Technische Universität Dresden
Institut für Wirtschaft und Verkehr
Lehrstuhl für Ökonometrie und Statistik, insbes. im Verkehrswesen

Skript zur Vorlesung

Verkehrsökonometrie

Methoden und Modelle

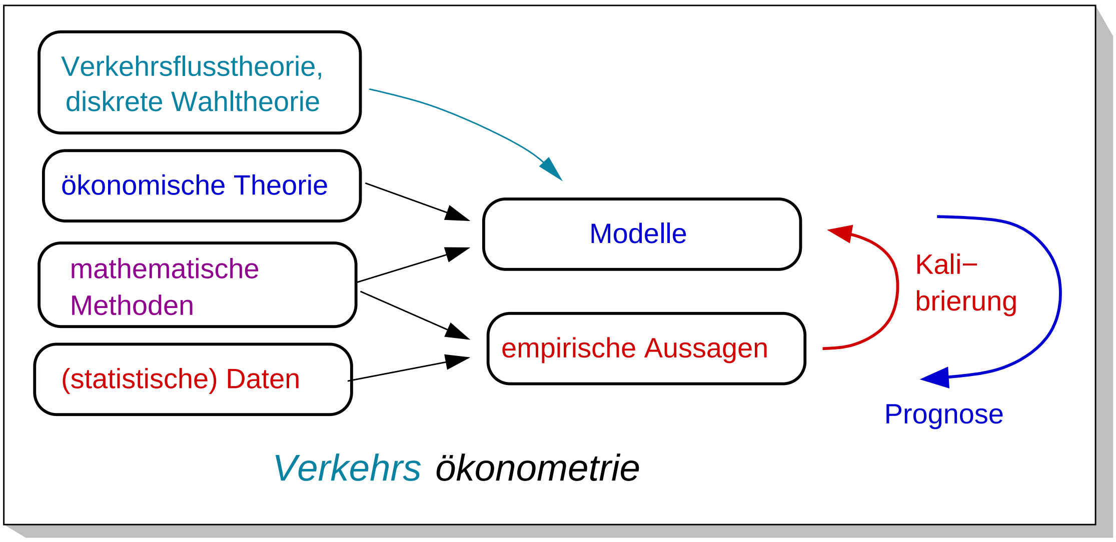
Dr. Martin Treiber

Wintersemester 2015/16

Inhaltsverzeichnis

1 Allgemeines

Das Wort **Ökonometrie** hat seinen Ursprung von der Ökonomie (Wirtschaft) und dem lateinischem Wort *metiri*, welches "messen" bedeutet. Damit beinhaltet Ökonometrie nicht nur *quantitative* wirtschaftliche Theorien und Konzepte, sondern vor allem deren **empirische** (also auf Messung, Beobachtung und Erfahrung beruhende) *Überprüfung*. Bei den durch Messung und Beobachtung gewonnen Daten handelt es sich fast immer um **statistische Daten**, also um eine Beschreibung von *Massenphänomenen*. Damit spielt neben der quantitativen Formulierung wirtschaftlicher Theorien durch **mathematische Modelle** auch die Auswertung der Daten durch **statistische Modelle** eine Rolle. Das Zusammenspiel der verschiedenen Elemente der Ökonometrie wird in folgendem Flussdiagramm deutlich.



In der Verkehrsökonometrie werden diese allgemeinen Konzepte auf den Verkehrssektor spezialisiert:

> Die **Verkehrsökonometrie** umfasst die Gesamtheit mathematischer Modelle und statistischer Verfahren, um auf einer empirischen Grundlage den Verkehr und seine volkswirtschaftlichen Auswirkungen quantitativ zu analysieren und zu prognostizieren.

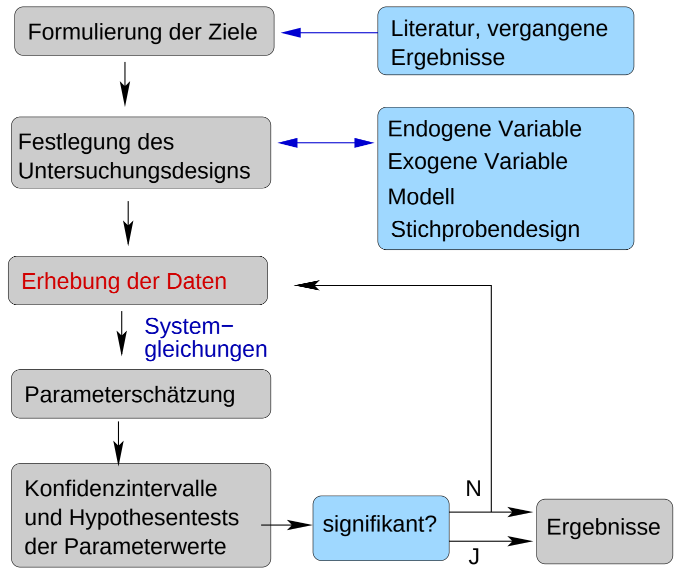


Abbildung 1.1: Allgemeiner Ablauf einer ökonometrischen Untersuchung.

1.1 Ablauf einer ökonometrischen Untersuchung

Im Prinzip läuft eine ökonometrische Untersuchung nach dem in Abb. 1.1 dargestellten Schema ab. Nach Formulierung des Untersuchungsziels und dem Abgleich mit vergangenen Untersuchungen kommt die wichtige Phase der Festlegung des Erhebungsdesigns und des Modells. Diese beiden Komponenten müssen zusammenpassen. Will man beispielsweise ein Logit-Modell anwenden, muss der Erhebungsfragebogen derart gestaltet sein, dass er eine endliche Anzahl von Alternativen enthält, welche (i) *alle* möglichen Antworten enthalten (Vollständigkeit), (ii) jeweils nur eine wählbar ist (Exklusivität) und (iii) wesentlich voneinander verschieden sind (sonst müsste man ein verallgemeinertes Entscheidungsmodell anwenden). Ferner müssen die Anzahl der Variablen und ihre Skalierung (siehe unten) zur Modellspezifikation kompatibel sein.

Nach der dann folgenden Durchführung der Datenerhebung (Abschnitt 3) werden die Modellparameter anhand der Daten geschätzt, also ihre wahrscheinlichsten Werte und ihr Streubereich festgestellt. Die eigentliche Erkenntnis der ökonometrischen Untersuchung ergibt sich neben der deskriptiven Auswertung der Erhebung vor allem anhand der Schätzung der Modellparameter. Damit lassen sich unter Anderem wichtige von unwichtigen Einflussfaktoren trennen sowie Maßzahlen verkehrswirtschaftlicher Sachverhalte gewinnen wie eine Bevorzugung bestimmter Verkehrsmittel ("bei gleichem Zeit- und Kostenaufwand fahre ich lieber mit dem ÖV als dem Auto") oder implizite Zeitwerte ("eine Stunde weniger Reisezeit ist mir 20 Euro wert").

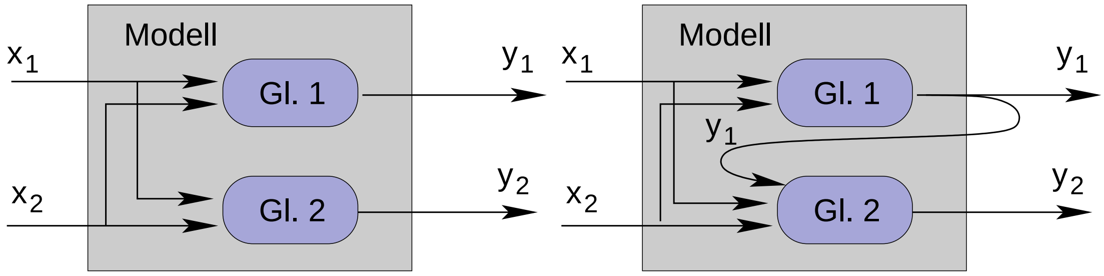


Abbildung 1.2: Flussdiagramm der exogenen und endogenen Variablen des allgemeinen ökonometrischen Modells (1.1). Gezeigt ist der Fall für je zwei exogene und endogene Variablen, $M = k = 2$ ohne (links) und mit (rechts) Kopplung der endogenen Variablen. Das Modell (eckige Box) besteht hier also aus zwei Gleichungen, welche zwei Eingleichungsmodellen (links) bzw. einem Mehrgleichungsmodell (rechts) entsprechen.

1.2 Modelle und Variablen der Verkehrsökonometrie

Da die Ökonometrie wirtschaftliche Zusammenhänge quantitativ beschreibt, besteht ihre Grundlage aus mathematischen Gleichungen. Generell kann man ökonometrische Modelle (und mathematische Modelle allgemein) als eine mathematische Abbildung betrachten, die gewisse Eingabegrößen auf Ausgabegrößen abbildet, wobei die Abbildung an den konkreten Sachverhalt durch Modellparameter angepasst wird (Abb. 1.3). Wir werden im Folgenden Gleichungen folgender Struktur betrachten (vgl. Abb. 1.2):

$$Y_k = f_k(x_1, ..., x_m, .., x_M, \beta_0, ..., \beta_j, ..., \beta_J) + \epsilon_k = f_k(\boldsymbol{x}, \boldsymbol{\beta}) + \epsilon_k. \tag{1.1}$$

1.2.1 Endogene Variablen

Die Y_k sind die **erklärten Variablen**. Eine solche Variable wird bisweilen auch als **Explanandum** (lat. "zu erklärende" bzw. erklärte Variable), als **endogene Variable** (Griechisch: von innen heraus, also aus dem Modell kommend) oder gemäß allgemeinen mathematischen Sprachgebrauch als **abhängige Variable** bezeichnet. In der Systemtheorie entsprechen die Y_k dem *Output* des jeweiligen Modells. Da viele Modelle stochastischer Natur sind, also Zufallsgrößen ϵ_k enthalten (vgl. Abschnitt 1.2.4), sind die Y_k im Allgemeinen Zufallsvariablen und werden, der allgemeinen Konvention der Statistik entsprechend, "groß" geschrieben.

Je nach der Anzahl und Kopplung der endogenen Variablen werden verschiedene Modellklassen unterschieden: von

- **Eingleichungsmodelle**: Nur eine endogene Variable $Y_1 = Y$,
- **Mehrgleichungsmodelle**: Mehrere endogene Variable Y_k, die miteinander gekoppelt sind (vgl. Abb. 1.2 rechts),

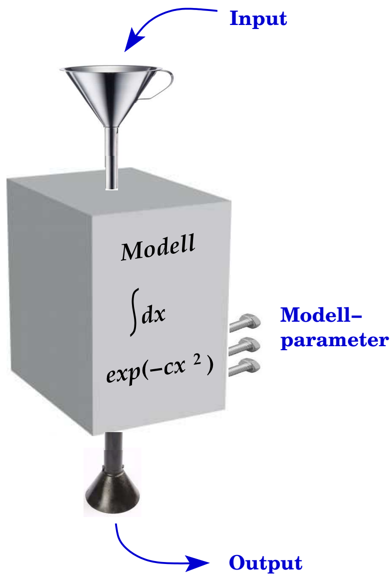


Abbildung 1.3: "Flussdiagramm" eines mathematischen Modells in einer anfderen Perspektive: Oben kommen die exogenen Variablen hinein, unten die endogenen heraus. Die Box stellt das eigentliche Modell dar und die "Stellschreauben" die Modellparameter.

- **Mehreren Eingleichungsmodelle**: Mehrere ungekoppelte endogene Variable wie im Flussdiagramm 1.2 links. Dies entspricht mehreren separat zu lösenden Eingleichungsmodellen.

Beispiel Verkehrsmittelwahl: Die Y_k sind die absoluten Häufigkeiten der gewählten Verkehrsmittel, z.B. Y_1=Zahl der Kfz-Fahrten, Y_2=Zahl der ÖPNV-Fahrten und Y_3, Y_4 = Zahl der Wege mit dem Rad bzw. zu Fuß.

1.2.2 Exogene Variablen

Die Größen x_m, $m = 1, \ldots, M$ sind die **erklärenden Variablen**, welche auch als **Explanans** (lat. "erklärend"), **exogene Variable** (Griechisch: von außen kommend) oder allgemein als **unabhängige Variable** bezeichnet werden. In der Systemtheorie entsprechen die x_m dem *Input* in das jeweilige Modell. In der Regel werden die x_m als deterministisch betrachtet und deshalb, der allgemeinen Konvention entsprechend, "klein" geschrieben.[1]

Je anch Anzahl der exogenen Variablen definiert man

[1] In der Praxis werden sowohl die x_m als auch die Y_k aus Erhebungen gewonnen, beide Variablenkategorien sind damit prinzipiell stochastischer Natur. Man kann aber ohne Einschränkung der Allgemeinheit die Stochastizität der Input-Größen auf die Zufallsanteile ϵ_k verschieben, da die Modelle ja stochastische *Funktionen* darstellen, also die exogenen Variablen sowieso variabel sind. Dies ist sogar geboten, da andernfalls die Stochastizität überbestimmt ist. Näheres in Abschnitt 1.2.4.

- **univariate Modelle**: $M = 1$,
- **multivariate Modelle**[2]: $M > 1$.

Beispiel Verkehrsmittelwahl: Die x_m sind die Reisezeiten und andere, die Verkehrsmittelwahl beeinflussende Faktoren (Kosten, Zuverlässigkeit etc)

1.2.3 Modellparameter

Die Größen β_j, $j = 0, \ldots, J$ sind die **Modellparameter**. Im Gegensatz zu den exogenen und endogenen Variablen, welche sich bei jedem System bzw. bei jeder Anwendung des Modells ändern, sollten die Modellparameter nach ihrer **Schätzung** bzw. **Modellkalibrierung** und einer solchermaßen bewirkten Anpassung des Modellverhaltens an die Wirklichkeit (vgl. Abschnitt 2.4) für alle Anwendungen einen konstanten Wert besitzen. *Diese Eigenschaft ist ein entscheidendes Kriterium für die Qualität eines Modells und seiner Aussage- und Prognosekraft!*

Beispiel Verkehrsmittelwahl: Modellparameter bestimmen die relative Gewichtung einzelner Einflussfaktoren (z.B. für jedes Verkehrsmittel den Wert der Zeit in €/h) sowie eine a-Priori-Bevorzugung bestimmter Verkehrsmittel gegenüber anderen (in € oder Minuten) bei eigentlich gleichem Nutzwert.

1.2.4 Zufallsanteile

Die ϵ_k beschreiben unbestimmte oder **nicht erklärte Anteile** welche meist durch Zufallsvariablen mit Erwartungswert $E(\epsilon_k) = 0$ modelliert werden[3] Insbesondere gibt es folgende Gründe für die Notwendigkeit eines Zufallsanteils ϵ_k:

- Das Modell kann nicht alle Einflussfaktoren berücksichtigen. Wichtig ist aber, zumindest alle Faktoren mit *systematischen Einfluss* zu berücksichtigen, da das Modell sonst **fehlspezifiziert** ist und falsche Aussagen liefert (vgl. Abschnitt 2.3).
- Die zur Modellkalibrierung verwendeten Messwerte sind fehlerbehaftet. Dieser Anteil von ϵ gibt direkt die kumulierten Messfehler wider.
- Der Mensch ist keine Maschine. Der entsprechende Anteil von ϵ spiegelt die Abweichung des in Wirklichkeit oft nichtrationalem Verhaltens vom Idealbild des *Homo Oeconomicus* wider.

Korrekt spezifizierte und eindeutig definierte Zufallsterme spielen eine wesentliche Rolle bei der Parameterschätzung (Abschnitt 2.4)

[2]Streng genommen sind die Termini Einfach- und Mehrfachregression in diesem Zusammenhang nicht korrekt, da sie eine spezielle Schätzmethode (die Regression) und nicht die Modelle als solches bezeichnen. Dies wird aber in der Literatur häufig ungenauerweise gleichgesetzt.

[3]Man sagt, dass "Zufallselemente nichts anderes als das Eingeständnis von Unwissen" seien. Die Bedingung $E(\epsilon_k) = 0$ ist keine Einschränkung, da man einen eventuellen Erwartungswert auf den deterministischen Teil verlagern kann.

Beispiel Verkehrsmittelwahl: ϵ_k entspricht dem *Zufallsnutzen*. Dieser sorgt dafür, dass – wie in der Wirklichkeit – mit geringerer Wahrscheinlichkeit auch ein Verkehrsmittel mit nichtmaximalen deterministischen, d.h. modellierten Nutzen gewählt wird.

1.2.5 Modellfunktionen

Die Funktionen $f_k(...)$ charakterisieren schließlich das eigentliche ökonometrische Modell. Wie bei den Parametern ist es für die Güte und Prognosekraft eines Modells wichtig, dass es nach seiner Entwicklung unverändert auf neue Instanzen des zu beschreibenden Sachverhalts mit unterschiedlichen exogenen Variablen angewandt werden kann.

Beispiel Verkehrsmittelwahl bei N Entscheidungen insgesamt:

$$P_k = \text{Prob(Altern } k \text{ gewählt)} = \frac{e^{V_k(\boldsymbol{x},\boldsymbol{\beta})}}{\sum_{k'} e^{V_{k'}(\boldsymbol{x},\boldsymbol{\beta})}}, \tag{1.2}$$

$$Y_k \sim \text{Multinom}(N, \boldsymbol{P}) \text{ - Verteilung.} \tag{1.3}$$

Dies ist ein verkettetes Modell (s. weiter unten) mit 2 Stufen:

- 1. Stufe: Deterministisches Modell (1.2) für die Auswahlwahrscheinlichkeiten
- 2. Stufe: Stochastisches Modell (1.3), welches die Zufallsvariablen "Zahl der Entscheidungen für Alternative k" mit Hilfe der Multinomialverteilung modelliert.

Die Nutzenfunktionen $V_k(\boldsymbol{x}, \boldsymbol{\beta})$ enthalten die Einflussfaktoren (exogenen Variablen) $\boldsymbol{x}$ (Reisezeit, Kosten,...) und die Modellparameter $\boldsymbol{\beta} = \{\beta_j\}$ (relative Gewichtungen und Bevorzugungen), während die zweite Stufe keine weiteren exogenen Variablen oder Parameter enthält. Näheres dazu im Abschnitt 4.

1.2.6 Das Prinzip der Sparsamkeit

Bei der Modellformulierung gilt das Sparsamkeitsprinzip (*principle of parsimony*), welches man, frei nach Einstein, folgendermaßen formulieren kann:

> Das Modell sollte so einfach wie möglich sein, aber nicht einfacher. Es sollte so wenig Parameter wie möglich enthalten, aber nicht weniger:
> *Mach' es so einfach wie möglich, aber nicht einfacher* (Albert Einstein)

1.3 Modellgleichungen

Die das ökonometrische Modell beschreibenden mathematischen Gleichungen kann man nach zwei Kriterien klassifizieren: Bezüglich der *Inhaltlichen* bzw. *semantischen* Struktur und bezüglich der *mathematischen* Struktur.

1.3.1 Inhaltliche Struktur

Hier unterscheidet man nach zwei Kategorien:

- Das **ökonometrische Modell im engeren Sinn** wie Gl. (1.1) beschreibt den *abstrakten funktionalen Zusammenhang*. Manchmal wird noch zwischen dem *allgemeinen Modell* (nicht spezifizierte Werte der Modellparameter) und dem *geschätzten Modell* (nach der Parameterschätzung) unterschieden. Im einfachstmöglichen nichttrivialen Fall eines linearen univariaten Eingleichungsmodells (eine exogene und eine endogene Variable) lautet das allgemeine ökonometrische Modell

$$Y(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x + \epsilon \tag{1.4}$$

  und das geschätzte Modell

$$\hat{Y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x \tag{1.5}$$

  Im *allgemeinen Modell* haben die Parameter feste, aber noch nicht bestimmte Werte und die Unsicherheit wird durch die Zufallsgröße ϵ ausgedrückt. Im *geschätzten Modell* gibt es keinen expliziten Zufallsterm mehr. Vielmehr sind die Zufallseinflüsse auf die geschätzten Parameterwerte $\hat{\beta}_0$ und $\hat{\beta}_1$ übergegangen, welche nun *ihrerseits* Zufallsvariablen darstellen.

  Im Falle eines linearen multivariaten Eingleichungsmodells (mehrerer exogene und eine endogene Variable) lautet das allgemeine ökonometrische Modell

$$Y(x) = \beta_0 + \sum_{m=1}^{M} \beta_m x_m + \epsilon \tag{1.6}$$

- Die **Systemgleichungen**, manchmal auch als **Messgleichungen** bezeichnet, beschreiben die *Anwendung* des ökonometrischen Modells auf konkrete Systeme bzw. auf die von diesen Systemen gemessenen Werte der exogenen und endogenen Variablen. Im Gegensatz zum abstrakten Modell hängen die Systemgleichungen vom konkreten System ab. Die zum ökonometrische Modell (1.4) gehörigen Systemgleichungen lauten

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n \tag{1.7}$$

  und werden auf n Datensätze $\{(x_i, y_i)\}$ angewandt. Die Systemgleichungen zum Mehrfachregressionsmodell (1.6) lauten

$$Y_i = \beta_0 + \sum_{m=1}^{M} \beta_m x_{mi} + \epsilon_i \tag{1.8}$$

  und werden auf n Datensätze $\{(x_{1i}, x_{2i}, ..., x_{Mi}, y_i)\}$ angewandt. Mit Hilfe der Systemgleichungen kann man die Modellparameter derart schätzen, dass eine Funktion $F(\{\epsilon_i\})$ der Modellierungsfehler ϵ_i, z.B. die Fehlerquadratsumme, minimal wird (Kapitel 2.4)

Merke: Wichtig ist es, bei einfach indizierten exogenen Variablen zu unterscheiden, ob es sich um ein abstraktes Mehrfachregressionsmodell der Art (1.6) oder um Systemgleichungen eines Einfachregressionsmodells der Art (1.7) handelt.

Verständnisfrage:
Warum sollte es bei den Systemgleichungen immer mehr Sätze von Messwerten geben als es der Zahl der Parameter entspricht ($n > J+1$)? Was passiert, wenn $n = J+1$ oder gar $n < J+1$?

1.3.2 Mathematische Struktur

Unterscheidung nach Linearität

Die Unterscheidung ist in Hinblick auf die Lösungsmethoden wichtig. Man kann folgende Kategorien unterscheiden:

- **Lineare Modelle**: Hier hängen die endogenen Variablen linear von den exogenen Variablen $\boldsymbol{x}$ und den Parametern $\boldsymbol{\beta}$ ab: Bei Eingleichungsmodellen haben sie also die Form
$$Y(x) = \beta_0 + \sum_{m=1}^{M} \beta_m x_m + \epsilon. \tag{1.9}$$
In vielen ökonometrischen Lehrbüchern werden ausschließlich lineare Modelle behandelt, in der Verkehrsökonometrie sind jedoch auch nichtlinieare Modelle wichtig.
Beispiel: Einfaches Modell für die mittlere Fahrleistung Y pro Person in einer Region:
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \tag{1.10}$$
mit den erklärenden Variablen
  - x_1: Mittleres Einkommen (Euro pro Jahr)
  - x_2: Treibstoffpreis (Euro/Liter)
  - x_3: Ausbau des Straßennetzes (Kilometer pro Einwohner)

  Als Aussagen lassen sich hier zum Beispiel Elastizitäten gewinnen wie
$$\epsilon_2 = \frac{x_2}{Y} \frac{\mathrm{d}Y}{\mathrm{d}x_2} = \frac{\beta_2 x_2}{Y}$$
Ein Wert $\epsilon_2 = -0.2$ sagt z.B. aus, dass bei einer Erhöhung der Treibstoffpreise um 10% nur 2% weniger Auto gefahren wird.

- **Reduzible nichtlineare Modelle**: Zwar hängt hier die endogene Variable nichtlinear von sowohl den exogenen Variablen $\boldsymbol{x}$ als auch den Parametern $\boldsymbol{\beta}$ ab, man kann die Modellgleichungen jedoch durch Transformation der exogenen und/oder endogenen Variablen in eine lineare Form bringen.

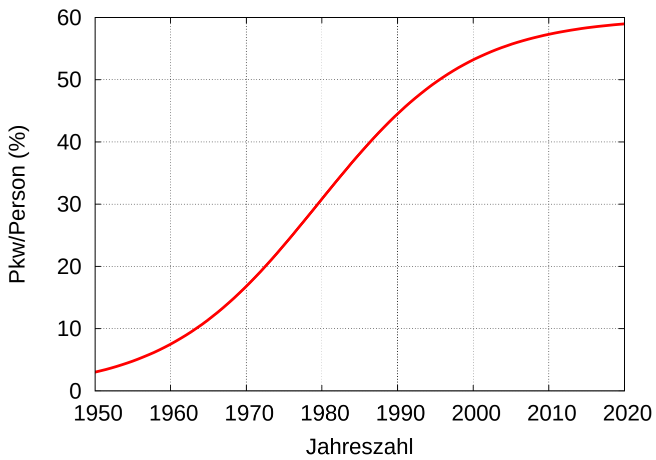


Abbildung 1.4: Graph des Modells (1.13).

Beispiel: Modell für das *unbeschränkte Wachstum* eines ökonomischen Prozesses (z.B. die anfänglichen Verkaufszahlen eines neu eingeführten Produkts). Mit der Zeit als exogenen Variablen lautet dieses

$$y(t) = y_0 e^{t/\tau+\epsilon} \tag{1.11}$$

Durch die Transformation $y = e^w$ und anschließende Logarithmierung wird dieses in ein lineares Eingleichungsmodell transformiert:

$$w(t) = \ln(y_0) + t/\tau + \epsilon = \beta_0 + \beta_1 t + \epsilon \tag{1.12}$$

Verständnisfrage:
Warum funktioniert dies bloß, wenn die Zufallsanteile im ursprünglichen Modell *multiplikativ* wirken?

- **Irreduzible nichtlineare Modelle**: Hier hängt die endogene Variable nichtlinear sowohl von den exogenen Variablen $\boldsymbol{x}$ als auch von den Parametern ab und es ist keine Transformation auf eine lineare Form möglich. Dies ist zwingend immer dann der Fall, wenn die endogenen Variablen *diskreter Natur* sind (z.B. Y = Zahl der Wege, die mit dem Rad zurückgelegt werden) oder wenn es sich gar um **qualitative** bzw. **nominalskalierte Variablen** handelt, z.B. Y = gewählter Beruf mit den Ausprägungen Maurer, Schreiner, Physiker etc. Aber auch im Bereich der stetigen Modelle gibt es manchmal die Notwendigkeit von nichtlinearen Modellen, wie im folgenden Beispiel.

Beispiel: Modell für das *Sättigungsverhalten* eines ökonomischen Prozesses (z.B. die Verkaufszahlen von Autos oder Mobiltelefonen seit Erfindung der jeweiligen Produkte) beschreiben will. Das klassische **Modell beschränkten Wachstums** mit der Zeit x als exogenen Variablen hat die Form (vgl. Abb. 1.4)

$$\hat{y}(x) = \frac{y_s}{1 + \left(\frac{y_s}{y_0} - 1\right) e^{-x/\tau}} \tag{1.13}$$

Verständnisfrage:

Diskutieren Sie die Bedeutung der Parameter τ, y_0 und y_s im Modell (1.13). Kann man das Modell auch in der Form $\hat{y}(x) = y_s/[1+e^{-(x-x_0)/\tau}]$ schreiben? Was ist dann die Bedeutung des neuen Parameters x_0 und wie hängt er mit den Parametern der Formulierung (1.13) zusammen?

- **Quasilineare** bzw. **parameterlineare Modelle**: In dieser Modellklasse sind die Modellgleichungen linear bezüglich der Parameter $\boldsymbol{\beta}$, aber nichtlinear bezüglich der exogenen Variablen $\boldsymbol{x}$. Solche Modelle kann man immer durch eine Transformation der exogenen Variablen linearisieren,[4] so dass sie *völlig äquivalent* zu den linearen Modellen sind.

Alle Methoden linearer Modelle sind identisch auch auf die transformierten quasilinearen Modelle anwendbar. Im weiteren Verlauf dieses Skriptes sind deshalb bei linearen Modellen immer auch quasilineare Modelle mit eingeschlossen.

Beispiel: Einfaches Modell für die mittlere PKW-Fahrleistung Y eines PKW-Besitzers pro Jahr:

$$Y(\boldsymbol{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon \tag{1.14}$$

mit den exogenen Variablen

- x_1: Mittleres Einkommen (Euro pro Jahr)
- x_2: Treibstoffpreis (Euro/Liter)

Mit den Transformationen

$$z_1 = x_1, \quad z_2 = x_2, \quad z_3 = x_1 x_2$$

[4] Im Gegensatz zu den reduziblen nichtlinearen Modellen gibt es hier auch keine Probleme mit den Zufallsanteilen.

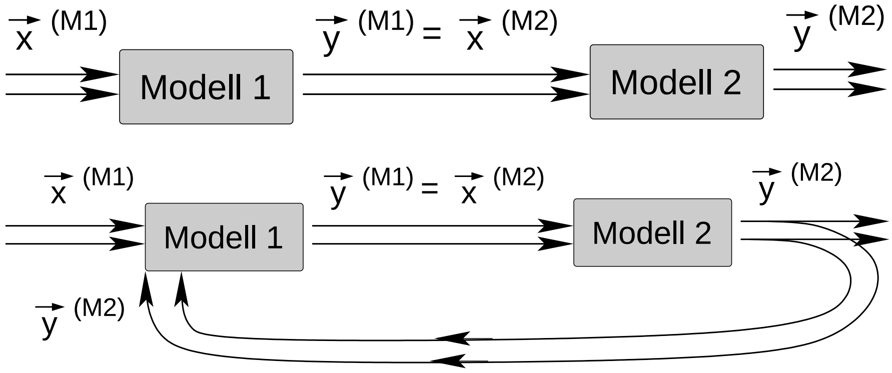


Abbildung 1.5: Flussdiagramm der exogenen und endogenen Variablen für verkette (oben) und rückgekoppelte (unten) Modell-Systeme. Das "Innenleben" der Modelle (also die je zwei Gleichungen, vgl. Abb. 1.2) ist nicht mehr gezeigt. Die Vektorpfeile symbolisieren den ganzen Satz jeweiliger endogener und exogener Variablen.

wird dieses Modell in ein lineares Modell mit *drei* exogenen Variablen transformiert:

$$Y(\boldsymbol{z}, \boldsymbol{\beta}) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \epsilon \tag{1.15}$$

Die zugehörigen Anstiegsparameter bedeuten

- β_1: Anstieg der Fahrleistung mit dem Einkommen (üblicherweise $\beta_1 > 0$)
- β_2: Preissensitivität der PKW-Nutzung (üblicherweise $\beta_2 < 0$)
- β_3: Reduktion der Preissensitivität mit dem Einkommen (üblicherweise $\beta_3 > 0$)

Verständnisfrage:
Machen Sie sich die Vorzeichen der drei Parameter in diesem Beispiel klar. Warum ist bei Berücksichtigung auch extrem hoher Einkommen ein irreduzibel nichtlineares Modell notwendig? Zeigen Sie dies anhand der dann nicht plausiblen Aussagen des Modells (1.15).

Unterscheidung nach weiteren mathematischen Kriterien

Weitere in der Ökonometrie relevante Unterscheidungsmerkmale sind

- Existenz eines Zufallsanteils: **Deterministische** *vs.* **stochastische** Modelle.

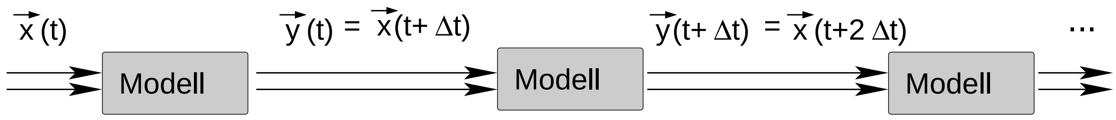


Abbildung 1.6: Flussdiagramm der Verkettung im Falle dynamischer Modelle. Das Modell selbst ist in der Regel in allen Schritten dasselbe, oft mit unveränderten Parametern (*autonomes* dynamisches Modell), manchmal mit von der Zeit abhängigen (*nichtautonomes* Modell).

- Zahl der exogenen Variablen: Eine bei univariaten und mehrere bei multivariaten Modellen.

- Zahl der endogenen Variablen: Eine bei Eingleichungsmodellen, mehrere bei Mehrgleichungsmodellen.

- Skalierung der endogenen Variablen: **Entscheidungsmodelle** haben diskrete bzw. qualitative/nominalskalierte endogene Variablen, während **kontinuierliche Modelle** stetige sowie verhältnisskalierte endogene Variable aufweisen. Diese beiden Kategorien bedingen grundsätzlich verschiedene Herangehensweisen bei der Modellierung, beispielsweise sind Entscheidungsmodelle immer nichtlinear.

  Hingegen ist die Skalierung der exogenen Variablen nicht so bedeutsam, da man qualitative bzw. nominalskalierte exogene Variable durch kardinalskalierte **Pseudovariablen** der Art 1=Maurer, 2=Schreiner, 3=Physiker etc ausdrücken kann.

- Existenz von Verkettung oder Rückkopplung (siehe Abb. 1.5 und Abb. 1.7):
  - Bei **verketteten Modellen** sind die endogenen Variablen einer Modellstufe die exogenen der nächsten. Wichtigstes Beispiel dafür ist das Vier-Stufen-Schema der klassischen Verkehrsplanung (Abb. 1.7).
  - Bei **rückgekoppelten Modellen** koppeln die endogenen Variablen eines Modells einer späteren Verkettungsstufe auf die exogenen Variablen eines Modells einer früheren Verkettungsstufe zurück. Beim Vier-Stufen-Modell der Abb. 1.7 können als Folge der Routenwahl einzelne Netzelemente überlastet, also verstaut werden. Die damit verbundenen Resiezeitverlängerungen beeinflussen wiederum die Ziel- und ggf. die Verkehrsmittelwahl.

- Dynamische *vs.* statische Modelle: **dynamische Modelle**, also Modelle mit expliziten dynamischen Zeitbezug, stellen einen Sonderfall der Verkettung im Zeitbereich dar: Die Entwicklung des aktuellen Zeitschrittes hängt von den Ergebnissen des vergangenen Zeitschrittes bzw. der vergangenen Zeitschritte ab (Abb. 1.6). Im Rahmen der Vorlesung werden wir einen einfachen Typ dieser Modelle kennen lernen, das **dynamische Input-Output-Modell** (Abschnitt 5).

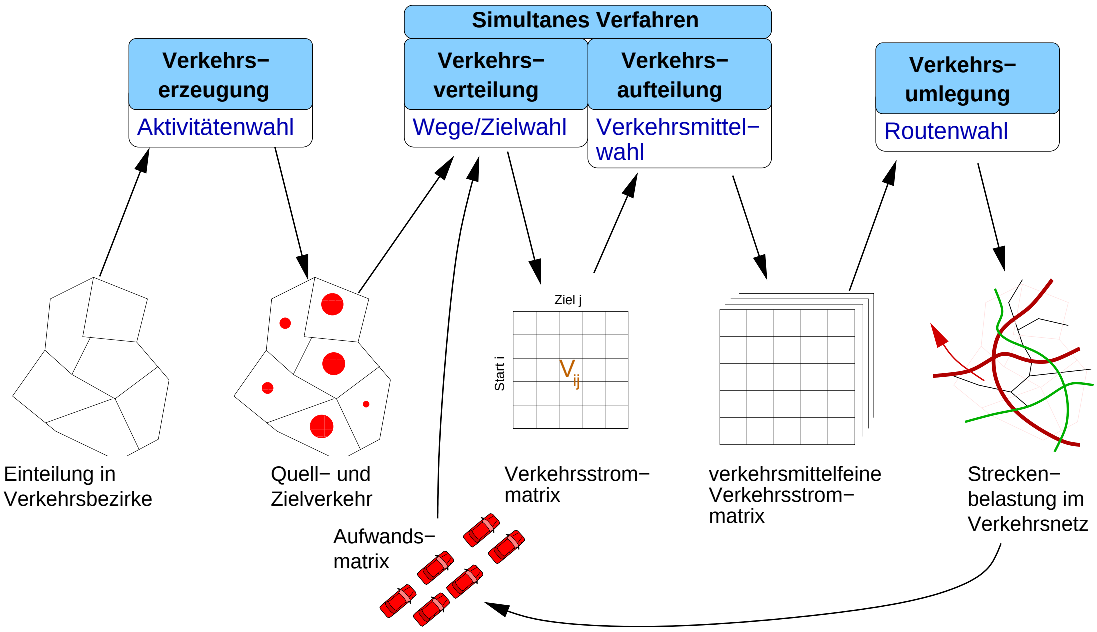


Abbildung 1.7: Das klassische Vierstufenmodell der Verkehrsplanung als Beispiel der Verkettung und Rückkopplung von Modellen. Die endogenen Variablen der ersten Modellstufe "Aktivitätenwahl", die im Verlauf eines Tages anfallenden Verkehrsströme von und nach jedem Bezirk (Quell- bzw. Zielsummen), spielen gleichzeitig die Rolle von exogenen Variablen bei der Modellierung der "Zielwahl". Die endogenen Variablen der Zielwahl (Quelle-Ziel-Relationen bzw. Wege zwischen jeweils zwei Bezirken) sind wiederum die exogenen Variablen der Verkehrsmittelwahl und in weiterer Folge der Routenwahl. Das Ergebnis der Routenwahl führt ggf. zu einer Überlastung einzelner Netzelemente und zu einer Erhöhung der Reisezeit auf einigen Relationen. Diese sogenante "Reisezeitmatrix" bzw. "Aufwandsmatrix" ist nun wiederum eine weitere exogene Variable der Zielwahl ("wenn der Weg von Bezirk i nach j immer verstopft ist, gehe ich vielleicht doch woanders einkaufen"), so dass die Routenwahl auf die Zielwahl rückkoppelt.

1.4 Ein Anwendungsbeispiel: Verkehrsmittelwahl

Zur Veranschaulichung der in diesem Kapitel beschriebenen Konzepte soll das folgende Beispiel der Modellierung einer Verkehrsmittelwahl dienen.

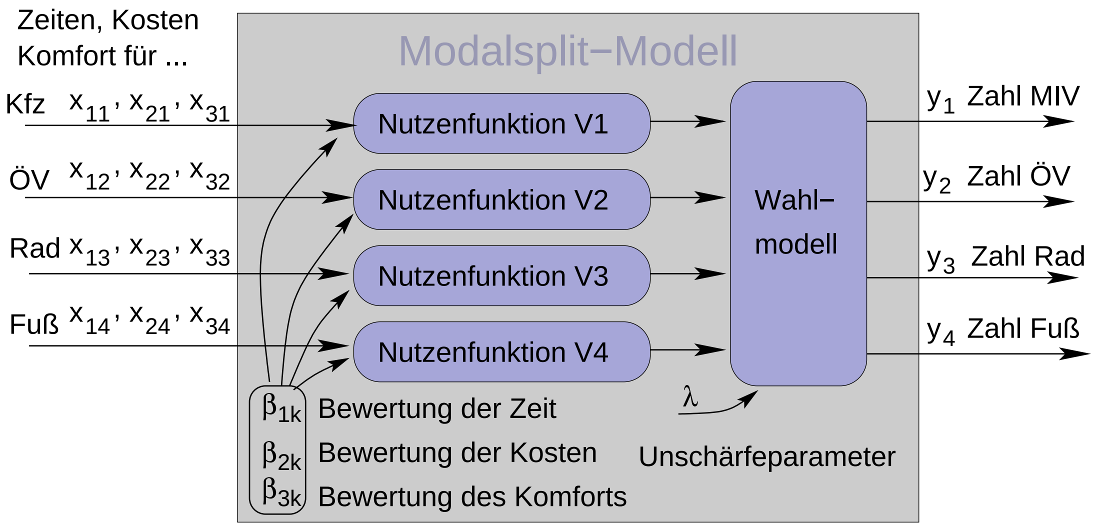


Abbildung 1.8: Beispiel eines zweistufigen ökonometrischen Wahlmodells. Zunächst werden in vier voneinander unabhängigen stetigen Eingleichungsmodellen die deterministischen Nutzenfunktionen V_k der verschiedenen Verkehrsmittel k modelliert. Diese wiederum spielen die Rolle von exogenen Variablen im eigentlichen Wahlmodell. Dieses diskrete Mehrgleichungsmodell liefert letztendlich als endogene Variablen die Zahlen Y_k der Entscheidungen für Verkehrsmittel k.

1.4.1 Modellstruktur

Das Beispiel enthält als Verkettung die wichtigsten beiden der in dieser Vorlesung behandelten Modellklassen (Abb. 1.8):

- In der ersten Stufe gibt es als Teilmodelle vier stetige Eingleichungsmodelle, welche aus den exogenen Variablen
  - Zeitbedarf X_{1k} des Verkehrsmittels k,
  - Kosten X_{2k} des Verkehrsmittels k,
  - Komfort X_{3k} des Verkehrsmittels k

  den deterministischen Nutzen V_k des jeweiligen Modus als endogene Variable ausgeben.

- Das eigentliche *Wahlmodell* enthält als verkettete exogene Variablen die deterministischen Nutzenfunktionen V_k der vier Alternativen und liefert die prognostizierten Entscheidungszahlen $\hat{Y}_k$ für jede Alternative. In seiner Eigenschaft als diskretes Mehrgleichungsmodell ist das Wahlmodell notwendigerweise nichtlinear.

1.4.2 Modellgleichungen und Parameter

Die vier Modelle für die Nutzenfunktionen werden als voneinander unabhängige lineare (oder auch quasilineare) Eingleichungsmodelle formuliert:

$$U_k(\boldsymbol{x}_k, \boldsymbol{\beta}_k) = \beta_{0k} + \sum_{m=1}^{3} \beta_{mk} x_{mk} + \epsilon_k \tag{1.16}$$

Hierbei bedeuten:[5]

- β_{0k}: Globaler Bonus des Verkehrsmodus k gegenüber Modus 1 (damit ist offensichtlich $\beta_{01} = 0$),
- β_{1k}: Zeitbewertung bei Verkehrsmodus k (beispielsweise kann man die Zeit zu Fuß unangenehmer empfinden als die Zeit im Kfz, dann gilt $0 > \beta_{11} > \beta_{14}$),
- β_{2k}: Geldbewertung bei Verkehrsmodus k,
- β_{3k}: Bewertung des Komforts bei Verkehrsmodus k.

Von diesen Nutzenfunktionen wird der deterministische Anteil $\hat{U}_k = V_k$ sowie die *Standardabweichung* β_0 des Zufallsnutzens an das eigentliche Wahlmodell weitergegeben. Bei insgesamt N vorzunehmenden Entscheidungen hat das einfachste und am weitesten verbreitete Wahlmodell, das Multinomial-Logit-Modell, die Modellgleichung

$$Y_k(\boldsymbol{V}, \beta_0) = N P_k(\boldsymbol{V}, \beta_0) = N \frac{e^{\beta_0 V_k}}{\sum_{k'} e^{\beta_0 V_{k'}}}. \tag{1.17}$$

1.4.3 Systemgleichungen

Da die verketteten intermediären Variablen V_k nicht messbar sind, muss man die Verkettung der Teilmodelle (1.16) und (1.17) als Black-Box betrachten. Damit gelangt man zu einem Modell der Form (1.3) mit messbaren exogenen Variablen x_{mk} und endogenen Variablen Y_k.

Um dieses Modell zu schätzen, benötigt man n Messungen jeweils aller exogenen und endogenen Variablen. Eine “Messung” ist hier der technische Ausdruck einer Befragung einer Person: Von jeder Person i müssen also folgende Werte erfragt werden:[6]

$$\{\underbrace{x_{mki}}_{\substack{\text{Eigenschaften}\\ \text{der Modi}}}, \underbrace{y_{ki}}_{\substack{\text{Zahl Entscheidungen}\\ \text{für Alternative } k}}\}$$

[5] Die Parametrisierung ist nicht vollständig, es fehlen z.B. die sozioökonomischen Variablen (siehe Kapitel 4). Als einführendes Beispiel dient das vereinfachte Modell aber hier seinen Zweck.

[6] Die Werte der exogenen Variablen haben einen Dreifachindex, da die Menge der exogenen Variablen bereits doppelt indiziert ist.

Nach einer erfolgreichen Befragung hat der Interviewer eine Tabelle, welche folgendermaßen aussehen kann:

Größe	Zeitbedarf Auto	Kosten Auto	Subj. Komfort Auto (1-6)	Kompl. Reisezeit ÖPNV	...	Wahlentsch. Auto	Wahlentsch. ÖPNV	...
Variable	x_{11i}	x_{21i}	x_{31i}	x_{12i}	...	y_{1i}	y_{2i}	...
Person 1	20 min	2.50 €	3	30 min	...	0	1	...
Person 2	11 min	2.00 €	1	20 min	...	1	0	...
Person 3	34 min	4.00 €	1	15 min	...	0	1	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Anhand der Tabelle und der Zahl N_i der Entscheidungen, welche Person i insgesamt trifft, werden die Systemgleichungen formuliert,

$$\hat{Y}_{ki} = N_i \frac{e^{V_k(\boldsymbol{x}_i, \boldsymbol{\beta})}}{\sum_{k'} e^{V_{k'}(\boldsymbol{x}_i, \boldsymbol{\beta})}}, \quad i = 1, \ldots, n. \tag{1.18}$$

Hierbei fasst der Vektor $\boldsymbol{x}_i = \{x_{11i}, \ldots, x_{34i}\}$ die für Person i relevanten exogenen Variablen zusammen. Anhand der Systemgleichungen wird der (für alle Personen gleiche!) Parametervektor $\boldsymbol{\beta}$ durch Vergleich der modellierten endogenen Variablen $\hat{Y}_{ki}$ mit den gemessenen (erfragten) endogenen Variablen Y_{ki} geschätzt.

Aggregierung

Grundsätzlich sind Modelle der Verkehrsmittelwahl **mikroskopisch**, betrachten also jede Entscheidung einzeln. Damit kann auch nur eine der endogenen Variablen y_k für Alternative k gleich 1 sein, während alle anderen $y_{k'}$, $k' \neq k$, gleich 0 sein müssen. In drei Fällen ist eine *Aggregierung*, also eine **makroskopische Hernagehensweise** möglich, sinnvoll oder sogar notwendig:[7]

- *personenbezogene Aggregierung:* Ein und dieselbe Person wird in vergleichbaren Situationen (gleiche exogene Variable) mehrfach befragt. Geht es z.B. um die Verkehrsmittelwahl zur Arbeit und wurden die Entscheidungen den letzten zwei Wochen erhoben, kann man die 10 Entscheidungen zusammenfassen.

- *situationsbezogene Aggregierung:* Wählt man die exogenen Variablen so, dass sie für mehrere Personen *und alle Alternativen* etwa gleiche Werte besitzen, kann man diese Personen zu Gruppen bzw. Klassen zusammenfassen. In der Praxis funktioniert dies meist nur, wenn man nur Variable wählt, welche für alle Alternativen

[7] Hier dient die Aggregierung hauptsächlich zu "pädagogischen" Zwecken, da sie die Darstellung vereinfacht. In Realität wird man bei Vorliegen mikroskopischen Datenmaterials immer das mikroskopische Vorgehen wählen.

etwa gleich sind, also entweder sozioökonomische Variable (Alter, Geschlecht, Autoverfügbarkeit), externe Variable (Wetter) oder spezielle generische Variable wie die Entfernung.

- *datenerzwungene Aggregierung:* Liegen Daten nur aggregiert vor, kann man natürlich nur makroskopisch vorgehen.

Beispiel: mikroskopische Daten, welche eine Aggregierung zulassen (situationsbezogene Aggregierung)

	Entf. Fuß	Entf. Rad	Entf. ÖPNV	Entf. MIV	Wahl-entsch. Fuß	Wahl-entsch. Rad	Wahl-entsch. ÖPNV	Wahl-entsch. MIV
Variable	x_{11}	x_{12}	x_{13}	x_{14}	y_1	y_2	y_3	y_4
Person 1	2 km	2 km	3 km	3 km	0	1	0	0
Person 2	4 km	4 km	4.5 km	∞	0	0	1	0
Person 3	3 km	3 km	∞	3.5 km	0	0	0	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Aggregierung dieser Daten in Entfernungsklassen

Entfernungsklasse	Wahl-entsch. Fuß	Wahl-entsch. Rad	Wahl-entsch. ÖPNV	Wahl-entsch. MIV	Pers. in Klasse
x_1 = Klassenmitte	y_1	y_2	y_3	y_4	n_i
0-1 km	5	1	0	0	6
1-2 km	2	0	5	0	7
2-3 km	0	2	11	0	13
3-5 km	0	0	7	1	8
5-10 km	0	0	5	0	5
10-20 km	0	0	1	1	2
> 20 km	0	0	3	0	3
$\sum$	7	3	32	2	44

1.5 Verwendete Symbole

Y	Abhängige bzw. erklärte bzw. endogene Variable. Bei Mehrgleichungsmodellen bekommen diese den Index k.
x_m, $m = 1, \ldots, M$	Unabhängige bzw. erklärende bzw. exogene Variablen
$\boldsymbol{x}$	Gesamtheit der exogenen Variablen, als Spaltenvektor geschrieben
y_i, x_{mi}, $i = 1, \ldots, n$	Wert der abhängigen bzw. der m-ten unabhängigen Variablen bei der i-ten Messung bzw. dem i-ten Element der Stichprobe.
β_0	Achsabschnitt (engl. *intercept*)
β_j, $j = 1, \ldots, J$	Lineare Anstiegsparameter
$\boldsymbol{\beta}$	Gesamtheit der Modellparameter, als Spaltenvektor geschrieben
$\hat{\beta}_j$, $\hat{Y}$	Das "Dach" ist ein Symbol für geschätzte Größen, also geschätzte Parameterwerte, exogene Variable des geschátzten Modells, etc.
ϵ	Ein additiver i.i.d. Zufallsterm mit Erwartungswert 0 und der Varianz σ^2
z_m, w	Transformierte exogene und endogene Variablen (mit dem Ziel, ein lineares Modell in den transformierten Variablen zu erhalten)
V_k	Deterministische Nutzenfunktion der Alternative k bei Wahlmodellen
U_k	Gesamtnutzen $V_k + \epsilon_k$ der Alternative k bei Wahlmodellen
n	Zahl der Stichprobenelemente bzw. Messungen bzw. befragte Personen (Laufindex $i = 1, \ldots, n$)
N	Zahl der Entscheidungen bei Wahlmodellen, $N = \sum_k Y_k$
N_i	Zahl der Entscheidungen von Person i bei Wahlmodellen, $N_i = \sum_k y_{ki}$