

ZfTM-Work in Progress Nr. 75:

Ordinale Regression

– Eine anwendungsorientierte Einführung –

Torsten J. Gerpott*/Ilaha Mahmudova**

© 2007

* Univ.-Prof. Dr. Torsten J. Gerpott, Lehrstuhl Unternehmens- und Technologieplanung, Schwerpunkt Telekommunikationswirtschaft, Universität Duisburg-Essen, Lotharstr. 65, 47057 Duisburg.

** Dipl.-Kff. Ilaha Mahmudova, Lehrstuhl Unternehmens- und Technologieplanung, Schwerpunkt Telekommunikationswirtschaft, Universität Duisburg-Essen, Lotharstr. 65, 47057 Duisburg.



Zentrum für
Telekommunikations- und
Medienwirtschaft

Work in Progress

für die Telekommunikations- und Medienwirtschaft

ZFTM-Work in Progress ist eine Schriftenreihe des Förderkreises Zentrum für Telekommunikations- und Medienwirtschaft e.V. Alle Angaben erfolgten nach bestem Wissen, jedoch ohne Gewähr für Richtigkeit und Vollständigkeit. Für Fehler im Text und in Grafiken wird keinerlei Haftung übernommen. Aus der Veröffentlichung kann nicht geschlossen werden, daß die beschriebene Lösung oder die verwendete Bezeichnung frei von gewerblichen Schutzrechten ist. Nachdruck oder sonstige Reproduktion (auch Auszüge) nur mit ausdrücklicher Genehmigung des Vorstandes des Förderkreises. Herausgeber (presserechtlich verantwortlich): Prof. Dr. Torsten J. Gerpott, Dipl.-Wirtsch.-Ing. Andreas Walter, Dipl.-Kff. Sandra Thomas.

Kontakt:

Förderkreis Zentrum für Telekommunikations- und Medienwirtschaft e.V.
Am Freischütz 6
47058 Duisburg
Tel 0203-3793109
Fax 0203-3792656
Internet www.zftm.de
Email kontakt@zftm.de

Vorspann

In der empirischen Wirtschaftsforschung geht es häufig darum, Zusammenhänge zwischen abhängigen Kriterien mit mindestens drei Ausprägungen, die sich in eine Rangordnung bringen lassen, und unabhängigen Erklärungsvariablen zu schätzen. Das hierfür geeignete Verfahren ist die ordinale Regression. Dieser Beitrag erläutert einführend die Prinzipien, die Beurteilung der Anpassungsgüte und die Interpretation von ordinalen Regressionsrechnungen.

1. Einordnung des Verfahrens

In den Wirtschafts- und Sozialwissenschaften wird man oft mit kategorialen Variablen bzw. Merkmalen konfrontiert. Ein kategoriales Merkmal ist durch eine endliche Anzahl von Ausprägungen, sog. Responsekategorien, gekennzeichnet (vgl. *Tutz*, 2000, S. 1). Beispiele für solche Merkmale sind Kaufverhaltenstypen (Regelmäßig-, Häufig-, Gelegenheits- und Seltenkäufer), arbeitsrechtlicher Mitarbeiterstatus (Arbeiter, Angestellter, Beamter usw.) oder Flugkategorie (Economy, Business und First Class).

Im Folgenden sind solche Fälle von Interesse, bei denen das kategoriale Merkmal als abhängige Variable auftritt. Hier muss man zwischen Variablen mit ungeordneten und geordneten Kategorien unterscheiden. Ist die abhängige Variable ungeordnet, wie z.B. der arbeitsrechtliche Mitarbeiterstatus, spricht man von einer **nominal skalierten Variablen**. Bei einem derartigen Kriterium stellt die **Diskriminanzanalyse** ein geeignetes Verfahren dar, um empirisch aufzuzeigen, durch welche unabhängigen Merkmale sich die Ausprägungskategorien der abhängigen Variablen unterscheiden (vgl. *Backhaus et al.*, 2003, S. 156-199). Lassen sich die Ausprägungen der abhängigen Variablen jedoch in eine Rangfolge bringen, wie bei Kaufverhaltenstypen oder Flugkategorien, dann liegt eine **ordinal skalierte Variable** vor. Da jedoch die Abstände zwischen den einzelnen Ausprägungen nicht zwangsläufig identisch sind, stellt die herkömmliche Regressionsanalyse für intervallskalierte Kriterien kein geeignetes Verfahren zur Erkundung von Variablenassoziationen dar. In diesem Fall sollte man die **ordinale Regression** zur Analyse von Zusammenhängen zwischen unabhängigen und abhängigen Variablen verwenden.

Weist die kategoriale abhängige Variable lediglich zwei Ausprägungen auf, kann man die **logistische Regression** einsetzen. Da dieses Verfahren in der Literatur sehr gut dokumentiert ist, wird es in der Praxis auch beim Vorliegen von mehr als zwei Ausprägungen angewendet (vgl. z.B. *Krafft*, 1997; *Reise/Bierend*, 1999). Hierzu ist es erforderlich, die mehr als zwei vorhandenen Kategorien zu insgesamt zwei Kategorien zusammenzufassen. Nachteil dieses Vorgehens ist ein erheblicher Informationsverlust. Beispielsweise gehen bei der Zusammenfassung von Fluggästen der Business und First Class zu

einer Gruppe und ihrer Gegenüberstellung mit Fluggästen der Economy Class Informationen wie der Preisunterschied zwischen den beiden zuerst genannten Flugkategorien verloren, obwohl er in der Praxis oft bedeutsam ist.

Das multivariate Verfahren der ordinalen Regression weist den Informationsverlustnachteil nicht auf, ihm wurde aber bislang in der Methodenliteratur ebenso wie in empirischen Forschungsarbeiten wenig Beachtung geschenkt. Deshalb ist es Ziel dieses Beitrags, die Prinzipien der ordinalen Regressionsanalyse einführend zu erläutern.

2. Prinzipien der ordinalen Regression

2.1. Modellschätzung

Die grundlegende Vorgehensweise der ordinalen Regression wird im Folgenden anhand eines **Fallbeispiels** erklärt: In fünf Flughäfen der Bundesrepublik Deutschland wurden jeweils Kunden der Fluggesellschaft *AZFLY* beobachtet, von denen 55 (34,4%) Economy Class, 56 (35,0%) Business Class und 49 (30,6%) First Class fliegen. Die abhängige Variable „Flugkategorie“ wird als ordinal skaliert betrachtet, d.h. die Responsekategorien lassen sich in eine Rangreihe (z.B. bezüglich der Flugticketflexibilität oder des Ticketpreises) bringen. Den oben genannten drei Kategorien werden Zahlen „0“ für die Kategorie „Economy Class“, „1“ für die Kategorie „Business Class“ und „2“ für die Kategorie „First Class“ zugeordnet. Hier ist nur die Rangfolge, nicht aber der numerische Abstand zwischen den Kategorien bzw. Werten sinnvoll interpretierbar. In einer kurzen Befragung wurde bei den Kunden ihr Einkommen (genaue Angabe der Einkommenshöhe) und der Anlass des Fluges (privat = „0“ und dienstlich = „1“) erhoben. Mit Hilfe einer ordinalen Regression soll untersucht werden, wie stark sich das Einkommen und der Fluganlass auf die gebuchte Flugkategorie auswirken.

Hierzu ist zunächst zu entscheiden, ob das Schwellenwertmodell (kumulative Modell) oder das sequenzielle Modell der ordinalen Regression verwendet werden soll. Das **sequenzielle Modell** geht davon aus, dass die Kategorien der abhängigen (= endogenen) Variablen nur sukzessiv erreichbar sind (vgl. *Fahrmeir/Hamerle/Tutz*, 1996, S. 274). Dieses Modell ist jedoch in marktgan-