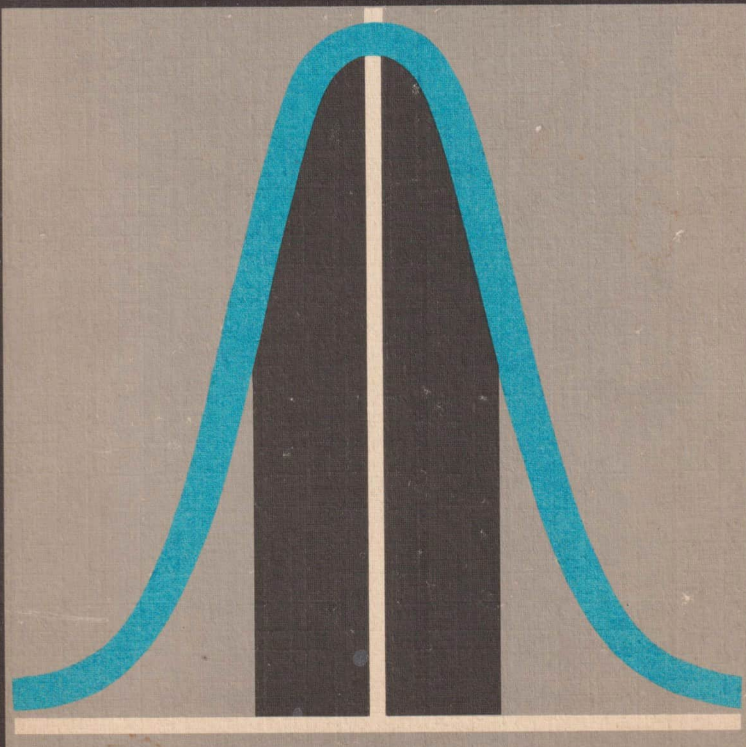
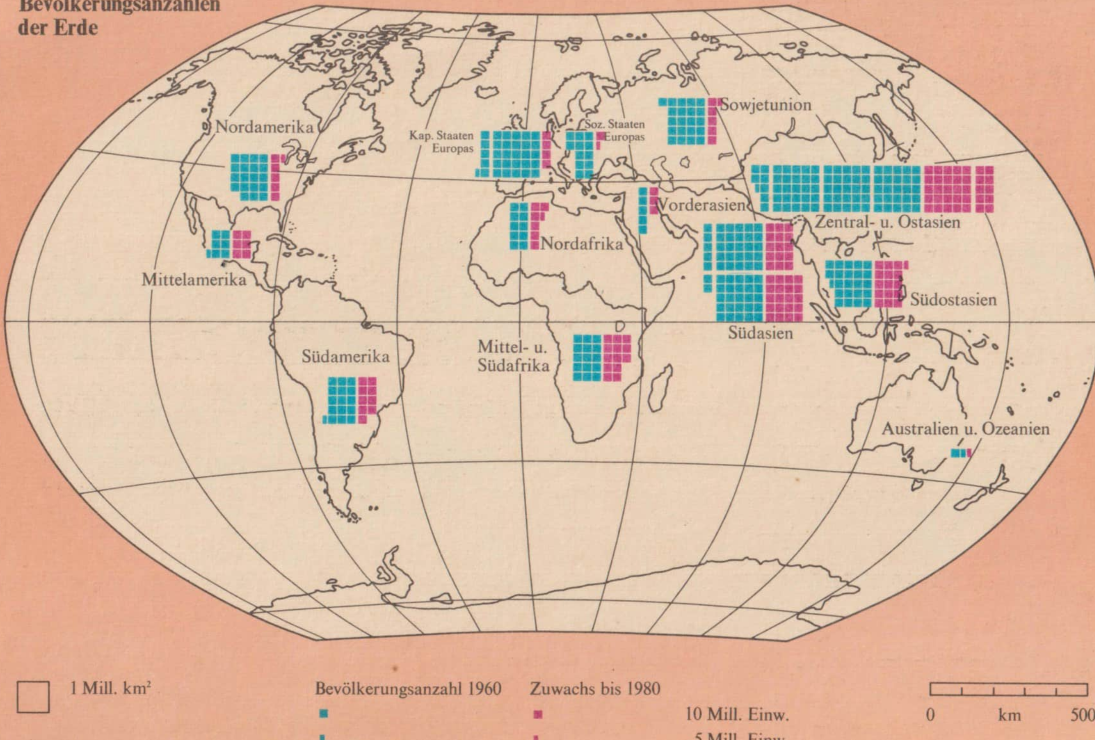
Elementare Statistik

Heinz Lohse



Bevölkerungsanzahlen der Erde



Nach „horizont", 12. Jahrgang/1979, Nr. 48

Elementare Statistik

Heinz Lohse

Volk und Wissen
Volkseigener Verlag Berlin
1983

Autor:
Prof. Dr. sc. Heinz Lohse

Dieses Buch wurde durch das Ministerium für Volksbildung der Deutschen Demokratischen Republik als Lehrmaterial für den fakultativen Kurs Elementare Statistik in den Klassen 9 und 10 bestätigt und unter Nr. 94 in die mathematische Schülerbücherei aufgenommen.

1. Auflage
Lizenz Nr. 203/1000/83 (E 001716-1) · VWV 36/82
LSV 1071
Redaktion: Heinz Junge, Karlheinz Martin
Illustrationen: Harri Förster
Zeichnungen: Heinrich Linkwitz
Einband: atelier vwv, Wolfgang Lorenz
Typographie: atelier vwv, Karl-Heinz Bergmann
Printed in the German Democratic Republic
Gesamtherstellung: INTERDRUCK Graphischer Großbetrieb Leipzig,
Betrieb der ausgezeichneten Qualitätsarbeit, III/18/97
Schrift: 9/10 p Times-Monotype
Redaktionsschluß: 20. September 1982
Bestell-Nr.: 707 153 5
DDR 6,00 M

Inhalt

Vorwort

Die Statistik hat in den letzten Jahren erheblich an Bedeutung gewonnen. Sie ist zum unentbehrlichen Hilfsmittel der Leitung, Planung und Forschung in der gesamten Volkswirtschaft, in Wissenschaft und Technik geworden. Es gibt kaum noch einen Bereich, der auf die Statistik – direkt oder indirekt – verzichten kann.
Mit dem Titel „Elementare Statistik" wird in erster Linie das Ziel verfolgt, Teilnehmern und Leitern entsprechender fakultativer Kurse ein Material zur Verfügung zu stellen, das ihnen das Erreichen der im Rahmenprogramm ausgewiesenen Ziele erleichtert. Darüber hinaus wird das Buch allen denen eine Hilfe sein, Zugang zur Statistik zu finden, die mit der vorhandenen Literatur nur unter Mühen zurechtkommen.
Das Buch ist so angelegt, daß im Zuge der Stoffvermittlung Beispiele genannt und oft ausführlich behandelt werden, die den Lernenden befähigen, die zahlreichen Aufgaben selbständig zu lösen und im Kollektiv beratene Forschungsaufträge oder Untersuchungen selbst oder im Rahmen einer Schülergruppe auszuführen. Die Beispiele und Aufgaben sind vielen Bereichen entnommen; sie können zur selbständigen Stellung von Aufgaben, die den spezifischen Bedingungen des Territoriums entsprechen, anregen. Parallel zur Bearbeitung des Stoffes mit dem Buch sollte ein mit dem Schülerkollektiv zu lösender Forschungsauftrag erfüllt werden, der rechtzeitig mit einem Betrieb, einer Landwirtschaftlichen Produktionsgenossenschaft oder einer Institution abgestimmt wurde. Die Arbeit auf dem Gebiet der Statistik macht vor allem dann Spaß, wenn man sich von Anfang an, d. h. beginnend mit der Planungsphase, mit den zu untersuchenden Problemen befaßt. Man sollte in den Arbeitsgemeinschaften selbsterfaßte Daten bearbeiten und das aufbereitete und ausgewertete Material sinnvoll nutzen, z. B. für Wandzeitungen, für Lernkonferenzen, für die Messe der Meister von Morgen. In diesem Zusammenhang sei jedoch darauf verwiesen, daß die Datenerfassung gemäß den Richtlinien in den staatlichen Verfügungen zu erfolgen hat. Man orientiere sich rechtzeitig in

- der Verordnung über Rechnungsführung und Statistik – GBl I Nr. 31, Seite 585,
- der Zweiten Verordnung über Rechnungsführung und Statistik – GBl I Nr. 22, Seite 215 und
- der Dritten Verordnung über Rechnungsführung und Statistik – GBl I Nr. 6, Seite 125.

Über die Anregungen hinaus, die das Buch vermittelt, können folgende Sachverhalte aufgegriffen werden:

- Auswertung von Klassenarbeiten und Leistungskontrollen zur Vorbereitung einer Lernkonferenz;
- Unterstützung der Wehrerziehung;
- Anzahl der Schwimmer/Nichtschwimmer und die Schwimmstufenverteilung in einzelnen Klassen oder Klassenstufen;
- Entwicklung der Schulsportgemeinschaft, Spartakiadebewegung;

- Verwendung finanzieller Mittel an den Schulen;
- Solidaritätsaufkommen;
- Zahlenmaterial der Altstoffsammlungen;
- Meinungsumfrage über Filme;
- Gesundheitserziehung, z. B. Problem Raucher/Nichtraucher.

Aktivität und Selbständigkeit der Schüler können dabei erheblich gefördert werden. Es wird bei diesem Vorgehen auch bewußt, daß die Arbeit mit statistischem Material nicht losgelöst vom Gesamtzusammenhang erfolgen darf, daß Statistik integrativer, vom Anfang der Problemstellung an zu berücksichtigender Bestandteil von Untersuchungen und Aufträgen ist.

Der Verfasser dankt allen, die das Entstehen des Buches förderten. Wertvolle Unterrichtserfahrungen, Aufgaben und Anregungen der Herren Dr. Eilhauer, Oehme, Dr. Richter, Dr. Rohn, Dr. Sill, Dipl.-Päd. Stockhause, Dr. Thamm, Dr. Tischer und Dr. Wienke sowie von Frau Dr. Moldenhauer konnten aufgegriffen und zum Teil verwendet werden.
Alle Vorschläge zur weiteren Verbesserung des Materials sind jederzeit willkommen.

Leipzig, 1982 Heinz Lohse

Hinweise

Definitionen (D), Erklärungen (E), Beispiele (B) sowie Aufgaben und Aufträge sind jeweils durch das ganze Buch fortlaufend numeriert und werden durch folgende Zeichen besonders hervorgehoben:

▶ Definitionen und Erklärungen (z. B. ▶ D 1, ▶ E 2);
■ Beispiele;
● . Aufgaben und Aufträge.

Zu Aufgaben, die mit einem Stern versehen sind, finden sich im Anhang die Lösungen.
Mit dem Zeichen „↗" (siehe) werden Bildverweise und Seitenhinweise eingeleitet.

Wertvolles Zahlenmaterial kann den Statistischen Jahrbüchern der DDR und den ebenfalls jährlich erscheinenden Statistischen Taschenbüchern der DDR entnommen werden. Das Literaturverzeichnis am Ende des Buches vermittelt zahlreiche weitere Anregungen.

Zu beachten ist: Bei Durchführung größerer Untersuchungen ist unbedingt eine Abstimmung entweder mit der betreffenden Bezirksstelle der Zentralverwaltung für Statistik der Deutschen Demokratischen Republik oder mit der jeweiligen Kreisstelle der Zentralverwaltung bzw. der betreffenden Fachabteilung des Rates des Kreises erforderlich.

Zu den Aufgaben und zur Bedeutung der Statistik

1

1.1. Was ist und was kann die Statistik?

Das Wort „Statistik" hört und liest man heute sehr häufig. Schlagen wir eine Tageszeitung auf, so finden wir beispielsweise die neueste „Fußball-Statistik" oder eine graphische Darstellung statistischer Daten (↗ Bild 1) oder den ausführlichen Bericht der Staatlichen Zentralverwaltung für Statistik über die Durchführung des Volkswirtschaftsplanes.

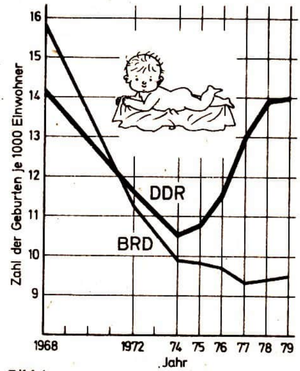


Bild 1

Schon daraus können wir entnehmen, welch große Bedeutung der Statistik in einem modernen Industriestaat zukommt. Das gilt erst recht für die entwickelte sozialistische Gesellschaft, in der Planung und statistische Datenauswertung eine dominierende Rolle spielen. Ohne aktuelle und zuverlässige Informationen über den Stand und den Verlauf aller Produktions-, Transport-, Bildungs- und anderer wichtiger Prozesse ist die Leitung und Planung der sozialistischen Volkswirtschaft nicht denkbar. Dennoch sind damit die Aufgaben der Statistik bei weitem noch nicht erschöpft. In jedem Industriebereich, in der Landwirtschaft und in vielen Wissenschaften ist ein Voranschreiten, ist wissenschaftlich-technisches Höchstmaß ohne Anwendung statistischer Methoden und Verfahren heute nicht mehr möglich. Immer mehr Werktätige benötigen die Statistik für ihre tägliche Arbeit.

Der bekannte sowjetische Wissenschaftler B. V. Gnedenko hat das einmal mit folgenden Worten zum Ausdruck gebracht:

„Statistische Auffassungen und Gesetzmäßigkeiten sind in unseren Tagen nicht nur für irgendwelche besonderen Spezialisten erforderlich, sondern buchstäblich für alle – den Arbeiter und den Arzt, den Ingenieur und den Lehrer, für den Ökonomen und den Offizier, den Biologen und Agronomen, den Baufachmann und den Produktionsorganisator."

Dabei geht es keineswegs nur um das Festhalten, Auswerten und Veranschaulichen bestimmter Daten, sondern es geht oft auch um das Prüfen wissenschaftlicher Hypothesen, um das Prüfen von Annahmen also, das zur Erweiterung und Vervollständigung unserer Erkenntnisse über Natur und Gesellschaft führt.

● 1 Überlegen Sie, wo Ihnen das Wort „Statistik" schon überall begegnet ist!

● 2* In welchen Unterrichtsfächern waren Sie bereits mit der Erfassung und graphischen Darstellung statistischer Zahlenangaben beschäftigt?

● 3 Entnehmen Sie dem Bild 1, wie viele Geburten je 1000 Einwohner 1979 in der DDR und wie viele in der BRD zu verzeichnen waren! Worauf führen Sie den Unterschied zurück?

Was heißt nun „Statistik"? Aus dem Wort allein können wir die heutige Bedeutung nicht erfassen. „Statistik" geht auf das italienische Wort ‚statista' zurück, das soviel wie „Staatskundiger" bedeutet. In der Mitte des 17. Jahrhunderts bildete sich in Italien eine – z. T. auf die Erfassung von Zahlen gerichtete – Staatslehre als selbständiger Wissenschaftszweig heraus, den man etwa als „Lehre von den Staatsmerkwürdigkeiten" beschreiben könnte. Es ging dabei um die Zusammensetzung der Bevölkerung, der Armee und des Gewerbes. Zu Beginn des 18. Jahrhunderts hielt M. SCHMEITZEL an der damals neugegründeten Universität Halle eine Vorlesung unter dem Thema „collegium politico-statisticum" und wurde so zum Namensgeber der Statistik im deutschsprachigen Raum. Natürlich ist das Bemühen der Menschen um das Erfassen und Auswerten statistischer Zahlenangaben historisch schon viel früher nachweisbar. Die Zählung von Personen, Vieh oder Vorräten geht zurück bis ins Altertum.

■ **B 1** Schon im alten Reich Ägyptens (2650–2190 v. u. Z.) wurden alle zwei Jahre Zählungen des Geldes und der Felder durchgeführt.
Im alten Rom bestand seit 433 v. u. Z. ein „Volkszählungsbüro", das regelmäßige Erhebungen von Bevölkerungsdaten veranlaßte. So fanden bis zum Jahre 37 u. Z. 69 Volkszählungen statt.
Zur Zeit der Herrschaft KARLS des GROSSEN (768–814) wurden Güter- und Besitzverzeichnisse angelegt, in denen neben Personen auch Wohnräume aller Art, Getreidebestände und das Vieh – getrennt nach Arten und Alter – erfaßt wurden.

Die praktische Statistik blickt also auf eine viereinhalb Jahrtausende währende Geschichte zurück. Dagegen ist die theoretische Statistik – heute als mathematische Statistik bezeichnet – relativ jung. Als eigenständiges Wissenschaftsgebiet gibt es sie erst seit etwa 100 Jahren, und eine Vielzahl bedeutender statistischer Verfahren wurde gar erst in den letzten Jahrzehnten entwickelt. Dieses rasche Voranschreiten hält noch immer an. So ist die Statistik eine sehr alte und sehr junge Wissenschaft zugleich.

Was verstehen wir heute unter Statistik, und welche Teile dieser Wissenschaft gilt es zu unterscheiden? Alle statistischen Angaben haben ihren Ursprung in unserem täglichen Leben und in unserer Arbeit. So enthält z. B. das Klassenbuch Zusammenstellungen verschiedener Angaben, häufig in Form von Zahlen (z. B. Zensuren). In Betrieben werden die Produktionsergebnisse in Listen erfaßt.

■ **B 2** Von 6 Arbeitern einer Brigade eines Werkzeugmaschinenbetriebes wurden für den Zeitraum einer Woche folgende Zahlenangaben erfaßt und berechnet. (Die Norm betrage 20 Stück je Stunde.)

Arbeiter	Arbeitszeit (in h)	Produktion (in St.)	Normerfüllung (in %)
A	48	1250	130,2
B	43	830	96,5
C	44	1040	...
D	44	920	...
E	48	1100	...
F	40	860	...

● 4* Berechnen Sie die fehlenden Prozentangaben in der Spalte „Normerfüllung“!

● 5* Als wichtige Kennzahl wird oft der Durchschnittswert, das arithmetische Mittel, bestimmt. Überlegen Sie, wie man vorgehen müßte, um die durchschnittliche Normerfüllung aller 6 Arbeiter zu berechnen!
Verzagen Sie nicht, wenn Sie diese Aufgabe nicht auf Anhieb lösen! Wir kommen später darauf zurück.

Alle Arbeitsergebnisse, die täglich in den Betrieben und Werkstätten anfallen, werden auf Formblättern festgehalten und abgerechnet. Eine Vielzahl von Zahlen wird erfaßt, wir bezeichnen sie meist als **Daten**. In der weiteren Verarbeitung werden daraus z. B. Material- oder Lohnabrechnungen gewonnen. Das kann unter Zuhilfenahme von Elektronischen Datenverarbeitungsanlagen (EDVA) geschehen.
Die Daten werden mit anderen zusammengefaßt oder verglichen, es geht also in der Statistik nie um eine Einzelerscheinung, sondern stets um das Erfassen und Untersuchen von Massenerscheinungen. Unter **Massenerscheinungen** verstehen wir Tatbestände, deren wesentlichste Eigenschaft die Wiederholung ist und bei denen (in stärkerem oder schwächerem Maße) Zufallseinflüsse auftreten.

Mit dem für die Statistik wichtigen Begriff des Zufalls werden wir uns im Abschnitt 1.2. ausführlicher beschäftigen. Es gilt aber schon hier zu erkennen, daß wir mit dem Alltagsbegriff des „Zufälligen“ nicht viel anfangen können.

Die Untersuchung von Massenerscheinungen in Natur und Gesellschaft führt zu verallgemeinerungsfähigen Erkenntnissen über das Untersuchungsmerkmal, z. B. über die Körpergröße 15jähriger Mädchen in der DDR, die nicht gewonnen werden könnten, wenn wir nur die Einzelerscheinung (die Körpergröße eines einzelnen 15-jährigen Mädchens) messen würden. Aus der Messung einer relativ großen Anzahl von Mädchen lassen sich aber zuverlässige Schlüsse ziehen über das Auftreten und die Häufigkeit dieser und jener Körpergrößen und damit über die Konfektionsgrößen (z. B. Hosen- oder Rocklängen), die in der Textilindustrie benötigt werden.
Ein Hauptanliegen der Statistik besteht also darin, aus den erhobenen Daten Kennzahlen abzuleiten, die für weite Bereiche von Wirtschaft, Industrie, Bildung usw. von großer Bedeutung sind.

Wir sind jetzt in der Lage, eine Definition der Statistik zu geben.

► **D 1 Die *Statistik* ist die Wissenschaft von der zahlenmäßigen Erfassung und Untersuchung von Massenerscheinungen in Natur und Gesellschaft, bei denen Zufallseinflüsse wirken.**

● 6* Es wurden erfaßt und untersucht
a) die Qualität von Fernsehbildröhren, die im Laufe eines Monats von einem VEB produziert wurden,
b) der Brustumfang (in cm) von 1000 jungen Männern im Alter von 18 Jahren,
c) die Ursache der Krankheit eines zweijährigen Kleinkindes,
d) der Durchmesser (in mm) automatisch gefertigter Wellen einer bestimmten Größe und
e) die tägliche Fördermenge (in t) an Braunkohle im Tagebau Schleenhain.
Eines dieser fünf Beispiele gehört *nicht* zum Untersuchungsgegenstand der Statistik. Nennen Sie das Beispiel, und begründen Sie Ihre Aussage!

Es darf nicht übersehen werden, daß das Wort „Statistik" in der Alltagssprache oft in einer anderen Bedeutung verwendet wird, nämlich als Bezeichnung für in Tabellenform vorliegende statistische Erhebungen oder in Tabellenform gebrachte Ergebnisse solcher Erhebungen (z. B. Fußball-Statistik, Temperaturstatistik).
In den weiteren Darlegungen wird „Statistik" nur in dem oben definierten Sinne verstanden, also als Bezeichnung einer Wissenschaftsdisziplin.
Unsere Definition der Statistik ist sehr umfassend. Die Statistik setzt sich aus mehreren Teilen zusammen. Zwei wichtige seien genannt:
In der **beschreibenden Statistik** erfolgen Erfassung, Aufbereitung und Darstellung statistischer Daten sowie das Berechnen von Kennzahlen (statistischer Maßzahlen wie Mittelwerte und Streuungsmaße, auf die wir im 2. Abschnitt zurückkommen werden), von Zusammenhängen zwischen Zufallsgrößen und anderen Ausdrucksmitteln der Statistik. In der **Prüfstatistik** steht das Schätzen unbekannter Parameter und das Prüfen von Hypothesen zur Stützung wissenschaftlicher oder technischer Erkenntnisse im Vordergrund.

■ **B 3** Das Bild 1 (↗ S. 6) zeigt die Anwendung einer Methode der beschreibenden Statistik – der graphischen Darstellung – auf einen bevölkerungsstatistischen Sachverhalt.

■ **B 4** Das Bild 2 zeigt eine Anwendung der beschreibenden Statistik in einem Teilgebiet der Wirtschaftsstatistik. Beachten Sie, daß man sich bei einer solchen Darstellung stets erst vergewissern muß (an der Überschrift und an den Bezeichnungen), welcher Sachverhalt der objektiven Realität jeweils wiedergegeben wird.

● **7** Welche Erkenntnisse können Sie dem Bild 2 entnehmen? Diskutieren Sie eventuelle Ursachen!

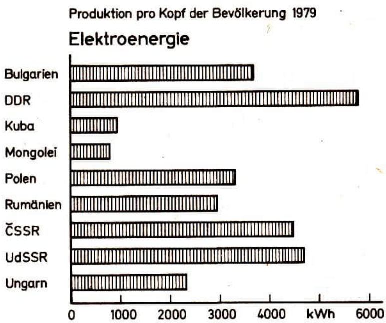


Bild 2

Das Beispiel B 4 mit seinen Schlußfolgerungen soll verdeutlichen, daß die Statistik für die Leitung von Wirtschaftsprozessen eine große Unterstützung ist und daß sie der Gewinnung neuer Erkenntnisse dient.
W. I. Lenin hat in zahlreichen Arbeiten auf die Bedeutung der Statistik sowohl für die theoretischen Untersuchungen als auch für die Planung und Kontrolle im sozialistischen Staat hingewiesen. Er fordert immer wieder die korrekte Anwendung statistischer Methoden. So schreibt er in seinem Aufsatz „Statistik und Soziologie":
„Tatsachen sind, nimmt man sie in ihrer Gesamtheit, in ihrem Zusammenhang, nicht nur ‚hartnäckige', sondern auch unbedingt beweiskräftige Dinge. Nimmt man aber einzelne Tatsachen losgelöst vom Ganzen, losgelöst aus ihrem Zusammenhang, sind die Daten lückenhaft, sind sie willkürlich herausgegriffen, dann ist dies eben nur ein Jonglieren mit Daten oder etwas noch Schlimmeres." [23]
Diese für den Umgang mit statistischem Datenmaterial grundlegende Erkenntnis

kann nicht hoch genug eingeschätzt werden. Aus ihr geht die große Verantwortung hervor, die der Forscher oder Praktiker bei der Erhebung, Auswertung, Analyse und Interpretation statistischer Daten trägt.
In den letzten Jahrzehnten hat die Statistik einen gewaltigen Aufschwung genommen. Es waren insbesondere sowjetische Wissenschaftler wie A. N. KOLMOGOROV (↗ Bild 3), A. J. CHINTSCHIN und B. W. GNEDENKO, die wesentliche Beiträge zur Weiterentwicklung der mathematischen Statistik, der theoretischen Grundlage der gesamten Statistik, und ihrer Anwendung in der Praxis lieferten.

Bild 3

Die Statistik ist heute aus dem Leben der Menschen und der Völker, aus der Entwicklung der Staaten und ihrer Wirtschaft nicht mehr wegzudenken. Die Erfolge beim Aufbau und Voranschreiten der sozialistischen Gesellschaft in der DDR spiegeln sich wider im Statistischen Jahrbuch der Deutschen Demokratischen Republik, das jährlich erscheint und von der Staatlichen Zentralverwaltung für Statistik beim Ministerrat der DDR herausgegeben wird. Die immer enger werdende Verflechtung der Volkswirtschaft der Mitgliedsländer des Rates für gegenseitige Wirtschaftshilfe (RGW) erfordert einheitlich gestaltete und damit vergleichbare statistische Informationen aus den Mitgliedsländern. So besteht im Rahmen des RGW eine „Ständige Kommission für Statistik". Auch die Organisation der Vereinten Nationen (UNO) könnte sich nur ein unvollständiges Bild von der Bevölkerungsentwicklung, von den Ernteergebnissen, von der Industrieproduktion usw. in den einzelnen Ländern und Regionen der Welt machen, wenn nicht eine regelmäßige Erhebung statistischer Daten erfolgen würde. Beispiele werden mit dem Bild 4 und mit den Tabellen 1 und 2 angefügt.

Tabelle 1 **Schiffsbestand der Handelsflotte der DDR**

Jahr	Anzahl	Wasserverdrängung BRT	Tragfähigkeit tdw
1952	1	917	1250
1957	21	35416	46859
1962	82	350648	482424
1967	162	755824	1021245
1972	194	1027671	1463869
1977	200	1259074	1853914
(1980	192	1305084	1877370)

(Quelle: Statistisches Jahrbuch der DDR 1981)

Tabelle 2 **Berufstätige in der DDR (in Millionen)**

Jahr	Industrie	Handel	Land- und Forstwirtschaft
1950	2,098	0,674	2,005
1960	2,768	0,890	1,304
1970	2,855	0,858	0,997
1980	3,128	0,850	0,878

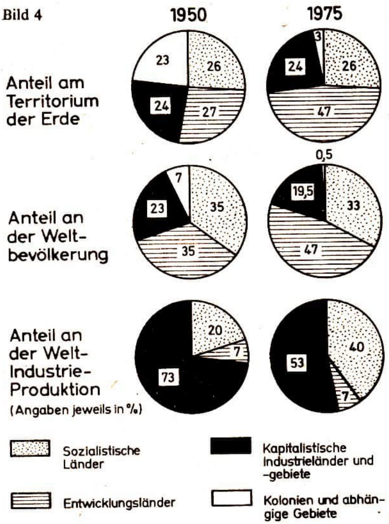
(Quelle: Statistisches Jahrbuch der DDR 1981)

● **8** Weisen Sie anhand der beiden unteren Kreisdiagramme des Bildes 4 nach, wie sich die Industrieproduktion auf unserer Erde in 25 Jahren zugunsten des sozialistischen Lagers verändert hat!

● **9*** Auf wieviel Prozent ist der Schiffsbestand der Handelsflotte der DDR in den 23 Jahren von 1957 bis 1980 gestiegen? (Der Bestand 1957 werde mit 100% angesetzt.) Auf wieviel Prozent stieg in der gleichen Zeit die Tragfähigkeit der Flotte?

● **10** Welche weiteren Erkenntnisse entnehmen Sie dem Bild 4 sowie den Tabellen 1 und 2 auf der Seite 10?

Bild 4



Im folgenden werden wir uns nicht auf die Erörterung graphischer und tabellarischer Darstellungen beschränken. Wir wollen vielmehr auch lernen, selbständig statistische Daten zu erfassen, aufzubereiten und auszuwerten.

Zusammenfassung des Abschnitts 1.1.

> Die Statistik ist ein wichtiges Instrument für die Leitung, Planung und Erforschung von Erscheinungen und Prozessen in der sozialistischen Gesellschaft sowie zum Kenntlichmachen von Entwicklungen in der Welt.
> Sie stellt Methoden bereit, Daten aus allen Bereichen der Volkswirtschaft, Wissenschaft und Technik zu erfassen, aufzubereiten, darzustellen, zu analysieren und daraus Schlußfolgerungen für die erfolgreiche tägliche Arbeit des Einzelnen sowie für die weitere Gestaltung der entwickelten sozialistischen Gesellschaft als Ganzes abzuleiten.

1.2. Notwendigkeit und Zufall

Wenn uns ein Gegenstand entgleitet, so fällt er zu Boden. Das ist fest vorausbestimmt – streng determiniert, wie man auch sagt. Der Grund dafür ist in der Erdanziehungskraft zu suchen, die auf alle Körper einwirkt und zum Mittelpunkt der Erde hin gerichtet ist. Die Erdanziehungskraft folgt dem Massenanziehungsgesetz, einem allgemeinen Naturgesetz.
Man kann also mit voller Gewißheit voraussagen, in welcher Richtung sich der

entglittene Gegenstand bewegt. Der Vorgang ist eindeutig bestimmt; man sagt, er verläuft mit **Notwendigkeit**.

Betrachten wir den Landevorgang eines Raumschiffes. Es gilt, sämtliche Faktoren, die die Bahn in der letzten Phase des Unternehmens beeinflussen, mit hoher Genauigkeit zu erfassen. Dazu gehören:

- der Zustand der atmosphärischen Schichten, die das Raumschiff durchquert,
- die Massenanziehungskräfte,
- Öffnungszeitpunkt und wirksame Flächen des Landefallschirmes usw.

Eine genaue Messung aller Gegebenheiten der Luftmassen in der Atmosphäre bereitet jedoch große Schwierigkeiten und ist nicht mit absoluter Verläßlichkeit möglich. Auch bei diesem Rückkehrvorgang könnten *prinzipiell* alle auf das Landeaggregat wirkenden Einzelkräfte berücksichtigt, berechnet und damit angegeben werden, wohin es sich exakt bewegt, aber das ist aufgrund der Vielzahl der beteiligten Kräfte und der Komplexheit ihres Zusammenwirkens nahezu unmöglich.

Damit erscheint es uns als **Zufall**, an welchem Punkt in der Nähe des Zielgebietes das Auftreffen der Landekapsel letztendlich erfolgt.

Der Zufall ist also immer dann im Spiel, wenn die Bedingungen und Voraussetzungen für den betrachteten Vorgang nur unzureichend bekannt oder nicht eindeutig bestimmbar sind, das heißt, wenn mehrere Möglichkeiten für den Ausgang des Geschehens bestehen. Wir müssen uns aber der Tatsache bewußt sein, daß es den absoluten Zufall nicht gibt; auch das Zufällige ist ursächlich bedingt. Damit ist der Zufall die *Ergänzung* der Notwendigkeit.

Notwendigkeit und Zufall bilden eine Einheit; sie sind dialektisch miteinander verbunden. Und das gilt nicht nur für Vorgänge in der Natur, sondern ebenso für Vorgänge in der Gesellschaft.

Wann beim einzelnen Menschen der Tod eintritt, ist in den meisten Fällen eine zufällige Erscheinung, *daß* er eintritt, geschieht mit Notwendigkeit. Das Sterbealter des einzelnen Menschen ist zufällig, betrachtet man jedoch eine große Anzahl von Menschen und erfaßt deren Sterbealter in Listen, so ergeben sich Gesetzmäßigkeiten, die für den Versicherungsmathematiker eine wichtige Berechnungsgrundlage darstellen. Es gibt also **statistische Gesetzmäßigkeiten**, die erst bei Massenerscheinungen, beim Beteiligtsein vieler Individuen[1]) oder Objekte offenbar werden. Der Begriff der statistischen Gesetzmäßigkeit – wir werden ihm immer wieder begegnen – besagt also, daß es sich nicht um die Gesetzmäßigkeit einer Einzelerscheinung handelt, sondern vielmehr um die gesetzmäßigen Zusammenhänge einer *Gesamtheit gleichartiger Erscheinungen.* Diese sind im allgemeinen mit dem Einzelfall nicht identisch und können von ihm her auch niemals restlos erfaßt werden. Mit dem Begriff der statistischen Gesetzmäßigkeit wird zum Ausdruck gebracht, daß sich die Notwendigkeit – als Inbegriff aller zum *Wesen* der Erscheinung gehörenden Bedingungen – in einer Anzahl von Zufällen – als der Gesamtheit aller hinzutretenden Bedingungen – durchsetzt. Das sei an einem einfachen Beispiel anschaulich dargestellt.

■ **B 5 (1)** Beim Würfeln mit einem echten Spielwürfel erscheint möglicherweise zuerst eine „3", dann eine „1", dann vielleicht eine „4", wiederum eine „3" usw. Wir hoffen auf eine „6"; diese erscheint aber eine ganze Weile nicht. Doch – wie es der Zufall so will – kommt sie plötzlich dreimal hintereinander. Die Einzelerscheinung hat also zufälligen Charakter.

Nun betrachten wir nicht nur einige Würfe, sondern viele: 50 Würfe, dann 100, weiterhin 500, schließlich 1000 und gar 5000 Würfe. Dabei kommt es uns jedesmal darauf an, wie oft die Augenzahl „6" erscheint, mit welcher Häufigkeit sie auftritt.

[1]) Individuum (lat.) – das Einzelwesen (im Verhältnis zur Gemeinschaft)

In der folgenden Tabelle wird eine solche Untersuchung dargestellt. Dabei wurde für verschiedene Anzahlen n der Würfe die Häufigkeit h, mit der eine 6 erscheint, ermittelt. Außerdem wurde in einer dritten Spalte die sogenannte „relative Häufigkeit" vermerkt. Das ist die Häufigkeit h, bezogen auf die Anzahl n der Würfe, also der Quotient $\frac{h}{n}$.

Anzahl n der Würfe	Häufigkeit h, mit der „6" erscheint	relative Häufigkeit $\frac{h}{n}$
50	7	0,140
100	18	0,180
500	82	0,164
1000	168	0,168
5000	833	0,167
…	…	…

BROWNsche Bewegung eines Moleküls

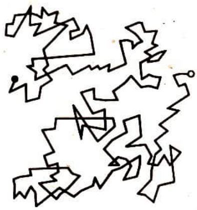


Bild 5

Was fällt uns auf? Je mehr Würfe wir durchführen, um so mehr nähert sich die relative Häufigkeit einem konstanten Wert, um so deutlicher tritt eine Gesetzmäßigkeit hervor – hier die Annäherung an den Wert $\frac{1}{6} = 0{,}166\ldots$ In der Vielzahl von Zufällen (den Einzelwürfen) setzt sich die Notwendigkeit (in Form der statistischen Gesetzmäßigkeit) durch.

● **11*** Zum Beispiel 5(1): Welche relative Häufigkeit für das Auftreten der Augenzahl „6" würde sich aller Voraussicht nach bei 10000 Würfen einstellen?

● **12** Werfen Sie einen Würfel 25mal, bestimmen Sie Häufigkeit und relative Häufigkeit, mit der die Augenzahl „1" erscheint, und erörtern Sie das Resultat!

Das oben Angeführte zeigt deutlich, daß der Zufall stets allgegenwärtig ist. Die Wirkung des Zufälligen in Natur, Technik und Gesellschaft ist groß. Einige Beispiele seien dazu angeführt.

■ **B 6** **a)** Die Entstehung des Lebens wie auch die Veränderungen (Mutationen) im Entwicklungsprozeß des Lebens stellen Zufallserscheinungen dar. Diesen ist es u. a. zu verdanken, daß sich der Mensch von primitiven Vorstufen bis zu seiner heutigen Leistungsfähigkeit entwickeln konnte.

b) Der Zufall ist beherrschendes Element der BROWNschen Molekularbewegung, die sich in festen, flüssigen und gasförmigen Körpern vollzieht und durch die Molekülstöße hervorgerufen wird (↗ Bild 5).
In der statistischen Physik ist es möglich, aus den Eigenschaften dieser Massenerscheinung über Mittelwertbildung auf wichtige physikalische Größen wie Druck, Temperatur, elektrische Leitfähigkeit usw. des betrachteten Körpers zu schließen.

c) Auch Psychologen und Militärwissenschaftler, die oft über das bestmögliche Verhalten in Konfliktsituationen zu entscheiden haben, erkannten die Bedeutung des Zufalls. Sie überzeugten sich davon, daß häufig das an Zufallsgesetzen orientierte Verhalten am vorteilhaftesten ist.

● **13** Stellen Sie das dialektische Wechselverhältnis von Notwendigkeit und Zufall an einem selbstgewählten Beispiel dar! Beziehen Sie Erscheinungen, die

Sie im naturwissenschaftlichen Unterricht kennengelernt haben, in diese Betrachtung ein!

Zusammenfassung des Abschnitts 1.2.

> Notwendigkeit und Zufall bilden eine dialektische Einheit. Auch der Zufall ist eine objektive Erscheinung. Das gilt für Vorgänge in der Natur wie in der Gesellschaft.
> Auch den zufälligen Erscheinungen und Prozessen wohnen Gesetzmäßigkeiten inne, sogenannte statistische Gesetzmäßigkeiten. Das sind gesetzmäßige Zusammenhänge einer Gesamtheit gleichartiger Erscheinungen, die bei Betrachtung einer großen Zahl solcher Erscheinungen deutlich werden.
> Diese statistischen Gesetzmäßigkeiten gilt es aufzudecken.

1.3. Zufallsgrößen

Wir betrachten im folgenden einige Berufe und mit diesen verbundene Aufgaben, zu deren Bewältigung statistische Kenntnisse benötigt werden. In der letzten Spalte sind die jeweils zu untersuchenden Größen angeführt.

Beruf Betrieb oder Einrichtung	Kennzeichnung der „statistischen Aufgabe“	Zu untersuchende Größe
Schlosser Maschinenfabrik	Ermitteln der Toleranz einer Bohrung Ermitteln des Durchmessers automatisch gefertigter Wellen	Toleranz Durchmesser
Ökonom Kombinat	Untersuchen der Normerfüllung Ermitteln von Ausfallzeiten	Normerfüllung Ausfallzeit je Maschine
Oberarzt-Sekretärin Kinderklinik	Analyse von Körpergröße und Körpergewicht von Neugeborenen Bestimmen der Säuglingssterblichkeit	Körpergröße Körpergewicht Säuglingssterblichkeit
Facharbeiter für Elektronik Fernsehgerätewerk	Qualitätskontrolle Messen der Arbeitszeit bei Arbeiten am Band	Qualität von Bildröhren Arbeitszeit/Arbeitsgang
Facharbeiter für Pflanzenproduktion LPG	Ermitteln des durchschnittl. Hektarertrags Untersuchen des Aussaattermins	Hektarertrag Zeitpunkt der Aussaat
Verkehrspolizist Volkspolizeikreisamt	Untersuchen der Verkehrsdichte einer Kreuzung Feststellen der Blutalkoholkonzentration	Zahl der Fahrzeuge je Stunde Blutalkoholkonzentration
Meteorologe Wetterwarte	Untersuchen der Niederschlagsmenge und Sonnenscheindauer	Niederschlagsmenge Sonnenscheindauer

In dieser Übersicht werden zu jedem der angeführten Berufe stets nur ein oder zwei mit statistischen Mitteln zu lösende Aufgaben genannt. In Wirklichkeit sind es natürlich viel mehr. Bei bestimmten Aufgaben ist es auch erforderlich, den Zusammenhang und die Wechselwirkung zwischen zwei oder mehreren Größen zu ermitteln, um so zu neuen Erkenntnissen in Wissenschaft, Technik oder Praxis zu gelangen.
Wir wollen vorerst zu jeder Aufgabe jeweils nur *eine* Größe betrachten.

● **14** Ergänzen und erweitern Sie die Tabelle, indem Sie zu den angeführten Berufen weitere Aufgaben und die jeweils zu untersuchende Größe nennen und indem Sie andere Berufe anfügen und einige mit statistischen Mitteln zu lösende Arbeitsaufgaben sowie die dabei zu untersuchenden Größen aufschreiben! Fragen Sie auch Ihre Eltern und deren Bekannte!

Was ist nun kennzeichnend für all diese Aufgaben und für die entsprechenden Größen? Eines wissen wir schon: Die Aufgaben lassen sich mit Hilfe der Statistik lösen, also handelt es sich um die Erfassung, Aufbereitung und Auswertung von Massenerscheinungen. Die Größen sind Längen, Zeiten, Qualitäten, Erträge, Anzahlen usw. Sie werden gemessen, beobachtet, geschätzt oder gezählt; und zwar jeweils *an ausgewählten oder an allen* Elementen einer entsprechenden Grundmenge.

■ **B 7** **a)** Der Wellendurchmesser wird gemessen an *ausgewählten* Wellen der insgesamt in der betreffenden Serie gefertigten.
b) Am Ende des Schuljahres wird der Leistungsstand *aller* Schüler ermittelt.

Die Elemente werden **Untersuchungsobjekte** genannt, das sind Werkstücke, Schüler, Bildröhren usw.
Die Grundmenge heißt Grundgesamtheit.

► **D 2 Unter der *Grundgesamtheit* verstehen wir eine nichtleere Menge gleichartiger Objekte, an denen eine bestimmte Größe untersucht werden soll.**

■ **B 8** **a)** Grundgesamtheit im Beispiel 7a) sind alle Wellen der betreffenden Serie;
Untersuchungsobjekte sind die für die Messung ausgewählten Wellen.
b) Grundgesamtheit im Beispiel 7b) sind alle Schüler;
Untersuchungsobjekte sind alle Schüler.

Die Anzahl N der Elemente der Grundgesamtheit wird **Umfang der Grundgesamtheit** genannt.

An den Untersuchungsobjekten erfolgen nun – *unter konstantgehaltenen äußeren Bedingungen* – die Messungen, Beobachtungen, Zählungen usw. Es liegt in der Natur der Sache, daß bei Wiederholung des Untersuchungsvorganges (Messen des Durchmessers automatisch gefertigter Wellen, Ermitteln der Zeit für einen Arbeitsgang usw.) nicht stets ein und derselbe Wert erscheint, sondern verschiedene Werte auftreten. Solche Größen, die sich erst durch konkretes Messen, Beobachten, Zählen ergeben, heißen **Zufallsgrößen** (auch: Zufallsvariable).

► **D 3 Eine *Zufallsgröße* ist eine Funktion, die ihre Werte in Abhängigkeit vom Zufall annimmt.
Die Werte heißen *Realisationen der Zufallsgröße*.**

Ein äußeres Kennzeichen einer jeden Zufallsgröße ist die Menge der möglichen Werte, die sie annehmen kann; welchen ihrer Werte sie aber bei Durchführung einer entsprechenden Messung oder Beobachtung annehmen wird, kann nicht mit Bestimmtheit vorausgesagt werden, sondern nur mit einer gewissen Wahrscheinlichkeit (↗ Abschnitt 3).
In den Bereichen der Praxis trifft man statt des Begriffs Zufallsgröße oft den Begriff **Merkmal** an.

▶ **E 1 Merkmale** sind bestimmte Charakteristika des Untersuchungsobjekts. Merkmale fassen wir als Zufallsgrößen auf. Die Einzelwerte heißen **Merkmalswerte** oder **Beobachtungswerte**, die überhaupt möglichen (voneinander verschiedenen) Merkmalswerte sind die **Merkmalsausprägungen**.
Diese sind **quantitativer Art**, wenn sie durch Zahlen dargestellt werden oder dargestellt werden können, andernfalls sind sie **qualitativer Art**.

Wir werden diesen Begriffen so oft begegnen, daß sie uns bald in Fleisch und Blut übergehen. Zufallsgrößen (Merkmale) findet man bei fast allen Meß- und Beobachtungsvorgängen, sie dienen der direkten oder indirekten Kennzeichnung eines bestimmten Sachverhalts. Sämtliche in der letzten Spalte der Tabelle (S. 14) dargestellten Größen sind Zufallsgrößen (Merkmale).

Es ist üblich, Zufallsgrößen und damit auch Merkmale mit großen lateinischen Buchstaben zu kennzeichnen, meist mit X, Y, Z, U, V, W, ihre Realisationen (Merkmalswerte) mit kleinen lateinischen Buchstaben, die mit Indizes versehen werden, also $x_1, x_2, x_3, \ldots$ für X; $u_1, u_2, u_3, \ldots$ für U usw.

Wir betrachten zwei konkrete Beispiele, die uns die Begriffe näherbringen werden:

■ **B 9 (1)** Um die Leistung von zwei Lehrlingen am Ende des ersten Ausbildungsjahres einschätzen zu können, wurden die von ihnen zur Montage einer bestimmten Sorte von Steckern benötigten Arbeitszeiten ermittelt. Diese Zeiten sollen mit der Facharbeiternorm (45 s) verglichen werden. Bei jedem Lehrling wurde dazu die pro Stecker benötigte Arbeitszeit gemessen, und zwar hintereinander an 50 Steckermontagen.
Man erhielt folgende Werte (in s):

Lehrling A:

50	44	42	43	45	43	43	51	42	41
43	44	43	43	42	45	47	45	46	44
43	40	44	45	47	45	44	44	41	49
47	43	43	42	44	41	45	46	46	42
41	41	43	45	44	45	45	44	43	42

Lehrling B:

44	43	43	44	42	41	46	44	45	43
45	45	44	45	45	44	43	42	43	43
41	41	44	43	42	41	42	43	42	42
42	41	42	41	40	41	43	44	43	43
41	44	44	44	41	42	42	46	42	41

▶ **E 2** Eine Auflistung der gewonnenen Daten so, wie sie beobachtet oder gemessen werden, nennt man eine **Urliste**.

Hier liegen also zwei Urlisten vor. In einem späteren Abschnitt werden wir die Werte ordnen.
Vorerst ist für uns wichtig zu wissen, was es mit dieser Untersuchung auf sich hat.

Untersuchungsobjekte sind hier die Steckermontagen.
Zufallsgrößen oder untersuchte *Merkmale* sind hier: X = Arbeitszeit des Lehrlings A je Steckermontage und Y entsprechend für Lehrling B (jeweils in s).
Als *Meßinstrument* werden Stoppuhren eingesetzt.
Die überhaupt möglichen Meßergebnisse für „Arbeitszeit je Steckermontage" sind die *Merkmalsausprägungen* [z. B. das Intervall (35 s; 60 s)].
Die Meßergebnisse sind auf volle Sekunden gerundet.
Art der Merkmalsausprägungen: quantitativ
Die tatsächlich erhaltenen Meßergebnisse sind die *Merkmalswerte* oder *Daten* (↗ Urlisten).

■ **B 10 (1)** Um die Verkehrsbelegung einer großen ampelgeregelten Kreuzung in den Abendstunden zu ermitteln, wurde die Zahl aller Fahrzeuge festgestellt, die die Kreuzung während einer vollen Ampelphase (hier: 3 min, sonst oft kürzer) befahren. Es wurde an 3 Werktagen, jeweils in der Zeit von 18.30 bis 19.00 Uhr, gezählt. Die Untersuchung ist Teil einer Entscheidungsvorlage für das Abschalten der Ampel in den Abendstunden. Es ergaben sich folgende Werte (Fahrzeuge/Ampelphase):

1. Tag:	14	11	27	8	21	17	13	18	9	12
2. Tag:	20	14	15	23	7	14	18	16	11	13
3. Tag:	17	13	13	21	19	17	7	12	20	14

Untersuchungsobjekte: Ampelphasen an einer Kreuzung

Das könnte Verwunderung hervorrufen, es stimmt aber. Nicht an den Fahrzeugen untersuchen wir ein Merkmal, sondern wir untersuchen das Merkmal „Anzahl der Fahrzeuge ..." an (oder während) einer Ampelphase. Die Ampelphasen sind also die Untersuchungsobjekte.

Zufallsgröße (untersuchtes Merkmal): Z = Anzahl aller Fahrzeuge, die je Ampelphase die Kreuzung in einem bestimmten Zeitabschnitt ($\frac{1}{2}$ Stunde) queren
Meßinstrument: Zählende Personen
Merkmalsausprägungen: 0 bis 100 Fahrzeuge/Ampelphase
Art der Merkmalsausprägungen: quantitativ
Merkmalswerte oder *Daten:* Urliste

An diesen Beispielen wird deutlich, wie exakt man schon bei der Daten*erfassung* vorgehen muß; genauso sorgfältig wie bei allen weiteren statistischen Teilarbeiten. Es ist nicht ratsam, nur eine Person zählen zu lassen, sondern es sollten zwei oder drei sein, die *unabhängig* voneinander beobachten (möglichst von verschiedenen Stellen der Kreuzung aus) und die Zahl der Fahrzeuge feststellen. Die Einzelwerte sind miteinander zu vergleichen, und nur bei Übereinstimmung dürfen sie weiter verwendet werden. Sonst ist die **Objektivität** der Untersuchung nicht gesichert.

Wenn Sie selbst an Untersuchungen beteiligt sind, gilt es also zu beachten:

Exakt messen oder beobachten!
Werte sorgfältig notieren!
Werte mit denen der anderen Beobachter vergleichen und sie nur bei Übereinstimmung weiterverwenden! Andernfalls Untersuchung wiederholen!

(Näheres zur Untersuchungsstrategie ↗ unter 2.1.)
Es ist notwendig, sich zu Beginn einer jeden Untersuchung – möglichst schon in der Planung vor der eigentlichen Untersuchung – einen Überblick über Untersuchungs-

objekte, Zufallsgrößen, Meßinstrumente, Merkmalsausprägungen und Art der Merkmalsausprägungen zu verschaffen, so, wie es in den Beispielen 9 und 10 geschehen ist. Vor allem auf die genaue Kennzeichnung der Zufallsgröße (des Merkmals) kommt es an!
Nicht minder wichtig ist die Überlegung, ob eine **Einheit** zu der gerade betrachteten Zufallsgröße gehört und, wenn ja, welche.

■ **B 11** Die Zufallsgröße (das Merkmal) „Geschlecht" hat keine Einheit.
Die Zufallsgröße „Arbeitszeit je Steckermontage" im Beispiel 9 trägt die Einheit s; die Zufallsgröße „Anzahl aller Fahrzeuge ..." im Beispiel 10 trägt die Einheit Fahrzeuge/Ampelphase.

● **15*** Geben Sie jeweils die Einheiten zu folgenden Zufallsgrößen an:
Bolzenlänge, Hektarertrag, Normerfüllung, Niederschlagsmenge!

● **16*** Geben Sie die Ausprägungen und deren Art für die folgend aufgeführten Merkmale an!
Geschlecht, Lebensalter des Menschen, Zahl der Ferkel je Wurf, Benzinverbrauch des Pkw „Trabant", ausgeübter Beruf von Werktätigen

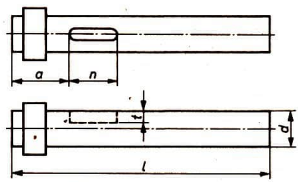
● **17*** Bei der Fertigung einer Welle ist eine Nut vorgesehen, die eine Tiefe von 3,0 mm bis 3,5 mm (zulässige Toleranz) haben darf (↗ Bild 6). Beim Nachmessen von 40 fertiggestellten Wellen ergaben sich folgende Werte für die Nutentiefe.

Urliste:

3,0	4,1	3,5	4,9	3,0	3,2	3,0	3,2	3,0	2,9
2,9	3,9	3,1	2,9	3,3	3,5	3,9	3,1	3,2	3,0
3,0	3,9	3,0	3,1	2,9	2,9	3,2	2,7	2,9	3,2
3,0	3,3	3,0	3,3	2,9	3,0	3,0	3,5	3,1	3,4

Welle mit Nut

Bild 6



l Gesamtlänge
d Wellendurchmesser
n Nutenlänge
t Nutentiefe
a Nutenabstand

a) Geben Sie für diese Untersuchung an:
Untersuchungsobjekte, Zufallsgröße (Merkmal) mit Einheit, Meßinstrument, Merkmalsausprägungen (in den hier wirklich auftretenden Grenzen) und Art der Merkmalsausprägungen!
b) Wieviel Prozent aller Werte liegen im zulässigen Toleranzbereich?

● **18*** Für einen sportlichen Leistungsvergleich wird die Körpergröße von 15 Schülern eines Jahrgangs durch die Wörter „klein", „mittelgroß", „groß" oder „sehr groß" gekennzeichnet. Zur Vereinfachung der Schreibarbeit wird folgende Zuordnung (Codierung) genutzt: „K" bedeutet klein, „M" mittelgroß, „G" groß und „S" sehr groß.

Urliste:

G S K G M K M S G S G K M M G

a) Geben Sie hierfür an: Untersuchungsobjekte, untersuchtes Merkmal mit Einheit und die Merkmalsausprägungen!
b) Die Angabe des Meßinstruments wird Ihnen hier sicher schwerfallen. Warum wohl?

Vergleichen wir jetzt einmal die Zufallsgrößen der Beispiele 9 und 10 sowie der Aufgaben 17 und 18 miteinander! Es werden Unterschiede deutlich, die über das bisher Gesagte hinausgehen.

	B 9	B 10	Aufg. 17	Aufg. 18
Zufallsgröße (Merkmal)	Arbeitszeit je Steckermontage	Anzahl der Fahrzeuge je Ampelphase	Nutentiefe	Körpergröße
Einheit	s	Fahrzeuge/Ampelphase	mm	eigentlich cm

Wieso „eigentlich cm"? Betrachten wir noch einmal die Aufgabe 18. Zwar handelt es sich um die Zufallsgröße „Körpergröße", die eigentlich mit einem Längenmaß ermittelt wird, hier geht es jedoch um Ungefähr-Angaben, und dafür wird die Körpergröße nicht *gemessen*, sondern nur *geschätzt* und aufgrund der Schätzung einer der vier Merkmalsausprägungen K, M, G und S zugeordnet.

Wir erhalten also wirkliche **Meßwerte** (im physikalischen Sinne) nur dann, wenn uns ein exaktes Meßinstrument, dessen Maßstab gleichgroße Abstände zwischen den Teilstrichen aufweist, zur Verfügung steht und wir es auch nutzen. Bei einem solchen Maßstab ist gewährleistet, daß die *Differenz* zwischen je zwei benachbarten Strichen *an jeder Stelle* des Maßstabes *gleich* ist. Die Intervallgröße ist konstant, man spricht von einer **Intervallskala** oder **metrischen Skala**.

Diese liegt bei den Meßinstrumenten, die im Beispiel 9 und in der Aufgabe 17 angewandt wurden, sicher vor.

Im Beispiel 10 wurden die Daten durch Zählen ermittelt, was ebenfalls zu den Meßverfahren gehört. So sind auch die hier anfallenden Daten (27 Fahrzeuge/Ampelphase, 8 Fahrzeuge/Ampelphase usw.) Meßwerte.

Es fällt uns nur deshalb etwas schwerer, hier von einer Intervallskala zu sprechen, weil die Fahrzeuge, die die Kreuzung passieren, große Unterschiede aufweisen. Zum Zwecke der Ermittlung der Verkehrsbelegung gilt jedoch jedes Fahrzeug als eine Einheit, unabhängig von Art, Typ oder Größe.

In einigen Anwendungsbereichen sprechen wir auch dann noch von einer Intervallskala, wenn die Forderung nach Gleichabständigkeit der Teilstriche nicht voll gewährleistet ist.

Punkte (für Leistungen in der Schule, im Sport usw.) können nur dann als Meßwerte angesehen werden, wenn jedem Punkt die *gleiche Wertigkeit* zukommt.

Eine Intervallskala mit einem *eindeutig bestimmten absoluten Nullpunkt* bezeichnet man als **Verhältnisskala**.

Wo immer möglich, sollte man Meßwerte anstreben; sie haben den höchsten Informationswert und lassen sich mit statistischen Verfahren am besten auswerten.

In der Aufgabe 18 stehen die Merkmalsausprägungen „klein", „mittelgroß" usw. in einer Ordnungsrelation zueinander. „Sehr groß" ist mehr als „groß", „groß" mehr als „mittelgroß" usw. Daten mit dieser Anordnungseigenschaft heißen **Rangdaten.** Auch hier kann man sich die Merkmalsausprägungen auf einer Skala angeordnet denken. Dieser liegt immer die „Größer-kleiner-Beziehung" zugrunde. Deshalb bezeichnet man sie als **Rang-** oder **Ordinalskala.**

Die Rangskala ist oft dort anzutreffen, wo es um Wertungen oder Schätzungen geht, die durch Zensuren, Platznummern im Sport o. ä. ausgedrückt werden.

Für die Anwendung statistischer Verfahren ist es bei Rangdaten nicht unwichtig, zwischen *Rangwerten* und *Rangplätzen* zu unterscheiden (↗ [24]).

Können *weder* Meß- *noch* Rangwerte erhoben werden, so lassen sich die Daten nur noch in **Kategorien** erfassen. An die Stelle des Messens oder Verwendens einer Rangordnung tritt hier das Kategorisieren, das ist das Einordnen qualitativer Ausprägungen eines Merkmals in Gruppen oder Klassen (Kategorien).

■ **B 12** **a)** Bei 20 Jugendlichen wird die Zufallsgröße „Mitgliedschaft im DTSB" ermittelt. Merkmalsausprägungen sind hier die beiden möglichen Angaben „Mitglied" und „Nichtmitglied". Das sind Kategorien.

b) Zur Zufallsgröße „Familienstand" gehören die qualitativen Merkmalsausprägungen „ledig", „verheiratet", „verwitwet", „geschieden". Hier gibt es also vier Kategorien.

Diese kann man in beliebiger Weise anordnen und auf eine Skala schreiben. Die Anordnung ist nicht zwingend, man kann die Kategorien in beliebiger Reihenfolge angeben. Die Skalenwerte sind nicht mehr durch Zahlen gekennzeichnet, sondern nur noch durch den *Namen* der Ausprägung. Eine solche Skala heißt deswegen **Nominalskala.**

Für die Auswertung auf EDVA werden die Bezeichnungen für die Kategorien verschlüsselt (codiert), also mit natürlichen Zahlen belegt, z. B.: ledig 1, verheiratet 2, verwitwet 3, geschieden 4. Das ändert aber absolut nichts an dem Gesagten, es sind und bleiben Kategorien, eine Rangordnung liegt nicht zugrunde und darf auch nicht hineingedeutet werden.

Wir können nun Meßwerte, Rangdaten und Kategorien voneinander unterscheiden. Wir wissen, daß es Zufallsgrößen gibt, die nur durch Kategorien erfaßt werden können, und andere, die sich mit Hilfe von Meßwerten, Rangdaten oder Kategorien darstellen lassen. Bei den letzteren haben die Daten, die mittels einer Intervallskala gemessen wurden, den höchsten und diejenigen, die mit einer Nominalskala erfaßt wurden, den niedrigsten Informationswert. Der Übergang von einer Datenart zu einer anderen in der Richtung

Meßwerte → Rangdaten → Kategorien

ist bei zahlreichen Zufallsgrößen also möglich, aber mit einem Informationsverlust verbunden.

Wir wollen Zufallsgröße/Merkmal, Merkmalsausprägungen und jeweilige Skala, mit der die Merkmalswerte erfaßt werden, anhand einiger Beispiele einander zuordnen.

■ **B 13**

Zufallsgröße/ Merkmal	Merkmalsausprägungen	Skala
Leistungsstand in Russisch	Noten 1, 2, 3, 4, 5	Ordinalskala
Brustumfang	(40 cm; 150 cm)	Intervallskala
Zahl der Ferkel je Wurf	4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16	Intervallskala
Güte der FS-Bildröhre	standardgerecht, nicht standardgerecht	Nominalskala

● **19*** In Vorbereitung der Radfernfahrt für den Frieden erhalten alle Teilnehmer eine Startnummer, die sie während der ganzen Dauer der Friedensfahrt behalten und auf ihrer Rennhose tragen.
Beim Zieleinlauf wird die Zeitdauer erfaßt, die sie für die Bewältigung der Tagesetappe benötigten, ferner wird festgehalten, an wievielter Position sie die Ziellinie erreichen.
Welche Datenart liegt bei den drei erwähnten Daten vor?

Wir haben an den Beispielen und Aufgaben der letzten Seiten gesehen: Die quantitative und qualitative Erfassung der verschiedenen Zufallsgrößen erfolgt auf unterschiedliche Weise:

In den Naturwissenschaften, in der Technik und *in vielen Praxisbereichen* werden weitläufig bekannte Meßinstrumente verwendet, denen Intervallskalen zugrunde liegen und die eigentliches Messen zulassen.

■ **B 14**

Zufallsgröße (Merkmal)	Meßinstrument
Länge Zeit Masse	Metermaß, Feinmeßlehrschraube Uhr, Stoppuhr Waage

● **20*** Geben Sie die Meßinstrumente für die Merkmale „Temperatur", „Luftdruck" und „Stromstärke" an!

In anderen Bereichen – z. B. in der Ökonomie, Soziologie, Pädagogik, Teilgebieten der Psychologie, der Medizin, Landwirtschaft und des Sports – fehlen derartige Meßinstrumente. Hier gelingt es durch Skalierungsverfahren, geeignete Methoden zur Erfassung der Zufallsgrößen bereitzustellen.

■ **B 15**

Zufallsgröße (Merkmal)	Erfassungsinstrument
Einstellung zum Beruf Konzentrationsfähigkeit Leistung in Mathematik	Erfassungsfragebogen Konzentrationstest Leistungskontrolle mit Punktbewertung

Nachdem wir eine Reihe von neuen Begriffen kennengelernt und die Zusammenhänge zwischen ihnen aufgedeckt haben, ist es uns jetzt möglich, die Tabelle von Seite 19 zu vervollständigen.

	B 9	B 10	Aufg. 17	Aufg. 18
...	...	...	...	...
Merkmalsausprägungen	(35 s; 60 s)	0 bis 100 Fahrzeuge	(2,7 mm; 4,9 mm)	sehr groß, groß, mittelgroß, klein
Art der Merkmalsausprägungen	quantitativ	quantitativ	quantitativ	quantitativ
Meß- oder Erfassungsinstrument	Stoppuhr	Zählpersonen	Tiefenmeßschraube	Schätzskala
Skala	Intervallskala	Intervallskala	Intervallskala	Rangskala
Datenart	Meßwerte	Meßwerte	Meßwerte	Rangdaten

● **21*** Welche Skala liegt der Ermittlung der Daten bei den Merkmalen „Sprungweite", „Lieblingsfach in der Schule", „Blutdruck" und „Geschlecht" jeweils zugrunde?

● **22*** Bei der Vermarktung von 60 Schweinen, die aus industrieller Aufzucht stammen, wurden folgende Massen (gemessen in kg) ermittelt:

122	113	141	101	156	119	121	127	139	119
123	139	138	128	144	146	149	136	146	144
124	116	152	131	149	152	131	138	134	143
143	132	133	135	139	131	127	141	122	108
134	131	155	138	175	155	139	114	123	103
136	138	114	127	155	158	135	136	137	116

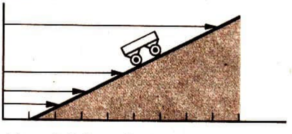
Geben Sie hierzu an: Zufallsgröße (Merkmal), Einheit, Merkmalsausprägungen, Art der Merkmalsausprägungen, Meßinstrument, Skala, Datenart!

● **23*** Geben Sie zur Zufallsgröße „Familienstand" an: Einheit (sofern möglich), Merkmalsausprägungen, Art der Merkmalsausprägungen, Erfassungsinstrument, Skala und Datenart!

Auf eine wichtige Eigenschaft von Zufallsgrößen (Merkmalen) sind wir bisher nicht eingegangen. Jede Zufallsgröße läßt sich einer der beiden Klassen „stetig" oder „diskret" zuordnen.

► **E 3** Eine Zufallsgröße (ein Merkmal) ist **stetig**, wenn sie (es) *jeden beliebigen Wert* eines bestimmten Intervalls der Zahlengeraden annehmen kann. Eine Zufallsgröße (ein Merkmal) ist dagegen **diskret**, wenn sie (es) im betrachteten Intervall *nur endlich* (oder abzählbar) *viele Zahlenwerte* annehmen kann oder wenn sie (es) *nur in n Kategorien* ($n \geqq 2$) angebbar ist.

Die stetige Zufallsgröße würde – um einen Vergleich aus der Physik heranzuziehen – einer *geneigten Ebene* entsprechen, die diskrete Zufallsgröße einer Treppe (↗ Bild 7).



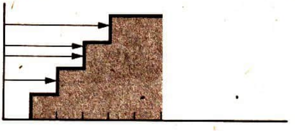


Bild 7

■ **B 16 a)** „Arbeitszeit je Steckermontage" ist eine *stetige* Zufallsgröße, weil – unabhängig von der Präzision des verwendeten Meßinstruments – im Intervall (35 s; 60 s) prinzipiell jeder reelle Zahlenwert angenommen werden kann. Es sind also auch die Meßwerte 41,52 s oder 56,923 s möglich.
Die Zeit ist immer ein stetiges Merkmal – wie fast alle physikalischen Größen.

b) „Anzahl der Fahrzeuge, die je Ampelphase eine Kreuzung befahren" ist eine *diskrete* Zufallsgröße, weil prinzipiell nur natürliche Zahlen für den Zählvorgang in Frage kommen. Die Beobachtungswerte „$17\frac{1}{2}$ Fahrzeuge je Ampelphase" oder „11,72 Fahrzeuge je Ampelphase" sind unmöglich.

c) Die Zufallsgröße „Familienstand" ist *diskret*, weil sie nur in den $n = 4$ bekannten Kategorien angegeben werden kann. Zwischenwerte wie „halb ledig, halb verheiratet" sind für statistische Belange sinnleer.

d) Die Zufallsgröße „Körpergröße eines Menschen" ist *stetig*, denn in

einem gewissen Intervall – sagen wir zwischen 40 cm und 250 cm – kann jeder reelle Zahlenwert Realisation der Zufallsgröße sein.
An dieser Aussage ändert sich nichts, wenn für bestimmte Zwecke – z. B. für die Eintragung im Personalausweis – eine Zusammenfassung aller möglichen Meßwerte in nur vier Merkmalsausprägungen erfolgt.

Beachten Sie bitte: Die Eigenschaft „stetig“ oder „diskret“ *kommt der Zufallsgröße* (dem Merkmal) *selbst zu* und hängt nicht davon ab, ob wir die Zufallsgröße mit einem feinen Maßstab messen oder mit einem groben nur schätzen!

● **24*** Ordnen Sie folgende Zufallsgrößen den beiden Klassen stetig/diskret zu! Nutentiefe, Mitgliedschaft im DTSB, Masse, Geschwindigkeit, Leistungsvermögen in Mathematik, Hektarertrag, Schiffsbestand der Handelsflotte der DDR, Niederschlagsmenge

Die Unterscheidung in stetige und diskrete Zufallsgrößen ist bedeutsam aus mehreren Gründen:
- Die wissenschaftliche Behandlung der mathematischen Statistik erfordert die Trennung in diese beiden Klassen von Zufallsgrößen, da die Wahrscheinlichkeitsverteilung der stetigen Zufallsgröße eine ganz andere ist als die der diskreten.
- Graphische Darstellungen der Häufigkeitsverteilung intervallskalierter Zufallsgrößen haben unterschiedliches Aussehen, je nachdem, ob die betrachtete Zufallsgröße stetig oder diskret ist.
- Nicht wenige statistische Verfahren verlangen als Voraussetzung ihrer Anwendbarkeit das Vorliegen stetiger Zufallsgrößen.

Zusammenfassung des Abschnitts 1.3.

Aus einer nichtleeren Menge gleichartiger Objekte, der **Grundgesamtheit**, werden mehrere Elemente entnommen, an denen eine **Zufallsgröße** (ein Merkmal) untersucht wird.
Die Zufallsgröße ist eine Funktion, die ihre Werte (Realisationen) in Abhängigkeit vom Zufall annimmt. In der Praxis spricht man meist von Merkmalen und deren **Merkmalswerten.** Die überhaupt möglichen, voneinander verschiedenen Werte sind die **Merkmalsausprägungen.** Sie können – je nach dem Wesen des Merkmals – **quantitativer Art** (durch Zahlen darstellbar) oder **qualitativer Art** sein.
Die durch die Untersuchung anfallenden **Daten** sind je nach möglicher Skala, die aufgrund des Merkmals oder der Meß- bzw. Erfassungsvorschrift angewandt werden kann: **Meßwerte (Intervallskala), Rangdaten (Rangskala)** und **Kategorien (Nominalskala).**
Zufallsgrößen (Merkmale) sind entweder **stetig** (*alle* Zahlenwerte in einem bestimmten Intervall können angenommen werden) oder **diskret** (als Merkmalsausprägungen treten nur *gewisse* Zahlen oder Kategorien auf).

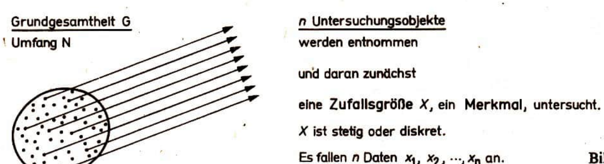


Bild 8

Daten

gewonnen mit

Intervallskala	**Ordinalskala**	**Nominalskala**
↓	↓	↓
Meßwerte	**Rangdaten**	**Kategorien**
Zahlen, die mit einem *Maßstab* (Meßinstrument) gewonnen wurden	*Zahlen* oder *Eigenschaften*, die der Größe oder Wertigkeit nach in eine *bestimmte Ordnung* oder *Reihenfolge* gebracht werden können	*Eigenschaften*, die nicht in eine bestimmte Ordnung gebracht, sondern in ihrer *Reihenfolge vertauscht* werden können
Zwischen den Teilstrichen des Maßstabs besteht wertmäßig der *gleiche Abstand.*	Zwischen den Zahlen oder Eigenschaften besteht wertmäßig *kein gleicher Abstand.*	Abstand zwischen den Eigenschaften wertmäßig *nicht existent*
$\underbrace{\text{Merkmalsausprägungen stets } \textbf{quantitativ}}$ (Intervallskala und Ordinalskala)		Merkmalsausprägungen stets **qualitativ**

2 Datenerfassung, Datenaufbereitung, tabellarische und graphische Darstellung von Verteilungen

2.1. Einige prinzipielle Bemerkungen zur Durchführung von Untersuchungen und zur Arbeit mit statistischem Material

Die Durchführung von Untersuchungen und der Umgang mit statistischem Material, die Erfassung, Bearbeitung und Auswertung der Daten sind verantwortungsvolle Tätigkeiten, die ein hohes Maß an Wissen und Können, aber auch gute Charaktereigenschaften erfordern. Umsicht und Sorgfalt bei der Vorbereitung, Exaktheit und Gründlichkeit bei der Datenerfassung, Ehrlichkeit und Gewissenhaftigkeit im Bearbeiten der Daten, Sauberkeit und Übersichtlichkeit in der Darstellung der Ergebnisse sowie kritisches Herangehen bei der Einschätzung aller Teilschritte sind einige solcher Eigenschaften. *Halten wir uns stets vor Augen, daß das zu erfassende statistische Material wertvolle Grundlage für eine zu treffende Entscheidung sein kann!* Eine solche Untersuchung ist deshalb oft auch nicht die Sache eines einzelnen, sondern bedarf der kollektiven Zusammenarbeit in einer Gruppe. Dabei muß sich jeder auf den anderen verlassen können.
Um sich von vornherein einer richtigen **Untersuchungsstrategie** zu bedienen, ist es nützlich, sich den Gesamtzusammenhang, der bei der Bewältigung einer solchen Aufgabe zu beachten ist, vor Augen zu halten.
Die folgende Übersicht zeigt die Teilschritte, die bei jeder Untersuchung zu gehen sind, in Form von Fragen. Das erleichtert uns das Vorgehen und die Einordnung der gerade zu lösenden Teilaufgabe in das große Ganze. Neben den Kästchen sind die Teilergebnisse angeführt, die sich bei Ausführung des jeweiligen Untersuchungsschrittes ergeben. Diese Teilergebnisse sind *schriftlich zu fixieren*, von der Problemstellung bis zu den nach der Untersuchung noch offenen Fragen.
Es versteht sich von selbst, daß in der Übersicht Begriffe auftauchen, die wir erst in den weiteren Abschnitten gründlich kennenlernen werden. Im Gesamtzusammenhang sind sie durchaus auch ohne nähere Erklärung zu verstehen. Und das nachfolgende Beispiel wird zum Verständnis noch beitragen.

Zur Untersuchungsstrategie

1.	Was untersuchen wir? Welches Problem wollen wir lösen?	Ziel der Untersuchung und Problemstellung
	↓	
2.	Wurde das Ziel mit den verantwortlichen Stellen oder gesellschaftlichen Kräften abgestimmt?	Abstimmung
	↓	

3.	Für welche Grundgesamtheit wollen wir Aussagen gewinnen?	Grundgesamtheit
4.	Welche Zufallsgröße(n) gilt es zu untersuchen?	Zufallsgröße(n)
5.	Wie wählen wir die zu untersuchenden Elemente (Objekte) aus?	Untersuchungsobjekte und Stichprobe
6.	Welche Meß- oder Erfassungsinstrumente verwenden wir für die Zufallsgröße(n)?	Geeignetes Instrumentarium
7.	Wie erfassen wir die Daten? Was gilt es dabei zu beachten?	Datenerfassung, Besonderheiten
8.	Wodurch wird die Objektivität der Datenerfassung gesichert?	Sicherung der Objektivität
9.	Wie bereiten wir die Daten auf, und wie werten wir sie aus?	Datenaufbereitung und -auswertung
10.	Wie lauten die Ergebnisse unserer Untersuchung, und wie stellen wir sie dar?	Ergebnisse und Ergebnisdarstellung
11.	Welche Erkenntnisse können wir daraus ableiten a) für die Praxis, b) für die Theorie?	Problemlösung und Schlußfolgerungen für Theorie und Praxis
12.	Welche Teilprobleme konnten von uns nicht gelöst werden?	Offene Fragestellungen

■ **B 17** Wir wollen uns die zwölf Einzelschritte anhand des Beispiels 10 (↗ S. 17) vor Augen führen.

Zu 1. Wir untersuchen: Verkehrsbelegung einer bestimmten ampelgeregelten Kreuzung in den Abendstunden.
Problem: Können die Ampeln abends abgeschaltet werden?

Zu 2. Abstimmung erfolgte mit Volkspolizei-Kreisamt, Abteilung Verkehrspolizei, und dem Direktor der Schule.

Zu 3. Grundgesamtheit: Alle Ampelphasen an dieser Kreuzung.

Zu 4. Zufallsgröße: *Z* „Anzahl aller Fahrzeuge, die je Ampelphase die Kreuzung in einem bestimmten Zeitabschnitt ($\frac{1}{2}$ Stunde) queren".

Zu 5. Untersuchungsobjekte: Ampelphasen an der Kreuzung.
Auswahl der Stichprobe: Wir nehmen an drei Werktagen von 18.30 bis 19.00 Uhr die Ampelphasen (jeweils 10) in die Stichprobe auf.

Zu 6. Erfassungsinstrumente: 3 Zählpersonen.

Zu 7. Datenerfassung: Erfolgt mit Strichlisten. Die Zahl der Fahrzeuge je Ampelphase wird registriert. Besonderheiten: Die drei Zählpersonen erfassen *alle* Fahrzeuge. Es wird nicht nach Art, Typ oder Größe der Fahrzeuge unterschieden. Die Ampelphase beginnt jeweils mit Grün für Nord-Süd-Richtung.
Die Zählpersonen erfassen die Daten an drei aufeinanderfolgenden Werktagen, um mögliche Unterschiede von Tag zu Tag ausschalten und gegebenenfalls auch erkennen zu können.

Zu 8. Sicherung der Objektivität: Die Zählpersonen beobachten und registrieren – an verschiedenen Stellen der Kreuzung stehend – unabhängig voneinander die passierenden Fahrzeuge. Nur wenn – im nachfolgenden Vergleich – mindestens zwei Beobachtungswerte je Ampelphase übereinstimmen, werden sie weiterverwendet.

Zu 9. Datenaufbereitung und -auswertung: Über alle drei Werktage hinweg wird eine Häufigkeitsverteilung erarbeitet sowie tabellarisch und graphisch dargestellt.
Zur Verdichtung der Daten werden Klassen gebildet, für die so entstehende Verteilung erfolgt ebenfalls die tabellarische und graphische Darstellung. Die Berechnung statistischer Maßzahlen schließt sich an. Für einen der Werktage nehmen wir eine gesonderte Auswertung vor.

Zu 10. Ergebnisse und Ergebnisdarstellung: Wir fassen die Ergebnisse in einem Bericht zusammen und bereiten eine Schautafel für die Sichtagitation vor.

Zu 11. (Mögliche) Problemlösung und Schlußfolgerungen: Die Verkehrsbelegung der Kreuzung in der Zeit von 18.30 bis 19.00 Uhr ist so gering, daß das Abschalten der Ampeln zu vertreten ist.
Alle Untersuchungsmaterialien (von der Problemstellung bis zur Problemlösung) werden der Volkspolizei zur Verfügung gestellt. Diese prüft, ob die Untersuchung für die zu treffende Entscheidung ausreicht.

Zu 12. Offene Fragen:
- War es richtig, nur die *Zahl* der Fahrzeuge und nicht auch deren *Art* zu registrieren? Hätten nicht PKWs, Motorräder, Fahrräder, Straßenbahnen und sonstige Fahrzeuge getrennt aufgeführt werden müssen?
- Wegen der Ausgedehntheit der Kreuzung waren drei Zählpersonen für den Gesamtverkehr je Ampelphase zu wenig. Wäre es nicht besser, die einzelnen Verkehrsströme getrennt erfassen zu lassen?
- Ist es vertretbar, daß die die Kreuzung passierenden Fußgänger nicht mit in die Untersuchung einbezogen wurden?

Nach der Untersuchung sieht man manches klarer und kritischer als *vor* der Untersuchung. Dann stellt man – meist zu spät – fest, daß man dieses oder jenes noch hätte anders anlegen oder ausführen können. Deshalb anhand kleiner Stichproben **Voruntersuchungen** durchführen, aber auch während der **Hauptuntersuchung** daran denken:

- Lieber ein paar Angaben mehr notieren als eine zu wenig!
- Auch Zeitpunkt und Ort der Untersuchung sowie Namen der Untersuchenden aufschreiben.
- Vor allem aber Besonderheiten und Mängel bei der Datenerfassung nicht verschweigen, sondern offen darlegen!

● **25** Überdenken Sie die 12 Einzelschritte, die in der Übersicht auf Seite 25f. aufgezeigt wurden, für eine mögliche Untersuchung aufgrund der Angaben aus Auftrag 17 (Nutentiefe von 40 Wellen, ↗ S. 18)!

2.2. Datenaufbereitung

2.2.1. Von der Datenerfassung zur Datenaufbereitung

Die Datenerfassung ist eine wichtige – und wie wir gesehen haben – mit größter Sorgfalt auszuführende Etappe bei der Verarbeitung statistischen Materials. Durch sie werden die zur Auswertung benötigten Informationen bereitgestellt.

► **E 4** Die **Datenerfassung** dient der Gewinnung von Informationen in Form von Zahlen (in erster Linie), aber auch von Buchstaben und Worten zur späteren Auswertung.

Wie die Datenerfassung erfolgt, konnten wir an einigen Beispielen des Abschnitts 1.3. bereits kennenlernen. Sie kann von Menschen vorgenommen werden (*manuell*), aber auch durch geeignete Instrumente selbsttätig erfolgen (*automatisch*).

■ **B 18** Manuelle Datenerfassung: Tägliches Ablesen und Notieren der Wassertemperatur der Ostsee vor Kühlungsborn; Erfassen von Beobachtungswerten in einer Strichliste.
Automatische Datenerfassung: Ständiges Aufzeichnen von Erschütterungen der Erdkruste am Geophysikalischen Observatorium Collm.

Im Ergebnis der Datenerfassung entsteht ein **Datenträger**, der die gewonnenen Informationen in irgendeiner Form speichert. Man unterscheidet einfache, *nicht-maschinenlesbare* Datenträger (handschriftlicher Beleg, Urliste, Versuchsprotokoll usw.) von *maschinenlesbaren* Datenträgern (Lochkarte, Lochband, Magnetband usw.). Diese Datenträger können in elektronischen Datenverarbeitungsanlagen (EDVA) direkt verwendet werden und bilden die Grundlage für eine schnelle und effektive Auswertung der Daten. Das Bild 9 zeigt eine moderne EDVA der ESER-Reihe.

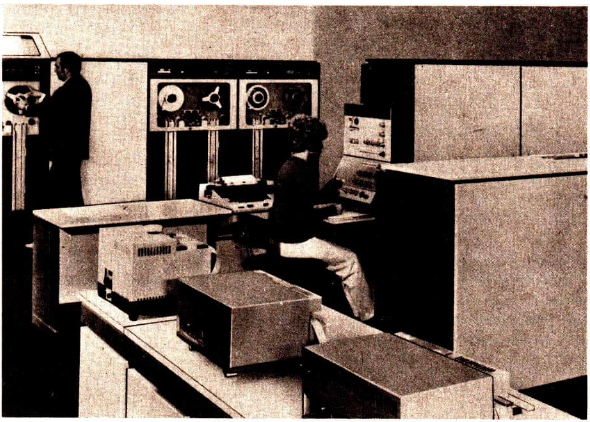


Bild 9

Im Zusammenhang mit der Volks-, Berufs-, Wohnraum- und Gebäudezählung in der DDR am 31. 12. 1981 wurden die Angaben der Bürger (Daten) durch Zählinstrukteure als Strichmarkierungen auf Markierungsbelegen, einer besonders vorteilhaften Art maschinenlesbarer Datenträger, festgehalten. Diese Markierungen gestatten mit Hilfe von modernen elektronischen Markierungslesegeräten ein unmittelbares Einlesen der Informationen in die EDVA. Der Einsatz von Markierungsbelegen bei der Volkszählung machte die Herstellung von rund 27 Millionen Lochkarten überflüssig. Dadurch konnten also Arbeit, Kosten und Zeit gespart werden.

Nun folgt die Datenaufbereitung als zweite Etappe einer empirischen Untersuchung.

▶ **E 5** Unter **Datenaufbereitung** verstehen wir das Ordnen und Verdichten der Daten in Form von Tabellen.

Bei der Datenerfassung waren das *Messen* und *Beobachten* die entscheidenden Tätigkeiten. Jetzt, für die Aufbereitung der Daten, steht das *Zählen* im Vordergrund.
Das kann wiederum auf unterschiedliche Weise geschehen: *manuell* oder *maschinell.*
Bei Untersuchungen kleineren Umfangs kann die Datenaufbereitung manuell ohne besondere Hilfsmittel erfolgen (wir kommen im Abschnitt 2.2.2. darauf zurück), bei Untersuchungen mittleren Umfangs bedient man sich des Kerbkartenverfahrens, das ebenfalls zur manuellen Aufbereitung gerechnet wird.
Bei größeren Untersuchungen dagegen kommt man ohne maschinelle Aufbereitung nicht aus. Dann steht entweder das Lochkartenverfahren oder – vorteilhafteste Form – die Aufbereitung mittels einer EDVA zur Verfügung. Beim Lochkartenverfahren übernehmen Sortier- und/oder Tabelliermaschinen wesentliche Teile der Aufbereitungsarbeiten. Es wird nach Merkmalsausprägungen sortiert und dann tabelliert. Bei großer Anzahl der zu untersuchenden Objekte und Merkmale – wie es z. B. bei der Volkszählung der Fall ist – kann allein die Aufbereitung mit Hilfe einer elektronischen Datenverarbeitungsanlage zu schnellen und verläßlichen Ergebnissen führen. Hierbei werden die Daten mittels einer Leseeinheit (für Lochkarten, Lochbänder, Markierungsbelege o. ä.) auf einen externen Speicher der Anlage (Magnetband, Magnettrommel o. ä.) gebracht, von dem aus sie jederzeit abgerufen und äußerst schnell verarbeitet werden können.
Bei Verwendung von maschinenlesbaren Datenträgern hängen **Erfassung, Aufbereitung** und **Auswertung** des statistischen Materials eng miteinander zusammen. Darin liegt ein weiterer überaus bedeutsamer Vorteil des Einsatzes von EDVA, deren Leistungsfähigkeit durch den Einzug der Mikroelektronik noch gesteigert wird.

2.2.2. Häufigkeitsverteilung und deren Darstellung

Die in einer Untersuchung anfallenden und in einem Urbeleg erfaßten Originaldaten (auch: Urdaten) gestatten aufgrund ihres Ungeordnetseins keinerlei Einblicke in das Wesen der Erscheinung. Erst durch die Datenaufbereitung, das *Ordnen* und *Verdichten* der Einzeldaten, wird das Beobachtungsmaterial überschaubar. Damit gelingt es, Typisches der untersuchten Erscheinung sichtbar zu machen.
Wir wollen hier einfache Möglichkeiten der manuellen Aufbereitung kennenlernen, die jeder von uns nachvollziehen kann.

Urbeleg – Strichliste – Häufigkeitsverteilung

Im Urbeleg – das kann die Urliste, ein Fragebogen oder ein Versuchsprotokoll sein – haben wir die Daten des untersuchten Merkmals so erfaßt, wie sie anfallen.

Jetzt beginnen wir mit der Ordnung der Daten. Ziel soll sein, alle Werte zusammenzufassen, die die *gleiche Merkmalsausprägung* tragen. Dazu ist es notwendig, die in der Untersuchung vorkommenden Merkmalsausprägungen untereinander anzuordnen. Sind die Merkmalsausprägungen *quantitativer Art*, so beginnt man mit der Angabe der *kleinsten* wirklich auftretenden Merkmalsausprägung (x_{min}) und endet mit der *größten* wirklich auftretenden (x_{max}), läßt aber dazwischen keine der möglichen (nach Maßstab oder gewählter Genauigkeit) Ausprägungen aus. Sind die Merkmalsausprägungen *qualitativer Art*, so sind alle vorgesehenen Kategorien anzuführen, deren Anordnung ist jedoch willkürlich.

Merkmalsausprägung x_k $(k = 1, 2, \dots, m)$	Striche
4	II
5	
6	III
7	~~IIII~~ II
8	~~IIII~~ ~~IIII~~ I
...	...

Wir kennzeichnen in einer 2. Spalte hinter den Merkmalsausprägungen *jeden* im Urbeleg vorhandenen Wert durch einen senkrechten Strich, nur den jeweils 5. Strich legen wir waagerecht durch die vier bereits vorhandenen. Es entsteht die **Strichliste.**

► **E 6** Unter der **Strichliste** verstehen wir die (geordnete) Folge der Merkmalsausprägungen einer Untersuchung mit Angabe der jeweils vorhandenen Anzahl der Beobachtungswerte als Striche.

Die Strichliste faßt also alle *gleichen* Meß- oder Beobachtungswerte in *einer* Zeile zusammen. Es ist günstig, sich die Daten des Urbelegs in der dort gegebenen Reihenfolge ansagen zu lassen, das geht rasch und verläuft meist fehlerfrei. In zahlreichen statistischen Untersuchungen erweist es sich als sinnvoll, die Daten unmittelbar in Strichlisten zu erfassen, denken wir nur an die Zählung der Fahrzeuge, die je Ampelphase eine Kreuzung queren.

Zunächst kommen wir aber auf unser Beispiel 9 (↗ S. 16) zurück und wiederholen die wichtigsten Angaben, ergänzt durch Zeit und Ort der Untersuchung. Es sollte nie versäumt werden, im Kopf der Listen festzuhalten, was im Zusammenhang mit der Untersuchung von Bedeutung sein kann.

■ **B 9 (2)** *Ziel:* **Vergleich der Leistung zweier Lehrlinge mit der Facharbeiternorm**[1])

Untersuchungsobjekte: Steckermontagen
Zufallsgrößen: X = Arbeitszeit des Lehrlings A je Steckermontage,
Y = Arbeitszeit des Lehrlings B je Steckermontage
Art der Merkmalsausprägungen: quantitativ
Untersuchungsort: BBS Fernmeldewerk Berlin
Untersuchungszeit: 12. 11. 81, 7.30 Uhr

Urlisten: ↗ S. 16

Wir untersuchen zunächst die Zufallsgröße X. Der Urliste ist zu entnehmen: $x_{min} = 40$ s; $x_{max} = 59$ s. Diese beiden Werte und alle dazwischenliegenden Merkmalsausprägungen (in Sekunden, da auf volle Sekunden gerundet wurde) werden in der Strichliste aufgeführt.

[1]) Mit 9(2) ist das zweite Auftreten des Sachverhaltes bezeichnet, der schon im Beispiel 9(1) auf Seite 16 untersucht wurde.

Strichliste zu *X*:

Merkmalsausprägung x_k ($k = 1, 2, \ldots, 12$) (s)	Striche	Häufigkeit h_k
40	ǀ	1
41	卌	5
42	卌ǀ	6
43	卌卌ǀ	11
44	卌ǀǀǀǀ	9
45	卌ǀǀǀǀ	9
46	ǀǀǀ	3
47	ǀǀǀ	3
48		0
49	ǀ	1
50	ǀ	1
51	ǀ	1

Jeder Strich repräsentiert ein Untersuchungsobjekt, d. h. hier eine Steckermontage, für die die in der 1. Spalte angegebene Zeit benötigt wurde. Die Anzahl der Striche in einer Zeile gibt Auskunft darüber, wie *häufig* eine bestimmte Zeit (in s) für die Steckermontage auftrat.

Infolgedessen ist es von der Strichliste zur **Häufigkeitstabelle** nur ein ganz kurzer Weg: Wir brauchen lediglich die Anzahl der Striche – deren Häufigkeit h_k – in einer weiteren Spalte festzuhalten. Das ist in der obenstehenden Tabelle bereits geschehen. Damit ist eine Häufigkeitstabelle entstanden, die tabellarische Darstellung einer Häufigkeitsverteilung.

► **D 4 Unter *Häufigkeit* einer Merkmalsausprägung verstehen wir die Anzahl *h* der Meß- oder Beobachtungswerte, die diese Merkmalsausprägung tragen. Dabei handelt es sich stets – wenn nichts anderes vermerkt wird – um *absolute Häufigkeiten.***

► **D 5 Die eindeutige Zuordnung von Häufigkeiten zu allen bei der betreffenden Untersuchung in Betracht kommenden Merkmalsausprägungen heißt *Häufigkeitsverteilung.***
Die *Häufigkeitstabelle* ist die Darstellung der Häufigkeitsverteilung in tabellarischer Form.

Die Summe aller Häufigkeiten $h_1 + h_2 + \ldots + h_m = \sum_{k=1}^{m} h_k$ [1]) führt stets auf die Gesamtzahl *n* aller Untersuchungsobjekte oder – was dasselbe ist – auf die Anzahl aller Meß- oder Beobachtungswerte. Dadurch ist eine Kontrollmöglichkeit gegeben.

$$\sum_{k=1}^{m} h_k = n \qquad (1)$$

[1]) Näheres zum Summenzeichen $\sum$ im Abschnitt 2.3.1., Seite 85

Die Häufigkeitstabelle hat gegenüber der Urliste folgende
Vorteile: Größere Übersichtlichkeit, geringerer Schreibaufwand, Besonderheiten der Häufigkeitsverteilung treten zutage.
Sie hat aber auch einen – allerdings unbedeutenden –
Nachteil: Aus der Häufigkeitstabelle ist nicht mehr erkennbar, bei welchem speziellen Objekt das Merkmal die betreffende Ausprägung besitzt.

■ **B 19** Beispiel für eine Häufigkeitstabelle bei einem Merkmal mit qualitativen Ausprägungen:
Häufigkeitsverteilung zum Merkmal „Geschlecht" für den Bezirk Karl-Marx-Stadt; Zeit 1978

Merkmalsausprägung x_k ($k = 1,2$)	Häufigkeit h_k
männlich	894463
weiblich	1054741
$\sum h_k$	$1949204 = n$

Dieses Beispiel zeigt auch, daß man bei so großen Häufigkeiten mit der manuellen Datenaufbereitung nicht weit käme.

● **26*** Stellen Sie die Häufigkeitstabelle für die bei der Untersuchung der Zufallsgröße Y aus Beispiel 9 (↗ S. 16; 30) gemessenen Werte auf!

● **27** Geben Sie die Häufigkeitsverteilung zu den im Auftrag 18 (↗ S. 18) erfaßten Daten in tabellarischer Form an!

Vergleiche zwischen absoluten Häufigkeiten bei unterschiedlichem Umfang n der Untersuchungsobjekte (bzw. unterschiedlichem Umfang N der Grundgesamtheit) erweisen sich als schwierig.

■ **B 20** Erweitern wir die Häufigkeitstabelle aus Beispiel 19, indem wir den Häufigkeiten des Bezirkes Karl-Marx-Stadt die entsprechenden des Bezirkes Suhl zur Seite stellen, so ergibt sich folgendes Bild:

Merkmalsausprägung x_k ($k = 1, 2$)	Häufigkeit h_k	
	Karl-Marx-Stadt	Suhl
männlich	894463	258082
weiblich	1054741	289265
$\sum h_k$	$1949204 = n_1$	$547347 = n_2$

Will man nun wissen, ob im bevölkerungsstärksten Bezirk der DDR, Karl-Marx-Stadt, verhältnismäßig mehr Frauen und Mädchen leben als im bevölkerungsschwächsten, Bezirk Suhl, dann ist das aus den absoluten Häufigkeiten nicht unmittelbar ablesbar.
In einem solchen Falle bedient man sich vorteilhafter der **relativen Häufigkeiten** (↗ Beispiel 5, S. 12f.). Diese erhält man sehr einfach dadurch, daß man die absoluten Häufigkeiten durch den jeweiligen Umfang n der Menge der Untersuchungsobjekte (bzw. Umfang N der Grundgesamtheit) dividiert.

▶ **D 6 Unter der *relativen Häufigkeit* einer Merkmalsausprägung verstehen wir das Verhältnis der vorliegenden absoluten Häufigkeit h zum Umfang n (bzw. N).**

Relative Häufigkeit $= \frac{h}{n} \left(\text{bzw. } \frac{h}{N}\right)$ **(2)**

Relative Häufigkeiten werden oft in Prozent angegeben. Man spricht dann auch von **prozentualer Häufigkeit.**
Die Häufigkeitstabelle im Beispiel 20 stellt sich unter Verwendung prozentualer Häufigkeiten folgendermaßen dar:

■ **B 21** Häufigkeitsverteilung zum Merkmal „Geschlecht" für den bevölkerungsstärksten und bevölkerungsschwächsten Bezirk der DDR im Vergleich. Zeit 1978

Merkmalsausprägung x_k ($k = 1, 2$)	Prozentuale Häufigkeit Karl-Marx-Stadt	Suhl
männlich	45,89%	47,15%
weiblich	54,11%	52,85%
$\sum \frac{h_k}{n}$ %	100,00%	100,00%

Jetzt ist sehr leicht zu erkennen, daß der Anteil der Frauen und Mädchen an der Gesamtbevölkerung im Bezirk Karl-Marx-Stadt größer ist als im Bezirk Suhl.

Die Summe der relativen Häufigkeiten muß *immer* 1, die der prozentualen Häufigkeiten 100% betragen. (Dabei sind allerdings geringfügige Ungenauigkeiten durch Rundungen möglich.)

$$\sum_{k=1}^{m} \frac{h_k}{n} = 1 \qquad \left(\text{bzw. } \sum_{k=1}^{m} \frac{h_k}{N} = 1\right) \tag{3}$$

Es bedeuten:
h_k Absolute Häufigkeit der k-ten Merkmalsausprägung;
m Anzahl der auftretenden Merkmalsausprägungen;
n Umfang der Menge der Untersuchungsobjekte;
N Umfang der Grundgesamtheit.

Mit (3) ist eine weitere Kontrollmöglichkeit gegeben.

● **28** Ergänzen Sie die Häufigkeitstabelle im Beispiel 9 (2) (↗ S. 31) um zwei Spalten, die der relativen und die der prozentualen Häufigkeiten (unter Weglassen der Strichspalte, die nur für die 1. Phase der Datenaufbereitung von Bedeutung ist)!

Lösungshinweis: Für die erste Zeile würde sich ergeben:

Merkmalsausprägung x_k ($k = 1, 2, \ldots, 12$) (s)	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$	Prozentuale Häufigkeit
40	1	0,02	2%

● **29*** Erweitern Sie die im Auftrag 26 gewonnene Häufigkeitstabelle für *Y* aus dem Beispiel 9 um die relativen und die prozentualen Häufigkeiten! Achten Sie auf die exakte Benennung der Kopfzeile!

● **30*** Besteht in den Tabellen der Beispiele 20 und 21 die Möglichkeit des Austausches der beiden Merkmalsausprägungen „männlich", „weiblich" mitsamt den ihnen entsprechenden Häufigkeiten?

Graphische Darstellung von Häufigkeitsverteilungen

Mit dem Ordnen der Daten und deren Darstellung in Form einer Häufigkeitstabelle hat das statistische Material schon an Übersichtlichkeit gewonnen. Die Besonderheiten der Verteilung treten jedoch erst durch eine graphische Darstellung deutlich zutage.

Es gibt drei Möglichkeiten der **Darstellung von Daten:**
- Tabelle (tabellarische Darstellung),
- Diagramm (graphische Darstellung),
- statistische Maßzahlen (Mittelwerte, Streuungsmaße).

Die **Tabelle** hat den Vorteil der genauen Vermittlung der Einzeldaten, ist aber wegen der Fülle des Zahlenmaterials, das sie bei größeren statistischen Erhebungen enthält, oft nicht im ganzen zu überblicken und daher weniger instruktiv als das Diagramm. Die **graphische Darstellung** gestattet aufgrund ihres hohen Grades an Anschaulichkeit einen schnellen Einblick in die Struktur des vorliegenden Sachverhalts, in den Zusammenhang von Erscheinungen oder in den Verlauf eines Entwicklungsprozesses. Graphische Darstellungen lernten wir im Schulunterricht bereits in vielfältiger Form kennen.

■ **B 22** Das Bild 10 stellt eine Temperaturkurve und ein Niederschlagsdiagramm für Berlin dar. Mit diesen wird die über viele Jahre hinweg gemittelte Temperatur (in °C) bzw. die Niederschlagsmenge (in mm) in Abhängigkeit vom jeweiligen Monat dargestellt.

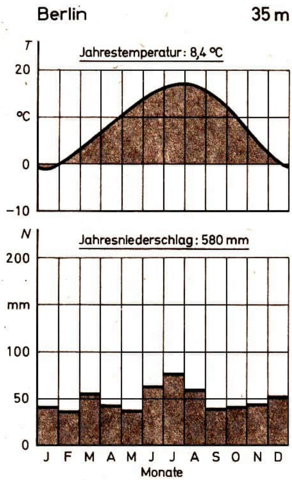


Bild 10

Wir kennen auch bereits Stabdiagramme, Balkendiagramme und Kartogramme – das sind geographische Karten, in die für bestimmte Regionen zutreffende Angaben eingetragen sind. (Zum Beispiel könnte in einer Karte die Standortverteilung der metallverarbeitenden Industrie Italiens eingetragen sein.) Ferner kennen wir einfache Linien- und Kreisdiagramme wie auch die graphische Darstellung von Wertepaaren und Funktionen im rechtwinkligen Koordinatensystem.
Graphische Darstellungen sind von großem Wert, wenn es darum geht, das Wesen einer Erscheinung, das Typische eines Sachverhalts, das Auffällige eines Zusammenhangs oder das Bemerkenswerte eines Entwicklungsprozesses schnell und verläßlich zu

erfassen. Das gilt für Wirtschaft und Industrie, Wissenschaft und Technik, Landwirtschaft und Medizin, Staat und Recht, für Schule und Beruf in gleicher Weise.

► **E 7** Die **graphische Darstellung** ist eine Zeichnung, in der Zahlen, Paare oder Tripel von Zahlen, Meß- oder Beobachtungswerten eindeutig durch Punkte, Strecken, Flächen oder Körper wiedergegeben werden.

Vorrangige Aufgabe der graphischen Darstellung für *statistische Zwecke* ist, Struktur, Verlauf oder Ergebnis der untersuchten Erscheinung anschaulich widerszuspiegeln. Wir werden im weiteren die wichtigsten Darstellungsarten behandeln, die wir für die *Auswertung statistischen Materials* benötigen. Wir haben zu unterscheiden zwischen Punkt-, Linien-, Flächen- und Körperdiagrammen sowie Kartogrammen.

Die **graphische Darstellung von Häufigkeitsverteilungen für eine Zufallsgröße** wird im Vordergrund der Betrachtungen stehen. Wir verwenden ein rechtwinkliges Koordinatensystem. Auf der Abszissenachse werden die Merkmalsausprägungen (Meßwerte, Rangdaten, Kategorien) aufgetragen. Sie heißt deshalb Merkmalsachse. Die zugehörigen Häufigkeiten stellt man auf der Ordinatenachse dar.
Dabei können sowohl absolute als auch relative oder prozentuale Häufigkeiten verwendet werden. Die Angabe von relativen oder prozentualen Häufigkeiten bietet den Vorteil eines Vergleichs mehrerer Häufigkeitsverteilungen zur gleichen Zufallsgröße bei ungleichem Umfang n der Menge der untersuchten Objekte (↗ Bild 11).

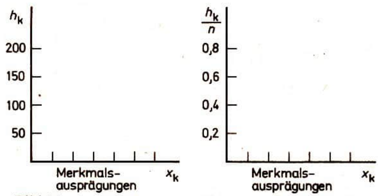


Bild 11

Oft wird es nützlich sein, zur Darstellung Millimeterpapier oder wenigstens Kästchenpapier zu verwenden, das erleichtert die Einteilung der Achsen und das Einzeichnen der Häufigkeiten.

Um eine ansprechende und vor allem korrekte Darstellung zu erhalten, gilt es zu beachten:

- Das Diagramm soll in etwa eine quadratische Form erhalten. Das läßt sich meist erreichen, beispielsweise durch Festlegen einer unterschiedlichen Länge der Einheiten auf den Achsen.
- Bei der Anordnung der Merkmalswerte auf der Abszissenachse sollten – sofern es möglich ist – *bessere* Leistungen stets nach rechts aufsteigend angeordnet werden.
- Die Achsenbezeichnungen dürfen nicht vergessen werden.
- Bei sehr großen absoluten Häufigkeiten wähle man die Einheit auf der Ordinatenachse entsprechend klein. Liegen die Häufigkeiten (bezüglich ihrer Zahlenwerte) dicht beieinander, so besteht die Möglichkeit der „Verkürzung" der Ordinatenachse und damit der Streifen oder Strecken.

Die Art der graphischen Darstellung hängt ab von der Klasse der Zufallsgröße (stetig/diskret) *und der in der Untersuchung verwendeten Datenart* (Meßwerte/Rangdaten/Kategorien).

Wir besprechen zunächst drei mögliche Fälle.

1. Fall: Es handelt sich um ein *stetiges Merkmal*, und es wurden *Meßwerte* erhoben. Zur graphischen Darstellung dieses Falles eignen sich Linienzug und Histogramm. Der **Linienzug** (auch Polygonzug) gehört zu den **Liniendiagrammen.**

> ▶ **E 8** Der **Linienzug** besteht aus aneinandergefügten Strecken, die die den geordneten Paaren $(x_k; h_k)$ der Häufigkeitsverteilung entsprechenden Punkte miteinander verbinden.

Hinweis: An den Enden der Verteilung wird der Linienzug bis zur nächsten Merkmalsausprägung auf der Achse heruntergezogen, sofern diese Ausprägungen noch einen Sinn (gemäß der Aufgabe) haben.

■ **B 9 (3)** Graphische Darstellung der Häufigkeitsverteilung zur Untersuchung der Zufallsgröße X (Arbeitszeit des Lehrlings A je Steckermontage) (↗ Beispiel 9(2), S. 30) mittels Linienzug (↗ Bild 12)

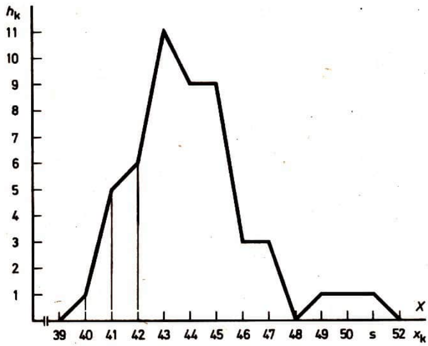


Bild 12

Anmerkungen zum Bild 12: Obwohl die Merkmalsausprägungen in der Häufigkeitstabelle (↗ S. 31) nur von 40 s bis 51 s reichen, haben wir die Merkmalswerte 39 s und 52 s (beides sind mögliche Meßwerte) auf der Abszissenachse mit zu berücksichtigen. (Nur so ist es möglich, den Linienzug an beiden Enden bis zur Merkmalsachse herunterzuziehen.)
Die kurzen Parallelstriche links vom Merkmalswert 39 sollen andeuten, daß die Merkmalsachse unterbrochen ist. Denn es gäbe eine unvorteilhafte Darstellung des Sachverhalts, wenn man sämtliche Merkmalsausprägungen (von 1 s bis 53 s) eintragen würde.

● **31** Überprüfen Sie das selbst am Beispiel „Arbeitszeit des Lehrlings A je Steckermontage“!

● **32** Stellen Sie die Häufigkeitsverteilung zur Zufallsgröße Y aus dem Beispiel 9 (↗ S. 16; 30f.) grafisch dar! Die Häufigkeitstabelle dazu finden Sie im Lösungsteil, Seite 125, Aufgabe 29.

● **33*** Was fällt Ihnen an der Form des Linienzuges, den Sie im Auftrag 32 erhalten, auf?

● **34*** Was können Sie über Unterschiede in der Leistung der Lehrlinge A und B aussagen, wenn Sie die graphische Darstellung im Bild 12 (↗ S. 36) mit der im Auftrag 32 ermittelten graphischen Darstellung vergleichen und sich daran erinnern, daß die Facharbeiternorm 45 s/Steckermontage betrug?

Das **Histogramm** – die zweite Möglichkeit der graphischen Darstellung von Häufigkeitsverteilungen *stetiger Merkmale*, die auf *Meßwerten* beruhen – gehört zu den **Flächendiagrammen.**
Das Histogramm wird zuweilen auch als Streifendiagramm bezeichnet, das kann aber zu Verwechslungen mit den eigentlichen Streifen- (oder Balken-) Diagrammen führen, die bei ordinal- oder nominalskalierten Daten angebracht sind.

▶ **E 9** Das **Histogramm** besteht aus unmittelbar nebeneinanderstehenden Rechtecken, deren auf der Merkmalsachse gelegene Seite durch die Breite eines Merkmalswertes und deren Höhe durch die zugehörige Häufigkeit bestimmt ist. Die *Fläche* des Rechtecks repräsentiert also die Häufigkeit des jeweiligen Meßwertes.

Das Zeichnen eines Histogramms geschieht auf folgende Weise:
Die Teilstriche auf der Merkmalsachse repräsentieren die Merkmalsausprägungen (Meßwerte). Wir halbieren die Abstände zwischen zwei benachbarten Teilstrichen und erhalten so die Begrenzungen der Merkmalsausprägung nach oben und unten. Im Abschnitt 2.2.3. lernen wir diese Begrenzungen als „exakte Klassengrenzen" kennen.
Da es sich hier um ein stetiges Merkmal handelt, umfaßt der Merkmalswert „43 s" das gesamte Intervall reeller Zahlen von 42,5 s bis 43,5 s (↗ Bild 13).
In diesen Begrenzungen errichten wir Senkrechte auf der Merkmalsachse, die die (im allgemeinen) längeren Seiten der zu zeichnenden Rechtecke bilden. Das dem jeweiligen x_k entsprechende h_k bestimmt die Höhe der Rechtecke.

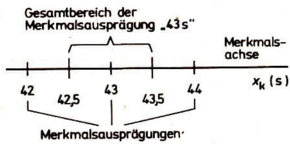


Bild 13

Beachten Sie: Die Rechtecke dürfen beim Histogramm nicht ausgemalt werden!

■ **B 9 (4)** Graphische Darstellung der Häufigkeitsverteilung von *X* (Arbeitszeit des Lehrlings A je Steckermontage) gemäß Beispiel 9 (↗ S. 16; 30) und 9(3) (↗ S. 36) mittels Histogramm (Bild 14)

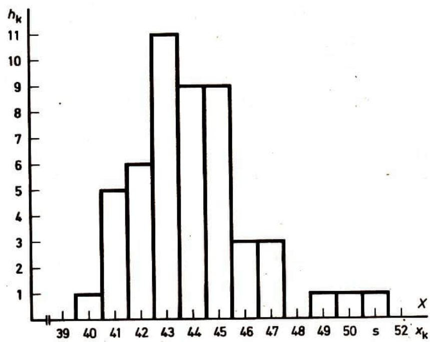


Bild 14

● **35** **a)** Vergleichen Sie die beiden Darstellungsarten Linienzug und Histogramm ein und derselben Häufigkeitsverteilung miteinander (Bilder 12 und 14)!
b)* Wie erhält man aus einem Histogramm den zugehörigen Polygonzug?

● **36** 20 Schüler beteiligen sich an einem Mathematik-Olympiade-Wettbewerb und erreichen folgende Punktzahlen (die vergebenen Punkte sind gleichwertig):

27	26	22	28	30	26	22	28	20	31
28	27	24	30	25	29	23	26	24	28

a)* Geben Sie hierzu an: Untersuchungsobjekte, Zufallsgröße (Merkmal), Art der Zufallsgröße (stetig/diskret), Einheit, Merkmalsausprägungen, Art der Merkmalsausprägungen, Skala, die den gegebenen Daten zugrunde liegt, Datenart!

b)* Stellen Sie in *einer* Tabelle dar:
Strichliste, Häufigkeitstabelle (absolut und relativ)!

c) Stellen Sie die Häufigkeitsverteilung auf Millimeterpapier graphisch dar, und zwar als Histogramm *und* als Linienzug in *einer* Zeichnung! Kennzeichnen Sie die Ordinatenachse für absolute und für relative Häufigkeiten!

d) Welche Aussage können Sie formulieren, wenn Sie die Gesamtfläche unter dem Linienzug (also zwischen Linienzug und Merkmalsachse) vergleichen mit der Gesamtfläche aller Rechtecke des Histogramms?

e) Was ist das Auffällige am Verlauf dieses Linienzuges?

2. Fall (↗ Ankündigung und 1. Fall auf Seite 35f.): Es handelt sich um ein *diskretes Merkmal*, und es wurden *Meßwerte* erfaßt.
Die graphische Darstellung solcher Merkmale erfolgt durch das **Streckendiagramm.** Es gehört zu den Liniendiagrammen.

► **E 10** Das **Streckendiagramm** besteht aus einzelnen senkrecht zur Merkmalsachse stehenden Strecken, die in den Merkmalsausprägungen errichtet werden und deren Länge durch die Häufigkeit h_k der betreffenden Ausprägung x_k bestimmt ist.

Das Zeichnen des Streckendiagramms ist einfach; es geht aus der Erklärung E 10 unmittelbar hervor. Zur besseren Heraushebung verstärkt man die Strecken ein wenig.

Beachten Sie: Es dürfen keine Streifen gezeichnet werden. Die Endpunkte der Strecken dürfen keinesfalls miteinander verbunden werden!

■ **B 23** Wir würfeln 100mal mit einem echten Spielwürfel und sind an der jeweils oben erscheinenden Augenzahl interessiert.
Die *Zufallsgröße X* heißt hier also: „Augenzahl je Wurf", sie ist *diskret*. Ihre *Merkmalsausprägungen* x_k sind: „1", „2", „3", „4", „5", „6".
Wir ermitteln dazu die Häufigkeitsverteilung und stellen sie graphisch dar. Wir beginnen mit der Strichliste und ordnen die Merkmalsausprägungen der Einfachheit halber nebeneinander an.

x_k	„1“	„2“	„3“	„4“	„5“	„6“	
Striche	𝍸𝍸𝍸𝍷𝍷	𝍸𝍸𝍷𝍷	𝍸𝍸𝍸𝍷𝍷𝍷	𝍸𝍸𝍸𝍷𝍷𝍷𝍷	𝍸𝍸𝍸𝍷𝍷𝍷	𝍸𝍸𝍸𝍷	
h_k	17	12	18	19	18	16	$\sum h_k = 100$

Es entsteht ein Streckendiagramm (↗ Bild 15).
Die metrische Einteilung der Merkmalsachse, wie wir sie von Linienzug und Histogramm her kennen, liegt auch hier vor. Der wesentliche Unterschied beim Streckendiagramm besteht aber darin, daß hier den Werten zwischen den Merkmalsausprägungen keinerlei Meßwertbedeutung zukommt. Die Augenzahl „1,7“ oder „4,5“ gibt es nicht.

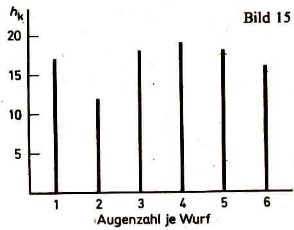


Bild 15

● **37** **a)** Wiederholen Sie die Aufgabe aus Beispiel 23, jedoch mit 150 Würfen!
b)* Welche Form würde die Häufigkeitsverteilung bei $h = 1000$ Würfen annehmen?

■ **B 10 (2)** Bild 16: Graphische Darstellung der Häufigkeitsverteilung zur Zufallsgröße Z „Anzahl der Fahrzeuge, die ...“ (↗ Beispiel 10(1), S. 17)

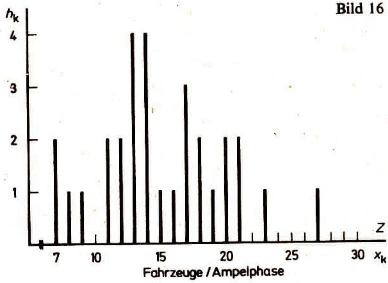


Bild 16

● **38** Was fällt Ihnen an dem Streckendiagramm im Bild 16 besonders auf? Was könnte man tun, um Gesetzmäßigkeiten der Verteilung besser hervortreten zu lassen und damit erkennbarer zu machen?

● **39*** **a)** Stellen Sie die Zufallsgröße Z für die Werte graphisch dar, die am 3. Beobachtungstage ermittelt wurden (↗ Beispiel 10(1), S. 17)!
b) Was wird an dieser Darstellung deutlich?

3. Fall: Es handelt sich um ein *stetiges oder diskretes Merkmal*, und es liegen *rang- oder nominalskalierte* Daten vor.
Für die graphische Darstellung dieser Art von Merkmalen stehen folgende Möglichkeiten offen:
Streifen- oder Balkendiagramm, Staffelbild, Kreisdiagramm.
Alle drei gehören zu den **Flächendiagrammen.**

Wir wenden uns zunächst dem **Streifendiagramm** zu, das manchmal auch als Balkendiagramm bezeichnet wird.

▶ **E 11** Das **Streifendiagramm** besteht aus *getrennt voneinander stehenden* (oder liegenden) Vollstreifen (ausgemalte oder schraffierte Rechtecke), die den rang- oder nominalskalierten Merkmalsausprägungen zugeordnet sind. Die Länge der Streifen ist durch die Häufigkeit h_k der betreffenden Ausprägung x_k bestimmt.

■ **B 24** In einer Werkstatt wurden die an einem Tage reparierten Fernsehgeräte – nach Typen unterschieden – erfaßt.
Zufallsgröße U: Güte eines Erzeugnisses – erfaßt nach Güteklassen

Häufigkeitstabelle:

Güteklasse u_k ($k = 1,2, \ldots, 5$)	Häufigkeit h_k
1	5
2	9
3	6
4	4
5	1

Graphische Darstellung (Bild 17)

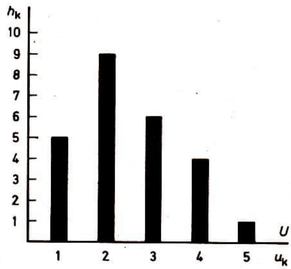


Bild 17

Beachten Sie beim Streifendiagramm besonders:
Die meist (auch im Bild 17) vorgenommene gleichmäßige Teilung der Merkmalsachse darf uns nicht zu dem Schluß verleiten, diese trage einen intervallskalierten Maßstab. Kategorien wie Rangdaten kommt keine Metrik zu, d. h., die Gleichabständigkeit der Streifen bedeutet keinesfalls Gleichabständigkeit der Merkmalsausprägungen! Das wird beim nächsten Beispiel noch deutlicher.

■ **B 25 (1)** Häufigkeitsverteilung des Merkmals „Familienstand" für die Bevölkerung der DDR (gesamt und weiblich) 1979

Häufigkeitstabelle, absolut und relativ:

Merkmalsausprägung	Absolute Häufigkeit gesamt	weiblich	Prozentuale Häufigkeit gesamt	weiblich
ledig	6135424	2941986	36,6%	33,0%
verheiratet	8386375	4200929	50,1%	47,2%
verwitwet	1518554	1293392	9,1%	14,5%
geschieden	703153	467155	4,2%	5,2%
Summe	16743506	8903462	100,0%	99,9%

Graphische Darstellung (prozentuale Häufigkeiten) für die Gesamtbevölkerung (Bild 18)

Bei Merkmalen mit nur qualitativen Ausprägungen – also beim Vorliegen von Kategorien – ist selbst die Anordnung der Merkmalsausprägungen auf der Merkmalsachse – vom Prinzip her – willkürlich wählbar.

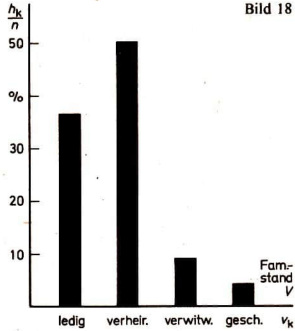


Bild 18

Aus einer solchen Darstellung mit prozentualen Häufigkeiten (Bild 18) läßt sich der verhältnismäßige Anteil der einzelnen Kategorien gut ablesen. Nicht zu entnehmen sind dem Bild *allein* irgendwelche Angaben über die zugrunde liegenden absoluten Häufigkeiten. Die im Bild 18 dargestellte Häufigkeitsverteilung könnte sich auch auf die Einwohner eines Dorfes oder auf die Bevölkerung eines ganzen Kontinents beziehen.

Beachten Sie also: In einer tabellarischen oder graphischen Darstellung mit *absoluten* Häufigkeiten sind *mehr Informationen* enthalten!

Die Darstellung mit *relativen* oder *prozentualen* Häufigkeiten ist jedoch von großem Wert, wenn es um *Vergleiche* von Grundgesamtheiten oder Stichproben mit unterschiedlichem Umfang geht. Im folgenden Beispiel wird gezeigt, wie man die Angaben im Beispiel 25 für die Gesamtbevölkerung und für die weibliche Bevölkerung in *einem* Diagramm anschaulich darstellen kann.

■ **B 25 (2)** Graphische Darstellung (prozentuale Häufigkeiten) des Familienstandes für die Gesamtbevölkerung der DDR und für den weiblichen Anteil 1979 (↗ Bild 19)

Hier sind also zwei Streifendiagramme (für die zwei Zufallsgrößen „Familienstand Gesamtbevölkerung" und „Familienstand weibliche Bevölkerung") ineinandergefügt, was den Vergleich zwischen den Verteilungen der beiden Zufallsgrößen wesentlich erleichtert.

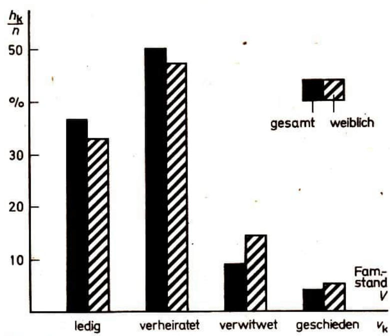


Bild 19

Zur besseren Unterscheidung sind die Streifen der einen Zufallsvariablen voll ausgemalt, die der anderen schraffiert. In einer Legende (Erläuterung zum Bild) werden die beiden Kennzeichnungsarten (voll und schraffiert) erklärt (↗ Bild 19 rechts oben).

Welche Einsichten gewinnen wir nun aus dem Vergleich?

– Der Anteil der Witwen an der weiblichen Bevölkerung ist größer als der Anteil der Verwitweten in der Bevölkerung überhaupt, damit auch größer als der Anteil der Witwer. Das beruht u. a. auf der Tatsache, daß die mittlere Lebenserwartung der Frauen höher liegt als die der Männer.

– Der Anteil der verheirateten Frauen ist geringer als der der Gesamtbevölkerung. Das läßt sich nicht so einfach erklären, denn es gibt (absolut) mehr Frauen als Männer. (Dies geht aus der letzten Zeile der Tabelle im Beispiel 25 (1) hervor.)

● **40*** Dürfen im Streifendiagramm die Mitten der oberen Rechteckseiten miteinander verbunden werden, so daß ein Linienzug entsteht?

● **41*** Ist im Bild 2 (↗ S. 9) ein Streifendiagramm dargestellt? Handelt es sich um die graphische Darstellung einer Häufigkeitsverteilung?

● **42*** Stellen Sie die Tabelle im Beispiel 20 (↗ S. 32) (Prozentuale Verteilung des Merkmals „Geschlecht" in den Bezirken Karl-Marx-Stadt und Suhl) mit Hilfe eines Streifendiagramms graphisch dar!

Als zweite Möglichkeit der Darstellung von Häufigkeitsverteilungen eines Merkmals mit Rangdaten oder Kategorien als Ausprägungen bietet sich das **Staffelbild** an.

► **E 12** Das **Staffelbild** ist ein Rechteck der Länge n oder N (Umfang der Stichprobe bzw. Grundgesamtheit) oder $1{,}0 = 100\,\%$ (bei relativen bzw. prozentualen Häufigkeiten), das aus so vielen Teilrechtecken besteht, wie Merkmalsausprägungen vorhanden sind (Bild 20).
Die Größe der Teilflächen entspricht der (absoluten oder relativen) Häufigkeit der jeweiligen Merkmalsausprägung.

Das Staffelbild kann sowohl vertikal (wie im Bild 20) als auch horizontal angeordnet werden.
Zur besseren Kennzeichnung der Teilrechtecke werden diese mit unterschiedlicher Schraffur oder Farbe versehen (dunklere Töne unten).
Oft werden relative oder prozentuale Häufigkeiten bevorzugt. Diese gestatten ein rasches Zeichnen des Diagramms und bringen den weiteren Vorteil, mehrere Häufigkeitsverteilungen – als Staffelbild neben- oder untereinander angeordnet – miteinander vergleichen zu können.

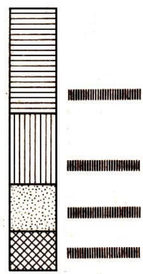


Bild 20

■ **B 26** Einem Geographielehrbuch wurden die folgenden Staffelbilder entnommen (↗ Bild 21):

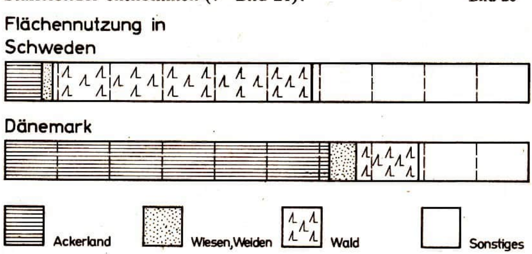


Bild 21

Solche Vergleiche von Staffelbildern sind von hoher Aussagekraft. Diese kann noch gesteigert werden durch Angabe der den Ausprägungen entsprechenden Zahlen (absolute oder prozentuale Häufigkeiten).

■ **B 27** Graphische Darstellung der sozialökonomischen Struktur der Berufstätigen der DDR (1980) mittels Staffelbild (Bild 22)

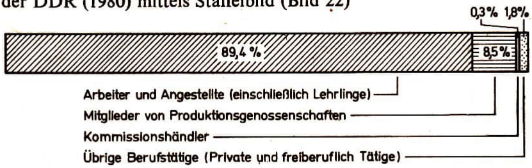


Bild 22

Bei der Gegenüberstellung von zwei oder mehr Staffelbildern geht es um den Vergleich eines Merkmals zwischen Ländern (↗ Bild 21), Bezirken oder Städten, um die Herausstellung des Unterschieds z. B. zwischen Export und Import oder um die Kennzeichnung von Entwicklungen (↗ Bild 23).

■ **B 28** Hier ist der Prozentmaßstab außen angebracht, eine sehr günstige Darstellungsweise.
Es muß jedoch mit Nachdruck betont werden, daß der Vergleich zweier oder mehrerer Staffelbilder leicht zu Fehlschlüssen führen kann, wenn man nicht weiß, worauf sich die 100% beziehen, oder wenn man die absoluten Häufigkeiten für die Merkmalsausprägungen nicht kennt.

Beachten Sie das bei der Lösung der folgenden Aufgaben!

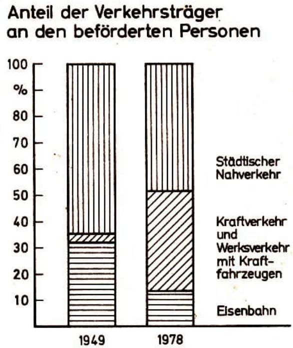


Bild 23 (Nach: Statistisches Taschenbuch der DDR 1979)

● **43*** Zu dem im Bild 21 dargestellten Sachverhalt folgen vier Aussagen. Eine davon ist falsch. Welche und warum?

A. Knapp die Hälfte der Fläche Schwedens ist mit Wald bedeckt.
B. Die Fläche aus Wiesen und Weiden ist in Dänemark größer als in Schweden.
C. In Dänemark ist der Anteil des Ackerlandes an der Gesamtfläche viel größer als in Schweden.
D. Der Ödanteil (Sonstiges) an der Gesamtfläche ist in Dänemark geringer als in Schweden.

● **44** **a)*** Kann man der Darstellung im Bild 23 entnehmen, ob die Zahl der 1978 mit der Eisenbahn beförderten Personen geringer war als die Zahl der 1949 beförderten?

b) Die exakte Angabe der prozentualen Häufigkeiten je Merkmalsausprägung ist am Bild 23 schwieriger zu realisieren als im Bild 22. Inwiefern?

● **45** **a)** Stellen Sie die prozentuale Häufigkeitsverteilung des Merkmals „Familienstand“ für die Gesamtbevölkerung der DDR (↗ Beispiel 25, S. 40) mittels (liegendem) Staffelbild graphisch dar!

b) Vergleichen Sie das Ergebnis mit der Darstellung im Bild 18 (↗ S. 41)!

● **46*** Für die Warenstruktur des Außenhandels der DDR 1980 liegen bezüglich der einzelnen Erzeugnisgruppen folgende Zahlen vor (Angaben in %):

	Export	Import
Brennstoffe, mineralische Rohstoffe, Metalle	14,8	36,7
Chemische Erzeugnisse	10,2	5,5
Andere Rohstoffe, Nahrung und Genuß	6,4	18 9
Industrielle Konsumgüter	14,8	5,0
Maschinen, Ausrüstungen, Transportmittel	51,3	30,8
Übrige Waren	2,5	3,1

Zeichnen Sie dazu zwei stehende Staffelbilder, wobei Sie mit der Erzeugnisgruppe „Brennstoffe, mineralische Rohstoffe, Metalle“ unten beginnen!

Die gleichen Aufgaben wie das Staffelbild kann auch das **Kreisdiagramm** erfüllen (↗ Bild 24). Es gehört ebenfalls zu den Flächendiagrammen.

► **E 13** Das **Kreisdiagramm** stellt sich als ein in Sektoren aufgeteilter Kreis dar. Die Flächen der Sektoren sind den Häufigkeiten der meist qualitativen Merkmalsausprägungen direkt proportional.

Die einzelnen Kreissektoren können mit unterschiedlicher Schraffur versehen oder verschiedenfarbig gestaltet werden. Die Erläuterungen erfolgen in einer Legende oder werden gleich an die Sektoren geschrieben, Zahlenangaben meist *in* die jeweilige Sektorenfläche.

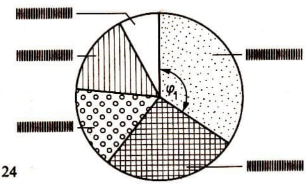


Bild 24

Die Aufteilung der Kreisfläche in die Sektoren geschieht über den Mittelpunktswinkel φ_k $(k = 1, \ldots, m)$. Der Vollwinkel 360° wird entsprechend den Häufigkeiten aufgeteilt. Aus der Verhältnisgleichung

$$\varphi_k : 360^\circ = h_k : n$$

folgt für den Mittelpunktswinkel φ_k in Grad

$$\varphi_k = \frac{h_k}{n} \cdot 360^\circ. \qquad (4)$$

Sind relative Häufigkeiten f_k gegeben, so vereinfacht sich (4) zu

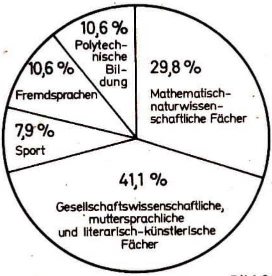
$$\varphi_k = f_k \cdot 360^\circ, \qquad (4a)$$

bei prozentualen Häufigkeiten ist das Produkt noch durch 100 zu dividieren.

■ **B 29** Anteile der Unterrichtsfächer in der Zehnklassigen Polytechnischen Oberschule der DDR an der Zahl der Unterrichtsstunden (↗ Bild 25)

Merkmalsausprägung x_k	Prozentuale Häufigkeit	Winkel φ_k des betr. Sektors
Mathematisch-naturwissenschaftl. Fächer	29,8 %	107,3°
Gesellschaftswissensch., muttersprachl. u. künstlerische Fächer	41,1 %	148,0°
Sport	7,9 %	28,4°
Fremdsprachen	10,6 %	38,2°
Polytechn. Bildung und Erziehung	10,6 %	38,2°
Summe	100,0 %	360,1°

Beachten Sie: Beim Zeichnen der Diagramme die prozentualen Häufigkeiten nicht mit den Mittelpunktswinkeln verwechseln!



(Nach: ND 258/80 vom 2.11.80)

Bild 25

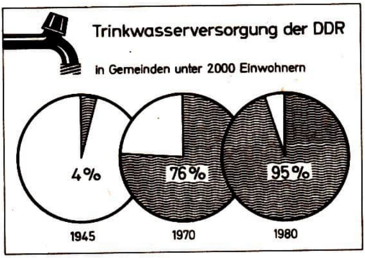


Bild 26

Zum Vergleich ein- und desselben Merkmals zu verschiedenen Zeitpunkten oder für unterschiedliche Bereiche zeichnet man zwei oder mehrere solcher Kreisdiagramme nebeneinander, wobei sich diese teilweise überlappen dürfen (↗ Bild 26).
In einem solchen Schaubild können Entwicklungen über Zeiträume hinweg oder/und unter Berücksichtigung mehrerer Aspekte sehr anschaulich und überzeugend dargestellt werden (↗ Bild 4, S. 11).

● **47** **a)** Stellen Sie die sozialökonomische Struktur der Berufstätigen der DDR (1980) (Angaben aus B 27, S. 43) mittels Kreisdiagramm graphisch dar!
b) Spricht Ihrer Ansicht nach das Staffelbild (Bild 22, S. 43) oder das Kreisdiagramm mehr an?

● **48** Erfassen Sie im Rahmen Ihrer Klasse oder Klassenstufe Daten zu den im weiteren angeführten Zufallsgrößen, und stellen Sie die sich ergebenden Häufigkeitsverteilungen korrekt und so zweckmäßig wie möglich (tabellarisch und graphisch) dar!

a) „Geschlecht der Schüler" für alle Schüler Ihrer Klasse
b) „Sozialökonomische Struktur aller werktätigen Eltern" (in den Kategorien: Arbeiter und Angestellte, Mitglieder von Produktionsgenossen-

schaften, Komissionshändler, übrige Berufstätige (Private und freiberuflich Tätige)

c) „Leistung in einer Mathematik-Kontrollarbeit", die mit Punkten bewertet wurde, von denen man annehmen kann, daß sie gleichgewichtig (gleichwertig) sind

d) „Hauptsächlich betriebene Sportart" für die Schüler Ihrer Klasse (Fußball, Schwimmen, Schach usw.)

e) Vergleich zwischen „Leistung in einer Physikarbeit" und „Leistung im etwa zur gleichen Zeit geschriebenen Klassenaufsatz" (es liegen jeweils Noten vor) für Ihre Klasse

f) Vergleiche von Klasse zu Klasse bezüglich der Merkmale „Anzahl gesammelter Flaschen" und „Masse Altpapier (in kg)" im Rahmen der Erfassung von Sekundärrohstoffen an Ihrer Schule

● **49*** In der Presse finden Sie nicht selten Darstellungen wie im Bild 27.
Erstellen Sie hierzu die Häufigkeitstabelle mit absoluten, relativen und prozentualen Häufigkeiten!

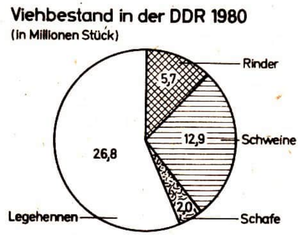


Bild 27

Um graphische Darstellungen von Häufigkeitsverteilungen jeweils richtig auszuführen, hat man den Einfluß der Datenart – bei Meßwerten – der Art der Zufallsgröße (stetig oder diskret) zu beachten. Das Bild 28 gibt eine Übersicht für Häufigkeitsverteilungen mit *einer* Zufallsgröße.

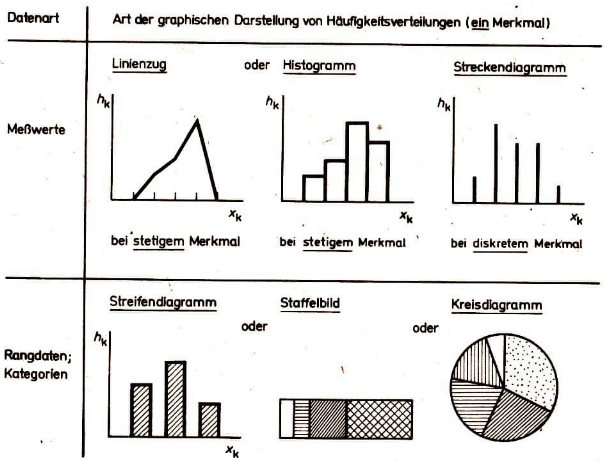


Bild 28

2.2.3. Klassenbildung

Die Überschrift darf uns nicht zu der Annahme verleiten, wir wollten jetzt die Klassen in den Schulen anders ordnen oder neue bilden.
Nein, nachdem wir die Daten geordnet haben, geht es jetzt um den zweiten wichtigen Schritt bei der Datenaufbereitung, um das **Verdichten der Daten.** Darunter verstehen wir die Gruppierung der Daten, genauer gesagt, die Zusammenfassung von mehreren Merkmalsausprägungen zu Gruppen oder Klassen. Wenn im Zusammenhang mit einer Untersuchung Schulklassen beteiligt sind, so wollen wir sie stets als Schulklassen bezeichnen, dadurch sind Verwechslungen weitgehend ausgeschlossen.
An einigen Beispielen des Abschnitts 2.2.2. konnten wir sehen, daß die Häufigkeitsverteilung von Merkmalen mit *quantitativen Ausprägungen* – also der Linienzug, das Histogramm oder (bei diskreten Merkmalen) das Streckendiagramm – ein sehr „unruhiges“, „zerrissenes“ Aussehen haben kann. Diese Form ist auf zufallsbedingte Unregelmäßigkeiten zurückzuführen und läßt die Gesetzmäßigkeit der untersuchten Erscheinung nicht deutlich hervortreten.
Das wird an der Verteilung der jetzt weiter zu verfolgenden Aufgabe 17 (Messung der Nutentiefe an 40 Wellen, Urliste ↗ S. 18) ganz augenscheinlich. Wir erarbeiten dazu die Häufigkeitstabelle und stellen die Verteilung graphisch dar.

■ **B 30** Häufigkeitstabelle zum stetigen Merkmal „Nutentiefe produzierter Wellen“

Merkmalsausprägung x_k ($k = 1, \ldots, 23$) (in mm)	Striche	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$
2,7	I	1	0,025
2,8		0	0
2,9	~~IIII~~ II	7	0,175
3,0	~~IIII~~ ~~IIII~~ I	11	0,275
3,1	IIII	4	0,100
3,2	~~IIII~~	5	0,125
3,3	III	3	0,075
3,4	I	1	0,025
3,5	III	3	0,075
3,6		0	0
3,7		0	0
3,8		0	0
3,9	III	3	0,075
4,0		0	0
4,1	I	1	0,025
4,2		0	0
4,3		0	0
4,4		0	0
4,5		0	0
4,6		0	0
4,7		0	0
4,8		0	0
4,9	I	1	0,025
Summe		40	1,000

Die graphische Darstellung (↗ Bild 29) zeigt einen Linienzug, der durch die Unterbrechungen und Zacken recht unruhig wirkt. Der weit außen liegende Wert von 4,9 mm ist übrigens ein sogenannter „Ausreißerwert".
Nun können die aus der Tabelle oder aus dem Bild 29 ablesbaren Häufigkeiten der Einzelwerte durchaus von Bedeutung sein, meist ist man jedoch an der Gesetzmäßigkeit interessiert, die sich hinter der beobachteten Erscheinung verbirgt. Dann erweisen sich die Unebenheiten als eine Art Störgröße, die man durch Klassenbildung weitgehend beseitigen kann.

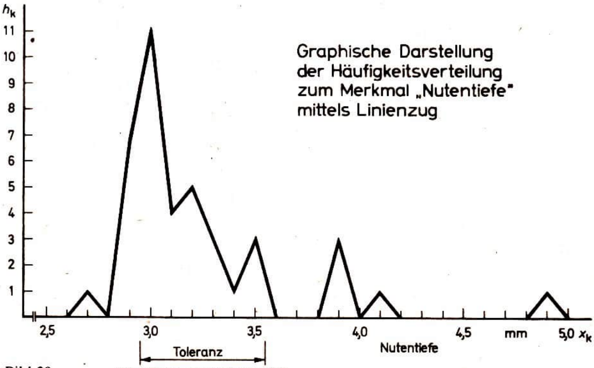


Bild 29

▶ **E 14 Klassenbildung** ist die Zusammenfassung zweier oder mehrerer benachbarter Merkmalsausprägungen zu Gruppen oder Klassen.
Die Klassenbildung dient dazu, das Charakteristische der untersuchten Verteilung deutlicher sichtbar zu machen.

Die Klassenbildung selbst ist nun gar nicht so einfach, wie es auf den ersten Blick erscheint. Sie darf vor allem *nicht formal* vollzogen werden. Die Klassenbildung kann nämlich zu verschiedenen Lösungen führen. Diese hängen ab von der Wahl der *Klassenbreite d* und des *Anfangswerts a der ersten Klasse.*
Grob ausgedrückt kennzeichnet die Klassenbreite die Breite des Intervalls, in welchem wir benachbarte Merkmalsausprägungen zu einer Gruppe zusammenfassen.

■ **B 30 (2)** So könnten wir bei unserem Beispiel „Nutentiefe" (↗ S. 47) eine Klassenbreite $d = 0{,}2$ oder $d = 0{,}3$ oder $d = 0{,}4$ (jeweils mm) vorsehen. Die daraus resultierende Anzahl der Klassen würde sich zwischen 12 und 6 bewegen, *d* ist also von nicht unbedeutendem Einfluß.
Entscheiden wir uns für eine Klassenbreite von $d = 0{,}3$ mm – Kriterien für gute Klassenbildung lernen wir noch kennen –, so können wir die erste Klasse mit dem Merkmalswert 2,7 beginnen lassen, aber auch mit den in der Tabelle auf Seite 47 gar nicht aufgeführten Merkmalswerten 2,5 oder 2,6 mm. Aus der Festlegung einer bestimmten Klassenbreite ergibt sich also wiederum eine bestimmte Anzahl von Varianten der Klasseneinteilung.

Wir betrachten zwei von den drei hier möglichen Varianten.
Variante A: $a = 2{,}7$ mm, eckige Klammern
Variante B: $a = 2{,}5$ mm, geschweifte Klammern (↗ Tabelle S. 47)
Die beiden Varianten führen zu folgenden Klasseneinteilungen:

Variante A			**Variante B**		
Klasse (in mm)	Klassenmitte x_k (in mm)	Absolute Häufigkeit h_k	Klasse (in mm)	Klassenmitte x_k (in mm)	Absolute Häufigkeit h_k
2,7 bis 2,9	2,8	8	2,5 bis 2,7	2,6	1
3,0 bis 3,2	3,1	20	2,8 bis 3,0	2,9	18
3,3 bis 3,5	3,4	7	3,1 bis 3,3	3,2	12
3,6 bis 3,8	3,7	0	3,4 bis 3,6	3,5	4
3,9 bis 4,1	4,0	4	3,7 bis 3,9	3,8	3
4,2 bis 4,4	4,3	0	4,0 bis 4,2	4,1	1
4,5 bis 4,7	4,6	0	4,3 bis 4,5	4,4	0
4,8 bis 5,0	4,9	1	4,6 bis 4,8	4,7	0
			4,9 bis 5,1	5,0	1
Summe		40	Summe		40

B ist – wie wir sehen werden – für die vorliegende Verteilung die günstigere Variante. Das wird sich auch an der graphischen Darstellung widerspiegeln.

Vorerst benötigen wir jedoch präzise Erklärungen der neu aufgetretenen und einiger weiterer Begriffe.
Betrachten wir die zweite der in Variante B dargestellten Klassen (2,8 bis 3,0 mm) und veranschaulichen wir uns deren Lage auf der Merkmalsachse (↗ Bild 30).

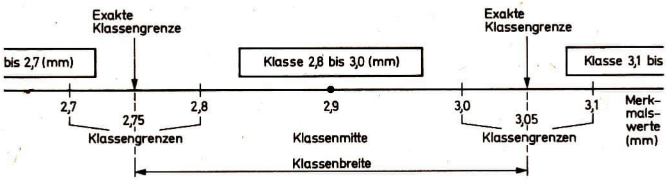


Bild 30

▶ **E 15** Die **Klassengrenzen** (oder Intervallgrenzen) werden durch den kleinsten und den größten *aufgeführten* Merkmalswert der betreffenden Klasse bestimmt.
Der kleinste Merkmalswert der Klasse heißt **untere Klassengrenze** x_u, der größte **obere Klassengrenze** x_o.

Bei *stetigen Merkmalen mit quantitativen Ausprägungen* – insbesondere bei Meßwerten im eigentlichen Sinne – ist es ratsam und oft nötig, **exakte Klassengrenzen** anzugeben und mit diesen zu operieren.

▶ **E 16 Exakte Klassengrenzen** umschließen jeweils ein Intervall, in welche *alle möglichen Merkmalswerte* der betreffenden Klasse hineinfallen, auch wenn Zwischenwerte gar nicht erfaßt werden.
Man unterscheidet zwischen **exakter unterer Klassengrenze** und **exakter oberer Klassengrenze,** x_{ug} bzw. x_{og}.

Beachten Sie: Kennzeichnet man Klassen durch ihre exakten Klassengrenzen, so ist die Klassenbezeichnung in folgender Weise vorzunehmen:

Klasse (in mm)	Absolute Häufigkeit
2,45 bis unter 2,75	1
2,75 bis unter 3,05	18
...	...

Würde das nicht so geschehen, wären Mißverständnisse möglich, z. B. könnte der Merkmalswert der exakten Klassengrenze *zwei* Klassen angehören.

■ **B 30(3)** Zu unserem Beispiel „Nutentiefe" (↗ S. 47f., 49 mit der Variante B und Bild 30) betrachten wir die *Klasse* 2,8 bis 3,0 mm. Hierbei sind:
untere Klassengrenze: $x_u = 2{,}8$ mm,
obere Klassengrenze: $x_o = 3{,}0$ mm
und, da Meßwerte erhoben wurden:
exakte untere Klassengrenze: $x_{ug} = 2{,}75$ mm,
exakte obere Klassengrenze: $x_{og} = 3{,}05$ mm.
Für die Klasse „4,3 mm bis 4,5 mm" sind $x_u = 4{,}3$ mm; $x_o = 4{,}5$ mm; $x_{ug} = 4{,}25$ mm; $x_{og} = 4{,}55$ mm.

● **50*** Geben Sie x_u, x_o, x_{ug} und x_{og} für die letzte Klasse der Variante B (↗ Tabelle auf S. 49) an!

▶ **E 17** Die **Klassenmitte** ist der die Mitte der Klasse kennzeichnende Wert. Sie wird berechnet als arithmetisches Mittel der Klassengrenzen *oder* der exakten Klassengrenzen.
Die Klassenmitte wird oft als *Repräsentant* (stellvertretend) *für die betreffende Klasse* verwendet und dann mit $x_k (k = 1, \ldots, m)$ bezeichnet.
Unter **Klassenbreite** d (auch Intervallbreite oder Klassenweite) verstehen wir die Differenz zwischen exakter oberer und exakter unterer Klassengrenze.

$$d = x_{og} - x_{ug} \qquad (6)$$

■ **B 30 (4)** Für die von uns im Beispiel 30 (3) betrachtete *Klasse* 2,8 bis 3,0 (mm) (↗ Bild 30) ist
die *Klassenmitte* $x_2 = 2{,}9$ mm (x_2, weil es sich um die zweite Klasse der betrachteten Verteilung handelt). Dieser Wert leuchtet hier sofort ein, in schwierigeren Fällen berechnet man die Klassenmitte über

$$x_2 = \frac{x_u + x_o}{2} = \frac{2{,}8 + 3{,}0}{2}\,\text{mm} = 2{,}9\,\text{mm} \qquad \text{oder}$$

$$x_2 = \frac{x_{ug} + x_{og}}{2} = \frac{2{,}75 + 3{,}05}{2}\ \text{mm} = 2{,}9\ \text{mm}.$$

Die *Klassenbreite d* ist hier $d = (3{,}05 - 2{,}75)\ \text{mm} = 0{,}3\ \text{mm}$.

Für die *Klasse* 4,3 bis 4,5 mm sind

$x_7 = 4{,}4\ \text{mm}$ und $d = 0{,}3\ \text{mm}$.

In den meisten Fällen sind alle Klassen gleich groß. Dann kann man die Klassenbreite d auch als Differenz zweier benachbarter Klassenmitten errechnen.

Anmerkung: Zuweilen wird die Klassenbreite auch als Anzahl der in der Klasse zusammengefaßten Werte erklärt. Das gilt aber nicht allgemein, sondern nur für Merkmale, die ganzzahlig erfaßt werden.

Auch Häufigkeitsverteilungen, mit denen keine (oder noch keine) Klassenbildung vorgenommen wurde, können mit den in E 16 und E 17 erklärten Begriffen belegt werden. Bei solchen Verteilungen liegt dann oft die Klassenbreite $d = 1$ vor.

■ **B 9 (5)** Wir wenden uns erneut dem Beispiel mit dem Merkmal X „Arbeitszeit des Lehrlings A je Steckermontage" (↗ S. 16, 30, 36 und 37) zu. In der Tabelle auf Seite 31 können wir die Merkmalsausprägungen als Klassen mit $d = 1$ s ansehen. Für die erste Merkmalsausprägung $x_1 = 40$ (= Klasse) gilt dann:

Exakte untere Klassengrenze: $x_{ug} = 39{,}5$ s,
exakte obere Klassengrenze: $x_{og} = 40{,}5$ s,
Klassenmitte: $x_1 = 40$ s,
Klassenbreite: $d = 1$ s.

Klassen, die nach oben oder nach unten nicht begrenzt sind, heißen **offene Klassen.**

■ **B 31** Zufallsgröße „Alter in Jahren" (im Rahmen einer Untersuchung einer größeren Anzahl von Menschen)

Häufigkeitstabelle nach Klassenbildung:

Klasse (Jahre)	Häufigkeit
Unter 20	350
20 bis unter 35	1527
35 bis unter 50	684
50 bis unter 65	75
65 und darüber	19
Summe	2655

Die erste und letzte Klasse sind hier offene Klassen.

Bei offenen Klassen hat die Angabe der Klassenmitte, der Klassenbreite und (bei nach oben offenen Klassen) auch der oberen Klassengrenze keinen Sinn.

● **51*** Es liege nebenstehender Ausschnitt aus einer Verteilungstabelle zur Zufallsgröße „Geschwindigkeit" (Einheit km/h) vor.
Geben Sie für die in der Mitte stehende Klasse an: Exakte obere Klassengrenze, Klassenbreite, untere Klassengrenze, Klassenmitte!

Klasse $\left(\frac{\text{km}}{\text{h}}\right)$	Häufigkeit
...	...
70 bis unter 75	7
75 bis unter 80	2
80 bis unter 85	3
...	...

Wir haben jetzt die wichtigsten Begriffe kennengelernt, die im Zusammenhang mit der Klassenbildung auftreten: Die Klassenbildung ist aber nicht Selbstzweck.
Das Ziel unserer Bemühungen besteht in diesem Abschnitt darin, die erhobenen Daten so zu verdichten, daß sich Einzelinformationen nicht zu sehr verwischen, daß sie andererseits aber auch nicht zu stark hervortreten und dadurch das Aufdecken der hinter der untersuchten Erscheinung verborgenen Gesetzmäßigkeit erschweren. Ganz sicher ist, daß der **Bestimmung der richtigen Klassenbreite** eine große Bedeutung zukommt.
Bei *zu großer Klassenbreite* erhalten wir *nur wenige Klassen, verwischen* aber nicht nur die Einzelheiten, sondern möglicherweise auch *die Besonderheiten* der Verteilung;
bei *zu kleiner Klassenbreite* erhalten wir *zu viele Klassen*, damit bleiben *zu viele Einzelheiten* erhalten; auch hier gelingt es nicht, das Besondere der Verteilung abzuheben.
Daß auch die Wahl des Anfangswertes a der ersten Klasse nicht ohne Einfluß ist, hatten wir bereits erkannt.
Anhand unseres Beispiels wollen wir je eine Darstellung für zu große und für zu kleine Klassenbreite einander gegenüberstellen. In den Tabellen und Diagrammen werden wir von nun an vorwiegend mit den Klassenmitten x_k arbeiten. Sie sind die kürzeste Darstellungsform für die betreffende Klasse.

● **52** Rechnen Sie einige dieser Klassenmitten nach!

■ **B 30 (5)** Wir betrachten nun zum bereits mehrfach erörterten Beispiel „Nutentiefe" (↗ Tabelle S. 47) die tabellarische Darstellung der Häufigkeitsverteilung für $d = 0{,}6$ mm und $a = 2{,}7$ mm (Fall 1) sowie für $d = 0{,}2$ mm und $a = 2{,}6$ mm (Fall 2). Die Bilder 31 und 32 zeigen (jeweils als Linienzug) die graphischen Darstellungen hierfür.

Fall 1:

Klasse (mm)	Klassenmitte x_k (mm)	Absolute Häufigkeit h_k
2,7 bis 3,2	2,95	28
3,3 bis 3,8	3,55	7
3,9 bis 4,4	4,15	4
4,5 bis 5,0	4,75	1
Summe		40

Zu große Klassenbreite, zu geringe Klassenanzahl

Fall 2:

Klasse (mm)	Klassenmitte x_k (mm)	Absolute Häufigkeit h_k
2,6 bis 2,7	2,65	1
2,8 bis 2,9	2,85	7
3,0 bis 3,1	3,05	15
3,2 bis 3,3	3,25	8
3,4 bis 3,5	3,45	4
3,6 bis 3,7	3,65	0
3,8 bis 3,9	3,85	3
4,0 bis 4,1	4,05	1
4,2 bis 4,3	4,25	0
4,4 bis 4,5	4,45	0
4,6 bis 4,7	4,65	0
4,8 bis 4,9	4,85	1
Summe		40

Zu kleine Klassenbreite, zu große Klassenanzahl

Wir markieren auf der Merkmalsachse die Klassenmitten, weil wir über diesen die Häufigkeit der jeweiligen Klasse aufzutragen haben. Da die Linienzüge an den Rändern der Verteilung bis zur Merkmalsachse gezogen werden, müssen wir zusätzlich zu den schon angeführten Klassenmitten auch die für die linke und rechte jeweils anschließende Klasse berechnen.

Für 2,1 bis 2,6 ergibt sich

z. B. $\frac{2,1 + 2,6}{2} = 2,35.$

Für 2,4 bis 2,5 ergibt sich 2,45.

Keine der hier vorgenommenen Klasseneinteilungen ist günstig. Das Bild 31 spiegelt den Sachverhalt zu grob wider, das Bild 32 noch zu fein!
Die beste Darstellung ist die in der Variante B auf Seite 49 verwirklichte ($d = 0,3$ mm; $a = 2,5$ mm), die aus der Tabelle im Beispiel 30 unter Beachtung der geschweiften Klammern entstanden ist.
Wir stellen unten den zugehörigen Linienzug dar (Bild 33).

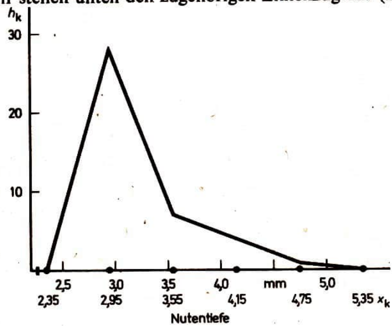


Bild 31

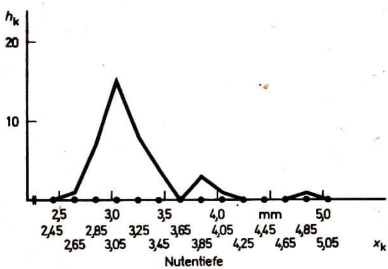


Bild 32

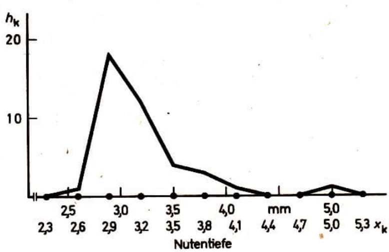


Bild 33

Warum ist das nun die gelungenste Darstellung für die in diesem Beispiel erhobenen Daten?
Folgende wichtige Hinweise liefern *Anhaltspunkte für die Bestimmung der richtigen Klassenanzahl und damit Klassenbreite sowie für einen günstigen Anfangswert der ersten Klasse:*

1. Ist n die Anzahl der untersuchten Objekte (Umfang der Stichprobe) und m die Anzahl der Klassen, so ist zu beachten:

Für $n =$	10	16	25	40	64	100	160	250	400	1000	
soll $m \leqq$	5	6	7	8	9	10	11	12	13	15	sein.

2. Als vorteilhaft erweisen sich die Klassenbreiten $d = 2, 3, 5, 7$ oder 10 (bzw. je nach Einheit Zehntel oder Hundertstel davon). Auch andere Klassenbreiten sind möglich.
3. Der Anfangswert a der ersten Klasse und die Klassengrenzen sind (unter der Voraussetzung konstanter Klassenbreite) so festzulegen, daß *Merkmalsausprägungen mit starken Häufigkeiten möglichst in der Klassenmitte* liegen.

Dieser 3. Punkt ist besonders wichtig. Er läßt sich nicht immer ideal erfüllen (für $d = 2$ überhaupt nicht), man sollte sich aber bemühen, ihm weitgehend Rechnung zu tragen. Die geschweiften Klammern in der Tabelle auf Seite 47 entsprechen im großen und ganzen den hier formulierten Punkten. Infolgedessen sind die daraus abgeleiteten Darstellungen (Variante B, S. 49, und Bild 33, S. 53) die besten Realisierungen der Klassenbildung zum Beispiel „Nutentiefe".

Wer die Häufigkeitsverteilung nach Klassenbildung (Variante B) bezüglich der drei Anhaltspunkte prüft, wird feststellen, daß der 1. Punkt formal nicht erfüllt ist. Für $n = 40$ müßte $m \leqq 8$ (gemäß der Wertetabelle unter 1. auf Seite 54) sein. Die Zahl der Klassen in der Variante B ist aber 9. Diese Klasseneinteilung ist trotzdem gerechtfertigt, denn praktisch liegen hier nur 6 Klassen vor, sofern man den Ausreißerwert (4,9 mm) für diese Betrachtung einmal unberücksichtigt läßt. (In der Statistik gibt es Ausreißertests, die eine Entscheidung darüber herbeiführen, ob ein solcher Ausreißerwert weggelassen werden darf oder nicht.)
Es sei also nochmals betont: Nicht formal an die Klassenbildung herangehen!

Als Ergebnis der Datenaufbereitung und -darstellung ergibt sich für das Beispiel „Nutentiefe" also eine *linksschiefe Verteilung*. Das Maximum liegt links von der Verteilungsmitte und außerhalb des Toleranzbereichs (3,0 bis 3,5 mm). Mehrere der produzierten Wellen haben eine zu geringe Nutentiefe.

Im Hinblick auf die Qualität des Fertigungsprozesses wäre nebenstehendes Verteilungsbild günstiger gewesen (Bild 34).
Eine solche glockenförmig verlaufende, symmetrische Häufigkeitsverteilung nennt man *Normalverteilung*. Wir kommen im Abschnitt 2.2.4. darauf zurück.

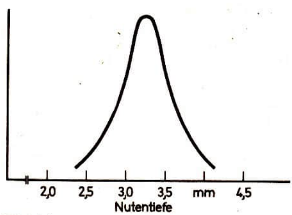


Bild 34

Ebenso wie bei der Häufigkeitstabelle *ohne* Klasseneinteilung sind auch in den Tabellen *mit* Klasseneinteilung **relative und prozentuale Häufigkeiten** (bei Vergleichen z. B.) oft günstiger als absolute.

■ **B 30 (6)** Merkmal „Nutentiefe". Wir betrachten die Tabelle zur Variante B auf Seite 49 unter Zusatz der relativen und prozentualen Häufigkeiten.

Klasse (mm)	Klassenmitte x_k (mm)	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$	Prozentuale Häufigkeit
2,5 bis 2,7	2,6	1	0,025	2,5%
2,8 bis 3,0	2,9	18	0,450	45,0%
3,1 bis 3,3	3,2	12	0,300	30,0%
3,4 bis 3,6	3,5	4	0,100	10,0%
3,7 bis 3,9	3,8	3	0,075	7,5%
4,0 bis 4,2	4,1	1	0,025	2,5%
4,3 bis 4,5	4,4	0	0	0%
4,6 bis 4,8	4,7	0	0	0%
4,9 bis 5,1	5,0	1	0,025	2,5%
Summe		40	1,000	100,0%

Der Klassenbildung kommt im Rahmen der Datenaufbereitung eine bedeutende Rolle zu. Wir greifen die beiden Beispiele 9 und 10 (↗ S. 16f.) erneut auf, führen diese mit der Klassenbildung fort und bedenken dabei Besonderheiten, die in dem einen oder anderen Falle auftreten.

■ **B 9 (6)** Das Beispiel 9 mit der Untersuchung der Zufallsgrößen X und Y „Arbeitszeit des Lehrlings A bzw. B je Steckermontage" beschäftigte uns schon an mehreren Stellen des Buches (↗ S. 30, 36, 37, 51 sowie in den Aufgaben 26, 29, 32, 34).
Im Zusammenhang mit der Klassenbildung gilt es, folgendes zu beachten: Da wir an einem *Vergleich* der Leistungen der Lehrlinge A und B interessiert sind, darf die Klassenbildung nicht getrennt am Datenmaterial für A und für B erfolgen, sondern muß an der zusammengefaßten Häufigkeitstabelle (für A und B) vorgenommen werden. Dabei dürfen die drei Anhaltspunkte für eine gute Klasseneinteilung (S. 54) nicht außer acht gelassen werden.

Merkmalsausprägung x_k (s)	Absolute Häufigkeit h_k für Lehrling A	für Lehrling B	insgesamt für A und B
40	1	1	2
41	5	10	15
42	6	11	17
43	11	11	22
44	9	10	19
45	9	5	14
46	3	2	5
47	3		3
48	0		0
49	1		1
50	1		1
51	1		1
Summe	50	50	100

Wir entscheiden uns also für eine Klasseneinteilung mit der Klassenbreite: $d = 2$ s. Als Anfangswert der ersten Klasse ergibt sich $a = 40$ s.

Klasse (s)	Klassenmitte x_k (s)	Absolute Häufigkeit h_k nach Klassenbildung für A	für B	Prozentuale Häufigkeit nach Klassenbildung für A	für B
40 bis 41	40,5	6	11	12%	22%
42 bis 43	42,5	17	22	34%	44%
44 bis 45	44,5	18	15	36%	30%
46 bis 47	46,5	6	2	12%	4%
48 bis 49	48,5	1	0	2%	0%
50 bis 51	50,5	2	0	4%	0%
Summe		50	50	100%	100%

Anmerkung: Bei der Klassenbezeichnung ist statt „40 bis 41" ebenso möglich „39,5 bis unter 41,5", usw.

Kontrolle des Erfülltseins der drei Punkte von S. 54:

Zu 1: Für $n = 40$ soll gelten $m \leqq 8$. Hier liegt $n = 50$ vor, $m = 6$ Klassen sind also gerechtfertigt. Eine andere Klassenbreite wäre nicht angebracht.

Zu 2: Erfüllt. $d = 2$ s ist eine vorteilhafte Klassenbreite.

Zu 3: Für $d = 2$ s nicht erfüllbar.

Graphische Darstellung der Häufigkeitsverteilungen (prozentuale Häufigkeiten) zu Merkmal X und Y nach Klassenbildung mittels Histogramm (↗ Bilder 35 und 36):

Anmerkungen: Auf der Merkmalsachse werden im allgemeinen *entweder* nur die Klassen *oder* die exakten Klassengrenzen *oder* – zumeist – nur die Klassenmitten angegeben und nicht alle drei Bezeichnungsarten wie hier.
Auf der Ordinatenachse hätten statt der prozentualen auch die absoluten Häufigkeiten aufgetragen werden können. Zuweilen erweist sich die Angabe beider Häufigkeiten als günstig.

Die genaue Untereinanderanordnung der beiden Histogramme erleichtert Vergleiche zwischen den Leistungen der Lehrlinge A und B. Es bestätigt sich, daß der Lehrling B bezüglich der Montagezeit bessere Leistungen bringt; die Mehrzahl seiner Werte liegt in oder unter der Norm. Beide Zufallsgrößen sind angenähert normalverteilt. X weist eine größere Streuung auf, was man an den Diagrammen erkennt, später (im Abschnitt 2.3.4.) aber auch rechnerisch nachgewiesen wird.

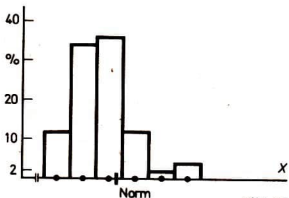


Bild 35

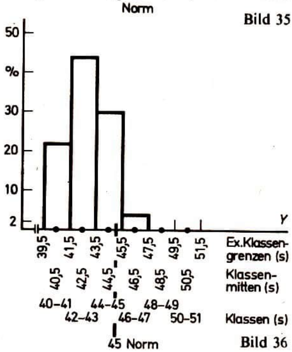


Bild 36

Verwendet man statt der Darstellungsart Histogramm den Linienzug, so lassen sich die Verteilungen beider Zufallsgrößen in *einem* Diagramm darstellen (↗ Bild 37).

● **53*** Es sollen wieder zwei Lehrlinge A und B verglichen werden, diesmal aber bezüglich der Zufallsgröße W „Qualität des bearbeiteten Werkstücks"

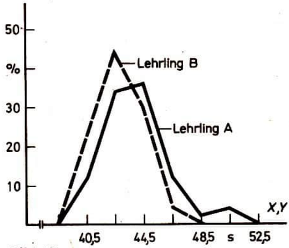


Bild 37

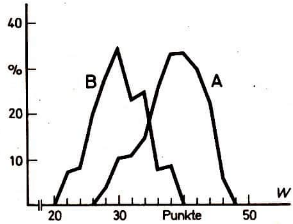


Bild 38

(gemessen in Punkten, denen eine Intervallskala zugrunde liegt). Das Ergebnis der Datenaufbereitung ist im Bild 38 dargestellt.

a) Der Linienzug von A liegt weiter rechts als der von B. Ist die Leistung von A höher einzuschätzen als die von B oder umgekehrt? Wer bringt mehr Qualität in seiner Arbeit?

b) Trifft der Tatbestand immer zu, daß die Leistung *der* Stichprobe besser ist, deren Linienzug (oder Histogramm oder Streckendiagramm) sich auf der Merkmalsachse weiter rechts befindet? Wenn nein, wo trifft das nicht zu?

Wir führen jetzt das Beispiel 10 von Seite 17 in Kurzfassung weiter.

■ **B 10 (3)** Häufigkeitsverteilung zur Untersuchung der Zufallsgröße Z „Anzahl der je Ampelphase eine Kreuzung passierenden Fahrzeuge". Die an drei Werktagen jeweils abends erhobenen Werte werden zusammengefaßt.
Da es sich hier um eine *diskrete* Zufallsgröße handelt, ist das Streckendiagramm die richtige Form der graphischen Darstellung (↗ Bild 16, S. 39).
Darauf ist auch bei der graphischen Darstellung *nach* Klassenbildung zu achten (↗ Bild 39).

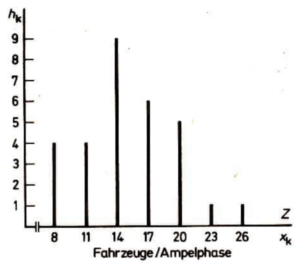


Bild 39

Häufigkeitstabelle mit Klassenbildung:

Merkmalsausprägung z_k (Fahrz./Ampelph.)	Absolute Häufigkeit h_k	Klasse (Fahrz./Ampelph.)	Klassenmitte z_k (Fahrz./Ampelph.)	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$
7	2				
8	1	7 bis 9	8	4	0,133
9	1				
10	0				
11	2	10 bis 12	11	4	0,133
12	2				

Merkmalsausprägung z_k (Fahrz./Ampelph.)	Absolute Häufigkeit h_k	Klasse (Fahrz./Ampelph.)	Klassenmitte z_k (Fahrz./Ampelph.)	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$
13	4				
14	4	13 bis 15	14	9	0,300
15	1				
16	1				
17	3	16 bis 18	17	6	0,200
18	2				
19	1				
20	2	19 bis 21	20	5	0,167
21	2				
22	0				
23	1	22 bis 24	23	1	0,033
24	0				
25	0				
26	0	25 bis 27	26	1	0,033
27	1				
Summe	30			30	0,999

Aufgaben zur Klassenbildung:

● **54** Vergleichen Sie die Bilder 16 (↗ S. 39) und 39 (↗ S. 57) miteinander, und überdenken Sie Sinn und Zweck der Klassenbildung!

● **55*** Untersuchung zur Zufallsgröße „Leistung im Mathematik-Olympiade-Wettbewerb" (schließt an Aufgabe 36, S. 38) an. Die Häufigkeitstabelle findet sich auf der Seite 125f.
Bilden Sie Klassen, und stellen Sie die Verteilung nach der Klassenbildung tabellarisch und graphisch (auf Millimeterpapier) dar! Geben Sie auch die relativen Häufigkeiten an!

● **56*** Untersuchung zur Zufallsgröße „Masse von Schweinen" (in kg). Die Urliste finden Sie innerhalb der Aufgabe 22 (↗ S. 22).

a) Stellen Sie in *einer* Tabelle dar:
Strichliste, Häufigkeitsverteilung (absolut, vor und nach Klassenbildung), prozentuale Häufigkeitsverteilung (nur nach Klassenbildung)!
Empfehlung für Klasseneinteilung: $d = 9$ kg; $a = 98$ kg. Zur Bezeichnung der Klassen in der Tabelle verwenden Sie die exakten Klassengrenzen!

b) Haben Sie die drei Anhaltspunkte geprüft, die bei der Klassenbildung zu beachten sind?

c) Geben Sie für die zweite der entstandenen Klassen die exakte untere Klassengrenze, die obere Klassengrenze und die Klassenmitte an!

d) Zeichnen Sie das Histogramm zur Häufigkeitsverteilung nach Klassenbildung, und beurteilen Sie deren Form!

● **57** Bei der Lösung der Aufgabe 48c) auf S. 46 entwickelten Sie eine Häufigkeitstabelle. Nehmen Sie eine geeignete Klassenbildung vor, und stellen Sie das Ergebnis graphisch dar!

2.2.4. Normalverteilung und andere Verteilungsformen

Die Ermittlung der Häufigkeit des Auftretens der einzelnen Merkmalsausprägungen und die Verdichtung der gewonnenen Daten in Merkmalsklassen – kurz: die Datenaufbereitung – bedeutet soviel, wie das **Verteilungsgesetz** des untersuchten Merkmals zu ergründen. Die Kenntnis des Verteilungsgesetzes ist oft eine wichtige Voraussetzung für die Lösung zahlreicher mit dem Untersuchungsproblem zusammenhängender Fragen.

Die Häufigkeitsverteilung einer bestimmten Zufallsgröße ist – bei nicht zu kleiner Zahl von Untersuchungsobjekten – nicht regellos, sondern weist typische Verteilungsformen auf, die Ausdruck des Wirkens bestimmter Gesetzmäßigkeiten sind. An zahlreichen Beispielen (wie etwa 10 ab Seite 17; 30 ab Seite 47) und Aufgaben (36, 55 u. a.) konnten wir sehen, daß unregelmäßige, „zerrissene" Linienzüge, Histogramme oder Streckendiagramme nach der Klassenbildung ein geschlosseneres, ein regelhaftes Aussehen annehmen. Das Verdichten der Daten ermöglicht das Herausheben des Wesentlichen der untersuchten Erscheinung.

Dabei war augenfällig, daß eine bestimmte Verteilungsform sehr häufig auftritt, die **Normalverteilung.** Sie ist die wichtigste theoretische Verteilung für stetige Zufallsgrößen.

► **D 7 Die *Normalverteilung* ist die theoretische Verteilung einer stetigen Zufallsgröße mit einer bestimmten (nicht gerade einfachen) Funktionsgleichung.**

Eine Normalverteilung ist in der Realität immer dann zu erwarten, wenn folgende vier Bedingungen erfüllt sind:

1. Es liegt eine sehr *große Zahl* von Beobachtungswerten vor.
2. Die Auswahl der Untersuchungsobjekte unterliegt ausschließlich dem *Zufall*, sie ist nicht vorausgelesen.
3. Die Zufallsgröße beruht auf dem *Zusammenspiel einer großen Zahl voneinander unabhängiger und etwa gleich stark wirksamer Faktoren.*
4. Die *relative Häufigkeit für das Auftreten gegensinniger Faktoren* ist *annähernd gleich groß.*

Schon das Fehlen einer dieser Bedingungen genügt, keine Normalverteilung (im exakten Sinne) entstehen zu lassen.

■ **B 32** Normalverteilung ist nicht zu erwarten, wenn man zur Bestimmung der mittleren Körpergröße von 50 achtzehnjährigen Jugendlichen 30 Basketballspieler einbezieht. Die Bedingungen 1 bis 4 sind nicht erfüllt.
Hier entstünde etwa eine Verteilungsform wie sie im Bild 40 dargestellt ist.

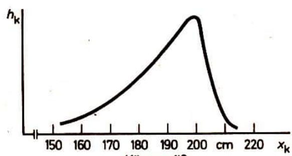


Bild 40

Das ist ein Beispiel für eine rechtsschiefe, fast J-förmige Verteilung.

● **58** Führen Sie in Ihrer Schulklasse, FDJ- oder Sportgruppe folgende kleine Untersuchung selbst durch:
Die beteiligten Schüler, Jugendfreunde oder Sportler – es sollten mehr als 20 sein – befinden sich im Ruhezustand. Jeder ertastet durch leichten Druck den eigenen Pulsschlag. Sie als Versuchsleiter nehmen eine gute Uhr mit Sekundenangabe als Meßinstrument zur Hand und lassen die Versuchspersonen während der Dauer einer Minute die Schläge des eigenen Pulses zählen. Nur Sie schauen auf die Uhr und markieren Anfang und Ende einer Minute durch einen Schlag mit der Hand auf den Tisch.
Erfaßt wird so die Zufallsgröße „Herzfrequenz“, allerdings nur grob als Zahl der Pulsschläge je Minute (Ps/min). Die Frequenz ist ein stetiges Merkmal.
Jeder Versuchsteilnehmer sagt seine Pulszahl an, gemeinsam legen Sie Strichliste und Häufigkeitstabelle an und bilden Klassen (Empfehlung: 46 Ps/min bis 55 Ps/min, 56 bis 65, 66 bis 75 usw.). Wenn nicht zu wenige Personen beteiligt sind, entsteht annähernd eine Normalverteilung (↗ Bild 41).

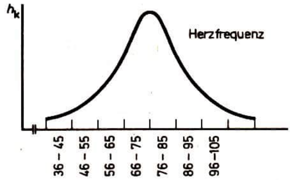


Bild 41

Bei der Untersuchung von Zufallsgrößen in Natur und Gesellschaft ist die Normalverteilung oft anzutreffen. Ihre Haupteigenschaft besteht darin, daß sich die Mehrzahl der Merkmalswerte um einen bestimmten Wert – den Mittelwert – anhäuft und links und rechts davon immer weniger Beobachtungswerte auftreten.
Das ist die *Erscheinungsform einer statistischen Gesetzmäßigkeit*. Die Normalverteilung ist also Ausdruck des *Wechselspiels von Notwendigkeit und Zufall* (vgl. mit Abschnitt 1.2., S. 11), sie stellt ein Modell dar, das die Wahrscheinlichkeit des Eintritts möglicher Ereignisse veranschaulicht.
Das Funktionsbild der Normalverteilung heißt **Normalkurve, Glockenkurve** oder **GAUSS-Kurve** (nach dem großen deutschen Mathematiker Carl Friedrich GAUSS, 1777 bis 1855). Die theoretische Normalkurve ist hauptsächlich gekennzeichnet durch die Eigenschaften

- *Symmetrie* und
- *glockenförmiger Verlauf.*

Formuliert man diese Eigenschaften mit größerer mathematischer Strenge, so würde das bedeuten:

Die Normalkurve ist *eingipflig* und *achsensymmetrisch* (zur Ordinate im Gipfelwert μ, den wir als Mittelwert erkennen werden). *Sie verläuft von* $-\infty$ *bis* $+\infty$, ihre *Ordinaten nehmen ab*, je weiter die Merkmalswerte von μ entfernt sind. Die Kurve *nähert sich der Merkmalsachse immer stärker an*, erreicht sie aber im Endlichen nicht. Die Kurve hat *zwei Wendepunkte* (↗ Bild 42).

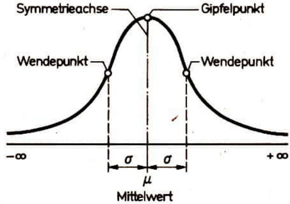


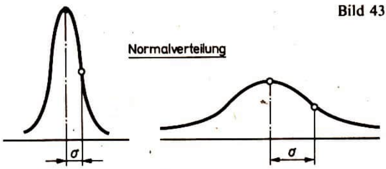
Bild 42

Die Normalkurve wird zuweilen auch als GAUSSsche Fehlerverteilungskurve bezeichnet. Grund: Für die Verteilung der Fehler (Abweichungen vom Sollmaß) ergeben sich bei vielen Meßfolgen (z. B. bei der Messung des Durchmessers von Wellen) Eigenschaften, die mit den oben genannten übereinstimmen:

- Gleich große positive und negative Fehler (Abweichungen) kommen etwa gleich häufig vor (sind „gleich wahrscheinlich“).
- Große Fehler (Abweichungen) sind weit seltener als kleine Fehler.

Wichtige statistische Kenngrößen, die die *Lage* und die *Form der Normalverteilung* charakterisieren, sind der **Mittelwert** μ (griech. Buchstabe, sprich: mü) und die **Streuung** σ (griech., sprich: sigma). Die Streuung ist ein Maß, aus dem die Lagerung aller Einzelwerte um den Mittelwert herum hervorgeht. Ist σ klein, liegt also eine geringe Streuung vor, so scharen sich die Merkmalswerte eng um den Mittelwert. Ist σ groß, liegt also eine große Streuung vor, so verteilen sich die Merkmalswerte breit längs der Merkmalsachse (↗ Bild 43).

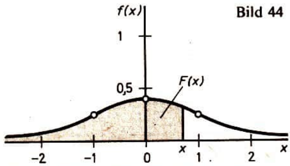
Bild 43



a) σ klein
geringe Streuung

b) σ groß
große Streuung

Bild 44



Standardisierte Normalkurve ($\mu = 0$; $\sigma = 1$)

σ ist der Abstand der Wendepunktsabszisse vom Mittelwert.

Einzelheiten zu Mittelwert und Streuung, insbesondere auch zur Berechnung dieser Kenngrößen, folgen im Abschnitt 2.3.

Bedeutungsvoll sind die **Ordinaten der Normalverteilung** und die **Flächenteile** unter der Normalkurve. Diese Angaben findet man in Tabellen (oft auch Tafeln genannt). Um nicht zu viele Tafeln handhaben zu müssen, wurde eine bestimmte Normalverteilung als Standard (Normgröße) festgelegt: die **standardisierte Normalverteilung.**

► **E 18** Die **standardisierte Normalverteilung** ist gekennzeichnet durch $\mu = 0$ und $\sigma = 1$.

In Tafelwerken findet man sowohl die Ordinaten der standardisierten Normalkurve als auch die Flächen unter derselben; mit diesen Tafeln kann man die GAUSS-Kurve exakt zeichnen und Aufgaben zur Normalverteilung lösen (↗ Bild 44).
Die Fläche unter der gesamten Normalkurve, also zwischen Normalkurve und x-Achse, beträgt 1 ($\triangleq$ 100%). Das ist ein schönes Beispiel dafür, daß eine *unendliche Fläche* einen *endlichen Flächeninhalt* haben kann.
Einige wenige Angaben aus den Tafeln zur Normalverteilung:

x	−2	−1,96	−1	−0,5	0	0,5	1	1,96	2	→ ∞
$f(x)$ = Ordinate	0,054	0,058	0,242	0,352	0,399	0,352	0,242	0,058	0,054	→ 0
$F(x)$ = Fläche	0,023	0,025	0,159	0,310	0,500	0,691	0,841	0,975	0,977	→ 1

● **59** Zeichnen Sie unter Verwendung der wenigen Werte aus der vorstehenden Wertetafel die standardisierte Normalkurve! (Verwenden Sie Millimeterpapier, und nehmen Sie als Einheit auf der x-Achse 2 cm, auf der $f(x)$-Achse 6 cm!)

Mit den Angaben für die Flächen unter der Normalkurve kann man auch Differenzen bilden.

■ **B 33** Die Fläche unter der Normalkurve zwischen $x = -1$ und $x = +1$ beträgt $0{,}682 \mathrel{\hat{=}} 68{,}2\,\%$.
Berechnung: $F(1) - F(-1)$
$= 0{,}841 - 0{,}159 = 0{,}682$
(↗ Bild 45)

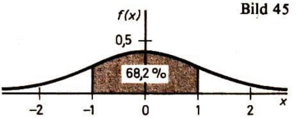


Bild 45

● **60*** Wie groß ist die Fläche unter der Normalkurve
a) zwischen $x = -1{,}96$ und $x = +1{,}96$;
b) zwischen $x = 1{,}96$ und $x \to +\infty$?
c) Veranschaulichen Sie den Flächenteil von b) in Ihrer graphischen Darstellung zu Aufgabe 59!

Diese Flächen geben die Wahrscheinlichkeit an, mit der die Zufallsgröße X bestimmte Werte annimmt (z. B. alle Werte zwischen $-1{,}96$ und $+1{,}96$).
Im wissenschaftlichen Erkenntnisprozeß ist das *Aufstellen und Prüfen von Hypothesen* (Annahmen, Vermutungen über einen bestimmten Sachverhalt) von entscheidender Bedeutung. (Im Abschnitt 3. kommen wir kurz darauf zurück.)
Dabei spielt die Normalverteilung eine wichtige Rolle. Die 95 %, die Sie in Aufgabe 60a) berechnet haben, sind Ausdruck der **statistischen Sicherheit,** mit der eine Hypothese – beispielsweise: Brigade A liefert bessere Qualität als Brigade B – anzunehmen oder zurückzuweisen ist.

Eine angenäherte Normalverteilung kann man auf technische Weise mit dem sogenannten GALTONschen Brett erzeugen (↗ Bild 46); benannt wurde es nach dem englischen Naturforscher FRANCIS GALTON (1822 bis 1911).
Beim GALTONschen Brett rollen Kugeln aus einem Trichter über ein Nagelbrett. Die Kugeln werden durch das Auftreffen auf die Nägel mehr oder weniger stark abgelenkt. In Fächern unter den Nägeln werden die Kugeln aufgefangen und bilden eine angenäherte Normalverteilung.

● **61** Bauen Sie ein GALTONsches Brett! Beachten Sie ein günstiges Verhältnis von Kugeldurchmesser und Nagelabstand!

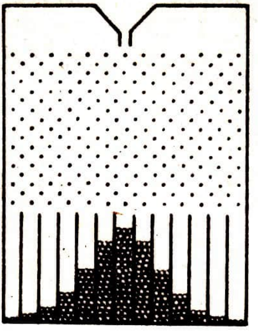


Bild 46

Genaugenommen entsteht am GALTONschen Brett eine **Binomialverteilung.** Das ist die theoretische Verteilung für zweigestufte, also diskrete Zufallsgrößen. Die graphische Darstellung erfolgt als Streckendiagramm.

Mit der Normal- und Binomialverteilung haben wir zwei wichtige **theoretische Verteilungen** kennengelernt. Sie sind das Sinnbild zweier bestimmter statistischer Gesetzmäßigkeiten. Das, was wir aus der Praxis – *aus der Erfahrung* mit unserer Umwelt, durch Beobachten und Forschen – gewinnen, findet seinen Niederschlag in **empirischen Verteilungen.** Sämtliche in den Abschnitten 2.2.2. und 2.2.3. behandelten Verteilungen sind empirische Verteilungen.
An einem weiteren Beispiel soll die enge Beziehung zwischen empirischer und theoretischer Verteilung aufgezeigt werden.

■ **B 34** „Ballweitwurf"; von 51 Schülern der Klassenstufe 8 erzielte Weiten (in m). Der beste von drei Würfen wurde gewertet.

Merkmalsausprägung x_k(m)	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$
15	1	0,020
16	3	0,059
17	2	0,039
18	6	0,118
19	6	0,118
20	10	0,196
21	9	0,176
22	5	0,098
23	2	0,039
24	4	0,078
25	1	0,020
26	2	0,039
Summe	51	1,000

$\sum \frac{h_k}{n} = 1{,}000$ ist gleichbedeutend damit, daß die Fläche unter der Normalkurve den Wert 1 hat.

Die theoretische Verteilung ist Ausdruck einer statistischen Gesetzmäßigkeit, die mit Hilfe der empirischen Verteilung aufgedeckt wird (↗ Bild 47).

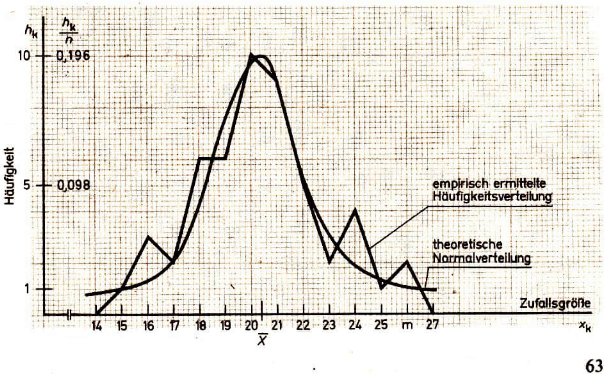


Bild 47

Wir wenden uns jetzt anderen *typischen Formen von Häufigkeitsverteilungen* zu. Dabei haben wir meist stetige Zufallsgrößen im Auge, zeichnen also i. allg. Kurvenzüge; die Bezeichnungen gelten aber gleichermaßen für diskrete Zufallsgrößen, wie wir es gleich zeigen.

Gleichverteilungen (↗ Bild 48)

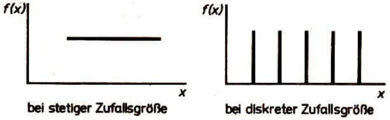


Bild 48

■ **B 35** Das Streckendiagramm nach Bild 15 (↗ S. 39) kommt bei 1000 Würfen einer Gleichverteilung nahe.

Links- und rechtsschiefe Verteilungen (↗ Bild 49)
Das sind eingipflige Verteilungen, bei denen der Gipfelwert links bzw. rechts von der Verteilungsmitte liegt.

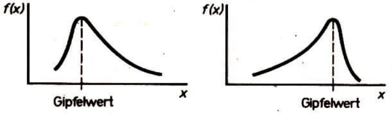


Bild 49

■ **B 36** Beispiele: Bild 14 (↗ S. 37); Bild 33 (↗ S. 53); Bild 40 (↗ S. 59)

J-förmige Verteilungen (↗ Bild 50)

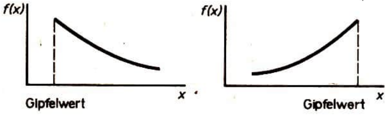


Bild 50

J-förmige Verteilungen erreichen ihren Gipfel an *einem* Ende der Verteilung.

Mehrgipflige Verteilungen (↗ Bild 51)
Es treten zwei oder mehr Gipfelpunkte auf.

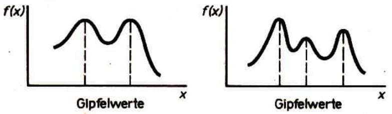


Bild 51

■ **B 37** Das Bild 32 (↗ S. 53) vermittelt eine zweigipflige Verteilung, wenn man den einen Ausreißerwert bei 4,85 mm außer acht läßt.

Beachten Sie: Bei Auftreten mehrgipfliger Verteilungen ist stets zu prüfen, ob es sich wirklich um eine mehrgipflige Verteilung handelt oder ob diese nur durch eine ungünstige Klasseneinteilung vorgetäuscht wird!

U-förmige Verteilungen (↗ Bild 52)
U-förmige Verteilungen sind zweigipflige Verteilungen mit den Gipfeln an *beiden* Enden der Verteilung.

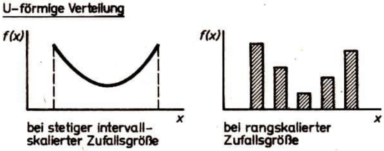


Bild 52

■ **B 38** Merkmal „Anzahl der Versäumnistage im Laufe eines Schuljahres" bei einer Gruppe von $n = 23$ Schülern

Klasse (Tage)	Häufigkeit h_k
1 bis 3	14
4 bis 6	3
7 bis 9	–
10 bis 12	2
13 und darüber	4
	$n = 23$

● **62** Stellen Sie die Häufigkeitsverteilung zur Zufallsgröße „Anzahl der Versäumnistage" gemäß der unter B 38 angegebenen Tabelle graphisch dar!

● **63*** Ein Schütze erzielt bei 50 Schuß auf die Zehnerscheibe folgende Treffer:

10	4	9	8	10	10	8	7	6	3
8	10	6	10	5	7	9	7	10	5
6	9	10	8	2	5	9	7	7	8
10	4	6	8	9	9	10	6	4	8
10	9	5	9	6	10	7	9	8	10

a) Stellen Sie die Häufigkeitsverteilung tabellarisch (ohne Klassenbildung) und graphisch dar!
b) Welche typische Verteilungsform liegt hier vor?

● **64*** Es liegen drei Häufigkeitstabellen vor. (Die x_k sind Klassenmitten.)

x_k	h_k
38,5	5
40,5	5
42,5	5
44,5	6
46,5	5
48,5	5
a)	31

x_k	h_k
0,07	11
0,10	9
0,13	6
0,16	4
0,19	3
0,22	1
b)	34

x_k	$\frac{h_k}{n}$
87	0,21
92	0,31
97	0,09
102	0,01
107	0,26
112	0,12
c)	1,00

Welchem Verteilungstyp gehören die drei empirischen Verteilungen jeweils an?

2.2.5. Summenverteilung und deren Darstellung

Häufigkeitsverteilungen für eine Zufallsgröße lassen noch eine ganz andere Art der Wiedergabe zu, und zwar als **Summenverteilung.** Der Vorteil der Summenverteilung besteht darin, daß man aus ihr direkt ablesen kann, wie viele Personen oder Objekte bezüglich der untersuchten Zufallsgröße *unterhalb* einer bestimmten Merkmalsausprägung liegen. Die Summenverteilung wird vorwiegend bei intervallskalierten Daten angewandt.

Wie gelangen wir zur Summenverteilung oder – wie man auch sagt – kumulativen Häufigkeitsverteilung? („kumulativ" heißt so viel wie „anhäufend" im Sinne von „schrittweise addierend" vom 1. Glied der Folge an.)

Wir betrachten dazu Zahlenfolgen und die jeweils entsprechende Summenfolge.

■ **B 39 a)** Zahlenfolge: 1 2 3 4 5 . . .
Summenfolge: 1 3 6 10 15 . . .
(entstanden aus: $1 + 2$ $3 + 3$ $6 + 4$ $10 + 5$ . . .)
Das k-te Glied der Summenfolge ergibt sich also als Summe des 1., 2., …, k-ten Gliedes der ursprünglichen Zahlenfolge.
$k = 4$ Das 4. Glied der Summenfolge ist 10, es entstand aus $1 + 2 + 3 + 4 = 10$.

b)

Zahlenfolge:	2	7	16	10	4	1
		+	+	+	+	+
Summenfolge:	2	9	25	35	39	40

Die Pfeile geben den Gang der Rechnung an.

● **65*** Bilden Sie die Summenfolge zur Zahlenfolge
a) 0,5 1 1,5 2 2,5 3 3,5;
b) 6 17 18 6 1 2 !

Wir gelangen von der empirischen Häufigkeitsverteilung zur Summenverteilung, indem wir den Begriff Summenfolge auf die Folge h_k der Häufigkeiten einer Verteilung anwenden.

Mit der Aufgabe 65b) haben Sie bereits absolute Summenhäufigkeiten berechnet, und zwar zu den Häufigkeiten, die wir bei Untersuchung der Zufallsgröße „Arbeitszeit je Steckermontage" für Lehrling A erhielten (↗ Tabelle S. 56).

► **E 19** Unter der **Summenhäufigkeit** sh_k verstehen wir die Summe der k Häufigkeiten einer Verteilung von der 1. bis zur k-ten.
Oder anders ausgedrückt:
Unter der Summenhäufigkeit verstehen wir die Summe der Häufigkeiten aller Werte, die kleiner oder gleich einer bestimmten Merkmalsausprägung (oder Merkmalsklasse) sind.

Aus der Folge der absoluten Häufigkeiten h_k	ergibt sich mittels Addition leicht die Folge der absoluten Summenhäufigkeiten
h_1	$sh_1 = h_1$
h_2	$sh_2 = h_1 + h_2$
h_3	$sh_3 = h_1 + h_2 + h_3$
…	…
h_m	$sh_m = h_1 + h_2 + h_3 + \ldots + h_m$

Nun ist der Begriff Summenverteilung ganz einfach zu verstehen.

▶ **D 8 Unter der *Summenverteilung* (auch: kumulativen Verteilung) verstehen wir die eindeutige Zuordnung von (absoluten oder relativen) Summenhäufigkeiten zu allen Ausprägungen oder Klassen des Merkmals (im betrachteten Intervall).**
Liegt Klasseneinteilung vor, so erfolgt die Zuordnung stets zur exakten oberen Klassengrenze der betreffenden Klasse.

Neben den absoluten Summenhäufigkeiten können wir also relative Summenhäufigkeiten berechnen, diese haben meist die größere Bedeutung.

■ **B 9 (7)** Zufallsgröße „Arbeitszeit je Steckermontage" für Lehrling A.
Tabellarische Darstellung der Häufigkeitsverteilung (aufgrund der Angaben aus der Tabelle auf Seite 56), ergänzt durch die absoluten und relativen Summenhäufigkeiten:

Klasse (s)	Klassenmitte x_k (s)	Absolute Häufigkeit h_k	Summenhäufigkeit sh_k	Relative Häufigkeit $\frac{h_k}{n}$	Summenhäufigkeit $\frac{sh_k}{n}$
40 bis 41	40,5	6	6	0,12	0,12
42 bis 43	42,5	17	23	0,34	0,46
44 bis 45	44,5	18	41	0,36	0,82
46 bis 47	46,5	6	47	0,12	0,94
48 bis 49	48,5	1	48	0,02	0,96
50 bis 51	50,5	2	50	0,04	1,00
Summe		50 ←	↑ Kontrolle	1,00 ←	↑ Kontrolle

Die Berechnung der relativen Summenhäufigkeiten kann auf zwei Wegen erfolgen:
Entweder wir summieren die relativen Häufigkeiten $\frac{h_k}{n}$ schrittweise auf oder wir berechnen zu den absoluten Summenhäufigkeiten jeder Merkmalsausprägung oder Klasse die entsprechende relative Häufigkeit.

Beachten Sie die in der Tabelle angegebenen *Kontrollmöglichkeiten:*
Die Summe der absoluten (bzw. relativen) Summenhäufigkeiten muß mit der absoluten (bzw. relativen) Summenhäufigkeit der letzten Klasse übereinstimmen; im ersteren Falle muß sich n ergeben (Umfang der Stichprobe), im zweiten Falle 1,00 (oder 100 % bei prozentualen Häufigkeiten)!

Mit Hilfe der Summenverteilung lassen sich Aufgaben der folgenden Art lösen:
- Wie viele Personen haben *weniger als x* Punkte erreicht?
- Für wie viele Steckermontagen wurden *weniger als x* Sekunden benötigt?
- Bei wie vielen Schweinen ist die Masse *geringer als x* Kilogramm?
- Bei wie vielen Ampelphasen passieren *weniger als x* Fahrzeuge die Kreuzung?

Allgemein: Wie viele Untersuchungsobjekte haben Merkmalswerte *kleiner als x* Einheiten?

Diese Fragestellung gilt bei Klasseneinteilung; dabei bezieht sich x stets auf die exakte obere Klassengrenze, also auf x_{og}, denn beim Aufsummieren werden ja sämtliche Werte der betreffenden Klasse erfaßt. Dieser Tatbestand kommt bereits im letzten Satz der Definition D 8 zum Ausdruck und ist besonders bei der graphischen Darstellung der Summenverteilung zu beachten.
Liegt keine Klasseneinteilung vor, so muß die allgemeine Frage so formuliert werden:

> Wie viele Untersuchungsobjekte haben Merkmalswerte, die *kleiner oder gleich* einem bestimmten Merkmalswert x sind?

■ **B 9 (8)** Zufallsgröße „Arbeitszeit je Steckermontage" für Lehrling A – benötigte Daten in der Tabelle des Beispiels 9 (7), Seite 67.
Für wie viele Steckermontagen wurden vom Lehrling A weniger als 45,5 s benötigt?
Antwort: Für 41 Steckermontagen, das entspricht 82 %.
Für wie viele Steckermontagen wurden vom Lehrling A weniger als oder gleich 50 s benötigt?
Diese Frage kann man anhand der Tabelle nicht beantworten, da „50" keine exakte Klassengrenze ist. Hier muß man auf die Häufigkeitsverteilung ohne Klassenbildung zurückgreifen (also auf die Tabelle der Seite 31 oder auf die Tabelle der Seite 55) und dort die Summenhäufigkeiten berechnen.
Antwort: Für 49 Steckermontagen, das entspricht 98 %.

Nun zur graphischen Darstellung der Summenverteilung.
Dazu dient bei *stetigen Merkmalen* mit *intervallskalierten* Daten die **Summenkurve.**

► **E 20** Die **Summenkurve** (oder der Summenpolygonzug) ist die graphische Darstellung der Summenverteilung für stetige Merkmale mit Meßwerten. Sie entsteht durch Verbinden der Endpunkte der an den exakten oberen Klassengrenzen abgetragenen Summenhäufigkeiten.

Summenkurve zum Beispiel 9 (9)

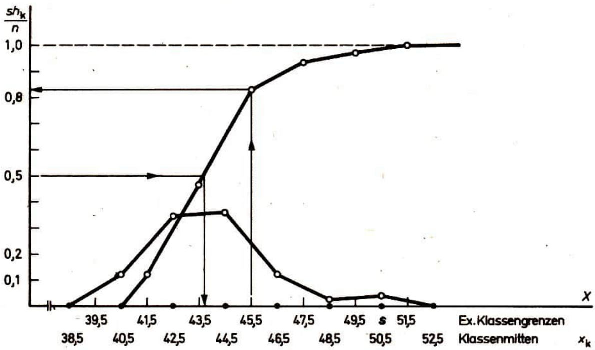


Bild 53

Die Summenkurve wird meist unter Verwendung der relativen Summenhäufigkeiten gezeichnet. Sie hat einen Verlauf, der einem stilisierten S ähnelt und erreicht stets die Parallele zur Merkmalsachse im Abstand 1,00 bzw. n (↗ Bild 53 zum Beispiel 9 (9), aufstrebende Kurve mit dem Maximalwert 1,0).

■ **B 9 (9)** Zufallsgröße X „Arbeitszeit je Steckermontage für Lehrling A" (Tabellarische Darstellung der Summenverteilung in der Tabelle auf Seite 67)

Beachten Sie die Lage der Punkte der Summenkurve zu den Klassen des untersuchten Merkmals! Diese Kurvenpunkte liegen über den *exakten Klassengrenzen*. Zusätzlich wurde in Bild 53 der Linienzug (für den gleichen Sachverhalt) eingezeichnet. Hierbei liegen die Kurvenpunkte über den *Klassenmitten*.

Anhand der graphischen Darstellung kann man die gleichen und ähnliche Fragen beantworten wie oben (S. 67) mittels der tabellarischen, z. B.:

– Bei wie vielen Steckermontagen wurden vom Lehrling A weniger als 45,5 s benötigt?

Zur Beantwortung dieser Frage mit Hilfe des Diagramms errichten wir im betreffenden Merkmalswert (hier 45,5 s) die Senkrechte bis zum Schnitt mit der Summenkurve. Durch diesen Schnittpunkt ziehen wir eine Parallele zur Merkmalsachse und lesen auf der Ordinatenachse die relative Summenhäufigkeit $\frac{sh_k}{n}$ ab.

Im Beispiel sind es rund 0,82 (↗ Bild 53). Antwort also: Bei 82% der Steckermontagen wurden weniger als 45,5 s benötigt.
Aber auch Fragen der folgenden Art lassen sich mit Hilfe der Summenkurve beantworten:

– Welche Arbeitszeit je Steckermontage maximal benötigt der Lehrling A für 50% aller Steckermontagen?

Auch das können wir aus dem Diagramm der Summenverteilung im Bild 53 leicht ablesen. Wir suchen 50% ≙ 0,5 auf der Ordinatenachse auf, ziehen die Parallele zur Merkmalsachse bis zum Schnittpunkt mit der Summenkurve. Von diesem fällen wir das Lot auf die X-Achse und lesen dort ab: rund 43,7 s.
Der Vorteil des Diagramms gegenüber der Tabelle liegt darin, daß wir aus der Zeichnung auch Zwischenwerte entnehmen können, was mit Hilfe der Tabelle nicht möglich ist.

Die Summenverteilung findet eine weitere wichtige Anwendung beim **Vergleich zweier oder mehrerer Gruppen** von Untersuchungsobjekten bezüglich der zum gleichen Merkmal gewonnenen Beobachtungsdaten. So könnten wir – bei Personen z. B. – am Vergleich zwischen einer *Versuchsgruppe* (sie wurde nach einem effektiven Verfahren unterrichtet oder mit einem bestimmten Medikament behandelt) und einer *Kontrollgruppe* (keine pädagogische oder medizinische Einwirkung) interessiert sein. Statistische Prüfverfahren erlauben uns gesicherte Aussagen darüber, ob zwischen den beiden Gruppen ein Unterschied besteht oder nicht.
Wir wollen die Leistungen der Lehrlinge A und B bezüglich des untersuchten Merkmals „Arbeitszeit je Steckermontage" miteinander vergleichen (hier sind bekanntlich die Steckermontagen die Untersuchungsobjekte), indem wir beide Summenverteilungen in *einem* Bild graphisch darstellen.

■ **B 9 (10)** Vergleich der Zufallsgrößen X und Y „Arbeitszeit je Steckermontage für Lehrling A bzw. B" mittels Summenkurven
Die Summenkurve zu X finden wir im Bild 53. Die Summenverteilung zu Y (Lehrling B) entwickeln wir aus der Tabelle auf Seite 56:

Klasse (s)	Klassen-mitte $y_k(s)$	Summenhäufigkeit absolute sh_k	relative $\frac{sh_k}{n}$
40 bis 41	40,5	11	0,22
42 bis 43	42,5	33	0,66
44 bis 45	44,5	48	0,96
46 bis 47	46,5	50	1,00
48 bis 49	48,5	50	1,00
50 bis 51	50,5	50	1,00

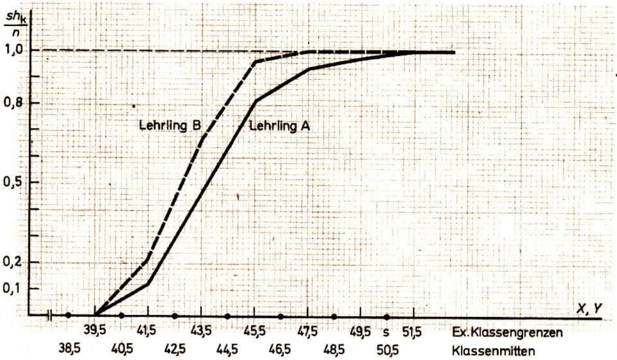


Bild 54

● **66** Vergleichen Sie dieses Diagramm mit dem Bild 37 (↗ S. 57)!
Dort kommt der gleiche Sachverhalt wie hier im Bild 54 zum Ausdruck, nur daß die Darstellung von X und Y im Bild 37 als Linienzug erfolgte.

Was können wir nun dem Bild 54 entnehmen?
Zunächst lassen sich die gleichen oder ähnliche Fragen, wie wir sie auf den Seiten 68 und 69 für die Verteilung von X stellten, jetzt auf beide Verteilungen anwenden. Daraus ergeben sich Aussagen über die Leistung der Lehrlinge A und B in bezug auf ganz bestimmte Montagezeiten.
Die Tatsache, daß die Summenkurve für Lehrling B vollständig links von der des Lehrlings A verläuft, läßt *hier* (die *besseren* Montagezeiten liegen auf der Merkmalsachse *links*) darauf schließen, daß Lehrling B bezüglich der beobachteten Montagezeit bessere Leistungen aufweist als Lehrling A. Eine solche Aussage dürfen wir aber zunächst nur als Hypothese (Vermutung) aufstellen. Es kann sich ja bei den Unterschieden zwischen den beiden Lehrlingen um zufällige Abweichungen handeln, ein wirkliches „Bessersein“ der Montageleistung des Lehrlings B muß damit nicht vorliegen.
Beachten Sie: Erst die Anwendung eines geeigneten statistischen Prüfverfahrens kann die Entscheidung darüber herbeiführen, ob wir es mit zufälligen Abweichungen oder mit einem statistisch *gesicherten Unterschied* zu tun haben, ob also B (bezüglich des untersuchten Merkmals) wirklich besser ist als A!
Beachten Sie ferner: Befinden sich die *besseren* Leistungen auf der Merkmalsachse

rechts (wie es im allgemeinen üblich ist), dann weist die weiter rechts verlaufende Summenkurve auf die bessere Leistungstendenz (der untersuchten Person, Gruppe oder des untersuchten Objektbereichs) hin!
Überschneiden die beiden Summenkurven einander, dann liegen diesem Sachverhalt ganz unterschiedliche Verteilungsformen oder Streuungen der empirischen Häufigkeitsverteilungen zugrunde.
Werfen wir noch einen kurzen Blick auf die *theoretische Summenverteilung* $F(x)$ zur standardisierten Normalverteilung. Sie heißt Gaussche **Summenkurve** oder **Ogive** und hat diesen Verlauf (↗ Bild 55).
Einige wichtige Eigenschaften der Ogive sind:

- $F(\to -\infty) = 0$
- $F(\to \infty) = 1$
- Ist $x_1 < x_2$, so ist $F(x_1) < F(x_2)$, d. h., die Kurve wächst streng monoton (steigt in ihrem ganzen Verlauf an).

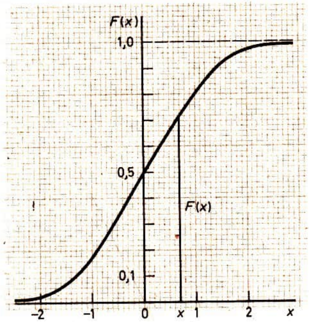


Bild 55 GAUSSsche Summenkurve

- Die Kurve weist keine Sprünge oder Unterbrechungen auf.

Die wichtigste Eigenschaft aber ist diese (sie gilt allgemein für stetige Zufallsgrößen):

- Der Zahlenwert der *Fläche unter der standardisierten Normalkurve* (Bild 44, ↗ S. 61) zwischen $-\infty$ und einem bestimmten Wert x der Zufallsgröße *ist gleich* dem Zahlenwert *der Ordinate der Gaussschen Summenkurve* an der Stelle x.

■ **B 40** Die Fläche unter der standardisierten Normalkurve zwischen $-\infty$ und $x = +1$ beträgt $F(1) = 0{,}841$ (nach der 3. Zeile der Wertetabelle auf Seite 61). Die Ordinate der Gaussschen Summenkurve an der Stelle $x = +1$ beträgt $F(1) = 0{,}84$ (als Ablesung im Bild 55).

● **67*** **a)** Lesen Sie $F(2)$ und $F(0)$ aus Bild 55 ab!
Vergleichen Sie Ihre Ablesewerte mit den Tafelwerten für $F(2)$ und $F(0)$ in der Wertetabelle auf Seite 61!
b) Welche Doppelbedeutung kommt dem Ausdruck $F(-1{,}96) = 0{,}025$ zu?

In der Praxis steht man oft vor der Frage, ob die aus einer Untersuchung gewonnene empirische Verteilung noch als näherungsweise normalverteilt anzusehen ist. Zur Lösung dieser Aufgabe steht ein zeichnerisches Verfahren zur Verfügung, das sich die Eigenschaften der Gaussschen Summenkurve zunutze macht (↗ [24]).

Auch bei *diskreten intervallskalierten* und bei *rangskalierten Merkmalen* ist die Bildung der Summenverteilung möglich.

■ **B 10 (4)** Summenverteilung zur Untersuchung der Zufallsgröße „Anzahl der Fahrzeuge je Ampelphase" (nach Klassenbildung)
Wir erweitern die Tabelle auf der Seite 57f. um die absoluten und relativen Summenhäufigkeiten.

Klasse (Fahrzeuge/ Ampelphase)	Klassen-mitte z_k (Fahrz./Aph.)	Absolute Häufig-keit h_k	Summen-häufigkeit sh_k	Relative Häufig-keit $\frac{h_k}{n}$	Summen-häufigkeit $\frac{sh_k}{n}$
7 bis 9	8	4	4	0,133	0,133
10 bis 12	11	4	8	0,133	0,266
13 bis 15	14	9	17	0,300	0,566
16 bis 18	17	6	23	0,200	0,766
19 bis 21	20	5	28	0,167	0,933
22 bis 24	23	1	29	0,033	0,966
25 bis 27	26	1	30	0,033	0,999
Summe		30 ←	↑ Kontr.	0,999 ←	↑ Kontr.

Die graphische Darstellung der Summenverteilung kann hier jedoch nicht mit der Summenkurve erfolgen (da ja Zwischenwerte zwischen den diskreten Merkmalsausprägungen gar nicht existieren), sondern nur mittels **Treppenpolygon.**

► **E 21** Das **Treppenpolygon** ist die graphische Darstellung der Summenverteilung für diskrete Merkmale, bei denen Meßwerte vorliegen, und für rangskalierte Merkmale. Es entsteht, indem wir über den Merkmalsausprägungen (oder Klassenmitten) die zugehörige Summenhäufigkeit abtragen und von dem erhaltenen Punkt parallel zur Merkmalsachse um eine Ausprägung (bzw. eine Klassenbreite) nach rechts gehen (↗ Bild 56).

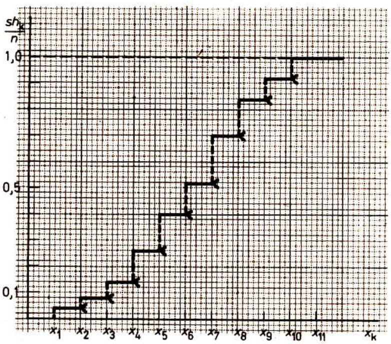


Bild 56

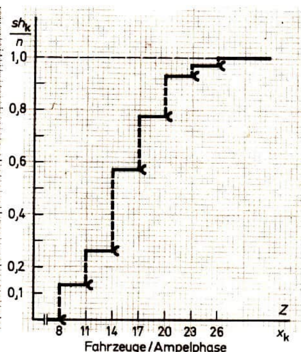


Bild 57

Anmerkung: „—c“ ist die graphische Kennzeichnung eines halboffenen Intervalls, also z. B. von $x_1 \leq x < x_2$.

■ **B 10 (5)** Treppenpolygon zum diskreten Merkmal Z „Anzahl der Fahrzeuge je Ampelphase" (nach Klassenbildung) (↗ Bild 57)

■ **B 10 (6)** Bei wie vielen Ampelphasen passieren weniger als 20 Fahrzeuge je Ampelphase die Kreuzung?
Ablesung aus dem Bild 57: Bei etwa 75% aller beobachteten Ampelphasen (genau sind es 76,6%, ↗ Tabelle auf S. 72).
Beachten Sie: Diese Aussage bezieht sich auf die Summenverteilung nach Klassenbildung! Bezogen auf die beobachtete (ungruppierte) Verteilung sind es 24 von 30, also 80%.

● **68*** Untersuchung zur Zufallsgröße „Leistung im Mathematik-Olympiade-Wettbewerb" (schließt an die Aufgabe 55, ↗ S. 58, mit Darstellung der Häufigkeitstabelle nach Klassenbildung, ↗ S. 128, an)

a) Stellen Sie die Summenverteilung (mit absoluten und relativen Summenhäufigkeiten) tabellarisch dar! Gestalten Sie die Kopfzeile der Tabelle analog zur Tabelle auf der Seite 67!

b) Stellen Sie die (absolute) Summenverteilung graphisch (auf Millimeterpapier) dar! Achten Sie auf die Zuordnung der Summenhäufigkeiten zu den exakten oberen Klassengrenzen!

c) Lesen Sie aus Ihrer Zeichnung ab: Wie viele Schüler erzielten weniger als 25 Punkte?

● **69*** Untersuchung zur Zufallsgröße „Masse von Schweinen" (kg, Fortsetzung der Aufgabe 22; ↗ S. 21 f. und 56, ↗ S. 58). Zu beachten ist besonders die Tabelle im Lösungsteil zur Aufgabe 56, S. 129.

a) Ist hier für die graphische Darstellung der Summenverteilung die Summenkurve oder das Treppenpolygon richtig? Begründen Sie Ihre Antwort!

b) Geben Sie die tabellarische und graphische Darstellung der Summenverteilung (mit absoluten und prozentualen Summenhäufigkeiten) an!

c) Wie viele Schweine haben geringere Masse als 151 kg?

● **70*** Zufallsgröße U „Güte eines Erzeugnisses", erfaßt in Güteklassen (B 24, S. 40): Stellen Sie die Summenverteilung (mit absoluten und relativen Summenhäufigkeiten) tabellarisch und graphisch dar!

● **71*** Beantworten Sie folgende Frage anhand des Bildes 54 (↗ S. 70): Bei wie vielen Steckermontagen wurden vom Lehrling A und vom Lehrling B weniger als 43 s benötigt?

2.2.6. Weitere Arten der graphischen Darstellung statistischer Sachverhalte

Wir haben uns bisher jeweils nur mit *einer* Zufallsgröße beschäftigt. Beobachten wir an den Untersuchungsobjekten das Auftreten *zweier* Merkmale und halten die entsprechenden Häufigkeiten fest, so entsteht eine *bivariable Häufigkeitsverteilung*.
Zur graphischen Darstellung einer bivariablen Häufigkeitsverteilung benötigen wir drei Dimensionen, und zwar für die Merkmale X (Merkmalswerte x_j) und Y (Merkmalswerte y_k), an den gleichen Untersuchungsobjekten ermittelt, sowie für die Häufigkeit h_{jk} des gemeinsamen Auftretens von x_j und y_k.
Dreidimensionale Sachverhalte erfordern drei Koordinatenachsen, zwei Merkmals- und eine Häufigkeitsachse. Die so entstehenden graphischen Darstellungen heißen **Stereogramme**.

Deren Abbildung in die Ebene ist mit Schwierigkeiten verbunden. Zum einen ist das Anfertigen von Stereogrammen im Zweidimensionalen zeitaufwendig, zum anderen kann es beträchtliche Informationsverluste mit sich bringen.
Um diesen Schwierigkeiten zu entgehen, bedient man sich zur Darstellung bivariabler Verteilungen besser der **Punktwolke.** Dabei werden die Häufigkeiten nicht als dritte Achse gezeichnet, sondern einfach durch Punkte und Striche im *XY*-Koordinatensystem kenntlich gemacht (↗ Bild 58).
Diese Darstellungsart – sie kann für stetige wie diskrete Merkmale und für jede Datenart angewandt werden – macht Zusammenhänge zwischen zwei Zufallsgrößen deutlich. Bei der graphischen Darstellung einer bivariablen Häufigkeitsverteilung ist darauf zu achten, daß die besseren Leistungen auf den Merkmalsachsen *nach rechts und oben* anzuordnen sind.

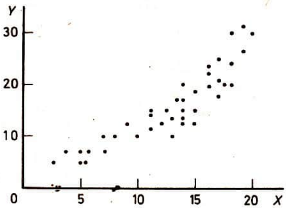


Bild 58

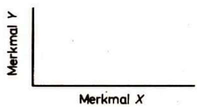


Bild 59

● **72*** Zeichnen Sie (auf Millimeterpapier) die Punktwolke für folgende Meßwertpaare von zwölf 13jährigen Schulkindern!

Schulkind (Versuchsperson)	A	B	C	D	E	F	G	H	I	K	L	M
Körpergröße in cm (Merkmal *X*)	160	172	155	165	162	168	161	169	170	173	160	157
Masse in kg (Merkmal *Y*)	46	55	38	51	45	53	39	45	50	58	48	43

Bisher hatten wir es stets mit Häufigkeitsverteilungen zu tun. Bei der statistischen Analyse von Sachverhalten (Erscheinungen, Zusammenhängen, Abläufen) begegnen uns öfter Fälle, in denen zwei oder mehrere Merkmale *miteinander in Beziehung gesetzt werden, ohne daß die Häufigkeit* des gemeinsamen Auftretens von Ausprägungen der Merkmale untersucht wird.

■ **B 41** Betrachten wir das Bild 2 (↗ S. 9)! Hier sind zwei Merkmale in Beziehung zueinander gesetzt worden. Eine Häufigkeitsverteilung liegt nicht vor.

● **73*** Um welche beiden Merkmale handelt es sich im Bild 2?

Die graphische Darstellung zweier Zufallsgrößen ohne Erfassung von Häufigkeiten für die einzelnen Ausprägungen ist durch den im Bild 59 dargestellten Haupttyp gekennzeichnet. Die Darstellung erfolgt dabei im rechtwinkligen Koordinatensystem. Die auf den Achsen einzutragenden Merkmalsausprägungen können wieder Meßwerte, Meßwertklassen, Rangdaten oder Kategorien sein. Den geordneten Paaren $(x_j;\ y_k)$ werden aber nicht Häufigkeiten zugeordnet (wie es bei der Punktwolke geschieht), sondern sie werden direkt als Punkte im *XY*-Koordinatensystem abgebildet. So entstehen Kurven, Linienzüge oder Streifendiagramme, denn je nach Datenart ist das entsprechende Darstellungsmittel zu wählen.

■ **B 42** Bei Zuordnung von Meßwerten zu Meßwerten ist die Kurve (oder der Linienzug) das richtige Darstellungsmittel (↗ Bild 60a), bei Zuordnung von Meßwertklassen zu Meßwertklassen der Linienzug,

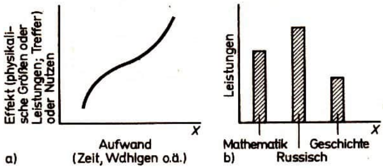


Bild 60

bei Zuordnung von Kategorien zu Meßwerten, Meßwertklassen oder Rangdaten (oder auch umgekehrt) das Streifendiagramm (↗ Bild 60b).

● **74*** Zu Bild 2 (↗ S. 9):
Welche Datenarten kommen den beiden hier beteiligten Merkmalen zu?
Ist das Streifendiagramm die richtige Art der graphischen Darstellung des Sachverhalts?

Wenn es sich um Entwicklungen über Zeiträume hinweg – über Wochen, Monate oder Jahre – handelt, wird die X-Achse als Zeitachse verwendet. Wir sprechen dann von **Entwicklungs-** oder **Zeitreihen.**

▶ **E 22** Eine **Zeitreihe** ist eine Folge von Merkmalswerten, die man zu bestimmten aufeinanderfolgenden Zeitpunkten oder Zeitabschnitten für ein in der Zeit veränderliches Merkmal erhält.

■ **B 43** Zeitreihe: Entwicklung des Agrarflugs der DDR am Beispiel der Düngung der Felder mit Stickstoff aus der Luft

Jahr	1974	1976	1978	1980
Jahresleistung in Mill. Hektar	1,19	1,82	2,25	2,39

Die Darstellung einer Zeitreihe erfolgt mit Hilfe einer Tabelle (↗ Beispiel 43) oder graphisch durch eine sogenannte **Entwicklungskurve.**

▶ **E 23** Die **Entwicklungskurve** ist die graphische Darstellung einer Zeitreihe. Die Abszissenachse ist die Zeitachse, die Ordinatenachse trägt die Ausprägungen des sich in der Zeit verändernden Merkmals (↗ Bild 61).

■ **B 43 (2)** Das Bild 62 zeigt die Entwicklungskurve zur Tabelle im Beispiel 43.

Die Darstellung der Entwicklungskurve kann als Linienzug (wie im Bild 61) oder als Streifendiagramm erfolgen. Zuweilen findet man beides nebeneinander, das Bild 63 zeigt es (Rückgang von Kinder-Infektionskrankheiten in der DDR).

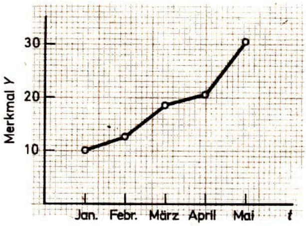


Bild 61

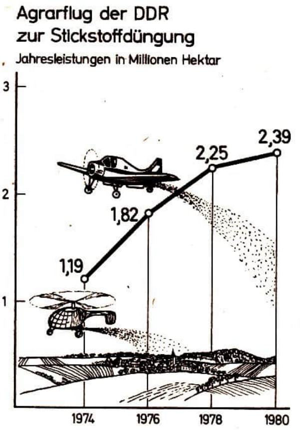


Bild 62

Der Linienzug ist vorteilhafter, wenn die statistischen Daten zeitlich lückenlos zu bestimmten Zeitpunkten oder für bestimmte Zeitabschnitte (z. B. Quartale, Jahre oder Fünfjahrplanzeiträume) erfaßt worden sind; das Streifendiagramm dagegen, wenn die Daten nicht regelmäßig, sondern nur in gewissen Zeitabständen erhoben wurden.

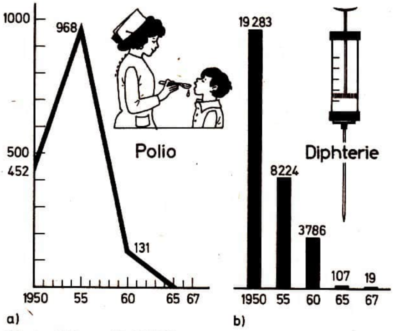


Bild 63 (Nach: LVZ vom 14. 6. 1969)

● **75** Zeitreihe für die Entwicklung der aktiven Tuberkulose in der DDR

Jahr	1950	1955	1960	1965	1970	1975	1980
Krankheitsfälle je Jahr	92760	47016	23418	15933	10306	6163	3962

a) Stellen Sie die Zeitreihe graphisch dar!
b) Wie würden Sie die Entwicklung charakterisieren (beschreiben)?

Entwicklungskurven sind durch einen **Trend** gekennzeichnet.

> **E 24** Der **Trend** ist die systematische Tendenz (Entwicklungs*richtung*) einer Zeitreihe.
> Der Trend kann durch eine besondere Kurvenform charakterisiert werden (↗ Bild 64).

Wichtig ist zu verstehen, daß der Trend die *systematische* Entwicklungsrichtung kennzeichnet; von zufälligen Abweichungen oder Schwankungen der Erscheinung wird abgesehen. Der tatsächliche Verlauf kann also mehr oder weniger vom Trend abweichen.
Wir unterscheiden mehrere typische **Trendformen** (↗ Bild 65).

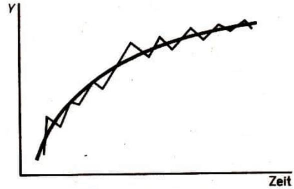


Bild 64 Tatsächlicher Verlauf (~) und Trend (⌒) einer Entwicklung

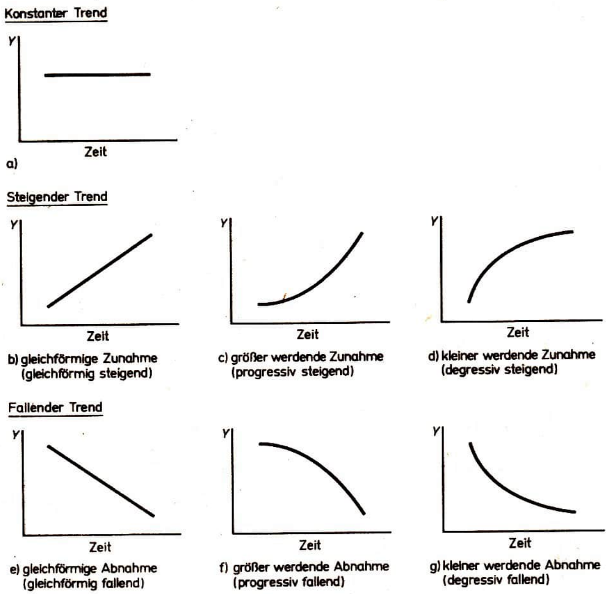


Bild 65

Von diesen Trendformen spricht man auch dann, wenn die graphische Darstellung der Zeitreihe als Streifendiagramm erfolgte.

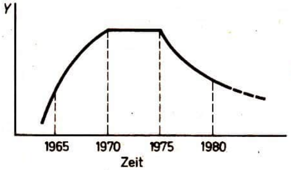
● 76* a) Kennzeichnen Sie den Trend zur Entwicklung der Diphterie in der DDR von 1950 bis 1967 (Bild 63b, S. 76)!

b) Ist bei der Analyse von Trends in Natur und Gesellschaft die Wertung richtig: Steigende Trendformen sind stets Ausdruck einer positiven Entwicklung?

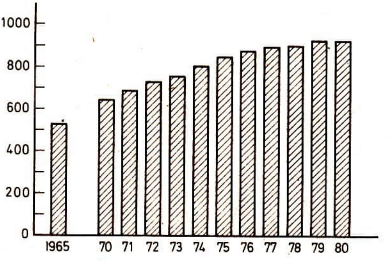
Die Kenntnis des Trends bietet die Möglichkeit, begründete Voraussagen für die untersuchte Erscheinung zu treffen. Damit bestätigt sich der große Nutzen der Statistik als wichtiges Arbeitsinstrument für die Leitung, Planung und Forschung. Das gilt für viele Bereiche der Volkswirtschaft, der Wissenschaft und der allgemeinen gesellschaftlichen Entwicklung.
Man darf aber nicht dem Fehler verfallen anzunehmen, eine bestimmte Trendform, die für relativ große Zeitspannen (z. B. für drei Fünfjahrpläne) Gültigkeit hatte, setze sich stets in der gleichen Weise fort. Gerade im Zusammenhang mit der immer schwieriger werdenden Rohstoffgewinnung müssen bei Trendberechnungen (die wir hier nicht behandeln) sowohl weltweit wirkende Einflüsse als auch Besonderheiten der Entwicklung in dem einen oder anderen Land berücksichtigt werden. Es wird also häufig notwendig sein, die Entwicklung einer Erscheinung durch mehrere Trendformen zu charakterisieren.

● 77* Die Entwicklung eines Merkmals Y soll durch den im Bild 66 dargestellten Verlauf gekennzeichnet sein. Welche Trendformen treten dabei auf?

Bild 66



Betreute Kinder in Einrichtungen der Vorschulerziehung je 1000 Kinder im Kindergartenalter



(Nach: Statistisches Jahrbuch der DDR 1981) Bild 67

● 78 Kennzeichnen Sie die Trendformen

a) zur Entwicklung des Schiffsbestandes der Handelsflotte der DDR von 1952 bis 1980 (↗ Tabelle S. 10);

b) zur im Bild 67 dargestellten Entwicklung!

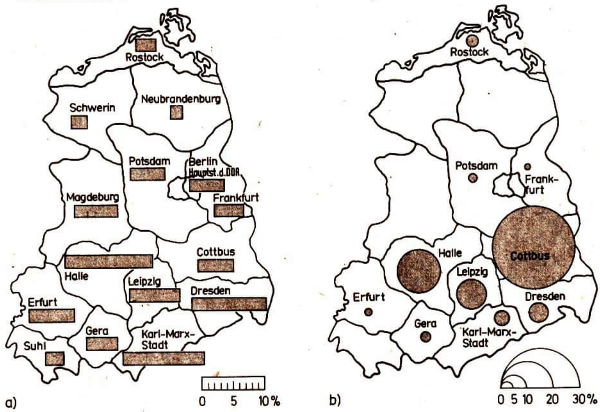
Am Schluß dieses Abschnittes wollen wir noch zwei Arten der graphischen Darstellung statistischer Sachverhalte kennenlernen, die von den bis jetzt behandelten erheblich abweichen, nämlich das **Kartogramm** und die **Bevölkerungspyramide.**

► **E 25** Das **Kartogramm** ist eine geographische Karte, in die für bestimmte geographische Regionen ermittelte statistische Daten eingetragen sind. Die Eintragungen erfolgen durch geometrische Gebilde (wie Quadrate, Kreise, Quader usw.), durch Symbole, durch Flächenschraffur oder unterschiedliche Farbgebung.

■ **B 44** Das Bild 68 zeigt zwei Kartogramme.
Hier sind die statistischen Angaben (in Prozent) durch Rechtecke bzw. Kreise dargestellt. Anhand eines jeweils beigefügten Maßstabs kann der prozentuale Anteil jedes Bezirks an der Gesamtindustrie bzw. Energie- und Brennstoffindustrie abgelesen werden. Unterschiede zwischen den Bezirken werden durch eine solche Art der Darstellung besonders deutlich.

Anteil der Bezirke an der industriellen Bruttoproduktion der DDR 1980

Gesamte Industrie — Energie- und Brennstoffindustrie



(Nach: Statistisches Jahrbuch der DDR 1981) **Bild 68**

● **79** Lesen Sie aus den Bildern 68a) und b) ab!

a) Wieviel Prozent der gesamten industriellen Bruttoproduktion der DDR erbrachte 1980 die Hauptstadt der DDR, Berlin, und wieviel Prozent der Bezirk Halle?

b) Mit welchem prozentualen Anteil an der Energie- und Brennstoffindustrie der DDR war 1980 der Bezirk Cottbus beteiligt und mit welchem Anteil der Bezirk Halle?

c) Stellen Sie (aus Bild 68b)) die Rangfolge des prozentualen Anteils der Bezirke an der Energie- und Brennstoffindustrie der DDR im Jahre 1980 auf! Beginnen Sie mit dem Bezirk mit dem stärksten Anteil an dieser Produktion, und führen Sie am Schluß auch die Bezirke auf, von denen kein (nennenswerter) Anteil erbracht wird!

Das Bild auf der 2. Umschlagseite zeigt ein Kartogramm bezüglich des untersuchten Merkmals „Bevölkerungszahl von Teilen der Erde 1960 und 1980". Die Quadrate stehen jeweils für 10 Millionen Menschen; die blauen Quadrate gelten für das Jahr 1960, die roten repräsentieren die Zunahme der Bevölkerung bis 1980.

● **80*** **a)** In welchem dauerbewohnten Erdteil ist in den 20 Jahren von 1960 bis 1980 der größte Bevölkerungszuwachs (absolut) zu verzeichnen, in welchem Erdteil der geringste?

b) In welchen drei Teilen der Erde ist in der Zeit von 1960 bis 1980 der größte prozentuale Zuwachs relativ zur 1960 errechneten Bevölkerungszahl eingetreten?

Die **Bevölkerungspyramide** oder **Alterspyramide** ist keine Pyramide im mathematischen Sinne des Wortes, sondern ein Flächendiagramm, das den Altersaufbau der Bevölkerung eines Landes wiedergibt, unterschieden nach den Geschlechtern, oft auch unter Einbeziehung des Familienstandes. Diese Darstellungsart vermittelt einen guten Einblick in die Bevölkerungsstruktur eines Landes oder einer Region und ist deshalb ein wichtiges Hilfsmittel der Bevölkerungsstatistik.
An einer solchen graphischen Darstellung ist ablesbar, wie viele Personen (männlichen und weiblichen Geschlechts) eines bestimmten Alters in dem untersuchten Lande leben. Personenverluste durch Kriegseinwirkungen und irgendwie bedingte Rückgänge in der Zahl der Geburten spiegeln sich in der Bevölkerungspyramide wider.

■ **B 45** Das Bild auf der 3. Umschlagsseite zeigt die Bevölkerungs- oder Alterspyramide von 1979 für die DDR-Bevölkerung.

● **81*** **a)** Was bedeuten die Einschnürungen bei etwa 30 und 60 Jahren, und worauf sind sie zurückzuführen?
b) Woraus begründet sich der symmetrische Aufbau der Pyramide zwischen 0 und 30 Altersjahren?

Die Alterspyramide setzt sich eigentlich aus mehreren Häufigkeitsverteilungen zur Zufallsgröße „Alter" zusammen.
Im Gegensatz zu den im Abschnitt 2.2.2. behandelten Häufigkeitsverteilungen eines Merkmals ist hier jedoch das Merkmal „Alter" auf der Ordinatenachse aufgetragen und die Häufigkeit der Männer oder Frauen oder verwitweten Frauen usw. auf der Abszissenachse, und zwar so, daß von der Mitte, von Null aus nach rechts *und* links, positive Werte abgetragen sind, nach links für den männlichen Bevölkerungsteil, nach rechts für den weiblichen.

Aufgrund dieses Maßstabs läßt sich ablesen, wie viele Männer (oder Frauen) eines bestimmten Alters es gibt oder wie groß die Anzahl der ledigen Frauen (oder verwitweten Männer oder ...) in einem bestimmten Alter ist.

■ **B 46** Anzahl der Frauen im Alter von 70 Jahren: über 110000
Anzahl der Männer im Alter von 70 Jahren: 62000
Anzahl der Witwen im Alter von 70 Jahren: 55000
Anzahl der Witwer im Alter von 70 Jahren: etwa 8000
(jeweils bezogen auf die DDR-Bevölkerung im Jahre 1979)

● **82** **a)** Etwa wie viele Frauen des Geburtsjahrganges 1900 gab es 1979 in der DDR? Wie viele davon waren verwitwet?
b) Wie groß war 1979 die Anzahl der Männer vom Geburtsjahr 1917?

Zusammenfassung des Abschnitts 2.2.

> Die **Datenerfassung** dient der Gewinnung von Informationen in Form von Zahlen, aber auch von Buchstaben und Worten, zur späteren Auswertung.
> Als unmittelbares Ergebnis der Datenerfassung entsteht ein **Datenträger**, dieser kann nicht-maschinenlesbar (Urliste, Fragebogen, Versuchsprotokoll) oder maschinenlesbar (Lochkarte, Magnetband) sein.
> Die **Datenaufbereitung** ist das Ordnen und Verdichten der Daten in Form von Tabellen. Ein wesentliches Ziel der Aufbereitung kann darin bestehen, das **Verteilungsgesetz** des untersuchten Merkmals aufzudecken.

Zunächst liegen jedoch **Originaldaten** (auch: Urdaten) im **Urbeleg** vor, sodann erfolgt die Ordnung der Daten nach Merkmalsausprägungen. Über die Strichliste gelangen wir zur (empirischen) **Häufigkeitsverteilung**, das ist die eindeutige Zuordnung von (absoluten oder relativen) Häufigkeiten zu allen bei der betreffenden Untersuchung in Betracht kommenden Merkmalsausprägungen. Die Häufigkeitsverteilung kann als **Tabelle** oder **graphisch** dargestellt werden (↗ Bild 28, S. 46).

Um die Übersichtlichkeit und Aussagekraft der Häufigkeitsverteilung zu erhöhen, bedient man sich der **Klassenbildung**, das heißt der Zusammenfassung zweier oder mehrerer benachbarter Merkmalsausprägungen zu Gruppen oder Klassen. Dabei gilt es, die Begriffe **Klassengrenzen, exakte Klassengrenzen, Klassenmitte, Klassenbreite, Anfangswert der ersten Klasse** (↗ Bild 30, S. 49) voneinander zu unterscheiden und die Anhaltspunkte für eine gute Klassenbildung (S. 54) zu berücksichtigen. Auch für die Häufigkeitsverteilung *nach* Klassenbildung sind alle möglichen Formen der graphischen Darstellung anwendbar, natürlich immer in Abhängigkeit von der Art des untersuchten Merkmals (stetig/diskret) und der vorliegenden Daten (intervall-, rang-, nominalskaliert).

Von herausragender Bedeutung ist die **Normalverteilung.** Sie ist die **theoretische Verteilung** einer stetigen Zufallsgröße und entsteht stets dann, wenn vier Bedingungen gleichzeitig wirken:

- große Anzahl von Beobachtungswerten;
- Zufallsauswahl der Untersuchungsobjekte;
- Zusammenspiel einer großen Zahl voneinander unabhängiger und etwa gleich stark wirksamer Faktoren;
- annähernde Übereinstimmung der relativen Häufigkeit für das Auftreten gegensinniger Faktoren.

Normalverteilung von Zufallsgrößen ist als Erscheinungsform einer statistischen Gesetzmäßigkeit in Natur und Gesellschaft häufig anzutreffen.

Die graphische Darstellung dazu heißt **Normal-** oder **Glockenkurve.** Die **standardisierte Normalverteilung** ist gekennzeichnet durch Mittelwert $\mu = 0$ und Streuung $\sigma = 1$. Teilflächen unter der standardisierten Normalkurve zwischen x_1 und x_2 geben die Wahrscheinlichkeit an, mit der die Zufallsgröße X einen Wert zwischen x_1 und x_2 annimmt, die Gesamtfläche unter der Normalkurve beträgt 1.

Neben der Normalverteilung existieren zahlreiche andere **typische Formen von Häufigkeitsverteilungen** (↗ Bilder 48 bis 52, S. 64f.).

Die **Summenverteilung** (oder kumulative Verteilung) ist die eindeutige Zuordnung von (absoluten oder relativen) Summenhäufigkeiten zu allen Ausprägungen oder Klassen des untersuchten Merkmals. Die graphische Darstellung der Summenverteilung für stetige Merkmale mit intervallskalierten Daten ist die **Summenkurve** (auch: Summenpolygonzug, bei Normalverteilung Ogive), für diskrete intervallskalierte und allgemein für rangskalierte Merkmale das **Treppenpolygon.**

Weitere Arten der graphischen Darstellung statistischer Sachverhalte:

die **Punktwolke** für bivariable Verteilungen,

die **Entwicklungskurve** zur Darstellung von Zeitreihen (dabei spielt der Trend als systematische Entwicklungsrichtung eine wichtige Rolle, typische Trendformen zeigt das Bild 65, S. 77),

das **Kartogramm** zur Darstellung von statistischen Daten für bestimmte geographische Regionen und

die **Bevölkerungs-** (oder Alters-) **pyramide** zur Charakterisierung der Bevölkerungsstruktur eines Landes zu einem bestimmten Zeitpunkt.

2.3. Datenauswertung

Nach der Datenerfassung (*Messen* und *Beobachten* von Sachverhalten, Gewinnen von Merkmalswerten) und der Datenaufbereitung (*Ordnen* und *Verdichten* der Daten, Gewinnen von Häufigkeitstabellen und graphischen Darstellungen) folgt jetzt die **Datenauswertung** als dritte Etappe bei der Durchführung empirischer Untersuchungen. Hier steht das *Berechnen statistischer Kenngrößen* im Vordergrund, insbesondere von Mittelwerten und Streuungsmaßen, die es gestatten, weitere Aufschlüsse über das untersuchte Material zu vermitteln, und die zugleich eine nochmalige Verdichtung der aufbereiteten Daten gewährleisten.

► **E 26** Unter **Datenauswertung** verstehen wir das Berechnen statistischer Kenngrößen (auch Maßzahlen genannt) aus den erfaßten oder aufbereiteten Daten.
Die Datenauswertung leitet über zur Prüfung von Hypothesen mit Hilfe mathematisch-statistischer Methoden.

Statistische Kenngrößen ermöglichen es, die Häufigkeitsverteilung, deren Aufstellung und Darstellung wir in vorangehenden Abschnitten behandelten, *durch wenige numerische Größen zu charakterisieren.*
Will man verschiedene Verteilungen miteinander vergleichen, so kann der Vergleich bei Berücksichtigung aller Daten erschwert werden. Häufig interessieren dabei auch nicht die zufälligen Unterschiede zwischen den Beobachtungswerten der Verteilungen, sondern es interessieren die Unterschiede zwischen markanten charakteristischen Kenngrößen, die die Verteilungen hinreichend beschreiben. Dadurch werden zufällige Unterschiede ausgeschaltet, und der Blick wird auf Wesentliches konzentriert. So gelingt es, eine verläßliche Wertung des Materials vorzunehmen und notwendige Maßnahmen zur Veränderung des erfaßten Zustandes einzuleiten.
Die Entscheidung darüber, ob eine Verteilung durch alle einzelnen Merkmalswerte oder durch wenige Kenngrößen charakterisiert werden soll, kann nur vom Ziel der Untersuchung abgeleitet werden.

■ **B 47** Die Entwicklung der Milchleistung der Kühe in der DDR soll über einen längeren Zeitraum untersucht werden. Hierzu interessieren nicht die zufälligen Unterschiede in den Milchleistungen der einzelnen Kühe, sondern nur die Durchschnittsleistung aller Milchkühe der DDR in den betrachteten Jahren. Aus diesen Zahlen kann der Trend der Leistungsentwicklung abgelesen werden.

Jahr	1950	1955	1960	1965	1970	1975	1980
Durchschnittl. Milchleistung je Kuh (in kg)	1891	2394	2646	2982	3314	3803	3923

Hier sind es also Mittelwerte, die miteinander verglichen werden. Es wäre unrationell und viel zu arbeitsaufwendig, unter der genannten Zielstellung eine Häufigkeitsverteilung für die Leistung der etwa 2 Millionen Milchkühe der DDR anzufertigen.
Will man dagegen die Futterzusammenstellung in einer großen Rindviehanlage auf wissenschaftlicher Grundlage vornehmen, dann muß die Milchleistung jeder einzelnen Kuh ermittelt werden, denn die Milchkühe werden

nach Leistung gefüttert, d. h., sie erhalten in Abhängigkeit von der Milchleistung eine unterschiedliche Menge und Qualität an Futtermitteln. Nur durch eine solche Differenzierung ist eine effektive Futternutzung gewährleistet.

An diesem Beispiel lernten wir einen Mittelwert als Kenngröße der untersuchten Häufigkeitsverteilung kennen. Es gibt aber noch andere Kenngrößen von Verteilungen. Wir unterscheiden

- **Mittelwerte** oder **Kenngrößen der Lage**
  Sie geben Auskunft darüber, wie die *Lage* der Verteilung auf der verwendeten Intervall- oder Rangskala ist (bezüglich der Normalverteilung ↗ Bild 42, S. 60). Für nominalskalierte Daten gibt es den Begriff Mittelwert im eigentlichen Sinne nicht.
- **Streuungsmaße** oder **Kenngrößen der Form**
  Sie erteilen Auskunft darüber, *wie* die Beobachtungswerte *um den Mittelwert verteilt* sind, ob sie dicht gedrängt oder weit auseinander liegen (bezüglich Normalverteilung ↗ Bild 43, S. 61). Streuungsmaße sind nur für Intervall- oder Rangskalen definierbar.
- **Kenngrößen,** die die **Gestalt** der Verteilung kennzeichnen, z. B. die Schiefe als Abweichung von der Symmetrie

Wir werden uns nur mit Mittelwerten und Streuungsmaßen beschäftigen. Es gibt mehrere verschiedene Mittelwerte und Streuungsmaße, wir behandeln nur die wichtigsten.
Mittelwerte haben – sollen sie ihren Zweck erfüllen – mehreren *Forderungen* zu genügen. Zwei haben wir schon genannt:

- Der Mittelwert soll das statistische Datenmaterial oder die Häufigkeitsverteilung eines Merkmals auf eine Kenngröße, die die mittlere Lage verdeutlicht, verdichten und damit das Material oder die Verteilung überschaubar und vergleichbar machen.
- Der Mittelwert soll das statistische Material oder die Verteilung insgesamt charakterisieren und die Einzelwerte möglichst gut repräsentieren.

Nicht weniger bedeutsam sind diese weiteren Forderungen:

- Der Mittelwert soll die Beurteilung von Einzelwerten innerhalb der erhobenen Daten ermöglichen.
- Die Bestimmung des Mittelwertes muß den Eigenschaften der Skala, auf der die Beobachtungswerte gemessen oder erfaßt werden, entsprechen.

Das bedeutet: Von Datenart *und* Verteilungsform hängt es ab, welchen der zu behandelnden Mittelwerte wir verwenden dürfen.
Scheuen Sie nicht die Mühe, die Berechnungen der in den Teilabschnitten folgenden Beispiel- und Übungsaufgaben selbständig vorzunehmen! Bei weiteren Aufgaben kann dann der elektronische Taschenrechner eine wertvolle Hilfe sein.

● **83*** **a)** Wenn Sie die Tabelle auf Seite 82 als Entwicklungskurve darstellen sollen, wie wären dann die Achsen zu bezeichnen?
b) Welche Darstellungsart wäre möglich?

● **84*** Dem Statistischen Taschenbuch der DDR von 1981 ist u. a. zu entnehmen:

Jahr	1950	1960	1970	1980
Hühnereier je Henne	92	135	168	205

Jahr	1950	1960	1970	1980
Honig (Tonnen)	2518	3190	5829	3907

Sind in diesen Tabellen Kenngrößen (statistische Maßzahlen) enthalten?

● **85*** Bei der Rückgabe einer Kontrollarbeit in einer Schulklasse erfährt jeder Schüler die von ihm erreichte Punktzahl.
Für die Auswertung einer solchen Kontrollarbeit im Kreis oder Bezirk sind von jeder Schule bestimmte Angaben weiterzuleiten. Welche sind das? Wird die Punktzahl jedes einzelnen Schülers mitgeteilt?

● **86** Für eine bedarfsgerechte Produktion der Bekleidungsindustrie ist das Erfassen verschiedener Daten erforderlich. Machen Sie Aussagen über Angaben, die für die Produktionsplanung auf diesem Gebiet zur Verfügung stehen müssen!

2.3.1. Das arithmetische Mittel

Das arithmetische Mittel ist der wichtigste und am häufigsten verwendete Mittelwert. Es muß aber von vornherein betont werden, daß er *nicht für jedes Datenmaterial und für jede Verteilung* einsetzbar ist.
Die Berechnung des arithmetischen Mittels ist nur dann gerechtfertigt, wenn
- Meßwerte vorliegen, der Messung also eine intervallskalierte Skala zugrunde liegt,
- die Häufigkeitsverteilung wenigstens annähernd normal verteilt ist.

Das arithmetische Mittel wird oft einfach als „Mittel", „Mittelwert" oder als „Durchschnitt" bezeichnet. In der 5. Klasse wurden Sie zum ersten Male mit dem arithmetischen Mittel vertraut gemacht. Sie lernten eine einfache Definition kennen und berechneten Durchschnittsweiten und Durchschnittserträge. Die dort verwendete Definition ist aber für unsere Zwecke nicht ausreichend, da sie auf ganze Zahlen beschränkt ist.

Die allgemeine, für uns zu nutzende Definition lautet:

► **D 9 Unter dem *arithmetischen Mittel* $\bar{x}$ (zu lesen: x quer) von n Zahlen (Meßwerten einer Zufallsgröße) $x_k (k = 1, 2, \ldots, n)$ verstehen wir den Quotienten aus der Summe der x_k und ihrer Anzahl n.**

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} \qquad (7)$$

Die Berechnung des arithmetischen Mittels ist also denkbar einfach: Man addiert alle Meßwerte und dividiert die Summe durch die Anzahl n der Meßwerte. Daraus folgt, daß das arithmetische Mittel stets die gleiche Einheit wie die einzelnen Meßwerte hat. Das trifft übrigens für alle Mittelwerte und Streuungsmaße zu. Das arithmetische Mittel kann also eine Zahl oder eine Größe sein.
Ergeben sich einige Meßwerte als 0, so ist deren Anzahl in n einzubeziehen.

■ **B 48 (1)** Im Zusammenhang mit einem chemischen Prozeß ist häufig die Temperatur der Kühlflüssigkeit zu prüfen. Acht Messungen ergeben die folgenden Temperaturwerte T (in Grad Celsius):

Messung k	1	2	3	4	5	6	7	8
T_k(°C)	+5	+1	−2	0	+4	+6	+2	0

Die Durchschnittstemperatur T ist zu bestimmen.

$$T = \tfrac{1}{8}(5 + 1 - 2 + 0 + 4 + 6 + 2 + 0)\,°\mathrm{C} = 2\,°\mathrm{C}$$

Kürzer läßt sich die Formel 7 mit Hilfe des **Summenzeichens** $\sum$ (groß sigma, es ist der griech. Buchstabe für S) ausdrücken:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k \tag{7a}$$

Der Ausdruck $\sum\limits_{k=1}^{n} x_k$ wird gelesen „Summe aller x_k für k gleich 1 bis n" und bedeutet ausführlich

$$\sum_{k=1}^{n} x_k = x_1 + x_2 + x_3 + \dots + x_n, \tag{8}$$

ist also eine Kurzform der Aufforderung: Summiere alle Zahlen oder Größen von x_1 bis x_n. k wird als Laufindex bezeichnet, 1 und n (mit $n > 0$, ganz) sind die Summationsgrenzen.

Das Summenzeichen bringt eine wesentliche Vereinfachung der Schreibarbeit, sofern zusammenhängende Glieder von Zahlenfolgen zu addieren sind.

■ **B 49** **a)** $\sum\limits_{k=1}^{4} x_k = x_1 + x_2 + x_3 + x_4$ **b)** $\sum\limits_{k=1}^{3} k^2 = 1^2 + 2^2 + 3^2 = 14$

Man hat also für den Laufindex k nacheinander die natürlichen Zahlen 1, 2, ... einzusetzen bis zu der Zahl, die über dem $\sum$ steht, und die so erhaltenen Glieder zu addieren.

Umgekehrt kann ein Summationsausdruck unter Verwendung des Summenzeichens $\sum$ verkürzt dargestellt werden.

■ **B 50** **a)** $y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 = \sum\limits_{k=1}^{7} y_k$

b) $1 + \dfrac{1}{2} + \dfrac{1}{3} + \dfrac{1}{4} = \sum\limits_{k=1}^{4} \dfrac{1}{k}$

Anstelle von k kann als Laufindex auch i oder j stehen. Die untere Summationsgrenze muß nicht immer 1 sein. In der Statistik ist es allerdings fast immer die Eins.

Wenn die Summationsgrenzen (1 und n oder m, z. B. bei Klasseneinteilung) feststehen und Mißverständnisse ausgeschlossen sind, so kann auf deren Angabe verzichtet werden, also

$$\sum_{k=1}^{n} x_k = \sum x_k.$$

● **87*** **a)** Stellen Sie in Kurzschreibweise dar und berechnen Sie die Summe der ersten sieben geraden Zahlen!

b) Schreiben Sie ohne Verwendung des Summenzeichens auf und berechnen Sie $\sum\limits_{j=0}^{5} 2^j$.

● **88*** Berechnen Sie!

a) $\sum\limits_{k=1}^{4} \sqrt{k-1}$

b) $\sum_{k=1}^{8} x_k$ für $x_1 = 5{,}0$ cm; $x_2 = 3{,}4$ cm; $x_3 = 6{,}0$ cm; $x_4 = 2{,}2$ cm; $x_5 = 7{,}5$ cm; $x_6 = 2{,}8$ cm; $x_7 = 4{,}4$ cm; $x_8 = 6{,}3$ cm

● **89*** Berechnen Sie das arithmetische Mittel

a) für die acht Meßwerte der Zufallsgröße „Länge" aus Aufgabe 88b);

b) der Punktzahlen, die die 20 am Mathematik-Olympiade-Wettbewerb beteiligten Schüler erzielten (Urdaten in Aufgabe 36, ↗ S. 38)!

● **90*** Welche mittlere (oder durchschnittliche) Ringzahl erreichte der Schütze, dessen Werte in Aufgabe 63 (↗ S. 65) angegeben sind?

Die Lösung der Aufgabe 90 ist unter Nutzung der Formeln (7) oder (7a) doch recht aufwendig. Beim *Vorliegen von Häufigkeitsverteilungen* (ohne und mit Klasseneinteilung) empfiehlt sich zur Verkürzung des Rechenaufwands die Anwendung der im weiteren aufgeführten Formeln (9) und (11).

$$\bar{x} = \frac{x_1 h_1 + x_2 h_2 + \ldots + x_m h_m}{n} \tag{9}$$

Unter Anwendung des Summenzeichens geht (9) über in

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{m} x_k h_k . \tag{9a}$$

Dabei bedeuten:

k Laufindex ($k = 1, \ldots, m$);
x_k k-te Merkmalsausprägung oder (bei Klasseneinteilung) k-te Klassenmitte;
h_k Häufigkeit der Meßwerte zur k-ten Merkmalsausprägung oder in der k-ten Klasse;
m Anzahl der verschiedenen Merkmalsausprägungen x_k oder Anzahl der Klassen;
n Anzahl der Meßwerte;
$\bar{x}$ arithmetisches Mittel der Verteilung.

Wir berechnen $\sum_{k=1}^{m} x_k h_k$, indem wir an die jeweils vorliegende Verteilungstabelle (mit den Spalten für x_k und h_k) rechts eine weitere Spalte für die Produkte $x_k h_k$ anfügen und diese summieren.

■ **B 9 (11)** Zum Merkmal X „Arbeitszeit des Lehrlings A je Steckermontage" ist das arithmetische Mittel der erfaßten Meßwerte (↗ Tabelle S. 16) gesucht. Zunächst fragen wir uns, ob wir berechtigt sind, das arithmetische Mittel zu berechnen. Die Frage ist zu bejahen, denn es liegen Meßwerte vor, die annähernd normal verteilt sind.

Merkmals-ausprägung x_k(s)	Häufigkeit h_k	Produkt $x_k h_k$
Spalte ①	②	③ = ① · ②
40	1	40
41	5	205
42	6	252
43	11	473
…	…	…
51	1	51
Summe	$50 \doteq n$	$2200 = \sum x_k h_k$

Nach (9a) ist $\bar{x} = \frac{1}{50} \cdot 2200\ \text{s} = 44{,}0\ \text{s}$.

Beachten Sie zur Resultatangabe: Wir geben stets eine Dezimalstelle mehr an, als mit der Messung des Merkmals erfaßt wurde! Hier wurde in vollen Sekunden gemessen, also $\bar{x} = 44{,}0$ s.

Man findet bei Berechnungen statistischer Kenngrößen oft die Unsitte, daß die Ergebnisse mit vier oder mehr Stellen nach dem Komma angegeben werden, obwohl die Meßwerte beispielsweise nur mit einer Stelle nach dem Komma erfaßt wurden. Begehen Sie nicht den gleichen Fehler! Täuschen Sie keine übertriebene Genauigkeit vor!

● **91*** Berechnen Sie das arithmetische Mittel der Arbeitszeit des Lehrlings B je Steckermontage! Häufigkeitstabelle dazu ↗ Tabelle zur Lösung der Aufgabe 26, S. 124.

Berechnen wir das arithmetische Mittel aus der Häufigkeitsverteilung *nach* Klassenbildung (mit Klassenbreite $d \geqq 2$), so ergeben sich für $\bar{x}$ geringfügige Abweichungen gegenüber der aus der ursprünglichen Häufigkeitsverteilung ermittelten Kenngröße. Das ist auch nicht verwunderlich, bringt doch die Klasseneinteilung immer einen gewissen Informationsverlust mit sich. (Aus der Tabelle nach Klassenbildung können wir ja nicht mehr entnehmen, mit welcher Häufigkeit die einzelnen Merkmalsausprägungen selbst auftreten.) Die Unterschiede sind aber im allgemeinen unbedeutend, so daß wir auch das $\bar{x}$ weiter verwenden dürfen, das sich aus der Häufigkeitsverteilung *mit* Klasseneinteilung (wo $d \geqq 2$) ergibt.

■ **B 9 (12)** Zum Merkmal X „Arbeitszeit des Lehrlings A je Steckermontage" ist das arithmetische Mittel zu berechnen, wobei, anders als in B 9 (11), dieses Mal die Häufigkeitsverteilung nach Klassenbildung (↗ Tabelle S. 56) verwendet werden soll.

Klasse (s)	Klassenmitte x_k (s)	Absolute Häufigkeit h_k	Produkt $x_k h_k$
Spalte ⓪	①	②	③ = ① · ②
40 bis 41	40,5	6	243
42 bis 43	42,5	17	722,5
44 bis 45	44,5	18	801
46 bis 47	46,5	6	279
48 bis 49	48,5	1	48,5
50 bis 51	50,5	2	101
Summe		$50 = n$	$2195{,}0 = \sum_{k=1}^{m} x_k h_k$

$$\bar{x} \approx \frac{2195}{50} = 43{,}9\ \text{s}$$

Gegenüber den 44,0 s, die wir aus der ursprünglichen (ungruppierten) Häufigkeitstabelle auf Seite 86 gewannen, wirklich ein geringer Unterschied! Allerdings ergeben sich hier bei der Berechnung über die Häufigkeitsverteilung *nach* Klassenbildung nur relativ geringe Vorteile. Das ist dadurch begründet, daß gegenüber den zwölf verschiedenen Merkmalsausprägungen, für die in der ersten Tabelle (↗ S. 86) Produkte zu bilden waren, jetzt die Produkte von immerhin noch sechs Klassen zu berechnen sind.

● **92*** Berechnen Sie die durchschnittliche Arbeitszeit des Lehrlings B je Steckermontage aus der Häufigkeitsverteilung nach Klassenbildung (↗ Tabelle S. 56), und vergleichen Sie das Ergebnis mit dem unter Aufgabe 91 gewonnenen!

Bei der Aufgabe zur Vermarktung von 60 Schweinen (untersuchtes Merkmal: Masse in kg; Urliste in Aufgabe 22, S. 22) ist der Weg zur Berechnung des arithmetischen Mittels über die Häufigkeitstabelle *nach* Klassenbildung schon wesentlich rationeller. Denn hier ist die Anzahl m der Klassen (mit 9) bedeutend geringer als die Anzahl der verschiedenen Merkmalsausprägungen.

● **93*** Berechnen Sie das arithmetische Mittel der Masse von 60 Schweinen aus der Häufigkeitstabelle nach Klasseneinteilung (↗ Tabelle zur Lösung der Aufgabe 56, S. 129)!

Sind der *Umfang n der Meßwerte wie auch deren Zahlenwerte recht groß*, so lohnt es sich, die manuelle Berechnung des arithmetischen Mittels noch weiter zu rationalisieren, und zwar mit dem **Verfahren des angenommenen Mittelwerts.** Das ist eine indirekte Methode zur Berechnung des arithmetischen Mittels, mit der es gelingt, die Zahlen, mit denen operiert werden muß, klein zu halten und so möglichen Rechenfehlern aus dem Wege zu gehen.
Der Grundgedanke dieses Verfahrens beruht auf folgendem:
Es soll beispielsweise das arithmetische Mittel der drei Zahlen 1019, 1011 und 997 berechnet werden. Wir operieren hier zweckmäßigerweise nicht mit diesen hohen Werten, sondern legen als angenommenen Mittelwert $x_a = 1000$ fest und berechnen das arithmetische Mittel der drei „Hilfswerte“ (eigentlich: Differenzen) 19, 11 und -3 $(= 997 - 1000)$. Wir erhalten $\frac{27}{3} = 9$. Infolgedessen ist $\bar{x} = 1009$.
Wie verfährt man nun bei Vorliegen der *Häufigkeitsverteilung* (mit oder ohne Klasseneinteilung)?
Wir gehen von einer möglichst zentral in der Verteilung gelegenen – im Grunde aber beliebigen – Merkmalsausprägung bzw. Klassenmitte x_k aus und setzen diese als *angenommenen Mittelwert* x_a. Statt mit den Merkmalswerten rechnen wir mit kleinen *Hilfswerten* x'_k, die wir mittels der Transformation (Umwandlungsgleichung)

$$x'_k = \frac{x_k - x_a}{d} \tag{10}$$

aus den Merkmalsausprägungen bzw. Klassenmitten x_k ($k = 1, 2, \ldots, m$; d ist die Klassenbreite) erhalten und die *stets ganzzahlig* sind. Das im besonderen ist es, was den Rechengang so vereinfacht.
Das aus diesen Hilfswerten x'_k gebildete arithmetische Mittel

$$\bar{x}_k = \frac{1}{n}(x'_1 h_1 + x'_2 h_2 + \ldots + x'_m h_m) \quad \text{bzw.} \quad \bar{x}_k = \frac{1}{n} \sum_{k=1}^{m} x'_k h_k$$

wird mit d multipliziert und zum angenommenen Mittelwert x_a addiert, so daß sich die *Bestimmung des arithmetischen Mittels einer Verteilung nach dem Verfahren des angenommenen Mittelwerts* ergibt aus

$$\bar{x} = x_a + \frac{d}{n}(x'_1 h_1 + x'_2 h_2 + \ldots + x'_m h_m) \tag{11}$$

bzw.

$$\bar{x} = x_a + \frac{d}{n} \sum_{k=1}^{m} x'_k h_k. \tag{11a}$$

Es bedeuten:
x_a angenommener Mittelwert;
d Klassenbreite (wenn keine Klasseneinteilung vorliegt, ist $d = 1$);
$x'_k = \frac{x_k - x_a}{d}$ Hilfswert für die k-te Merkmalsausprägung oder Klassenmitte;
h_k Häufigkeit der Meßwerte zur k-ten Merkmalsausprägung oder Klasse;
m Anzahl der verschiedenen Merkmalsausprägungen oder Anzahl der Klassen;
n Anzahl der Meßwerte;
$\bar{x}$ arithmetisches Mittel der Verteilung.

Die Teilschritte zur Anwendung des Verfahrens des angenommenen Mittelwerts sind:

1. Festlegen einer häufig vorkommenden Merkmalsausprägung bzw. Klassenmitte nahe der Verteilungsmitte als angenommener Mittelwert x_a;
2. Berechnen der Hilfswerte $x'_k = \frac{x_k - x_a}{d}$, diese x'_k sind *stets ganze Zahlen;*
3. Multiplizieren der x'_k mit dem entsprechenden h_k;
4. Summieren der Produkte $x'_k h_k$ (Vorzeichen beachten!);
5. Einsetzen in Formel (11a).

Auch hier erweist sich die Berechnung in Tabellenform als sehr günstig. Es treten fünf Spalten auf (0, 1, ..., 4). Darüber hinaus lassen wir Raum für eine weitere, die wir erst später benötigen.

■ **B 9 (13)** Zum Merkmal X „Arbeitszeit des Lehrlings A je Steckermontage" ist das arithmetische Mittel der Meßwerte nach Klassenbildung mit Hilfe des Verfahrens des angenommenen Mittelwerts zu berechnen (↗ Beispiel 9 (12), S. 87).

Klasse (s)	Klassenmitte x_k (s)	Absolute Häufigkeit h_k	Hilfswert x'_k	Produkt $x'_k h_k$	
⓪	①	②	③	④ = ③ · ②	⑤
40 bis 41	40,5	6	−2	−12 = (−2) · 6	
42 bis 43	42,5	17	−1	−17	
44 bis 45	44,5 = x_a	18	0	0	
46 bis 47	46,5	6	1	6	
48 bis 49	48,5	1	2	2	
50 bis 51	50,5	2	3	6	
Summe		50 = n		$-15 = \sum_{k=1}^{m} x'_k h_k$	

Wir haben uns für 44,5 als angenommenen Mittelwert x_a entschieden. (Wir hätten aber auch jede andere Klassenmitte dafür nehmen können.)

x'_3 ist gleich Null $\left(\text{wegen } x'_3 = \frac{44{,}5 - 44{,}5}{2} = 0\right)$.

Die anderen Hilfswerte sind ganze Zahlen; nach den niedrigen Merkmalsausprägungen zu ist es die Folge der negativen ganzen Zahlen, nach den höheren zu die Folge der positiven ganzen Zahlen. Man braucht also gar nicht alle x'_k zu berechnen, doch zur Kontrolle ist es für einige Klassen angebracht, z. B.

$$x'_1 = \frac{x_1 - x_a}{d} = \frac{40{,}5 - 44{,}5}{2} = -2.$$

Nach (11a) folgt aus der Tabelle

$$\bar{x} = \left(44{,}5 + \frac{2}{50} \cdot (-15)\right) \text{s} = (44{,}5 - 0{,}6)\,\text{s} = 43{,}9\,\text{s}.$$

Wir erhalten das gleiche Ergebnis wie im Beispiel 9 (12) (↗ S. 87).

Es gilt stets:

Das Verfahren des angenommenen Mittelwerts nach Formel (11) oder (11a) liefert die gleichen Resultate wie das Vorgehen nach Formel (9) bzw. (9a). Das Verfahren kann auf gruppierte wie auch ungruppierte Daten angewandt werden, ist also für jede angenähert normale Häufigkeitsverteilung aus Meßwerten geeignet.

● **94** Vollziehen Sie die auf S. 89 angegebenen Teilschritte zur Anwendung des Verfahrens des angenommenen Mittelwerts anhand des Beispiels 9 (13) nach!
Beachten Sie dabei: Beim Summieren in Spalte 4 sind die Vorzeichen zu berücksichtigen, infolgedessen ist auch $\sum x'_k h_k$ eine vorzeichenbehaftete Größe!

● **95*** Berechnen Sie die durchschnittliche Arbeitszeit des Lehrlings B je Steckermontage aus der Häufigkeitsverteilung nach Klassenbildung (↗ Tabelle S. 56) unter Anwendung des Verfahrens des angenommenen Mittelwerts, und vergleichen Sie das Ergebnis mit dem aus Aufgabe 92 gewonnenen!

Wir wenden uns jetzt den wichtigsten **Eigenschaften des arithmetischen Mittels** zu. Es gilt:

1. *Die Summe der Abweichungen aller Meßwerte von deren arithmetischem Mittel ist gleich Null.*
   Betrachten wir n Meßwerte $x_1, x_2 \ldots, x_n$, so ist also

$$\sum_{k=1}^{n} (x_k - \bar{x}) = 0. \tag{12}$$

Summe der Abweichungen aller x_k von $\bar{x}$	ist gleich Null.

■ **B 48 (2)** Temperatur der Kühlflüssigkeit bei chemischem Prozeß (↗ Beispiel 48, S. 84f.)
Erfaßt wurde:

Messung k	1	2	3	4	5	6	7	8
T_k (°C)	+5	+1	−2	0	+4	+6	+2	0

Berechnet wurde $\bar{T} = 2$ °C. Dann ergibt sich für die Differenzen $T - \bar{T}$ (Abweichungen der Meßwerte vom Mittelwert):

$T_k - \bar{T}$ (°C)	+3	−1	−4	−2	+2	+4	0	−2

Die Summe dieser Abweichungen ist $\sum_{k=1}^{8} (T_k - \bar{T}) = 0$.

Wurden die Meßwerte auf eine oder mehrere Dezimalstellen Genauigkeit erhoben, so können sich aufgrund von Rundungen geringfügige Abweichungen von Null ergeben.

2. *Die Summe der Abweichungsquadrate (kurz: SQ) aller Meßwerte von deren arithmetischem Mittel ist stets ein Minimum.*

$\sum (x_k - \bar{x})^2$	Min. (13)
Summe der Quadrate der Abweichungen aller x_k von $\bar{x}$	ist stets ein Minimum (kleinster Wert) gegenüber jedem anderen Wert, den man für $\bar{x}$ einsetzen würde.

■ **B 48 (3)** Wir schließen an das Beispiel 48 (2) an und quadrieren die Werte in der Zeile $T_k - \bar{T}$ (°C).

$T_k - \bar{T}$ (°C)	+3	−1	−4	−2	+2	+4	0	−2
$(T_k - \bar{T})^2$ (°C)^2	9	1	16	4	4	16	0	4

$$\sum_{k=1}^{8} (T_k - \bar{T})^2 = 54$$

Setzen wir jetzt für $\bar{T}$ statt des (richtigen) Mittelwertes 2 °C irgendwelche anderen Werte ein, z. B. 0 °C und dann 3 °C, so erhalten wir

im ersten Falle: $\sum (T_k - 0\,°\mathrm{C})^2 = 86$,
im zweiten Falle: $\sum (T_k - 3\,°\mathrm{C})^2 = 62$,

also in beiden Fällen einen höheren Wert als 54.
Und das ist stets so.

Nur $\sum_{k=1}^{8} (T_k - \bar{T})^2$ erweist sich als Minimum, als kleinster Wert.

So einfach das arithmetische Mittel und seine Berechnung auch ist, muß man sich doch einiger **Besonderheiten**, die ihm zu eigen sind, stets bewußt sein:

- Das arithmetische Mittel *stützt sich auf alle Meßwerte*. Jede Änderung an einem Meßwert, jede Zugabe oder Wegnahme eines Meßwertes bewirken im allgemeinen Veränderungen des arithmetischen Mittels der Meßreihe.
- In den bisherigen Beispielen und Aufgaben betrachteten wir Meßwerte und arithmetische Mittel von Stichproben. Zuweilen ist es aber erforderlich, den Mittelwert der Verteilung der Zufallsgröße in der Grundgesamtheit zu bestimmen. Diesen Mittelwert bezeichnen wir mit μ (zu lesen: mü, griech. Buchstabe), es ist der *wichtigste Lageparameter der Grundgesamtheit.*
  Auf zwei Wegen kann man ihn ermitteln:
  1. Anhand einer repräsentativen (durch Zufallsauswahl gewonnenen) Stichprobe wird deren arithmetisches Mittel $\bar{x}$ berechnet und von diesem auf den Parameter μ geschlossen (Schluß von der Stichprobe auf die Grundgesamtheit, ↗ Abschnitt 3.2., S. 117ff.).
  2. Man berechnet den Mittelwert der Zufallsgröße für die Grundgesamtheit, das heißt, man bezieht sämtliche Untersuchungsobjekte, die für die Grundgesamtheit in Frage kommen, in die Berechnung ein. Das kann jedoch sehr aufwendig sein.

  In bestimmten Fällen bedient man sich zur *Berechnung des Mittelwertes (Durchschnitts) der Zufallsgröße in der Grundgesamtheit* noch eines anderen Vorgehens.
  Wollen wir z. B. den Pro-Kopf-Verbrauch der DDR-Bevölkerung an Butter für ein bestimmtes Jahr ermitteln, so ist es nicht erforderlich, den individuellen Verbrauch eines jeden DDR-Bürgers an Butter über ein Jahr lang täglich zu registrieren und schließlich zu summieren. Wir dividieren vielmehr den Gesamtverbrauch, der als identisch angesehen wird mit der Summe der Einzelverbräuche, durch die

Einwohnerzahl der DDR und erhalten somit den durchschnittlichen (mittleren) Verbrauch eines jeden Bürgers.

Dieses eben geschilderte Vorgehen erweist sich in ähnlicher Weise als notwendig beim Vorliegen von Prozentwerten. Zu beachten ist: *Prozentwerte dürfen im allgemeinen nicht arithmetisch gemittelt werden.*
Erinnern wir uns des Beispiels B 2 und der sich daran anschließenden Aufgabe 5 (↗ S. 7f.). Dabei ging es um die Normerfüllung (in %) von sechs Arbeitern; zu berechnen war die durchschnittliche Normerfüllung.
Falsch wäre es, das arithmetische Mittel der Prozentwerte zu berechnen, richtig ist vielmehr, die Produktion aller sechs Arbeiter in Stück (6000) durch die Normstückzahl insgesamt (5340 Stück) zu dividieren. Richtiges Ergebnis: 112,4%.

Weitere Aufgaben zum arithmetischen Mittel:

● **96*** Berechnen Sie das arithmetische Mittel der Anzahl von Fahrzeugen, die in einer halben Stunde am Abend dreier Werktage je Ampelphase eine Kreuzung passieren, indem Sie

a) die Urliste im Beispiel 10 (↗ S. 17),
b) die Häufigkeitstabelle ohne Klasseneinteilung auf Seite 57f.,
c) die Häufigkeitstabelle mit Klassenbildung (↗ S. 57f., dieselbe Tabelle) verwenden!
d) Prüfen Sie, ob die Berechnung des arithmetischen Mittels zu a) bis c) überhaupt möglich und sinnvoll ist! Wenn ja, vergleichen Sie die Ergebnisse miteinander!

● **97*** Bestimmen Sie das arithmetische Mittel

a) der von 51 Schülern der Klassenstufe 8 im Ballweitwurf erzielten Weiten (Daten ↗ S. 63),
b) zum Merkmal „Familienstand" für die Bevölkerung der DDR (↗ S. 40)!

Begründen Sie in jedem dieser Fälle, ob die Anwendung des arithmetischen Mittels auch gerechtfertigt ist!

● **98*** Über die Produktionsplanerfüllung von acht Betrieben in einem Monat liegen folgende Angaben vor:

Betrieb	Monatsproduktion in 1000 M Plan	Ist	Planerfüllung in %
A	100	105	105
B	150	144	96
C	100	102	102
D	108	95	88
E	80	80	...
F	125	130	...
G	300	336	...
H	50	40	...

a) Berechnen Sie die Planerfüllung (in %) für die Betriebe E, F, G und H!
b) Berechnen Sie die durchschnittliche wertmäßige Planerfüllung für alle acht Betriebe!

● **99*** Einem Eisenbahnfahrplan ist folgender Ausschnitt entnommen (Städtenamen fiktiv):

km			D 1777	…
0	A-stadt	ab	7.15	
30	B-hausen	an	7.40	
		ab	7.45	…
90	C-burg	an	9.05	
		ab	9.10	
115	D-leben	an	9.35	

a) Berechnen Sie die *durchschnittliche Reisegeschwindigkeit* des D 1777 von A-stadt nach D-leben!

b) Ferner soll die *durchschnittliche Fahrgeschwindigkeit* des D 1777 von A-stadt nach D-leben bestimmt werden. Ein Reisender macht das auf folgende Weise:

Geschwindigkeit des D-Zuges für die Strecke

A-stadt bis B-hausen: $v_1 = \dfrac{30\ \text{km}}{25\ \text{min}} = 1{,}20\,\dfrac{\text{km}}{\text{min}}$

B-hausen bis C-burg: $v_2 = \dfrac{60\ \text{km}}{80\ \text{min}} = 0{,}75\,\dfrac{\text{km}}{\text{min}}$

C-burg bis D-leben: $v_3 = \dfrac{25\ \text{km}}{25\ \text{min}} = 1{,}00\,\dfrac{\text{km}}{\text{min}}$

Durchschnittliche Fahrgeschwindigkeit

$$\bar{v} = \frac{v_1 + v_2 + v_3}{3} = 0{,}983\,\frac{\text{km}}{\text{min}} = 59{,}0\,\frac{\text{km}}{\text{h}}$$

Ist dieses Vorgehen gerechtfertigt, und ist damit die Größe $59\,\dfrac{\text{km}}{\text{h}}$ richtig?

Wenn nein, warum nicht? Schlagen Sie einen anderen Rechenweg vor!

	Strichliste zur statistischen Auswertung von Meßergebnissen	Nr.:	Datum:
		Bemerkungen:	
Material:		Prüfbedingungen:	

Klassen-grenzen (mm)	absolute Häufigkeit Strichliste	Anzahl h_m	Klassen-nummer m	$m \cdot h_m$	$m^2 \cdot h_m$	Häufigkeit %	Σ %
(1)	(2)	(3)	(4)	(5) =(3)×(4)	(6) =(4)×(5)	(7)	(8)
28,5 b.u. 29,5	I	1	−3	−3	9	1,1	1,1
29,5 b.u. 30,5	卌 III	8	−2	−16	32	8,8	9,9
30,5 b.u. 31,5	卌 卌 II	12	−1	−12	12	13,2	23,1
31,5 b.u. 32,5	卌 卌 卌 卌 卌 卌	30	0	0	0	33,0	56,1
32,5 b.u. 33,5	卌 卌 卌 卌 III	23	1	23	23	25,3	81,4
33,5 b.u. 34,5	卌 IIII	9	2	18	36	9,9	91,3
34,5 b.u. 35,5	卌 I	6	3	18	54	6,6	97,9
35,5 b.u. 36,5	II	2	4	8	32	2,2	100,1

Bild 69

Zur **rationellen Aufbereitung und Auswertung statistischer Daten** kann ein Vordruck „Strichliste zur statistischen Auswertung von Meßergebnissen“ (↗ Bild 69, S. 93, hier sind die Daten aus einer Untersuchung bereits eingetragen) eine wertvolle Hilfe sein.

Es bedeuten in unseren Symbolen: Index $m = k$ (Spalten ③, ⑤ und ⑥); Klassennummer m = Hilfswert x_k' (Spalte ④).
Hier ist für empirische Häufigkeitsverteilungen (mit oder ohne Klasseneinteilung) nicht nur das Eintragen der Strichliste möglich, sondern auch der absoluten und relativen Häufigkeiten, der relativen Summenhäufigkeiten (hier unter Spalte ⑧ als $\sum$ %) sowie der Spalten zur Datenauswertung, nämlich der Berechnung des arithmetischen Mittels und der Streuung (auf deren Berechnung wir unter 2.3.4. zurückkommen).

Als wichtiges Charakteristikum für die Häufigkeitsverteilung wird das **arithmetische Mittel in die graphische Darstellung der Verteilung** (Linienzug/Histogramm, Streckendiagramm oder Summenkurve/Treppenpolygon) **eingetragen,** und zwar auf der Merkmalsachse. Man orientiert sich an der Achseneinteilung über die richtige Plazierung des Mittelwerts. (Einzelheiten und ein Beispiel dazu werden auf der Seite 109 mit dem Beispiel 9 (20) gegeben.)

Wiederholungsaufgabe:

● **100** **a)** Stellen Sie die im Bild 69 festgehaltene Häufigkeitsverteilung graphisch dar!
b) Zeichnen Sie die Summenkurve der Verteilung (möglichst auf Millimeterpapier)!
c)*Berechnen Sie das arithmetische Mittel der Verteilung nach dem Verfahren des angenommenen Mittelwerts!
d) Tragen Sie das berechnete arithmetische Mittel der Verteilung auf den Merkmalsachsen der beiden aus a) und b) gewonnenen Diagramme ein, und beschreiben Sie in beiden Fällen die Lage des Mittelwertes zur ganzen Verteilung!

2.3.2. Das gewogene arithmetische Mittel

Sollen mehrere arithmetische Mittelwerte gemittelt werden, mit anderen Worten: Soll der Durchschnitt einiger Werte berechnet werden, die selbst Durchschnittswerte sind, so können einem bei gedankenlosem Vorgehen leicht Fehler unterlaufen.
Die Anwendung des einfachen arithmetischen Mittels führt in diesen Fällen nicht zum Ziel, wenn die Einzelmittelwerte sich auf unterschiedliche Stichprobenumfänge, Wertungen, verschiedene Zeiten o. dgl. beziehen. An einem einfachen Beispiel sei das gezeigt.

■ **B 51(1)** Zwei Schülerarbeitsgemeinschaften „EDV“ einer Schule beteiligen sich an einem Wettbewerb, bei dem maximal 20 Punkte zu erreichen sind.
Die AG I besteht aus 5 Schülern, die AG II aus 11 Schülern. Es wurden folgende Leistungen erzielt:

Schüler	Leistung (in Punkten) AG I	AG II
1	7	13
2	3	17
3	8	19
4	5	10

Schüler	Leistung (in Punkten) AG I	AG II
5	2	8
6		16
7		16
8		14
9		20
10		15
11		17
Summe	25	165
Arithmet. Mittel $\bar{x}$	$\bar{x}_1 = 5{,}0$ Pkt.	$\bar{x}_2 = 15{,}0$ Pkt.

Die AG I schnitt also relativ schlecht, die AG II verhältnismäßig gut ab.
Bildet man von $\bar{x}_1$ und $\bar{x}_2$ das einfache arithmetische Mittel, so erhält man:

$$\bar{x} = \frac{\bar{x}_1 + \bar{x}_2}{2} = 10{,}0 \text{ Punkte.}$$

Mit diesem Wert darf keinesfalls gearbeitet werden, *dieser Wert ist falsch*!
Den richtigen arithmetischen Mittelwert für beide Gruppen erhalten wir, wenn wir sämtliche 16 Schüler zu einer Gruppe zusammenfassen und deren arithmetisches Mittel bestimmen.

$$\bar{x}_g = \frac{190}{16} \text{ Punkte} = 11{,}88 \text{ Punkte} \qquad \bar{x}_g = 11{,}9 \text{ Punkte}$$

Das ist doch ein erheblicher Unterschied, und bei großen Stichprobenumfängen kann sich das noch stärker auswirken.
Nun sind, wenn man Gruppenmittel zusammenfassen will, oft nicht mehr die Einzelwerte bekannt, sondern nur die Mittelwerte der Gruppen, für unser Beispiel also die Angaben

$$\text{AG I mit } n_1 = 5; \quad \bar{x}_1 = 5{,}0 \text{ Pkte;}$$
$$\text{AG II mit } n_2 = 11; \quad \bar{x}_2 = 15{,}0 \text{ Pkte.}$$

Doch das genügt. Man erhält das richtige Gesamtmittel $\bar{x}_g$, wenn man die Mittelwerte $\bar{x}_1$ und $\bar{x}_2$ mit den Umfängen n_1 bzw. n_2 der betreffenden Gruppen „wichtet“, also ansetzt

$$\bar{x}_g = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2}{n_1 + n_2}. \tag{14}$$

Das nennt man das **gewogene arithmetische Mittel.**

■ **B 51 (2)** In unserem Beispiel ergibt sich:

$$\bar{x}_g = \frac{5 \cdot 5 + 15 \cdot 11 \text{ Punkte}}{16} = 11{,}88 \text{ Punkte,}$$

$\bar{x}_g = 11{,}9$ Punkte.

Der oben berechnete Wert $\bar{x}_1$ bestätigt sich also.

Allgemein definiert man:

▶ **D 10 Das *gewogene arithmetische Mittel* $\bar{x}_g$ von l Zahlen x_i ($i = 1, \ldots, l$) ist ein arithmetischer Mittelwert, bei dem die x_i mit Wägungsfaktoren (Gewichten) n_i versehen sind.**

$$\bar{x}_g = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_l n_l}{n_1 + n_2 + \ldots + n_l} \tag{15}$$

Die x_i brauchen also nicht unbedingt selbst Mittelwerte zu sein, es können auch Noten sein, die nach sachlogischen Erwägungen, oder Geschwindigkeiten, die nach unterschiedlichen Zeiten „gewichtet" werden.
In den meisten Fällen sind es aber Mittelwerte $\bar{x}_i$ von Einzelstichproben mit den Umfängen n_i ($i = 1, \ldots, l$), die zu einem Gesamtmittel $\bar{x}_g$ vereinigt werden:

$$\bar{x}_g = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \ldots + \bar{x}_l n_l}{n_1 + n_2 + \ldots + n_l}. \tag{15a}$$

● **101*** Von drei Stichproben gleichen Umfangs sind zu ein und derselben Zufallsgröße arithmetische Mittel berechnet worden. Der durchschnittliche Wert aller drei Gruppen soll bestimmt werden.
Beweisen Sie, ausgehend von Formel (15a), daß man hier das einfache arithmetische Mittel anwenden kann!

● **102*** Die drei Kliniken einer Stadt geben die durchschnittlichen Geburtsmassen der in einem Monat bei ihnen zur Welt gekommenen männlichen Neugeborenen an:

Klinik	Zahl n_i der Neugeborenen	Durchschnittliche Geburtsmasse $\bar{x}_i$ (in g)
I	7	3600
II	68	3350
III	25	3550
	$100 = n$	

Berechnen Sie die durchschnittliche Geburtsmasse der im betrachteten Monat in den drei Kliniken zur Welt gekommenen Jungen!

● **103*** Von den 9 landwirtschaftlichen Betrieben eines Kreises werden die durchschnittlichen Milchleistungen (in l je Kuh und Jahr) und die Zahl ihrer Kühe wie folgt angegeben:

Betrieb i	Durchschnittliche Milchleistung $\bar{x}_i$ (in Liter je Kuh u. Jahr)	Zahl n_i der Kühe
1	2950	240
2	3600	300
3	3450	420
4	4100	460
5	3480	180

Betrieb i	Durchschnittl. Milchleistung $\bar{x}_i$ (in Liter je Kuh u. Jahr)	Zahl n_i der Kühe
6	4260	220
7	3600	560
8	4400	600
9	3810	400
		3380

Berechnen Sie die durchschnittliche Milchleistung aller Kühe des betreffenden Kreises!

2.3.3. Zentralwert und Modalwert

Neben dem arithmetischen und dem gewogenen arithmetischen Mittel gibt es weitere Mittelwerte, von denen wir in diesem Abschnitt zwei betrachten wollen: den Zentralwert (oder: Median) und den Modalwert (oder: das Dichtemittel).
Der **Zentralwert** ist anwendbar auf *intervall- und ordinalskalierte Daten*. Diese Kenngröße repräsentiert die mittlere Lage einer Verteilung insbesondere auch dann, wenn

- *die Verteilung nicht symmetrisch* (also linksschief oder rechtsschief) ist,
- *„Ausreißerwerte“* auftauchen,
- *nur eine kleine Zahl von Beobachtungswerten vorliegt*, bei der man über die Verteilungsform kaum etwas aussagen kann.

Dazu sei ein Beispiel angeführt.

■ **B 52 (1)** In einem Industrieinstitut liege folgende Einkommensverteilung vor:

Monatl. Bruttoeinkommen x_k (in M)	Häufigkeit h_k
300	4
400	6
500	3
700	2
1000	2
2000	1
2500	1
4000	1
	$20 = n$

Ein Teil der beobachteten Werte liegt dicht beieinander, während andere nach rechts hin weit abstehen („Ausreißerwerte“), die Verteilung ist nicht symmetrisch, und sie ist zerrissen (↗ Bild 70, S. 98).
Jede dieser Eigenschaften für sich wäre schon Anlaß genug, von der Anwendung des arithmetischen Mittels Abstand zu nehmen.

Wenn wir es hier trotzdem tun, so nur, um deutlich zu machen, zu welchen Fehlschlüssen das Verwenden des arithmetischen Mittels bei einer solchen Verteilung führen kann.

$$\bar{x} = \frac{1}{20} \cdot 17000 \text{ M} = 850 \text{ M}$$

Würde man zur Charakterisierung des Einkommens der Mitarbeiter dieses Instituts die Aussage formulieren, daß sie im Durchschnitt 850 M monatlich verdienen, so würden 15 von den 20 Mitarbeitern berechtigt Einwände dagegen erheben, denn ihr Einkommen liegt – zum Teil erheblich – unter diesem „Mittelwert".

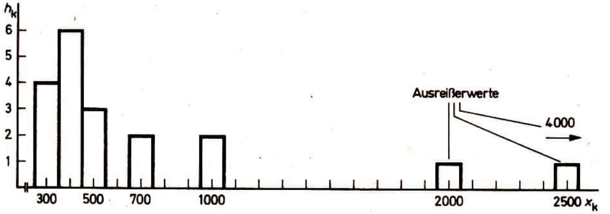


Bild 70

Ein treffenderes Bild von der zentralen Lage derartigen statistischen Materials liefert der **Zentralwert.**

▶ **D 11 Der *Zentralwert* (oder *Median*) Z ist derjenige Wert in der nach der Größe geordneten Folge der Meß- oder Rangwerte, der die mittlere Lage aller Werte einnimmt.**

Bei einem stetigen Merkmal kann man bezüglich der Fläche unter der Verteilungskurve formulieren: 50 % der Beobachtungswerte liegen unterhalb, 50 % oberhalb des Zentralwerts (↗ Bild 71).

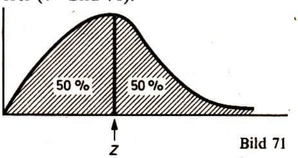


Bild 71

■ **B 53** Für die Beobachtungsfolge 3; 4; 4; $\underset{Z}{7}$; 8; 9; 9 ist 7 der Zentralwert.

■ **B 52 (2)** Für die Einkommensverteilung

300; 300; 300; 300; 400; 400; 400; 400; 400; 400; $\underset{Z}{}$ 500; 500; 500; 700; 700; 1000; 1000; 2000; 2500; 4000

ist $Z = 450$ M. 10 Werte liegen unterhalb von Z, 10 Werte oberhalb.

Die Bestimmung des Zentralwerts ist auf zwei Wegen möglich: durch *Auszählen* (wie wir es in den Beispielen soeben taten; bei großer Zahl von Beobachtungswerten ist das aber nicht rationell) oder durch *Berechnen.*

Auszählen:

Wir ordnen die beobachteten Werte der Größe nach und stellen fest, ob ihre Anzahl gerade oder ungerade ist.

Für *n ungerade* ist Z der in der Mitte der geordneten Datenfolge stehende Wert (↗ Beispiel 53);

für *n gerade* ist Z das arithmetische Mittel der beiden in der Mitte der Folge stehenden Werte [↗ Beispiel 52 (2)].

Berechnung des Zentralwerts:

Hierzu geht man von der Häufigkeitstabelle der betreffenden Verteilung (mit oder ohne Klasseneinteilung) aus, sie soll möglichst auch die Summenhäufigkeiten enthalten.

Als erstes hat man den sogenannten **Eingriffsspielraum für** Z zu bestimmen.

► **E 27** Der **Eingriffsspielraum für** Z ist die Merkmalsausprägung oder Klasse, in die der Zentralwert eingreift, d. h., in der der 50%-Wert der Verteilung liegt.

Es gilt

$$Z = x_{ug} + d \cdot \frac{\frac{n}{2} - sh_u}{h_z}. \tag{16}$$

Dabei bedeuten:

x_{ug} exakte untere Klassengrenze des Eingriffsspielraumes für Z;
sh_u Summenhäufigkeit bis x_{ug}, d. h. Summe der Häufigkeiten h_k unterhalb x_{ug};
h_z Häufigkeit im Eingriffsspielraum;
d Klassenbreite;
n Anzahl der Beobachtungswerte, Umfang der Stichprobe.

Beweisanmerkung: Die Formel (16) läßt sich aus dem Interpolationsansatz

$x_z : d = \left(\frac{n}{2} - sh_u\right) : h_z;\quad Z = x_{ug} + x_z$ herleiten.

■ **B 9 (14)** Zur Verteilung des Merkmals X „Arbeitszeit des Lehrlings A je Steckermontage" soll der Zentralwert berechnet werden, und zwar aufgrund der Häufigkeitstabelle nach Klassenbildung. Z ist mit $\bar{x}$ (↗ Beispiel 9 (12), S. 87, oder Beispiel 9 (13), S. 89f.) zu vergleichen.

Klasse (s)	Klassenmitte x_k(s)	Absolute Häufigkeit h_k	Summenhäufigkeit sh_k
40 bis 41	40,5	6	6
42 bis 43	42,5	17	$23 = sh_u$
$\boxed{\text{44 bis 45}}$	44,5	$18 = h_z$	41
46 bis 47	46,5	6	47
48 bis 49	48,5	1	48
50 bis 51	50,5	2	50
Summe		$50 = n$	

Hier ist $n = 50$. Wir suchen die Klasse, die den 50%-Wert der Verteilung, das ist also der 25. Beobachtungswert der geordneten Wertefolge, enthält. Aus der letzten Spalte der Tabelle entnehmen wir, daß die eingerahmte Klasse den $\boxed{\text{Eingriffsspielraum für } Z}$ bildet.

Es ist $x_{ug} = 43{,}5$; $h_z = 18$; $sh_u = 23$; $d = 2$.

Aus (16) folgt

$$Z = \left(43{,}5 + 2 \cdot \frac{25 - 23}{18}\right) \text{s} = 43{,}7 \text{ s}.$$

Für $\bar{x}$ ergab sich aus dem Beispiel 9 (12) oder 9 (13): $\bar{x} = 43{,}9$ s.
Der geringe Unterschied zwischen Z und $\bar{x}$ sowie die Ungleichung $\bar{x} > Z$ lassen auf eine leicht linksschiefe Verteilung schließen, was uns ja schon aus den Bildern 12 (↗ S. 36) und 14 (↗ S. 37) hinlänglich bekannt ist.

Verteilungen, für die die Berechnung von $\bar{x}$ möglich ist, lassen natürlich auch die Bestimmung von Z zu. Die Umkehrung dieses Satzes gilt nicht.
Sind beide Mittelwerte bekannt, so sind folgende Aussagen möglich:
Symmetrische Verteilungen liegen dann und nur dann vor, wenn gilt $\bar{x} = Z$ (Bild 72a)).
Entsprechend gilt
für *linksschiefe Verteilungen:* $\bar{x} > Z$ (Bild 72b)),
für *rechtsschiefe Verteilungen:* $\bar{x} < Z$ (Bild 72c)).

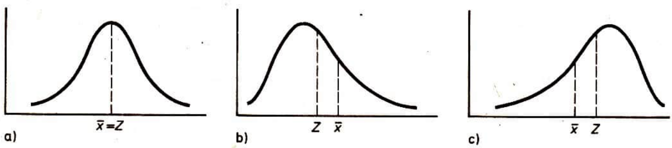


Bild 72

Eigenschaften des Zentralwerts:

1. *Die Summe der absoluten Beträge der Abweichungen aller Einzelwerte vom Zentralwert ist ein Minimum.*

$$\sum_{k=1}^{n} |x_k - Z| \text{ min.} \tag{17}$$

2. *Der Zentralwert ist von extremen Werten der empirischen Verteilung unabhängig.* Das ist in gewissen Fällen von großem Vorteil.
3. *Z kann auch dann bestimmt werden, wenn (eine oder beide) Randklassen der Verteilung nach unten und/oder oben offen sind.* Offene Klassen sind z. B. „unter 40 (s)" oder „über 51 (s)".
4. *Bei stetigen Merkmalen ist Z der Fußpunkt derjenigen Ordinate, welche die Fläche unter dem Linienzug in zwei flächengleiche Teile zerlegt* (↗ Bild 71, S. 98).

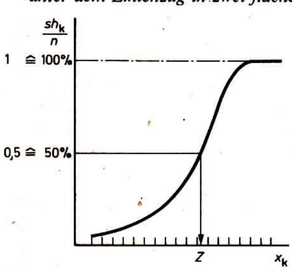


Bild 73

All diese Eigenschaften sind in der Praxis von Bedeutung.

Aus der 4. Eigenschaft folgt, daß sich der Zentralwert in einfachster Weise aus der Summenkurve der Häufigkeitsverteilung ablesen läßt. Man zieht durch die 0,5 ($\triangleq$ 50 %)-Marke auf der Ordinatenachse eine Parallele zur Merkmalsachse bis zum Schnitt mit der Summenkurve. Lotet man von diesem Schnittpunkt auf die Merkmalsachse, so liest man dort unmittelbar den Zentralwert ab (Bild 73).

● **104*** Berechnen Sie den Zentralwert der zum Merkmal „Arbeitszeit des Lehrlings B je Steckermontage" erfaßten Häufigkeitsverteilung (mit Klasseneinteilung, ↗ Tabelle S. 56), und vergleichen Sie Z mit $\bar{x}$ (aus der Lösung zur Aufgabe 92 oder 95)!

● **105*** Bestimmen Sie durch Auszählen oder Berechnen den Zentralwert

a) für die Wertefolge
11; 12; 13; 13; 13; 14; 14; 14; 16; 16; 17; 18; 20 (in mm),

b) für die folgende Verteilung!

Klassenmitte x_k(g)	0,13	0,16	0,19	0,22	0,25	0,28	Summe
h_k	1	0	2	4	12	11	$30 = n$
sh_k	1	1	3	7	19	30	

c) Wäre für a) und b) auch die Berechnung von $\bar{x}$ angebracht? Wenn nein, warum nicht?

● **106*** Wir greifen unser Beispiel zum stetigen Merkmal „Nutentiefe" wieder auf (Urliste: ↗ Aufgabe 17, S. 18; Häufigkeitstabelle nach Klassenbildung: ↗ S. 49, Variante B; graphische Darstellung: ↗ Bild 33, S. 53).

a) Ist das arithmetische Mittel $\bar{x}$ oder der Zentralwert Z der dieser Verteilung angemessenere Mittelwert? Begründen Sie Ihre Antwort!

b) Berechnen Sie $\bar{x}$ und Z (möglichst rationell, also in *einer* Tabelle), und beurteilen Sie deren Lage im Häufigkeitsdiagramm (Bild 33)!

● **107*** Merkmale X und Y „Arbeitszeit je Steckermontage für Lehrling A bzw. B" Lesen Sie aus dem Bild 54 (S. 70) auf einfache Weise den Zentralwert für die Häufigkeitsverteilung von X (Lehrling A) und von Y (Lehrling B) ab! Vergleichen Sie die von Ihnen abgelesenen Werte mit den berechneten (unter Beispiel 9 (14), S. 99f., und Lösung zu Aufgabe 104, S. 137)!

Der **Modalwert** nimmt in zweierlei Hinsicht eine Sonderstellung unter den Kenngrößen der Lage ein:

- Er ist *bei bestimmten Verteilungsformen* (mehrgipfligen, insbesondere U-förmigen, aber auch bei J-förmigen Verteilungen) *gar kein „Mittelwert"* im eigentlichen Sinne des Wortes.
- Der Modalwert ist die *einzige Kenngröße*, die auch *bei nominalskalierten Daten* Anwendung finden kann.

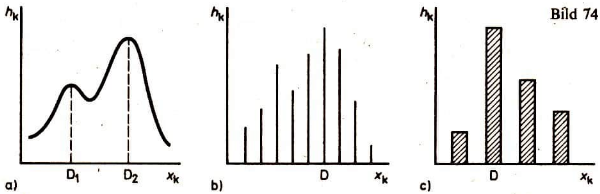
► **D 12 Unter einem *Modalwert* (oder *Dichtemittel*) D einer Verteilung versteht man diejenige Merkmalsausprägung, die bezüglich der benachbarten Ausprägungen *am häufigsten* vorkommt.**

Bei zwei- und mehrgipfligen Verteilungen gibt es also zwei bzw. mehrere Modalwerte.

■ **B 54** Beispiele für die Anzahl und Lage von Modalwerten bei einer

zweigipfligen Verteilung einer stetigen Zufallsgröße mit intervallskalierten Daten	eingipfligen Verteilung einer diskreten Zufallsgröße	eingipfligen Verteilung einer diskreten Zufallsgröße mit ordinal- oder nominalskalierten Daten

Bild 74



Die Angabe des Dichtemittels ist also bei Verteilungen *jeder Datenart* (von Meßwerten, Rangdaten oder Kategorien) möglich. Der Modalwert wird meist der tabellarischen Darstellung der Häufigkeitsverteilung entnommen.
Man *beachte*, daß man für D die Merkmalsausprägung(en) x_k angibt, zu der der (oder die) häufigste(n) Wert(e) gehören, und nicht etwa die Häufigkeit h_k selbst!

■ **B 9 (15)** Verteilung des Merkmals X „Arbeitszeit des Lehrlings A je Steckermontage" *vor* der Klassenbildung (↗ Tabelle zum Beispiel 9 (2), S. 31).
Modalwert $D = 43$ s, weil zu dieser Merkmalsausprägung die größte Häufigkeit ($h = 11$) gehört. Vollkommen falsch wäre: $D = 11$.

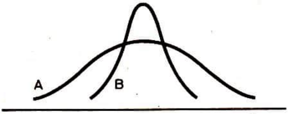
● **108*** Geben Sie den (oder die) Modalwert(e) folgender Häufigkeitsverteilungen an!
a) Zufallsgröße U „Güte eines Erzeugnisses", erfaßt nach Güteklassen (↗ Tabelle zum Beispiel 24, S. 40)
b) Zufallsgröße „Familienstand für DDR-Bevölkerung" (↗ Tabelle zum Beispiel 25 oder zum Bild 19, S. 40)
c) Zufallsgröße „Nutentiefe" *vor* (↗ Tabelle S. 47) und *nach* Klassenbildung (↗ Tabelle zu Variante B, S. 49)
d) Zufallsgröße X „Arbeitszeit des Lehrlings A ..." nach Klassenbildung (↗ Tabelle S. 56)

● **109*** a) Gilt für eine (wirkliche) Normalverteilung $\bar{x} = Z = D$?
b) Welche Beziehung (Ungleichung) besteht zwischen $\bar{x}$, Z und D bei einer rechtsschiefen Verteilung?

2.3.4. Streuungsmaße

Neben den Mittelwerten sind es insbesondere die Streuungsmaße, die zur Charakterisierung der Häufigkeitsverteilung durch wenige numerische Größen beitragen. Streuungsmaße sind *Kenngrößen der Verteilungsform*, sie geben an, wie stark die einzelnen Beobachtungswerte vom Mittelwert (der untersuchten Stichprobe oder Grundgesamtheit) abweichen, wie weit sie unter Bezugnahme auf den Mittelwert „streuen".

● **110*** Das Bild 75 zeigt die Linienzüge von zwei annähernd normalen empirischen Häufigkeitsverteilungen A und B einer stetigen Zufallsgröße.



Worin zeigen die beiden Verteilungen Übereinstimmung?
Was ist bei ihnen verschieden?

Bild75

Wie bei den Mittelwerten gibt es auch für die Streuung mehrere unterschiedliche Maßzahlen:

– Die Variationsweite w,
– die durchschnittliche Abweichung e,
– die Standardabweichung z usw.

Ihre Anwendung hängt nicht zuletzt wieder von der Form der Verteilung und der Art der erhobenen Daten ab. Die Standardabweichung ist das wichtigste Streuungsmaß, sie wird meist gemeinsam mit dem arithmetischen Mittel $\bar{x}$ berechnet und ist überall dort angebracht, wo auch $\bar{x}$ sinnvoll genutzt werden kann. Zum Zentralwert Z dagegen gehört das Streuungsmaß „Quartilabstand"; dieses behandeln wir hier nicht.
Die **Variationsweite** w ist das einfachste Streuungsmaß. Es gibt die Verteilungsbreite an und wird anhand der Häufigkeitstabelle *ohne* Klassenbildung berechnet.

► **D 13 Unter der *Variationsweite* (auch *Variationsbreite*) w einer Verteilung verstehen wir die Differenz zwischen deren kleinstem und größtem Merkmalswert.**

$$w = x_{max} - x_{min} \tag{18}$$

w kann beim Vorliegen intervall- oder rangskalierter Daten auf jede Verteilungsform angewandt werden.

■ **B 9 (16)** Zur Verteilung des Merkmals X „Arbeitszeit des Lehrlings A ..." (Häufigkeitstabelle vor Klassenbildung, ↗ S. 31) ist die Variationsweite $w = (51 - 40)$ s $= 11$ s, weil $x_{max} = 51$ s der größte und $x_{min} = 40$ s der kleinste in der Verteilung überhaupt vorkommende Merkmalswert ist.

Die Bedeutung der Variationsweite darf nicht unterschätzt werden.

■ **B 55** Der VEB Filmfabrik Wolfen liefert ORWOchrom-Filme in viele Länder. Für den Transport und die Lagerung der Filme in den Einsatzländern gilt es zu beachten, daß extreme Temperaturen auftreten können (man denke an Grönland und an Angola). Hier muß die Variationsweite bezüglich der Zufallsgröße „Temperatur" berechnet werden. Die Angabe eines anderen Streuungsmaßes oder gar der mittleren Temperatur wäre für diesen Zweck wertlos.

● **111*** Von einer Einheit der Nationalen Volksarmee soll ein Fluß überwunden werden. Dem Kommandeur wurden bezüglich der Tiefe des Flusses an der betreffenden Stelle zwei Angaben gemeldet:
Durchschnittliche Tiefe $\bar{x} = 1{,}50$ m; Tiefe reicht von 0,60 m bis 2,25 m.

a) Welche der beiden Angaben ist für den Kommandeur wichtiger?
b) Berechnen Sie die Variationsweite w!

● **112*** Berechnen Sie die Variationsweite zu folgenden Häufigkeitsverteilungen!

a) Zufallsgröße Y „Arbeitszeit des Lehrlings B ...“ (↗ S. 124, Lösung zur Aufgabe 26)

b) Zufallsgröße Z „Anzahl der Fahrzeuge ... an drei Abenden ...“ (↗ Urliste unter Beispiel 10, S. 17)

c) Zufallsgröße „Nutentiefe“ (↗ Tabelle auf S. 47)

Die Vorteile des Streumaßes Variationsweite haben wir kennengelernt. Sie hat aber auch Nachteile. Ein wesentlicher besteht darin, daß sie nur von den beiden Extremwerten der Verteilung ausgeht, die zwischen diesen liegenden Werte jedoch unberücksichtigt läßt.
Dieser Nachteil besteht bei der durchschnittlichen Abweichung e und bei der Standardabweichung s nicht, hier finden *alle* Meßwerte Eingang in die Berechnung. Außerdem wird bei e und s Bezug auf das arithmetische Mittel der Verteilung genommen. Die **durchschnittliche Abweichung *e*** charakterisiert also die Verteilung als Ganzes schon besser als w.

► **D 14 Die *durchschnittliche Abweichung* (oder: mittlere Abweichung) *e* ist das arithmetische Mittel aus den Absolutbeträgen der Abweichungen aller Meßwerte von deren arithmetischem Mittel.**

$$e = \frac{1}{n}(|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|)$$

$$= \frac{1}{n} \sum_{k=1}^{n} |x_k - \bar{x}| \qquad (19)$$

Formel (19) ist günstig, wenn Einzelmeßwerte gegeben sind.
Beim *Vorliegen von Häufigkeitsverteilungen* (ohne und mit Klassenbildung) empfiehlt sich die Anwendung von

$$e = \frac{1}{n} \sum_{k=1}^{m} |x_k - \bar{x}|\, h_k . \qquad (20)$$

(Zur Bedeutung der einzelnen Symbole vgl. mit den Angaben unter Formel (9a), S. 86)
Zu *beachten* ist, daß e nur in den Fällen berechnet werden darf, bei denen auch das arithmetische Mittel $\bar{x}$ angezeigt ist, also nur bei angenähert normalen Verteilungen mit intervallskalierten Daten. Wird von der Häufigkeitsverteilung mit Klasseneinteilung ausgegangen, so muß *das* arithmetische Mittel $\bar{x}$ verwendet werden, das aus dieser Verteilung gewonnen wurde! Zur Berechnung von e bewährt sich das Anlegen einer Tabelle. Wir bringen ein Beispiel zur Anwendung von Formel (19), ein weiteres zur Anwendung von (20).

■ **B 48 (4)** Temperatur der Kühlflüssigkeit bei chemischem Prozeß (↗ Beispiele 48 (1) und 48 (2), S. 84f. bzw. S. 90)
$\bar{T} = 2\,°C$

Temperatur T_k (°C)	Abweichung $T_k - \bar{T}$ (°C)	Absol. Betrag $\lvert T_k - \bar{T}\rvert$ (°C)
+5	+3	3
+1	−1	1
−2	−4	4

Temperatur T_k (°C)	Abweichung $T_k - \bar{T}$ (°C)	Absol. Betrag $\lvert T_k - \bar{T}\rvert$ (°C)
0	−2	2
+4	+2	2
+6	+4	4
+2	0	0
0	−2	2
Summe	0	$18 = \sum_{k=1}^{m} \lvert T_k - \bar{T}\rvert$

Nach (19): $e = \frac{1}{8} \cdot 18\ °C = 2{,}25\ °C$

$e \approx 2{,}2\ °C$

Das heißt: Die gemessenen Temperaturwerte weichen durchschnittlich 2,2 °C vom arithmetischen Mittel $\bar{T} = 2{,}0\ °C$ ab.

Das Ergebnis einer derartigen Berechnung wird dann auch oft in der Form $\bar{x} \pm e$ festgehalten. Für das Beispiel heißt das: (2,0 ± 2,2) °C.

■ **B 9 (17)** Für das Merkmal X „Arbeitszeit des Lehrlings A ..." ist die durchschnittliche Abweichung aus der Häufigkeitstabelle mit Klasseneinteilung (↗ Tabelle S. 56) zu berechnen. Das arithmetische Mittel ergab sich zu $\bar{x} = 43{,}9$ s (↗ Beispiel 9 (12), S. 87). (Nicht etwa $\bar{x} = 44{,}0$ s verwenden, da dieser Wert aus der ungruppierten Häufigkeitsverteilung gewonnen wurde)

Klasse (s)	Klassenmitte x_k (s)	Absolute Häufigkeit h_k	Abweichung $x_k - \bar{x}$ (s)	Absol. Betrag $\lvert x_k - \bar{x}\rvert$ (s)	Produkt $\lvert x_k - \bar{x}\rvert \cdot h_k$ (s)
40 bis 41	40,5	6	−3,4	3,4	20,4
42 bis 43	42,5	17	−1,4	1,4	23,8
44 bis 45	44,5	18	0,6	0,6	10,8
46 bis 47	46,5	6	2,6	2,6	15,6
48 bis 49	48,5	1	4,6	4,6	4,6
50 bis 51	50,5	2	6,6	6,6	13,2
Summe		50 = n			88,4

Nach (20): $e = \frac{1}{50} \cdot 88{,}4\ \text{s} = 1{,}77\ \text{s}$

$\bar{x} \pm e = (43{,}9 \pm 1{,}8)\ \text{s}$

● **113*** Berechnen Sie die durchschnittliche Abweichung der Häufigkeitsverteilung zur Zufallsgröße Y „Arbeitszeit des Lehrlings B ..."

a) unter Zugrundelegung der ungruppierten Daten (also Häufigkeitsverteilung *ohne* Klassenbildung (↗ S. 124, Lösung zur Aufgabe 26; $\bar{x} = 42{,}8$ s)) und

b) unter Verwendung der gruppierten Daten (also *nach* Klassenbildung (↗ Tabelle auf der Seite 56; $\bar{x} = 42{,}8$ s))!

c) Vergleichen Sie die Ergebnisse aus a) und b)!

Zur Kennzeichnung der Streuung einer Verteilung ist die **Standardabweichung** am besten geeignet. In diese Kenngröße gehen nicht nur alle Einzelwerte – wie bei e – ein, sondern aufgrund des Quadrierens der Differenzen $x_k - \bar{x}$ werden geringe Abweichungen der Meßwerte x_k vom Mittelwert $\bar{x}$ noch kleiner und beeinflussen die Kenngröße kaum, andererseits werden größere Abweichungen (> 1) noch größer und beeinflussen die Kenngröße stärker. Gerade ein solches Verhalten verlangen wir aber von einem guten Streuungsmaß.

► **D 15 Die *Standardabweichung* (oder: *mittlere quadratische Abweichung*) s einer Verteilung von n Zahlen (Meßwerten einer Zufallsgröße) x_k ($k = 1, \ldots, n$) ist die Quadratwurzel aus der durch $(n - 1)$ dividierten Summe der Quadrate der Abweichungen der x_k vom arithmetischen Mittel $\bar{x}$.**

$$s = \sqrt{\frac{1}{n-1}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2\right]} \qquad (21)$$

oder unter Hinzuziehung des Summenzeichens

$$s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2} \qquad (21\,a)$$

Beim Vorliegen *intervallskalierter Daten* in *angenähert normaler Verteilung* ist die Standardabweichung das geeignetste Maß für die Streuung der Einzelwerte um ihren Mittelwert.

■ **B 48 (5)** Temperatur der Kühlflüssigkeit bei chemischem Prozeß (↗ Beispiel 48 (4), S. 104f.)
Für die gemessenen Werte ist die Standardabweichung zu berechnen.
Zur Anwendung der Beziehung (21) brauchen wir die Zahlen der mittleren *oder* letzten Spalte der Tabelle im Beispiel 48(4) lediglich zu quadrieren und zu summieren.

$$9 + 1 + 16 + 4 + 4 + 16 + 0 + 4 = 54 = \sum_{k=1}^{n}(T_k - \bar{T})^2$$

$$s = \sqrt{\frac{1}{7} \cdot 54\,°C} = 2{,}78\ °C \approx 2{,}8\ °C$$

Bezüglich der Zahl der Stellen, die für das Ergebnis sinnvoll sind, ist immer wieder zu *beachten*, daß höchstens eine Dezimalstelle mehr angegeben werden darf, als in den Meßwerten vorliegt (↗ S. 87).

Das Gesamtresultat halten wir auf diese Weise fest:

$\bar{x} \pm s$. Im Beispiel: $(2{,}0 \pm 2{,}8)$ °C

Die Standardabweichung trägt also stets die gleiche Einheit wie das arithmetische Mittel.
Berechnet man s aus der *Häufigkeitstabelle* (ohne oder mit Klasseneinteilung), so sind die Beziehungen

$$s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{m}(x_k - \bar{x})^2 h_k} \qquad (22)$$

und – bei Nutzung des Verfahrens des angenommenen Mittelwerts –

$$s = d\sqrt{\frac{1}{n-1}\left(A - \frac{B^2}{n}\right)} \qquad (23)$$

mit $A = \sum_{k=1}^{m} x_k'^2 h_k$ und $B = \sum_{k=1}^{m} x_k' h_k$

anzuwenden. Als sehr vorteilhaft erweist sich insbesondere Formel (23), obwohl deren Aufbau Bedenken ob deren „Einfachheit" auslösen könnte. Die in den Formeln auftretenden Variablen sind auf Seite 89 oben erläutert.

■ **B 9 (18)** Für das Merkmal X „Arbeitszeit des Lehrlings A ..." ist die Standardabweichung aus der ursprünglichen Häufigkeitstabelle (also ohne Klasseneinteilung, ↗ Tabelle auf S. 31) zu berechnen. $\bar{x}$ ergab sich im Beispiel 9 (11) (↗ S. 87) zu $\bar{x} = 44{,}0$ s.
Folgende Tabelle führt zum Ziel.

Merkmals-ausprägung x_k (s)	Absolut. Häufigkeit h_k	Abweichung $x_k - \bar{x}$ (s)	Quadrat $(x_k - \bar{x})^2$	Produkt $(x_k - \bar{x})^2 h_k$
①	②	③	④	⑤ = ② · ④
40	1	−4	16	16
41	5	−3	9	45
42	6	−2	4	24
43	11	−1	1	11
44	9	0	0	0
45	9	1	1	9
46	3	2	4	12
47	3	3	9	27
48	0	4	16	0
49	1	5	25	25
50	1	6	36	36
51	1	7	49	49
Summe	50 = n			254 = $\sum (x_k - \bar{x})^2 h_k$

Aus (22) folgt:

$$s = \sqrt{\frac{1}{49} \cdot 254\,\text{s}} = 2{,}28\ \text{s}; \qquad s \approx 2{,}3\ \text{s} \qquad \bar{x} \pm s = (44{,}0 \pm 2{,}3)\ \text{s}$$

Man sagt dazu auch: Der Bereich der einfachen Streuung (Standardabweichung) erstreckt sich von $\bar{x} - s = 41{,}7$ s bis $\bar{x} + s = 46{,}3$ s. Der Rechenaufwand reduziert sich bei Häufigkeitstabellen *mit* Klasseneinteilung unter Verwendung von Formel (23) wesentlich. Das zeigt folgendes Beispiel.

■ **B 9 (19)** Für das gleiche Merkmal X (↗ Beispiel 9 (18)) ist die Standardabweichung aus der Häufigkeitstabelle nach Klassenbildung mit Hilfe des Verfahrens des angenommenen Mittelwerts zu berechnen.
Wir benötigen für Formel (23) die Summen

$$B = \sum_{k=1}^{m} x'_k h_k \text{ und } A = \sum_{k=1}^{m} x'^2_k h_k.$$

Produkt $x'_k \cdot x'_k h_k$
⑤ = ③ · ④
24 = (−2) · (−12)
17
0
6
4
18
69 = $\sum_{k=1}^{m} x'^2_k h_k = A$

Dazu schlagen wir die Tabelle zum Beispiel 9 (13) (↗ S. 89) auf, lesen die Summe B direkt ab ($B = \sum x'_k h_k = -15$) und brauchen für die Summe A nur noch die letzte Spalte auszufüllen, indem wir das Produkt aus den beiden voranstehenden Spalten bilden und die Produkte summieren.
Nach (23) folgt:

$$s = 2\sqrt{\frac{1}{49}\left(69 - \frac{(-15)^2}{50}\right)} \qquad s = 2{,}29\ \mathrm{s}; \qquad s \approx 2{,}3\ \mathrm{s}$$

$$\bar{x} \pm s = (43{,}9 \pm 2{,}3)\ \mathrm{s}$$

Anmerkung: Der Unterschied zu dem aus der Verteilung *ohne* Klassenbildung berechneten Wert für s ist minimal. Hier ist $\bar{x} = 43{,}9$ s zu verwenden, das arithmetische Mittel also, das sich aus der Häufigkeitsverteilung *mit* Klasseneinteilung ergab. Es sei jedoch an dieser Stelle noch einmal darauf hingewiesen, daß das Verfahren des angenommenen Mittelwertes auch bei Häufigkeitsverteilungen ohne Klasseneinteilung eingesetzt werden kann.

Bei der Berechnung von s mit dem Verfahren des angenommenen Mittelwerts sind also die ersten vier Teilschritte genau so auszuführen, wie auf S. 89 angegeben; es folgt dann:

5. Multiplizieren von x'_k mit $x'_k h_k$ (also der Spalten ③ und ④). Diese Produkte sind stets positiv.
6. Summieren der Produkte $x'^2_k h_k$. Das führt auf A.
7. Einsetzen der Summen $A = \sum x'^2_k h_k$ und $B = \sum x'_k h_k$ (also der Summen von Spalte ⑤ bzw. ④) in Formel (23).

● **114*** Berechnen Sie die Standardabweichung
a) für die acht Meßwerte der Zufallsgröße „Länge" aus Aufgabe 88 b) (↗ S. 86),
b) der Punktwerte, die die 20 am Mathematik-Olympiade-Wettbewerb beteiligten Schüler erzielten [Urdaten in Aufgabe 36 (↗ S. 38)]!

● **115*** Darf die Standardabweichung für die in der Tabelle auf der Seite 47 dargestellte Häufigkeitsverteilung zum stetigen Merkmal „Nutentiefe produzierter Wellen" berechnet werden?

● **116*** Berechnen Sie die Standardabweichung zur Zufallsgröße Y „Arbeitszeit des Lehrlings B je Steckermontage"
a) unter Verwendung der Tabelle zur Lösung der Aufgabe 26 (↗ S. 124) – Verteilung ohne Klassenbildung – und
b) unter Verwendung der Tabelle zur Lösung der Aufgabe 95 (↗ S. 135), also aus der Tabelle mit Klasseneinteilung, unter Anwendung des Verfahrens des angenommenen Mittelwerts!
Vervollständigen Sie dazu einfach die vorliegende Tabelle, und setzen Sie die entsprechenden Summen in Formel (23) ein!
c) Vergleichen Sie beide Ergebnisse!

● **117*** Berechnen Sie die Standardabweichung der Verteilung
a) zum Merkmal „Anzahl der Fahrzeuge ..." (ohne Klasseneinteilung; ↗ Tabelle auf S. 57f.),
b) zum gleichen Merkmal, aber *mit* Klasseneinteilung (↗ ebenda),
c) zum Merkmal „Weiten im Ballweitwurf" (↗ Tabelle auf S. 63),
d) zum Merkmal „Familienstand" für die Bevölkerung der DDR (↗ Tabelle auf S. 40)!

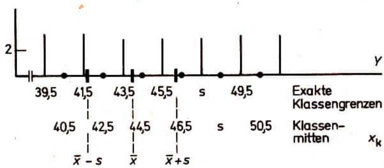
● **118*** Geben Sie das Intervall der einfachen Streuung

$\bar{x} - s \ldots \bar{x} + s$ an

a) zu Aufgabe 116a), **b)** 116b), **c)** 117a), **d)** 117c)!

Die für eine Häufigkeitsverteilung maßgebenden Kenngrößen, also der geeignete Mittelwert und das zugehörige Streuungsmaß, unter Umständen auch noch ein weiterer Mittelwert, werden in die graphische Darstellung der Verteilung (Linienzug, Histogramm, Streckendiagramm oder auch Summenkurve) eingetragen. Die Kennzeichnung erfolgt auf der Merkmalsachse.

■ **B 9 (20)** Merkmal X „Arbeitszeit des Lehrlings A …"
Wir berechneten für die Häufigkeitsverteilung nach Klassenbildung $\bar{x} = 43{,}9$ s; $s = 2{,}3$ s. Bild 35 (↗ S. 56) zeigt das Histogramm dieser Verteilung.
Wir geben hier (aus Platzgründen) nur die Merkmalsachse noch einmal an.
Die Lage von $\bar{x}$, von $\bar{x} - s$ sowie von $\bar{x} + s$ wurde eingezeichnet (Bild 76).



$\bar{x} - s = 41{,}6$ s

$\bar{x} + s = 46{,}2$ s

Bild 76

Größte Sorgfalt beim Eintragen der Werte ist angebracht.

● **119** Nehmen Sie die entsprechenden Eintragungen vor in den Bildern
a) 36, S. 56 (Histogramm zum Merkmal Y „Arbeitszeit des Lehrlings B …"),
b) 87, S. 128 (Linienzug zum Merkmal „Mathematikleistung"),
c) 90, S. 131 (Darstellung der Summenkurve zum gleichen Merkmal),
d) 39, S. 57 (Streckendiagramm zum Merkmal „Anzahl der Fahrzeuge …")!

Gute Vergleichsmöglichkeiten zwischen den Leistungen mehrerer Lehrlinge oder anderer Adressaten (für die Verteilung als Ganzes, speziell aber für Mittelwert und Streuung) ergeben sich aus Zusammenstellungen, wie wir sie in den Bildern 35 und 36 vornahmen.

Zusammenfassung des Abschnitts 2.3.

Wesentlicher Bestandteil der **Datenauswertung** ist das Berechnen statistischer Kenngrößen – insbesondere von Mittelwerten und Streuungsmaßen – aus den erfaßten oder aufbereiteten Daten. Die Kenngrößen *charakterisieren* die Häufigkeitsverteilung, Mittelwerte deren Lage, Streuungsmaße deren Form (↗ Bild 77).
Verteilung I: Normal mit *niedrigem* Mittelwert und *geringer* Streuung
Verteilung II: Normal mit *hohem* Mittelwert und *starker* Streuung

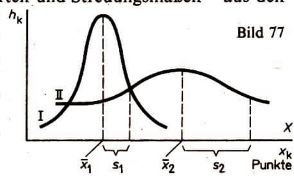


Bild 77

Übersicht über die verschiedenen Mittelwerte
(in Abhängigkeit von Datenart und Verteilungsform)

Datenart	Form der Häufigkeitsverteilung Eingipflig symmetrisch	Eingipflig nicht symmetrisch	Mehrgipflig
Meßwerte	Arithm. Mittel $\bar{x}$ Zentralwert Z Modalwert D	Zentralwert Z Modalwert D Bei leichter Schiefe: Arithm. Mittel $\bar{x}$	Modalwerte D_i Zentralwert Z
Rangdaten	Zentralwert Z Modalwert D	Zentralwert Z Modalwert D	Modalwerte D_i Zentralwert Z
Kategorien	Modalwert D	Modalwert D	Modalwerte D_i

Formelzusammenstellung

Mittelwerte

Arithmetisches Mittel

für n Zahlen (Meßwerte) — für Häufigkeitsverteilung (mit oder ohne Klassenbildung)

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \quad (7\text{a}) \qquad \bar{x} = \frac{1}{n}\sum_{k=1}^{m} x_k h_k \quad (9\text{a})$$

über das Verfahren des angenommenen Mittelwerts

$$\bar{x} = x_a + \frac{d}{n}\sum_{k=1}^{m} x'_k h_k \quad (11\text{a}) \qquad \text{mit} \quad x'_k = \frac{x_k - x_a}{d} \quad (10)$$

Gewogenes arithmetisches Mittel

$$\bar{x}_g = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_l n_l}{n_1 + n_2 + \ldots + n_l}, \quad (15)$$

wo die $x_1, x_2, \ldots x_l$ meist Mittelwerte von Einzelstichproben sind

Zentralwert

$$Z = x_{ug} + d \cdot \frac{\frac{n}{2} - sh_u}{h_z} \quad (16)$$

Streuungsmaße

Variationsweite $w = x_{max} - x_{min}$ (18) (außer für Kategorien stets einsetzbar)

Durchschnittliche Abweichung für n Zahlen (Meßwerte)

$$e = \frac{1}{n}\sum_{k=1}^{n} |x_k - \bar{x}| \quad (19)$$

(einsetzbar dort, wo $\bar{x}$ sinnvoll)

Standardabweichung (ebenfalls einsetzbar dort, wo $\bar{x}$ sinnvoll ist)

für n Zahlen (Meßwerte) — für Häufigkeitsverteilung (mit oder ohne Klassenbildung)

$$s = \sqrt{\frac{1}{n-1} \sum_{k=1}^{n} (x_k - \bar{x})^2} \tag{21a}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{k=1}^{m} (x_k - \bar{x})^2 h_k} \tag{22}$$

über das Verfahren des angenommenen Mittelwertes

$$s = d \sqrt{\frac{1}{n-1} \left(A - \frac{B^2}{n} \right)} \tag{23}$$

mit $A = \sum_{k=1}^{m} x_k'^2 h_k$ und $B = \sum_{k=1}^{m} x_k' h_k$

Zum Schließen in der Statistik (Kurzer Abriß)

3

Das 2. Kapitel vermittelte Kenntnisse und Erkenntnisse zur beschreibenden Statistik. Mit dem Erfassen, Aufbereiten, Darstellen und Auswerten von Daten haben wir eine notwendige Grundlage geschaffen. Weitere Gebiete der Statistik bauen darauf auf. Bedeutsam für viele Wissenschaften und für die Praxis ist die **Prüfstatistik.** Dabei geht es um das Schätzen von Parametern und um das Aufstellen und Prüfen wissenschaftlicher Hypothesen, das heißt um das Bestätigen oder Ablehnen von begründeten Vermutungen über das Untersuchungsergebnis. Der Schluß von der Stichprobe auf die Grundgesamtheit spielt dabei eine wichtige Rolle, und die Wahrscheinlichkeitsrechnung wird zum tragenden Element.
Dieses umfangreiche Gebiet Prüfstatistik läßt sich in den Schülerkollektiven nicht umfassend behandeln, wohl aber ist die Darlegung von einigen Teilproblemen möglich, die den Blick des Lernenden dafür weiten, was eigentlich alles zur Statistik gehört. Dafür stehen zwei Bücher zur Verfügung, die auch dem Schüler (in Bibliotheken oder Buchhandlungen) zugänglich und für ihn verständlich sind ([26] und [25]):

MAIBAUM, Gert: Wahrscheinlichkeitsrechnung, Nr. 63 der Mathematischen Schülerbücherei. Volk und Wissen Volkseigener Verlag, Berlin 1980;
LOHSE, Heinz/Rolf LUDWIG: Prüfstatistik. Ein programmierter Lehrgang. VEB Fachbuchverlag, Leipzig 1982.

Das erste bringt eine ausgezeichnete Einführung in die Grundlagen der Wahrscheinlichkeitsrechnung und enthält zahlreiche interessante Beispiel- und Übungsaufgaben; das zweite schließt unmittelbar an das hier in der „Elementaren Statistik" Dargestellte an, ist auf Grund seiner programmierten Form leicht verständlich und ermöglicht die Auswahl des einen oder anderen in sich abgeschlossenen Teilgebietes (z. B. t-Test zum Prüfen von Mittelwertunterschieden oder KOLMOGOROV-SMIRNOV-Test zum Vergleich zweier empirischer Verteilungen).
Hier kann mit den Abschnitten 3.1. und 3.2. nur ein ganz kurzer Einblick in die Prüfstatistik und deren Grundlagen vermittelt werden.

3.1. Zum Begriff der Wahrscheinlichkeit

Wir wissen, was eine „Zufallsgröße" ist. Dieser Ausdruck beherrscht das ganze Buch. Wir haben absolute und relative Häufigkeiten von Merkmalsausprägungen erfaßt in allen möglichen statistischen Erhebungen, aber auch im Zusammenhang mit Experimenten (↗ Würfelversuch in B 5, S. 12f., und B 23, S. 38f.). Wir haben über das dialektische Verhältnis gesprochen, in dem Zufall und Notwendigkeit zueinander stehen (↗ Abschnitt 1.2., S. 11ff.).
Jetzt soll der Wahrscheinlichkeitsbegriff präzisiert und definiert werden. Die Wahrscheinlichkeitsrechnung untersucht die Gesetzmäßigkeiten **zufälliger Ereignisse.**

▶ **D 16 Ein Versuchsergebnis, das eintreten kann, aber nicht unbedingt einzutreten braucht, dessen Eintreten also mit einer gewissen Wahrscheinlichkeit erfolgt, heißt ein *zufälliges Ereignis A*.**

■ **B 56** Beispiele für zufällige Ereignisse

a) Augenzahl „1", „2", ..., „6" beim Werfen eines Spielwürfels
b) „Fehlschuß", Ringe „1", „2", ..., „10" beim Schießen auf eine Zehnerscheibe
c) „Knabe", „Mädchen" bei der Geburt
d) „Fehlerfrei", „Ausschuß" beim Prüfen der Qualität produzierter Fernsehbildröhren

● **120 (1)*** Ein Versuch bestehe im Hochwerfen einer Münze. Welche zufälligen Ereignisse treten dabei auf?

● **120 (2)** Werfen Sie eine Münze 40mal etwa einen Meter hoch, und registrieren Sie, wie oft das zufällige Ereignis „Zahl" und wie häufig das Ereignis „Wappen" eintritt! Setzen Sie diese absoluten Häufigkeiten ins Verhältnis zur Anzahl der Versuche (Münzwürfe)!

Wir erinnern uns des Beispiels 5, S. 12f. Dort hielten wir jeweils die Häufigkeit h fest, mit der die Augenzahl „6" beim 50maligen, beim 100maligen usw. Werfen mit einem Spielwürfel erscheint. Dann berechneten wir die **relative Häufigkeit** $\frac{h}{n}$ des **zufälligen Ereignisses** A.

▶ **D 17 Unter der *relativen Häufigkeit* $\frac{h(A)}{n}$ *des zufälligen Ereignisses A* verstehen wir das Verhältnis der absoluten Häufigkeit des Eintretens von A zur Anzahl *n* der durchgeführten Versuche.**

● **121** **a)** Vergleichen Sie diese Definition mit D 6, S. 33, und lesen Sie die dort folgende Seite!
b)* Versuch A: 60maliges Werfen eines Spielwürfels
Es trat 11mal die Augenzahl „3" auf. Wie groß ist die relative Häufigkeit des zufälligen Ereignisses „3"?

● **120 (3)*** Berechnen Sie die relative Häufigkeit des zufälligen Ereignisses „Z" und die des zufälligen Ereignisses „W" aus den unter Aufgabe 120 (2) gewonnenen Werten! Was stellen Sie für $\frac{h(Z)}{n} + \frac{h(W)}{n}$ fest?

Für ein abgeschlossenes Ereignissystem $A_1, \ldots, A_m$ gilt also

$$\sum_{k=1}^{m} \frac{h(A_k)}{n} = 1, \tag{3a}$$

eine Beziehung, die der Formel (3), S. 33, entspricht.

Die relative Häufigkeit $\frac{h(A)}{n}$ eines zufälligen Ereignisses A ist u. a. gekennzeichnet durch folgende Eigenschaft:

Die relative Häufigkeit eines zufälligen Ereignisses liegt stets im Intervall zwischen 0 und 1.

$$0 \leqq \frac{h(A)}{n} \leqq 1 \tag{24}$$

Das läßt sich leicht bestätigen: $h(A)$ kann mindestens gleich 0, höchstens gleich n sein, also

$$0 \leqq h(A) \leqq n \qquad |:n \neq 0$$

$$0 \leqq \frac{h(A)}{n} \leqq 1$$

Anmerkung: $\frac{h(A)}{n}$ kann auch die Randwerte 0 oder 1 annehmen; so kann beim 10maligen Werfen einer Münze stets „Zahl" erscheinen. Dann ist $\frac{h(Z)}{10} = 1$.

Von der relativen Häufigkeit $\frac{h(A)}{n}$ eines zufälligen Ereignisses A gelangen wir zur **Wahrscheinlichkeit** $P(A)$ des zufälligen Ereignisses A, indem wir n sehr groß werden lassen.

► **D 18 Die *Wahrscheinlichkeit* $P(A)$ für das Eintreten des zufälligen Ereignisses A ist gleich dem Verhältnis aus der absoluten Häufigkeit $h(A)$ des Eintretens von A zur Anzahl n der Versuche, wenn n gegen Unendlich strebt.**

$$P(A) = \frac{h(A)}{n} \quad \text{bei} \quad n \to \infty \tag{25}$$

Das ist die **statistische Definition** der Wahrscheinlichkeit.

■ **B 5 (2)** Je mehr Würfel wir nach Beispiel 5 (1), S. 12f., durchführen (…; $n = 500$; $n = 1000$; $n = 5000$), um so mehr nähert sich die jeweilige relative Häufigkeit (…; 0,164; 0,168; 0,167) dem *konstanten Wert* $\frac{1}{6} = 0{,}1666\ldots$, der *Wahrscheinlichkeit* des Auftretens der Augenzahl „6" beim Würfeln.

Allgemein: Je mehr Versuche einer Versuchsserie wir durchführen, desto mehr stabilisiert sich das Verhältnis $\frac{h(A)}{n}$, und zwar in Richtung auf den Wert $P(A)$. Wir sprechen von der *relativen Stabilität* der relativen Häufigkeit des zufälligen Ereignisses A. *Bei einer endlichen Anzahl von Versuchen ist die relative Häufigkeit* $\frac{h(A)}{n}$ *also stets ein Näherungswert für die Wahrscheinlichkeit* $P(A)$ *des zufälligen Ereignisses* A.

$$P(A) \approx \frac{h(A)}{n} \quad \text{bei endlichem } n \tag{25a}$$

● **120 (4)* a)** Wie groß ist die Wahrscheinlichkeit des Eintretens des Ereignisses „Zahl" und des Ereignisses „Wappen" beim Hochwerfen einer Münze?
b) Bestimmen Sie $P(Z) + P(W)$.

Aufgrund der Beziehung $P(A) = \frac{h(A)}{n}$ bei $n \to \infty$ wird jedem zufälligen Ereignis A eine reelle Zahl $P(A)$ zugeordnet, eben die Wahrscheinlichkeit für das Eintreten von A:

Analog zu der genannten Eigenschaft der relativen Häufigkeit $\frac{h(A)}{n}$ [Formel (24)] gilt für die Wahrscheinlichkeit $P(A)$ folgende Eigenschaft:

$$0 \leqq P(A) \leqq 1. \tag{26}$$

■ **B 57** Versuch: Einmaliges Werfen eines Spielwürfels

a) Die Wahrscheinlichkeit, beim Werfen eines Spielwürfels eine „3" zu erhalten, ist $P(„3") = \frac{1}{6} = 0{,}1666\ldots$

b) Die Wahrscheinlichkeit, eine der Zahlen „1", „2", ..., „6" (Ereignis $A_{1\,\text{bis}\,6}$) zu werfen, ist $P(A_{1\,\text{bis}\,6}) = 1$.

Es sei betont, daß wir uns bei der Verwendung des Wahrscheinlichkeitsbegriffs in der Mathematik und in der Statistik nicht mit den Vorstellungen der Alltagssprache belasten dürfen. Dort hat „wahrscheinlich" die Bedeutung von „vermutlich", „vielleicht". (Beispiel: „Ich komme wahrscheinlich später.")

Hier in der Statistik müssen wir uns an exakte Definitionen halten. *Eine* Definition zum Begriff „Wahrscheinlichkeit" haben wir kennengelernt (D 18). Es gibt aber noch weitere. Für den systematischen Aufbau der mathematischen Statistik bildet die axiomatische Definition des sowjetischen Mathematikers KOLMOGOROV die Grundlage. Für praktische Belange nützlich ist die **klassische Definition** der Wahrscheinlichkeit.

► **D 19 Die *Wahrscheinlichkeit* für das Eintreten des zufälligen Ereignisses A in einem Versuch ist gleich dem Verhältnis aus der Zahl g der für das Ereignis A *günstigen* Resultate zur Anzahl n aller *möglichen* Resultate des Versuchs.**

$$P(A) = \frac{g}{n} \tag{27}$$

Manchem fällt es anfangs schwer, die hier gemeinte Bedeutung von „günstig" und „möglich" zu erfassen. Wir geben dazu ein Beispiel.

■ **B 58** Versuch: Einmaliges Werfen eines Spielwürfels

Gesucht: P („3")

Auf dem Würfel befinden sich sechs verschiedene Augenzahlen, eine davon liegt nach dem Wurf oben, *jede* der sechs ist möglich (also: $n = 6$, Zahl der möglichen Resultate). *Eine* davon ist die gesuchte „3" (also: $g = 1$, Zahl der günstigen Resultate).

Nach (27): $P(A) = \frac{1}{6} = 0{,}1666\ldots$ (damit Bestätigung des schon unter B 57a) Gefundenen)

Der Wahrscheinlichkeitsbegriff findet vielfältige **Anwendungen** in der Statistik. Auf eine wollen wir im folgenden hinweisen.

Wir erinnern uns (↗ Abschnitt 2.3.4., S. 102ff.), daß zur Kennzeichnung einer Verteilung, die annähernd normal ist und der intervallskalierte Daten zugrunde liegen, Mittelwert $\bar{x}$ und Standardabweichung s dienen. Wir erinnern uns ferner (↗ Abschnitt 2.2.4., S. 59ff.), daß Flächen unter der Normalkurve unter Zuhilfenahme der Tabelle auf Seite 61 leicht berechnet werden können. Bringen wir beides in Verbindung miteinander, so läßt sich angeben, mit welcher Wahrscheinlichkeit die Zufallsgröße X einen Wert zwischen $\bar{x} - s$ und $\bar{x} + s$ annimmt, oder, mit anderen Worten,

welcher Anteil aller Meßwerte der untersuchten empirischen Häufigkeitsverteilung im Intervall $[\bar{x} - s;\ \bar{x} + s]$ *liegt.*
Das haben wir schon berechnet, und zwar unter B 33 (S. 62) für die standardisierte Normalverteilung. Es ergab sich:
Die Fläche unter der standardisierten Normalkurve zwischen $x = -1$ und $x = +1$ beträgt 68,2%. Da die standardisierte Normalverteilung durch $\mu\ (\triangleq \bar{x}) = 0$ und $\sigma(\triangleq s) = 1$ gekennzeichnet ist (↗ E 18, S. 61), entspricht diesem Intervall $[-1;\ +1]$ für eine allgemeine Normalverteilung gerade unser Intervall $[\bar{x} - s;\ \bar{x} + s]$. Das heißt: Bei jeder normalen empirischen Häufigkeitsverteilung liegen *im Intervall* $[\bar{x} - s;\ \bar{x} + s]$ *rund 68,2% aller Beobachtungswerte* (Bild 78).
Wir überprüfen das an unserer Beispielaufgabe zum Merkmal X „Arbeitszeit des Lehrlings A ...“. Wir wissen, daß hier nur eine angenäherte Normalverteilung vorliegt.

■ **B 9 (21)** Wir berechneten aus der ursprünglichen Häufigkeitstabelle $\bar{x} \pm s = (44{,}0 \pm 2{,}3)$ s (↗ B 9 (18), S. 107). Der Bereich der einfachen Streuung erstreckt sich also von 41,7 s bis 46,3 s.
Wir entnehmen der Häufigkeitstabelle (z. B. der Tabelle auf S. 107), wie viele Beobachtungswerte in diesem Intervall liegen. Es sind $6 + 11 + 9 + 9 + 3 = 38$. Das ergibt nach (27) $P(A) = \frac{38}{50} = 0{,}76 \triangleq 76\%$.
Es bestätigt sich, daß mit dieser Häufigkeitsverteilung nur eine angenäherte Normalverteilung gegeben ist.

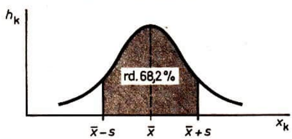


Bild 78

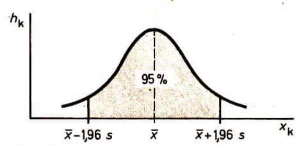


Bild 79

● **122*** Berechnen Sie die Anzahl der Beobachtungswerte, die im Intervall $[\bar{x} - s;\ \bar{x} + s]$ liegen für die empirische Häufigkeitsverteilung

a) zum Merkmal Y „Arbeitszeit des Lehrlings B ...“ (benötigte Werte unter der Lösung zu Aufgabe 118 b, S. 139, und aus der Tabelle unter der Lösung zu Aufgabe 26, S. 124);

b) zum Merkmal „Leistung im Mathematik-Olympiade-Wettbewerb“ $(\bar{x} \pm s) = (26{,}2 \pm 3{,}0)$ Punkte, Häufigkeitstabelle ↗ Tabelle unter der Lösung zu Aufgabe 36b), S. 125f.!

In Aufgabe 60a), S. 62, berechneten wir den Flächenanteil unter der standardisierten Normalkurve im Intervall $[-1{,}96;\ +1{,}96]$ und erhielten 95,0%.
$[-1{,}96;\ +1{,}96]$ ist gleichbedeutend mit dem Intervall $[\bar{x} - 1{,}96\ \text{s};\ \bar{x} + 1{,}96\ \text{s}]$ bei einer empirischen Häufigkeitsverteilung (↗ Bild 79).
Diese 95% heißen **Vertrauenswahrscheinlichkeit** (oder statistische Sicherheit), die restlichen 5% (rechts und links von der gerasterten Fläche) stellen die sogenannte **Irrtumswahrscheinlichkeit** dar. Bei den meisten Untersuchungen verwendet man eine Vertrauenswahrscheinlichkeit von 95%, nur in Ausnahmen (z. B. in der Medizin, Pharmakologie usw.) ist eine Vertrauenswahrscheinlichkeit von 99% nötig.

3.2. Der Schluß von der Stichprobe auf die Grundgesamtheit

In D 2 (↗ S. 15) definierten wir den Begriff **Grundgesamtheit** als eine nichtleere Menge gleichartiger Objekte, an denen eine oder mehrere Zufallsgrößen untersucht werden.

■ **B 59 (1)** Im VEB Fernsehgerätewerk Staßfurt werden große Serien an Fernsehbildröhren hergestellt. Die gesamte Monatsproduktion eines Typs bildet die Grundgesamtheit. In der statistischen Qualitätskontrolle werden u. a. folgende Zufallsgrößen untersucht:
„Güte der Röhre", „Lebensdauer der Röhre".

Die Grundgesamtheit ist meist sehr groß. Sämtliche Objekte der Grundgesamtheit in die Untersuchung einzubeziehen, also eine Vollerhebung vorzunehmen, ist unökonomisch (kostet einen unvertretbaren hohen Aufwand an Arbeitskraft, Zeit und Geld) oder zuweilen unmöglich.

■ **B 59 (2) a)** Beispiel zu „unökonomisch": Zufallsgröße „Güte der Röhre". Es würde viel zu viel Kosten und Zeit in Anspruch nehmen, würde man jede der gefertigten Fernsehbildröhren auf deren Qualität prüfen.
b) Beispiel zu „unmöglich": Zufallsgröße „Lebensdauer der Röhre". Würde man *jede* Röhre auf deren Lebensdauer untersuchen, sie also in der Qualitätskontrolle so lange laufen lassen, bis sie ihren Dienst versagt, so könnte man keine Fernsehbildröhren mehr ausliefern.

Wir sind also darauf angewiesen, eine *gewisse Auswahl von Untersuchungsobjekten*, nur eine Teilmenge der Menge aller Objekte in die Untersuchung einzubeziehen, m. a. W., eine Teilerhebung vorzunehmen.
Diese Teilmenge bezeichnen wir als **Stichprobe** S.

▶ **D 20 Die für eine bestimmte Untersuchung aus der Grundgesamtheit G ausgewählte Teilmenge von Objekten heißt *Stichprobe* S aus der Grundgesamtheit G.**

S ist echte Teilmenge von G; $S \subset G$ (↗ Bild 80).
Die Anzahl der Elemente der Grundgesamtheit wird mit N bezeichnet, die Anzahl der Elemente der Stichprobe mit n. Es gilt $n < N$.

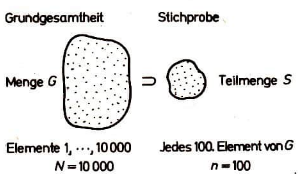
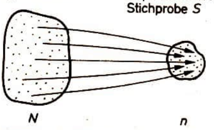


Bild 80

■ **B 59 (3)** Nehmen wir an, bei der Produktion von Fernsehbildröhren wird jede 10. Bildröhre genau geprüft. Dann ist jede 10. Röhre Element der Stichprobe S.

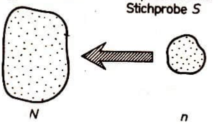
Im Kapitel 2. hatten wir es sehr oft mit Stichproben zu tun, auch wenn wir sie nicht immer als solche bezeichnet haben. Wir untersuchten *nicht alle* Montagezeiten, sondern nur eine Auswahl von 50; wir zählten die Fahrzeuge, die eine ampelgeregelte Kreuzung passieren, nicht an allen Abenden, sondern nur an drei zufallsmäßig ausgewählten; wir bestimmten die Masse nicht von allen Schweinen, sondern nur von 60, die die Gesamtmenge der Schweine einer LPG oder eines Territoriums möglichst gut *repräsentieren* (für die Grundgesamtheit etwas Wesentliches aussagen) sollen. Diese Auswahl der Elemente der Stichprobe kann nicht *irgendwie* erfolgen, sondern muß möglichst *zufallsmäßig* geschehen mit dem Ziel, zu **repräsentativen Stichproben** zu gelangen. Die repräsentative Stichprobe muß ein wirkliches Abbild, ein getreues Modell der Grundgesamtheit sein.

► **D 21 Unter einer *repräsentativen Stichprobe* der Grundgesamtheit G verstehen wir eine Stichprobe, die ein getreues Modell der Grundgesamtheit ist. Wir erhalten sie durch *zufallsmäßige Entnahme* von Elementen der Grundgesamtheit nach einem Auswahlplan** (↗ Bild 81).



Zufallsmäßige Entnahme nach einem Auswahlplan

Bild 81



Schluß von der Stichprobe auf die Grundgesamtheit

Bild 82

Der Auswahlplan kann besagen, daß man jedes 10. oder jedes 100. Element in die Stichprobe aufnimmt (systematische Auswahl) oder jeden 2. Schüler aus dem Klassenbuch (ebenfalls systematische Auswahl) oder jeden Sportler, dessen Name mit „L" beginnt (Buchstabenauswahl). Wichtig ist dabei stets, daß jedes Element von G die *gleiche Chance* hat, in die Stichprobe aufgenommen zu werden. Sonst liegt keine Zufallsauswahl vor.

Mit der Erzeugung einer repräsentativen Stichprobe verfolgen wir das Ziel, *von* Erkenntnissen, die wir in der überschaubaren und prüfbaren *Stichprobe* gewinnen, *auf* die Gegebenheiten der *Grundgesamtheit zu schließen* (↗ Bild 82).

Wir schließen vom Teil aufs Ganze. Aber: Nur wenn wir von einer repräsentativen Stichprobe ausgehen, können wir zuverlässige Aussagen über die Grundgesamtheit gewinnen.

■ **B 59 (4)** Jede 10. Fernsehbildröhre wird genau geprüft. Es wird also nach der systematischen Auswahl verfahren. Die so zufallsmäßig entnommenen Elemente der Grundgesamtheit bilden die repräsentative Stichprobe. Von der Qualität dieser Elemente (der Stichprobe) wird geschlossen auf die Qualität aller produzierten Bildröhren (Grundgesamtheit).

Dieser Schluß von der Stichprobe auf die Grundgesamtheit ist stets mit einem gewissen Risiko behaftet, er kann nie mit 100%iger Sicherheit erfolgen.

■ **B 59 (5)** Es kann sein, daß jede 10. Bildröhre in Ordnung ist, während die 5., 17., 33., 58., ... nicht den Qualitätsnormen entsprechen.

Die Statistik bietet aber nun gerade die Möglichkeit, dieses *Risiko berechenbar zu machen und in engen Grenzen zu halten*, so daß wir mit einer *hohen Vertrauenswahrscheinlichkeit* und damit *geringer Irrtumswahrscheinlichkeit* arbeiten, prüfen oder untersuchen können.

● **123*** Eine Thüringer PGH stellt täglich 50 Thermofenster her. Diese haben aufgrund einer luftverdünnten Schicht zwischen den zwei Glasscheiben besonders gute Wärmedämmeigenschaften. Für die ausführliche Qualitätskontrolle eines Fensters werden 30 min benötigt. Warum können nicht alle produzierten Fenster genau geprüft werden?

● **124** Eine Gruppe von Schülern aus der DDR wird zur Mathematik-Olympiade nach Moskau delegiert. Sind die n Schüler eine repräsentative Stichprobe der Anzahl N aller Schüler der DDR?

Diese Aufgaben machen uns auf etwas aufmerksam, das es zu bedenken gilt. Nicht selten werden in der Statistik die Begriffe Grundgesamtheit und Stichprobe nicht auf die Untersuchungsobjekte selbst bezogen, sondern *auf die an den Objekten erfaßten Merkmalswerte*. Zu jeder Grundgesamtheit von Objekten können entsprechend den verschiedenen Merkmalen mehrere solcher Grundgesamtheiten definiert werden.

■ **B 60** An einer repräsentativen Stichprobe von $n = 60$ Schülern aus der Grundgesamtheit aller Schüler der Schule ($N = 1200$) kann ich sehr viele verschiedene Zufallsgrößen untersuchen:
- die Schülermeinung zur Pausendisziplin;
- die Körpergröße der Schüler;
- ihre Einstellung zur Jugendtouristik;
- die Anzahl der Stunden, die sie wöchentlich am Fernsehbildschirm zubringen.

Bleiben wir beim zweiten Merkmal: Zur Körpergröße erhalten wir also 60 Meßwerte, und diese stellen die Stichprobe aus der Grundgesamtheit der möglichen 1200 Meßwerte dar. Die Begriffe Grundgesamtheit und Stichprobe beziehen wir also hier auf das Merkmal „Körpergröße“.
Und gleiches gilt für die anderen zu untersuchenden Merkmale. Aus der *einen* Grundgesamtheit von Objekten sind also vier Grundgesamtheiten *bezüglich der vier zu untersuchenden Merkmale* geworden.
So haben wir das eigentlich schon im gesamten Buch gehandhabt. Beim Beispiel B 9 (1) (↗ S. 16, ↗ auch B 9 (2), S. 30) waren die Steckermontagen die Untersuchungsobjekte. An diesen untersuchten wir die Arbeitszeit (in s), die der Lehrling A (bzw. B) für jede Montage benötigt. Wir gewannen eine Stichprobe von je 50 Arbeitszeiten (in s) aus der Menge aller Arbeitszeiten (der Grundgesamtheit) des Lehrlings. Und von den Einsichten, den Erkenntnissen, die wir anhand dieser Stichprobe gewinnen, schließen wir auf die Arbeitszeit, die er allgemein, die er z. B. im Mittel für 1000 Steckermontagen benötigt.

Grundgesamtheit und Stichprobe stehen in einem engen Verhältnis zueinander. Wir haben nicht nur zwischen den Umfängen n und N zu unterscheiden, sondern auch zwischen $\bar{x}$ und μ, zwischen s und σ (↗ Bild 83, S. 120).
Die aus der Stichprobe ermittelten Kenngrößen sind *Schätzwerte* für die entsprechenden Parameter der Grundgesamtheit, $\bar{x}$ ist Schätzwert für μ und s Schätzwert für σ.

Wichtige Aufgaben der Prüfstatistik bestehen nun darin,
- von den Kenngrößen der repräsentativen Stichprobe zu schließen auf die Parameter der Grundgesamtheit **(Parameterschätzung)**;

	Grund-gesamtheit	Stichprobe
Bezeichnung	G	S
Umfang	N	n
Statistische Maßzahlen	Parameter	Kenngrößen
Mittelwert	μ	$\bar{x}$
Standardabweichung	σ	s

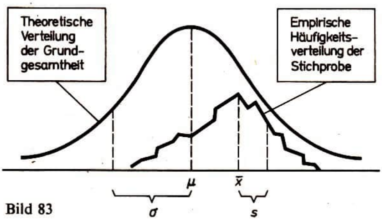


Bild 83

- von den Kenngrößen zweier repräsentativer Stichproben zu prüfen, ob sie einer oder zwei verschiedenen Grundgesamtheiten angehören, d. h. mit anderen Worten, ob die Kenngrößen als gleich angesehen werden können oder als signifikant (bedeutend) verschieden voneinander **(Hypothesenprüfung).**

Wir wollen uns hier nur mit der erstgenannten Aufgabe beschäftigen. Entnehmen wir einer (möglichst großen) Grundgesamtheit mehrere Stichproben, so werden deren Kenngrößen (z. B. Mittelwerte $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_l$) variieren, sie bilden ihrerseits eine Häufigkeitsverteilung.

► **E 28** Die Verteilung der Mittelwerte $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_l$ wird **Stichprobenverteilung des Mittelwerts** $\bar{x}$ genannt.

Für diese Stichprobenverteilung gilt nun ein wichtiger Satz:
Die Verteilung der Mittelwerte $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_l$ folgt unter der Voraussetzung eines genügend großen l stets einer ***Normalverteilung*** *mit dem Mittelwert μ und der Standardabweichung $\sigma_{\bar{x}}$.*
Der Schätzwert $s_{\bar{x}}$ der Standardabweichung $\sigma_{\bar{x}}$ der Stichproben-Mittelwerte – oft auch Standardfehler des Mittelwerts genannt – wird nach der Formel

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{28}$$

berechnet.

■ **B 61** Wir ermitteln den Schätzwert $s_{\bar{x}}$ der Standardabweichung der Stichprobenmittelwerte bei zwei Stichproben gleicher Streuung ($s_1 = s_2 = 5$ Einheiten), aber ungleichen Umfangs ($n_1 = 100$; $n_2 = 1024$).
Aus (28) ergibt sich

für Stichprobe 1: $s_{\bar{x}} = \dfrac{5}{10} = 0{,}5$ Einheiten,

für Stichprobe 2: $s_{\bar{x}} = \dfrac{2}{32} = 0{,}06$ Einheiten.

Wir erkennen: Je größer die Stichprobe, um so kleiner wird bei etwa gleicher Streuung der Standardfehler des Mittelwerts.

● **125** Berechnen Sie den Schätzwert der Standardabweichung der Stichproben-Mittelwerte

a) zu unserer Beispielaufgabe 9 (1) (Merkmal „Arbeitszeit des Lehrlings A ..."; $\bar{x} = 44{,}0$ s; $s = 2{,}3$ s);

b) zur Aufgabe mit dem Merkmal „Leistung im Mathematik-Olympiade-Wettbewerb", $\bar{x} = 26{,}2$ Punkte; $s = 3{,}0$ Punkte)!

Wir wissen: Das aus einer Stichprobe berechnete arithmetische Mittel $\bar{x}$ ist ein *Schätzwert* für den Mittelwert μ der betreffenden Grundgesamtheit. Zu diesem Schätzwert läßt sich nun ein Intervall angeben, das sich über die benachbarten Werte erstreckt und mit hoher Wahrscheinlichkeit auch den Mittelwert μ der Grundgesamtheit enthält.
Dieses Intervall um die Kenngröße, das den Parameter mit einschließen soll, heißt **Vertrauensintervall** oder **Vertrauensbereich.**

► **D 22 Ein *Vertrauensbereich* ist ein geschätztes Intervall, welches den wahren Wert des unbekannten Parameters mit vorgegebener Vertrauenswahrscheinlichkeit $(100 - \alpha)$ % überdeckt. Dabei ist α die Irrtumswahrscheinlichkeit.**

Die Vertrauensbereiche lassen sich angeben aufgrund der Berechnung von **Vertrauensgrenzen.** Das sind die sich bei der Vertrauensintervallschätzung ergebenden Grenzen für den unbekannten Wert eines Parameters.
Die Vertrauensgrenzen für den Mittelwert μ einer normalverteilten Grundgesamtheit werden aufgrund der Formeln

$$\bar{x} - z \cdot s_{\bar{x}} \quad (29\text{a}) \qquad \text{und} \qquad \bar{x} + z \cdot s_{\bar{x}} \quad (29\text{b})$$

bestimmt, wobei $\bar{x}$ der Mittelwert der untersuchten Stichprobe ist und z der jeweilige Standardwert der Normalverteilung für die entsprechende Vertrauenswahrscheinlichkeit $1 - \alpha$ (↗ die folgende Tabelle; vgl. auch mit S. 116).

Irrtumswahrscheinlichkeit α (%)	Vertrauenswahrscheinlichkeit $1 - \alpha$ (%)	z
5	95	1,96
1	99	2,58

Das **Vertrauensintervall für den Mittelwert μ einer normalverteilten Grundgesamtheit** ist dann – unter der Voraussetzung, daß die Stichprobe einen Umfang von $n > 40$ hat –

$$\left[\bar{x} - z \cdot \frac{s}{\sqrt{n}};\quad \bar{x} + z \cdot \frac{s}{\sqrt{n}}\right]. \tag{30}$$

Durch Veränderung der Irrtumswahrscheinlichkeit α und damit der Vertrauenswahrscheinlichkeit $(1 - \alpha)$ läßt sich festlegen, *wie sicher* die Aussage ist, daß das Vertrauensintervall den Parameter der Grundgesamtheit auch enthält.
Das wird an folgendem Beispiel deutlich.

■ **B 9 (22)** Für die Untersuchung zum Merkmal „Arbeitszeit des Lehrlings A" ergeben sich

bei Zugrundelegen der Irrtumswahrscheinlichkeit $\alpha = 5\,\%$ die Vertrauensgrenzen

$$\bar{x} - z \cdot \frac{s}{\sqrt{n}} = \left(44{,}0 - 1{,}96 \cdot \frac{2{,}3}{\sqrt{50}}\right) \text{s} = 43{,}4 \text{ s}$$

und

$$\bar{x} + z \cdot \frac{s}{\sqrt{n}} = 44{,}6 \text{ s},$$

bei Zugrundelegen der Irrtumswahrscheinlichkeit $\alpha = 1\,\%$ die Vertrauensgrenzen

$$\bar{x} - z \cdot \frac{s}{\sqrt{n}} = \left(44{,}0 - 2{,}58 \cdot \frac{2{,}3}{\sqrt{50}}\right) \text{s} = 43{,}2 \text{ s}$$

und

$$\bar{x} + z \cdot \frac{s}{\sqrt{n}} = 44{,}8 \text{ s}.$$

Wir können also sagen: Mit der Vertrauenswahrscheinlichkeit oder **statistischen Sicherheit** von 95 % wird der Parameter μ vom Konfidenzintervall [43,4 s; 44,6 s] überdeckt, mit der statistischen Sicherheit von 99 % vom Konfidenzintervall [43,2 s; 44,8 s].

Oder allgemein:

Je kleiner die Irrtumswahrscheinlichkeit, je größer also die statistische Sicherheit ist, desto größer ist das Vertrauensintervall für den betreffenden Parameter der Grundgesamtheit.

● **126** Berechnen Sie die Vertrauensbereiche unter Zugrundelegen der Vertrauenswahrscheinlichkeiten von 95 % und 99 % für die empirische Häufigkeitsverteilung zum Merkmal *Y* „Arbeitszeit des Lehrlings B ..."!

2. Mathematik, Physik, Geographie, Staatsbürgerkunde, ESP

4. *Normerfüllung (in %)*

C: $\frac{1\,040 \text{ St.}}{44 \cdot 20 \text{ St.}} \cdot 100\,\% = 118{,}2\,\%$;

D: 104,5; E: 114,6; F: 107,5

5. Richtig ist, die Produktion aller 6 Arbeiter in Stück (6000) durch die Normstückzahl (5340 St.) zu dividieren. Man erhält 112,4%.
Falsch wäre, das arithmetische Mittel der Normerfüllung (in %) zu berechnen.

6. **c)** Hier wird ein Einzelfall untersucht und keine Massenerscheinung.

9. Auf 914,3% bzw. auf 4006,4%

11. Die relative Häufigkeit 0,1666; auf drei Dezimalstellen gerundet 0,167

15. Bolzenlänge (mm), Hektarertrag (dt/ha), Normerfüllung (ohne Einheit, wird in Prozent angegeben), Niederschlagsmenge (mm/Zeitraum)

16.

Merkmal	*Merkmalsausprägungen*	*Art der Merkmalsausprägungen*
Geschlecht	männlich, weiblich	qualitativ
Lebensalter	0 bis 150 Jahre	quantitativ
Zahl der Ferkel je Wurf	4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16	quantitativ
Benzinverbrauch des Trabant	(3,5 l; 10 l)	quantitativ
Ausgeübter Beruf von Werktätigen	Dreher, Maler, Chemiefacharbeiter usw.	qualitativ

17. **a)** *Untersuchungsobjekte:* 40 Wellen mit Nut; *Zufallsgröße (Merkmal):* Nutentiefe in mm; *Meßinstrument:* Tiefenmeßschraube; *Merkmalsausprägungen:* [2,7 mm; 4,9 mm], gerundet auf Zehntelmillimeter; *Art der Merkmalsausprägungen:* quantitativ
b) Im Toleranzbereich 67,5% aller Werte

18. **a)** *Untersuchungsobjekte:* Schüler eines Jahrganges; *Merkmal* mit *Einheit:* Körpergröße, eigentlich cm (↗ S. 19 oben);

Merkmalsausprägungen: klein, mittelgroß, groß, sehr groß oder: K, M, G, S

b) Ein wirkliches Meßinstrument wird hierzu nicht verwendet; die Körpergröße wird geschätzt und einer der vier Merkmalsausprägungen zugeordnet.

19. *Startnummer:* Kategorien; *Zeitdauer:* Meßwerte in h:min:s; *Positionen:* Rangdaten

20. Thermometer, Barometer, Amperemeter

21. Intervallskala, Nominalskala, Intervallskala, Nominalskala (in dieser Reihenfolge)

22. *Zufallsgröße* (*Merkmal*): Masse; *Einheit:* kg; *Merkmalsausprägungen:* [101 kg; 175 kg], gerundet auf ganze Kilogramm; *Art der Merkmalsausprägungen:* quantitativ; *Meßinstrument:* Viehwaage; *Skala:* Intervallskala; *Datenart:* Meßwerte

23. *Zufallsgröße:* Familienstand; *Einheit:* entfällt; *Merkmalsausprägungen:* verheiratet, ledig, geschieden, verwitwet (Reihenfolge ohne Belang); *Art der Merkmalsausprägungen:* qualitativ; *Erfassungsinstrument:* Personenstandsbuch oder Ehebuch beim Standesamt (beste Antworten), auch richtig: Personalausweis, Urkunden; *Skalenart:* Nominalskala; *Datenart:* Kategorien

24.

Stetige Zufallsgröße	*Diskrete Zufallsgröße*
Nutentiefe	Mitgliedschaft im DTSB
Masse	Schiffsbestand der Handelsflotte der DDR
Geschwindigkeit	
Leistungsvermögen in Mathematik	
Hektarertrag	
Niederschlagsmenge	

Anmerkung: Ordneten Sie „Leistungsvermögen in Mathematik“ falsch ein, so wiederholen Sie das Beispiel 16d) (↗ S. 23) und den nachfolgenden Satz!

26.

Merkmalsausprägung x_k ($k = 1, \ldots, 7$) (s)	Striche	Häufigkeit h_k
40	|	1
41	卌 卌	10
42	卌 卌 |	11
43	卌 卌 |	11
44	卌 卌	10
45	卌	5
46	||	2
Summe		$50 = n_2$

29.

Merkmalsausprägung x_k ($k = 1, \ldots, 7$) (s)	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$	Prozentuale Häufigkeit
40	1	0,02	2%
41	10	0,20	20%
42	11	0,22	22%
43	11	0,22	22%
44	10	0,20	20%
45	5	0,10	10%
46	2	0,04	4%
Summe	$50 = n$	1,00	100%

30. Ja, natürlich. Bei Vorliegen einer Nominalskala ist die Anordnung der Merkmalsausprägungen willkürlich, also dem Untersucher überlassen.

33. Der Linienzug hat die Form einer Glockenkurve, ist fast symmetrisch.

34. Die Leistung der Lehrlinge A und B ist hoch einzuschätzen, liegen sie doch beide gut im Vergleich zur Facharbeiternorm.
Die Leistung des Lehrlings B ist jedoch noch besser als die von A.
Gründe: B übererfüllt die Norm (der Linienzug – richtiger: die Fläche unter dem Linienzug – liegt in wesentlichen Teilen links von 45 s), er arbeitet beständiger, mit gleichmäßig guter und sehr guter Zeit (der Linienzug ist schmaler und hochgipfliger).
A erfüllt die Norm im wesentlichen, arbeitet aber nicht so zuverlässig wie B (es kommen Zeiten vor von 49 s, 50 s und 51 s, der Linienzug ist nicht so geschlossen, breiter gelagert).

35. **b)** Indem man die Mitten der oberen Rechteckseiten miteinander verbindet. An den Rändern der Verteilung ist die Mitte der oberen Rechteckseiten mit der nächstgelegenen Merkmalsausprägung auf der Achse zu verbinden.

36. **a)** *Untersuchungsobjekte:* $n = 20$ Schüler; *Zufallsgröße (Merkmal):* Leistung im Mathematik-Olympiade-Wettbewerb; *Art der Zufallsgröße:* stetig; *Einheit:* Punkte; *Merkmalsausprägungen:* [20 Pkte.; 31 Pkte.], gerundet auf ganze Punkte; *Art der Merkmalsausprägungen:* quantitativ; *Skala:* Intervallskala; *Datenart:* Meßwerte

b)

Merkmalsausprägung x_k ($k = 1, \ldots, 12$) (Pkte.)	Striche	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$
20	I	1	0,05
21		0	0,00
22	II	2	0,10
23	I	1	0,05
24	II	2	0,10
25	I	1	0,05
26	III	3	0,15
27	II	2	0,10
28	IIII	4	0,20

Merkmalsausprägung x_k ($k = 1, \ldots, 12$) (Pkte.)	Striche	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$
29	|	1	0,05
30	||	2	0,10
31	|	1	0,05
Summe		$20 = n$	1,00

e) Der Linienzug wirkt „unruhig", „zerrissen".
Das Wesen der Erscheinung vermag nicht deutlich hervorzutreten.

37. **b)** Alle sechs Strecken wären fast gleich lang mit $h_k \approx 167 \left(\hat{=} \frac{h_k}{n} \approx \frac{1}{6}\right)$.
Es würde eine *Gleichverteilung* entstehen.

39. **a)**

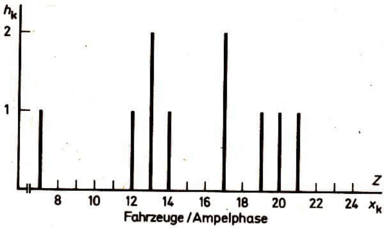


Bild 84

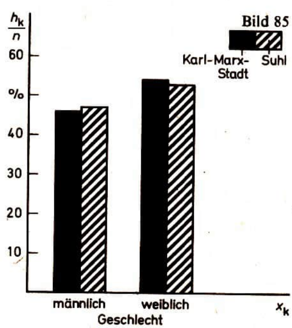
b) Das in einer halben Stunde eines Tages gewonnene Datenmaterial reicht bei weitem nicht aus, um Regelhaftigkeiten der Verkehrsbelegung (der betr. Kreuzung in den Abendstunden) aufzudecken. Nur eine Erweiterung der Datenerfassung (längere Zeitspanne in den Abendstunden mehrerer Tage) kann hier weiterhelfen. Bild 84 (für sich betrachtet) könnte Tatbestände vortäuschen, die in Wirklichkeit gar nicht existieren.

40. Nein, keinesfalls. Es würden Zwischenwerte vorgetäuscht, die es gar nicht gibt.

41. Streifendiagramm: Ja, ein liegendes
Graphische Darstellung einer Häufigkeitsverteilung: Nein. Hier sind zwei Merkmale („Sozialistische Länder" und „Elektroenergieerzeugung") zueinander in Beziehung gesetzt, ohne daß die Häufigkeit von Merkmalsausprägungen vorläge.

42. Prozentuale Verteilung des Merkmals „Geschlecht" in den Bezirken Karl-Marx-Stadt und Suhl mittels Streifendiagramm (↗ Bild 85)

Bild 85



43. *B* ist falsch. Die Aussage *B* enthält eine Absolutangabe (Die Fläche ... ist ... größer als ...), die dem Bild 21 gar nicht entnommen werden *kann*. Wer diese Aussage formuliert, zieht nicht in Betracht, daß Schweden (448000 km²) mehr als zehnmal so groß ist wie Dänemark (43000 km²). Die Fläche der Wiesen und Weiden beträgt in Dänemark etwa 2500 km², in Schweden rund 9000 km².

44. **a)** Nein, das kann man nicht. Die Angaben in Bild 23 vermitteln den *prozentualen Anteil* der einzelnen Verkehrsträger, sie vermitteln keine absoluten Häufigkeiten. Die Gesamtzahl der beförderten Personen lag 1978 wesentlich höher als 1949, das trifft speziell auch für die Eisenbahn zu. Aus Bild 23 ist das allerdings nicht ablesbar, erst recht nicht Sachverhalte, die in diesem Staffelbild gar nicht angesprochen sind (z. B. durch Flugzeug oder mit eigenem PKW beförderte Personen).

46.

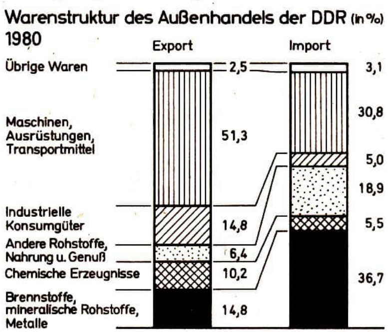


Bild 86

49.

Merkmalsausprägung x_k	Absolute Häufigkeit h_k (in Millionen)	Relative Häufigkeit $\frac{h_k}{n}$	Prozentuale Häufigkeit
Rinder	5,7	0,120	12,0%
Schweine	12,9	0,272	27,2%
Schafe	2,0	0,042	4,2%
Legehennen	26,8	0,565	56,5%
Summe	47,4	0,999	99,9%

50. $x_u = 4{,}9$ mm; $x_o = 5{,}1$ mm; $x_{ug} = 4{,}85$ mm; $x_{og} = 5{,}15$ mm

51. $80\,\frac{\text{km}}{\text{h}}$; $5\,\frac{\text{km}}{\text{h}}$; $75\,\frac{\text{km}}{\text{h}}$; $77{,}5\,\frac{\text{km}}{\text{h}}$

Anmerkungen: Haben Sie auch die Angabe der Einheit nicht vergessen? „75 bis unter 80" heißt *nicht* „75 bis 79", sondern „75 bis 79,999 ..." (jeweils $\frac{\text{km}}{\text{h}}$). Zur Berechnung von Klassenmitte und Klassenbreite wird der Wert 80 verwendet. Wenn – wie hier – als Bezeichnung für die Klassen die exakten Klassengrenzen verwendet werden, fallen untere Klassengrenze und exakte untere Klassengrenze zusammen, dann ist also $x_u = x_{ug}$ und entsprechend $x_o = x_{og}$.

53. **a)** Die Leistung von A ist höher einzuschätzen. A bringt mehr Qualität.
b) Nein, das trifft nicht immer zu, z. B. dann nicht, wenn das untersuchte Merkmal die „Zeit" ist und kürzere Zeiten bessere Leistung bedeuten. Es trifft auch nicht zu, wenn – bei diskreten Merkmalen – eine geringere Anzahl eine bessere Leistung anzeigt.

55.

Merkmalsausprägung (Pkte.)	Absolute Häufigkeit h_k	Klasse (Pkte.)	Klassenmitte x_k (Pkte.)	Absolute Häufigkeit h_k	Relative Häufigkeit $\frac{h_k}{n}$
20	1	18 bis 20	19	1	0,05
21	0				
22	2	21 bis 23	22	3	0,15
23	1				
24	2				
25	1	24 bis 26	25	6	0,30
26	3				
27	2				
28	4	27 bis 29	28	7	0,35
29	1				
30	2				
31	1	30 bis 32	31	3	0,15
Summe	20			20	1,00

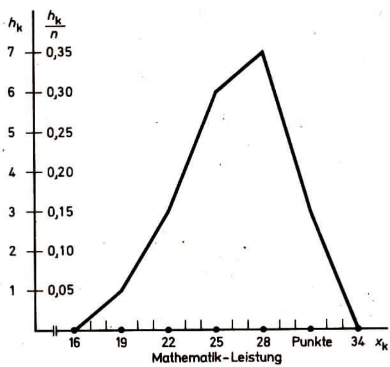


Bild 87

Anmerkungen: Die graphische Darstellung als Histogramm wäre ebenso richtig. Falsch dagegen Streckendiagramm oder andere Darstellungsarten. Weil der Gipfel nach rechts tendiert, bezeichnet man eine derartige Verteilung als *leicht rechtsschief normalverteilt.* Möglich ist, daß Sie eine andere Klasseneinteilung gewählt haben, z. B.

mit $d = 2$ (Pkte.) oder mit $d = 2$ (Pkte.)
$a = 20$ (Pkte.) $a = 19$ (Pkte.)

Absolute Häufigkeiten wären dann

1 3 3 5 5 3 bzw. 1 2 3 4 6 3 1.

Ungünstig, da gegen Anhaltspunkt 3 (S. 54) verstoßend

Ungünstig, da gegen Anhaltspunkt 1 (S. 54) verstoßend (zu viele Klassen)

Auch bei diesen Varianten kommt die leicht rechtsschiefe Tendenz zum Ausdruck. Die oben dargestellte Klasseneinteilung ist jedoch die bestmögliche.

56. Untersuchung der Zufallsgröße „Masse von Schweinen" (kg)
Klasseneinteilung mit $d = 9$ kg, $a = 98$ kg

a)

Merkmalsausprägung (kg)	Striche	Absolute Häufigkeit	Klasse (kg)	Klassenmitte x_k (kg)	Absolute Häufigkeit h_k	Prozentuale Häufigkeit
101	I	1				
102			97,5			
103	I	1	bis unter 106,5	102	2	3,33 %
104						
105						
106						
107						
108	I	1				
109						
110			106,5			
111			bis unter 115,5	111	4	6,67 %
112						
113	I	1				
114	II	2				
115						
116	II	2				
117						
118						
119	II	2	115,5			
120			bis unter 124,5	120	10	16,67 %
121	I	1				
122	II	2				
123	II	2				
124	I	1				
.		.	124,5 b. u. 133,5	129	10	16,67 %
.		.	133,5 b. u. 142,5	138	18	30,00 %
.		.	142,5 b. u. 151,5	147	8	13,33 %
			151,5 b. u. 160,5	156	7	11,67 %
			160,5 b. u. 169,5	165	0	0,00 %
170						
171						
172			169,5			
173			bis unter 178,5	174	1	1,67 %
174						
175	I	1				
Summe		60			60	100,01 %

b) Prüfung der Anhaltspunkte (von S. 54)

1. Für $n = 60$ mit $m = 9$ Klassen erfüllt
2. Klassenbreite $d = 10$ kg hätte die Erfüllung des 3. Punktes erschwert.
3. Weitgehend erfüllt

c) $x_{ug} = 106{,}5$ kg; $x_o = 115$ kg; $x_z = 111$ kg.

d)

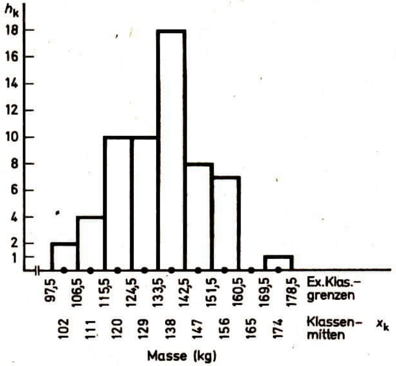


Bild 88

Es liegt eine angenäherte Normalverteilung vor mit einem Gipfelwert, der in die Klasse „133,5 bis unter 142,5" fällt (↗ Bild 88).

60. **a)** $F(1{,}96) - F(-1{,}96) = 0{,}975 - 0{,}025 = 0{,}950 \mathrel{\hat{=}} 95\%$
b) $F(\to \infty) - F(1{,}96) = 1{,}000 - 0{,}975 = 0{,}025 \mathrel{\hat{=}} 2{,}5\%$

63. Zufallsgröße „Zahl der Ringe je Schuß"

a)

Merkmals-ausprägung x_k	Absolute Häufigkeit h_k
2	1
3	1
4	3
.	.
.	.
.	.
10	12
$\sum h_k$	50

b)

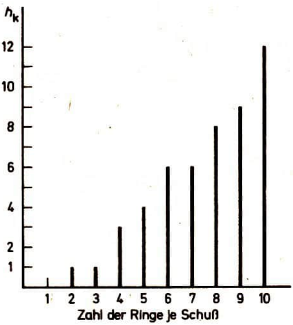


Bild 89

Als graphische Darstellung kommt nur das Streckendiagramm in Frage, weil die Zufallsgröße „Zahl der Ringe je Schuß" diskret ist.

c) Es liegt eine rechtsschief J-förmige Verteilung vor.

64. **a)** Gleichverteilung **b)** Linksschief J-förmige Verteilung
c) Zweigipflige Verteilung

65. **a)** 0,5 1,5 3 5 7,5 10,5 14
b) 6 23 41 47 48 50

67. **a)** Ablesung (↗ Bild 55): $F(2) \approx 0{,}98$; $F(0) = 0{,}50$
Tafelwerte (Tabelle S. 61): $F(2) = 0{,}977$; $F(0) = 0{,}500$
b) $F(-1{,}96) = 0{,}025$ ist einerseits der Zahlenwert der Fläche unter der standardisierten Normalkurve zwischen $-\infty$ und $x = -1{,}96$, andererseits der Zahlenwert der Ordinate der GAUSSschen Summenkurve an der Stelle $x = -1{,}96$.

68. **a)**

Klasse (Punkte)	Klassenmitte x_k	Absolute Häufigkeit h_k	Summenhäufigkeit sh_k	Relative Häufigkeit $\frac{h_k}{n}$	Summenhäufigkeit $\frac{sh_k}{n}$
18 bis 20	19	1	1	0,05	0,05
21 bis 23	22	3	4	0,15	0,20
24 bis 26	25	6	10	0,30	0,50
27 bis 29	28	7	17	0,35	0,85
30 bis 32	31	3	20	0,15	1,00
Summe		20 ←	Kontrolle	1,00 ←	Kontrolle

b) Summenkurve (↗ Bild 90)

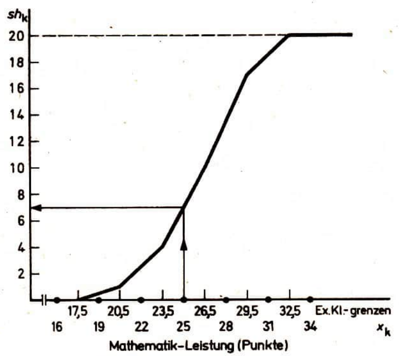


Bild 90

c) Ablesung: 7 Schüler erzielten weniger als 25 Punkte.
Wir beachten: Diese Angabe resultiert aus der Summenverteilung *nach* Klassenbildung. Wir wissen, daß mit der Klassenbildung zwar eine Informations*verdichtung* erreicht wird, andererseits aber auch ein Informations*verlust* verbunden ist.
Für die ungruppierten Daten (ohne Klassenbildung) würde die Antwort lauten: 6 Schüler.

69. **a)** Summenkurve, da es sich um ein *stetiges* Merkmal mit *intervallskalierten* Daten handelt (↗ Bild 91)

Beim Zeichnen der Summenkurve galt es wiederum, darauf zu achten, daß die Summenhäufigkeiten über den exakten oberen Klassengrenzen der jeweiligen Klasse aufzutragen sind.

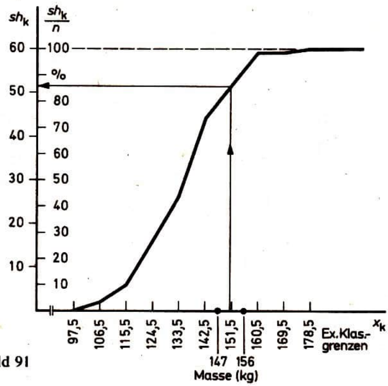


Bild 91

b)

Klasse (kg)	Absolute Häufigkeit h_k	Summenhäufigkeit sh_k	Prozentuale Häufigkeit	Prozentuale Summenhäufigkeit
97,5 bis unter 106,5	2	2	3,33%	3,33%
106,5 bis unter 115,5	4	6	6,67%	10,00%
115,5 bis unter 124,5	10	16	16,67%	26,67%
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
169, 5 bis unter 178,5	1	60	1,67%	100,01%
Summe	60		100,00%	

c) 52 Schweine ≙ 86,67%

70. Tabellarische Darstellung

Güteklasse u_k $(k = 1, \ldots, 5)$	Absolute Häufigkeit h_k	Absolute Summenhäufigkeit sh_k	Relative Summenhäufigkeit $\frac{sh_k}{n}$
1	5	5	0,20
2	9	14	0,56
3	6	20	0,80
4	4	24	0,96
5	1	25	1,00
Summe	25 ←	Kontrolle	

Graphische Darstellung (↗ Bild 92): Treppenpolygon (da es sich um eine rangskalierte Zufallsgröße handelt)

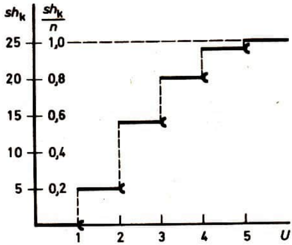


Bild 92

71. Ablesung am Bild 54, S. 70: Lehrling A benötigte bei 38 % der Steckermontagen weniger als 43 s, Lehrling B bei 54 % der Steckermontagen (Bild 93; Ausschnitt aus Bild 54). Das sind absolut: 19 Steckermontagen (Lehrling A) bzw. 27 Steckermontagen (Lehrling B).

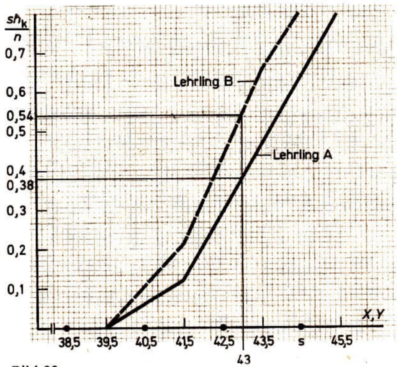


Bild 93

72.

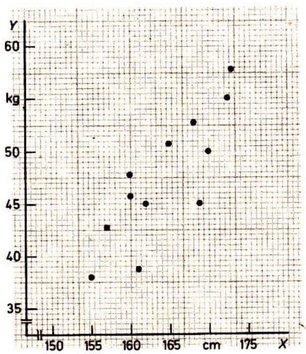


Bild 94

73. Merkmal X: Elektroenergie (in kWh)
Merkmal Y: RGW-Land (oder: Sozialistische Länder)

74. Datenart zu X: Meßwerte; Datenart zu Y: Kategorien
Ja, Streifendiagramm ist richtig.

76. **a)** Degressiv fallend oder (mit anderen Worten) kleiner werdende Abnahme
b) Nein, keinesfalls
Bei der Zurückdrängung von Krankheiten (Bild 63, ↗ S. 76), bei der Einsparung von Material und Kosten usw. sind *fallende* Trendformen Ausdruck einer positiven Entwicklung.

77. 1965 bis 1970: degressiv steigend; 1970 bis 1975: konstant; 1975 bis 1980: degressiv fallend; 1980 bis ?: gleichförmig fallend

80. **a)** Absolut größter Zuwachs (mit je 350 Mill. Menschen) sowohl in Zentral- und Ostasien als auch in Südasien
Absolut geringster Zuwachs (mit 5 Mill. Menschen nach Angabe des Kartogramms) in Australien und Ozeanien
b) Größter prozentualer Zuwachs von 1960 bis 1980, bezogen auf die Bevölkerungszahl von 1960:

1. Mittelamerika	mit	6:6,5	≙ 92,3 %	Zuwachs
2. Mittel- und Südafrika	mit	13:15	≙ 86,7 %	Zuwachs
3. Südostasien	mit	15,5:22,5	≙ 68,9 %	Zuwachs

81. **a)** Die Einschnürungen bedeuten eine erhebliche Verminderung der Bevölkerungszahl in dem betreffenden Alter. Sie sind auf die beiden Weltkriege zurückzuführen.
1976 − 30 = 1946 Auswirkungen des II. Weltkrieges (Gefallene, Getötete, zerrissene Familien, deshalb wenige Geburten);
1976 − 60 = 1916 Auswirkungen des I. Weltkrieges
b) Aus den drei Jahrzehnten friedlichen Aufbaus in der DDR seit Ende des II. Weltkrieges

83. **a)** Abszissenachse: Zeit (Jahre)
Ordinatenachse: Durchschnittliche (oder mittlere) Milchleistung je Kuh (kg)
b) Linienzug oder Streifendiagramm

84. Ja, in der 2. Zeile handelt es sich um mittlere Stückzahlen je Henne.

85. Es werden Mittelwerte je Klasse oder je Klassenstufe weitergeleitet, die Punktzahl jedes einzelnen Schülers nicht.

87. **a)** $\sum_{k=1}^{7} 2k = 2 + 4 + 6 + 8 + 10 + 12 + 14 = 56$

b) $2^0 + 2^1 + 2^2 + 2^3 + 2^4 + 2^5 = 1 + 2 + 4 + 8 + 16 + 32 = 63$

88. **a)** $\sum_{k=1}^{4} \sqrt{k-1} = \sqrt{0} + \sqrt{1} + \sqrt{2} + \sqrt{3} = 4{,}146$

b) $\sum_{k=1}^{8} x_k = 37{,}6$ cm (Einheit nicht vergessen!)

89. **a)** $\bar{x} = \frac{1}{8} \sum_{k=1}^{8} x_k = 4{,}70$ cm (Einheit nicht vergessen!)

b) $\bar{x} = \frac{1}{20} \cdot 524$ Punkte $= 26{,}2$ Punkte

90. $\bar{x} = \frac{1}{50} \cdot 376$ Ringe $= 7{,}5$ Ringe

91. $\sum x_k h_k = 2142; \quad \bar{x} = \frac{1}{50} \cdot 2142 \text{ s} = 42{,}84 \text{ s} \approx 42{,}8 \text{ s}$

Anmerkung: Da in der Untersuchung ganze Zahlen erfaßt wurden, berechnen wir $\bar{x}$ auf zwei Dezimalstellen und runden dann auf eine.

92. $\sum x_k h_k = 2141; \quad \bar{x} = \frac{1}{50} \cdot 2141 \text{ s} = 42{,}82 \text{ s} \approx 42{,}8 \text{ s}$

93.

Klassenmitte x_k (kg)	Absolute Häufigkeit h_k	Produkt $x_k h_k$
Spalte ①	②	③ = ① · ②
102	2	204
111	4	444
120	10	1200
.	.	.
.	.	.
.	.	.
174	1	174
Summe	$60 = n$	$8064 = \sum x_k h_k$

$\bar{x} = \frac{1}{60} \cdot 8064 \text{ kg} = 134{,}4 \text{ kg}$

Anmerkung: Der exakte Wert aus der ungruppierten Häufigkeitstabelle ist 134,2 kg. (Dieser Wert war nicht gefordert.)

95.

Klasse (s)	Klassenmitte x_k (s)	Absolute Häufigkeit h_k	Hilfswerte x'_k	Produkt $x'_k h_k$	
⓪	①	②	③	④ = ③ · ②	⑤
40 bis 41	40,5	11	−1	−11	
42 bis 43	$42{,}5 = x_a$	22	0	0	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
50 bis 51	50,5	0	4	0	
Summe		$50 = n$		$8 = \sum x'_k h_k$	

Nach (11a) folgt: $\bar{x} = \left(42{,}5 + \frac{2}{50} \cdot 8\right) \text{s} = (42{,}5 + 0{,}32) \text{ s}; \qquad \bar{x} = 42{,}8 \text{ s}$

Wir erhalten das gleiche Ergebnis wie in Aufgabe 92. Durch Anwendung des Verfahrens des angenommenen Mittelwerts tritt also kein Informationsverlust ein. An Stelle von 42,5 s hätte auch jede andere Klassenmitte als x_a gesetzt werden können.

96. **a)** Nach Formel (7a), S. 85: $\bar{x} = \frac{1}{30} \cdot 454$ Fahrzeuge je Ampelphase $= 15{,}1$ Fahrzeuge je Ampelphase

b) Mit $\sum_{k=1}^{21} z_k h_k = 454$ ergibt sich nach Formel (9a), S. 86: $\bar{x} = \frac{1}{30} \cdot 454$ Fahrzeuge je Ampelphase $= 15{,}1$ Fahrzeuge je Ampelphase.

c) Mit $\sum_{k=1}^{7} z_k h_k = 453$ erhält man anhand der Tabelle auf Seite 57f. wiederum nach (9a), jetzt aber auf die Klassen bezogen, $\bar{x} = \frac{1}{30} \cdot 453$ Fahrzeuge je Ampelphase $= 15{,}1$ Fahrzeuge je Ampelphase.

d) Ja, die Berechnung des arithmetischen Mittels ist in allen drei Fällen möglich und sinnvoll, denn es liegen intervallskalierte Daten vor (Gleichabständigkeit der Marken ist gewährleistet), und die Verteilung ist – wie u. a. dem Bild 39 (↗ S. 57) zu entnehmen ist – angenähert normal. Daß das untersuchte Merkmal diskret ist, ändert an diesen Gegebenheiten nichts. Alle drei Ergebnisse stimmen gut überein.

97. **a)** Mit $\sum_{k=1}^{12} x_k h_k = 1037$ (aus der Tabelle zum Beispiel 34, ↗ S. 63) ergibt sich nach Formel (9a), S. 86: $\bar{x} = \frac{1}{51} \cdot 1037\,\text{m} = 20{,}33\,\text{m}$; $\bar{x} = 20{,}3\,\text{m}$.

b) Bestimmung von $\bar{x}$ absolut unmöglich, da hier Nominalskala zugrundeliegt

98. **a)** 100; 104; 112; 80

b) Durchschnittliche wertmäßige Planerfüllung für die acht Betriebe:

$$\frac{\text{Gesamtist}}{\text{Gesamtsoll}} = \frac{1032\,\text{TM}}{1013\,\text{TM}} \mathrel{\hat{=}} 101{,}9\,\%$$

Falsch wäre das Bilden des arithmetischen Mittels der Planerfüllung der acht einzelnen Betriebe; 98,4 % ist also falsch.

99. **a)** $v = \frac{s}{t} = \frac{\text{Weg}}{\text{Zeit}}$

$$\text{Durchschnittl. } \textit{Reise}\text{geschw.} = \frac{\text{Gesamtweg}}{\text{Gesamtzeit}} = \frac{115\,\text{km}}{\frac{7}{3}\,\text{h}} = 49{,}29\,\frac{\text{km}}{\text{h}}$$

b) Das Vorgehen des Reisenden war ungerechtfertigt, der Wert $59\,\frac{\text{km}}{\text{h}}$ ist falsch. Er beachtete nicht, daß er Geschwindigkeiten einfach arithmetisch mittelte, die sich auf unterschiedlich lange Teilfahrzeiten bezogen. Das ist genau so unstatthaft wie das Bilden des arithmetischen Mittels von Mittelwerten, die sich auf unterschiedlich große Stichproben beziehen.
Korrektes Vorgehen:

Entweder: $\frac{\text{Gesamte Fahrstrecke}}{\text{Gesamte Fahrzeit}} = \frac{115\,\text{km}}{(25 + 80 + 25)\,\text{min}}$ und damit durchschnittliche *Fahr*geschwindigkeit $= 0{,}885\,\frac{\text{km}}{\text{min}} = 53{,}1\,\frac{\text{km}}{\text{h}}$ oder: über das *gewogene* arithmetische Mittel bezüglich der Zeit (↗ Abschnitt 2.3.2., S. 94ff.)

100. c) $x_a = 32$; $d = 1$; $n = 91$; $\sum_{k=1}^{m} x'_k h_k = \sum m h_m = 36$

$$\bar{x} = \left(32 + \frac{1}{91} \cdot 36\right) \text{mm} = (32 + 0{,}396)\ \text{mm} \approx 32{,}40\ \text{mm}$$

101. $\bar{x}_g = \dfrac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \bar{x}_3 n_3}{n_1 + n_2 + n_3}$; $n_1 = n_2 = n_3$ nach Voraussetzung

$\bar{x}_g = \dfrac{n_1(x_1 + x_2 + x_3)}{3 n_1}$; Division durch $n_1 \neq 0$ führt auf Formel (7) des einfachen arithmetischen Mittels

102. Aus (15a) folgt für $l = 3$

$$\bar{x}_g = \frac{3600 \cdot 7 + 3350 \cdot 68 + 3550 \cdot 25}{100}\ \text{g} = 3417{,}5\ \text{g}$$

Falsch ist: $\dfrac{10500}{3}$ g $= 3500$ g.

103. $\sum_{k=1}^{9} \bar{x}_i n_i = 12866600$; $\bar{x}_g = 3806{,}7\ l$

Falsch ist: $\dfrac{33650}{9}\ l = 3738{,}9\ l$.

104.

Klasse (s)	Klassenmitte x_k (s)	Absolute Häufigkeit h_k	Summenhäufigkeit sh_k
40 bis 41	40,5	11	11
$\boxed{\text{42 bis 43}}$	42,5	22	33
44 bis 45	44,5	15	48
46 bis 47	46,5	2	50
Summe		$50 = n$	

$\dfrac{n}{2} = 25$; Eingriffsspielraum für Z: 42 bis 43, folglich $x_{ug} = 41{,}5$; $sh_u = 11$; $h_z = 22$; $d = 2$

Nach (16): $Z = (41{,}5 + 1{,}27)$ s; $Z \approx 42{,}8$ s

Zum Vergleich mit $\bar{x}$ ist es hier angebracht, $\bar{x}$ und Z auf zwei Stellen nach dem Komma anzugeben:

$\bar{x} = 42{,}82$ s $> Z = 42{,}77$ s, d. h., es liegt eine leicht linksschiefe Verteilung vor.

105. a) $Z = 14$ mm

b) $n = 30$; $\dfrac{n}{2} = 15$; Eingriffsspielraum für Z ist Klasse mit Klassenmitte 0,25. Diese Klasse hat die exakten Grenzen 0,235 ... 0,265, also:
$x_{ug} = 0{,}235$; $sh_u = 7$; $h_z = 12$; $d = 0{,}03$.
Nach (16): $Z = (0{,}235 + 0{,}020)$ g $= 0{,}255$ g
(Hier liegt eine stark rechtsschiefe Verteilung vor. Das zeigt auch der Vergleich mit $\bar{x}$:
$\bar{x} = 0{,}249$ g $< Z = 0{,}255$ g. Dieser Vergleich war nicht gefordert.)

c) Nur für a)

$\bar{x}$ ist bei c) unangebracht wegen starker Schiefe der Verteilung und Auftretens eines Ausreißerwertes.

106. **a)** Z ist angemessener; wegen Schiefe der Verteilung und wegen Ausreißerwerten.
b) $\bar{x} \approx 3{,}19$ mm; $Z = 3{,}08$ mm
$\bar{x} > Z$; das bedeutet: linksschiefe Verteilung.

107. Ablesung für Merkmal X (Lehrling A): $Z_x = 43{,}7$ s;
für Merkmal Y (Lehrling B): $Z_y = 42{,}8$ s
Nur bei sehr exaktem Zeichnen kann man so genau ablesen.

108. **a)** $D = 2$ **c)** *vor* Klassenbildung: $D = 3{,}0$ mm
b) $D =$ verh. *nach* Klassenbildung: $D = 2{,}9$ mm
d) $D = 44{,}5$ s (Angabe der Einheit nicht vergessen)

109. **a)** Ja **b)** $\bar{x} < Z < D$

110. Übereinstimmung im (arithmetischen) Mittelwert.
Deutlicher Unterschied in der Streuung.

111. **a)** Die Angabe des Streubereichs der Tiefe
b) $w = (2{,}25 - 0{,}60)$ m $= 1{,}65$ m

112. **a)** $w = (46 - 40)$ s $= 6$ s
b) $w = 20$ Fahrzeuge je Ampelphase
c) $w = (4{,}9 - 2{,}7)$ mm $= 2{,}2$ mm

113. **a)**

Merkmals-ausprägung x_k (s)	Absolute Häufigkeit h_k	Abweichung $x_k - \bar{x}$ (s)	Absoluter Betrag $\lvert x_k - \bar{x}\rvert$ (s)	Produkt $\lvert x_k - \bar{x}\rvert\, h_k$
40	1	−2,8	2,8	2,8
41	10	−1,8	1,8	18,0
42	11	−0,8	0,8	8,8
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
46	2	3,2	3,2	6,4
Summe	$50 = n$		12,2	61,2

Nach (20): $e = \frac{1}{50} \cdot 61{,}2 \text{ s} \approx 1{,}2 \text{ s}$

b) Nach Entwicklung einer Tabelle gemäß der Tabelle auf S. 105 ergibt sich, wiederum aufgrund von (20):

$$e = \frac{1}{50} \cdot 64{,}8 \text{ s} \approx 1{,}3 \text{ s}.$$

c) Kein bedeutsamer Unterschied

114. **a)** $s = 1{,}84$ cm **b)** $s = 2{,}98$ Punkte $\approx 3{,}0$ Punkte

115. Nein, da keine Normalverteilung vorliegt

116. **a)** $s = 1{,}476$ s $\approx 1{,}5$ s

b) Produkt

$x_k'^2 h_k$
⑤ = ③ · ④
11
0
15
8
$34 = \sum_{k=1}^{m} x_k'^2 h_k = A$

Nach (23):

$$s = 2 \cdot \sqrt{\frac{1}{49} \cdot \left(34 - \frac{8^2}{50}\right)}\ \mathrm{s}$$
$$= 1{,}63\ \mathrm{s}$$
$$s \approx 1{,}6\ \mathrm{s}$$

c) Unterschied auch hier unbedeutend, wenn man bedenkt, daß die Meßwerte in ganzen Zahlen (volle Sekunden) erhoben wurden

117. a) $s = 4{,}77$ Fahrzeuge je Ampelphase; $s \approx 4{,}8$ Fahrzeuge je Ampelphase

b) $s = 4{,}56$ Fahrzeuge je Ampelphase; $s \approx 4{,}6$ Fahrzeuge je Ampelphase

c) Auch hier ist zu empfehlen, das Verfahren des angenommenen Mittelwertes zu verwenden.

$s = 2{,}54\ \mathrm{m} \approx 2{,}5\ \mathrm{m}$

d) Bestimmung von s wäre unsinnig; für nominalskalierte Daten gibt es kein Streuungsmaß.

118. a) $\bar{x} = 42{,}8$ s (aus Lösung zur Aufgabe 91); $s = 1{,}5$ s

$\bar{x} - s \ldots \bar{x} + s = [41{,}3;\ 44{,}3]$ s

b) $\bar{x} = 42{,}8$ s (aus Lösungen zu Aufgabe 92 oder 95); $s = 1{,}6$ s

$\bar{x} - s \ldots \bar{x} + s = [41{,}2;\ 44{,}4]$ s

c) $\bar{x} = 15{,}1$ Fahrzeuge/Ampelphase; $s = 4{,}8$ Fahrzeuge/Ampelphase

$\bar{x} - s \ldots \bar{x} + s = [10{,}3;\ 19{,}9]$ Fahrzeuge/Ampelphase

d) $\bar{x} = 20{,}3$ m; $s = 2{,}5$ m

$\bar{x} - s \ldots \bar{x} + s = [17{,}8;\ 22{,}8]$ m

120 (1). „Wappen“ und „Zahl“

120 (3). $\frac{h(Z)}{n} + \frac{h(W)}{n} = 1$

120 (4). a) $P(Z) = 0{,}50$; $P(W) = 0{,}50$

b) $P(Z) + P(W) = 1$

121. b) $\frac{h(„3“)}{n} = \frac{11}{60} = 0{,}1833\ldots$

122. a) $[\bar{x} - s;\ \bar{x} + s] = [41{,}3;\ 44{,}3]$ s

$g = 11 + 11 + 10 = 32$ Meßwerte

Nach (27), S. 115: $P(A) = \frac{32}{50} = 0{,}64 \mathrel{\hat{=}} 64\,\%$

b) $[\bar{x} - s; \bar{x} + s] = [23{,}2; 29{,}2]$ Pkte.

$g = 2 + 1 + 3 + 2 + 4 + 1 = 13$ Werte

$$P(A) = \frac{13}{20} = 0{,}65 \mathrel{\widehat{=}} 65\,\%$$

123. Es würden durch das Prüfen zu viele Arbeitskräfte gebunden.

Literatur

* Für die AG(R) „Elementare Statistik" besonders zu empfehlen
** Auch für Schüler geeignet
MSB „Mathematische Schülerbücherei"
VWV Volk und Wissen Volkseigener Verlag, Berlin

[1] ADAM, J.: Einführung in die Biostatistik, Reaktionskinetik und EDV. VEB Verlag Volk und Gesundheit, Berlin 1972

[2]* Autorenkollektiv: Ausgewählte Kapitel der Mathematik für Ingenieur- und Fachschulen. VEB Fachbuchverlag, Leipzig 1967

[3] Autorenkollektiv: Mathematik für die Berufsausbildung „Facharbeiter für Datenverarbeitung". Verlag Die Wirtschaft, Berlin 1977

[4]* Autorenkollektiv: Mathematik für Wirtschaftswissenschaften. Fachschullehrbuch, Teil 3. Verlag Die Wirtschaft, Berlin 1978

[5] Biometrisches Wörterbuch. 2 Bände. VEB Deutscher Landwirtschaftsverlag, Berlin 1968

[6]* CLAUSS, G./H. EBNER: Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen. VWV 1974

[7] DDR-Standard Statistische Qualitätskontrolle. TGL 14 499/Symbole und Begriffe

[8] Demographic Yearbook 1977. UNO, New York 1978

[9] Der Morgen, Zentralorgan der LDPD, Nr. 219/80 vom 16. 9. 1980

[10] Die DDR stellt sich vor. PANORAMA DDR. Verlag Zeit im Bild, Dresden 1981

[11]* DONDA/HERDE/KUHN/STRUCK: Statistik. Verlag Die Wirtschaft, Berlin 1974

[12]* DRUSHIN: Mathematische Statistik in der Ökonomie. Verlag Die Wirtschaft, Berlin 1974

[13]* ENGELS, F.: Dialektik der Natur. Dietz Verlag, Berlin 1952

[14]** GILDE, W.: Gespiegelte Welt. VEB Fachbuchverlag, Leipzig 1979

[15]* GNEDENKO, B. W./A. J. CHINTSCHIN: Elementare Einführung in die Wahrscheinlichkeitsrechnung. VEB Deutscher Verlag der Wissenschaften, Berlin 1971

[16]** GÖRKE, L./K. ILGNER/G. LORENZ/G. PIETZSCH/M. REHM: Rund um die Mathematik. Kinderbuchverlag, Berlin 1969 (MSB 34)

[17]** GÖTTNER, R./P. FISCHR/R. KRIEG: Was ist – was kann die Statistik? Urania Verlag, Leipzig/Jena/Berlin 1975 (MSB 86)

[18]* GRENANDER, U.: Einführung in das Studium der mathematischen Statistik. VEB Fachbuchverlag, Leipzig 1969

[19]* HERFURTH, G.: Umgang mit Zufallsgrößen. BSB B. G. Teubner, Leipzig. Teil 1: Fehler- und Ausgleichsrechnung, 1970. Teil 2: Wahrscheinlichkeitsrechnung und mathematische Statistik, 1969

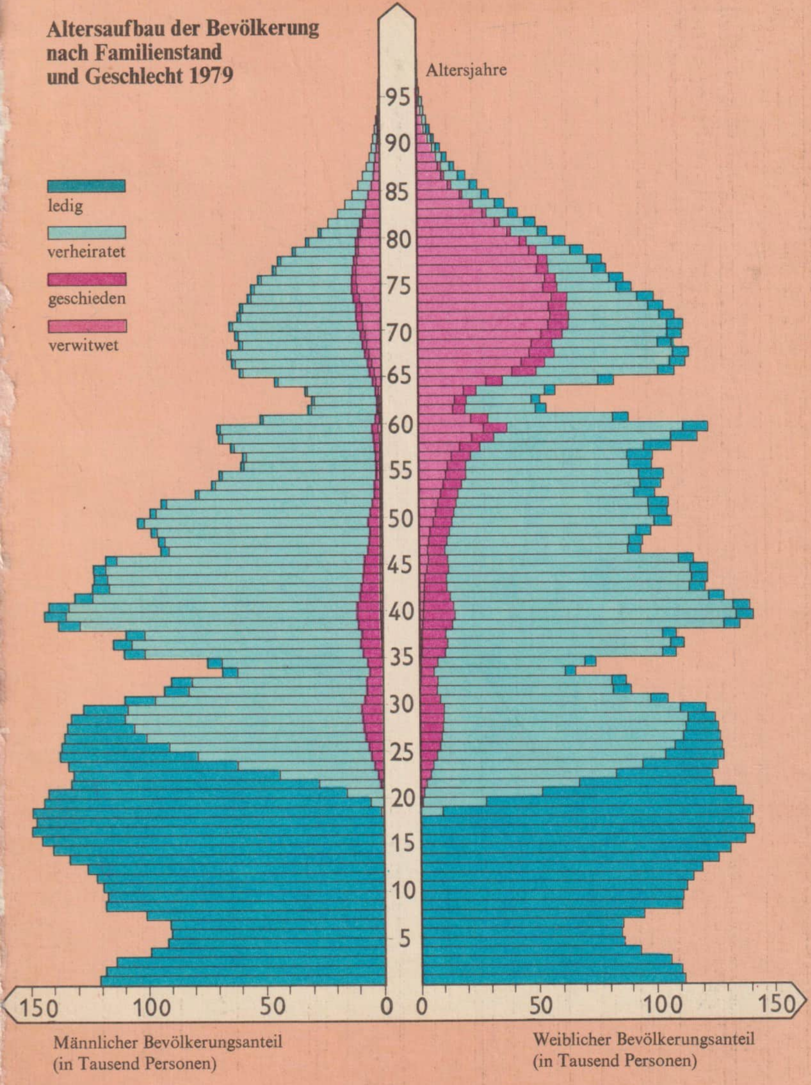
[20] HONECKER, M.: Der gesellschaftliche Auftrag unserer Schule. Referat des Ministers für Volksbildung auf dem VIII. Pädagogischen Kongreß. Dietz Verlag, Berlin 1978
[21] HÖRZ, H.: Zufall – Eine philosophische Untersuchung. Akademie-Verlag, Berlin 1980
[22] Horizont, Sozialistische Wochenzeitung für internationale Politik und Wirtschaft, Nr. 48/79 (12)
[23]* LENIN, W. I.: Statistik und Soziologie. In: Werke, Bd. 23. Dietz Verlag, Berlin 1957
[24]* LOHSE, H./R. LUDWIG: Statistik für Forschung und Beruf. Ein programmierter Lehrgang. Erfassung, Aufbereitung und Darstellung statistischer Daten. VEB Fachbuchverlag, Leipzig 1973
[25]* LOHSE, H./R. LUDWIG: Prüfstatistik. Programmierter Lehrgang 2. VEB Fachbuchverlag, Leipzig 1982
[26]* MAIBAUM, G.: Wahrscheinlichkeitsrechnung, 3. Auflage. VWV 1980 (MSB 63)
[27] Meyers Neues Lexikon, 2. Auflage (16 Bände). VEB Bibliographisches Institut, Leipzig 1978
[28] Neues Deutschland, Zentralorgan der SED, Nr. 258/80 vom 1./2. 11. 1980
[29]* Rahmenprogramm für die Arbeitsgemeinschaft der Klassen 9 und 10 „Elementare Statistik". VWV 1977
[30]* RASCH, D.: Elementare Einführung in die mathematische Statistik, 2. Auflage. VEB Deutscher Verlag der Wissenschaften, Berlin 1970
[31]** RASTRIGIN, L. A.: Zahl oder Wappen? Ein Buch über den Zufall. Verlag MIR, Moskau/Urania Verlag, Leipzig 1973
[32]* SABELUS, H.: Zur Planung und methodischen Gestaltung in der AG(R) „Elementare Statistik" unter dem Aspekt von Aktivität und Erkenntnis. Pädagogische Lesung 5975/1980, Berlin
[33] SCHMIDT, H.-D.: Empirische Forschungsmethoden der Pädagogik. Einführung. VWV 1961
[34] SCHWARZ, H./R. STRUCK/H. WASCHKAU: Allgemeine Statistik. Aufgabensammlung mit ausführlichen Lösungswegen. Verlag Die Wirtschaft, Berlin 1964
[35] Statistisches Jahrbuch der Deutschen Demokratischen Republik 1977. Staatsverlag der DDR, Berlin 1977
[36] Statistisches Jahrbuch der Deutschen Demokratischen Republik 1979. Staatsverlag der DDR, Berlin 1979
[37] Statistisches Jahrbuch der Deutschen Demokratischen Republik 1981. Staatsverlag der DDR, Berlin 1981
[38]* Statistisches Taschenbuch der Deutschen Demokratischen Republik 1979. Staatsverlag der DDR, Berlin 1979
[39]* STORM, R.: Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle, 5. Auflage. VEB Fachbuchverlag, Leipzig 1976
[40] SWESCHNIKOW, A. A.: Wahrscheinlichkeitsrechnung und Mathematische Statistik in Aufgaben. BSB B. G. Teubner Verlagsgesellschaft, Leipzig 1970
[41]* THAMM, R./K. WIENKE: Problemhafte Prozeßgestaltung in der AG(R) „Elementare Statistik". Wissenschaftliche Zeitschrift der Pädagogischen Hochschule „Liselotte Hermann", Güstrow (1979) 1
[42]* TISCHER, E.: Zur Gestaltung der AG(R) „Elementare Statistik" unter Einbeziehung betriebstechnischer und -ökonomischer Aufgabenstellungen. Pädagogische Lesung 5338/1979, Berlin
[43]* WALTHER/SCHOOP/UHLMANN: Statistik im Unterricht. VWV 1966
[44] WEBER, E.: Grundriß der biologischen Statistik, 7. Auflage. VEB G. Fischer Verlag, Jena 1972

[45]* WIENKE, K.: Mittelwerte, Streuung und Streuungsmaße. Manuskript. Güstrow 1979

[46] X. Parteitag der SED, 11. bis 16. April 1981 in Berlin: Bericht des ZK der Sozialistischen Einheitspartei Deutschlands an den X. Parteitag der SED. Berichterstatter: E. HONECKER. Dietz Verlag, Berlin 1981

Zur 3. Umschlagseite: Nach „Statistisches Jahrbuch der DDR“, 1981

Altersaufbau der Bevölkerung nach Familienstand und Geschlecht 1979



Kurzwort: 00 17 16 Elem. Statistik
DDR 6,00 M