

Linder / Berchtold

Elementare  
statistische  
Methoden

---

UTB

Birkhäuser

---

Uni-Taschenbücher

# UTB

Eine Arbeitsgemeinschaft der Verlage

Birkhäuser Verlag Basel und Stuttgart

Wilhelm Fink Verlag München

Gustav Fischer Verlag Stuttgart

Francke Verlag München

Paul Haupt Verlag Bern und Stuttgart

Dr. Alfred Hüthig Verlag Heidelberg

Leske Verlag + Budrich GmbH Opladen

J. C. B. Mohr (Paul Siebeck) Tübingen

C. F. Müller Juristischer Verlag – R. v. Decker's Verlag Heidelberg

Quelle & Meyer Heidelberg

Ernst Reinhardt Verlag München und Basel

K. G. Saur München · New York · London · Paris

F. K. Schattauer Verlag Stuttgart · New York

Ferdinand Schöningh Verlag Paderborn

Dr. Dietrich Steinkopff Verlag Darmstadt

Eugen Ulmer Verlag Stuttgart

Vandenhoeck & Ruprecht in Göttingen und Zürich



Arthur Linder  
Willi Berchtold

# Elementare statistische Methoden

Birkhäuser Verlag, Basel, Boston, Stuttgart

A. Linder, Professor em. Universität Genf und ETH Zürich. Honorary Fellow, Royal Statistical Society.

W. Berchtold, Oberassistent für Biometrie an der Eidgenössischen Technischen Hochschule (ETH) in Zürich.

CIP-Kurztitelaufnahme der Deutschen Bibliothek

**Linder, Arthur**

Elementare statistische Methoden/Arthur Linder:

Willi Berchtold. – Basel, Boston, Stuttgart: Birkhäuser, 1979.

(Uni-Taschenbücher; 796)

ISBN 3-7643-0905-9

NE: Berchtold, Willi:

Nachdruck verboten. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen und der Reproduktion auf photostatischem Wege oder durch Mikrofilm, vorbehalten.

© Birkhäuser Verlag Basel, 1979

ISBN 3-7643-0905-9

# Vorwort

Mit dieser Einführung in die elementaren statistischen Methoden soll einem breiten Kreis von Biologen, Medizinern, Ingenieuren, Wirtschaftswissenschaftlern und Soziologen ermöglicht werden, sich die am häufigsten verwendeten statistischen Hilfsmittel zu erarbeiten. Der Leser wird angeleitet, aus Beobachtungen oder Versuchen erhaltene Daten richtig auszuwerten.

Nach den einleitenden Kapiteln 1 und 2 folgen in den Kapiteln 3 und 4 Verfahren zur Auswertung von *Anzahlen* (oder Häufigkeiten), im Kapitel 5 solche für die Auswertung von *Messwerten*. Die einfacheren Methoden der Korrelation und Regression finden sich im Kapitel 6. Neben den klassischen, werden jeweils auch nichtparametrische Methoden angegeben. Im Kapitel 7 findet der Leser eine Uebersicht der mathematischen Grundlagen, wobei wir – der Anlage des ganzen Buches entsprechend – die Beweise nicht ausführen. Diese finden sich beispielsweise bei LINDER (1964).

Wir beabsichtigen keineswegs, alle in der Fachliteratur publizierten Methoden zusammenzutragen. Gestützt auf langjährige Erfahrung haben wir vielmehr jene Verfahren ausgewählt, welche wir als die nützlichsten und allgemein zweckmässigsten halten.

Es ist unsere Absicht, in einem zweiten Band weitergehende statistische Verfahren wie die Streuungserlegung, die mehrfache Regression und die sogenannten multivariaten Methoden darzustellen.

Unser Dank geht an alle, die mit Rat, Kritik oder Beispielen zu diesem Buche beigetragen haben, insbesondere an Herrn Prof. H. L. Le Roy, der unsere Arbeit grosszügig unterstützt hat. Bei der Vorbereitung des Manuskripts für den Druck war Frl. M. Schneeberger wesentlich beteiligt.

Genf und Zürich, März 1978

A. LINDER  
W. BERCHTOLD



# Inhaltsverzeichnis

<b>1</b>	<b>Häufigkeitsverteilung, Lage- und Streuungsmasse</b>	<b>13</b>
1.1	Häufigkeitsverteilung	13
1.2	Lage- und Streuungsmasse	17
1.21	Lagemasse	17
1.22	Streuungsmasse	19
1.3	Durchschnitt und Streuung	22
1.31	Berechnung mit Hilfe der Einzelwerte	22
1.32	Lineare Transformation der Daten	24
1.33	Berechnung bei gruppierten Daten	28
<b>2</b>	<b>Schätz- und Prüfverfahren</b>	<b>32</b>
2.1	Stichprobe und Grundgesamtheit	32
2.2	Schätzen von Parametern	34
2.21	Problemstellung	34
2.22	Schätzkriterien	35
2.23	Maximum-Likelihood-Schätzungen	37
2.3	Prüfen von Hypothesen	41
2.31	Grundbegriffe	41
2.32	Fehler 1. und 2. Art, Trennschärfe	44
2.33	Normalverteilung und $t$ -Test von «Student»	49
2.4	Vertrauensgrenzen	53
2.41	Genauigkeit eines Schätzwertes	53
2.42	Vertrauensgrenzen zu einem Durchschnitt	54
2.43	Vertrauensgrenzen zu einem Anteil und zum Median	54
<b>3</b>	<b>Beurteilen von Anzahlen</b>	<b>56</b>
3.1	Binomialverteilung	56
3.11	Verteilungsfunktion, Erwartungswert, Varianz	56
3.12	Schätzen des Parameters $\pi$	57
3.13	Vertrauensgrenzen	58
3.14	Vergleich zweier Anteile	59
3.15	Binomialtest und Vorzeichentest	60
3.2	Poissonverteilung	62
3.21	Verteilungsfunktion, Erwartungswert, Varianz	62

3.22	Schätzen des Parameters $\lambda$ . . . . .	63
3.23	Vertrauensgrenzen . . . . .	64
3.24	Vergleich zweier Parameter $\lambda_1$ und $\lambda_2$ . . . . .	65
3.3	Negative Binomialverteilung . . . . .	66
3.31	Verteilungsfunktion, Erwartungswert, Varianz . . . . .	66
3.32	Schätzen der Parameter $\kappa$ und $\pi$ . . . . .	68
3.33	Standardabweichungen der Schätzungen $k$ und $p$ . . . . .	72
<b>4</b>	<b>Das Chiquadrat-Prüfverfahren</b> . . . . .	<b>73</b>
4.1	Prüfen von Anzahlen auf gegebene Verhältnisse . . . . .	74
4.11	Allgemeine Formel und Spezialfall von zwei Klassen . . . . .	74
4.12	Orthogonale Zerlegung von $\chi^2$ . . . . .	76
4.2	Unterschied zwischen beobachteter und theoretischer Verteilung . . . . .	80
4.21	Binomialverteilung . . . . .	80
4.22	Poissonverteilung . . . . .	83
4.23	Negative Binomialverteilung . . . . .	86
4.3	Unabhängigkeit bei qualitativen Merkmalen . . . . .	87
4.4	Unabhängigkeit in der Vierfeldertafel . . . . .	91
4.41	Prüfen mit $\chi^2$ ; grosse Anzahlen . . . . .	92
4.42	Prüfen mit $\chi^2$ ; kleinere Anzahlen (Korrektur von Yates) . . . . .	94
4.43	Prüfen mit dem «exakten» Test von R. A. Fisher. . . . .	96
4.5	Weitere Anwendungen von $\chi^2$ . . . . .	99
4.51	Regression mit Anzahlen . . . . .	100
4.52	Prüfen auf Symmetrie . . . . .	102
<b>5</b>	<b>Beurteilen von Durchschnitten und Streuungen</b> . . . . .	<b>105</b>
5.1	Normalverteilung . . . . .	105
5.11	Wahrscheinlichkeitsnetz . . . . .	105
5.12	Prüfen auf Normalität mit dem 3. und 4. Moment. . . . .	108
5.13	Normalitätstest von A. Linder . . . . .	112
5.2	Prüfen von Durchschnitten . . . . .	117
5.21	Abweichung eines Durchschnitts von seinem theoretischen Wert . . . . .	117
5.22	Unterschied zweier Durchschnitte aus unabhängigen Stichproben . . . . .	119
5.23	Vergleich zweier verbundener Stichproben . . . . .	122
5.24	Nichtparametrische Verfahren . . . . .	123
5.3	Prüfen von Streuungen . . . . .	127

5.31	Abweichung einer Streuung von ihrem theoretischen Wert . . . . .	128
5.32	Unterschied zweier Streuungen . . . . .	130
5.4	Vergleich mehrerer Durchschnitte . . . . .	132
5.41	Einfache Streuungszersetzung . . . . .	132
5.42	Orthogonale Vergleiche . . . . .	138
5.43	Mehrfache Vergleiche zwischen Durchschnitten . . . . .	144
5.44	Kruskal-Wallis-Test . . . . .	148
<b>6</b>	<b>Regression und Korrelation . . . . .</b>	<b>151</b>
6.1	Allgemeines . . . . .	151
6.2	Einfache lineare Regression . . . . .	153
6.21	Einführendes Beispiel . . . . .	153
6.22	Berechnung der Regressionsgeraden . . . . .	156
6.23	Prüfen des Regressionskoeffizienten . . . . .	160
6.24	Linearität der Regression . . . . .	162
6.25	Residuen, Normalität und Ausreisser . . . . .	165
6.26	Vertrauensgrenzen . . . . .	172
6.27	Regression durch den Nullpunkt . . . . .	175
6.3	Korrelation . . . . .	177
6.31	Schätzen des Korrelationskoeffizienten . . . . .	178
6.32	Tests und Vertrauensgrenzen . . . . .	180
6.33	Nichtparametrische Zusammenhangsmasse . . . . .	183
<b>7</b>	<b>Theoretische Grundlagen . . . . .</b>	<b>189</b>
7.1	Allgemeines über Wahrscheinlichkeiten und Verteilungen. . . . .	189
7.11	Addition und Multiplikation von Wahrscheinlichkeiten	189
7.12	Wahrscheinlichkeitsverteilung. . . . .	190
7.13	Erwartungswert, Varianz, Kovarianz . . . . .	190
7.2	Einige Wahrscheinlichkeitsverteilungen . . . . .	191
7.21	Binomialverteilung . . . . .	192
7.22	Poissonverteilung . . . . .	192
7.23	Negative Binomialverteilung . . . . .	194
7.24	Multinomialverteilung . . . . .	194
7.25	Normalverteilung . . . . .	195
7.3	Prüfverteilungen . . . . .	198
7.31	Chiquadrat-Verteilung . . . . .	198
7.32	<i>t</i> -Verteilung . . . . .	199
7.33	<i>F</i> -Verteilung . . . . .	200

7.34	Beziehungen zwischen $\chi^2$ -, $t$ - und $F$ -Verteilung . . . . .	200
7.4	Grundlagen für die Prüfverfahren . . . . .	201
7.41	Verteilung von Durchschnitt und Streuung . . . . .	201
7.42	Verteilung des Regressionskoeffizienten . . . . .	202
7.43	Beziehung zwischen $\chi^2$ - und Poisson-Verteilung . . . . .	204
7.44	Beziehung zwischen $F$ - und Binomial-Verteilung . . . . .	204
7.45	Beziehung zwischen $\chi^2$ - und Multinomial-Verteilung . . . . .	205
7.5	Statistische Prüfverfahren . . . . .	206
7.51	Prüfen von Durchschnitten . . . . .	206
7.511	Prüfen eines Durchschnitts . . . . .	206
7.512	Unterschied zwischen zwei Durchschnitten. . . . .	207
7.513	Einfache Streuungszerlegung . . . . .	209
7.52	Prüfen von Streuungen . . . . .	211
7.521	Prüfen einer Streuung . . . . .	211
7.522	Unterschied zwischen zwei Streuungen . . . . .	212
7.53	Prüfen der einfachen linearen Regression . . . . .	212
7.531	Prüfen des Regressionskoeffizienten . . . . .	212
7.532	Linearität der Regression . . . . .	213
7.533	Prüfen der Bestimmtheitsmasses . . . . .	215
7.534	Prüfen des Korrelationskoeffizienten . . . . .	216
7.6	Unterschied zwischen beobachteten und theoretischen Häufigkeiten . . . . .	217
7.61	Prüfen auf gegebene Verhältnisse . . . . .	217
7.62	Unterschied zwischen beobachteter und theoretischer Verteilung . . . . .	217
7.63	Prüfen auf Unabhängigkeit in Mehrfeldertafeln . . . . .	218
7.7	Maximum Likelihood Schätzungen . . . . .	219
7.71	Das Schätzprinzip. . . . .	219
7.72	Näherungslösungen . . . . .	220
7.73	Likelihood für grosses $N$ . . . . .	221
7.74	Mehrere Parameter . . . . .	222
<b>8</b>	<b>Tafeln . . . . .</b>	<b>225</b>
<b>I</b>	<b>Normale Verteilung . . . . .</b>	<b>225</b>
<b>II</b>	<b>Verteilung von <math>\chi^2</math> . . . . .</b>	<b>226</b>
<b>III</b>	<b>Verteilung von <math>t</math> . . . . .</b>	<b>227</b>
<b>IV</b>	<b>Verteilung von <math>F</math> . . . . .</b>	<b>228</b>
<b>V</b>	<b>Vorzeichentest . . . . .</b>	<b>234</b>
<b>VI</b>	<b>Wilcoxontest . . . . .</b>	<b>235</b>
<b>VII</b>	<b>Mann-Whitney-Test . . . . .</b>	<b>236</b>

VIII	Faktoren $q$ zum Test von Newman-Keuls . . . . .	238
9	Verzeichnis der Beispiele . . . . .	239
10	Literatur . . . . .	241
	Namenverzeichnis . . . . .	245
	Sachverzeichnis . . . . .	246



# 1 Häufigkeitsverteilung, Lage- und Streuungsmasse

Eine erste Aufgabe der Statistik besteht darin, Ergebnisse von Beobachtungen derart zusammenzufassen, dass sie auf einfache Art dargestellt werden können. Weiter hat die Statistik zur Aufgabe, diese Ergebnisse in möglichst knapper, aber trotzdem das Wesentliche erfassender Art zahlenmässig zu kennzeichnen; dies geschieht durch die statistischen Masszahlen, von denen der Durchschnitt und die Streuung am häufigsten verwendet werden.

## 1.1 Häufigkeitsverteilung

Die Beobachtungsergebnisse werden in der Regel in chronologischer Folge in Hefte eingetragen oder sonstwie sorgfältig niedergelegt. Man nennt dieses Rohmaterial der statistischen Bearbeitung die *Urliste*.

In der Regel können die beobachteten Werte als eine *Stichprobe* aufgefasst werden, die uns Aufschluss geben soll über eine *Grundgesamtheit*, aus der sie stammen. Die Beziehungen aufzufinden und zu untersuchen, die zwischen Stichprobe und Grundgesamtheit bestehen, ist eine der Hauptaufgaben der Statistik, die wir in den späteren Kapiteln erörtern. Hier betrachten wir vorerst einzig eine Stichprobe, ohne uns um die Grundgesamtheit zu kümmern.

*Beispiel 1.* Urliste der Einzelgewichte von 100 zweiwöchigen Küken in g (Institut für Tierzucht an der ETH, Zürich).

---

107	117	105	106	114	105	113	88	119	116
108	98	104	126	102	100	120	121	87	110
111	114	121	114	104	94	101	94	95	114
101	82	111	108	100	109	92	96	108	108
97	92	112	105	112	100	108	105	97	119
113	102	103	100	94	102	104	110	127	102
109	100	76	101	95	96	118	91	118	107
105	112	92	99	118	100	130	112	110	103
116	115	96	125	97	114	111	101	101	90
122	106	109	116	103	134	86	124	107	107

---

Wenn wir in dieser Urliste auszählen, wie oft jedes einzelne Gewicht vorkommt, erhalten wir die *Häufigkeitsverteilung*. Das leichteste der 100 Küken wiegt 76 g, das schwerste 134 g. Man findet folgende Häufigkeiten:

Gewicht	Häufigkeit	Gewicht	Häufigkeit	Gewicht	Häufigkeit
76	1	96	3	116	3
77	—	97	3	117	1
78	—	98	1	118	3
79	—	99	1	119	2
80	—	100	6	120	1
81	—	101	5	121	2
82	1	102	4	122	1
83	—	103	3	123	—
84	—	104	3	124	1
85	—	105	5	125	1
86	1	106	2	126	1
87	1	107	4	127	1
88	1	108	5	128	—
89	—	109	3	129	—
90	1	110	3	130	1
91	1	111	3	131	—
92	3	112	4	132	—
93	—	113	2	133	—
94	3	114	5	134	1
95	2	115	1		

Diese Häufigkeitsverteilung gibt noch kein einprägsames Bild der Gesamtheit der Werte der Stichprobe; die Besonderheiten zeigen sich weit besser, wenn wir Gewichtsklassen von je 5 g bilden. Man kann die neuen Häufigkeiten aus der vorangehenden Häufigkeitsverteilung durch Addition von je fünf aufeinanderfolgenden Häufigkeiten bilden.

Wenn man die neue Häufigkeitsverteilung aus der Urliste unmittelbar ableiten will, so geschieht dies am einfachsten durch *Stricheln*. Ein Wert der Urliste nach dem andern wird auf der entsprechenden Zeile der nachstehenden Uebersicht durch einen Strich vermerkt, was zu folgendem Ergebnis führt:

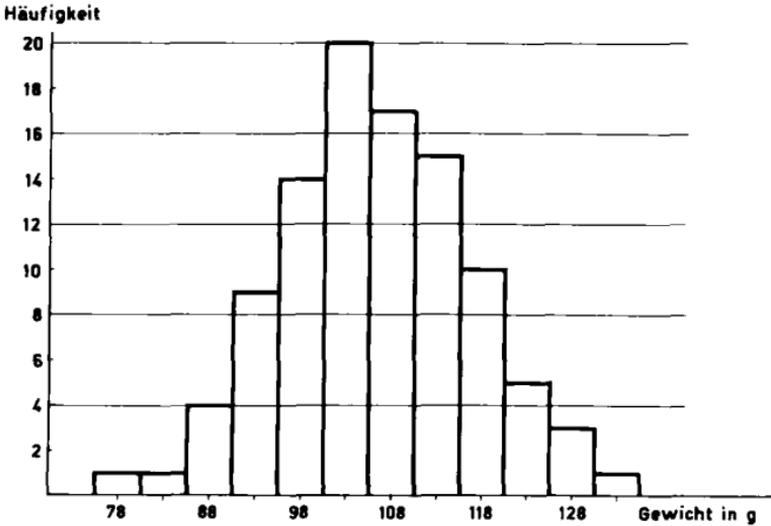
Gewicht in g	Häufigkeit	$f_i$
76- 80		1
81- 85		1
86- 90		4
91- 95		9
96-100		14
101-105		20
106-110		17
111-115		15
116-120		10
121-125		5
126-130		3
131-135		1
Summe		100

Das Stricheln führt nicht nur rasch und einfach zum Ziel, es bietet zudem den Vorteil einer übersichtlichen Darstellung der Häufigkeitsverteilung.

Anstelle der Anzahlen pro Klasse, den *absoluten* Häufigkeiten, rechnet man auch mit den *relativen* Häufigkeiten, d. h. dem Anteil der Beobachtungen, der in die betreffende Klasse fällt. Dies vor allem dann, wenn Verteilungen graphisch dargestellt werden oder miteinander zu vergleichen sind. Drückt man die relativen Häufigkeiten in Prozenten aus, so erhält man im Beispiel 1 für die absoluten und die relativen Häufigkeiten gleiche numerische Werte.

Die Häufigkeitsverteilung stellt man als *Rechteckdiagramm* oder *Histogramm* dar. Zu unserem Beispiel gehört Figur 1; man erhält ein übersichtliches, wenn auch wegen der Zusammenfassung in Klassen von je 5 g, etwas schematisiertes Bild der beobachteten Werte.

Eine andere Art der Darstellung, deren Nutzen im Abschnitt 5.11 ersichtlich wird, bietet die sogenannte *Summenhäufigkeitsverteilung*. Die Summenhäufigkeiten geben an, wieviele der beobachteten Werte kleiner oder gleich einem



Figur 1. Häufigkeitsverteilung der Gewichte von 100 zweiwöchigen Küken.

bestimmten Gewicht sind. Man findet die Summenhäufigkeiten durch fortgesetzte Addition der absoluten Häufigkeiten wie in der folgenden Zusammenstellung:

Gewicht in g	Häufigkeit	Summenhäufigkeit
76- 80	1	1
81- 85	1	2
86- 90	4	6
91- 95	9	15
96-100	14	29
101-105	20	49
106-110	17	66
111-115	15	81
116-120	10	91
121-125	5	96
126-130	3	99
131-135	1	100

Die letzte der Summenhäufigkeiten entspricht selbstverständlich der Gesamtzahl der Beobachtungen bzw. 100%, wenn mit den relativen Häufigkeiten gerechnet wird.

Die Summenhäufigkeiten lassen sich graphisch darstellen wie dies Figur 2 zeigt. In dieser Figur sind im Grunde nur die