



## Inhaltsverzeichnis

5.1	Problemstellung .....	294
5.2	Vorgehensweise .....	301
5.2.1	Modellformulierung .....	303
5.2.1.1	Das lineare Wahrscheinlichkeitsmodell (Modell 1) .....	304
5.2.1.2	Logit-Modell mit gruppierten Daten (Modell 2) .....	306
5.2.1.3	Logistische Regression mit Individualdaten (Modell 3) .....	308
5.2.1.4	Klassifizierung .....	310
5.2.1.5	Multiple logistische Regression (Modell 4) .....	317
5.2.2	Schätzung der logistischen Regressionsfunktion .....	319
5.2.3	Interpretation der Regressionskoeffizienten .....	322
5.2.4	Prüfung des Gesamtmodells .....	330
5.2.4.1	Likelihood-Ratio-Statistik .....	331
5.2.4.2	Pseudo-R-Quadrat-Statistiken .....	333
5.2.4.3	Beurteilung der Klassifizierung .....	335
5.2.4.4	Prüfung auf Ausreißer .....	336
5.2.5	Prüfung der geschätzten Koeffizienten .....	340
5.2.6	Durchführung einer binären Logistischen Regression mit SPSS .....	342
5.3	Multinomiale logistische Regression .....	346
5.3.1	Das multinomiale logistische Modell .....	349
5.3.2	Beispiel und Interpretation .....	351
5.3.3	Das Baseline-Logit-Modell .....	353
5.3.4	Gütemaße .....	356
5.3.4.1	Pearson-Gütemaß .....	357
5.3.4.2	Devianz-Gütemaß .....	359
5.3.4.3	Informationskriterien für die Modellauswahl .....	361
5.4	Fallbeispiel .....	363
5.4.1	Problemstellung .....	363
5.4.2	Durchführung einer Multinomialen Log. Regression mit SPSS .....	364

5.4.3	Ergebnisse .....	369
5.4.3.1	Blockweise Logistische Regression .....	369
5.4.3.2	Schrittweise Logistische Regression .....	377
5.4.4	SPSS-Kommandos .....	379
5.5	Modifikationen und Erweiterungen .....	380
5.6	Anwendungsempfehlungen .....	384
	Literatur .....	386

## 5.1 Problemstellung

Bei vielen Problemstellungen in Wissenschaft und Praxis treten immer wieder die folgenden Fragen auf:

- Welcher von zwei oder mehreren alternativen Zuständen *liegt vor* oder welches Ereignis *wird eintreffen*?
- Welche Faktoren *beeinflussen* das Zustandekommen von bestimmten Ereignissen und welche *Wirkung* haben sie?

Häufig geht es dabei nur um zwei alternative Zustände oder Ereignisse, z. B. hat ein Patient eine bestimmte Krankheit oder nicht? Wird er überleben oder nicht? Wird ein Kreditnehmer seinen Kredit zurückzahlen oder nicht? Wird ein potenzieller Käufer ein Produkt kaufen oder nicht? In anderen Fällen geht es um mehr als zwei Alternativen, z. B. welche Marke wird ein potenzieller Käufer wählen oder welcher Partei wird ein Wähler seine Stimme geben?

Zur Beantwortung derartiger Fragen kann die *logistische Regression* angewandt werden. Die logistische Regression gehört zur Klasse der *strukturen-prüfenden Verfahren*. Sie bildet, wie schon der Name erkennen lässt, eine Variante der Regressionsanalyse. Im Allgemeinen befasst sich die logistische Regressionsanalyse mit Problemen der folgenden Art:

$$Y = f(X_1, X_2, \dots, X_J)$$

wobei die abhängige Variable (Response-Variable)  $Y$  eine kategoriale Variable ist. Die unabhängigen Variablen (Prädiktoren) können metrische oder auch kategoriale Variablen sein. Die logistische Regressionsanalyse zählt heute zu den wichtigsten Methoden zur Analyse von Problemen mit kategorialen Phänomenen.

Bei der abhängigen Variablen  $Y$  handelt es sich um eine kategoriale Variable, deren Ausprägungen ( $g = 1, \dots, G$ ) die möglichen Alternativen (Gruppen, Response-Kategorien) repräsentieren. Da das Eintreffen von Ereignissen meist mit Unsicherheit behaftet ist, wird  $Y$  als eine Zufallsvariable betrachtet, und es werden die Wahrscheinlichkeiten für die Kategorien von  $Y$  prognostiziert. Das Ziel der logistischen Regressionsanalyse besteht dann darin, Wahrscheinlichkeiten für die Prognose von Ereignissen abzuschätzen: