Edition Psychologie

Herausgegeben von
Dr. Arno Mohr

Bisher erschienene Werke:

Güttler, Sozialpsychologie, 3. Auflage
Mayer, Einführung in die Wahrnehmungs-, Lern- und Werbe-Psychologie
Sanns · Schuchmann, Lineare und loglineare Modelle in Psychologie und Sozialwissenschaften
Schuchmann, Probabilistische Testtheorie

Lineare und loglineare Modelle in Psychologie und Sozialwissenschaften

Mit MS Excel-Programmen

Buch mit Diskette

Von
Dipl.-Math. Werner Sanns
Dipl.-Math. Marco Schuchmann

R. Oldenbourg Verlag München Wien

Die Informationen in diesem Buch und dem beiliegenden Datenträger wurden mit großer Sorgfalt erstellt. Fehler können jedoch nicht ausgeschlossen werden. Für fehlerhafte Angaben und deren Folgen werden weder juristische Verantwortung noch irgendeine Haftung übernommen.

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Sanns, Werner:
Lineare und loglineare Modelle in Psychologie und Sozialwissenschaften
: mit MS-Excel-Programmen ; Buch mit Diskette / von Werner Sanns und Marco Schuchmann. – München ; Wien : Oldenbourg, 2000
(Edition Psychologie)
ISBN 3-486-25503-7

Rosenheimer Straße 145, D-81671 München
Telefon: (089) 45051-0
www.oldenbourg-verlag.de

Gedruckt auf säure- und chlorfreiem Papier
Gesamtherstellung: Huber KG, Dießen

ISBN 3-486-25503-7

0. Vorwort

Bei den linearen und loglinearen Modellen der Statistik handelt es sich um moderne Analysemethoden für qualitative Daten, welche in Psychologie und den Sozialwissenschaften weitaus häufiger anfallen als quantitative Daten. Für letztere gibt es viele bekannte Analyseverfahren, während für rein qualitative Daten die Anzahl der gängigen Verfahren geringer ist. Am bekanntesten ist dabei der Chi-Quadrat Test auf Unabhängigkeit, mit dessen Hilfe man die Unabhängigkeit zweier qualitativer Variablen überprüfen kann.
Lineare bzw. loglineare Modelle gehen jedoch viel weiter. Mit linearen und loglinearen Modellen kann man versuchen, bei qualitativen Daten die Stärke von Einflüssen und Abhängigkeiten zahlenmäßig zu erfassen, ähnlich wie es in der Regressionsanalyse für quantitative Daten getan wird. Man kann die Daten überdies auf Wechselwirkungen untersuchen und diese gegebenenfalls statistisch nachweisen. Allerdings sind lineare Modelle für qualitative Daten nicht in allen kommerziellen Statistikpaketen realisiert, oder müssen zumindest selbst in dem jeweiligen System programmiert werden. Die Literatur und Systemhandbücher hierzu sind meist von so hoher mathematischer Anforderung, daß diese Modelle nur selten von empirisch arbeitenden Psychologen und Sozialwissenschaftlern eingesetzt werden.

Unser Anliegen war es daher mit diesem Buch und dem beigefügten Datenträger folgendes zu realisieren:

- Die Darstellung der Theorie der genannten mathematischen Modelle, soweit sie für den Psychologen und Sozialwissenschaftler notwendig ist.
- Die Erstellung eines Programms, das auf einem weit verbreiteten Softwaresystem (Excel) einsetzbar ist, und mit welchem, über die Beispiele des Buches hinaus, in der Praxis leicht gearbeitet werden kann.

- Die Berechnungen und ihre Resultate sollen verständlich erklärt werden.
- Die erzeugten Ergebnisse von Berechnungen sollen grafisch dargestellt werden.
- Der überwiegend an der praktischen Durchführung der Verfahren interessierte Anwender soll seine Daten direkt in die ihn interessierenden Kreuztabellen eingeben können und sofort die Resultate erhalten.

Zu diesem Zweck haben wir auf dem Datenträger vier Exceldateien im Format von Excel 97 abgelegt, die somit auch in Excel 2000 eingelesen werden können. Diese werden zu Beginn des Buches und auch vor der Ausführung der jeweiligen Beispiele genau erklärt. Mit Hilfe dieser Programme werden im Buch alle Beispiele zu den Verfahren durchgerechnet. Die Verfahren sind so programmiert, daß der Leser eigene Datenanalysen für eine große Anzahl praktisch auftretender Datenstrukturen selbst durchführen kann. Er erhält somit ein gut getestetes Werkzeug zur Analyse qualitativer Daten. Für konstruktive Kritik und freundliche Hinweise zu unserem Buch möchten wir uns bei unseren Lesern im voraus herzlich bedanken.
Die Autoren sind erreichbar unter folgender e-mail Adresse:
m.schuchmann@t-online.de

Werner Sanns, Marco Schuchmann

Inhalt

1 Daten

Statistisch auszuwertende Daten können ganz verschiedener Natur sein. Wir wollen daher zunächst statistische Daten nach verschiedenen Kriterien klassifizieren.
Zuerst unterscheiden wir, ob die Daten stetig oder diskret sind. Stetige Daten sind solche, die innerhalb eines gewissen Intervalls jeden reellen Zahlenwert annehmen können. Zum Beispiel liefern die Messungen von Längen, Temperaturen, Gewichten etc. stetige Daten (obwohl die Messgenauigkeit schließlich doch dazu führt, daß Zahlenwerte auf eine feste Stellenzahl gerundet werden).
Im Gegensatz dazu können bei diskreten Daten die Werte nicht ein ganzes reelles Intervall überdecken, sondern es können nur endlich viele (oder höchstens „abzählbar unendlich viele") Werte als Wertebereich zugrunde liegen. Daten müssen keineswegs immer Zahlen sein, sondern können auch Begriffe, wie Farben, Stimmungen etc., repräsentieren. Wir sprechen dann von qualitativen oder kategoriellen Daten.

Unabhängig von dieser Einteilung wird auch zwischen metrischen und nichtmetrischen Daten unterschieden. Mit metrischen Daten können numerische Berechnungen sinnvoll vorgenommen werden. So hat zum Beispiel die Körpergröße metrisches Datenniveau, denn es ist durchaus sinnvoll die Differenz zwischen zwei Körpergrößen zu berechnen. Außerdem ist es bei metrischen Daten zulässig den Mittelwert zu berechnen. Dies ist bei nichtmetrischen Daten anders, denn es ist beispielsweise sinnlos, den Mittelwert zu berechnen, wenn die Daten die Augenfarben von Personen darstellen, auch wenn diese Farben durch Zahlen kodiert würden. Allgemein sind numerische Operationen, wie Addition, Subtraktion usw. bei qualitativen Daten nicht sinnvoll. Bei nichtmetrischen Daten können wir noch zwischen nominal und ordinal skalierten Daten unterscheiden. Die Augenfarbe wäre zum Beispiel nominal skaliert, da es zwischen den einzelnen Augenfarben keine

Rangfolge gibt, wie dies für ordinale der Fall ist. Erfaßt man dagegen die Antwort auf die Frage „Wie ist Ihr Befinden heute?“ mit den Antwortmöglichkeiten „schlecht“, „mittelmäßig“, „gut“, „sehr gut“, so genügt die Variable, die das Befinden erfaßt, einer ordinalen Skala, denn hier liegt eine Rangfolge vor. Auch Schulnoten sind ordinale Daten. Sie werden mit den Zahlen 1,2,..bis 6 kodiert, dahinter stehen aber Beurteilungen von Leistungen, die man zwar ordnen kann, mit denen aber nicht sinnvoll gerechnet werden kann. (In Klassenarbeiten ist zum Beispiel der Bereich einer Leistung, die mit "gut" bewertet wird, meist weit enger, als ein Leistungsbereich der mit "ausreichend" bewertet wird). Daß Mittelwertbildung mit diesen kodierten Werten nicht sinnvoll ist, sehen Sie sofort ein, wenn Sie die Noten durch beliebige andere geordnete Zahlen verschlüsseln, die jedoch andere Abstände besitzen, und Sie dann den „Mittelwert“ bilden.
Abhängig vom Datenniveau gibt es zahlreiche statistische Verfahren zur Datenanalyse. Wir geben hier einen Überblick über die gängigsten Verfahren bei den verschiedenen Datenniveaus:

	abhängige Variable	
Unabhängige Variablen	*quantitativ*	*qualitativ*
quantitativ	Regression	Logistische Regression
qualitativ	Varianzanalyse	Lineare / loglineare Modelle

Für den gemischten Fall, bei dem die unabhängigen Variablen teils qualitativer teils quantitativer Natur sind und die abhängige Variable quantitativ ist, steht die Kovarianzanalyse zur Verfügung. Sind auch die abhängigen Variablen von gemischter Natur ist die logistische Kovarianzanalyse angebracht. Zum Thema Regressions- bzw. Varianzanalyse gibt es zahlreiche Bücher und fertige Programme. Unser Buch deckt den in der Literatur und in Statistikprogrammpaketen weit weniger häufig zu findenden Themenkomplex ab, der in der letzten Spalte

der oberen Tabelle zu sehen ist: Logistische Regression und lineare/loglineare Modelle für qualitative Daten.

Die Programme, die Sie ab dem nächsten Kapitel kennenlernen werden, sind teilweise für sogenannte dichotome Variablen ausgelegt, das heißt für Variablen mit zwei Kategorien. Daher möchten wir an dieser Stelle noch einige Bemerkungen zur Dichotomisierung von Variablen machen. Theoretisch könnte man auch lineare Modelle rechnen, bei denen die Variablen mehr als zwei Kategorien aufweisen. Damit unser System übersichtlich bleibt, haben wir uns aber bei den linearen Modellen auf dichotome Variablen festgelegt. Falls Ihre Variablen also mehr Kategorien aufweisen, und Sie ein lineares Modell anwenden wollen, so müssen Sie Variablen dichotomisieren, das heißt, Sie müssen aus 3,4,5 oder mehr Kategorien zwei Kategorien bilden. Das hat zunächst den Anschein, als ginge dabei Information für die Modellrechnung verloren. Hierzu ist jedoch folgendes zu sagen: Wir gehen davon aus, daß Sie als empirisch arbeitender Psychologe bzw. Psychologin Ihre Daten anhand eines Fragebogens gewonnen haben. Diese Daten haben im allgemeinen nicht die Eigenschaften exakter Meßwerte, wie zum Beispiel Meßdaten in der Physik. Nehmen wir einmal die Fragestellung „Wie geht es Ihnen nach der Therapie ?" mit den Antwortmöglichkeiten „schlecht", „mittel" , „gut" und „sehr gut". Falls Sie nun im Rahmen der Auswertung mit dem linearen Modell die Kategorien „schlecht" und „mittel" zusammenfassen und „gut" mit „sehr gut" zusammenfassen, so lassen sich durchaus „trennschärfere" Aussagen machen, nicht zuletzt deshalb, weil die Antworten auf die Frage ohnehin von kurzzeitigen Gemütsschwankungen der Patienten abhängt. Durch ein positives Erlebnis am Tag der Befragung, kann der Patient, der sich eigentlich nicht gut fühlt auch in die Kategorie „mittel" gelangen. Es lassen sich daher keine genauen Aussagen über die Wahrscheinlichkeiten für die Zugehörigkeit zu einer bestimmten Gruppe machen. Dichotomisiert man dagegen, so wird der Patient einer Gruppe zugeordnet, die diese kurzzeitigen Schwankungen eher nicht berücksichtig sondern „ausglättet". In dieser Hinsicht werden die Vorhersagen mit dem Modell eher genauer. Außerdem werden die

Ergebnisse ohne Dichotomisierung schlecht interpretierbar, da sie meist viele Fallunterscheidungen treffen. Für jede Variable im Modell steigt mit der Anzahl der Kategorieren die Anzahl der Parameter. Dadurch werden zum einen die rechnerischen Ergebnisse und die Grafiken unübersichtlicher, zum anderen hat die Kontingenztafel bei vielen Kategorien oft viele gering besetzte oder sogar überhaupt nicht besetzte Zellen. Zum Beispiel gibt es bei 2 Variablen mit je 4 Kategorien bereits 4*4=16 Zellen in der Kontingenztafel. Bei 3 Variablen mit je 4 Kategorien gibt es bereits 64 Zellen. Die erforderliche Anzahl an Daten, um eine solche Kontingenztafel in allen Zellen mit einem Mindestmaß an Werten aufzufüllen, geht daher sehr rasch in die Höhe.
Aus diesem Grund haben wir in praktischen Studien in Zusammenarbeit mit mehreren Psychologen bisher meist ein Modell mit dichotomen Variablen angewandt, was in der Praxis gut geeignet war.

2 Die Handhabung der Excel-Arbeitsmappen

Sie finden auf dem Datenträger zu diesem Buch 4 Excel-Arbeitsmappen. Diese sind im Format von Excel 97 (unter Windows 95/98 oder NT) als Datei abgespeichert und sind somit auch in Excel 2000 einlesbar. Alle in den folgenden Kapiteln beschriebenen Verfahren können mit den Excel-Arbeitsmappen LINMOD, LOGLIN und LOGIT durchgeführt werden. Dabei wird bei diesen Arbeitsmappen davon ausgegangen, daß die zu untersuchenden Daten bereits in der Form von Kreuztabellen vorliegen. Für diejenigen Leser, die kein Statistikprogramm zur Erzeugung von Kreuztabellen aus ihren Rohdaten besitzen, haben wir eine vierte Arbeitsmappe KONTINGENZ auf der Diskette gespeichert, mit deren Hilfe solche Kreuztabellen erzeugt werden können.

Sie sollten aus Gründen der Ausführungsgeschwindigkeiten zuerst die 4 Dateien LINMOD.XLS, LOGLIN.XLS, LOGIT.XLS und KONTINGENZ.XLS von dem beigefügten Datenträger auf ein von Ihnen erstelltes neues Verzeichnis der Festplatte kopieren, sagen wir zum Beispiel in ein Verzeichnis namens „Lineare Modelle".

Zunächst erklären wir ganz allgemein, was die vier Programme leisten. Ihre praktische Benutzung wird dann erklärt, wenn sie bei den Beispielen im Buch zum Einsatz kommmen.
Starten Sie Excel (Version 97 oder 2000) und öffnen Sie die Datei LINMOD.XLS, die Sie auf ihre Festplatte kopiert haben. Mit LINMOD können lineare Modelle der kategoriellen Datenanalyse mit einer, zwei oder drei dichotomen Faktorvariablen (d.h. unabhängigen Variablen) und einer dichotomen Responsevariablen (d.h. abhängigen Variable) berechnet werden. Zusätzlich kann ein lineares Modell mit zwei Faktorvariablen und mit Wechselwirkungen berechnet werden.