Statistik

von Univ.-Prof. Dr. Karl Mosler
Universität zu Köln

Der Begriff „Statistik“ hat im Deutschen mehrere Bedeutungen: Erstens das Ergebnis einer Datenerhebung und -auswertung, zweitens die entsprechende Aktivität, drittens die Gesamtheit der mit dieser Aktivität befassten Institutionen, und viertens die wissenschaftliche Disziplin Statistik. Um letztere hauptsächlich geht es im Folgenden.

Statistik als wissenschaftliche Disziplin ist die Lehre von der *methodischen Erhebung und Auswertung von Daten*. Zu ihren Aufgaben gehören

- die methodische Erhebung und Bereinigung der Daten,
- die graphische Darstellung,
- das Charakterisieren durch Kennzahlen,
- das Schätzen unbekannter Parameter,
- das Testen von Hypothesen,
- die Prognose künftiger Entwicklungen.

Sie gliedert sich in zwei große Bereiche, die beschreibende Statistik und die schließende Statistik. Die letzten drei Aufgaben und Teile der ersten werden der schließenden Statistik, die übrigen der beschreibenden Statistik zugerechnet.

Die betriebswirtschaftliche Theorie macht Aussagen über ökonomische Größen und ihre Beziehungen untereinander. Diese Aussagen beziehen sich auf reale Sachverhalte. Ihre Gültigkeit kann anhand von Beobachtungen des wirtschaftlichen Geschehens überprüft und quantifiziert werden. Beobachtungen müssen zunächst einmal beschrieben und gemessen werden. Die Beobachtung und Messung des wirtschaftlichen Geschehens und die Sammlung der so gewonnenen Daten sind die Aufgaben der *Wirtschaftsstatistik*. Die *beschreibende Statistik*, auch *deskriptive Statistik* genannt, dient dazu, die Daten unter bestimmten Aspekten zu beschreiben und die in den Daten enthaltene Information auf ihren – für eine gegebene Fragestellung – wesentlichen Kern zu reduzieren. Die *schließende Statistik* stellt darüber hinaus Methoden bereit, um Aussagen der Theorie anhand von Beobachtungsdaten als Hypothesen zu widerlegen oder zu bestätigen. Außerdem umfasst sie Methoden, um unbekannte Modellparameter zu schätzen und um künftige Entwicklungen zu prognostizieren.

Eine Grundaufgabe der beschreibenden Statistik ist die Charakterisierung der Daten durch *Kennzahlen*. Die wichtigsten Kennzahlen sind Maße der Lage, der Streuung, und der Schiefe sowie, bei mehreren Merkmalen, des Zusammenhangs. Eine weitere Grundaufgabe der beschreibenden Statistik besteht darin, die Daten in Graphiken übersichtlich und anschaulich darzustellen. Der Datenerhebung voraus geht die sinnvolle Auswahl der Beobachtungseinheiten. Die Erkennung und etwaige Elimination von extremen oder untypischen Beobachtungen, so genannten *Ausreißern*, ist ebenfalls eine Aufgabe der Statistik. Bei Zeitreihendaten ist ggf. die *Saisonfigur* zu bestimmen und die *Zeitreihe* um die Einflüsse der Saison zu bereinigen.

Die *schließende Statistik* nennt man auch *induktive Statistik* oder *statistische Inferenz*. Sie bietet eine spezielle Art von Logik, die es erlaubt, aus Beobachtungsdaten Schlüsse zu ziehen.

Allerdings gelten die betreffenden Folgerungen nicht mit Sicherheit, sondern nur „mit großer Wahrscheinlichkeit“ und unter bestimmten Annahmen über die Entstehung der Daten. In der statistischen Inferenz werden die beobachteten Daten als Ergebnisse von Zufallsvorgängen angesehen und im Rahmen von Wahrscheinlichkeitsmodellen analysiert. Die →*Wahrscheinlichkeitsrechnung* bildet deshalb die Grundlage der statistischen Inferenz.

Ausgangspunkt jeder statistischen Untersuchung ist die Festlegung einer →*Grundgesamtheit* (Beispiel: Börsentage eines Jahres, Betriebe einer Branche) und eines oder mehrerer beobachtbarer Merkmale, über die etwas ausgesagt werden soll. Grundlegend für die Auswertung ist das *Skalenniveau* der Merkmalswerte (=Daten). Man spricht von *metrisch skalierten Daten*, wenn die Merkmalswerte Zahlen sind und ihre Differenzen sinnvoll verglichen werden können (Beispiel: Temperatur, Beschäftigtenzahl), von *ordinal skalierten Daten*, wenn die Merkmalswerte sinnvoll der Größe nach geordnet werden können; ansonsten von *nominal skalierten Daten*.

Bei einer *Totalerhebung* werden die Daten aller Einheiten der Grundgesamtheit beobachtet, bei einer *Teilerhebung* nur eine Stichprobe davon. Eine solche Stichprobe muss nach allgemeinen Prinzipien ausgewählt werden: Bei der *reinen Zufallsauswahl* hat jede Einheit der →Grundgesamtheit die gleiche Chance, in die Stichprobe zu kommen. Bei der *geschichteten Zufallsauswahl* wird die Grundgesamtheit zunächst in Schichten (bspw. die Wohnbevölkerung in In- und Ausländer) zerlegt und dann in jeder Schicht eine Zufallsauswahl durchgeführt. Die *Quotenauswahl* ist eine nichtzufällige, systematische Auswahl, bei der bestimmte vorgegebene Anteile (bspw. an Männern und Frauen und an Altersgruppen) eingehalten werden.

Wenn die Daten eigens für die Untersuchung erhoben werden, liegt eine *Primärerhebung* vor, ansonsten eine *Sekundärerhebung*. *Längsschnittdaten* beziehen sich auf einen Merkmalsträger zu verschiedenen Zeiten, *Querschnittsdaten* auf mehrere Merkmalsträger zu einem Zeitpunkt; *Paneldaten* stellen eine Kombination von Längsschnitt- und Querschnittdaten dar.

Ein erster Schritt der Auswertung ist die *Häufigkeitstabelle*; sie enthält alle möglichen Werte eines Merkmals und die absoluten (oder relativen) Häufigkeiten, mit denen sie in den Daten vorkommen; sie heißt auch *diskrete Klassierung* und lässt sich graphisch auf vielerlei Weise (etwa als *Säulendiagramm*) veranschaulichen. Wenn lediglich die Häufigkeiten gezählt werden, mit denen die Merkmalswerte in bestimmte Intervalle fallen, spricht man von *stetiger Klassierung*; sie ist mit einem Informationsverlust gegenüber den ursprünglichen Daten verbunden. Ein *Lageparameter* spiegelt Verschiebungen der Daten sowie Änderungen der Maßeinheit wieder. Wichtigster Lageparameter für metrische Daten ist das arithmetische Mittel (→Mittelwerte), für ordinale Daten der →Median, für nominale der →Modus. Ein Skalenparameter ändert sich nicht bei einer Verschiebung der Daten, jedoch bei einer Änderung der Maßeinheit. Wichtigste Skalenparameter für metrische Daten sind *Standardabweichung* und *Spannweite*, für ordinale Daten der *Quartilsabstand* (→ Median). Weitere Kennzahlen der Verteilung eines Merkmals sind Schiefeparameter, die die Asymmetrie, sowie Wölbungsparameter, die die Masse auf den äußeren Flanken der Verteilung messen (→empirische Momente).

Werden zwei Merkmale zugleich betrachtet, bildet man zuerst eine (zweidimensionale) Tabelle der gemeinsamen Häufigkeiten, auch *Kontingenztafel* genannt. Ihre Zeilen- bzw. Spaltensummen bilden die Randhäufigkeiten, das sind die gewöhnlichen Häufigkeiten der einzelnen Merkmale. Der Zusammenhang zweier Merkmale wird durch den *Korrelationskoeffizienten* gemäß Bravais-Pearson (bei metrischen Daten), den

Rangkorrelationskoeffizienten gemäß Spearman (bei ordinalen Daten) und den *Kontingenzkoeffizienten* (bei nominalen Daten) gemessen. Dabei ist zu beachten, dass der Korrelationskoeffizient nur etwaige lineare Zusammenhänge der Merkmale misst; ein nicht-linearer, etwa quadratischer Zusammenhang bleibt unbeachtet.

Wichtigste Methode, den Zusammenhang zweier Merkmale zu quantifizieren, ist die *Regression*. Die lineare Regression unterstellt einen linearen Zusammenhang zwischen dem Merkmal Y (der „erklärten Variablen“) und dem Merkmal X (der „erklärenden Variablen“), was graphisch einer Geraden in der XY-Ebene entspricht. In der multiplen linearen Regression wird ein Merkmal Y in entsprechender Weise durch mehrere Merkmale $X_1, ..., X_m$ erklärt.

Im Unterschied zur beschreibenden Statistik bezieht die schließende Statistik Begriffe und Modelle der Wahrscheinlichkeitsrechnung in ihre Methoden ein. Man unterstellt, dass die erhobenen Daten Ergebnisse eines Zufallsvorgangs, das heißt, Realisationen von Zufallsvariablen sind. Dabei werden Annahmen über diesen Zufallsvorgang – etwa, dass es sich um eine einfache Zufallsstichprobe handelt – und über die in Frage kommenden Wahrscheinlichkeitsverteilungen der Zufallsvariablen getroffen. Aus den beobachteten Daten zieht man dann mit Hilfe der Wahrscheinlichkeitsrechnung Folgerungen über die konkret vorliegende Verteilung, insbesondere über Erwartungswert, Varianz und anderen Parametern dieser Verteilung. Weiterhin ist es möglich, Wahrscheinlichkeitsaussagen über die Genauigkeit von Schätzern und die Gültigkeit von Hypothesen zu treffen.

Etwa im Fall der linearen Regression unterstellt man, dass die Abweichungen zwischen dem beobachteten Y und dem erklärten linearen Zusammenhang normalverteilt sind. Dann lässt sich mit Hilfe eines Signifikanztests entscheiden, ob die Hypothese, dass die Steigung der Geraden von Null verschieden ist, statistisch gesichert werden kann. Ebenso lässt sich für jeden der beiden Parameter der Regression, Steigung und Ordinatenachsenabschnitt, ein Konfidenzintervall angeben, durch das der unbekannte Parameter mit einer vorgegebenen Wahrscheinlichkeit (z.B. von 95%) überdeckt wird.

Statistische Methoden sind universell; sie hängen grundsätzlich nicht vom Gebiet ihrer Anwendung ab. Einige Spezialgebiete der Statistik sind jedoch von besonderer Bedeutung für die Betriebswirtschaftslehre: Zeitreihenanalyse, Prognoseverfahren, Regressions- und Korrelationsanalyse, Clusteranalyse, Ereignisanalyse, Statistische Qualitätskontrolle; s.u. eine Auswahl aus der Literatur.

Siehe auch →statistische Software, →Ausbildung in Statistik, →Institutionen der Statistik, →Deutsche Statistische Gesellschaft,

Allgemeine Einführungen in die Statistik für Betriebswirte (Auswahl):
Bamberg, G., Baur, F. (2002): Statistik. Oldenbourg, München, 12. Auflage.
Lippe, P. von der (1996): Wirtschaftsstatistik. Amtliche Statistik und Volkswirtschaftliche Gesamtrechnungen. Lucius & Lucius, Stuttgart, 5. Auflage.
Mosler, K., Schmid, F. (2005): Beschreibende Statistik und Wirtschaftsstatistik. Springer, Berlin, 2. Auflage.
Mosler, K., Schmid, F. (2005): Wahrscheinlichkeitsrechnung und schließende Statistik. Springer, Berlin, 2. Auflage.

Rinne, H., (1996): Wirtschafts- und Bevölkerungsstatistik: Erläuterungen, Erhebungen, Ergebnisse. Oldenbourg, München, 2. Auflage.
Schira, J.(2005): Statistische Methoden der VWL und BWL. Pearson Studium, München, 2. Auflage.

Einführungen in relevante Spezialgebiete (Auswahl):

Blossfeld, H.-P., Hamerle, A., Meyer, K.U. (1986): Ereignisanalyse. Campus, Frankfurt.
Fahrmeir, L., Hamerle, A. Tutz, G. (1996): Multivariate statistische Verfahren, DeGruyter, Berlin, 2. Auflage.
Pokropp, F. (1996): Stichproben: Theorie und Verfahren. Oldenbourg, München.
Rinne, H. Mittag, H.-J. (1991): Statistische Methoden der Qualitätssicherung, Hanser, München.

Internetseite des Autors: http://www.uni-koeln.de/wiso-fak/wisostatsem/mosler

ERKLÄRUNGSSTICHWÖRTER

Ausbildung in Statistik

Im Rahmen des Studiums der Betriebswirtschaftslehre an den deutschsprachigen Universitäten und Fachhochschulen gehört eine Grundausbildung in →*Statistik* zum Pflichtprogramm. Sie umfasst in der Regel sowohl beschreibende als auch schließende Methoden. Grundständige Studiengänge des Fachs Statistik bieten die Universitäten Dortmund und München sowie die Berliner Freie Universität gemeinsam mit der Humboldt-Universität Berlin. An den meisten Hochschulen ist innerhalb des Studiums der Mathematik eine Vertiefung in statistischen Methoden möglich.

Institutionen der Statistik

Zur *amtlichen Statistik* in Deutschland gehört das Statistische Bundesamt (http://destatis.de), die statistischen Ämter der Länder und Kommunen sowie die statistischen Abteilungen weiterer Behörden, etwa der Deutschen Bundesbank (http://www.bundesbank.de), der Bundesagentur für Arbeit (http://www.arbeitsagentur.de) und des Bundesministeriums der Finanzen (http://www.bundesfinanzministerium.de). Zur *nichtamtlichen Statistik* zählt man die unabhängigen Wirtschaftswissenschaftlichen Institute: Das Institut für Weltwirtschaft in Kiel (http://www.uni-kiel.de/ifw), das Deutsche Institut für Wirtschaftsforschung in Berlin (http://www.diw.de), das Hamburger Weltwirtschaftliche Archiv (http://www.hwwa.de), das IFO-Institut für Wirtschaftsforschung (http://www.ifo.de), das Rheinisch-Westfälische-Institut für Wirtschaftsforschung (http://www.rwi-essen.de) und das Institut für Wirtschaftsforschung Halle (http://www.iwh-halle.de). Zur nichtamtlichen Statistik gehören auch die Wirtschaftsforschungsinstitute von Interessenverbänden sowie unabhängige Institutionen wie der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (http://www.sachverstaendigenrat-wirtschaft.de/), die Monopolkommission (http://www.monopolkommission.de) und Umfrageinstitute wie INFAS (http://www.infas.de) und das Institut für Demoskopie Allensbach (http://www.ifd-allensbach.de). Weitere nützliche Datenquellen bieten die internationalen Institutionen, insbesondere EUROSTAT (http://www.europa.eu.int/comm/eurostat), die OECD (http://www.oecd.org) und die Vereinten Nationen (http://www.un.org).

Statistische Software

Viele einfache statistische Berechnungen lassen sich bereits mit einem Tabellenkalkulationsprogramm wie EXCEL durchführen; eine Einführung dazu bietet das Lehrbuch Mosler/Schmid, Beschreibende Statistik und Wirtschaftsstatistik, Berlin 2005.
Die kommerziellen Programmsysteme SPSS (http://www.spss.com/de) und S-Plus (http:www.insightful.com) enthalten eine Vielzahl von über Menüs zugänglichen höheren

statistischen Verfahren; ihr sinnvoller Einsatz setzt beim Benutzer allerdings eine genaue Kenntnis der verwendeten Methoden und ihrer spezifischen Voraussetzungen voraus. Das Gleiche gilt für die Programmsysteme Stata (http://www.stata.com) und SAS (http://www.sas.de), die weitere flexible Möglichkeiten der statistischen Analyse bieten. Speziell für die Analyse von Zeitreihen und ökonometrischen Problemen wurde das Programmsystem Eviews (http://www.eviews.com) entwickelt. Eine nichtkommerzielle, kostenlose Alternative zu S-Plus bietet das Projekt R (http://www.r-project.org).

Deutsche Statistische Gesellschaft

Die Deutsche Statistische Gesellschaft (http://www.dstatg.de) wurde 1911 von Georg von Mayr gegründet. Sie umfasst etwa 800 Statistiker aller Fachrichtungen als Mitglieder. Ihr Ziel ist die Förderung der statistischen Wissenschaften in Theorie und Praxis. Sie versteht sich dabei als Bindeglied zwischen den Produzenten und den Nutzern statistischer Methoden. Wissenschaftliche Zeitschrift der Gesellschaft ist das vierteljährlich erscheinende *Allgemeine Statistische Archiv.*

Wahrscheinlichkeitsrechnung

Die Wahrscheinlichkeitsrechnung dient der Quantifizierung des möglichen Auftretens von Ereignissen. Grundmodell ist der *Zufallsvorgang*, das ist ein Vorgang, dessen Ergebnis im Voraus nicht feststeht. Jedem von endlich vielen möglichen Ergebnissen wird eine Zahl zwischen Null und Eins, seine *Wahrscheinlichkeit*, zugeordnet; die Wahrscheinlichkeiten aller Ergebnisse addieren sich zu Eins. Im Rahmen der Wahrscheinlichkeitsrechnung werden abhängige und unabhängige Zufallsvorgänge, Folgen von Zufallsvorgängen und numerische Ergebnisse von Zufallsvorgängen (→Zufallsvariable) untersucht. Es werden Wahrscheinlichkeiten abgeleiteter Ereignisse berechnet und Wahrscheinlichkeitsverteilungen von Zufallsvariablen durch geeignete Parameter charakterisiert. Die Wahrscheinlichkeitsrechnung erstreckt sich auch auf Zufallsvariable mit unendlich vielen Ergebnissen. Er erlaubt approximative Berechnungen von Wahrscheinlichkeiten im Rahmen von sogenannten Grenzwertsätzen.

Einführende Literatur zur Wahrscheinlichkeitsrechung für Betriebswirte:

Bamberg, G., Baur, F. (2002): Statistik. Oldenbourg, München, 12. Auflage.
Mosler, K., Schmid, F. (2005): Wahrscheinlichkeitsrechnung und schließende Statistik. Springer, Berlin, 2. Auflage.

Zufallsvariable

Eine Zufallsvariable X ist das numerische Ergebnis eines Zufallsvorgangs (→Wahrscheinlichkeitsrechnung). Sie kann endlich oder unendlich viele Werte auf der Zahlengeraden annehmen. Die *Verteilungsfunktion* F(x) einer Zufallsvariablen X gibt für jede Zahl x die Wahrscheinlichkeit an, mit der X den Wert x nicht übersteigt. Wichtigste

Parameter einer Zufallsvariablen X sind ihr *Erwartungswert* und ihre *Varianz*. Falls X nur endlich viele Werte $x_1, \ldots, x_k$ mit Wahrscheinlichkeiten $p_1, \ldots, p_k$ annimmt, ist sein Erwartungswert $E[X] = \sum_{i=1}^{k} x_i p_i$, also das mit den Wahrscheinlichkeiten gewichtete Mittel der Zufallsvariablen. Die Varianz $V[X]$ ist die - ebenfalls mit den Wahrscheinlichkeiten gewichtete - quadrierte Abweichung der Zufallsvariablen von ihrem Erwartungswert: $V[X] = \sum_{i=1}^{k} (x_i - E[X])^2 p_i$. Der Erwartungswert charakterisiert die Lage, die Varianz die Streuung einer Zufallsvariablen. Die Quadratwurzel aus der Varianz heißt *Standardabweichung*, $\sigma[X] = \sqrt{V[X]}$.

Mittelwerte

In der beschreibenden Statistik bezeichnen Mittelwerte die allgemeine Lage von metrisch skalierten (→Statistik) Daten $x_1, x_2, \ldots, x_N$. Der am weitesten verbreitete Mittelwert ist das *arithmetische Mittel* $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}(x_1 + x_2 + \ldots + x_N)$. Es wird auch kurz als *Mittelwert* oder *Durchschnitt* der Daten bezeichnet. Eine Verallgemeinerung des arithmetischen Mittels bildet das *gewichtete Mittel*. Es hat die Form $\bar{x}_w = \sum_{i=1}^{N} w_i x_i$ mit Gewichten $w_i \geq 0$ für alle i und $\sum_{i=1}^{N} w_i = 1$. Man nennt $\bar{x}_w$ gewichtetes Mittel zum Gewichtsvektor $w = (w_1, w_2, \ldots, w_n)$. Die Gewichte sind für die jeweilige Anwendung geeignet zu wählen. Speziell, wenn alle Gewichte gleich sind, erhält man das arithmetische Mittel. Wenn man besonders große und besonders kleine Werte weglässt, und zwar sowohl oben wie unten einen Anteil α der Daten $(0 < \alpha < 0{,}5)$, und das arithmetische Mittel aus den verbleibenden Daten berechnet, erhält man das *α-getrimmte Mittel* $\bar{x}_\alpha$. Das *harmonische Mittel* $\bar{x}_H$ ist der Kehrwert des arithmetischen Mittels der Kehrwerte der Daten, $\bar{x}_H = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i}\right)^{-1}$. Das *geometrische Mittel* $\bar{x}_G$ wird vor allem zur Berechnung von durchschnittlichen Wachstumsfaktoren und Wachstumsraten benötigt (Voraussetzung: alle $x_i > 0$), es gilt $\bar{x}_G = (x_1 \cdot x_2 \cdot \ldots \cdot x_N)^{1/N}$.

Median

Falls die Daten $x_1, \ldots, x_N$ ordinal skaliert (→Statistik) sind, wird ihre Lage durch den Median beschrieben. Um ihn zu berechnen, ordnet man zunächst die Werte in aufsteigender Weise. Falls N ungerade ist, ist er der an die Stelle $(N+1)/2$ geordnete Wert, falls N gerade ist, ist er der an die Stelle $N/2$ geordnete Wert. Offenbar liegt rechts und links vom Median jeweils etwa die Hälfte der Daten. In ähnlicher Weise werden weitere *Quantile* berechnet. Das untere *Quartil* teilt die Daten im Verhältnis $1/4$ zu $3/4$, das obere Quartil im Verhältnis $3/4$ zu $1/4$. *Quintile* unterteilen die Daten in Fünftel, *Perzentile* in Hundertstel.

Grundgesamtheit

Die Grundgesamtheit ist die Gesamtheit der Einheiten, über die eine statistische Untersuchung etwas aussagen soll. Sie ist eine Menge im Sinne der Mengenlehre. Ihre Elemente heißen *Untersuchungseinheiten*, *statistische Einheiten* oder *Merkmalsträger*. Die Grundgesamtheit einer statistischen Untersuchung muss in sachlicher, räumlicher und zeitlicher Hinsicht genau abgegrenzt sein. (Beispiele: Personen mit deutscher Staatsangehörigkeit am 01.01.2005, in Deutschland im Jahre 2004 produzierte Personenkraftwagen).

Eine Grundgesamtheit (oder einen Teil davon) bezeichnet man auch als *statistische Masse*. Man spricht von einer *Bestandsmasse*, wenn sie durch Angabe eines Zeitpunkts abgegrenzt wird, und von einer *Bewegungsmasse*, wenn sie durch Angabe eines Zeitraums bestimmt ist. Unter einem *Merkmal* versteht man eine Eigenschaft der Merkmalsträger, die statistisch untersucht wird. Ein Merkmal hat verschiedene mögliche Merkmalswerte. Grundgesamtheit und Merkmale müssen zu Beginn einer jeden statistischen Untersuchung präzise festgelegt werden. Die zu untersuchenden ökonomischen Größen sind zu operationalisieren, d.h. um eine Vorschrift zu ergänzen, die ihre konkrete numerische Beobachtung bei den Merkmalsträgern ermöglicht. Die beobachteten Werte eines Merkmals in einer Grundgesamtheit werden in einer *Datenmatrix* zusammengefasst: deren Zeilen entsprechen den Untersuchungseinheiten, die Spalten den – ggf. mehreren – erhobenen Merkmalen.

Indexzahlen

Die zeitliche Veränderung einer einzelnen ökonomischen Größe (Beispiel: Preis eines Markenprodukts) von einem Zeitpunkt auf einen anderen wird durch eine *Messzahl*, den Quotienten der Größe zu den beiden Zeitpunkten, dargestellt. Um die zeitliche Änderung mehrerer Größen mit einer Zahl zu messen, muss man deren einzelne Messzahlen geeignet aggregieren. Eine Indexzahl ist ein gewichtetes Mittel (→Mittelwerte) von Messzahlen. Dies sei am Beispiel eines *Preisindexes* für Konsumgüter verdeutlicht: Hier werden die Gewichte nach den wertmäßigen Anteilen der einzelnen Güter am Konsum bemessen. Man betrachtet einen *Warenkorb* (das ist eine Kollektion von bestimmten Gütern) zu zwei verschiedenen Zeiten, der *Basiszeit* und der *Berichtszeit*. Ein Preisindex ist ein gewichtetes Mittel von Preismesszahlen, d.h. der Preisverhältnisse zwischen Basis- und Berichtszeit der einzelnen Güter. Als Gewichte werden die Anteile der Umsätze (Preis mal Menge) der einzelnen Güter am Gesamtumsatz verwendet. Der Preisindex vom Typ *Laspeyres* ist ein gewichtetes arithmetisches Mittel, gewichtet mit den Umsätzen der Basiszeit. Der Preisindex vom Typ *Paasche* ist ein gewichtetes harmonisches Mittel, gewichtet mit den relativen Umsätzen der Berichtszeit. In entsprechender Weise werden *Mengenindizes* und *Wertindizes* gebildet. Das Statistische Bundesamt (→Institutionen der Statistik) berechnet und veröffentlicht regelmäßig Preisindizes und andere Indexzahlen für die verschiedenen Sektoren der Wirtschaft.

Empirische Momente

Betrachtet werden Daten $x_1,...,x_N$ eines Merkmals. Als erstes empirisches Moment der Daten bezeichnet man das arithmetische Mittel $\overline{x}$ (→Mittelwerte) als zweites empirisches Moment die empirische Varianz $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2$. Die Wurzel aus der empirischen Varianz heißt empirische Standardabweichung. Während $\overline{x}$ ein Lageparameter ist, ist $\sigma_X = \sqrt{\sigma_X{}^2}$ ein Skalenparameter. Das dritte empirische Moment, die *Schiefe* $g_X = \sum_{i=1}^{N}\left(x_i - \overline{x}\right)^3 / N\sigma^3$ beschreibt Abweichungen der Daten von der Symmetrie; bei Symmetrie gilt $g_X = 0$. Das vierte empirische Moment ist die *Wölbung* oder *Kurtosis*; sie ist analog der Schiefe mit der vierten statt der dritten Potenz definiert. Die Wölbung einer symmetrischen Verteilung von Daten ist ein Maß für die Stärke der „Flanken“ der Verteilung, d.h. des relativen Auftretens extremer Werte.

KURZSTICHWÖRTER

Varianz 1.in der beschreibenden →Statistik: siehe →Empirische Momente, 2. in der Wahrscheinlichkeitsrechnung: siehe →Zufallsvariable.

Modus 1**. (=***Modalwert*) von Daten: ein am häufigsten vorkommender Merkmalswert (→Grundgesamtheit), 2. einer stetigen →Zufallsvariablen: Stelle des steilsten Anstiegs der Verteilungsfunktion.

Spannweite von Daten ist die Differenz aus dem größten und dem kleinsten vorkommenden Wert (→Statistik).