesenswert:

Zum einen wird der Anteil österreichischer Experten an der Entwicklung der Statistik in Deutschland dokumentiert. In der Liste der Gründungsmitglieder der DStatG (Anhang 2) scheinen mehrere Österreicher auf. In der keineswegs langen Liste der Ehrenmitglieder sind mit Viktor von Mataja, Walter Breisky, Otto von Zwiedeneck-Südenhorst und Wilhelm Winkler mehrere Österreicher vertreten. Österreicher traten nicht zuletzt als (Mit-) Herausgeber von Einzelschriften der DStatG auf, wie im Falle des Bandes “Anwendungen der Matrizenrechnung auf wirtschaftliche und statistische Probleme” 1959.

Die Lektüre lohnt aber zum anderen vor allem deshalb, weil immer wieder jene Grundthemen angesprochen werden, welche sich wie ein roter Faden auch durch die Geschichte der Österreichischen Statistischen Gesellschaft (ÖSG) ziehen und die jeden an Statistik Interessierten beschäftigen. Die wichtigsten dieser Problembereiche sind:

Das Verhältnis zwischen der primär methodisch orientierten, akademischen Statistik und der angewandten Substanzstatistik, die oft amtliche Statistik ist. Die mathematisch-statistisch arbeitenden Statistiker fühlten sich lange in der DStatG nicht ausreichend repräsentiert (s. Beitrag von Strecker, Bassenge-Strecker), erst 1972 wurde das erste Mal ein Hochschullehrer zum Vorsitzenden gewählt, es kam zu einer verbesserten Partnerschaft, die Kooperation mit anderen Disziplinen wurde intensiviert

(s. Beiträge von Grohmann, und Kauermann, Mosler). 2005 wurde dann die Deutsche Arbeitsgemeinschaft Statistik als Plattform wissenschaftlicher Gesellschaften, die jeweils eine große Zahl von Statistikern zu ihren Mitgliedern zählen, gegründet (s. Kauermann, Mosler).

Das nicht immer reibungslose Verhältnis spiegelte sich vor allem in der Zeitschrift der DStatG wider, die älter als die Gesellschaft selbst ist. Eine Balance zwischen methodisch/theoretisch ausgerichteten und anwendungsorientierten Artikeln zu gewährleisten, die eine Gruppe für die Arbeiten der anderen Gruppe zu interessieren, wurde offensichtlich immer schwieriger. Die Aufspaltung der Zeitschrift in zwei unterschiedlich ausgerichtete Journale (s. Beitrag von Rinne) illustriert die Schwierigkeit, dem gesamte Spektrum der statistischen Aktivitäten Raum zu geben und die Sprachlosigkeit zwischen den Teilbereichen nicht zu groß werden zu lassen.

Welche Früchte andererseits eine gute Kooperation erbringen kann, wird nicht zuletzt durch die Beispiele im Beitrag von Krug, Schmidt und Wiegert zur Methodik und Qualität statistischer Erhebungen und jenem von Wilrich zur Statistik in Naturwissenschaft und Technik illustriert.

Ausbildung; Kampf dem Innumeratentum Der fachlichen Ausbildung als zentrale Aufgabe der Gegenwart ist ein eigenes Kapitel (Krämer, Schmerbach) gewidmet. Die Notwendigkeit, die allgemeine “Bildung in Statistik” zu verbessern wird darüber hinaus in mehreren Beiträgen angesprochen und im Artikel von Kauermann und Mosler zu Recht als eine der großen Herausforderungen bezeichnet, der sich eine Statistische Gesellschaft zu stellen hat.

Verbesserung der Akzeptanz der Statistik in der Öffentlichkeit Ein breites fachlich fundiertes Verständnis für die Möglichkeiten und Grenzen statistischer Aussagen ist eine Voraussetzung zur Schaffung einer realistischen Einordnung statistischer Resultate in der Öffentlichkeit. Derzeit schwankt die Einschätzung der Statistik unreflektiert zwischen “Lüge und Manipulation” (wie es Kauermann und Mosler formulieren) und einer leichtfertigen Überinterpretation, wie sie etwa im Falle der beliebten Rankings geschieht.

Stellung der Statistik im Staat; Unabhängigkeit der amtlichen Statistik Die Folgen des Missbrauchs statistischer Daten für Verwaltungszwecke und die Konsequenzen des Fehlens der Unabhängigkeit wird in den Beiträgen zur Rolle der Statistik während der Nazidiktatur (Wilke) und in der DDR (im Beitrag zu den Herausforderungen aus der Wiedervereinigung von Stäglin) verdeutlicht. Die stets präsente Spannung zwischen dem Informationsbedürfnis einerseits und der informationellen Selbstbestimmung der Bürger ist auch Gegenstand des Kapitels über die Volkszählungen von Grohmann.

Alle diese (und viele andere) Themen werden vor allem im jeweiligen historischen Kontext angesprochen. Mancher Leser hätte sich von einer solchen Festschrift zusätzlich eine grundsätzlichere Auseinandersetzung, eine systematische Diskussion erhofft.

Dies gilt auch für die mögliche Rolle einer Statistischen Gesellschaft. Auf die im Vorwort aufgeworfene Frage “Was kann und will die Statistik, was trägt sie zum Wohl unseres Gemeinwesens und zum Funktionieren unserer Wirtschafts- und Sozialgestaltung

im Besonderen bei. Und was konkret war und ist dabei die Rolle der Deutschen Statistischen Gesellschaft?“ wird in mehreren Beiträgen immer wieder punktuell eingegangen. Ein eigener programmatischer Beitrag - etwa in der Art, wie ihn der ehemalige Vorsitzende der DStatG Prof. Reiner Stäglich in einem Vortrag im April 2010 im Rahmen der ÖSG versucht hat — fehlt aber.

Anmerken verdient auch, welche der immer wiederkehrenden “großen Themennicht oder kaum behandelt werden. Auf die eminente und unmittelbare Bedeutung statistischer Ergebnisse für das Alltagsleben (z.B. von Preisindizes in Wertsicherungsklauseln) und in der Politik (Ergebnisse von Volkszählungen für Mandatszahlen und Finanzausgleich auf nationaler Ebene, Resultate der Volkswirtschaftlichen Gesamtrechnung für EU-Beitragszahlungen, Maastricht-Kriterien auf internationaler Ebene, um nur wenige Beispiele zu nennen), wird nicht eingegangen.

Auch der Wandel, der die amtliche Statistik in Europa in den letzten Jahrzehnten erfasst hat, findet keine adäquate Würdigung. Neben dem Zwang vermehrt Verwaltungsdaten heranzuziehen, wird die Reorientierung “auf Europa hin nicht entsprechend behandelt. Wohl werden einige Vorgaben durch EU Rechtsnormen kurz erwähnt (so u.a. in den Beiträgen von Kauermann, Mosler; Grömling, Scheinost, Brachinger und Grohmann), der nicht gut informierte Leser könnte dennoch den irrigen Eindruck erhalten, über die Konzepte der wichtigsten Wirtschafts- und Sozialstatistiken wäre nach wie vor auf nationaler Ebene zu befinden. Das Ausblenden einer Entwicklung, die zu einem markanten Funktionswandel weiter Bereiche der amtlichen Statistik geführt hat, verwundert um so mehr, als ein ehemaliger Stellvertretender Vorsitzender der DStatG, Walter Rademacher als Generaldirektor des Statistischen Amtes der Europäischen Union (EUROSTAT) amtiert.

Trotz dieser kleinen Einwände bietet die Festschrift auch für den weniger an historischen Hintergründen Interessierten genug Lesenswertes und viele Anregungen. Zu hoffen bleibt, dass die noch intensivere, systematische Befassung mit den in der Festschrift nur andiskutierten Grundfragen nicht bis zum nächsten großen Jubiläums - wie etwa 100 Jahre ÖSG – aufgeschoben wird.

Josef Richter, Wien

Johann BACHER, Andreas PÖGE und Knut WENZIG (2010). **Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren** (3. Auflage). Oldenbourg Verlag München, 538 Seiten, Broschur, ISBN 978-3-486-58457-8 (€ 39.80)

Das vorliegende Buch gibt eine Einführung in Dimensionsreduktion und Clusteranalyse für Wirtschafts- und Sozialwissenschaftler. Im Vergleich zur 15 Jahre alten zweiten Auflage hat sich der Erstautor vier Koautoren „an Bord geholt“, davon zwei für das gesamte Buch und zwei weitere für zwei konkrete Kapitel. Drucktechnisch wurde das Manuskript auf L^AT_EX umgestellt.

Nach einem Einleitungs- und Übersichts Kapitel beschäftigt sich Teil 1 des Buches sehr ausführlich auf rund 100 der insgesamt ca. 500 Seiten mit „unvollständigen Clusterverfahren“, die in der Literatur je nach wissenschaftlicher Community oft auch als „graphische

Clusterverfahren“ bezeichnet werden: Die Daten werden durch Methoden der Dimensionsreduktion ins zweidimensionale projiziert wo die Cluster dann manuell identifiziert werden. Als Dimensionsreduktionsverfahren werden wie üblich multiple Korrespondenzanalyse, nichtmetrische Skalierung, Haupt- und Faktorenanalyse vorgeschlagen.

Rund die Hälfte des Buches ist dann in Teil 2 den klassischen hierarchischen und partitionierenden Clusterverfahren gewidmet. Hier finden sich Kapitel zu Gewichtung und Transformation von Variablen, Distanz- und Ähnlichkeitsmaße für unterschiedliche Skalenniveaus, sowie die wichtigsten Algorithmen wie verschiedene Linkage Methoden für hierarchisches Clustern oder das k-Means Verfahren.

Teil 3 beschäftigt sich mit modellbasierter Clusteranalyse sowohl für multivariat normalverteilte Variablen wie auch nominale und ordinale Meßniveaus. Auch finite Mischungen von Regressionsmodellen werden betrachtet, methodisch orientiert sich dieser Teil eng an der Software Latent GOLD. Der abschließende 4. Teil behandelt spezielle Anwendungsfragen, etwa wann welches Verfahren zum Einsatz kommen soll.

Das Buch ist sehr praxisorientiert, alle Verfahren werden anhand von konkreten Beispielen eingeführt und diskutiert, die entsprechenden Datensätze können von der Homepage des Buches heruntergeladen werden. Bei allen Verfahren werden immer auch deren Limits mitbesprochen, Kennzahlen und Verfahren zur Evaluation der Güte einer Clusterlösung ziehen sich wie ein roter Faden durch das Buch. In anderen Lehrbüchern wird gerne verschwiegen, daß nur weil Software eine Partition findet, diese in hochdimensionalen Datensätzen noch lange keine sinnvolle Klassifikation der Daten darstellen muß.

Als Schwäche des Buches empfinde ich, daß in mehreren Bereichen durch die vielen Details und Beispiele der große Blick für das Ganze etwas verloren geht. So werden die verschiedenen Linkage-Methoden in mehreren Einzelkapitel getrennt eingeführt. Daß Linkage-Methoden letztlich nur unterschiedliche Distanzen zwischen Mengen von Beobachtungen darstellen, geht dabei unter. Manche als „Neuerungen der letzten 15 Jahre“ verkauften Ergänzungen des Buches sind schon viel länger bekannt, z.B. Start von k-Means mit mehreren zufälligen Initialisierungen oder Verallgemeinerung des k-Means-Algorithmus für nicht-Euklidische Distanzen. Die einzige Neuerung hier ist die breitere Verfügbarkeit in Statistiksoftware, weil durch die massiv gestiegene Leistungsfähigkeit der Computer dies auch im Alltag gerechnet werden kann. Befremdlich ist auch der Vergleich verschiedener Software-Pakete. Es werden ordinale Daten simuliert, für Wirtschafts- und Sozialwissenschaften ohne Frage ein sehr wichtiges Skalenniveau. Geclustert wird dann aber mit zahlreichen auf metrische Variablen ausgelegten Softwarepaketen, in R z.B. das auf Normalverteilungen spezialisierte Erweiterungspaket *mclust*. Den Vergleich „gewinnt“ dann wenig überraschend SPSS Latent GOLD, das als einziges Verfahren *im Vergleich* explizite Modelle für Ordinaldaten enthält.

Trotz dieser vereinzelt Schwächen kann das Buch aber durchaus als Einstieg in die Welt der Clusteranalyse empfohlen werden. Die wichtigsten Verfahren werden alle detailliert und praxisorientiert vorgestellt, auf kritische Distanz zu vorschneller Interpretation von Softwareausgaben und sorgfältige Modelldiagnostik wird großer Wert gelegt.

Friedrich Leisch
Institut für angewandte Statistik und EDV
Universität für Bodenkultur Wien