

# 1 Einführung

Kein Wissenschaftler, kein Politiker und kein Unternehmer kann es sich heute leisten, seine Entscheidungen ohne statistisch belegte Informationen zu treffen. Dabei hat die Statistik in der breiten Bevölkerung nicht unbedingt einen guten Ruf. Als trockene Beschäftigung mit vielen Zahlen verschrien, wird ihr zusätzlich noch ein unredlicher Charakter nachgesagt. Der ehemalige britische Premierminister Benjamin Disraeli soll das einmal so auf den Punkt gebracht haben: „There are three kinds of lies: lies, damned lies, and statistics.“ Hinzu kommt, dass durch unsachgemäßes Interpretieren statistischer Erkenntnisse häufig auch unbeabsichtigt falsche Schlussfolgerungen gezogen werden. Nicht immer sagt einem der gesunde Menschenverstand, dass da etwas nicht stimmen kann, wie in dieser nicht wirklich passierten Episode: Ein Brummifahrer hält sich in einer Raststätte gerade mit einem Kännchen Kaffee munter, als durch das Radio die Meldung kommt: „Neueste statistische Veröffentlichungen besagen, dass bei 10 % aller Verkehrsunfälle Alkohol im Spiel war.“ Der Fahrer überlegt: „Das bedeutet doch, dass bei 90 % aller Unfälle ...? Herr Wirt, bitte zwei Doppelte!“

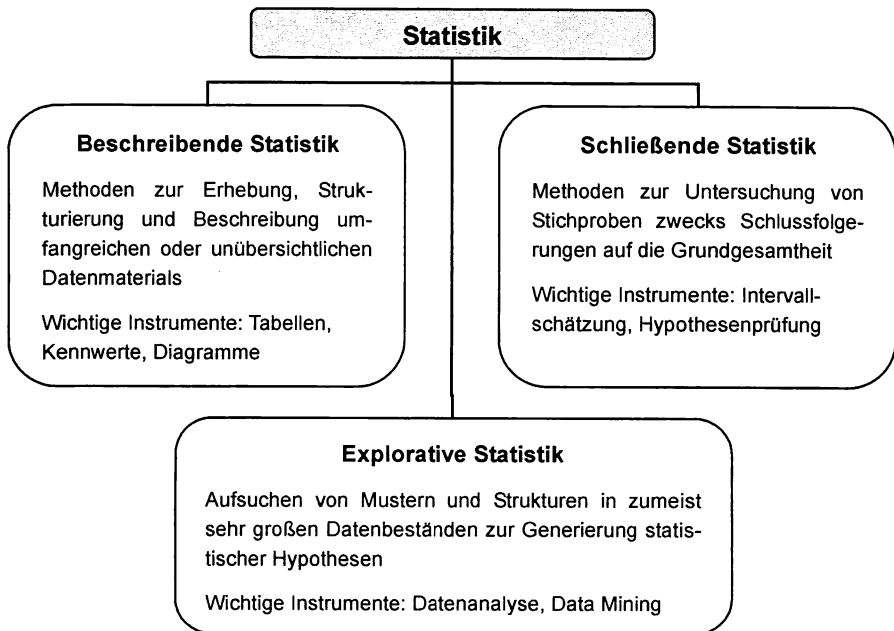
Der Ursprung des Wortes Statistik liegt im lateinischen „status“ (Zustand, Staat) und dem italienischen „statista“ (Staatskundiger, Politiker). So bedeutete Statistik in der Mitte des 17. Jahrhunderts die Lehre von den Staatsmerkwürdigkeiten. Damit waren z. B. die Bevölkerung, das Heer, die landwirtschaftlichen Flächen und die Gewerbe gemeint. Erst im 19. Jahrhundert erhielt das Wort die heutige Bedeutung des Sammelns und Analysierens von Daten.

Obwohl die Statistik eine relativ junge Wissenschaft ist, reichen ihre Ursprünge doch mindestens fünftausend Jahre zurück. Die Babylonier hinterließen kleine Tontafeln mit Tabellen über landwirtschaftliche Erträge und getauschte Waren. Die Ägypter nahmen 2500 v. Chr. alle zwei Jahre Zählungen des Geldes, der Felder und der Bevölkerung vor. Im alten Rom fanden innerhalb von 500 Jahren insgesamt 69 Volkszählungen statt, also im Durchschnitt alle sieben Jahre eine. Zum Vergleich, die letzte Volkszählung in der Bundesrepublik Deutschland war im Jahre 1987. Karl der Große ließ im 8. Jahrhundert Güter- und Besitzverzeichnisse anlegen. Später fällt dann allerdings auf, dass im Europa des Mittelalters keine Volkszählungen vorgenommen wurden. Die Ursache kann man im Alten Testament nachlesen. Zwar werden im 4. Buch Moses die Ergebnisse zweier Volkszählungen ausführlich beschrieben. Im 2. Buch Samuel jedoch wird König David durch den Herrn schwer bestraft, weil er sein Kriegsvolk gezählt hatte. Ihr modernes Gesicht erhielt die Statistik Ende des 19. und Anfang des 20. Jahrhunderts mit der zunehmenden Anwendung mathematischer Methoden, vor allem der Wahrscheinlichkeitsrechnung, in der Physik, der Biologie und den anderen

Naturwissenschaften. Bahnbrechenden Einfluss auf die Entwicklung der Statistik zur modernen Wissenschaft hatten die Engländer K. Pearson<sup>1</sup> und R. A. Fisher<sup>2</sup>.

Die Statistik wird gegenwärtig gern in drei Teilbereiche gegliedert (vgl. Abbildung 1-1): die **beschreibende** (auch: deskriptive), die **schließende** (auch: induktive) und die **explorative** Statistik, die Methoden sowohl der beschreibenden als auch der schließenden Statistik benutzt.

Abbildung 1-1: Teilbereiche der Statistik



<sup>1</sup> Karl Pearson, 1857 - 1936.

<sup>2</sup> Sir Ronald Aylmer Fisher, 1890 - 1962.

# Vorwort

In der Praxis werden umfassendere statistische Auswertungen heute ausnahmslos mittels entsprechender Software durchgeführt. Dabei muss man nicht zwangsläufig auf spezielle Statistikprogramme zurückgreifen, da selbst in vielen Standardanwendungen, wie beispielsweise in Microsoft Excel, entsprechende Methoden implementiert sind. Dies hat für den Anwender dieser Software zwei entscheidende Vorteile: Zum einen wird ihm die Rechenarbeit abgenommen und zum anderen können aufgrund heutiger Rechnerleistungen auch immense Datenmengen problemlos analysiert werden. Dennoch muss der Anwender selbst entscheiden, welche Methode zur Auswertung der Daten geeignet ist, welche Prämissen gegebenenfalls zu beachten und wie die Analyseergebnisse zu bewerten und im Hinblick auf die zugrundeliegende Problemstellung zu interpretieren sind. Genau an dieser Stelle setzt das vorliegende Buch an. Bei der Darstellung der Methoden wird besonderer Wert darauf gelegt, dass auch die jeweiligen Voraussetzungen, das Anwendungsspektrum und die entsprechenden Ergebnisinterpretationen nicht zu kurz kommen. Darüber hinaus werden die Methoden anhand zahlreicher Beispiele erläutert, um das grundlegende Verständnis zu vertiefen.

Die in der europäischen Hochschullandschaft zurzeit stattfindende Umstellung auf das gestufte Bachelor-Master-System verlangt, Studierende schon in sechs bis sieben Semestern für ihren angestrebten Beruf zu qualifizieren. Die erfordert eine Straffung der Ausbildung und die Konzentration auf grundlegende inhaltliche Schwerpunkte in den neuen Bachelorstudiengängen. Das vorliegende Lehrbuch soll einen Beitrag dazu leisten. Es richtet sich vor allem an Studierende wirtschafts- und sozialwissenschaftlicher sowie verwandter Bachelorstudiengänge, in denen grundlegende Kenntnisse statistischer und datenanalytischer Methoden benötigt werden.

In diesem Buch werden neben der deskriptiven und induktiven Statistik insbesondere

analyse in Form der Klassifikation, Repräsentation und Identifikation auseinander. Abschließend geht der Teil 4 des Buches noch auf den Bereich des Data Mining ein. Mit den Kapiteln 16 bis 18 werden dabei der Gegenstand des Data Mining, der Ansatz der Assoziationsanalyse sowie Entscheidungsbäume vorgestellt. Vorausgesetzt werden dabei mathematische und wahrscheinlichkeitstheoretische Grundkenntnisse, wie Sie üblicherweise in allen oben angesprochenen Studiengängen in den ersten beiden Semestern vermittelt werden. Dennoch wurden insbesondere wahrscheinlichkeitstheoretische Grundlagen im Anhang A dieses Buches zusammengestellt, auf die an den entsprechenden Textstellen auch verwiesen wird. Dadurch soll die Lektüre erleichtert werden, da die notwendigen Grundkenntnisse direkt nachgeschlagen und vertieft werden können. Die unverzichtbaren statistischen Tafeln, die von uns mit Hilfe von Microsoft Excel erstellt wurden, befinden sich im Anhang B.

Wir möchten dieses Vorwort nicht schließen, ohne uns bei all denjenigen recht herzlich zu bedanken, die an der Entstehung dieses Buches mitgewirkt haben. An erster Stelle ist hier Herr Dipl.-Kfm. Christian Kornprobst zu nennen, der das Kapitel 18 verfasst und uns bei der Erstellung der Kapitel 16 und 17 unterstützt hat. Besonderer Dank geht auch an den Gabler-Verlag und in diesem Zusammenhang vor allem an Frau Kramer und Frau Hinrichsen, die mit uns bis zur endgültigen Fertigstellung des Manuskripts verständnisvoll und jederzeit hilfsbereit zusammengearbeitet haben.

Udo Bankhötter  
Jürgen Vogel