

Krafft, Manfred

Working Paper — Digitized Version

Der Ansatz der logistischen Regression und seine Interpretation

Manuskripte aus den Instituten für Betriebswirtschaftslehre der Universität Kiel, No. 402

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Institute of Business Administration

Suggested Citation: Krafft, Manfred (1996) : Der Ansatz der logistischen Regression und seine Interpretation, Manuskripte aus den Instituten für Betriebswirtschaftslehre der Universität Kiel, No. 402, Universität Kiel, Institut für Betriebswirtschaftslehre, Kiel

This Version is available at:

<https://hdl.handle.net/10419/149034>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Manuskripte
aus den
Instituten für Betriebswirtschaftslehre
der Universität Kiel

Nr. 402

Manfred Krafft

**Der Ansatz der Logistischen Regression
und seine Interpretation**



Nr. 402

Manfred Krafft

**Der Ansatz der Logistischen Regression
und seine Interpretation**

August 1996

Dr. Manfred Krafft

**Lehrstuhl für Marketing
Christian-Albrechts-Universität zu Kiel
Olshausenstraße 40
24098 Kiel**

**Tel.: (0431) 880 - 1553
Fax: (0431) 880 - 1166
E-mail: krafft@bwl.uni-kiel.de**

Inhaltsverzeichnis

Überblick.....	III
1. Dichotome Probleme in der betriebswirtschaftlichen Forschung.....	1
2. Das Modell der Logistischen Regression.....	1
3. Zur Güte des Schätzmodells.....	4
4. Prüfung und Interpretation einzelner Koeffizienten.....	8
4.1. Partielle Ableitungen als Interpretationshilfe.....	8
4.2. Interpretation mit Hilfe von Elastizitäten.....	10
4.3. Sensitivitätsanalysen.....	12
5. Prognose mit Logistischen Regressionsmodellen.....	14
6. Abschließende Beurteilung der Logistischen Regression.....	15
Literatur.....	16
Zusammenfassung.....	17
Summary.....	17

Der Ansatz der Logistischen Regression und seine Interpretation

Von Manfred Krafft

Überblick

- Bei der Analyse betriebswirtschaftlicher Probleme stößt man häufig auf dichotome Fragestellungen, bei denen herkömmliche statistische Untersuchungsverfahren nicht angewendet werden können.
- Statt der sonst üblichen Zwei-Gruppen-Diskriminanzanalyse wird in jüngster Zeit zunehmend die Logistische Regression eingesetzt, die im vorliegenden Beitrag anhand einer empirischen Applikation vorgestellt wird.
- Neben der Darstellung des Modells der Logistischen Regression wird gezeigt, wie die Koeffizienten einzelner Variablen auf Signifikanz geprüft und inhaltlich interpretiert werden können.
- Aufgrund der geringen Anforderungen an das zu analysierende Datenmaterial, der asymptotischen Test-Statistiken und der Robustheit der Logistischen Regression bietet dieses Verfahren gegenüber herkömmlichen Methoden zwar nennenswerte Vorzüge, sollte aber stets auf seine inhaltliche Anwendbarkeit überprüft werden.

Eingegangen:

Dr. Manfred Krafft, Wissenschaftlicher Assistent am Lehrstuhl für Marketing,
Christian-Albrechts-Universität zu Kiel, 24098 Kiel, e-mail: krafft@bwl.uni-kiel.de
Schwerpunkte: Marketing-Controlling, Neue Institutionenlehre, Verbundeffekte im
Marketing, Verkaufsmanagement

1. Dichotome Probleme in der betriebswirtschaftlichen Forschung

In der Betriebswirtschaftslehre wird sowohl der Praktiker als auch der Forscher häufig mit qualitativen Fragestellungen konfrontiert. Aus den einzelnen Funktionsbereichen seien beispielhaft die Fragen der Eigenerstellung oder des Fremdbezugs (Beschaffung und Produktion), des Kaufs oder Nichtkaufs von Produkten (Marketing) und der Kreditwürdigkeit von Schuldnern (Finanzierung) genannt. Wenn man sich diesen Problemen methodisch nähern will, stellt man schnell fest, daß die herkömmliche Regressionsanalyse ungeeignet ist, solche dichotomen abhängigen Variablen zu analysieren, da eine binäre abhängige Variable zur Verletzung der Prämisse normalverteilter Residuen führt und folglich inferenzstatistische Aussagen unmöglich macht (vgl. *Aldrich und Cnudde*, 1975; *Aldrich und Nelson*, 1984, S. 13 f. und *Urban*, 1993, S. 16 ff.).

Neben der bislang standardmäßig eingesetzten Zwei-Gruppen-Diskriminanzanalyse (vgl. *Backhaus et al.*, 1994, S. 90 ff.) wird in der jüngeren Vergangenheit die Logistische Regression vermehrt zur Analyse von dichotomen Fragestellungen verwendet. Insbesondere die Integration dieses Verfahrens in statistische Standard-Software hat zu seiner Verbreitung beigetragen. Da eine gestraffte Darstellung der Logistischen Regression in der betriebswirtschaftlichen Literatur bisher fehlt, soll in diesem Beitrag zunächst gezeigt werden, wie dieses Schätzverfahren spezifiziert ist. Anhand einer Anwendung auf das Wahlproblem „Handelsvertreter oder Reisende“ wird im folgenden Abschnitt gezeigt, wie das Modell der Logistischen Regression spezifiziert ist. Während im 3. Abschnitt Maße zur Beurteilung der Güte der geschätzten Modelle diskutiert werden, stellt der Autor im 4. Abschnitt mit partiellen Ableitungen, Elastizitäten und Sensitivitätsanalysen drei Verfahren zur Interpretation der geschätzten Koeffizienten vor. Der Beitrag endet mit einem Beispiel zur Prognose auf der Basis Logistischer Regressionsmodelle und abschließenden Bemerkungen.

2. Das Modell der Logistischen Regression

In aktuellen Beiträgen wird für die Analyse dichotomer Fragestellungen vermehrt die Logistische Regression eingesetzt, die sich als robustes Schätzverfahren erwiesen hat und auch Analysen ermöglicht, bei denen die Prämissen der Zwei-Gruppen-Diskriminanzanalyse verletzt sind, so bei der Analyse nicht-metrischer Einflußfaktoren (*Aldrich und Cnudde*, 1975, S. 579 ff.; *Hosmer und Lemeshow*, 1989, S. 20 und *Press und Wilson*, 1978, S. 700). Die wesentlichen Voraussetzungen für den Einsatz der Logistischen Regression sind dabei, daß die Regressoren voneinander unabhängig, also nicht multikollinear sind, und keine Autokorrelation vorliegt (*Aldrich und Nelson*, 1984, S. 49).

Ein weiterer, für empirische Analysen besonders wichtiger Vorteil der Logistischen Regression besteht darin, daß für die mit ihrer Hilfe geschätzten Koeffizienten asymptotisch t -verteilte Statistiken angegeben werden können. Die Konfidenzintervalle der Diskriminanzanalyse sind dagegen nicht interpretierbar (vgl. *Anderson und Coughlan*, 1987, S. 77; *Crask und Perreault*, 1977, S. 60 ff. und *Morrison*, 1969, S. 157 ff.).

Das Schätzverfahren der Logistischen Regression erfährt zwar eine zunehmende Verbreitung, wird aber in deutschsprachigen statistischen Standardwerken kaum erörtert. Daher soll in diesem Beitrag ein kurzer Überblick über das Schätzverfahren der Logistischen Regression, die Beurteilung des gesamten Modells und die Möglichkeiten zur Interpretation der Koeffizienten gegeben werden. Zur Veranschaulichung wird auf die Studie von *Krafft* (1995b) zur Absatzformwahl („Handelsvertreter versus Reisende“) zurückgegriffen, um das Verfahren der Logistischen Regression anhand eines Beispiels zu verdeutlichen. Dabei nimmt die zu erklärende Variable zwei Werte an:

$$(1) \quad y = \begin{cases} 1 & \text{falls Reisende eingesetzt werden} \\ 0 & \text{falls Handelsvertreter eingesetzt werden.} \end{cases}$$

Zur Analyse einer solchen dichotomen abhängigen Variablen kann als einfachstes Verfahren das bekannte Kleinste-Quadrate-Schätzverfahren („linear probability model“) eingesetzt oder eine lineare Diskriminanzfunktion geschätzt werden, worauf hier nicht näher eingegangen werden soll (vgl. dazu *Aldrich und Nelson*, 1984). Im weiteren wird die Alternative betrachtet, daß eine nicht beobachtete („latente“) Variable y^* existiere. Was jedoch beobachtet wird, ist

$$(2) \quad y = \begin{cases} 1 & \text{falls } y^* > 0 \\ 0 & \text{sonst.} \end{cases}$$

Dieses Vorgehen ist Ausgangspunkt der Logistischen Regression. Es wird im einzelnen davon ausgegangen, daß folgendes Regressionsmodell gegeben sei:

$$(3) \quad y_i^* = \beta_0 + \sum_{k=1}^K \beta_k \cdot x_{ik} + u_i, \text{ wobei}$$

- y_i^* : Nicht beobachtete Variable beim i -ten Objekt ($i \in I$),
- β_0 : Konstante,
- β_k : Koeffizient der k -ten unabhängigen Variablen x_{ik} ($k \in K$),
- x_{ik} : Ausprägung der k -ten unabhängigen Variablen ($k \in K$) beim i -ten Objekt ($i \in I$),
- u_i : Störterm,
- I : Indexmenge der Objekte,
- K : Indexmenge der unabhängigen Variablen.

Im Gegensatz zum Kleinste-Quadrate-Schätzverfahren gehen wir somit davon aus, daß eine latente Variable y_i^* existiert, für die eine dichotome Realisierung y_i beobachtet wird. Im Anwendungsfall ist die beobachtete Variable der (Nicht-) Einsatz von Reisenden, während y_i^* als Wahrscheinlichkeit oder Neigung anzusehen ist, daß Reisende als Absatzform gewählt werden. Aus *Gleichung (2)* und *(3)* kann nun abgeleitet werden

$$(4) \quad P_i = \text{Prob}(y_i = 1) = \text{Prob}[u_i > -(\beta_0 + \sum_{k=1}^K \beta_k \cdot x_{ik})] = 1 - F[-(\beta_0 + \sum_{k=1}^K \beta_k \cdot x_{ik})],$$

wobei F : Kumulierte Verteilung des Störterms u .

Sofern die Verteilung von u symmetrisch ist, kann *Gleichung (4)* vereinfacht werden zu

$$(5) \quad P_i = F\left(\beta_0 + \sum_{k=1}^K \beta_k \cdot x_{ik}\right).$$

Da die betrachteten Werte y_{ik} Realisierungen eines binomialen Prozesses mit der Wahrscheinlichkeit P_i (siehe *Gleichung (4)*) darstellen, die in Abhängigkeit von x_{ik} von Beobachtung zu Beobachtung variieren, können wir folgende Likelihood-Funktion aufstellen:

$$(6) \quad L = \prod_{y_i=1} P_i \cdot \prod_{y_i=0} (1 - P_i).$$

Die Maximierung dieser Likelihood-Funktion wird in Statistikpaketen wie SAS oder SPSS mit Hilfe des Newton-Raphson-Algorithmus erreicht. Dabei führen zwar beliebige Ausgangswerte zum Maximum der konkaven Funktion, üblicherweise werden aber Schätzer aus dem „linear probability model“ als Startwerte eingesetzt. Verbal beschrieben besteht das Prinzip des Maximum-Likelihood (ML)-Schätzverfahrens darin, in einem schrittweisen Iterationsverfahren die Schätzungen der Parameter so zu wählen, daß der Beobachtung für diesen Schätzwert eine maximale Wahrscheinlichkeit zukommt (vgl. *Fahrmeir und Hamerle, 1984, S. 59*).¹ Je nachdem, welche Annahmen über den Störterm in *Gleichung (3)* getroffen werden, ergeben sich unterschiedliche funktionale Formen von F in *Gleichung (4)*. Sofern die Störterme u_i einer Normalverteilung folgen, handelt es sich um eine Probit- (oder Normit-) Analyse, die hier nicht näher betrachtet wird.² Bei Unterstellung einer logistischen Verteilung von u_i ergibt sich das Logit-Modell

$$(7) \quad F(Z_i) = \frac{e^{Z_i}}{1 + e^{Z_i}} = \frac{1}{1 + e^{-Z_i}}, \text{ wobei}$$

Z_i : Linearer Prädiktor des Logistischen Modells für das i -te Objekt ($i \in I$),
d.h. $Z_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \dots + \beta_k \cdot x_{ik} + \dots + \beta_K \cdot x_{iK}$ ($k \in K$).

Daraus folgt

$$(8) \quad \log \frac{F(Z_i)}{1 - F(Z_i)} = Z_i \text{ bzw.}$$

$$(9) \quad \log \frac{P_i}{1 - P_i} = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \dots + \beta_k \cdot x_{ik} + \dots + \beta_K \cdot x_{iK}.$$

In der betrachteten Studie ist dabei die Wahrscheinlichkeit, daß Reisenden-Verkaufsaußendienste eingesetzt werden, die abhängige Variable P_i . Der logistische Funktionsverlauf

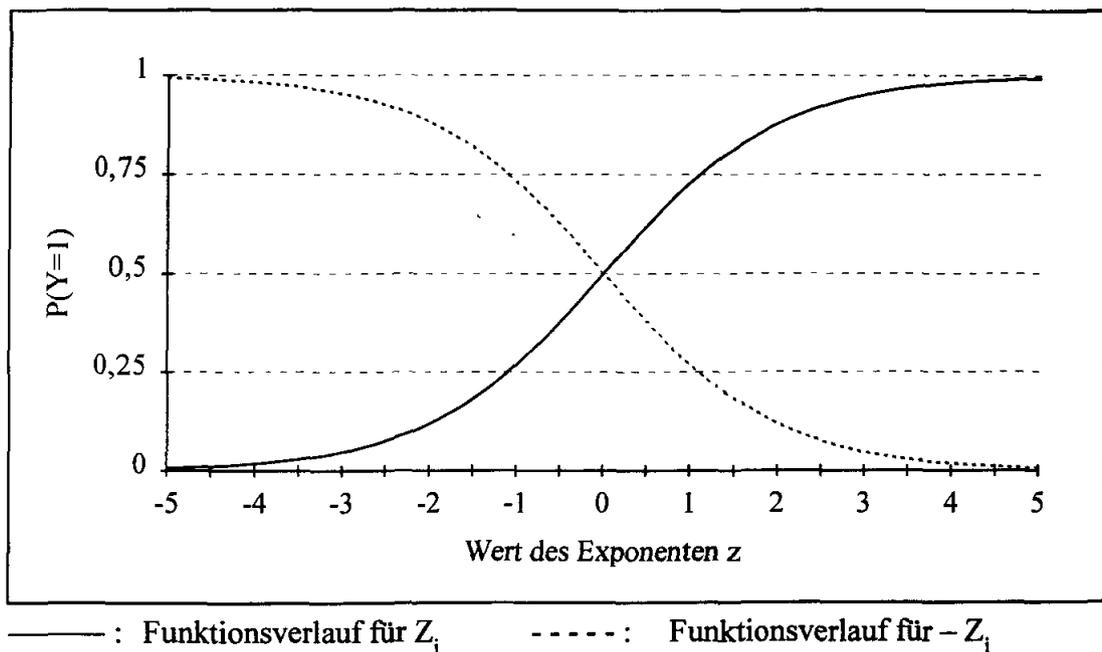
¹ Für den Fall *stetiger* Beobachtungen wird eine maximale Wahrscheinlichkeitsdichte angestrebt. Bei den hier betrachteten Fragestellungen liegen jedoch diskrete Beobachtungen vor.

² Aufgrund der unwesentlichen Unterschiede der beiden Verteilungsannahmen (der Logit ist an den Verteilungsextrema etwas höher), der häufigeren Anwendung der Logit-Analyse und ihrer guten Prognoseleistung wird sie hier dem Probit-Verfahren vorgezogen. Vgl. *Aldrich und Nelson, 1984, S. 31 ff.*; *Gessner et al., 1988, S. 63*; *Urban, 1993, S. 105*.

entspricht einem nichtlinearen, S-förmigen Verlauf der abhängigen Variablen. Die nichtlinearen Eigenschaften von Gleichung (7) werden in Abb. 1 veranschaulicht, indem der Exponent Z_i systematisch zwischen -5 und $+5$ variiert wird. Der Term $\ln(P_i/(1-P_i))$, also der natürliche Logarithmus des Quotienten aus Wahrscheinlichkeit und Gegenwahrscheinlichkeit, wird als "Logit" ("Logistic Probability Unit") bezeichnet. Dieser Logit hat zur Bezeichnung "Logistische Regression" geführt.

Gegenüber einer linearen Regression weist dieses Verfahren den Vorteil auf, daß selbst für Werte des linearen Prädiktors Z_i von $\pm\infty$ nie Werte von P_i außerhalb des Intervalls $[0, 1]$ resultieren. Eine weitere wesentliche Eigenschaft des Funktionstyps ist, daß ein Wert von 50%, also Indifferenz, vorhergesagt wird, sofern Z_i einen Wert von 0 annimmt. Außerdem ist die logistische Funktion symmetrisch bei einem Wendepunkt von $P_i=0,5$. Es läßt sich zeigen, daß die Konstante β_0 des linearen Prädiktors Z_i die Funktion horizontal verschiebt, während höhere Koeffizienten β_k zu einem steileren Verlauf der logistischen Funktion führen. Negative Vorzeichen der Koeffizienten β_k ändern den Ursprung der Kurve, was dem gestrichelten Funktionsverlauf in Abb. 1 entspricht (vgl. Green et al., 1977, S. 54; Menard, 1995, S. 13).

Abb. 1: Verläufe der logistischen Funktionskurve im Intervall von -5 bis $+5$



Voraussetzung zur Anwendbarkeit des ML-Prinzips und damit der Logistischen Regression ist aufgrund der asymptotischen Eigenschaften eine möglichst große Stichprobe, wobei als Faustregel eine Beobachtungszahl von mindestens 50 genannt wird (Urban, 1993, S. 1 und S. 13). Einen strengeren Maßstab setzen Aldrich und Nelson (1984), die 100 Freiheitsgrade als Minimum ansehen. In unserem Beispiel werden 149 Beobachtungen zur Schätzung von 23 Parametern verwendet, so daß der strengere Maßstab erfüllt ist (vgl. Krafft, 1995b, S. 12).

3. Zur Güte des Schätzmodells

Bevor mit der Interpretation einzelner Koeffizienten begonnen werden kann, ist zuerst zu klären, ob ein überprüftes Modell insgesamt verwendbar ist. Bei dieser Beurteilung der Güte

der Schätzmodelle kann aber nicht auf die herkömmlichen Maße und Tests der Linearen Regressionsanalyse zurückgegriffen werden, da die Koeffizienten der Logistischen Regression mit Hilfe der ML-Schätzmethode bestimmt werden. Eine perfekte Anpassung der Parameter ist dabei mit einer Likelihood von 1 verbunden. Es ist üblich, als Gütemaß der Anpassung des geschätzten Modells mit $-2 \cdot \log(\text{Likelihood})$ zu rechnen, das im folgenden als $-2LL$ oder *Devianz* bezeichnet wird (vgl. *Aldrich und Nelson*, 1984, S. 59; *Hosmer und Lemeshow*, 1989, S. 138 f.; *Maddala*, 1983, S. 40). $-2LL$ wird eingesetzt, da die mit -2 multiplizierte logarithmierte Likelihood asymptotisch χ^2 -verteilt ist mit $(I-p)$ Freiheitsgraden, wobei I die Zahl der Beobachtungen und p die Anzahl der Parameter angibt. Gute Modelle, die eine „hohe“ Likelihood nahe 1 aufweisen, resultieren in einer Devianz, die nahe 0 ist, während schlechte Anpassungen mit hohen Devianz-Werten verbunden sind. Mit dem Wert der Devianz kann die Hypothese H_0 getestet werden, daß das Modell eine perfekte Anpassung aufweist. Dabei bedeuten geringe Devianz-Werte bzw. hohe Signifikanzen, daß H_0 nicht abgelehnt werden kann und das Modell eine gute Anpassung aufweist. Insbesondere sind Devianzen nahe 0 bzw. Signifikanzen nahe 1 ein Indikator für sehr gute Modellanpassungen (*Menard*, 1995, S. 19 ff.; *Urban*, 1993, S. 64 f.).

Bei der Absatzformwahl-Studie sind 126 Freiheitsgrade ($N=149$, $p=23$) gegeben, und das Modell weist eine Devianz von 61,909 auf. Dieser relativ niedrige Wert³ und das Signifikanzniveau der Devianz von 1,000 zeigen, daß die Modellanpassung als sehr gut einzustufen ist.

Ein weiteres Gütemaß ist die *Goodness-of-Fit-Statistik* \hat{C} , die folgendermaßen definiert ist (*Hosmer und Lemeshow*, 1989, S. 140 ff.):

$$(10) \quad \hat{C} = \sum_{i \in I} \frac{(Y_i - P_i)^2}{P_i \cdot (1 - P_i)}, \text{ wobei}$$

- Y_i : Beobachteter Wert der abhängigen Variablen für die i -te Beobachtung ($i \in I$),
- P_i : Geschätzter Wert der abhängigen Variablen für die i -te Beobachtung ($i \in I$),
- I : Indexmenge der Beobachtungen.

Wie schon die Devianz ist \hat{C} asymptotisch χ^2 -verteilt. Ein niedriger χ^2 -Wert der \hat{C} -Statistik und ein hohes Signifikanzniveau implizieren wiederum eine gute Anpassung des Modells. Bei Anwendung der Logistischen Regression auf die Frage „Handelsvertreter oder Reisende“ ergibt sich in der Analyse von *Krafft* (1995b, S. 25) ein \hat{C} von 115,896 bei einem Signifikanzniveau von 0,7297, so daß auch nach diesem Maß eine gute Modellanpassung zu konstatieren ist.

Neben der Goodness-of-Fit-Statistik und der Devianz gibt es mit dem *Likelihood-Ratio-Test* und McFaddens R^2 zwei weitere Gütemaße, die die Verbesserung der Anpassung durch das Modell gegenüber einem Null-Modell überprüfen, das üblicherweise nur aus einem Schätzwert für die Konstante besteht. Die absolute Differenz der Devianz des Null-Modells und des endgültigen Modells stellt einen asymptotisch χ^2 -verteilten Wert dar, der gegen die

³ Die Devianz ist im positiven Bereich unbeschränkt. Ob eine Devianz als relativ "hoch" oder "niedrig" einzuschätzen ist, hängt daher jeweils von der betrachteten Stichprobe und Analyse ab.

Nullhypothese getestet werden kann, daß die Koeffizienten des endgültigen Modells nicht signifikant verschieden von 0 sind. Damit liegt ein Likelihood-Ratio-Test vor, der dem F-Test der Regressionsanalyse vergleichbar ist. Diese Teststatistik wird auch als "Model Chi-Square" bezeichnet. Hohe χ^2 -Werte und niedrige Signifikanzniveaus deuten darauf hin, daß die Koeffizienten des endgültigen Modells signifikant verschieden von 0 sind, wobei als kritisches Signifikanzniveau üblicherweise das 5%-Intervall angesetzt wird (*Hosmer und Lemeshow*, 1989, S. 31 ff. und *Menard*, 1995, S. 20 f.). In der Absatzformwahl-Studie weist das Logistische Modell einen Likelihood-Ratio-Wert von 87,764 auf, der hoch signifikant ist ($< 0,0001$). Somit wird gegenüber dem Null-Modell durch Einschluß der 23 Parameter eine signifikant bessere Anpassung des Modells an die beobachteten Wahrscheinlichkeiten, Reisende als Absatzform zu wählen, erreicht.

Die Kenntnis der Devianz des endgültigen sowie des Null-Modells kann zusätzlich genutzt werden, um ein dem Bestimmtheitsmaß der Regressionsanalyse ähnliches R^2 nach McFadden zu berechnen. *McFaddens R^2* ergibt sich aus folgender Gleichung (*McFadden*, 1974):

$$(11) \quad \text{McFaddens } R^2 = 1 - \frac{LL_1}{LL_0}, \text{ wobei}$$

$LL_1 =$ Natürlicher Logarithmus der Likelihood des endgültigen Modells,

$LL_0 =$ Natürlicher Logarithmus der Likelihood des Null-Modells.

Aufgrund der Eigenschaften der Logistischen Regression fällt *McFaddens R^2* generell niedriger aus als das R^2 in Linearen Regressionsanalysen, da die beobachteten Werte nur den Wert 0 oder 1 aufweisen, während die geschätzte abhängige Variable eine Wahrscheinlichkeit zwischen 0 und 1 annimmt. Modelle werden daher bereits als gut eingeschätzt, wenn *McFaddens R^2* im Intervall 0,2 bis 0,4 liegt (*Urban*, 1993, S. 62 f.). In unserem Beispiel weist das endgültige Modell für LL_1 bzw. LL_0 Werte von -30,955 bzw. -74,837 auf, so daß ein vergleichsweise hoher Wert von 0,5864 resultiert. Somit unterstreicht auch *McFaddens R^2* die Güte dieses Modells.

Zur Einschätzung der Vorhersagegenauigkeit logistischer Modelle wird ergänzend die *Klassifikationsmatrix* (auch "Confusion-Matrix" genannt) betrachtet, in der aus den Zeilen die beobachtete Gruppenzugehörigkeit hervorgeht, während die Klassifizierung aufgrund des Logistischen Modells in den Spalten vermerkt wird (*Menard*, 1995, S. 25 ff.). Die richtig klassifizierten Elemente stehen dann in der Hauptdiagonalen, die falsch klassifizierten Beobachtungen außerhalb der Hauptdiagonalen. Der Anteil der mit Hilfe der Logistischen Regression korrekt klassifizierten Elemente, der als "Trefferquote" bezeichnet wird, sollte dabei höher sein als die Trefferquote, die sich aus einer zufälligen Zuordnung ergibt. *Tab. 1* gibt die Klassifikationsmatrix aus der Studie von *Krafft* (1995b, S. 24) wieder. Einschränkend ist hier anzumerken, daß die Trefferquote des Modells überhöht ist, da die Parameterschätzung der Logistischen Regression und die Berechnung der Trefferquote im angeführten Beispiel auf Basis derselben Stichprobe erfolgen (siehe *Backhaus et al.*, 1994, S. 116; *Morrison*, 1969, S. 158). Bei Anwendung der geschätzten Koeffizienten auf andere Stichproben ist nämlich

mit geringeren Trefferquoten zu rechnen. Für relativ große Stichproben ist dieser Stichprobeneffekt in der Regel als gering einzuschätzen. Bei etwa gleich großen Gruppen der abhängigen Variablen wird als Referenz zur Beurteilung der Klassifizierungsgüte das "*maximum chance criterion*" (MCC) empfohlen, das dem Anteil der größeren Gruppe an der gesamten Stichprobe entspricht (vgl. Morrison, 1969, S. 158). Da im angeführten Beispiel mit 119 Reisenden bzw. 30 Handelsvertreter-Außendienstern (79,9% bzw. 20,1% der Stichprobe) sehr unterschiedlich große Gruppen vorliegen und zudem die richtige Klassifizierung beider Gruppen angestrebt wird, erweist sich das MCC als weniger geeignet. Nach diesem Kriterium würde nämlich eine Lösung als gut eingeschätzt, bei der alle Beobachtungen der Reisenden-Gruppe und wenige der Handelsvertreter-Gruppe richtig klassifiziert werden. Es ist aber eine möglichst gleichmäßige richtige Klassifizierung über beide Gruppen anzustreben.

Tab. 1: *Klassifikationsmatrix des Logistischen Regressionsmodells zur Absatzformwahl*

Beobachtete Gruppenzugehörigkeit	Prognostizierte Gruppenzugehörigkeit		Korrekte Klassifizierung
	Handelsvertreter-Verkaufsaußendienste	Reisenden-Verkaufsaußendienste	
Handelsvertreter-Verkaufsaußendienste	22	8	73,33%
Reisenden-Verkaufsaußendienste	5	114	95,80%
Insgesamt			91,28%

In der einschlägigen Literatur wird deshalb empfohlen, das "*proportional chance criterion*" (PCC) anzuwenden, insbesondere wenn -wie in unserem Beispiel- zwei sehr unterschiedlich große Gruppen analysiert werden. Das PCC entspricht einer zufälligen Trefferquote von $\alpha^2 + (1-\alpha)^2$, wobei α der Anteil einer Gruppe an der Gesamtzahl der Beobachtungen ist (vgl. Morrison, 1969, S. 158). Für unser Beispiel ist $\alpha = 0,799$ und somit das PCC 67,84%. Mit 91,28% richtig klassifizierten Beobachtungen ($[22+114]/149$) ist die Zuordnung gegenüber dem PCC als gut zu beurteilen, nicht zuletzt da auch die Beobachtungen der kleinen Handelsvertreter-Gruppe mit einer Trefferquote von 73,33% vergleichsweise gut klassifiziert werden.

Als letztes Gütekriterium des Modells der Logistischen Regression wird das *Histogramm der geschätzten Wahrscheinlichkeiten* diskutiert. Das Histogramm, das dem Klassifizierungsdiagramm der Diskriminanzanalyse entspricht, ist insbesondere daraufhin zu inspizieren, ob Beobachtungen mit Hilfe des Modells der Logistischen Regression eindeutig der richtigen Kategorie zugeordnet wurden. Zahlreiche falsch klassifizierte Fälle nahe der indifferenten Wahrscheinlichkeit von 0,5 deuten darauf hin, daß die Klassifikationsgüte besser zu beurteilen ist, als es durch die Dichotomie "richtig/falsch" zum Ausdruck kommt (vgl. Menard, 1995, S. 32 ff.; Norušis, 1993, S. 8 f.). Tab. 2 faßt die hier erörterten Gütekriterien der Logistischen Regression im Überblick zusammen.

Tab. 2: Akzeptable Wertebereiche der Gütemaße der Logistischen Regression

Kriterium	Akzeptabler Wertebereich
Devianz (-2LL)	-2LL nahe 0; Signifikanzniveau nahe 100%
Goodness-of-Fit (\hat{C})	\hat{C} nahe 0; Signifikanzniveau nahe 100%
Likelihood-Ratio-Test	Möglichst hoher χ^2 -Wert, Signifikanzniveau < 5%
McFaddens R^2	Akzeptabel, sofern größer 0,2
Klassifikation	Klassifikation mindestens besser als das "proportional chance criterion" $\{\alpha^2 + (1-\alpha)^2; \alpha: \text{relative Größe einer Gruppe}\}$

4. Prüfung und Interpretation einzelner Koeffizienten

Wenn die Beurteilung der Logistischen Regression wie in unserem Beispiel nicht zu einer generellen Ablehnung des geschätzten Modells geführt hat, kann mit der Prüfung der Koeffizienten hinsichtlich ihrer Signifikanz, Richtung und relativen Bedeutung begonnen werden. Dabei ist zu beachten, daß Parameterschätzungen einer Logistischen Regression wesentlich schwerer zu interpretieren sind als in der Linearen Regression. Dort entspricht der Koeffizient der absoluten Änderung der abhängigen Variablen bei Erhöhung der unabhängigen Variablen um eine Einheit. Bei der Logistischen Regression repräsentieren die Koeffizienten dagegen die Änderung des Logit der abhängigen Variablen bei einer Änderung der unabhängigen Variablen um eine Einheit (vgl. *Aldrich und Nelson*, 1984, S. 41 f.; *DeMaris*, 1992, S. 44; *Hair et al.*, 1992, S. 61; *Maddala*, 1983, S. 23 f.). Der Logit als natürlicher Logarithmus der "Gewinnchance", also des Verhältnisses der Wahrscheinlichkeit, daß die abhängige Variable gleich 1 ist, zu dessen Gegenwahrscheinlichkeit, stellt aber keine leicht verständliche abhängige Größe dar.

Direkt interpretierbar sind dagegen die im Rahmen der Logistischen Regression geschätzten Koeffizienten zum einen hinsichtlich der *Signifikanz* ihres Einflusses auf die abhängige Variable und zum anderen bezüglich der Richtung dieses Einflusses. Das Konfidenzintervall der einzelnen Koeffizienten kann aufgrund der χ^2 -verteilten Wald-Statistik bestimmt werden, die sich aus dem Quadrat des Quotienten von Koeffizient und Standardfehler einer Variablen errechnet. Diese Formel gilt nur für metrische Variablen mit einem Freiheitsgrad. Für kategoriale Variablen ist zusätzlich die Zahl der Variablen-Freiheitsgrade zu berücksichtigen (vgl. *Hosmer und Lemeshow*, 1989, S. 17).

Sofern Variablen signifikante Koeffizienten aufweisen, kann auch deren *Einfluß* interpretiert werden. Wie dabei schon aus *Abb. 1* ersichtlich wurde, bedeuten negative Vorzeichen, daß die Wahrscheinlichkeit P_i sinkt, während positive Vorzeichen steigende Wahrscheinlichkeiten implizieren.

4.1. Partielle Ableitungen als Interpretationshilfe

Außer der Signifikanz und der Richtung des Einflusses einzelner Variablen ist keine direkte Interpretation der Koeffizienten aus Logistischen Regressionsmodellen möglich. Zur

Interpretation der Koeffizienten aus Logistischen Regressionsmodellen werden hier drei alternative Formen diskutiert. Als eine Methode wird die *partielle Ableitung* der Logit-Funktion nach einer unabhängigen Variablen x_k vorgeschlagen. Die partielle Ableitung hat die Form (siehe *Hruschka*, 1991, S. 345; *LeClere*, 1992, S. 771 f.; *Petersen*, 1985, S. 130 f.):

$$(12a) \quad \frac{\partial P_i}{\partial x_k} = \frac{e^{-(\beta_0 + \sum_{k \in K} \beta_k \cdot x_k)}}{\left(1 + e^{-(\beta_0 + \sum_{k \in K} \beta_k \cdot x_k)}\right)^2} \cdot \beta_k \quad (k \in K), (i \in I).$$

Die Gleichung kann vereinfacht werden zu

$$(12b) \quad \frac{\partial P_i}{\partial x_k} = \beta_k \cdot P_i \cdot (1 - P_i) \quad (k \in K), (i \in I).$$

Üblicherweise wird für alle unabhängigen Variablen deren Mittelwert in *Gleichung (12a)* eingesetzt und die partielle Ableitung als marginale Wahrscheinlichkeit P_i hinsichtlich der Änderung der jeweils betrachteten Variablen um eine Einheit angesehen. Diese Interpretation ist ungenau, da die Höhe der partiellen Ableitung eine Funktion der jeweiligen Wahrscheinlichkeit P_i und des Koeffizienten β_k ist, der aber von der Skalierung der Variablen abhängt. Obwohl die partielle Ableitung die wahre marginale Wahrscheinlichkeit gegebenenfalls approximiert, sollte sie dieser aufgrund der Skalenabhängigkeit nicht gleichgesetzt werden (vgl. *LeClere*, 1992, S. 771 f.; *Petersen*, 1985, S. 130 f.). Zur Veranschaulichung werden die in der Studie von *Krafft* (1995b, S. 25) berichteten signifikanten Koeffizienten herangezogen. In *Tab. 3* werden neben diesen Koeffizienten auch die Mittelwerte der Handelsvertreter- bzw. Reisenden-Beobachtungen wiedergegeben.

Die partiellen Ableitungen der Parameter werden ermittelt, indem die Mittelwerte der Handelsvertreter-Beobachtungen in die *Gleichungen (7)* und *(12a)* eingesetzt werden. Diese Mittelwerte wurden gewählt, da die Reagibilität der Wahrscheinlichkeit P , Reisende zu wählen, in diesem Punkt groß ist. Bei Verwendung der Reisenden-Mittelwerte sind demgegenüber nur geringe absolute Änderungen von P zu beobachten, da für die Reisenden-Mittelwerte eine extrem hohe prognostizierte Wahrscheinlichkeit von 99,93% resultiert, daß Reisende gewählt werden. Wie aus *Abb. 1* bekannt ist, sind in diesem Extrembereich der Logistischen Funktion große Änderungen der Prädiktoren nötig, um die geschätzte Wahrscheinlichkeit zu reduzieren. Sofern die Mittelwerte der Handelsvertreter-Beobachtungen in die Logistische Regressionsfunktion eingesetzt werden, resultiert nach Maßgabe des Modells eine Wahrscheinlichkeit von $P=2,199\%$, daß Reisenden-Außendienste gewählt werden.⁴

Der für die Ermittlung der partiellen Ableitung einer Variablen benötigte Term $P \cdot (1-P)$ beträgt folglich in diesem Punkt der logistischen Funktion 0,0215. Die partielle Ableitung

⁴ Der exakte Wert von P_i ist 0,021987362 und ergibt sich aus dem Einsetzen der Mittelwerte der Handelsvertreter-Beobachtungen in *Gleichung 7*. Die Mittelwerte und die Koeffizienten der Variablen wurden *Krafft* (1995b, S. 25) entnommen. Der Wert der Konstanten beträgt -8,378.

einer Variablen in diesem Punkt wird nun berechnet als Produkt aus dem Koeffizienten und dem Term $P^*(1-P)$. Die jeweiligen partiellen Ableitungen der einzelnen Variablen sind in *Tab. 3* berichtet. Die Einflußgröße "Risikoaversion" weist für $P=2,199\%$ mit Abstand die absolut höchste partielle Ableitung auf, gefolgt von der Variablen "Reisezeitenanteil". Da die signifikanten Variablen sehr unterschiedliche Skalierungen und empirische Bandbreiten aufweisen, ist eine Interpretation der partiellen Ableitungen als Maß der relativen Bedeutung der Variablen problematisch.

4.2. Interpretation mit Hilfe von Elastizitäten

Das Problem der Abhängigkeit partieller Ableitungen von der Skalierung der unabhängigen Variablen kann nun zum einen durch Standardisierung der unabhängigen Variablen gelöst werden. Zum anderen können die direkten *Elastizitäten* der Variablen zur Interpretation der Koeffizienten der Logistischen Regression herangezogen werden. Da Elastizitäten dimensionslos sind und im logistischen Modell die prozentuale Änderung der Wahrscheinlichkeit P_i bei einer Änderung der jeweiligen unabhängigen Variablen um 1% bezeichnen, sind sie besser interpretierbar als die nicht skaleninvarianten partiellen Ableitungen der Koeffizienten. Die Elastizität der Wahrscheinlichkeit P_i bezüglich marginaler Änderungen von x_k ergibt sich dabei aus folgender Gleichung (siehe *Hruschka*, 1991, S. 345; *LeClere*, 1992, S. 772; *Petersen*, 1985, S. 131):

$$(13a) \quad \varepsilon_{k,i} = \frac{x_k}{P_i} \cdot \frac{\partial P_i}{\partial x_k} = \frac{x_k}{P_i} \cdot \frac{e^{-\left(\beta_0 + \sum_{k \in K} \beta_k \cdot x_k\right)}}{\left(1 + e^{-\left(\beta_0 + \sum_{k \in K} \beta_k \cdot x_k\right)}\right)^2} \cdot \beta_k \quad (k \in K), (i \in I).$$

Die Gleichung kann vereinfacht werden zu

$$(13b) \quad \varepsilon_{k,i} = \frac{x_k}{P_i} \cdot \frac{\partial P_i}{\partial x_k} = x_k \cdot (1 - P_i) \cdot \beta_k \quad (k \in K), (i \in I).$$

Die Elastizität resultiert somit aus der Multiplikation der partiellen Ableitung der Wahrscheinlichkeit P_i mit der Ausprägung der unabhängigen Variablen x_k , wobei häufig der Mittelwert der Variablen \bar{x}_k in der Stichprobe verwendet wird (vgl. *LeClere*, 1992, S. 773 f.), und Division durch P_i . Wie bei der partiellen Ableitung gilt zwar einschränkend, daß die Elastizitäten je nach Ausgangssituation P_i und Variablenausprägung x_k unterschiedlich hoch ausfallen. Die Dimensionslosigkeit von Elastizitäten ermöglicht aber direkte Vergleiche der relativen Einflüsse verschiedener unabhängiger Variablen auf die Wahrscheinlichkeit P_i .

Tab. 3: Partielle Ableitungen und Elastizitäten sowie Sensitivitätsanalyse der signifikanten Variablen des Hypothesentests zur Absatzformwahl

Unabhängige Variable	Logit-Koeffizient	Mittelwerte		Partielle Ableitung ^{a)}	Elastizität ^{a)}	Geschätzte Wahrscheinlichkeit, Reisende zu wählen, wenn der ... in das Logistische Modell eingesetzt wird		
		Handelsvertreter	Reisende			Mittelwert der HV-Stichprobe	Mittlere Wert aus HV- und R-Stichprobe ^{b)}	Mittelwert der R-Stichprobe ^{b)}
Kundenzahl je Mitarbeiter	-0,0077	878,67	171,63	-0,000166	-6,61700	2,199 %	25,484 %	83,878 %
Reisezeitenanteil	+0,1660	19,27	32,11	+0,003570	+3,12849	2,199 %	6,126 %	15,928 %
Echte Non-Selling-Aktivitäten	-0,1046	12,37	5,98	-0,002249	-1,26545	2,199 %	3,045 %	4,202 %
Produktspezifische Kenntnisse	+0,0001	11.108,67	25.023,88	+0,000002	+1,08644	2,199 %	4,314 %	8,290 %
Risikoaversion	-0,4097	3,00	2,83	-0,008810	-1,20208	2,199 %	2,275 %	2,354 %
Bildungsniveau	+0,0146	108,20	165,14	+0,000314	+1,54499	2,199 %	3,295 %	4,909 %

HV: Handelsvertreter

R: Reisende

a): errechnet für $P_i = 0,02199$

b): Es wurde jeweils nur für die betrachtete Variable der Mittelwert der HV-Stichprobe ersetzt (ceteris-paribus-Betrachtung)

Zur Ermittlung der jeweiligen Elastizität in unserem Beispiel wird wiederum von den Mittelwerten der Handelsvertreter-Beobachtungen ausgegangen. Elastizitäten bieten nun den Vorteil, daß sie von der absoluten Höhe direkt miteinander verglichen werden können. Ebenso wie die partiellen Ableitungen besitzen sie jedoch nur für den betrachteten Punkt der logistischen Wahrscheinlichkeitsfunktion Gültigkeit und variieren mit dem Verlauf der logistischen Funktion.⁵ Alle Elastizitäten, die in *Tab. 3* wiedergegeben sind, weisen absolute Werte größer 1 auf, d.h. der Einfluß der exogenen Variablen auf die Wahrscheinlichkeit, Reisende einzusetzen, ist vergleichsweise hoch. Zwischen den Elastizitäten bestehen substantielle Unterschiede: Der betragsmäßig stärkste Einfluß geht von der "Kundenzahl je Mitarbeiter" aus, gefolgt von den Variablen "Reisezeitenanteil" und "Bildungsniveau". Die von den partiellen Ableitungen abweichende Reihenfolge der Variablen ist dabei auf die Skaleninvarianz der Elastizitäten zurückzuführen.

Für den Fall, daß $P=2,199\%$ ist, kann festgestellt werden, daß mit einer einprozentigen Erhöhung der Kundenzahl je Mitarbeiter (also von 878,7 auf 887,5 Kunden) die Wahrscheinlichkeit, Reisenden-Außendienste einzusetzen, um 6,62% auf etwa 2,05% sinkt.⁶ Sofern sich der Reisezeitenanteil von 19,27% um 1% auf etwa 19,46% erhöht, resultiert nach Maßgabe des Logistischen Modells eine Erhöhung von P_i um 3,13% auf 2,27%. Für die weiteren fünf signifikanten Variablen des Beispiels kann die Wirkung einer einprozentigen Änderung der Einflußfaktoren analog berechnet werden.

4.3. Sensitivitätsanalysen

Die wohl eingängigste Form der Interpretation Logistischer Regressionsmodelle ist die *Sensitivitätsanalyse*, bei der die Reaktion der Wahrscheinlichkeit P_i auf unterschiedliche Ausprägungen der unabhängigen Variablen betrachtet wird. Zumeist wird dabei von den Stichproben-Mittelwerten der unabhängigen Variablen ausgegangen und P_i für diese Konstellation errechnet. Der Wert einer unabhängigen Variablen wird nun systematisch variiert, während die anderen Variablen konstant gehalten werden. Die Differenz der ursprünglich geschätzten zur resultierenden neuen Wahrscheinlichkeit P_i spiegelt den Einfluß von Änderungen signifikanter unabhängiger Variablen auf P_i wider. Der Vorteil von Sensitivitätsanalysen gegenüber Elastizitäten liegt in der Veranschaulichung des absoluten Effekts von unterschiedlichen Ausprägungen der unabhängigen Variablen auf die Wahrscheinlichkeit P_i (vgl. *LeClere*, 1992, S. 772 f.; *Urban*, 1993, S. 46 ff.).

⁵ Alternativ wurden die Elastizitäten ermittelt, die sich ergeben, wenn die Mittelwerte der Reisenden-Beobachtungen eingesetzt werden. Die Rangfolge der Variablen hinsichtlich der Intensität der Elastizitäten bleibt generell ähnlich, einzig die Variable "Kundenzahl je Mitarbeiter" zeigt eine deutliche Veränderung: Während dieser Faktor unter Verwendung der Handelsvertreter-Mittelwerte die absolut größte Elastizität aufweist, zeigt er für die Reisenden-Mittelwerte nur den vierthöchsten Elastizitätswert.

⁶ Diese Verringerung ist keine Änderung in Prozentpunkten, sondern in Prozent vom Ausgangswert, d.h. 6,62% von 0,02199. Die Wahrscheinlichkeit, Reisenden-Verkaufsaußendienste einzusetzen, sinkt somit in unserem Beispiel von 2,199% auf etwa 2,05%.

Bezogen auf das hier analysierte Absatzformwahl-Problem stellen sich Verkaufsaußendienstleiter die Frage, wie unterschiedliche Ausprägungen der unabhängigen Variablen die Wahrscheinlichkeit beeinflussen, daß Reisenden-Außendienste als Absatzform gewählt werden. Für die Sensitivitätsanalyse werden wiederum die Variablen-Mittelwerte der Handelsvertreter-Beobachtungen als Ausgangssituation gewählt. Es werden nun für jede signifikante unabhängige Variable die geschätzten Wahrscheinlichkeiten ermittelt, die sich ergeben, wenn der Wert der betrachteten Variablen durch den mittleren Wert⁷ aus Handelsvertreter- und Reisenden-Beobachtungen bzw. den Mittelwert der Reisenden-Beobachtungen ersetzt wird, wobei gleichzeitig die Werte der anderen Variablen konstant gehalten werden.

Die in *Tab. 3* berichteten geschätzten Wahrscheinlichkeiten bestätigen im wesentlichen die Interpretation nach Maßgabe der Elastizitäten: Der absolut größte Einfluß auf die Wahrscheinlichkeit, Reisende als Absatzform zu wählen, geht von der "Kundenzahl je Mitarbeiter" aus. Sofern die Zahl der vom einzelnen Mitarbeiter betreuten Kunden von 878,67 (Mittelwert der Handelsvertreter-Außendienste) auf 171,63 (Mittelwert der Reisenden-Außendienste) sinkt, steigt die geschätzte Wahrscheinlichkeit von 2,199% auf 83,878%. Auch der Reisezeitenanteil übt eine erhebliche, wenngleich deutlich geringere Wirkung auf die Änderung der Wahrscheinlichkeit aus, Reisenden-Außendienste einzusetzen. Sofern der Reisezeitenanteil 32,11% (Mittelwert Reisenden-Außendienste) statt 19,27% beträgt, steigt die Wahrscheinlichkeit, eine Reisenden-Vertriebsorganisation als Absatzform zu wählen, von 2,199% auf 15,928%. Während auch vom Umfang produktspezifischer Kenntnisse ein erheblicher Einfluß auf die geschätzte Änderung der abhängigen Variablen ausgeht, zeigen die weiteren drei Faktoren eine verhältnismäßig geringe Wirkung, was hinsichtlich der Variablen "Risikoaversion" auch auf die relativ geringe Variation dieser unabhängigen Variablen zwischen den beiden Gruppen zurückzuführen ist. Unterschiede in der Rangfolge der Variablen zwischen der Elastizitäts- und Sensitivitätsanalyse sind somit auf Ausprägungsunterschiede der betrachteten Variablen zurückzuführen, sofern Mittelwerte der Teilstichproben Gegenstand der Sensitivitätsanalyse sind. Falls Ausprägungen nur in relativ kleinen Bandbreiten variiert werden, ergeben sich keine Unterschiede zu den Ergebnissen der Elastizitätsanalyse. Der wesentliche Vorteil der Sensitivitätsanalyse ist somit die Veranschaulichung der absoluten Effekte von Variablen auf die betrachtete Wahrscheinlichkeit P_i .

⁷

Der mittlere Wert wird als ungewichteter Durchschnitt errechnet, d.h. $\frac{1}{2} * (\text{Mittelwert Handelsvertreter-Beobachtungen} + \text{Mittelwert Reisenden-Beobachtungen})$.

5. Prognose mit Logistischen Regressionsmodellen

Die Logistische Regression kann auch zu Prognosezwecken eingesetzt werden: Bezogen auf unser Beispiel könnten Unternehmen, die eine Vertriebsorganisation aufbauen bzw. die derzeit gewählte Absatzform wechseln wollen, die Befunde des hier diskutierten Modells als konkrete Entscheidungshilfe nutzen. In einem ersten Schritt wäre anhand der betrachteten Einflußgrößen zu untersuchen, wie sich die Situation im Persönlichen Verkauf voraussichtlich gestalten wird bzw. welche Konstellation tatsächlich gegeben ist. Diese Ausprägungen der Einflußfaktoren sind dann in das Modell der Logistischen Regression einzusetzen, um nach Maßgabe des für die untersuchte Stichprobe ermittelten üblichen Verhaltens zu schätzen, wie wahrscheinlich die Wahl von Reisenden-Außendiensten als Absatzform ist. Zur Veranschaulichung wird ein Unternehmen der Stichprobe herangezogen, für das Daten zu allen Variablen des Logistischen Modells vorliegen, das aber nicht zur Schätzung des Logistischen Regressionsmodells herangezogen wurde.

Tab. 4: Daten des Beispielunternehmens zur Absatzformwahl

Unabhängige Variable	Logit-Koeffizient	Ausprägung laut Fragebogen
Unsicherheit	0,014	19 a)
Volatilität	0,0204	10 a)
Kundenzahl je Mitarbeiter	-0,01	25
Risiko des Marktaustritts	-0,6542	6
Substituierbarkeit der Produkte	-0,1497	13 a)
Anzahl Hotelübernachtungen	-0,5573	4
Reisezeitenanteil	0,2657	20
Echte Non-Selling-Aktivitäten	-0,0228	10
Langfristig wirkende Non-Selling-Aktivitäten	0,0124	10
Anzahl Besuche bis Abschluß	-0,5533	3
Produktspezifische Kenntnisse	0,0003	24.500 b)
Risikoaversion	-0,3224	1
Erfahrung im Verkauf vor Eintritt in das Unternehmen	0,1899	4,8 b)
Unternehmenszugehörigkeit	0,1016	3,48 b)
Effektivitätsindex des Alters	4,0949	0,38 b)
Bildungsniveau	-0,0004	273 b)
Branchenübliches Einkommen	0,00004	100.443
Output als Leistungsmaß ungeeignet, Input geeignet	10,9786	0,65 a)

a): Zur Berechnung dieser Kompositskala siehe *Krafft*, 1995a, S. 251 ff.

b): Zur Berechnung dieser Meßvariablen siehe *Krafft*, 1995a, S. 224 ff.

In *Tab. 4* werden die Antworten des Beispielunternehmens wiedergegeben, die zusammen mit den Koeffizienten des Modells der Logistischen Regression in *Gleichung (7)*

eingesetzt wurden. Die geschätzte Wahrscheinlichkeit, daß ein Reisenden-Außendienst als Absatzform gewählt wird (P_i), beträgt nach Maßgabe des Logistischen Modells 99,997%. Da das betrachtete Unternehmen zur Zeit eine Reisenden-Vertriebsorganisation einsetzt, entspricht die gewählte Absatzform der in der Stichprobe beobachteten üblichen Praxis.

6. Abschließende Beurteilung der Logistischen Regression

Für den hier betrachteten Fall dichotomer Fragestellungen wurde mit der Logistischen Regressionsanalyse eine Methode vorgestellt, die gegenüber geläufigen Verfahren wie etwa der Zwei-Gruppen-Diskriminanzanalyse nennenswerte Vorteile aufweist. So erfordert dieser Ansatz nicht die für die Diskriminanzanalyse häufig kritische Prämisse homogener Varianzen in den Gruppen, und es können explizit auch kategoriale unabhängige Variablen einbezogen werden. Des weiteren können für die Koeffizienten der Logistischen Regressionsmodelle inferenzstatistische Aussagen zum Einfluß einzelner Variablen getätigt werden, was in der Diskriminanzanalyse nicht möglich ist.

Im vorliegenden Beitrag wurden einzelne Gütemaße für Logistische Regressionsmodelle vorgestellt und -soweit möglich- kritische Werte oder Bandbreiten angegeben. Anhand eines Logistischen Regressionsmodells von *Krafft* (1995b) wurde gezeigt, wie die Koeffizienten der gefundenen Lösung mit Hilfe von partiellen Ableitungen, Elastizitäten und der Sensitivitätsanalyse interpretiert werden können.

Die Logistische Regression stellt als äußerst robustes Schätzverfahren eine Alternative zu herkömmlichen Methoden der Analyse dichotomer endogener Variablen dar. Insbesondere in Bereichen wie der Analyse der Kreditwürdigkeit von Schuldnern (Credit-Scoring-Modelle) und des Konsumentenverhaltens (Kauf/Nichtkauf) bietet sich eine Anwendung dieses Verfahrens an. Dabei ist jedoch zu beachten, daß die Logistische Regressionsanalyse nicht zwangsläufig das dominant bessere Analyseinstrument darstellt. So ist insbesondere inhaltlich zu prüfen, ob der unterstellte logistische Verlauf der Wahlwahrscheinlichkeit P_i für den Untersuchungsgegenstand angemessen ist. Sofern beispielsweise eine lineare Änderung von P_i adäquater ist, sollte vom logistischen Modell abgesehen werden. Dann ist vielmehr zu prüfen, ob lineare Ansätze wie das LPM („linear probability model“) oder die Diskriminanzanalyse anzuwenden sind (vgl. *Aldrich und Nelson*, 1984).

Die Logistische Regression und insbesondere die mit ihrer Hilfe geschätzten Modelle sind keineswegs trivial. Da im vorliegenden Beitrag gezeigt wurde, daß die Modelle und Koeffizienten gut interpretierbar sind, hofft der Autor, daß dieses Verfahren in Zukunft bei der Analyse dichotomer Zusammenhänge häufiger in Betracht gezogen wird. Insbesondere werden methodisch interessierte Forscher aufgefordert, die relative Stärke alternativer Verfahren (Diskriminanzanalyse, Lineare Regressionsanalyse, Neuronale Netze) zu untersuchen (vgl. *Kumar et al.*, 1995).

Literatur

- Aldrich, J.; Cnudde, C. F. (1975): Probing the Bounds of Conventional Wisdom: A Comparison of Regression, Probit, and Discriminant Analysis, in: American Journal of Political Science, 19, S. 571 - 608.*
- Aldrich, J. H.; Nelson, F. D. (1984): Linear Probability, Logit, and Probit Models, Beverly Hills and London.*
- Anderson, E.; Coughlan, A. T. (1987): International Market Entry and Expansion via Independent or Integrated Channels of Distribution, in: Journal of Marketing, 51, (January), S. 71 - 82.*
- Backhaus, K., et al. (1994): Multivariate Analysemethoden: Eine anwendungsorientierte Einführung, 7. Auflage, Berlin et al.*
- Crask, M. R.; Perreault, W. D., Jr. (1977): Validation of Discriminant Analysis in Marketing Research, in: Journal of Marketing Research, 14, S. 60 - 68.*
- DeMaris, A. (1992): Logit Modeling: Practical Applications, Newbury Park, CA.*
- Fahrmeir, L.; Hamerle, A. (1984): Grundlegende multivariate Schätz- und Testprobleme, in: Fahrmeir, L. und Hamerle, A. (Hrsg.): Multivariate statistische Verfahren, Berlin/New York, S. 49 - 82.*
- Gessner, G., et al. (1988): Estimating Models with Binary Dependent Variables: Some Theoretical and Empirical Observations, in: Journal of Business Research, 16, (1), S. 49 - 65.*
- Green, P. E., et al. (1977): On the Analysis of Qualitative Data in Marketing Research, in: Journal of Marketing Research, 14, S. 52 - 59.*
- Hair, Jr., Joseph F., et al. (1992): Multivariate Data Analysis with Readings, New York.*
- Hosmer, J., David W.; Lemeshow, S. (1989): Applied Logistic Regression, New York et al.*
- Hruschka, H. (1991): Marktreaktionsfunktionen mit Interaktionen zwischen Marketing-Instrumenten, in: Zeitschrift für Betriebswirtschaft, 61, S. 339 - 356.*
- Krafft, M. (1995a): Außendienstentlohnung im Licht der Neuen Institutionenlehre, Wiesbaden.*
- Krafft, M. (1995b): Handelsvertreter oder Reisende? Eine Überprüfung von Hypothesen der Neuen Institutionenlehre zur Absatzformwahl, in: Manuskripte aus den Instituten für Betriebswirtschaftslehre der Universität Kiel, Nr. 370, erscheint in: Die Betriebswirtschaft, 56. Jg. (1996).*
- Kumar, A., et al. (1995): An Empirical Comparison of Neural Network and Logistic Regression Models, in: Marketing Letters, 6, S. 251-263.*

- LeClere, M. J.* (1992): The Interpretation of Coefficients in Models with Qualitative Dependent Variables, in: *Decision Sciences*, 23, S. 770 - 776.
- Maddala, G. S.* (1983): *Limited-dependent and qualitative variables in econometrics*, Cambridge et al.
- McFadden, D.* (1974): The Measurement of Urban Travel Demand, in: *Journal of Public Economics*, 3, S. 303 - 328.
- Menard, S.* (1995): *Applied Logistic Regression Analysis*, Thousand Oaks, CA.
- Morrison, D. G.* (1969): On the Interpretation of Discriminant Analysis, in: *Journal of Marketing Research*, 6, S. 156 - 163.
- Norušis, M. J.* (1993): *SPSS for Windows Advanced Statistics Release 6.0*, Chicago, IL.
- Petersen, T.* (1985): A Comment on Presenting Results from Logit and Probit Models, in: *American Sociological Review*, 50, (February), S. 130 - 131.
- Press, S. J.; Wilson, S.* (1978): Choosing Between Logistic Regression and Discriminant Analysis, in: *Journal of the American Statistical Association*, 73, S. 699 - 705.
- Urban, D.* (1993): *Logit-Analyse: Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen*, Stuttgart/Jena/New York.

Zusammenfassung

Zahlreiche betriebswirtschaftliche Fragestellungen haben dichotome abhängige Variablen zum Gegenstand, für deren empirische Analyse bisher standardmäßig die Diskriminanzanalyse eingesetzt wird. In diesem Beitrag wird mit der Logistischen Regression ein alternatives Verfahren vorgestellt, daß zunehmend Verbreitung findet. Anhand eines Beispiels wird das Schätzverfahren vorgestellt und gezeigt, wie das gesamte Modell und einzelne Koeffizienten hinsichtlich ihrer statistischen Güte zu beurteilen sind. Danach zeigt der Autor, wie partielle Ableitungen, Elastizitäten und Sensitivitätsanalysen zur Interpretation von Koeffizienten genutzt und Prognosen aufgrund Logistischer Regressionsmodelle aufgestellt werden können.

Summary

A substantial body of empirical accounting, finance, management, and marketing research deals with problems with binary dependent variables. The purpose of this paper is to introduce logistic regression as an alternative approach to analyze such problems. After explaining the estimation procedure in logistic regression, the author presents three methods of interpreting the coefficients in logistic regression and how it can be used for predictions. The estimation technique, the goodness-of-fit measures, and the interpretation of coefficients are illustrated with an example. It is shown that logistic regression has both strengths and weaknesses that must be clearly understood.