

Einleitung

Wenn man von Statistik spricht, denkt man zunächst an das Gebiet der **beschreibenden Statistik**. Dieser älteste Zweig der Statistik beschäftigt sich vorwiegend mit der übersichtlichen Darstellung großer Datenmengen. Graphische und tabellarische Beschreibungsmethoden bilden das Instrumentarium, um umfangreiche Datenmengen auf einige wesentliche Informationen oder Maßzahlen zu reduzieren. Der Name "Statistik" geht auf Achenwall zurück, der im 18. Jahrhundert Vorlesungen über Staatenkunde ("notitia politica vulgo statistica") hielt. Seine Lehre von den "Staatsmerkwürdigkeiten" enthielt allerdings kaum quantitative Aussagen und hatte noch wenig mit der heutigen beschreibenden Statistik zu tun. Die bevölkerungsstatistische Arbeit von Süßmilch (1741) und die sozialwissenschaftlichen Studien von Quetelet (1796–1874) über einen sog. "mittleren Menschen" zielten schon mehr in eine quantitative Betrachtungsweise und Methodik.

Heute wird in allen modernen Staaten beschreibende Statistik betrieben. In der Bundesrepublik werden auf Bundes- und Länderebene im Rahmen der "Amtlichen Statistik" wirtschafts-, sozialstatistische und agrarstatistische Datenerhebungen von den Statistischen Bundes- und Landesämtern aufgrund bestehender gesetzlicher Grundlagen durchgeführt. Die Ergebnisse dieser Erhebungen werden in den Statistischen Jahrbüchern festgehalten. Aber auch nichtstaatliche Stellen wie Markt- und Meinungsforschungsinstitute, Versicherungen, Wirtschaftsinstitute und Betriebe beschäftigen sich mit beschreibender Statistik. Statistiken über Geburts- und Todesfälle, Verkehrsunfälle, Konsumgewohnheiten, Börsenkurse u.a. sind Beispiele für die Zielrichtungen deskriptiver Statistik. Man kann sagen, Aufgabe der beschreibenden Statistik ist es, große Datenmengen auf einige wenige Maßzahlen zu reduzieren, um damit komplexe Sachverhalte übersichtlich darzustellen. Wenn man jedoch über eine solche deskriptive Datenaufbereitung in Form von Tabellen und Diagrammen hinaus auch noch Rückschlüsse auf Entwicklungen und Trends ziehen will, Abhängigkeiten zwischen verschiedenen Erscheinungen oder Merkmalen herauskristallisieren will oder ganz allgemein Rückschlüsse von einem Teil auf das Ganze ziehen will, dann verläßt man das Gebiet der beschreibenden Statistik und begibt sich auf das der beurteilenden oder mathematischen Statistik.

Die **mathematische Statistik** basiert auf der Wahrscheinlichkeitsrechnung. Diese ist eine rein mathematische Theorie und operiert mit Grundbegriffen, die "axiomatisch" gefordert werden. Die Wahrscheinlichkeitsrechnung wurde im 17. und 18. Jahrhundert von Pascal, den Brüdern Bernoulli sowie Laplace entwickelt und in den dreißiger Jahren unseres Jahrhunderts von dem russischen Mathematiker Kolmogoroff auf ein modernes Fundament gestellt. Die Wahrscheinlichkeitstheorie erklärt, wie man mit Wahrscheinlichkeiten rechnet, wenn man alle Prämissen eines zufällig ablaufenden Experiments vollständig kennt. Sie stellt damit das Rüstzeug für die angewandte arbeitende Statistik dar, denn die Statistik lehrt, wie man aufgrund empirischer Beobachtungen Wahrscheinlichkeiten bestimmt oder schätzt, wie man theoretische Modelle empirischen Sachverhalten anpassen kann. Damit ist man schon beim Kernproblem der mathematischen Statistik, von der man gelegentlich sagt, daß sie sich mit sog. "Massenerscheinungen" beschäftigt. Das sind Erscheinungen, die sich auf eine große

Zahl von Objekten oder Versuchseinheiten (Tiere, Pflanzen, Patienten usw.) beziehen. Während nun eine Einzelercheinung völlig unregelmäßig verläuft, kann man den Massenerscheinungen im Großen gewisse modellmäßige Gesetzmäßigkeiten zuordnen. Man denke z.B. an die statistische Behandlungsweise in der Thermodynamik. Über ein einzelnes Molekül eines Gases läßt sich nichts aussagen. Der Gesamtheit aller Moleküle kann man jedoch gewisse statistische Parameter zuordnen. Als weiteres Beispiel denke man an einen Würfel. Das Ergebnis eines einzelnen Wurfs ist vollkommen zufällig. Wenn man dagegen eine große Anzahl von Würfeln betrachtet, so wird man feststellen, daß die Zahlen von eins bis sechs in etwa gleich häufig vorkommen, sofern der Würfel symmetrisch gebaut ist. Anstelle von Massenerscheinungen sagt man auch **Grundgesamtheit** oder **Population**.

Die statistisch zu untersuchende Grundgesamtheit weist immer eine Variabilität in einer oder mehreren Richtungen auf. Wäre dies nicht der Fall, so hätte es keinen Sinn, sie zum Gegenstand statistischer Untersuchungen zu machen, denn determinierte Vorgänge sind statistisch uninteressant. Eine Tier- oder Pflanzenpopulation weist in einem bestimmten Merkmal (z.B. Körpermaß, Milchleistung, Ertrag usw.) gewisse durch den Zufall zu erklärende Variabilitäten auf. Diese sind zu unterscheiden von Meßfehlern bei der Messung dieser Merkmale. Die Variabilität, die durch physikalische Meßfehler hervorgerufen wird, ist zwar ein Teil der gesamten beobachteten Variabilität, aber selbst wenn man diese Meßfehler eliminiert, bleibt eine zufällige Variabilität übrig, welche die des Meßfehlers in der Regel weit übersteigt, und die man im Falle einer Tier- oder Pflanzenpopulation als "biologische Variabilität" bezeichnet. Es muß jedoch nicht immer eine existente Grundgesamtheit von Individuen vorliegen, sondern man läßt auch hypothetische Grundgesamtheiten zu. So kann man eine einzelne Beobachtung bei einem zufälligen Versuch als Versuchseinheit und die in Gedanken unbegrenzte Wiederholung des Versuchs als grundgesamtheiterzeugend betrachten. Im Sinne der Wahrscheinlichkeitstheorie versteht man unter einer Grundgesamtheit die Menge aller möglichen Resultate eines nach einem festen mathematischen Modell ablaufenden Zufallsexperiments.

In der Praxis ist nun nicht die ganze Grundgesamtheit in ihren Eigenschaften bekannt, sondern nur eine kleine zufällig ausgewählte Teilmenge, denn man kann in der Regel nur einige Versuche z.B. mit einem neuen Medikament, einem Futtermittel oder einer neuen Weizensorte durchführen. Oder man kennt nur einige Exemplare einer bestimmten Tier- oder Pflanzenspezies. Diese zufälligen Teilmengen nennt man auch **Stichproben**. Die wichtigste Aufgabe der mathematischen Statistik ist es nun, Rückschlüsse auf die im Ganzen unbekannte Grundgesamtheit aufgrund der bekannten, beobachteten Stichprobe zu ziehen. Der Anwender statistischer Methoden befindet sich bezüglich der Grundgesamtheit im Ungewissen. Der statistische Rückschluß auf Eigenschaften der Grundgesamtheit soll ihm eine Entscheidungshilfe in einer Situation der unvollständigen Information geben, soll ihm z.B. die Frage beantworten: Ist die Sorte *A* besser als die Sorte *B* oder ist das Präparat *X* wirksamer als das Präparat *Y*?

Man kann also zusammenfassend die folgende Definition von Statistik geben: Die mathematische oder beurteilende Statistik ist eine methodische Wissenschaft. Sie stellt wissenschaftliche Verfahren zur Verfügung, um Daten zu gewinnen, auszuwerten und zu interpretieren, mit dem Ziel, vernünftige, objektiv vergleichbare Entscheidungen im Falle von Ungewißheit zu treffen.

Es existieren vielfältige Anwendungsgebiete statistischer Methoden in Industrie, Technik und Wissenschaften, z.B. Physik, Meteorologie, Medizin, Pharmazie, Biologie, Agrarwissenschaften, Gartenbauwissenschaften, Ernährungswissenschaften, Umweltwissenschaften, Ökologie, Verkehrswesen, Psychologie, Soziologie, Volks- und Betriebswirtschaft, Demoskopie u.v.m, kurz alle technischen, biologischen, ökonomischen und soziologischen Bereiche. In der Industrie benötigt man beispielweise statistische Methoden bei der Materialprüfung, in der Fertigungssteuerung sowie bei der End- und Abnahmekontrolle. In der Medizin und Pharmazie möchte man z.B. die Wirksamkeit eines neuen Medikaments erproben. Die Agrarwissenschaft fragt möglicherweise nach dem Einfluß verschiedener Fütterungsarten auf die Gewichtszunahmen von Tieren oder möchte wissen, ob sich die Erträge mehrerer Getreidesorten "wesentlich" oder nur zufällig unterscheiden. In der Ökologie interessiert z.B. wie sich bestimmte Tier- oder Pflanzenarten in der Fläche verteilen.