

**Frank Lammers**

**Statistik I: deskriptive und  
explorative Statistik**

**Lehr- und Übungsbuch**

**2004**

**Verlag der Gesellschaft für Unternehmens-  
rechnung und Controlling m.b.H.**

## Vorwort zur zweiten Auflage

Mit der zweiten Auflage wurden eine Reihe von kleineren Fehlern, die sich eingeschlichen hatten, korrigiert und Ungenauigkeiten in der Darstellung verbessert. Außerdem wurde der Inhalt um einige statistische Auswertungswerkzeuge erweitert. Für jeden Hinweis auf weiterhin vorhandene oder neu hinzugekommene Fehler oder Ungenauigkeiten ist der Autor sehr dankbar.

Wernigerode im Dezember 2004

Frank Lammers

## Vorwort

Herzlich Willkommen in der Statistik! Nachfolgend möchte ich Sie mit einem Bereich vertraut machen, der einer der wichtigsten in unserer modernen, an Zahlen und Fakten orientierten Gesellschaft ist. Wo es möglich ist, werden heutzutage Vorgänge, Sachverhalte, Bewertungen, Einstellungen etc. quantifiziert. Die anfallenden Daten werden gespeichert und schließlich irgendwann ausgewertet. Nicht die Beschaffung von Daten stellt gegenwärtig das zentrale Problem dar, sondern das Herausfiltern sinnvoller Informationen aus riesigen Datenmengen. Im Englischen spricht man in diesem Zusammenhang bereits von „data mining“. Dies assoziiert das Bild von kleinen Mengen wertvoller Rohstoffe, wie beispielsweise Diamanten, die in gigantischen Mengen von Schutt und Abraum versteckt sind und geborgen werden müssen. Die Statistik ist der Königsweg zu diesen Diamanten.

Im Lehrbuch „Statistik I“ beschäftigen Sie sich mit der deskriptiven und explorativen Statistik. Die *deskriptive Statistik* beschreibt Daten und verdichtet sie zu aussagekräftigen Kennwerten. Darüber hinaus stellt sie graphische Darstellungen zur Verfügung, mit deren Hilfe Sie wichtige Aussagen in Ihren Daten sichtbar machen können. Gerade bei großen Datenmengen sind dies wertvolle Hilfsmittel, um sich einen Überblick zu verschaffen oder die Daten zu „begreifen“.

Die Erfassung von Daten erfolgt in der Regel nicht „L’art pour l’art“, um ihrer selbst Willen, sondern weil durch die Daten Fragen beantwortet werden sollen. Hat man keine Theorie zur Hand, auf deren Grundlage sich die Fragen mehr oder minder von selbst ergeben würden, muss man in einem ersten Schritt nach Strukturen in den Daten suchen. Erst solche Auffälligkeiten erlauben es, im nächsten Schritt Hypothesen zu generieren und damit dann die Fragen zu beantworten.

Die *explorative Statistik* hilft, derartige Strukturen oder Auffälligkeiten zu finden. Sie stellt eine Erweiterung des Methodenarsenals der deskriptiven Statistik im Hinblick auf komplexere Auswertungsverfahren dar. Man kann ihren Beginn in etwa auf das Erscheinen des Buchs von Tukey „Exploratory data analysis“ (1977) datieren. Aber erst die Bereitstellung von billiger Rechenkapazität an nahezu jedem Arbeitsplatz durch die PCs hat diesen Verfahren zu ihrem endgültigen Durchbruch verholfen. Im weiteren Verlauf des Buches wird nicht mehr zwischen Verfahren aus der deskriptiven und der explorativen Statistik unterschieden. In den jeweiligen Kapiteln wird zusammengetragen, was beide Bereiche zu bieten haben.

Im Lehrbuch „Statistik II“ gehen wir zur *Inferenzstatistik* über. Immer noch steht im Vordergrund, Fragen über Zusammenhänge, Unterschiede oder Ähnliches zu beantworten. Nachdem wir uns mit den Verfahren der „Statistik I“ einen Überblick über die Daten verschafft haben und möglicherweise in der Lage waren, konkrete Hypothesen zu generieren, geht es in der Folge darum, hinsichtlich der Hypothesen Entscheidungen zu fällen. Haben wir die richtigen Hypothesen aufgestellt, lassen sich mit den Entscheidungen die ursprünglichen Fragen beantworten.

In der Regel können wir nicht alle Objekte vermessen oder alle Personen befragen, die uns interessieren. Denken Sie nur an die Überprüfung der Lebensdauer von Glühlampen oder an alle Autofahrer in Deutschland zwischen 18 und 60 einer bestimmten Automarke. Wir müssen dementsprechend mit einer Teilmenge vorlieb nehmen. Dann aber stellt sich die Frage, ob das, was wir in dieser Teilmenge gefunden haben, auch für die Menge als Ganzes gilt: also der Schluss von einigen auf alle oder von unserer Stichprobe auf die Grundgesamtheit. (Mit der Erläuterung dieser Begriffe werden wir im ersten Kapitel dieses Buches beginnen.) Der Inferenzstatistik kommt bei diesen Schlussfolgerungen eine tragende Rolle zu. Begonnen werden soll aber mit der deskriptiven und explorativen Statistik.

# Inhaltsverzeichnis

<b>Vorwort</b>	<b>I</b>
<b>Inhaltsverzeichnis</b>	<b>II</b>
<b>1 Einführung</b>	<b>2</b>
1.1 Einige wichtige Grundbegriffe	2
1.2 Beispieldatensätze	2
1.3 Skalenniveaus und Typen von Merkmalen oder Variablen	2
1.4 Methoden der Datengewinnung	2
1.5 Übungsaufgaben zum Kapitel 1	2
<b>2 Häufigkeiten und ihre Darstellungen</b>	<b>2</b>
2.1 Absolute und relative Häufigkeiten	2
2.2 Häufigkeiten bei Klassenbildung	2
2.3 Graphische Darstellungen	2
2.3.1 Darstellung diskreter Merkmale mit wenigen Ausprägungen	2
2.3.2 Darstellung stetiger Merkmale oder diskreter Merkmale mit vielen Ausprägungen	2
2.4 Die empirische Verteilungsfunktion	2
2.5 Die empirische Verteilungsfunktion bei Klassenbildung	2
2.6 Übungsaufgaben zum Kapitel 2	2
<b>3 Lagemaße deskriptiver Verteilungen</b>	<b>2</b>
3.1 Das arithmetische Mittel	2
3.2 Der Median	2
3.3 Die Quantile	2
3.4 Der Modus	2
3.5 Weitere Maße der zentralen Tendenz	2
3.6 Übungsaufgaben zum Kapitel 3	2
<b>4 Streuungsmaße deskriptiver Verteilungen</b>	<b>2</b>
4.1 Empirische Varianz, Standardabweichung und Variationskoeffizient	2
4.2 Spannweite, Interquartilsabstand und Box-Plot	2
4.3 Übungsaufgaben zum Kapitel 4	2
<b>5 Schiefe und Wölbung</b>	<b>2</b>
5.1 Die Begriffe „Schiefe“ und „Wölbung“	2
5.2 Maßzahlen der Schiefe	2
5.3 Maßzahlen der Wölbung	2
5.4 Übungsaufgaben zum Kapitel 5	2

<b>6</b>	<b>Analyse der Zusammenhänge in bivariaten Datensätzen</b>	<b>2</b>
6.1	Graphische Darstellung von Zusammenhängen	2
6.1.1	Darstellungen stetiger Merkmale	2
6.1.2	Darstellungen diskreter Merkmale	2
6.2	Zusammenhangsmaße bei metrischen Merkmalen	2
6.3	Zusammenhangsmaße bei ordinalen Merkmalen	2
6.3.1	Rangkorrelationskoeffizient nach Spearman	2
6.3.2	Konkordanzkoeffizient nach Kendall	2
6.4	Zusammenhangsmaße bei nominalen Merkmalen	2
6.4.1	Kontingenztabellen und bedingte Häufigkeiten	2
6.4.2	Zusammenhangsmaße in Kontingenztabellen	2
6.5	Die Interpretation von Korrelationen und ihre Tücken	2
6.5.1	Darstellung anhand binomialer Effektgrößen	2
6.5.2	Tücken der Interpretation von Korrelationen	2
6.6	Übungsaufgaben zum Kapitel 6	2
<b>7</b>	<b>Lineare Regressionsanalyse</b>	<b>2</b>
7.1	Problemstellung und Lösungsansatz	2
7.2	Die Methode der kleinsten Quadrate	2
7.3	Kriterien für die Beurteilung der gefundenen Lösung	2
7.4	Übungsaufgaben zum Kapitel 7	2
	<b>Anhang</b>	<b>2</b>
A.	Beispieldatensätze	2
B.	Lösungen zu den Übungsaufgaben	2
B.1	Lösungen zu Kapitel 1	2
B.2	Lösungen zu Kapitel 2	2
B.3	Lösungen zu Kapitel 3	2
B.4	Lösungen zu Kapitel 4	2
B.5	Lösungen zu Kapitel 5	2
B.6	Lösungen zu Kapitel 6	2
B.7	Lösungen zu Kapitel 7	2
	<b>Literaturverzeichnis</b>	<b>2</b>
	<b>Einige Standardlehrbücher der Statistik</b>	<b>2</b>
	<b>Sachregister</b>	<b>2</b>