Marketing

Eine managementorientierte Einführung

Bearbeitet von
Prof. Dr. Franz-Rudolf Esch, Prof. Dr. Andreas Herrmann, Prof. Dr. Henrik Sattler

5. Auflage 2017. Buch. XX, 500 S. Kartoniert
ISBN 978 3 8006 5470 3
Format (B x L): 17,0 x 24,0 cm

Quelle	df	Abweichungsquadratsumme	Mittel der Quadrate	F-Wert	p-Wert
Regression	3	$\sum_i (\hat{Y}_i - \bar{Y})^2 = 51233{,}6$	17 077,9	16,51	<0,0001
Residuen	16	$\sum_i (Y_i - \hat{Y}_i)^2 = 16550{,}4$	1034,4		
Gesamt	19	$\sum_i (Y_i - \bar{Y})^2 = 67784{,}0$			

Abbildung C.27: ANOVA des Regressionsmodells

Zur Bestimmung des F-Werts werden die jeweiligen Abweichungsquadratsummen durch die Anzahl der Freiheitsgrade (df) geteilt, um so die mittlere Quadratsumme zu erhalten. Der Quotient aus der mittleren Quadratsumme, die durch das Regressionsmodell erklärt wird, und der nicht erklärten mittleren Quadratsumme ergibt den empirischen F-Wert. Die Freiheitsgrade entsprechen der Anzahl der Parameter minus 1 (im Beispiel: $k - 1 = 3$) sowie der Anzahl der Fälle abzüglich der Anzahl zu schätzender Parameter ($n - k = 16$). In diesem Fall wäre das gesamte Regressionsmodell mit einem empirischen F-Wert von 16,51 hoch signifikant ($p < 0{,}0001$). Dies bedeutet, dass die Nullhypothese, nach welcher keiner der Variablen Preis, Werbebudget und Handzetteleinsatz den Absatz erklären kann (H_0: b=c=d=0), mit einer Irrtumswahrscheinlichkeit von 0,01 % verworfen werden kann.

Aus den Quadratsummen der Varianzanalyse lässt sich außerdem ein zentrales Gütekriterium für das Modell ableiten. Das **Bestimmtheitsmaß R^2** basiert ebenfalls auf dem Varianzzerlegungssatz und entspricht dem Anteil der durch das Regressionsmodell erklärten Abweichungsquadratsumme bezogen auf die gesamte Quadratsumme von Y.

> Zentrales Gütekriterium der Regression ist das Bestimmtheitsmaß, welches den Anteil der durch das Regressionsmodell erklärten Varianz an der Gesamtvarianz von Y angibt.

Wird die gesamte Varianz der abhängigen Variablen durch die unabhängigen Variablen erklärt, ergibt sich somit ein Bestimmtheitsmaß von 1. Im Beispiel berechnet es sich, wie aus Abbildung C.27 ersichtlich, als $R^2 = 51233{,}6/67784{,}0 = 0{,}756$. Die Unterschiede im Absatz lassen sich also zu 75,6 % durch den Preis, das Marketingbudget und die Handzettelverteilung erklären. Da jedoch das Bestimmtheitsmaß mit jeder zusätzlich aufgenommenen Variablen zunimmt, unabhängig davon, ob die Variable für das Modell relevant ist oder nicht, wird häufig zusätzlich ein korrigiertes R^2 angegeben. Dieses korrigiert den ursprünglichen Wert wie folgt:

$$R^2_{korr} = R^2 - \frac{(1 - R^2) \cdot (k - 1)}{(n - k)}$$

Insgesamt muss bei der linearen Regression beachtet werden, dass ihr Einsatz nur dann gerechtfertigt ist, sofern verschiedene Voraussetzungen erfüllt sind, die für das Modell bzw. die Schätzung gelten. Zu den Prämissen gehören (z. B. Lehmann/Gupta/Steckel, 1998, S. 482 ff.):

- **Linearität**, d. h. ein linear-additiver Zusammenhang zwischen den Variablen. Ein linearer Zusammenhang kann dabei auch approximativ für einen bestimmten Wertebereich gelten oder durch Transformation hergestellt werden.
- **Homoskedastizität**, d. h. die beobachteten Residuen dürfen nicht von den Prädiktorvariablen und von der Reihenfolge der Beobachtungen abhängen.
- **Fehlende Autokorrelation** bzw. Unkorreliertheit der Residuen, d. h. die Verteilung der Residuen muss zufällig sein. Die Residuen dürfen also nicht voneinander abhängen. Autokorrelation tritt vor allem bei Zeitreihenanalysen auf.
- **Normalverteilung** der Residuen. Diese Annahme ist u. a. für das Schätzen auf Signifikanz erforderlich.

Problematisch für die Schätzung und Interpretation der Ergebnisse ist es weiterhin, wenn die unabhängigen Variablen untereinander hoch korreliert sind. Treten hohe Abhängigkeiten der Variablen, d. h. **Multikollinearität** auf, führt dies dazu, dass die Koeffizienten der korrelierten Variablen verzerrt werden und somit keine reliablen Aussagen möglich sind.

Durch die Quadrierung der Residuen mit der Methode der kleinsten Quadrate haben weiterhin Ausreißer, d. h. Extremwerte, einen großen Einfluss auf die Schätzergebnisse. Treten diese Fälle auf, sind sie sorgfältig zu berücksichtigen und ggf. aus dem Datensatz zu eliminieren.

Zur Überprüfung der Annahmen und zur Identifikation von möglichen Problemen können grafische Betrachtungen oder geeignete Testverfahren herangezogen werden (z. B. Lehmann/Gupta/Steckel, 1998, S. 482 ff.).

Logistische Regression

Die logistische Regression untersucht analog zur linearen Regression Beziehungen zwischen einer abhängigen Variablen (Y) und einer oder mehreren unabhängigen Variablen (X). Der Unterschied ist, dass die **abhängige Variable** bei der logistischen Regression **binär** ausgeprägt ist, d. h. lediglich zwei Werte annehmen kann (kodiert als 0 und 1). Die unabhängigen Variablen können, wie bei der linearen Regression, metrisch oder nicht-metrisch skaliert sein (z. B. Lehmann/Gupta/Steckel, 1998, S. 695 ff.; Frenzen/Krafft, 2008).

> Die logistische Regression dient dazu, eine binär ausgeprägte Variable durch eine oder mehrere unabhängige Variablen zu erklären bzw. zu prognostizieren.

Anstelle logistischer Regressionen wurden bislang vielfach Diskriminanzanalysen eingesetzt. Aufgrund spezifischer statistischer Vorteile der logistischen Regression

(Frenzen/Krafft, 2008) werden Diskriminanzanalysen mittlerweile jedoch kaum noch eingesetzt.

Beispiele für die Anwendungen der logistischen Regression sind die Untersuchung der Determinanten für den Kauf oder Nicht-Kauf von Produkten, die Wahl zwischen Handelsvertretern und Reisenden oder die Wahl einer Hersteller- oder Handelsmarke. Mithilfe der logistischen Regression lässt sich beispielsweise analysieren, unter welchen Rahmenbedingungen (unabhängige Variablen wie z. B. Ausmaß der Reisetätigkeiten oder Anzahl der zu vertreibenden Produkte) die Wahl eines Handelsvertreters gegenüber dem Einsatz eines Reisenden (abhängige Variable) vorteilhaft ist.

Die logistische Regression ist der linearen Regression ähnlich. Geht man zunächst vom linear-additiven Zusammenhang aus, gilt analog zur linearen Regression $\hat{Y} = a + b\,X$. Diese Modellierung einer Variablen, die nur die Werte 0 oder 1 annehmen kann, ist allerdings nicht angemessen, da hier Y beliebige Werte annehmen könnte.

Anstatt direkt die Werte der abhängigen Variablen Y zu schätzen, geht die logistische Regression dazu über, die Wahrscheinlichkeit des Eintritts von Y, p(Y), zu schätzen.

Somit ergibt sich: $\hat{p}(Y) = a + b \cdot X$ (1)

Da die geschätzte Wahrscheinlichkeit $\hat{p}(Y)$ allerdings auch keine beliebigen Werte annehmen kann, sondern nur zwischen den Werten 0 und 1 definiert ist, muss eine weitere Modifikation vorgenommen werden:

$$\ln\left(\frac{\hat{p}(Y)}{1-\hat{p}(Y)}\right) = a + b \cdot X \qquad (2)$$

Der Term $\hat{p}(Y)/(1-\hat{p}(Y))$, also die Wahrscheinlichkeit zur Gegenwahrscheinlichkeit, wird als **odds** bezeichnet. Die odds sind definiert für das Intervall [0,+∞[. Die Logarithmierung der odds (log odds bzw. **Logit**) ergibt dann die gewünschte lineare Beziehung im erwünschten Wertebereich von]–∞,+∞[. Die Wahrscheinlichkeit des Eintretens von Y als abhängige Variable ergibt sich nun durch Rücktransformation:

$$\hat{p}(Y) = \frac{\exp(a + b \cdot X)}{(1 + \exp(a + b \cdot X))} = \frac{1}{(1 + \exp(-1 \cdot (a + b \cdot X)))} \qquad (3)$$

Nach Transformation wird durch (3) die S-förmige logistische Funktion beschrieben, die auf das Intervall zwischen 0 und 1 definiert ist. Ein Beispiel für eine solche logistische Funktion ist in Abbildung C.28 aufgeführt.

Während bei der linearen Regression die Schätzer der unabhängigen Variablen den direkten Einfluss auf die abhängige Variable angeben, ist die Interpretation der Schätzer bei der logistischen Regression schwieriger. Sie geben den Einfluss auf die log odds an (Gleichung (2)). Dieser Wert sowie das weitere Vorgehen bei der logistischen Regression sollen an einem Beispiel erläutert werden.

Das Beispiel in Abbildung C.28 stellt eine Untersuchung der Teilnahme an fünf Wellen einer Online-Panelumfrage (Y, mit 0 = keine Teilnahme, 1 = Teilnahme) in Abhängigkeit von der Entlohnung der Panelmitglieder (E: Wert in Euro) sowie des Besitzes einer Internet-Flatrate (F, mit 0 = nicht vorhanden, 1 = vorhanden) dar. Der dazugehörige Datensatz umfasst 110 Werte. Wird zunächst nur der Einfluss des Entgelts (E) auf die Wahrscheinlichkeit der Teilnahme p(Y) betrachtet, ergibt sich das in Abbildung C.28 angegebene Streuungsdiagramm. Die Werte wurden dabei in horizontaler Richtung grafisch gestreckt. Das Strecken ist lediglich als Hilfsmittel zu verstehen, sodass alle Datenpunkte sichtbar werden.

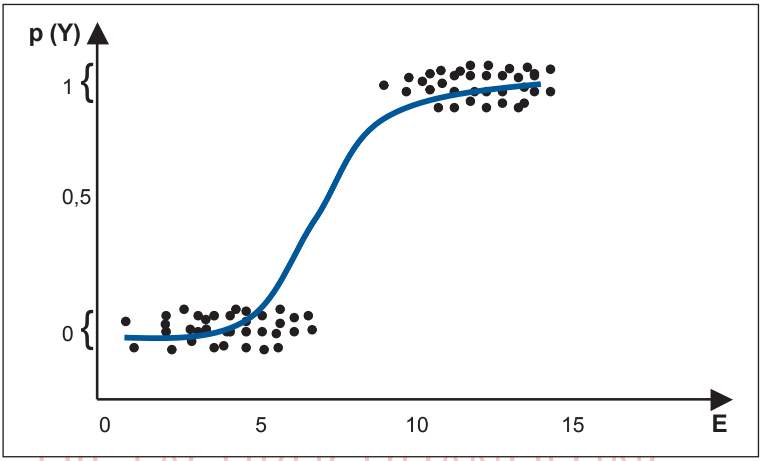

Abbildung C.28: Logistischer Funktionsverlauf

Die Grafik lässt vermuten, dass die Höhe des Entgelts (E) einen positiven Einfluss auf die Teilnahmewahrscheinlichkeit hat, da bei niedrigerer Entlohnung mehr Werte bei 0 liegen und umgekehrt. Um nun die in der Grafik dargestellte logistische Funktion zu schätzen, wird die Maximum-Likelihood-Schätzung (ML) eingesetzt. Sie wählt die Koeffizienten so, dass die Wahrscheinlichkeit der Beobachtung der empirischen Daten maximiert wird (Frenzen/Krafft, 2008, S. 629). Sie liefert für das multiple Modell das in Abbildung C.29 aufgeführte Ergebnis.

Variablen	Koeffizient	Exp (Koeffizient)	Standard-fehler	t-Wert	p-Wert
E	1,335	3,800	0,144	9,298	< 0,001
F	0,743	2,102	0,290	2,562	0,012
Konstante	–9,175		1,031	–8,901	< 0,001

Abbildung C.29: Schätzergebnisse für die logistische Regression

Die p-Werte des t-Tests (analog zur linearen Regression) weisen für alle Parameter signifikante Werte auf. Die Wahrscheinlichkeit der Teilnahme am Panel wird demnach laut (3) beschrieben durch:

$$\hat{p}(Y) = \frac{1}{(1 + \exp(-1 \cdot (-9{,}175 + 1{,}335 \cdot E + 0{,}743 \cdot F)))}$$

Mit den Schätzwerten lassen sich nun die Wahrscheinlichkeiten der Teilnahme am Panel berechnen. Für einen Probanden, der keine Internet-Flatrate besitzt (F = 0) und für die Teilnahme 5 Euro (E = 5) bekommt, gilt:

$$\hat{p}(Y \mid E = 5, F = 0) = \frac{1}{(1 + \exp(-1 \cdot (-9{,}175 + 1{,}335 \cdot 5 + 0{,}743 \cdot 0)))} = 0{,}076$$

Dieser Wert entspricht den odds von 0,076/(1–0,076) = 0,082.

Wird das Entgelt auf 6 Euro erhöht (c. p.), so ergibt sich analog eine vorhergesagte Wahrscheinlichkeit der Teilnahme von $\hat{p}(Y \mid E = 6, F = 0) = 0{,}238$ bzw. ein odds von 0,238/(1–0,238) = 0,312.

Wird der Quotient aus beiden odds gebildet, erhält man das so genannte **odds ratio** (OR).

Das odds ratio gibt den Faktor an, um den sich die Chance der Teilnahme ändert, sofern die unabhängige Variable um eine Einheit erhöht wird und alle weiteren Variablen konstant gehalten werden.

Für das Beispiel erhält man ein odds ratio von OR = 0,312/0,082 = 3,8. Eine Erhöhung des Entgelts für die Teilnahme um einen Euro steigert demnach die Chance, dass der Proband an der Umfrage teilnimmt um den Faktor 3,8. Dieser Wert entspricht exp(Koeffizient) und wird wegen der besseren Interpretation häufig in den Schätzergebnissen mit aufgeführt. Für Probanden, die eine Internet-Flatrate besitzen, kann eine um den Faktor 2,1 höhere Chance an allen Wellen des Panels teilzunehmen, angenommen werden (Abbildung C.30).

Die **Güte** einer logistischen Regression kann zum einen anhand der Likelihood-Werte der Schätzung beurteilt werden. Ein Gütekriterium, das auf den Likelihood-Werten basiert, ist z. B. McFaddens R^2 (Frenzen/Krafft, 2008, S. 634). Es kann näherungsweise analog zum

		Prognostiziert		
		Teilnahme	Keine Teilnahme	Korrekte Klassifizierung
Beobachtet	Teilnahme	49	16	75,4 %
	Keine Teilnahme	14	31	68,9 %
		Gesamtprozentsatz		**72,7 %**

Abbildung C.30: Klassifikationsmatrix

Bestimmtheitsmaß der linearen Regressionsanalyse interpretiert werden. Ein weiteres Kriterium zur Beurteilung der Güte stellt der Anteil der durch das Modell richtig klassifizierten Fälle dar. Werden Wahrscheinlichkeitswerte von p(Y) > 0,5 als Teilnahme und Werte darunter als Nicht-Teilnahme interpretiert, so lassen sich diese Vorhersagen mit den tatsächlichen Werten vergleichen. Es ergibt sich die in Abbildung C.30 dargestellte Klassifikationsmatrix (allgemein Frenzen/Krafft, 2008, S. 635). Der Anteil insgesamt korrekt klassifizierter Fälle, der auch als **hit ratio** bezeichnet wird (Frenzen/Krafft, 2008, S. 635), entspricht im Beispiel demnach (49 + 31) /110 = 0,727, also 72,7 %.

Strukturgleichungsmodelle

Wie die Regressionsanalyse ist auch die Strukturgleichungsmodellierung ein Verfahren zur empirischen Überprüfung von **Ursache-Wirkungs-Zusammenhängen** (Hildebrandt/Homburg, 1998; Weiber/Mühlhaus, 2010). Im Vergleich zur Regressionsanalyse weisen Strukturgleichungsmodelle jedoch insbesondere die folgenden Vorteile auf:

- Während die Regressionsanalyse eine fehlerfreie Messung der Variablen unterstellt, erlaubt die Strukturgleichungsmodellierung die explizite Berücksichtigung von Messfehlern.
- Strukturgleichungsmodelle ermöglichen im Gegensatz zur Regressionsanalyse die Überprüfung komplexer Abhängigkeitsstrukturen, wie z. B. wechselseitige Abhängigkeiten (z. B. Markeneinstellung ↔ Markenwahl) oder kausale Ketten (z. B. Werbung → Markenbekanntheit → Kaufabsicht).
- Korrelationen können zwischen den unabhängigen Variablen explizit im Rahmen der Modellformulierung und -schätzung berücksichtigt werden. Sie stellen somit im Gegensatz zur Regressionsanalyse (Multikollinearitätsproblem) kein prinzipielles Problem dar.

> Strukturgleichungsmodelle überwinden verschiedene Restriktionen der Regressionsanalyse, indem sie Messfehler explizit berücksichtigt, komplexe Abhängigkeitsstrukturen überprüft und Korrelationen zwischen den unabhängigen Variablen prinzipiell zulässt.

Zur Veranschaulichung der Strukturgleichungsmodellierung soll auf eine Studie zurückgegriffen werden, die sich mit den Erfolgsfaktoren von Markentransfers beschäftigt (Völckner, 2003; siehe hierzu genauer Kapitel E. 3.). Die Studie basiert auf einer Konsumentenbefragung (Quotenstichprobe, n = 2426). Die Aufgabe der Befragten bestand darin, eine Reihe von Markentransfers anhand verschiedener Merkmale zu beurteilen. Dabei wurden unter anderem die Größen Muttermarkenstärke, Marketingunterstützung des Transferprodukts und subjektiv wahrgenommener Erfolg des Transferprodukts berücksichtigt. Die genannten Größen stellen **komplexe Konstrukte** dar, die nicht direkt beobachtet und gemessen werden können.

Strukturgleichungsmodelle sind in der Lage, Wirkungszusammenhänge zwischen **komplexen Konstrukten** zu überprüfen. Eine grundlegende Besonderheit von Strukturgleichungsmodellen liegt in der Unterscheidung von **beobachteten** (d. h. direkt messbaren) Variablen (so genannten Indikatorvariablen bzw. Items, z. B. die subjektiv wahrgenommene Qualität der Muttermarke) und **latenten Variablen** (z. B. das Konstrukt Muttermarkenstärke). Letztere stellen komplexe Größen dar, die nicht direkt gemessen werden können, sondern nur indirekt über die Indikatorvariablen erfasst werden können.

Die Grundidee von Strukturgleichungsmodellen besteht nun darin, auf der Basis empirisch gemessener Varianzen und Kovarianzen von Indikatorvariablen Rückschlüsse auf die Wirkungszusammenhänge zwischen den zugrunde liegenden latenten Variablen zu ziehen (Homburg/Pflesser/Klarmann, 2008, S. 557). Die Analyse kann in vier Ablaufschritte unterteilt werden, die im Folgenden näher betrachtet werden sollen.

1. Schritt: Modellformulierung

Ein vollständiges Strukturgleichungsmodell besteht aus jeweils einem Messmodell für die latenten exogenen (d. h. erklärenden) und latenten endogenen (d. h. durch die Kausalstruktur erklärten) Variablen und aus einem Strukturmodell.

Im Rahmen des **Messmodells** wird festgelegt, welche Konstrukte betrachtet und wie diese gemessen werden sollen. In der Regel wird eine einzelne latente Variable über mehrere Indikatorvariablen (Items) gemessen. Dabei wird unterstellt, dass jeder Indikator eine fehlerbehaftete Messung der zugrunde liegenden latenten Variablen darstellt (Hildebrandt/Homburg, 1998). Das Messmodell des Strukturgleichungsmodells entspricht damit dem Grundgedanken der konfirmatorischen Faktorenanalyse (siehe hierzu auch Kapitel C. 4.2). In Abbildung C.31 ist das Messmodell für das Konstrukt Muttermarkenstärke des Anwendungsbeispiels dargestellt. Auf einer 7-stufigen Ratingskala wurde jeweils der Grad der Zustimmung zu den einzelnen Items erfasst. Analog sind Messmodelle für die übrigen Konstrukte zu formulieren.

Ein Strukturgleichungsmodell kann 'reflektive' und 'formative' Konstrukte enthalten. Das Strukturgleichungsmodell in Abbildung C.31 umfasst ausschließlich **reflektive Konstrukte**. Ein reflektives Konstrukt ist dadurch gekennzeichnet, dass es die ihm zugeordneten Indikatoren verursacht. Die Beziehungen zwischen Konstrukt und zugehörigen Indikatoren werden hier über Faktorladungen beschrieben. Daneben gibt es **formative Konstrukte**, deren wesentliches Merkmal darin zu sehen ist, dass sie durch ihre Indikatoren verursacht werden. Die Indikatoren sind nicht austauschbar, sondern in ihrer Gesamtheit zur vollständigen Erfassung des betrachteten Konstrukts erforderlich (Chin, 1998).

Das **Strukturmodell** beschreibt die hypothetischen Beziehungen zwischen den betrachteten Konstrukten (also den latenten Variablen). Die entsprechenden Hypothesen des Anwendungsbeispiels zwischen den Konstrukten sind in Abbildung C.31 dargestellt. Beispielsweise wurden hier die Hypothesen aufgestellt, dass die Marketingunterstützung einen positiven Einfluss auf die Handelsakzeptanz ausübt (Hypothese 6) und sich die wahrgenommene Ähnlichkeit zwischen Muttermarke und Transferprodukt positiv auf den Erfolg eines Transferprodukts auswirkt (Hypothese 4). Im Strukturmodell finden sich somit Grundgedanken der Regressionsanalyse wieder.

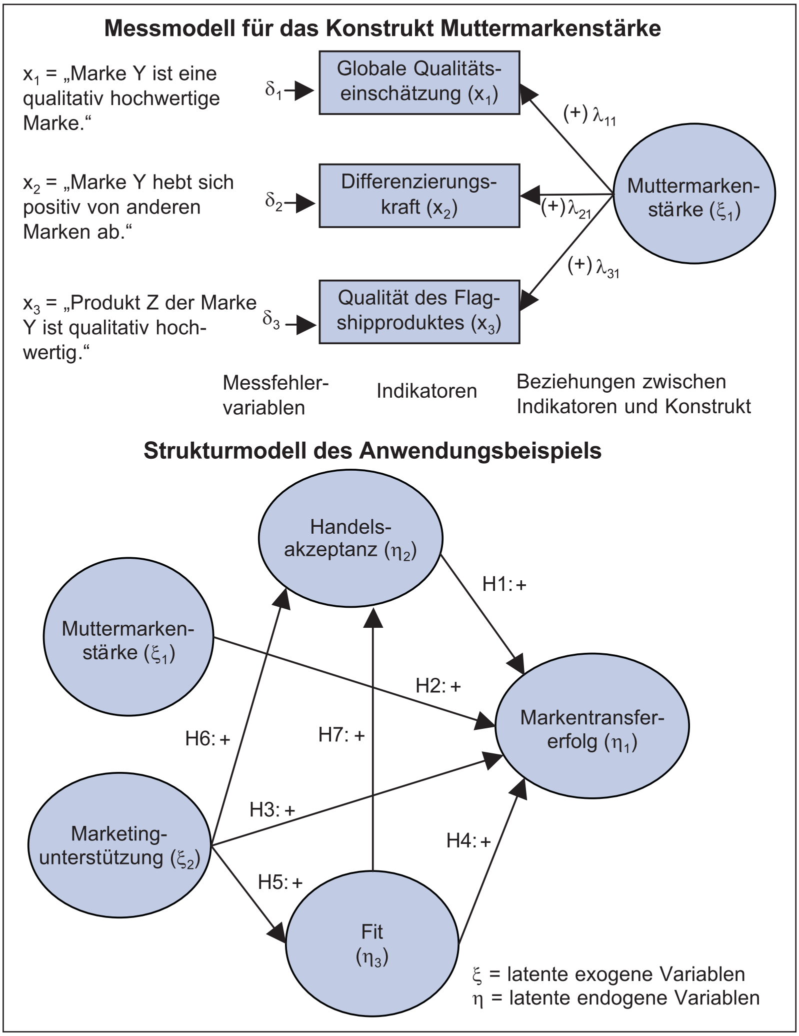

Abbildung C.31: Messmodell und Strukturmodell des Strukturgleichungsmodells

Reflektive und formative Konstrukte

Die Unterscheidung zwischen reflektiven und formativen Konstrukten soll anhand des Konstrukts „Trunkenheit“ veranschaulicht werden (Chin, 1998). Möchte man das Ausmaß der „Trunkenheit“ einer Person messen, so könnte man z. B. die Indikatoren Blutalkohol, Fahrtüchtigkeit und das Abschneiden der Person bei Kopfrechenaufgaben erfassen. „Trunkenheit“ stellt in diesem Fall ein reflektives Konstrukt dar. Denn je größer das Ausmaß der „Trunkenheit“ ist, desto schlechter wird z. B. die Fahrtüchtigkeit der Person oder ihr Abschneiden beim Kopfrechnen sein, d. h. das Konstrukt verursacht die zugrunde liegenden Indikatoren. Umgekehrt verursacht z. B.