

Allgemeine Methodenlehre der Statistik

von

Dr. phil. Johann Pfanzagl

Prof. an der Universität Köln

I

Elementare Methoden unter besonderer
Berücksichtigung der Anwendungen in den
Wirtschafts- und Sozialwissenschaften

3., neubearbeitete Auflage

Mit 50 Abbildungen



Sammlung Göschen Band 746/746a

Walter de Gruyter & Co. · Berlin 1966

vormals G. J. Göschen'sche Verlagshandlung · J. Guttentag,
Verlagsbuchhandlung · Georg Reimer · Karl J. Trübner · Veit & Comp.

Die Gesamtdarstellung umfaßt folgende Bände:

- Band I: Elementare Methoden unter besonderer Berücksichtigung der Anwendungen in den Wirtschafts- und Sozialwissenschaften
(Sammlung Göschen Band 746/746 a)
- Band II: Höhere Methoden unter besonderer Berücksichtigung der Anwendung in Naturwissenschaft, Medizin und Technik
(Sammlung Göschen Band 747/747 a)



Copyright 1965 by Walter de Gruyter & Co., vormals G. J. Göschen'sche Verlagshandlung — J. Cuttentag Verlagsbuchhandlung — Georg Reimer — Karl J. Trübner — Veit & Comp., Berlin 30. — Alle Rechte, einschl. der Rechte der Herstellung von Photokopien und Mikrofilmen, vom Verlag vorbehalten. — Archiv-Nr. 75 70 657. — Satz und Druck: Thormann & Goetsch, Berlin 44. — Printed in Germany.

Inhaltsverzeichnis

	Seite
1 Grundbegriffe	7
1.1 Einleitung	7
1.2 Die Statistik in den Naturwissenschaften	7
1.3 Die Statistik in den Sozialwissenschaften	8
1.4 Gegenüberstellung von Natur- und Sozialwissenschaften	10
2 Häufigkeitsverteilungen	13
2.1 Artmäßige und zahlenmäßige Merkmale	13
2.2 Diskrete und stetige Merkmale	15
2.3 Extensive und intensive Merkmale	16
2.4 Häufigkeitsverteilungen	16
3 Parameter	19
3.1 Lage-Parameter	20
3.2 Streuungs-Parameter	26
3.3 Der Variationskoeffizient	31
3.4 Konzentrationsmaße	33
4 Allgemeine Theorie der Maßzahlen	37
4.1 Die Konstruktion von Maßzahlen	37
4.2 Verhältniszahlen	41
4.3 Maßzahlen	42
4.4 Gliederungszahlen	50
4.5 Beziehungszahlen	53
4.6 Der Mittelwert von Verhältniszahlen	57
4.7 Standardisierung	59
4.8 Transplantation	66
5 Die Berechnung von Indexzahlen	67
5.1 Grundsätzliches	67
5.2 Die Formeln von Laspeyres und Paasche	70
5.3 Sub-Indizes	73
5.4 Repräsentation	74
5.5 Erweiterung des Index-Schemas	75
5.6 Substitution	78
6 Einige Beispiele für Indexzahlen	81
6.1 Index der Großhandelspreise	81
6.2 Index der Einzelhandelspreise	83
6.3 Index der Verbraucherpreise	83

	Seite
6.4 Kaufkraftvergleiche	87
6.5 Lohnindizes	89
6.6 Reallohnvergleiche	91
6.7 Volumenindizes und Umsatzindizes	92
6.8 Produktionsindizes	93
6.9 Produktivitätsindizes	96
7 Bestandsmassen — Bewegungsmassen	97
7.1 Grundbegriffe	97
7.2 Der Durchschnittsbestand	100
7.3 Analyse der Verweildauer	104
7.4 Die mittlere Verweildauer	105
7.5 Abgangsordnung und Verteilung der Verweildauer	111
7.6 Konstruktion stationärer Massen	114
8 Die Analyse von Zeitreihen	119
8.1 Ausschaltung von irregulären Schwankungen	119
8.2 Gleitende Durchschnitte	120
8.3 Methodische Bemerkungen zur Berechnung der gleitenden Durchschnitte	123
8.4 Die Differenzenmethode	128
8.5 Die Ausschaltung von Saisonschwankungen	129
8.6 Methoden für konstante Saisonfigur	133
8.7 Methoden für Saisonfigur mit variabler Amplitude	138
8.8 Methoden für bewegliche Saisonfigur	139
8.9 Harmonische Analyse	142
8.10 Schlußbemerkungen	143
9 Stichproben	144
9.1 Grundbegriffe	144
9.2 Der Zufallsfehler des Mittelwertes	146
9.3 Der Zufallsfehler der relativen Häufigkeit	150
9.4 Die Bestimmung des Stichprobenumfanges	153
9.5 Fiktive Gesamtheiten	156
9.6 Die geschichtete Stichprobe	159
9.7 Die optimale Stichprobe	162
9.8 Die proportionale Stichprobe	166
9.9 Die Auswahl der Schichten	171
9.10 Vergleich zwischen den Schichten	173
9.11 Die Klumpenstichprobe	175
9.12 Vergleich von Schichten und Klumpen	177
9.13 Die mehrstufige Stichprobe	178
9.14 Die Technik der Zufallsauswahl	179
9.15 Das Problem der Nichtbeantwortung	187
9.16 Verhältnis-Schätzung	191
10 Statistische Fehler	193
10.1 Zufällige und systematische Fehler	193
10.2 Fehler bei der Erfassung der Gesamtheit	195
10.3 Fehler bei der Erfassung der Daten	199
10.4 Die Bedeutung systematischer und zufälliger Fehler bei der Stichprobenerhebung	203
Die Interpretation fehlerbehafteter Zahlen	205

Technischer Anhang

11 Die Gewinnung des Zahlenmaterials	211
11.1 Die Abgrenzung der Erhebungsmasse und die Festlegung der Erhebungsmerkmale	211
11.2 Die Auswahl der Erhebungseinheiten und des Erhebungs- verfahrens	214
11.3 Die Aufbereitung	216
12 Die rechnerische Behandlung des Zahlenmaterials	221
12.1 Die Berechnung von Mittelwert und Varianz	221
12.2 Die Berechnung von Median und durchschnittlicher Ab- weichung	225
12.3 Das Rechnen mit fehlerbehafteten Zahlen	228
12.4 Die Kompensation unabhängiger Fehler	229
13 Die Darstellung des Zahlenmaterials	230
13.1 Die graphische Darstellung von Häufigkeitsverteilungen	230
13.2 Die Gestaltung von Tabellen	233
13.3 Kurvendiagramme	236
13.4 Stabdiagramme, Flächendiagramme	240
13.5 Darstellung einer Struktur	242
13.6 Kartogramme	243
13.7 Bildstatistiken	244
Tabelle	249
Literatur	251
Namen- und Sachverzeichnis	262

1 Grundbegriffe

1.1 Einleitung

Vielfach wird die Statistik als die Lehre von der Analyse von Massenerscheinungen definiert. Während sich früher die Anwendung der Statistik tatsächlich auf *große* Mengen bezog, wie durch das Wort „Masse“ suggeriert wird, tritt gerade in der neueren Statistik die Analyse von Stichproben kleinen Umfanges in den Vordergrund, so daß das Wort „Masse“ in der obigen Definition heute nur als „Mehrheit von Individuen“ interpretiert werden darf. Natürlich ist nicht *jede* Analyse von Massenerscheinungen Statistik. Das charakteristische Kennzeichen der *statistischen* Analyse ist, daß sie sich formaler Modellvorstellungen bedient. Die statistischen Methoden sind insbesondere zur Analyse von Erscheinungen geeignet, die unkontrollierten Einflüssen ausgesetzt sind. Dabei spielen sog. stochastische (d. h. auf der Wahrscheinlichkeitstheorie aufbauende) Modelle eine besondere Rolle.

1.2 Die Statistik in den Naturwissenschaften

Wenn der Physiker alle Bedingungen eines Experimentes mit beliebiger Präzision festlegen kann, braucht er keine Statistik, um die Ergebnisse dieses Experimentes auszuwerten. Er braucht die Statistik jedoch dann, wenn gewisse unkontrollierte Einflüsse da sind, welche eine merkliche „Streuung“ der Meßergebnisse verursachen. Dies trifft in der Physik oder Chemie bei Laboratoriumsexperimenten seltener zu als etwa bei biologischen oder psychologischen Experimenten, bei denen sich schon aus der Natur des Materials heraus unkontrollierbare Einflüsse ergeben. Es gilt jedoch vor allem dann, wenn man die Sphäre des Laboratoriumsexperimentes verläßt und sich in die Fabrik, auf das Feld oder in das Krankenhaus

begibt: Hier sind die Möglichkeiten des „Experimentierens“ sehr beschränkt. Viele maßgebliche Bedingungen sind vorgegeben, können nicht nach Belieben variiert werden — und trotzdem müssen aus den Daten ähnliche Schlüsse gezogen werden wie aus den Laboratoriumsexperimenten, bei denen man einen viel größeren Teil der Bedingungen nach Belieben festlegen kann. Wenn man eine Untersuchung darüber anstellt, ob — und wie — die Haltbarkeit von Kokillen von der chemischen Zusammensetzung des gegossenen Eisens abhängt, kann man nicht mit einem großen Experiment beginnen und Eisensorten mit den verschiedensten chemischen Zusammensetzungen unter genormten Bedingungen gießen. Man ist praktisch darauf angewiesen, die Abhängigkeit aus den verfügbaren Aufzeichnungen „herauszurechnen“. Und in diesen Aufzeichnungen spiegeln sich natürlich alle Einflüsse des normalen Betriebsgeschehens wider und verdecken den Einfluß der chemischen Zusammensetzung vor dem oberflächlichen Betrachter. Oder: Es soll ein neues Mittel zur Behandlung der Rhesus-Inkompatibilität erprobt werden. Auch hier ist man gezwungen, die Erprobung an einem sehr inhomogenen Material vorzunehmen: Frauen, die bereits frühere Geburten hinter sich haben, Frauen, bei denen es sich um die erste Geburt handelt. Die früheren Geburten können mit oder ohne Komplikationen verlaufen sein. Es kann eine Reihe von Interruptionen vorliegen usw. usw. Alle diese Faktoren können auf den Behandlungseffekt von Einfluß sein. Auch wenn man eine gewisse Auslese trifft, verbleibt stets eine große Zahl unkontrollierbarer Einflüsse, die unter Umständen den Effekt der Behandlungsmethode bei oberflächlicher Betrachtung verdecken können.

1.3 Die Statistik in den Sozialwissenschaften

Was hier für Technik und Medizin gesagt wurde, gilt in noch wesentlich stärkerem Ausmaß für die Analyse wirtschaftlicher und soziologischer Phänomene. Auch hier

ist natürlich die Möglichkeit von Experimenten gegeben. Wenn eine große Werbekampagne vorbereitet wird, kann man zwei einander möglichst ähnliche, aber voneinander isolierte Städte auswählen und in *einer* der beiden Städte eine Probe-Werbekampagne durchführen, um sodann die Umsatzentwicklung beider Städte zu vergleichen. Man hat auf diese Art verschiedene Einflüsse, wie die saisonale und konjunkturelle Entwicklung, unter Kontrolle gebracht, wenn sich saisonale und konjunkturelle Einflüsse in beiden Städten stets in gleicher Weise auswirken. Jeder Unterschied in der Entwicklung zwischen beiden Städten wird daher mutmaßlich auf den Einfluß der Werbekampagne zurückzuführen sein. (Wesentlich zuverlässigere Ergebnisse erhält man natürlich, wenn man dieses Experiment nicht mit einem einzelnen Städtepaar, sondern mit mehreren durchführt.) Die Möglichkeit, unter „kontrollierten“ Bedingungen zu arbeiten, ist jedoch bei wirtschaftlichen und soziologischen Analysen beschränkter als in den meisten anderen Wissenschaften*).

Eine ähnlich ungünstige Situation wie in den Sozialwissenschaften finden wir übrigens auch in der Meteorologie, und es ist nicht uninteressant, festzustellen, daß die statistischen Methoden in der Meteorologie nicht unähnlich denen der Sozialwissenschaften sind: emsiges Sammeln von Urmaterial, Korrelationsberechnungen, Zeitreihenanalysen usw.

Die typische Aufgabe der Wirtschaftsstatistik ist die Analyse von Daten, die von sehr vielen unkontrollierten Faktoren beeinflusst sind. Hier ein Beispiel:

Es ist zu beurteilen, ob das „Niveau“ der Konsumentenpreise eine steigende Tendenz aufweist oder nicht. Die Fakten: In den letzten Monaten haben sich bei verschiedenen Waren Preissteigerungen, vereinzelt auch Preisenkungen, ergeben. Hier und da ist eine Ware in einer

*) Wie *Morgenstern* betont, sind allerdings auch im Bereiche der Wirtschaft weit mehr Experimentiermöglichkeiten gegeben, als tatsächlich ausgenutzt werden.

völlig neuen Qualität auf den Markt gekommen und hat — trotz höheren Preises — das Interesse der Konsumenten auf sich gezogen. Einzelne der Preissteigerungen sind lediglich saisonbedingt — vielleicht im konkreten Fall wegen eines kalten Frühjahres besonders ausgeprägt. Andere Preissteigerungen dürften auf Panikkäufe zurückzuführen sein und sind daher mutmaßlich nur vorübergehender Natur. Wir haben hier eine sehr komplexe Situation vor uns, welche von einer großen Zahl unkontrollierter Faktoren beeinflusst ist. Es ist zu beurteilen, ob das „Preisniveau“ tatsächlich in Bewegung ist, ob deflatorische Maßnahmen notwendig erscheinen oder nicht. Hier ist natürlich ein Experiment undenkbar: Man kann nicht das Geschehen der letzten Monate nochmals abrollen lassen unter „kontrollierten“ Bedingungen, also den saisonalen Einfluß (oder zumindest die abnorm niedrige Temperatur des Frühjahrs) und die Panikkäufe eliminieren und beobachten, wie sich das von diesen Störungen befreite „Preisniveau“ entwickelt hätte. Man ist vielmehr gezwungen, die Schlüsse aus den das tatsächliche Geschehen betreffenden Daten zu ziehen, vor allem durch Zuhilfenahme von Vergleichsdaten aus anderen Perioden (durch Studium der Auswirkungen saisonaler Einflüsse in den vorhergehenden Jahren, durch Studium der Auswirkungen von Panikkäufen in früheren Zeitpunkten oder zur gleichen Zeit in anderen Ländern u. ä.).

1.4 Gegenüberstellung von Natur- und Sozialwissenschaften

Die hier skizzierten Beispiele zeigen eine große Variation in der Aufgabe, die der Statistik gestellt ist. Das eine Extrem: das Experiment, bei dem ein großer Teil der Einflußfaktoren beliebig festgelegt werden kann, so daß die Auswirkungen der übrigen — unkontrollierten — Einflüsse mehr oder minder den Charakter von Zufallsschwankungen tragen. Hier besteht die Aufgabe der Statistik darin, bei der Planung des Experimentes mitzuwirken (die Größe der kontrollierten Einflußfaktoren festzulegen usw.) und bei der Auswertung die Größe der zu-

fälligen Einflüsse zu beurteilen. Das andere Extrem: die Analyse von Daten, welche eine unkontrollierte Situation betreffen. Hier ist zu entscheiden: Welche Einflußfaktoren sind bei der Erhebung der Daten zu berücksichtigen? Die sorgfältige Aufzeichnung der verschiedenen tatsächlich auftretenden Kombinationen von Einflußfaktoren muß jene Informationen liefern, die man im Experiment durch planmäßige Festlegung der Einflußfaktoren erhält.

Noch ein zweiter Umstand beeinflusst den Charakter und die Anwendbarkeit der statistischen Methoden entscheidend: „... in allen Fällen wird die statistische Arbeitsweise dadurch charakterisiert, daß sich der Statistiker von vornherein von einem beträchtlichen Teil jener Kenntnisse lossagt, über die er für jede einzelne Beobachtungseinheit verfügt oder wenigstens verfügen könnte, und daß er sich nur mit jenen Angaben über die Einheiten begnügt, die über diese bei der Erhebung registriert werden.“ (*Anderson*, „Statistische Methodenlehre“, S. 6). An die Stelle der tatsächlich existierenden Einheit, des tatsächlichen Geschehens, tritt der „statistische Schatten“, wie *Anderson* sagt, in dem nur einzelne Merkmale der Wirklichkeit registriert sind. Der Einzelfall verliert seine Individualität, er geht in einer Gruppe von Fällen auf, die alle die gleiche Merkmalskombination aufweisen, deren „statistische Schatten“ identisch sind. Dieser Verlust der Individualität muß keine schwerwiegende Einbuße an Informationen mit sich bringen: Die Unfallversicherung weiß nicht, *wer* im nächsten Jahr einen Unfall erleiden wird und welches die näheren Begleitumstände dieses Unfalles sein werden, — aber sie weiß aus der Erfahrung vergangener Jahre, *wieviele* Unfälle es ungefähr sein werden, wieviele davon tödlich verlaufen werden usw. Für die Zwecke der Versicherung genügen diese Informationen. Entscheidend ist nicht der Verlust der Individualität, sondern die Frage, ob es möglich ist, alle relevanten Umstände im „statistischen Schatten“ zu berücksichtigen. Vielfach kann man wirklich alle relevanten Merkmale registrieren: Wenn genetische Unter-

suchungen mit irgendwelchen Fliegen durchgeführt werden, ist es irrelevant, ob die Tiere gefangen oder gezüchtet wurden, wie alt und wie groß sie sind, ob sie 6 Beine haben oder eines davon durch einen Unglücksfall einbüßten: Wesentlich ist der Genotypus, und dieser kann durch seinen „Schatten“ genau beschrieben werden. Die Vernachlässigung soundso vieler Merkmale bringt hier keinen Verlust an relevanten Informationen mit sich.

Bedenken wir nun anderseits die Probleme, die sich bei einer Untersuchung über den Zusammenhang von Wohlstand und Kinderzahl ergeben. Hier gibt es eine Reihe psychologischer Faktoren (Wunsch nach Ungebundenheit, Angst vor Kriegsgefahr usw.), welche den zu untersuchenden Sachverhalt maßgeblich beeinflussen, die aber in den „statistischen Schatten“ kaum aufgenommen werden können, die statistisch schwer erfassbar sind.

Die Vernachlässigungen, zu denen die Statistik bei der Behandlung wirtschaftlicher, soziologischer oder psychologischer Fragen gezwungen ist, sind bedeutend relevanter als in der Technik, Biologie oder Medizin. Daraus und aus dem unterschiedlichen Grad der „Kontrollierbarkeit“ ergibt sich eine starke Differenzierung in der Art der zur Anwendung gelangenden statistischen Methoden. So sind es bei sozialwissenschaftlichen Untersuchungen eher die elementaren Methoden, welche sinnvolle Anwendung finden, während die fruchtbarsten Anwendungsgebiete der verfeinerten Methoden in den Naturwissenschaften (Technik, Biologie, Medizin usw.) liegen.

Daraus ergibt sich eine natürliche Unterteilung des Stoffgebietes: Band I der vorliegenden Methodenlehre wendet sich vor allem an den Statistiker in den Sozialwissenschaften, Band II an den Statistiker in den Naturwissenschaften. Diese Trennung ist jedoch keineswegs streng: Der Statistiker in den Sozialwissenschaften wird immer wieder auf Fragen stoßen, welche verfeinerte Methoden erfordern, die erst in Band II behandelt werden. Umgekehrt wird der an den höheren Methoden inter-

essierte naturwissenschaftliche Statistiker auf wesentliche Teile des I. Bandes als Vorbereitung angewiesen sein.

2 Häufigkeitsverteilungen

Wie bereits eingangs ausgeführt wurde, besteht das Wesen der statistischen Methoden darin, daß von jeder bei einer Erhebung erfaßten Einheit gewisse Merkmale registriert werden. Es soll daher zunächst ein systematischer Überblick über die Natur der Merkmale gegeben werden.

2.1 Artmäßige und zahlenmäßige Merkmale

Herkömmlicherweise unterscheidet man zwischen *artmäßigen* (qualitativen) und *zahlenmäßigen* (quantitativen) Merkmalen. Typische Beispiele für artmäßige Merkmale sind das Geschlecht, der Familienstand und der Beruf; im Falle der Außenhandelsstatistik: die Art einer Ware oder das Bestimmungsland. Zahlenmäßige Merkmale sind etwa das Alter oder das Einkommen einer Person, die exportierte Menge u. ä. Selbstverständlich wird ein artmäßiges Merkmal nicht bereits dadurch zu einem zahlenmäßigen, daß man den einzelnen Ausprägungen Zahlen zuordnet: So ist die Steuergruppe, deren Ausprägungen durch die Zahlen 1, 2, 3 charakterisiert werden, natürlich ein artmäßiges Merkmal und kein zahlenmäßiges (es hätte wenig Sinn, festzustellen, daß sich die durchschnittliche Steuergruppe von 2,1 auf 2,2 verschoben hat).

Die Unterscheidung in artmäßige und zahlenmäßige Merkmale erweist sich für die meisten Zwecke der Sozialstatistik als ausreichend. Für die Anwendung verfeinerter statistischer Methoden ist sie jedoch zu oberflächlich. Neben artmäßigen Merkmalen wie Geschlecht oder Beruf, deren einzelne Ausprägungen wirklich verschiedenartig sind, gibt es auch solche, bei denen sich die einzelnen Ausprägungen im wesentlichen nur durch ihre *Intensität* unterscheiden. Das einfachste Beispiel liefert die Notenskala: „sehr gut“, „gut“, „befriedigend“, „genügend“,

„nicht genügend“. Hier ist eine eindeutige Ordnung (Reihenfolge) vorgegeben, zum Unterschied etwa von dem Merkmal „Beruf“, für dessen einzelne Ausprägungen (Tischler, Schneider, Verkäufer usw.) es keine natürliche Anordnung gibt. Ebenso kann man bei der Qualitätskontrolle die Produkte nach ihrer Fehlerhaftigkeit in Gruppen (1., 2., 3. Wahl) einteilen. Bei diesen Merkmalen ist eine natürliche Ordnung gegeben, die man selbstverständlich auch irgendwie durch Zahlen ausdrücken kann, bei denen diese Zahlen aber eben bloß diese Ordnung angeben und nicht mehr. Die Berechnung von Durchschnittswerten hat in diesem Falle wenig Sinn. (Dies gilt auch für die Schulnoten!) Selbst unter den zahlenmäßigen Merkmalen gibt es eine ganze Reihe, die im Grunde genommen bloß eine Ordnung zum Ausdruck bringen, so z. B. die Dioptrie — aufgefaßt als Maß für die Sehschärfe des Auges. Wir haben für diese Merkmale zwar eine konventionell festgelegte Zahlenskala; sie unterscheiden sich jedoch von den weiter unten besprochenen echten zahlenmäßigen Merkmalen dadurch, daß ihre Skala beliebigen monotonen Transformationen unterworfen werden kann, ohne daß sich deren Aussagegehalt ändert. Solche Skalen werden als *topologische* Skalen bezeichnet.

Geordnete artmäßige Merkmale und zahlenmäßige Merkmale, deren Skala beliebigen monotonen Transformationen unterworfen werden kann, gehören logisch gesehen in *eine* Klasse: Es sind Merkmale, für die es eine natürliche Ordnung gibt. Ob diese Ordnung durch eine Zahlenskala repräsentiert wird oder nicht, ist im Grunde genommen nebensächlich.

Daneben gibt es die sogenannten *metrischen* Skalen, die wir vor allem in der Physik finden: die Länge, das Gewicht, der elektrische Widerstand usw. Ein Widerstand von $200\ \Omega$ ist doppelt so groß wie ein Widerstand von $100\ \Omega$, denn zwei Widerstände von je $100\ \Omega$ in Serie geschaltet ergeben tatsächlich einen Widerstand von $200\ \Omega$. Hingegen wäre es sinnlos, festzustellen, daß eine Kurzsichtigkeit von 4 Dioptrien doppelt so stark ist wie eine

von 2 Dioptrien. Metrische Skalen dürfen nicht mehr beliebigen monotonen Transformationen unterworfen werden: Sie sind in der Regel eindeutig bis auf die Wahl der Maßeinheit.

2.2 Diskrete und stetige Merkmale

Ein zweiter Einteilungsgesichtspunkt ist die Unterscheidung von *diskreten* und *stetigen* Merkmalen. Bei diskreten Merkmalen sind nur ganz bestimmte Ausprägungen möglich. Diskrete Merkmale basieren in der Regel auf einem *Zählvorgang*. Beispiele für solche Merkmale sind: Anzahl der Kinder, die eine Frau geboren hat, Anzahl der Beschäftigten eines Betriebes usw. Stetige Merkmale sind solche, die jeden beliebigen Wert — zumindest innerhalb eines gewissen Intervalls — annehmen können. Stetige zahlenmäßige Merkmale basieren in der Regel auf einem *Messvorgang*. Beispiele für solche Merkmale sind: das Alter einer Person, ihre Körpergröße, ihr Gewicht usw.

Die Unterscheidung zwischen stetigen und diskreten Merkmalen wird in der Regel nur auf zahlenmäßige Merkmale angewendet. Sie ist jedoch auch bei artmäßigen Merkmalen sinnvoll, sofern eine natürliche Ordnung gegeben ist.

Selbstverständlich ist die Unterscheidung zwischen diskreten und stetigen Merkmalen eher graduell als prinzipiell. Das Gewicht einer Person ist ein stetiges, ihr Einkommen jedoch ein diskretes Merkmal, denn es kann — in Pfennigen ausgedrückt — nur ganzzahlige Werte annehmen. In der methodischen Behandlung besteht jedoch zwischen so fein abgestuften diskreten Merkmalen wie dem Einkommen (oder gar dem Umsatz eines Betriebes) und einem stetigen Merkmal kein Unterschied. Hingegen besteht ein ausgeprägter Unterschied in der methodischen Behandlung zwischen einem stetigen Merkmal und einem diskreten Merkmal mit einem kleinen Wertevorrat, wie es etwa die Kinderzahl darstellt. Umgekehrt wird ein stetiges Merkmal sehr oft „gruppiert“ und dadurch für die statistische Behandlung zu einem

diskreten Merkmal: So wurden in Tab. 3, S. 19, die Neugeborenen nach ihrer Größe in Gruppen zusammengefaßt. Ebenso wird oft einfach zwischen Rauchern und Nichtrauchern unterschieden, anstatt die Intensität des Rauchens zu erfassen.

2.3 Extensive und intensive Merkmale

Im Bereiche der zahlenmäßigen Merkmale kann man nach *Flaskämper* („Allgemeine Statistik“, Teil I, S. 40) weiter unterscheiden zwischen *extensiven* und *intensiven* Merkmalen. Extensive Merkmale betreffen den Umfang einer Erscheinung, intensive Merkmale ihren Inhalt. Formal gesehen sind extensive Merkmale solche, bei denen die Summe einen Sinn hat, intensive Merkmale solche, bei denen nur Mittelwerte sinnvoll sind. Extensive Merkmale sind z. B. die Anzahl der Einwohner eines Landes, denn man kann sie mit den analogen Zahlen anderer Länder zu einer größeren Einheit zusammenfassen, z. B. die Einwohnerzahlen der europäischen Länder zu einer Einwohnerzahl Europas; ferner Wertangaben, etwa der Umsatz im Monat Januar, der mit den Umsätzen anderer Monate zu einem Jahresumsatz zusammengefaßt werden kann. Intensive Größen sind beispielsweise die Preise: Man kann zwar einen mittleren Preis errechnen, aber die Summe von Preisen für sich hat keine Bedeutung. Extensiv-intensiv schließen einander jedoch nicht aus. Der Lohn beispielsweise kann sowohl als extensives als auch als intensives Merkmal aufgefaßt werden.

2.4 Häufigkeitsverteilungen

Um einen Überblick über die Struktur einer Masse zu geben, kann man bei einem diskreten (art- oder zahlenmäßigen) Merkmal einfach angeben, wieviele Fälle (oder welcher Anteil oder Prozentsatz) auf jede einzelne Ausprägung desselben entfallen. Dies nennt man *Häufigkeitsverteilung*. Oft wird die Häufigkeitsverteilung nicht nur in Form einer Tabelle angegeben, sondern auch graphisch dargestellt (Bild 1).

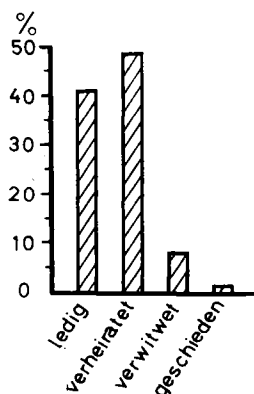


Bild 1. Wohnbevölkerung der Bundesrepublik Deutschland am 6. 6. 1961, aufgegliedert nach dem Familienstand.

Tabelle 1

Wohnbevölkerung der BRD am 6. 6. 1961
nach dem Familienstand

Familienstand	Häufigkeit	
	in 1 000	in %
ledig	22 264	41,2
verheiratet	26 375	48,9
verwitwet	4 530	8,4
geschieden	796	1,5
Wohnbevölkerung insgesamt	53 965	100,0

Quelle: Wirtschaft und Statistik 1963, S. 530.

Die in Tab. 1 angegebenen %-Zahlen werden *relative Häufigkeiten* genannt. Es handelt sich dabei um die Besetzungszahlen der einzelnen Ausprägungen, bezogen auf

die Gesamtzahl der Fälle. Diese wird entweder gleich 1 gesetzt oder gleich 100, wie in unserem Beispiel.

Ein weiteres Beispiel einer Häufigkeitsverteilung über einem diskreten Merkmal gibt Tabelle 2.

Bei den stetigen Merkmalen liegt die Situation etwas anders: Um die Struktur der Masse beschreiben zu können, muß man hier erst eine künstliche Gruppeneinteilung schaffen (siehe Tabelle 3, S. 19).

Tabelle 2
Aufgliederung der Münchener Familien
nach der Kinderzahl am 13. 9. 1950

(Familienhaushalte mit mindestens 2 Personen ohne Anstaltshaushalte
und Hausgemeinschaften familienfremder Personen)

Kinderzahl <i>i</i>	Häufigkeit	
	absolut	relativ
0	131 651	0,5863
1	59 224	0,2637
2	24 767	0,1103
3	6 582	0,0293
4	1 668	0,0074
5	468	0,0022
6	134	0,0006
7	39	0,0002
8	9	0,0000
9 u. mehr	7	0,0000
zusammen	224 549	1,0000

Quelle: Stat. Handbuch der Stadt München, 1954, S. 45.

Dabei erweist es sich oft als zweckmäßig, die Breite der einzelnen Gruppen unterschiedlich zu wählen, z. B. bei einer Verteilung, bei der die Randgruppen sonst zu schwach besetzt wären (siehe Tabelle 3, S. 19).

Tabelle 3
Aufgliederung neugeborener Kinder nach der Größe

Gruppengrenzen (Die obere Grenze zählt nicht zur Gruppe)	Häufigkeit	
	absolut	in %
30—35	31	0,28
35—40	75	0,68
40—42	57	0,52
42—44	70	0,64
44—46	238	2,17
46—47	253	2,31
47—48	485	4,42
48—49	899	8,20
49—50	1612	14,69
50—51	2506	22,84
51—52	2090	19,05
52—53	1413	12,88
53—54	721	6,57
54—55	337	3,07
55—56	123	1,12
56—58	58	0,53
58—60	3	0,03
insgesamt	10 971	100,00

Quelle: Unveröffentlichte Seminararbeit des Stat. Instituts
der Universität Wien.

Die graphische Darstellung von Häufigkeitsverteilungen wird im Anhang, S. 230 ff., behandelt.

3 Parameter

Bei vielen Problemen ist es zweckmäßig, nicht die Verteilung selbst zu betrachten, sondern nur eine (oder einige) ihrer charakteristischen Größen. Solche charakte-

ristische Größen sind der Mittelwert, der Median, die Varianz, die Schiefe etc. Diese Größen werden als *Parameter* bezeichnet. Ihre Definition setzt voraus, daß es sich um zahlenmäßige Merkmale handelt. Wichtig ist die Unterscheidung zwischen Lage-Parametern und Streuungs-Parametern.

3.1 Lage-Parameter

a) *Das arithmetische Mittel*: Gegeben seien n Zahlen: $a_1, a_2, \dots, a_n$. Unter dem arithmetischen Mittel $\bar{a}$ dieser n

Zahlen verstehen wir: $\bar{a} = \frac{1}{n} (a_1 + a_2 + \dots + a_n)$ oder, unter

Verwendung des Summationszeichens: $\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i$.

Das arithmetische Mittel hat vor allem bei extensiven Größen eine anschauliche Bedeutung, da dort auch die Summe eine anschauliche Bedeutung hat: Der durchschnittliche Seifenverbrauch — das ist der gesamte Seifenverbrauch gleichmäßig auf alle Personen aufgeteilt — oder umgekehrt: Der Durchschnitts-Verbrauch multipliziert mit der Anzahl der Personen gibt den Gesamt-Verbrauch. Allgemein: Das arithmetische Mittel multipliziert mit der Anzahl der Fälle ergibt die Summe. Dies ist eine der wichtigsten Eigenschaften des arithmetischen Mittels.

Neben seiner anschaulichen Bedeutung gibt es auch gewisse theoretische Gründe, die dem arithmetischen Mittel eine ausgezeichnete Stellung einräumen: So ist es beispielsweise möglich, aus dem arithmetischen Mittel zweier Gruppen und den zugehörigen Besetzungszahlen (d. h. der Anzahl der Einheiten in den einzelnen Gruppen) das gemeinsame arithmetische Mittel zu berechnen:

Beispiel: Auf Grund von Erhebungen des Statistischen Bundesamtes in Industriebetrieben mit 10 und mehr Beschäftigten betrug die im Jahre 1960 pro Arbeiter ausbezahlte Lohnsumme im Durchschnitt 5649 DM. Die im Jahre 1960 pro Angestellten ausbezahlte Gehaltssumme betrug 8851 DM. Die Zahl der Arbeiter betrug 6 311 000, die Zahl der Angestellten 1 465 000. Aus diesen Angaben können wir leicht

die Lohn- und Gehaltssumme pro Beschäftigten errechnen, die uns Aufschluß über das durchschnittliche Jahreseinkommen der in der Industrie beschäftigten Personen gibt:

Lohn- summe pro Arbeiter	×	Anzahl der Arbeiter	=	Lohn- summe	$5\,649 \text{ DM} \times 6\,311\,000 = 35\,651 \text{ Mill. DM}$
Gehalts- summe pro Ange- stellten	×	Anzahl der Ange- stellten	=	Ge- halts- summe	$8\,851 \text{ DM} \times 1\,465\,000 = 12\,967 \text{ Mill. DM}$
Lohn- und Gehaltssumme:					48 618 Mill. DM
Anzahl der Beschäftigten:					7 776 000
Lohn- und Gehaltssumme pro Beschäftigten:					6 252 DM

Die Lohn- und Gehaltssumme wird also als *gewogenes* arithmetisches Mittel aus der Lohnsumme pro Arbeiter und der Gehaltssumme pro Angestellten berechnet: *Gewichte* sind die Zahl der Arbeiter bzw. die Zahl der Angestellten. Das Ergebnis ist natürlich verschieden von dem ungewogenen arithmetischen Mittel

$$\frac{5649 + 8851}{2} = 7250.$$

Allgemein versteht man unter einem gewogenen arithmetischen Mittel einen Ausdruck der Form

$$\frac{G_1 a_1 + G_2 a_2 + \dots + G_n a_n}{G_1 + G_2 + \dots + G_n}.$$

Die Größen $G_1, G_2, \dots, G_n$ heißen Gewichte. Ein gewogenes arithmetisches Mittel läßt sich auch noch in folgender Form anschreiben:

$$g_1 a_1 + g_2 a_2 + \dots + g_n a_n,$$

wobei $g_i = \frac{G_i}{\sum_{i=1}^n G_i}$, daher $g_1 + g_2 + \dots + g_n = 1$.

(Manche Autoren beschränken den Ausdruck „Gewichte“ auf die Größen g_i , die aufaddiert die Summe 1 ergeben.)

Zerfällt eine statistische Masse in mehrere Teilmassen, so kann man das auf die gesamte Masse bezogene arithmetische Mittel als Mittel aus den arithmetischen Mitteln der Teilmassen durch Gewichtung mit ihren Umfängen darstellen (siehe obiges Beispiel).

Oft sind die Teilmassen in sich homogen (d. h. alle Einheiten einer Teilmasse haben den gleichen Merkmalswert). Dann erhält man das arithmetische Mittel der gesamten Masse, indem man den Merkmalswert jeder Teilmasse mit dem Umfang dieser Masse gewichtet.

Beispiel: Ein Betrieb hat ein Sortiment bestehend aus den Positionen A, B, C und D (Tabelle 4). Die Betriebsspanne errechnet sich aus den Stückspannen der einzelnen Positionen durch Gewichtung mit dem Umsatz. Sie beträgt $40,2 : 5,0 = 8,04\%$. (Eine Gewichtung mit der Stückzahl wäre unrichtig, da sich die in % ausgedrückte Spanne auf den Wert bezieht.)

Tabelle 4

Berechnung der Betriebsspanne aus den einzelnen Stückspannen

Position	Umsatz in Mill. DM (1)	Stückspanne in % (2)	(3) = (1) · (2)
A	0,7	8	5,6
B	1,3	12	15,6
C	0,5	8	4,0
D	2,5	6	15,0
Σ	5,0	8,04	40,2

Die Frage der möglichst zweckmäßigen numerischen Berechnung des arithmetischen Mittels wird im Anhang, S. 221 ff., behandelt.

b) *Der Median:* Gegeben seien n Zahlen $a_1, a_2, \dots, a_n$. Wir denken uns diese Zahlen der Größe nach geordnet und durchnummeriert. (Sind zwei Zahlen gleich, so werden diesen beiden Zahlen die Nummern nach Belieben zu-

geordnet.) Ist n *ungerade* ($n = 2k + 1$), so verstehen wir unter dem Median $\bar{a}$ das Element a_{k+1} . Der Median $\bar{a}$ ist also so beschaffen, daß gleichviele Zahlen kleiner und gleichviele größer als $\bar{a}$ sind. Ist n *gerade* ($n = 2k$), so kann man jeden beliebigen Wert zwischen a_k und a_{k+1} als Median ansehen. Vielfach wird $\bar{a} = \frac{1}{2}(a_k + a_{k+1})$ gesetzt.

Bei *stetigen* Merkmalen hat der Median eine sehr anschauliche Bedeutung: Es gibt genau einen Wert, der die Masse halbiert (d. h. unterhalb und oberhalb dessen je die Hälfte der Masse liegt). Bei der auf S. 19, Tab. 3, angegebenen Aufgliederung der Neugeborenen nach der Größe liegt der Median etwa bei 50,7 cm: Die Hälfte aller Neugeborenen ist kleiner, die Hälfte größer als 50,7 cm.

Bei *diskreten* Merkmalen ist die Bedeutung des Medians nicht so anschaulich wie bei stetigen. Bei der auf S. 18, Tabelle 2, angegebenen Aufgliederung nach der Kinderzahl ist der Median bei Kinderzahl 0. Die Median-Gruppe selbst umfaßt 59 % aller Fälle.

Liegt eine topologische Skala vor, so ist nur der Median, nicht aber das arithmetische Mittel bestimmbar: Die Zahlenangaben der topologischen Skala sind nicht eindeutig: Jede Skala, die aus der gegebenen durch eine beliebige monotone Transformation hervorgeht, hat genau denselben Aussagegehalt. Nun ist zwar der Median invariant gegenüber solchen Transformationen, nicht aber das arithmetische Mittel.

Ausführungen über die zweckmäßige numerische Bestimmung des Medians befinden sich im Anhang, S. 225 ff.

c) *Häufigster Wert*^{*)}. Manches Mal wird auch der häufigste Wert — der Wert, bei dem die Häufigkeitsdichte ihr Maximum hat — als Lage-Parameter verwendet. Er ist allerdings nur dann ein *plausibler* Lage-Parameter, wenn die Verteilung eingipfelig (unimodal) ist, nicht jedoch bei mehrgipfeligen, abfallenden oder U-förmigen Verteilungen.

^{*)} In Anlehnung an die englische Terminologie wird vielfach auch der Ausdruck „Modalwert“ gebraucht.



Bild 2. Eingipfelige Verteilung.

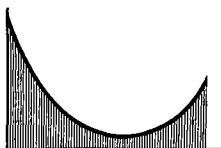


Bild 3. U-förmige Verteilung.

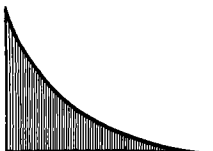


Bild 4. Abfallende Verteilung.

Zum Unterschied vom arithmetischen Mittel und dem Median kann der häufigste Wert nur dann mit brauchbarer Genauigkeit geschätzt werden, wenn die Stichprobe verhältnismäßig groß ist. Ähnlich wie das arithmetische Mittel ist auch der häufigste Wert nicht invariant gegenüber beliebigen monotonen Transformationen: Der transformierte häufigste Wert wird im allgemeinen nicht mehr der häufigste Wert der transformierten Verteilung sein.

d) *Das geometrische Mittel:* Neben den bisher genannten Lage-Parametern wird gelegentlich auch das geometrische Mittel verwendet. Unter dem geometrischen Mittel der Zahlen $a_1, a_2, \dots, a_n$ verstehen wir die Zahl

$$\sqrt[n]{a_1 a_2 \dots a_n}.$$

Die Verwendung des geometrischen Mittels ist auf Fälle beschränkt, in denen die Zahlen a_i nicht negativ sein können, da sonst u. U. die Wurzel nicht existiert. Eine sinnvolle Anwendung des geometrischen Mittels ist z. B. möglich bei der Mittelung relativer Änderungen. Früher wurde gelegentlich die Anwendung des geometrischen Mittels bei der Berechnung von Preisindexzahlen empfohlen.

Beispiel: Das geometrische Mittel der Zahlen in der letzten Spalte der Tabelle 5 beträgt 1,0188. Es sagt uns, daß

Tabelle 5
Preisindex für die Lebenshaltung (Mittlere Arbeitnehmer-
Haushalte, Basis 1950) in der BRD

Jahr	Preisindex für die Lebenshaltung	
	bez. auf 1950 = 1	bez. auf Vorjahr = 1
1950	1,000	
1951	1,078	1,078
1952	1,100	1,020
1953	1,081	0,983
1954	1,082	1,001
1955	1,100	1,017
1956	1,129	1,026
1957	1,152	1,020
1958	1,177	1,022
1959	1,188	1,009
1960	1,205	1,014

Quelle: Stat. Jahrbuch der BRD 1962, S. 506

eine konstante jährliche Steigerung von 1,88 % im Zeitraum 1950/60 zur gleichen Preiserhöhung geführt hätte.

Als grobe Faustregel könnte man angeben, daß die Verwendung des geometrischen Mittels dort sinnvoll ist, wo Unterschiede nicht durch die *Differenz*, sondern durch das *Verhältnis* charakterisiert sind (so wie gleiche Preissteigerung gleiche *relative* Preissteigerung bedeutet), also dort, wo auch die Verwendung einer logarithmischen Skala angemessen ist. Dieser Verwandtschaft entspricht folgende Relation: Entlogarithmiert man das arithmetische Mittel der Logarithmen, so erhält man das geometrische Mittel.

Lageregeln: Das geometrische Mittel ist stets kleiner als das arithmetische Mittel. (Grenzfall: Sind alle a_i gleich a_0 , dann sind auch arithmetisches und geometrisches Mittel gleich a_0 .)

Ist die Verteilung symmetrisch, so sind arithmetisches Mittel und Median identisch: Sie fallen mit dem Symme-

trie-Zentrum zusammen. Ist die Verteilung außerdem eingipfelig, so liegt auch der häufigste Wert im Symmetrie-Zentrum. Ist die Verteilung eingipfelig und schief nach rechts, so haben häufigster Wert (h), Median ($\tilde{\mu}$) und arithmetisches Mittel (μ) meist die in Bild 5 dargestellte Anordnung. (Bei einer nach links schiefen Verteilung ist die Reihenfolge umgekehrt.)

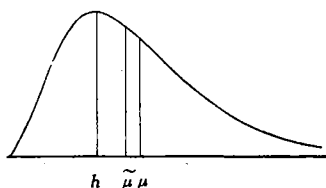


Bild 5. Die Anordnung von h , $\tilde{\mu}$ und μ bei einer nach rechts schiefen Verteilung.

3.2 Streuungs-Parameter

Ein Lage-Parameter gibt eine wichtige Information über die Verteilung, beschreibt diese aber keineswegs erschöpfend. Neben der Angabe eines Lage-Parameters ist vor allem eine Angabe darüber wichtig, wie stark die Einheiten um den Wert des Lage-Parameters streuen. Für manche Zwecke ist die Angabe eines Streuungs-Parameters sogar wichtiger als die eines Lage-Parameters.

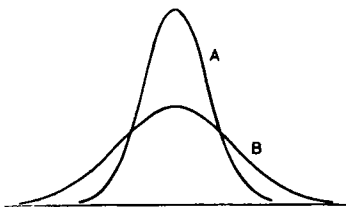


Bild 6. Zwei Verteilungen mit gleichem Mittelwert und verschiedener Streuung.

1. **Beispiel:** Betrachten wir als Beispiel die Reißfestigkeit des Garnes: Wichtig ist, daß bei der weiteren Verarbeitung des Garnes möglichst wenige Fadenbrüche auftreten. Die Zahl der Fadenbrüche hängt davon ab, an wie vielen Stellen des Garnes die Reißfestigkeit einen gewissen kritischen Wert unterschreitet. Dies hängt jedoch nicht nur von der *mittleren* Reißfestigkeit ab, sondern auch davon, wie stark die Reißfestigkeiten der einzelnen Stellen von diesem Mittelwert abweichen. m. a. W.: von der Streuung.

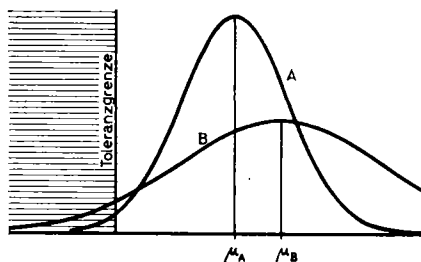


Bild 7. Obwohl Verteilung B ein größeres arithmetisches Mittel hat als A, wird bei ihr die Toleranzgrenze dennoch häufiger unterschritten.

2. **Beispiel:** Im allgemeinen wird man die Eignung eines Arbeiters für eine bestimmte Arbeit an Hand der Durchschnittszeit beurteilen, die er für deren Verrichtung braucht. Wenn es sich jedoch um einen Arbeitsvorgang handelt, der in einem bestimmten Rhythmus verrichtet werden soll (z. B. Arbeit am Fließband), ist auch die Streuung von Bedeutung. Eine zu große Streuung der Einzelzeiten könnte hier auch bei guter Durchschnittszeit zu Störungen führen.

3. **Beispiel:** Beim Bau von Staudämmen ist es nicht so sehr die durchschnittliche Niederschlagsmenge, die gefährliche Überschwemmungen hervorruft: Wo gewöhnlich viele Niederschläge fallen, hat sich sozusagen von selbst eine Regulierung ergeben. Die Gefahr von Überschwemmungen besteht dort, wo starke Unregelmäßigkeiten bestehen, d. h. wo die Niederschlagsmenge stark streut.

4. Beispiel: Sehr drastisch wird die Bedeutung der Streuung durch folgendes Beispiel illustriert*): Der Kongreß-abgeordnete John Jennings führte am 6. Juni 1946 vor dem Amerikanischen Kongreß aus, daß die durchschnittliche Tiefe des Tombigbee River zu gewissen Zeiten nur 1 Fuß betrage. „Mit anderen Worten, man kann den Fluß von der Mündung bis zur Quelle durchwaten.“ Ein Fluß mit einer durchschnittlichen Tiefe von 1 Fuß kann jedoch in Wirklichkeit an manchen Stellen sehr tief sein. Die *durchschnittliche* Tiefe sagt darüber, ob man den Fluß durchwaten kann, gar nichts.

Nun besprechen wir einige Streuungs-Parameter:

a) *Standardabweichung, Varianz*: Gegeben seien n Zahlen $a_1, a_2, \dots a_n$. Unter der *Varianz* dieser Zahlen verstehen wir:

$$s^2 = \frac{\sum_{i=1}^n (a_i - \bar{a})^2}{n - 1}.$$

Dabei ist $\bar{a}$ das arithmetische Mittel von $a_1, a_2, \dots a_n$. Die Größe s wird als *Standardabweichung* bezeichnet.

Es ist eher die Standardabweichung als die Varianz, die dem anschaulichen Begriff der „Streuung“ entspricht: Wird die Abweichung jedes Einzelwertes vom Mittelwert verdoppelt, so verdoppelt sich die Standardabweichung, während sich die Varianz vervierfacht. Die Standardabweichung hat die gleiche Dimension wie die Ursprungswerte. (Haben diese die Benennung cm, so hat die Standardabweichung gleichfalls die Benennung cm, die Varianz hingegen die Benennung cm².) Die Standardabweichung kann daher — zum Unterschied von der Varianz — sinnvoll mit irgendwelchen Lage-Parametern verglichen werden (siehe Abschnitt 3.3). Trotzdem wird in der mathematischen Statistik fast ausschließlich die Varianz verwendet, weil sich die meisten Formeln der Stichprobentheorie mit der Varianz übersichtlicher ausdrücken lassen als mit der Standardabweichung.

*) Zitiert nach Wallis und Roberts, „Statistics — a new approach“, S. 80.

Berechnet man die Standardabweichung (oder die Varianz) nicht für die Abweichungen vom arithmetischen Mittel $\bar{a}$, sondern von irgendeinem anderen Lage-Parameter, so erhält man stets einen größeren Wert. Das arithmetische Mittel hat also bei der Berechnung insofern eine besondere Stellung, als es unter allen denkbaren Lage-Parametern jener ist, der den kleinsten Wert der Standardabweichung ergibt. Dies ist die sog. „Minimumseigenschaft des arithmetischen Mittels“.

Die Frage der möglichst zweckmäßigen numerischen Berechnung von Varianz und Standardabweichung wird im Anhang, S. 221 ff. behandelt.

b) Durchschnittliche Abweichung: Unter der durchschnittlichen Abweichung der Zahlen $a_1, a_2, \dots, a_n$ verstehen wir:

$$d = \frac{\sum_{i=1}^n |a_i - a^*|}{n}$$

Für a^* wird in der Regel der Median $\tilde{a}$, manchmal auch das arithmetische Mittel $\bar{a}$ verwendet. Verwendet man bei der Berechnung der durchschnittlichen Abweichungen statt des Medians irgendeinen anderen Lage-Parameter, so erhält man stets einen größeren Wert. Der Median spielt für die durchschnittliche Abweichung eine ähnliche Rolle wie das arithmetische Mittel für die Standardabweichung: Er ist unter allen denkbaren Lage-Parametern jener, der den kleinsten Wert der durchschnittlichen Abweichung ergibt. (Minimumseigenschaft des Medians.)

Die durchschnittliche Abweichung hat — ähnlich der Standardabweichung — die gleiche Dimension wie die Ursprungswerte.

Die Frage der möglichst zweckmäßigen numerischen Berechnung der durchschnittlichen Abweichung wird im Anhang, S. 225 ff. behandelt.

Varianz (Standardabweichung) und durchschnittliche Abweichung sind nur dann sinnvolle Maße, wenn die Skala eindeutig bis auf lineare Transformationen ist.

c) *Quantile*: Es seien n Zahlen $a_1, a_2, \dots, a_n$ gegeben. Diese denken wir uns der Größe nach geordnet und durchnumeriert. Unter dem p -Quantil $\tilde{a}_p$ verstehen wir a_k , wobei k die auf np folgende ganze Zahl ist, sofern np nicht selbst ganzzahlig ist. Für ganzzahliges np kann als p -Quantil jeder Wert zwischen a_{np} und a_{np+1} genommen werden. Ein spezielles Quantil haben wir bereits kennengelernt: den Median $\tilde{a}$, in der hier eingeführten Schreibweise $\tilde{a}_{1/2}$. Der Median wird als Lage-Parameter verwendet. Man kann Quantile aber auch als Streuungs-Parameter verwenden. Die beiden Quartile $\tilde{a}_{1/4}, \tilde{a}_{3/4}$ oder auch die Dezile $\tilde{a}_{1/10}, \tilde{a}_{9/10}$ geben ein sehr anschauliches Bild von der Lage und Größe des Streubereiches. Als Streuungsmaß im eigentlichen Sinne dient ihre Differenz, vor allem der sog. *Quantilabstand* $\tilde{a}_{3/4} - \tilde{a}_{1/4}$.

Tabelle 6
Preise für Schweinefleisch in Wien und Graz
am 10. 1. 1959

Fleischsorte	Preise per kg in ö. S.					
	Wien			Graz		
	Min.	Max.	Spann- weite	Min.	Max.	Spann- weite
Schopfbraten, abgezogen mit Knochen	26	30	4	28	29	1
Schulter, abgezogen	30	36	6	31	36	5
Bauchfleisch mit Schwarten und Knochen	20	25	5	20	23	3
Schlegel	32	40	8	34	38	4

Quelle: Unveröffentlichte Statistik der Bundeskammer der gewerblichen Wirtschaft.