

Inhaltsverzeichnis

1. Datenerhebung und grafische Darstellungsformen	6
1.1 Datenerfassung, Verarbeitung und Interpretation	7
1.2 Datenaufbereitung	10
1.3 Absolute und relative Häufigkeit	15
1.4 Weiterführende Anwendungen	21
1.5 Weitere Diagrammdarstellungen	26
2. Lagemaße	28
2.1 Arithmetisches Mittel	29
2.2 Zusammenhang zwischen arithmetischem Mittel und relativer Häufigkeit	32
2.3 Median, Spannweite und Modalwert	35
2.4 Verteilung der Daten und Zusammenhang zwischen den Lagemaßen	39
3. Streumaße	43
3.1 Quartile und Quartilsabstand	44
3.2 Varianz und Standardabweichung	50
3.3 Normalverteilung	56
4. Zufallsexperimente	59
4.1 Wahrscheinlichkeit	60
4.2 Laplace-Experimente	63
4.3 Baumdiagramm und Pfadregeln	65
5. Häufigkeit und Wahrscheinlichkeit	70
5.1 Bedingte Wahrscheinlichkeit	71
5.2 Vierfeldertafel und Satz von Bayes	73
5.3 Inverses Baumdiagramm und Abschätzung von Wahrscheinlichkeiten	78
6. Hypothesentest	83
6.1 Aufstellen und Bewerten von Hypothesen	84
6.2 Binomialverteilung	88
Lösungsteil	92
Register	104

Liebe Leserinnen und Leser,

in der modernen Mathematik kennt man unterschiedliche Zweige. Einer davon behandelt den Themenbereich der angewandten und praktischen Mathematik. Und auch dieser Zweig unterteilt sich wieder in Teilgebiete, zu denen die Themenbereiche dieses Buches gehören, die Statistik und Stochastik.

Das Lernen im Telekolleg wird durch drei Bausteine unterstützt: durch die Fernsehsendung, die den eigentlichen Impuls für jedes Kapitel gibt, das Lehrbuch und den Unterricht an der Kollegtagschule. Alle drei Komponenten fügen sich zu einem hilfreichen Ganzen zusammen. Das Lehrbuch bietet Ihnen als Begleitmaterial zur Fernsehsendung die nötigen Vertiefungen und Erweiterungen des Lehrstoffes.

Von vielen Dingen im Leben wüssten wir gerne genaue Vorhersagen, wann sie eintreffen. Denn oftmals bleibt nur die Einsicht, dass sie zufällig geschehen. Für uns bedeutet dies, dass sie nicht mit Sicherheit vorhersagbar sind. Somit reichen uns oftmals schon ungefähre Vorhersagen aus, um sich auf kommende Ereignisse einstellen zu können. In diesem Buch über Zufall und Wahrscheinlichkeit betrachten wir den mathematischen Bereich von Zufallsexperimenten. Man benötigt dazu Beispiele, deren Ausgang nicht sicher beschrieben werden kann.

Für jedes Kapitel dieses Buches ist ein besonderes Beispiel gewählt, an dem die spezifisch wichtigen Sachverhalte der zugehörigen Lerneinheit erklärt werden. Dabei sind die Seiten in zwei Spalten aufgeteilt. Die linke Spalte dient zur Schilderung und Erklärung, die rechte Spalte zur Hervorhebung von Formeln und Gesetzmäßigkeiten sowie zur Darstellung von Grafiken und Tabellen. Durch die Spalteneinteilung soll Ihnen die Arbeit mit diesem Lehrbuch erleichtert werden. Im letzten Unterpunkt eines jeden Kapitels finden Sie Aufgaben zur Übung, deren Lösungswege und Lösungen im Schlusskapitel des Buches aufgeführt sind. Zum schnellen Auffinden von Teilgebieten und wichtigen Begriffen dient nicht nur das Inhaltsverzeichnis auf den vorderen Seiten, sondern auch das ausführliche Register am Ende des Buches.

In Ihrer Mathematikausbildung werden Sie nicht alles dem Zufall überlassen wollen, deshalb müssen Sie lernen und üben. Und hoher Lerneinsatz führt zur bedingten hohen Wahrscheinlichkeit des Lernerfolgs.

Lerne oft und lerne viel
Hab im Auge stets ein Ziel
Schlage oft das Lehrbuch auf
Dann schaltest Du den Zufall aus

Ich hoffe, dass dieses Lehrbuch zu Ihrem Lernerfolg im Telekolleg beitragen wird.

Dr. Barbara Fischer

1 Datenerhebung und grafische Darstellungsformen

Vor der Sendung

In dieser ersten Statistiklektion werden mathematische Sachverhalte zusammengestellt, die heute bereits in der Grundschule beginnend vermittelt werden.

Für die Rechenarbeit in der Statistik sind keine tiefgründigen mathematischen Vorkenntnisse notwendig. Es reicht im Grunde genommen die Beherrschung der Grundrechenarten aus.

Im Laufe der Geschichte hat sich die Statistik zu einem eigenständigen Teilgebiet der Mathematik entwickelt, das heute ein wichtiges Hilfsmittel für die Prognosen von Prozessen in Naturwissenschaft, Technik und Ökonomie ist.

Die "beschreibende Statistik" in diesem Buch basiert auf der Erhebung von Daten und ihren verschiedenen Möglichkeiten, diese grafisch darzustellen.

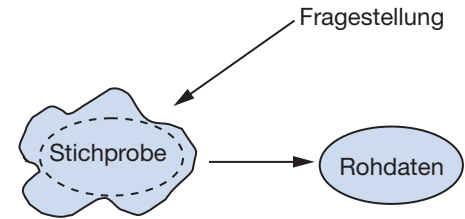
Übersicht

1. An einem einführenden Beispiel werden die Erfassung, die Verarbeitung und die Interpretation von Daten betrachtet.
2. Nach der Datenerfassung werden die Daten aufbereitet. Die Bezeichnungen der erforderlichen Listen und der daraus resultierenden Tabellen werden exemplarisch festgelegt.
3. Die auftretenden Begriffe "absolute und relative Häufigkeit" und ihre grafischen Darstellungen werden an Beispielen veranschaulicht.
4. Die unter Punkt 3 gefundenen Einsichten werden in ihrer Anwendung geübt.
5. Ein Ausblick auf die vielfältigen Möglichkeiten von Diagrammdarstellungen wird erstellt.

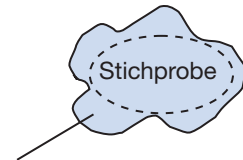
1.1 Datenerfassung, Verarbeitung und Interpretation

Für alle Sachverhalte ist es möglich, statistische Erhebungen durchzuführen. Der Ausgangspunkt für eine solche Erhebung ist die Frage über den zu betrachtenden Sachverhalt. Sie ist der Start für die Untersuchung. In der Regel wird dabei die Fragestellung nur an einen Teil der Gesamtheit gerichtet. Diesen ausgewählten Teil der Gesamtheit nennt man **Stichprobe**. Das Ergebnis der Stichprobe wird in einer **Urliste** festgehalten.

Alle abgefragten Daten nennt man **Rohdaten**. Es handelt sich dabei z. B. um Befragungsergebnisse, Messergebnisse oder Erfahrungswerte. Der Sammelbegriff dafür sind Beobachtungswerte. Die Eigenschaften der Elemente einer Stichprobe heißen Merkmale und die Lieferanten der Rohdaten Merkmalsträger.



Stichprobe: Teil der Gesamtheit



Gesamtheit

Urliste: Aufnahme der Rohdaten

Rohdaten: Befragungsergebnisse, Messergebnisse, Erfahrungswerte

Beispiel:

In einem Betrieb sind 420 Personen beschäftigt. Der Betriebsrat fordert für Neueinstellungen, dass mindestens 40% der Beschäftigten weiblichen Geschlechts sein sollen und dass die Altersstruktur aller Arbeitnehmer so beschaffen sein soll, dass höchstens 20% älter als 50 Jahre sind.

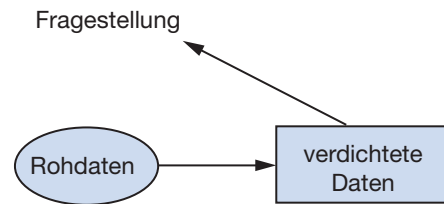
Um nicht alle 420 Personen befragen zu müssen, wird aus dieser Gesamtheit nur eine Stichprobe von 70 Mitarbeitern gewählt. Alter und Geschlecht sind die Merkmale der Befragung, also der Stichprobe. Die befragten 70 Beschäftigten sind die Merkmalsträger. Ihre Antworten, die in der Urliste erfasst sind, liefern die Rohdaten für die Merkmalsausprägung.

Merkmal	Merkmalsausprägung
Alter	35; 17; 58; ...
Geschlecht	männlich; weiblich

Urliste:

Beschäftigte	Alter	Geschlecht
1	35	m
2	17	w
...
70	43	w

Es folgt nun eine Verarbeitung der Rohdaten. Dabei werden diese nach Kriterien der Fragestellung sortiert und dann gebündelt zusammengefasst. Somit entsteht aus der Rohdatenmenge eine Menge von **verdichteten Daten**. Diese geordneten Daten, die nach der Datengewinnung durch die Datenaufbereitung entstanden sind, bilden die Grundlage zur Beantwortung der eingangs formulierten Fragestellung. Sie können in **Strichlisten** dargestellt werden.



verdichtete Daten: verarbeitete Rohdaten

Strichliste: Verdeutlichung der Merkmalsausprägung

Beispiel:

Von den 70 befragten Personen sind 40 männlichen Geschlechts und 30 weiblichen Geschlechts. Dies kann man sehr schön an der erstellten Strichliste für das Merkmal Geschlecht ablesen.

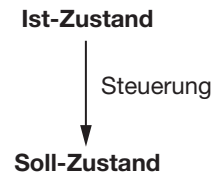
Strichliste (Merkmal Geschlecht)
 männlich |||| |||| |||| |||| |||| |||| ||||
 weiblich |||| |||| |||| |||| ||||

Für das Merkmal Alter ist es von Vorteil, spezifische Altersgruppen zu betrachten. Da in unserem Beispiel die zu berücksichtigende Altersgrenze bei 50 Jahren liegt, ist es sinnvoll, die Altersgruppen so zu wählen, dass eine Grenze bei 50 Jahren auftritt. Obwohl zwei Altersgruppen ("unter 50" und "über 50") ausreichen würden, ist es für die weitere Altersstrukturbetrachtung des Betriebs ratsam, eine detaillierte Einteilung vorzunehmen. Wir wählen deshalb für unsere Strichliste die Merkmalsausprägungen unter 30 Jahre, 30 bis 40 Jahre, 40 bis 50 Jahre, 50 bis 60 Jahre und über 60 Jahre. Auch aus der detaillierten Strichliste lässt sich ablesen, dass in diesem Betrieb zum jetzigen Zeitpunkt 42 Personen eine Merkmalsausprägung Alter unter 50 Jahren haben.

Strichliste (Merkmal Altersgruppe)
 unter 30 J. |||| ||
 30 – 40 J. |||| |||| ||||
 40 – 50 J. |||| |||| |||| ||||
 50 – 60 J. |||| |||| |||| ||||
 über 60 J. |||| ||||

Strichliste (Merkmal Alter)
 unter 50 J. |||| |||| |||| |||| |||| |||| |||| ||||
 über 50 J. |||| |||| |||| |||| ||||

Die Auswertung von Strichlisten führt zur Erkennung des **Ist-Zustandes** zum Zeitpunkt der Befragung. Möchte man darüber hinaus Prognosen für die weitere Entwicklung stellen, dann sind die Merkmalsausprägungen so zu steuern, dass sie den gestellten Anforderungen und somit dem erwünschten **Soll-Zustand** entsprechen. Die Interpretation der verdichteten Daten beantwortet also die eingangs festgelegte Fragestellung und schließt den Kreislauf.



Beispiel:

Der Ist-Zustand besagt, dass von 70 befragten Mitarbeitern 30 weiblichen und 40 männlichen Geschlechts sind. Dies entspricht einem Prozentsatz für die weiblichen Beschäftigten von ca. 43%. Da vom Betriebsrat gefordert wird, dass mindestens 40% der Beschäftigten weiblichen Geschlechts sein sollen, erfüllt der Ist-Zustand bereits den gewünschten Soll-Zustand. Es besteht also kein regulativer Handlungsbedarf.

Bezüglich der Altersstruktur zeigt sich ein Ist-Zustand von 42 unter 50-jährigen und 28 über 50-jährigen der insgesamt 70 befragten Personen aus unserer Stichprobe. Dies entspricht einem Prozentsatz von 40% für die über 50-jährigen Arbeitnehmer. In diesem Bereich besteht Handlungsbedarf, da die Forderung des Betriebsrats für den Soll-Zustand besagt, dass höchstens 20% der Mitarbeiter über 50 Jahre alt sein sollen.

Ergebnis: Bei Neueinstellungen kann bezüglich des Merkmals Geschlecht wie bisher verfahren werden, bezüglich des Merkmals Alter muss regulierend eingegriffen werden.

Ist-Zustand	(Stichprobenergebnis)
43%	weiblich
57%	männlich
40%	über 50 Jahre
60%	unter 50 Jahre

Soll-Zustand	(Betriebsratforderung)
≥ 40%	weiblich
< 60%	männlich
≤ 20%	über 50 Jahre
> 80%	unter 50 Jahre

1.2 Datenaufbereitung

Wenn man Daten gesammelt hat, gibt es eine Vielfalt von Möglichkeiten, sie darzustellen. Eine Art der Darstellung ist die Strichliste, mit der im Beispiel des Kapitels 1.1 bereits gearbeitet wurde. Es ist die einfachste Art, Daten nach bestimmten Merkmalen geordnet aufzubereiten. Möchte man dabei aber mehrere Merkmale zu einer Einheit zusammenfassen, so bildet man Merkmalsklassen. Aus diesen Merkmalsklassen lassen sich die Ausprägungen der Merkmale ablesen. Es ist also immer dann sinnvoll, Merkmalsklassen zu bilden, wenn Merkmale mit sehr vielen Merkmalsausprägungen untersucht werden.

Jede W-Frage (Frage nach einem bestimmten Zweck) liefert durch ihre Beantwortung eine Merkmalsausprägung.

Fragen wir zum Beispiel nach der Schneehöhe in Rabenberg für die Jahre 1950 bis 2000 jeweils am 24. Dezember, so liefern die Beobachtungswerte ein Merkmal mit unterschiedlichen Merkmalsausprägungen.

Die Liste ist durch ihre zeitliche Abfolge nach Jahrgängen bereits so weit aufbereitet, dass man sofort ablesen kann, wann in Rabenberg grüne Weihnacht war oder an welchem Weihnachtsabend der Schneeräumdienst ausrücken musste.

Hätte man die Liste nicht nach Jahrgängen aufsteigend, sondern nach Schneehöhen abfallend sortiert, so wären die Merkmalsausprägungen nach einer anderen eindeutigen Rangfolge geordnet. In jeder Darstellungsart sind die Daten so aufbereitet, dass wunschgemäß der Informationsgehalt erkennbar ist.

Merkmalsklasse:

Zusammenfassung von unterschiedlichen Merkmalen

Merkmalsausprägung:

spezifische Daten der Merkmale

Schneehöhe in Rabenberg

(am 24. Dezember)

Jahr	Schneehöhe in Metern
1950	1,20
1960	0,45
1970	0,00
1980	1,00
1990	1,10
2000	0,60

Schneehöhe in Metern	Jahr
1,20	1950
1,10	1990
1,00	1980
0,60	2000
0,45	1960
0,00	1970

Beispiel:

Am 01. April 2009 wurde bestimmt, dass jeder ein Gesundheitsbuch für die Einstufung seines Krankenkassenbeitrags zu führen hat. Neben der Beschreibung der Ess- und Trinkgewohnheiten, der Bewegungsaktivitäten und der Schlafzeiten wurde der Eintrag von Geschlecht, Körpergröße, Gewicht und Alter gefordert. Und genau diese zuletzt genannten Merkmale wurden in einer Telekollegklasse abgefragt. Es ergab sich nebenstehende nicht personalisierte aufbereitete Liste.

Liste

Geschlecht	m	m	...	w
Gewicht in kg	70	86		75
Körpergröße in m	1,84	1,86		1,78
Alter in Jahren	30	41		22

Tabelle:
geordnete Darstellung von Daten

Bei den qualitativen Merkmalen wie in unserem Beispiel das Geschlecht ist die Angabe einer Einheit nicht sinnvoll. Merkmale, die keine eindeutige Rangfolge besitzen, also nur Namen oder Bezeichnungen beinhalten, die nur der Kennzeichnung dienen, erfasst man in einer **Nominalskala**.

Ab der Datenaufbereitung in Skalen ist eine Rangordnung vorteilhaft. Aus einer ungeordneten **Liste** wird eine geordnete **Tabelle**.

Unsere befragte Telekollegklasse hat 18 Teilnehmer. Für das Merkmal Geschlecht kann eine Nominalskala mit den Merkmalsausprägungen männlich und weiblich gebildet werden.

Aus der Nominalskala ist ersichtlich, dass die Klasse von 10 weiblichen und 8 männlichen Teilnehmern frequentiert ist.

Nominalskala

m	w
	x
x	
x	
	x
	x
x	
	x
x	
	x
	x
	x
x	
x	
	x