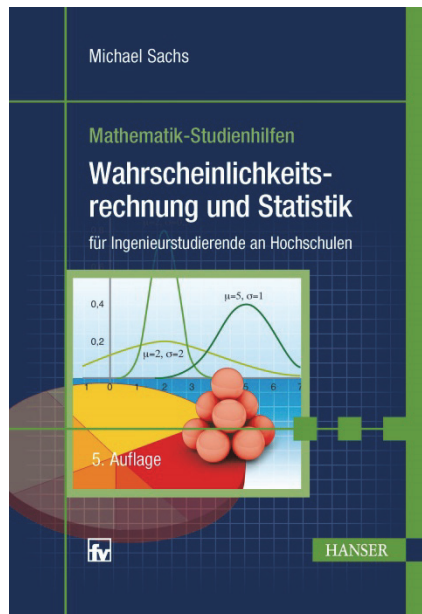


HANSER



Leseprobe

zu

„Wahrscheinlichkeitsrechnung und Statistik“

von Michael Sachs

ISBN (Buch): 978-3-446-45163-6

ISBN (E-Book): 978-3-446-45620-4

Weitere Informationen und Bestellungen unter
<http://www.hanser-fachbuch.de/978-3-446-45163-6>

sowie im Buchhandel

© Carl Hanser Verlag, München

Vorwort

„Wozu brauchen wir das alles?“

So lautet eine häufig gestellte Frage von Studierenden, besonders an Hochschulen für angewandte Wissenschaften, wo immer der Aspekt der Anwendung im Vordergrund steht. Der vorliegende Band aus der Reihe „Studienhilfen Mathematik“ versucht, für den Bereich der Statistik eine befriedigende Antwort auf diese Frage zu geben. Er richtet sich an Studierende der Ingenieurwissenschaften im Grundstudium. In allen technischen Studiengängen machen Studierende bei der Durchführung von Versuchen die Erfahrung von zufälligen Einflüssen. Die Erforschung von deren Gesetzen ist Gegenstand der **Wahrscheinlichkeitsrechnung**, ihre Anwendung Gegenstand der **schließenden Statistik**. Am Anfang steht ein Kapitel über **beschreibende Statistik**, das vollkommen ohne den Wahrscheinlichkeitsbegriff auskommt. Als Voraussetzung für das Verständnis ist hier nicht mehr nötig als die Kenntnis der vier Grundrechenarten und des Wurzelziehens. Trotz der vergleichsweise einfachen mathematischen Methoden sind Grundkenntnisse aus der beschreibenden Statistik unverzichtbar. Viele Begriffe aus der schließenden Statistik sind leichter zu verstehen, wenn man ihr Analogon aus der beschreibenden Statistik kennt. Verhältniszahlen und Zeitreihen werden wegen der gebotenen Kürze nicht behandelt.

Es kommen keine schwierigen Beweise oder umfangreiche Theorien vor. Ingenieure dürfen sich hier getrost auf die gesicherten Ergebnisse der Mathematiker verlassen und sollen vielmehr die Aussagen verstehen und richtig einschätzen lernen, die hinter solchen Sätzen stehen, sowie die Methoden sinnvoll anwenden und ihre Ergebnisse korrekt interpretieren können. Aus der höheren Mathematik werden Kenntnisse der elementaren Funktionen einer reellen Veränderlichen und ihrer Ableitungen sowie des Riemannschen Integrals vorausgesetzt.

Jeder Abschnitt stellt in Lehrsätzen und gelösten Aufgaben die Hilfsmittel bereit, mit denen man die anschließenden Übungsaufgaben lösen kann. Die Ergebnisse der Aufgaben sind zur Kontrolle im Anhang angegeben.

Für die 5. Auflage wurden Druckfehler korrigiert, Zeichnungen verschönert und Beispiele sowie das Literaturverzeichnis aktualisiert. Ich danke allen Kolleginnen und Kollegen, insbesondere Herrn Olaf LOTTER und Frau Anja NOSER für die zahlreichen Hinweise sowie den Mitarbeiterinnen vom Fachbuchverlag Leipzig für die stets angenehme Zusammenarbeit.

München, im Februar 2018

Michael Sachs

Inhaltsverzeichnis

1	Wozu Statistik?	7
2	Beschreibende Statistik	10
2.1	Grundbegriffe.....	10
2.2	Eindimensionale Häufigkeitsverteilungen.....	15
2.3	Kumulierte Häufigkeiten und empirische Verteilungsfunktion.....	20
2.4	Lageparameter.....	26
2.5	Streuungsparameter.....	39
2.6	Zweidimensionale Häufigkeitsverteilungen.....	46
2.7	Korrelationsrechnung.....	53
2.8	Regressionsrechnung.....	58
3	Wahrscheinlichkeitsrechnung	66
3.1	Kombinatorische Grundlagen.....	66
3.2	Zufall, Ereignisalgebra.....	68
3.3	Wahrscheinlichkeit und Satz von Laplace.....	76
3.4	Unabhängige Ereignisse und bedingte Wahrscheinlichkeit.....	83
3.5	Zufällige Variable und Wahrscheinlichkeitsverteilungen.....	91
3.6	Erwartungswert und Varianz einer Verteilung.....	101
3.7	Wichtige diskrete Verteilungen.....	112
3.8	Die Normalverteilung.....	120
4	Schließende Statistik	132
4.1	Problemstellung, Zufallsstichproben.....	132
4.2	Punktschätzungen.....	134
4.3	Intervallschätzungen.....	145
4.4	Hypothesentests.....	162
A	Tabellen	179
B	Lösungen der Übungsaufgaben	182
	Literaturverzeichnis	193
	Sachwortverzeichnis	194

2.6 Zweidimensionale Häufigkeitsverteilungen

In diesem Abschnitt betrachten wir bei einer statistischen Masse **zwei** Merkmale X und Y und die damit verbundenen Besonderheiten. Beispielsweise können wir bei Menschen die Merkmale „Alter“ und „Größe“ erheben, oder bei Werkstücken „Durchmesser“ und „Masse“. Diese zweidimensionale Datenmenge liegt zunächst als Urliste vor und kann in eine zweidimensionale Häufigkeitstabelle überführt und auch grafisch dargestellt werden.

Urliste, Häufigkeitstabelle und Streudiagramm

Die **Urliste** gibt zu jedem statistischen Element i die beiden Merkmalsausprägungen x_i von X und y_i von Y an nach folgendem Schema:

Element Nr.	1	2	...	i	...	n
Ausprägung von X	x_1	x_2	...	x_i	...	x_n
Ausprägung von Y	y_1	y_2	...	y_i	...	y_n

Bei größeren Datenmengen ist es aus Gründen der Übersichtlichkeit wieder sinnvoll, alle Elemente zusammenzufassen, die in beiden Ausprägungen übereinstimmen. Wir erhalten so eine **zweidimensionale Häufigkeitstabelle**, auch **Kontingenztafel** genannt. Dabei bezeichnen die a_j , $j = 1, \dots, l$, die verschiedenen Ausprägungen von X , die b_k , $k = 1, \dots, m$, die verschiedenen Ausprägungen von Y :

		Y						Verteilung von X
		b_1	b_2	...	b_k	...	b_m	
X	a_1	h_{11}	h_{12}	...	h_{1k}	...	h_{1m}	$h_{1.}$
	a_2	h_{21}	h_{22}	...	h_{2k}	...	h_{2m}	$h_{2.}$
	\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
	a_j	h_{j1}	h_{j2}	...	h_{jk}	...	h_{jm}	$h_{j.}$
	\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
	a_l	h_{l1}	h_{l2}	...	h_{lk}	...	h_{lm}	$h_{l.}$
Vert. von Y		$h_{.1}$	$h_{.2}$...	$h_{.k}$...	$h_{.m}$	n

Es gilt die in der Matrizenrechnung übliche Konvention für Doppelindizes: Der erste Index bezeichnet die Zeile, der zweite die Spalte. h_{jk} steht also in Zeile j , Spalte k , und es gilt somit

$$h_{jk} = \text{Anzahl Elemente mit } (X = a_j) \text{ und } (Y = b_k). \quad (2.46)$$

Die Werte am rechten und unteren Rand der Tabelle werden als **Randhäufigkeiten** bezeichnet. Sie geben die Häufigkeitsverteilungen von X allein bzw. Y allein an. $h_{j\cdot}$ ist die Summe über die Zeile j , dagegen $h_{\cdot k}$ die Summe über die Spalte k :

$$h_{j\cdot} = \text{Anzahl Elemente mit } (X = a_j) = \sum_{k=1}^m h_{jk} \quad \text{und} \quad (2.47)$$

$$h_{\cdot k} = \text{Anzahl Elemente mit } (Y = b_k) = \sum_{j=1}^l h_{jk}. \quad (2.48)$$

Die Summe aller Randhäufigkeiten muss jeweils gleich n sein:

$$\sum_{j=1}^l h_{j\cdot} = \sum_{k=1}^m h_{\cdot k} = n. \quad (2.49)$$

Bei stetigen Merkmalen und großem Umfang n können anstelle der a_j und b_k wieder Klassen K_j und L_k von Merkmalsausprägungen treten.

Die der zweidimensionalen Urliste angemessene grafische Darstellung von quantitativen Daten ist das **Streudiagramm**, aus nahe liegenden Gründen auch **Punktwolke** genannt: Jedes Paar von Ausprägungen (x_i, y_i) wird dabei als Punkt in einem (x, y) -Koordinatensystem repräsentiert.

Beispiel 2.24

Bei zwölf Kraftfahrzeugen eines Fuhrparks werden die Merkmale „Alter“ (in Jahren) und „gefahrte Kilometer“ (in 1 000 km) erhoben. Es ergibt sich folgende Urliste:

Nr.	Alter in Jahren	Strecke in 1 000 km	Nr.	Alter in Jahren	Strecke in 1 000 km
1	1,5	30	7	1,8	21
2	5,2	68	8	4,2	112
3	4,5	90	9	6,2	230
4	0,5	12	10	3,6	120
5	2,4	100	11	2,5	56
6	2,6	62	12	5,1	109

Man erstelle das dazugehörige Streudiagramm.

Lösung: Deutlich ist ein Zusammenhang erkennbar: Ältere Autos haben offenbar mehr gefahrene Kilometer, was plausibel ist. Man kann jedoch keine

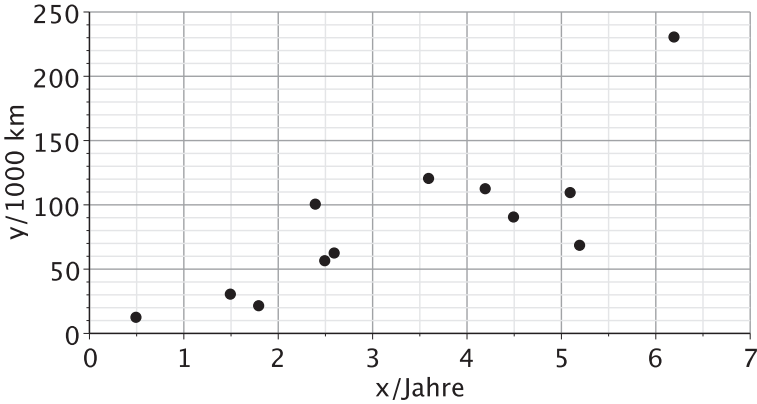


Bild 2.9: Streudiagramm für Alter und gefahrene Kilometer

allgemein gültige Formel angeben, mit deren Hilfe sich die gefahrenen Kilometer aus dem Alter ausrechnen lassen. Es gibt also wohl einen Zusammenhang, aber keine **funktionale Abhängigkeit** zwischen den beiden Merkmalen. ■

Arithmetisches Mittel und Varianz

Arithmetische Mittel und Varianzen für die einzelnen Merkmale X und Y werden wie in (2.15) bzw. (2.38) erklärt, wobei lediglich darauf zu achten ist, dass bei der Berechnung dieser Maßzahlen aus der zweidimensionalen Häufigkeitstabelle die entsprechenden Randhäufigkeiten $h_{j\cdot}$ bzw. $h_{\cdot k}$ einzusetzen sind:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{j=1}^l a_j h_{j\cdot}, \quad (2.50)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{k=1}^m b_k h_{\cdot k}, \quad (2.51)$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{j=1}^l (a_j - \bar{x})^2 h_{j\cdot}, \quad (2.52)$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{k=1}^m (b_k - \bar{y})^2 h_{\cdot k}. \quad (2.53)$$

Kovarianz

Die wirklich neue Maßzahl, an der beide Messreihen gleichzeitig beteiligt sind, ist die **Kovarianz**.

Definition 2.14

Die (empirische) **Kovarianz** der Wertepaare $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ist erklärt durch

$$\text{Cov}(X, Y) = s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \quad (2.54)$$

Ihre Einheit ist das Produkt der Einheiten der x_i und der y_i .

Man beachte, dass das Formelzeichen für die Kovarianz s_{xy} lautet und nicht etwa (wie bei der Varianz) s_{xy}^2 . Dies wäre auch sicherlich nicht sinnvoll, da die Kovarianz auch negativ werden kann, wie wir gleich sehen werden.

Für die praktische Berechnung mit Taschenrechner ist wieder die Umformung

$$s_{xy} = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right) \quad (2.55)$$

effizienter, die man durch Ausmultiplizieren der Klammern unter der Summe in (2.54) erhält.

Die Kovarianz lässt sich auch leicht aus der zweidimensionalen Häufigkeitstabelle errechnen:

$$s_{xy} = \frac{1}{n-1} \left(\sum_{j=1}^l \sum_{k=1}^m a_j b_k h_{jk} - n \bar{x} \bar{y} \right). \quad (2.56)$$

Um die Kovarianz auch inhaltlich zu verstehen, ist dagegen die Darstellung (2.54) am besten geeignet. Wir verwenden zur Illustration die Daten von Beispiel 2.24 mit den Lebensdauern und gefahrenen Kilometern von zwölf Kraftfahrzeugen und legen durch das Streudiagramm neue Koordinatenachsen, die genau durch \bar{x} und \bar{y} hindurchgehen (in Bild 2.10 gestrichelt). Dadurch wird das Streudiagramm in vier Quadranten eingeteilt, die folgende Eigenschaften haben:

- In I liegen die Punkte mit $x_i > \bar{x}$ und $y_i > \bar{y}$, in III die Punkte mit $x_i < \bar{x}$ und $y_i < \bar{y}$. In I und III gilt also: $(x_i - \bar{x})(y_i - \bar{y}) > 0$.

Sachwortverzeichnis

- Abhängigkeit 83, 103
 - von Ereignissen 83
 - von zufälligen Variablen 103
- Ablehnungsbereich 165
- absolute Häufigkeit 15, 18
- Alternativhypothese 163 ff.
- Anteilssatz 135
- arithmetisches Mittel 26 ff., 48
- Ausgleichsrechnung 58
- Ausreißer 34
- Axiome von Kolmogorov 78
- barometrische Höhenformel 65
- Bayes, Satz von 89
- bedingte Wahrscheinlichkeit 85 ff.
- Bernoulli-Experiment 112
- Bernoulli-Kette 112, 115
- beschreibende Statistik 7, 10 ff.
- Besetzungsdichte 18, 35
- Besetzungszahlen 175 f.
- Bestimmtheitsmaß 53 f.
- bimodale Verteilung 36
- Binomialkoeffizient 67
- Binomialverteilung 112 ff.
- binomische Formel 68
- χ^2 -Anpassungstest 174 ff.
- χ^2 -Verteilung 159 f., 176
- de Moivre, Grenzwertsatz von 129 f.
- de Morgansche Regeln 74
- deskriptive Statistik 7, 10 ff.
- Dichtefunktion 97 ff.
- diskrete zufällige Variable 93 ff.
- diskretes Merkmal 11
- Dreieckverteilung 111, 127 f.
- Effusiometer 140, 150, 161
- Einfallsklasse 25
- Elastizitätsmodul 144
- Elementarereignis 69, 79
- Elementarereignisraum 69
- empirische Regressionsgerade 60 ff.
- empirische Verteilungsfunktion 22 ff.
- Entscheidungsregel 164 ff.
- Ereignis(se) 70 ff.
 - , abhängige 83
 - , disjunkte 73
 - , komplementäres 72, 74, 79
 - , sicheres 70, 78
 - , unabhängige 83 ff.
 - , unmögliches 70, 78
- Ereignisalgebra 69, 72 ff.
- Ergebnisraum 69
- Erwartungswert 101 ff.
- Exponentialverteilung 98, 111
- Fakultät 66
- Fehler 1. Art 165 f.
- funktionale Abhängigkeit 48, 105
- Gauß, C. F. 59, 121
- Gaußsche Glockenkurve 121
- Gauß-Verteilung 121 ff.
- Geiger-Müller-Zählrohr 117
- geometrisches Mittel 26, 37 ff.
- Gesetz der großen Zahlen 110
- goodness-of-fit 174
- Gosset, W. S. 150
- Grad des Vertrauens 146
- Grenzwertsatz von de Moivre 129 f.
- Grundgesamtheit 113, 119, 132 ff.
- Güte(funktion) 167 f.
- Häufigkeit 15, 18, 46
 - , kumulierte 20 f.
- Häufigkeitstabelle 15, 46
 - , klassierte 27, 32, 42
 - , kumulierte 20 ff.
 - , unklassierte 27, 32, 42
- Häufigkeitsverteilung 15 ff., 46
 - , eindimensionale 15
 - , zweidimensionale 46
- häufigster Wert 26, 35 ff.
- Histogramm 19

- hoch-signifikantes Ergebnis 166
 hypergeometrische Verteilung 117 ff.
 Hypothese 162 ff.
 Hypothesentest 134, 162 ff.
 induktive Statistik 8, 132 ff.
 Interpolation 25
 Intervallschätzung 134, 145 ff.
 Irrtumswahrscheinlichkeit 165
 kartesisches Produkt 75
 Klassenbildung 17, 22
 Klassenbreite 18
 Klassendichte 18
 Klassenmitte 18
 klassierte Häufigkeitsverteilung 17
 Kolmogorov, A. N. 78
 Kombinatorik 66 ff.
 Komplement 72, 74 f., 79
 Konfidenzintervall 146 ff.
 - für die Steigung der
 Regressionsgerade 157 f.
 - für die Varianz 159 ff.
 - für μ 146 ff.
 - für p 152 f.
 - für zwei Erwartungswerte 153 f.
 - für zwei Wahrscheinlichkeiten 155 f.
 Kontingenztafel 46
 Korrelation 50 ff.
 Korrelationskoeffizient 53 f., 143
 Korrelationsrechnung 52, 53 ff.
 Kovarianz 49 ff.
 kritischer Bereich 165
 kumulierte Häufigkeiten 20 ff.
 Kurvenanpassung 58 ff.
 Lageparameter 26 ff.
 Laplace, Satz von 79
 Laplace-Annahme 79
 Leitfaden 139 f.
 lineare Regression 59 ff.
 lineare Transformation 30
 links-steile Verteilung 37
 Macht, Mächtigkeit 167 f.
 Median 31 ff.
 Merkmal 10
 Merkmalsausprägung 10
 Merkmalsträger 10
 Methode der kleinsten Quadrate 59 ff.
 Modalklasse 35 ff.
 Modalwert 35 ff.
 Modus 35 ff.
 Normalverteilung 98, 120 ff.
 Nullhypothese 163 ff.
 Parameter 133 ff.
 parametrischer Test 162
 Poisson, S. 115
 Poisson-Verteilung 115 ff.
 Power 167 f.
 Permutationen 66
 Primärstatistik 12
 Punktschätzung 134 ff.
 Punktwolke 47, 58, 64
 Quantil 24, 40, 147
 Quartil 40
 Quartilsabstand 39
 radioaktiver Zerfall 117 f.
 Randhäufigkeiten 47
 Rangkorrelationskoeffizient 57
 Realisationen 92, 125, 135
 Rechteckverteilung 98 f., 139
 rechts-steile Verteilung 37
 Regressionsgerade 60 ff., 141 ff., 157
 Regressionsparameter 62
 Regressionsrechnung 52, 58 ff.
 relative Häufigkeit 15, 18, 76 f.
 repräsentative Stichprobe 12
 Residuum 62
 Schätzfunktion 134 ff.
 Schätzprinzip 134
 Scheinkorrelation 55
 schließende Statistik 8, 132 ff.
 Sekundärstatistik 12
 signifikantes Ergebnis 165
 Signifikanzniveau 165
 Spannweite 39 f.
 Stabdiagramm 17, 94 f.

- Standardabweichung 39, 41 ff., 106 ff.
- einer statistischen Masse 41 ff.
- einer zufälligen Variablen 106 ff.
standardisierte zufällige Variable 109, 122
Standard-Normalverteilung 122 ff., 164
Standardunsicherheit 139
- des arithmetischen Mittels 139
statistische Elemente 10
statistische Masse 10
stetige Gleichverteilung 98 f.
stetige zufällige Variable 96 ff.
stetiges Merkmal 11, 17
Stichprobe 8, 12, 132 ff.
Stichprobenfunktion 134
Stichprobenmittel 134
Stichprobenvariable 132
Stichprobenvarianz 135
stochastische Konvergenz 110
Streudiagramm 46 ff.
Streuungsparameter 39 ff.
Studentsche t -Verteilung 150, 169
symmetrische Verteilung 104
 t -Test 168 f., 171 f.
 t -Verteilung 150, 169
Tabellengestaltung 13 f.
Teilerhebung 12
Testgröße 163 ff.
totale Wahrscheinlichkeit 88 f.
Totalerhebung 12
Trägheitsmoment 43 f.
Treppenfunktion 23, 94 f.
Unabhängigkeit 83 ff., 103 ff., 132
- von Ereignissen 83 ff.
- von zufälligen Variablen 103 ff., 132
unimodale Verteilung 36
unklassierte Häufigkeitsverteilung 17
Unsicherheit 138 ff., 143 ff.
- beim Messen 138 ff.
- der Steigung 143 f.
unverbundene Stichproben 172
unverfälschte Schätzfunktion 136 ff.
Urliste 15, 27, 31, 41, 46
Varianz 41 ff., 48, 101, 106 ff., 135
- einer zufälligen Variablen 101, 106 ff.
-, empirische 41 ff., 48
Variationskoeffizient 39, 44 f.
Venn-Diagramm 72 f.
verbundene Stichproben 172
Versuch 69
Verteilung 94
Verteilungsfunktion 22, 93, 97
- einer zufälligen Variablen 93, 97
-, empirische 22 ff.
Vertrauensintervall s. Konfidenzintervall
Vertrauensniveau 146
Vertrauenswahrscheinlichkeit 146
Verzinsungsfaktor 37
Wachstumsfaktor 37
Wachstumsrate 37
Wahrscheinlichkeit 76 ff.
-, bedingte 85 ff.
-, totale 88 f.
Wahrscheinlichkeitsbaum 87 f., 187
Wahrscheinlichkeitsdichte 97 ff.
Wahrscheinlichkeitselement 99
Wahrscheinlichkeitsmasse 94
Wahrscheinlichkeitsrechnung 9, 66 ff.
Wahrscheinlichkeitsverteilung 94
-, symmetrische 104
zentraler Grenzwertsatz 127 ff.
Zentralwert 26, 31 ff.
Ziehen mit Zurücklegen 81, 112 f.
Ziehen ohne Zurücklegen 81, 87, 113, 117
zufällige Variable 91 ff.
-, allgemeine 98
-, diskrete 93 ff.
-, standardisierte 109, 122
-, stetige 96 ff.
-, unabhängige 103, 108, 132
Zufall 68 ff.
Zufallsexperiment 68 f.
-, zusammengesetztes 75
Zufallsstichprobe 12, 132 ff.
Zusammenhang zweier Messreihen 55