



## Inhaltsverzeichnis

2.1	Problemstellung .....	64
2.2	Vorgehensweise .....	69
2.2.1	Modellformulierung .....	70
2.2.2	Schätzung der Regressionsfunktion .....	74
2.2.2.1	Einfache Regression .....	75
2.2.2.2	Multiple Regression .....	83
2.2.3	Prüfung der Regressionsfunktion .....	87
2.2.3.1	Standardfehler der Regression .....	88
2.2.3.2	Bestimmtheitsmaß (R-Quadrat) .....	89
2.2.3.3	Stochastisches Modell der Regression und F-Test .....	91
2.2.3.4	Overfitting und korrigiertes Bestimmtheitsmaß .....	95
2.2.4	Prüfung der Regressionskoeffizienten .....	97
2.2.4.1	Präzision der Regressionskoeffizienten .....	97
2.2.4.2	t-Test der Regressionskoeffizienten .....	99
2.2.4.3	Konfidenzintervall der Regressionskoeffizienten .....	102
2.2.5	Prüfung der Modellprämissen .....	103
2.2.5.1	Nichtlinearität .....	106
2.2.5.2	Vernachlässigung relevanter Variablen .....	109
2.2.5.3	Zufallsfehler in unabhängigen Variablen .....	115
2.2.5.4	Heteroskedastizität .....	118
2.2.5.5	Autokorrelation .....	120
2.2.5.6	Normalverteilung der Störgrößen .....	123
2.2.5.7	Multikollinearität und Präzision .....	124
2.2.5.8	Einflussreiche Ausreißer .....	128
2.3	Fallbeispiel .....	139
2.3.1	Problemstellung .....	139
2.3.2	Durchführung einer Regressionsanalyse mit Hilfe von SPSS .....	139
2.3.3	Ergebnisse .....	142

2.3.3.1	Ergebnisse der ersten Regressionsanalyse .....	142
2.3.3.2	Ergebnisse der zweiten Regressionsanalyse .....	144
2.3.3.3	Prüfung der Annahmen .....	146
2.3.3.4	Schrittweise Regression .....	149
2.3.4	SPSS-Kommandos .....	153
2.4	Modifikationen und Erweiterungen .....	153
2.4.1	Regression mit Dummy-Variablen .....	153
2.4.2	Regressionsanalyse mit Zeitreihendaten .....	156
2.4.3	Multivariate Regression .....	162
2.5	Anwendungsempfehlungen .....	162
	Literatur .....	163

---

## 2.1 Problemstellung

Die Regressionsanalyse ist eine der nützlichsten und daher am häufigsten verwendeten Methoden der statistischen Datenanalyse. Mithilfe der Regressionsanalyse können Beziehungen zwischen Variablen analysiert werden (Tab. 2.2). Dabei kann z. B. herausgefunden werden, ob eine bestimmte Variable durch eine andere Variable beeinflusst wird, und wenn ja, wie stark dieser Effekt ist.

Auf diese Weise kann man lernen, wie die Welt funktioniert. Die Regressionsanalyse kann bei der Suche nach Wahrheiten verwendet werden, was sehr spannend sein kann. Die Regressionsanalyse ist sehr nützlich, wenn man auf der Suche nach Erklärungen ist oder Entscheidungen oder Prognosen treffen will. Daher ist die Regressionsanalyse sowohl für alle empirischen Wissenschaften als auch für die Lösung praktischer Probleme von eminenter Bedeutung. In Tab. 2.1 sind Beispiele für die Anwendung der Regressionsanalyse in unterschiedlichen Anwendungsfeldern aufgeführt.

Unter den Methoden zur multivariaten Datenanalyse nimmt die Regressionsanalyse eine Sonderstellung ein. Die Erfindung der Regressionsanalyse durch Sir Francis Galton (1822–1911) im Zusammenhang mit seinen Studien zur Vererbung<sup>1</sup> kann als die Geburtsstunde der multivariaten Datenanalyse angesehen werden. Stigler (1997, S. 107) bezeichnet sie als „einen der großen Triumphe der Wissenschaftsgeschichte“. Von weiterer Bedeutung ist, dass die Regressionsanalyse eine Grundlage für zahlreiche andere Methoden bildet, die heute bei der Analyse großer Datenmengen und beim maschinellen Lernen eingesetzt werden. Für das Verständnis dieser anderen, oft komplexeren Methoden der multivariaten Datenanalyse ist die Kenntnis der Regressionsanalyse unverzichtbar.

Obwohl die Regressionsanalyse eine relativ einfache Methode im Bereich der multivariaten Datenanalyse darstellt, ist sie dennoch für Fehler und Missverständnisse anfällig: So kommt es häufig zu falschen Ergebnissen oder falschen Interpretationen der Ergebnisse der Regressionsanalyse. Dies betrifft insbesondere die dem Regressionsmodell zugrunde

---

<sup>1</sup>Galton (1886) untersuchte die Beziehung zwischen der Körpergröße von Eltern und ihren erwachsenen Kindern. Er regressierte die Körpergröße der Kinder auf die Körpergröße der Eltern.