

Hans-Joachim Mittag

# Statistik

Eine Einführung  
mit interaktiven Elementen

*5. Auflage*

freier  
Zugang  
zu interaktiven  
Elementen  
inklusive

EBOOK INSIDE



Springer Spektrum

# 1 Statistik, Daten und statistische Methoden

Anhand von Beispielen aus verschiedenen Lebensbereichen und Anwendungsfeldern wird illustriert, welche Bedeutung der Statistik heute zukommt. Statistik wird als eine Wissenschaft definiert, die Methoden zur Gewinnung von Daten und zum Lernen aus Daten bereit stellt. Es wird sichtbar gemacht, welches breite Aufgabenspektrum die Statistik umfasst und welche Teilbereiche sich unterscheiden lassen. Dabei wird deutlich, dass statistische Methodenkompetenz eine in immer mehr Arbeitsfeldern benötigte Schlüsselqualifikation darstellt, die auch im privaten Bereich nützlich ist.



Vorschau auf das Kapitel

## 1.1 Statistik im Alltag, in Politik und Gesellschaft

Die **Statistik** ist eine noch junge Wissenschaft, die alle Lebensbereiche durchdringt. Jeder von uns ist heute im Alltag mit einer Fülle von Daten und Visualisierungen von Daten konfrontiert, die uns über verschiedene Kanäle erreichen. Wenn wir am Morgen das Radio einschalten oder die Tageszeitung aufschlagen, erfahren wir etwas über die Entwicklung von Aktienkursen, über Trends auf dem Arbeitsmarkt oder über Ergebnisse der von der OECD getragenen PISA-Studien, die auf eine vergleichende Bewertung nationaler Bildungssysteme abzielen. Abends können wir im Fernsehen die Ziehung der Lottozahlen verfolgen oder uns über den Stand des aktuellen ZDF-Politbarometers informieren. Im Internet kann man gezielt nach Daten aller Art suchen, z. B. nach statistischen Informationen zu Migrationsströmen in Europa oder zur Entwicklung der Erwerbstätigkeit in Deutschland. Zugleich wird die Online-Präsentation von Daten immer benutzerfreundlicher. Dies gilt insbesondere für Daten der amtlichen Statistik – man studiere etwa die unter dem Etikett „Statistik anschaulich“ zusammengefassten interaktiven Anwendungen des **Statistischen Bundesamts**.

Statistische Daten im Alltag



Video von Eurostat zum Europäischen Statistiktage 2016

Die große gesellschaftliche Relevanz der Statistik spiegelt sich auch darin wider, dass das Jahr 2013 zum Internationalen Jahr der Statistik, der 20. Oktober 2015 zum Weltstatistiktage und der 20. Oktober 2016 zum Europäischen Statistiktage erklärt wurden. Zu diesen Anlässen gab es zahlreiche Veranstaltungen von Universitäten, Statistikämtern, Unternehmen oder internationalen Institutionen.

**Beispiel 1.1: Analyse der Wählerstimmung und Trendidentifikation**

Seit 1977 wird regelmäßig im Auftrag des ZDF eine Stichprobe von Wählern in Deutschland nach ihrer aktuellen Parteipräferenz, nach der Bewertung der bekanntesten Politiker und nach ihrer Haltung gegenüber aktuellen Entwicklungen in Politik und Gesellschaft befragt. Die Ergebnisse der als ZDF-Politbarometer bezeichneten Erhebung werden jeweils über Fernsehen und Internet verbreitet. Da die Personen in der Stichprobe so ausgewählt werden, dass sie als repräsentativ für die gesamte Bevölkerung anzusehen sind, können aus den Befragungsergebnissen Aussagen für alle Wähler in Deutschland abgeleitet werden. Aufgrund der Regelmäßigkeit der Befragungen gewinnt man nicht nur Aussagen für einen bestimmten Zeitpunkt, sondern Informationen zu langfristigen Trends und Veränderungen der politischen Stimmung.

Eine Frage des ZDF-Politbarometers, die sog. „Sonntagsfrage“, projiziert die aktuelle Parteipräferenz der befragten Wähler auf die nächste Bundestagswahl. Die Frage lautet: „Welche Partei würden Sie wählen, wenn am kommenden Sonntag Bundestagswahl wäre?“. Die Antworten zur „Sonntagsfrage“ vom 16. Oktober 2009 werden in diesem Manuskript mehrfach zur Illustration der Anwendung statistischer Konzepte herangezogen, etwa bei der Vorstellung grafischer Instrumente der beschreibenden Statistik oder bei der Analyse von Merkmalszusammenhängen.

Statistische Verfahren  
im Wirtschaftsleben

Die Statistik spielt auch für Unternehmen eine wichtige Rolle. Bei industriellen Fertigungsprozessen und im Dienstleistungsbereich werden statistische Verfahren schon in der Designphase eines Produkts oder einer Serviceleistung eingesetzt, um Fehler zu vermeiden und Kundenzufriedenheit zu sichern. Mängel können zu imageschädigenden Rückrufaktionen und Gewährleistungsprozessen führen und die Existenz selbst größerer Unternehmen bedrohen. Statistische Instrumente sind auch in der Markt- und Werbeforschung nicht mehr wegzudenken. Marktforschungsinstitute ermitteln Marktanteile und Marktpotenziale, etwa über computergestützte Telefoninterviews. Die Einschaltquoten für Radio- und Fernsehsender werden auf Stichprobenbasis geschätzt und determinieren dann die Preise von Werbespots. Banken setzen statistische Modelle bei Entscheidungen über die Vergabe von Krediten und bei der Analyse von Kapitalmarktdaten ein. Große Lebensmittelkonzerne werten die an den Kassen gesammelten Scannerdaten aus, um Auffälligkeiten zu identifizieren, etwa die aktuellen „Renner“ und Ladenhüter. Pharmahersteller benötigen statistische Testverfahren, um die bei der Zulassung neuer Medikamente geforderten Wirksamkeits- und Unbedenklichkeitsnachweise zu erbringen. Statistische Testverfahren werden auch eingesetzt, um die Wirksamkeit psychologischer Interventionen zu evaluieren, z. B. die Effekte psychotherapeutischer Maßnahmen.

Die Statistik erfüllt für viele Wissenschaften eine wichtige Servicefunktion. In der *Soziologie*, der *Psychologie* oder auch der *Medizin* stützen sich Publikationen in den Fachzeitschriften maßgeblich auf Daten und deren statistischer Analyse. Die Versuchsplanung, bei der es u. a. um die planmäßige Variation von Einflussfaktoren geht, ist ein weiteres Beispiel für den fächerübergreifenden Einsatz statistischer Methoden. Sie ist ein wichtiges Feld der experimentellen Psychologie und zugleich auch der *Ingenieurwissenschaften* – man denke an Experimente in der *Sozialpsychologie* zur Untersuchung von Motivationsstrukturen bei ehrenamtlich tätigen Personen oder an Belastungstests bei der Erforschung neuer Verbundwerkstoffe für Kraftfahrzeuge. Statistische Instrumente des Qualitätsmanagements werden in der *Bildungspädagogik* sowie in der *Gesundheitsökonomie* bei der Steuerung von Schulentwicklungen und Krankenhausbelegungen verwendet. Weitere Anwendungsfelder der Statistik sind die Beschreibung von Zufallsprozessen in der *Physik* (u. a. Brownsche Bewegung, radioaktiver Zerfall), die Berechnung von Lebensversicherungsprämien in der *Versicherungsmathematik*, die Verwendung von Zeitreihenmodellen in der *Kapitalmarktforschung*, die Analyse von Querschnitts- und Paneldaten in den *Wirtschaftswissenschaften*, die Modellierung von Wachstumsprozessen in der *Biologie* oder die Gewinnung empirisch fundierter Aussagen zum Zustand von Wäldern in den *Umweltwissenschaften*.

Statistik ist  
fachübergreifend



Video von SAS  
zum Jahr der  
Statistik 2013



Abb. 1.1: Qualitätskontrolle bei der Tensideherstellung (Säurezahlbestimmung und Eingabe für die statistische Auswertung); Quelle: Evonik Industries AG, Essen

Die Statistik spielt auch bei der *Politikplanung* und bei der Erfolgsbewertung von Politik eine gewichtige Rolle. Harmonisierte, d. h. über Ländergrenzen vergleichbare Daten, die **Eurostat**, das Statistische Amt Europas in Luxemburg, zusammenstellt und frei zugänglich macht, werden für nationale und europäische Politiken genutzt. So sind verlässliche Bevölkerungszahlen die Basis für Entscheidungen in der Gesundheits- und Bildungspolitik und werden für Abstimmungen des EU-Ministerrats nach

Statistik in der  
Politik

dem Grundlagenvertrag von Lissabon benötigt (Erfordernis der „doppelten Mehrheit“ mit 55 % der Staaten, die 65 % der EU-Bevölkerung repräsentieren). Auch der deutsche Beitrag zum Europäischen Stabilisierungsmechanismus (ESM), der die Stabilität des Euro sichern soll, hängt von Bevölkerungsdaten ab.

### Beispiel 1.2: Monitoring strategischer Ziele der amtlichen Statistik



Interaktives  
Lernobjekt

„Erwerbstätigkeit“

Im Jahr 2000 verständigten sich die Staats- und Regierungschefs der Länder der EU auf die sogenannte *Lissabon-Strategie*, die Entwicklungsziele für Europa bis 2010 definierte. Inzwischen ist eine mit dem Etikett *Europa 2020* versehene neue Strategie vereinbart, die wirtschaftliche und soziale Kernziele für Europa bis 2020 festlegt und anhand von acht Leitindikatoren operationalisiert. Ein Ziel ist z. B. die Erhöhung der Beschäftigungsquote der als erwerbsfähig geltenden EU-Bevölkerung auf 75 %. Der Erreichungsgrad dieses Ziels wird von Eurostat über den Indikator „Erwerbstätigenquote (Altersgruppe 20 - 64 Jahre)“ gemessen. Ein weiteres Ziel, das mit der Unterzeichnung des Kyoto-Protokolls (Zusatzvereinbarung zur Klima-Rahmenvereinbarung der Vereinten Nationen) in Zusammenhang steht, beinhaltet die Senkung der Emissionen von Treibhausgasen um 20 % gegenüber dem Stand von 1990.

Auch die *Vereinten Nationen* (UN) verfolgen globale Strategien und verknüpfen diese mit Indikatoren. Die UN wollen u. a. extreme Armut bekämpfen und gaben sich hier für den Zeitraum von 2000 bis 2015 Ziele vor, die aber nur teilweise erreicht wurden. Der Grad der Zielerreichung wurde anhand von acht Indikatoren, den *UN Millennium Development Goals*, in größeren Abständen quantifiziert. Seit Anfang 2016 gilt ein neuer, auf den *UN Millennium Development Goals* aufbauender Aktionsplan, die *UN Millennium Agenda 2030 für nachhaltige Entwicklung* mit 17 Kernzielen. Letztere sind nicht nur auf Entwicklungsländer und auf die Verminderung von Hunger und extremer Armut abgestellt. Vielmehr bieten sie einen für alle Länder gültigen Rahmen, der auf ein Umsteuern in Richtung nachhaltigen Wirtschaftens abzielt. Daten spielen auch hier eine Schlüsselrolle für das Politikmonitoring.



Interaktives  
Lernobjekt

„Emission von  
Treibhausgasen“

Aggregate aus verschiedenen Indikatoren, sog. zusammengesetzte Indikatoren, werden von internationalen Organisationen zur Beschreibung komplexer Entwicklungen eingesetzt, etwa zur Messung von Wohlfahrt oder Innovation.

### Exkurs 1.1: Datenjournalismus und Statistik-Blogs

Die zunehmende gesellschaftliche Relevanz der Statistik spiegelt sich auch darin wider, dass namhafte Zeitungen und Zeitschriften in ihren Häusern Ressorts eröffnet haben, die sich einer neuen Form des datengestützten Online-Journalismus widmen. Bei der als **Datenjournalismus** (engl: *data journalism*) bezeichneten Entwicklung steht die Verbindung interessanter Datensätze mit

interaktiven Grafiken, Landkarten, Animationen und erläuterndem Text (Analysen, Kommentierungen) sowie sozialen Netzwerken im Vordergrund. Bei der englischen Tageszeitung *The Guardian* werden die verwendeten Daten zudem in frei zugänglichen Datenarchiven dem Leserpublikum zur Verfügung gestellt und in Datenblogs diskutiert. Es entsteht ein neues interaktives Erzählformat, bei dem die im Brennpunkt stehenden Daten mit Datenbanken verknüpft sind (vgl. in Exkurs 1.2 den Link „Wohlstandsvergleich“).

Es gibt inzwischen viele Blogs zur Statistik, die nicht nur für Experten interessant sind. Beispielhaft genannt seien hier der *Stats Blog* und der *Blog about Stats*. Der erstgenannte Blog ist eine Plattform für den allgemeinen Informationsaustausch zwischen Statistikern mit Querverbindungen zu zahlreichen spezialisierten Statistik-Foren. Der letztgenannte Blog widmet sich vor allem neueren Entwicklungen im Bereich der Kommunikation amtlicher Daten.

## 1.2 Aufgaben und Teilbereiche der Statistik

Die Statistik ist also eine Disziplin mit vielfältigen Aufgaben und Anwendungsbereichen. Das Spektrum reicht von der Planung der *Erhebung von Daten* über die *Beschreibung und Visualisierung* der erhobenen Befunde über die *Identifikation von Auffälligkeiten* in den Daten bis zur *Ableitung von Schlüssen*, die über die vorliegenden Daten deutlich hinausgehen. Die Statistik ist demnach eine Wissenschaft, die Methoden zur Gewinnung und Analyse von Daten sowie zum Lernen aus Daten bereit stellt.

Aufgaben der  
Statistik

Umgangssprachlich wird Statistik oft anders verstanden, nämlich als eine schwer zugängliche, spröde Disziplin, die sich der Sammlung und Auswertung von Zahlenfriedhöfen verschrieben hat. Dieses Fehlverständnis reduziert die Statistik auf Tätigkeitsfelder, die für die heutige Statistik keinesfalls repräsentativ sind. Statistik ist eine faszinierende Wissenschaft mit vielfältigen Bezügen zur Praxis und interdisziplinärem Charakter.

Öffentliche  
Wahrnehmung des  
Fachs

Für Statistiker ist der Begriff „Statistik“ nicht eindeutig belegt. Sie verstehen hierunter einerseits ihre *Wissenschaft* als Ganzes. Sie verwenden den Begriff aber auch für *Kenngrößen*, die sich aus statistischen Daten ableiten (z. B. den Mittelwert). Im allgemeinen Sprachgebrauch wird auch häufig ein *Datensatz* als eine Statistik angesprochen, etwa ein Datensatz mit der Medaillenverteilung bei den Olympischen Sommerspielen. In diesem Manuskript wird „Statistik“ im Sinne von „Wissenschaft“ verwendet.

Mehrdeutigkeit des  
Begriffs „Statistik“

Innerhalb der Statistik lassen sich die beschreibende und die schließende Statistik unterscheiden. Die **beschreibende Statistik** oder **deskriptive Statistik** (engl.: *descriptive statistics*) umfasst numerische und grafische Verfahren zur Charakterisierung und Präsentation von Daten. Ziel ist die Reduktion der in den Daten enthaltenen statistischen Infor-

Teilbereiche der  
Statistik:

Beschreibende  
Statistik