

**Zusammenfassungsbox 1. Das Wichtigste zur Wahrscheinlichkeit****A. Definitionen**

Der Wahrscheinlichkeitsraum  $\langle \Omega, \mathfrak{A}, P \rangle$  repräsentiert das betrachtete Zufallsexperiment. Dabei sind:

- $\Omega$  die Menge der möglichen Ergebnisse
- $\mathfrak{A}$  die Menge der möglichen Ereignisse  $A \subset \Omega$  und
- $P: \mathfrak{A} \rightarrow [0, 1]$  ein Wahrscheinlichkeitsmaß auf  $\mathfrak{A}$

Axiome von Kolmogoroff:

- $P(A) \geq 0$ , für alle  $A \in \mathfrak{A}$
- wenn  $A_1, A_2, \dots$  eine Folge von paarweise disjunkten Mengen  $A_i \in \mathfrak{A}$  ist, dann gilt:  $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$
- $P(\Omega) = 1$

**B. Rechenregeln und Sätze**

- wenn  $B \subset A$ , dann  $P(A \setminus B) = P(A) - P(B)$  und  $P(A) \geq P(B)$
- $P(A \setminus B) = P(A) - P(A \cap B)$
- für  $\bar{A} := \Omega \setminus A$  gilt:  $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Wahrscheinlichkeitsraum  $\langle \Omega, \mathfrak{A}, P \rangle$

Nichtnegativität  
Additivität

Normierung

Eigenschaften von  $P$

$$P(C) = P(A \setminus B) = P(A) - P(A \cap B) = P(A) - P(\{\omega_2\}) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}.$$

Für das Ereignis  $D = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_6\}$  („der Würfel zeigt nicht 5 Augen“) gilt:

$$\begin{aligned} P(D) &= P(A \cup B) \\ &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(\{\omega_2\}) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}. \end{aligned}$$

**Beispiel 7.** Beim zweimaligen Werfen einer fairen Münze ist die Wahrscheinlichkeit, zweimal „Kopf“ zu werfen,  $P(\{\langle K, K \rangle\}) = 1/4$  (Auch hier setzen wir eine „faire“ Münze voraus.) Diese Wahrscheinlichkeit gilt auch für die drei anderen Elementarereignisse. Für das Ereignis  $\{\langle K, Z \rangle, \langle Z, K \rangle\}$ , dass genau einmal Kopf fällt, gilt:

$$\begin{aligned} P(\{\langle K, Z \rangle\} \cup \{\langle Z, K \rangle\}) &= P(\{\langle K, Z \rangle\}) + P(\{\langle Z, K \rangle\}) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned}$$

Entsprechend gilt für das Ereignis  $\{\langle K, K \rangle, \langle Z, Z \rangle\}$ , dass beide Male Kopf oder beide Male Zahl fällt:

$$\begin{aligned} P(\{\langle K, K \rangle\} \cup \{\langle Z, Z \rangle\}) &= P(\{\langle K, K \rangle\}) + P(\{\langle Z, Z \rangle\}) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned}$$

Zweifacher Münzwurf

*Anwendung des  
Satzes der totalen Wahrscheinlichkeit  
und des Bayes-Theorems  
in der Latent-class-Analyse*

#### Anwendungsbox 2

In der Latent-class-Analyse (s. z. B. Rost, 1996) spielen der Satz der totalen Wahrscheinlichkeit und das Bayes-Theorem eine zentrale Rolle. Die in Gleichung (3.11) vorkommenden Mengen  $A_i$  werden dort als latente Klassen interpretiert und das Ereignis  $B$  ist ein Antwortmuster auf eine Menge von Items mit zwei oder mehreren Antwortkategorien. Die Wahrscheinlichkeit  $P(B)$  kann in solchen Anwendungen direkt über relative Häufigkeiten geschätzt werden, wohingegen die auf der rechten Seite der Gleichung (3.11) vorkommenden theoretischen Wahrscheinlichkeiten mit relativ komplizierten Algorithmen mit entsprechenden Programmen (s. z. B. WinMira, von Davier, 1997) geschätzt werden können. Die Wahrscheinlichkeiten geben Aufschluss zum einen über die relativen Größen der latenten Klassen  $[P(A_i)]$  und zum anderen über die bedingten Wahrscheinlichkeiten für das jeweilige Antwortmuster  $[P(B | A_i)]$ . Man kann das Ereignis  $B$  auch als eine Antwort in einer bestimmten Kategorie eines jeweils betrachteten Items interpretieren und die bedingten Wahrscheinlichkeiten schätzen, in dieser Kategorie zu antworten, wenn man zur Klasse  $A_i$  gehört. (s. hierzu auch Beispiel 3).

Das Bayes-Theorem (3.12) kann dann verwendet werden, um aus einem bestimmten Antwortmuster  $B$  die Wahrscheinlichkeit der Zugehörigkeit zu jeder der Klassen  $A_i$  zu berechnen. Die ist die Voraussetzung der Klassifizierung der Personen in die einzelnen Klassen.

*Wahrscheinlichkeit  
der Schnittmenge  
von drei Ereignissen*

Setzen wir nun Gleichung (3.6) ein, erhalten wir

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2). \quad (3.8)$$

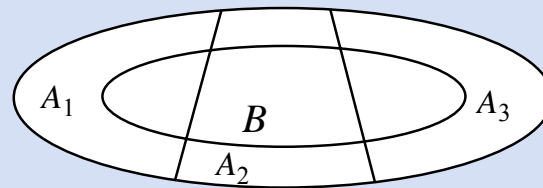
Setzen wir dieses Verfahren fort, erhalten wir die im folgenden Theorem formulierte Gleichung für die Schnittmenge  $A_1 \cap \dots \cap A_n$  beliebig vieler Ereignisse.

*Wahrscheinlichkeit  
der Schnittmenge  
beliebig vieler Ereignisse*

**Theorem 2.** Seien  $\langle \Omega, \mathfrak{A}, P \rangle$  ein Wahrscheinlichkeitsraum und  $A_1, \dots, A_n \in \mathfrak{A}$  Ereignisse. Ist  $P(A_1 \cap \dots \cap A_{n-1}) > 0$ , dann gilt:

$$\begin{aligned} &P(A_1 \cap \dots \cap A_n) \\ &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1}). \end{aligned} \quad (3.9)$$

Dieses Theorem spielt z. B. in der Theorie graphischer Modelle (s. z. B. Spirtes, Glymour & Scheines, 1993; Pearl, 2000) aber auch in der Item-response-Theorie (s. z. B. Boomsma, van Duijn & Snijders, 2001; Fischer & Molenaar, 1995; Rost, 1996; Steyer & Eid, 2001) und bei den log-linearen Modellen (s. z. B. Agresti, 1990, 1996; Andreß, Hagenaars & Kühnel, 1997; Pruscha, 1996) eine große Rolle. Dort werden die auf der rechten Seite vorkommenden Terme mit Hilfe bestimmter Modellannahmen (meist bestimmte bedingte Unabhängigkeitsannahmen) berechnet und man kann dann prüfen, ob diese Modellannahmen mit der empirischen Schätzung für die Wahrscheinlichkeit auf der linken Seite der Gleichung hinreichend übereinstimmen.



**Abbildung 2.** Venn-Diagramm zur Veranschaulichung des Satzes von der totalen Wahrscheinlichkeit.

### Anwendung in „Bayes-Statistik“

Benötigt werden dafür die bedingten Wahrscheinlichkeiten  $P(D|H_i)$  der Daten gegeben die jeweilige Hypothese  $H_i$  sowie die unbedingte Wahrscheinlichkeit der jeweiligen Hypothese  $H_i$ . Damit sind die so genannten „*A-priori-Wahrscheinlichkeiten*“ der Hypothesen gemeint, also deren Wahrscheinlichkeiten, bevor die Daten erhoben wurden. In der Regel kann man diese nur über subjektive Einschätzungen gewinnen. In dieser Anwendung charakterisieren Wahrscheinlichkeiten also nicht mehr nur Gesetzmäßigkeiten in einem Zufallsexperiment, sondern auch „subjektive Sicherheiten“.

### Latent-class-Modell

**Beispiel 3.** Angenommen, eine Population von Personen kann in zwei Klassen aufgeteilt werden, nämlich in diejenigen, die ein bestimmtes Wissensgebiet beherrschen und diejenigen, die es *nicht beherrschen*. In der ersten Klasse seien 80% aller Personen, in der zweiten Klasse 20%. Betrachten wir nun das Zufallsexperiment, zufällig (d. h. jede Person hat die gleiche Wahrscheinlichkeit gezogen zu werden) eine Person aus der Population auszuwählen und sie einige Aufgaben bearbeiten zu lassen. Dann ist die Wahrscheinlichkeit, dass die gezogene Person derjenigen Klasse angehört, die das Wissensgebiet beherrscht,  $P(A) = 0.8$  und der Klasse, die es nicht beherrscht  $P(\bar{A}) = 0.2$ . Für eine bestimmte Aufgabe, die sich auf das genannte Wissensgebiet bezieht, gelten die Lösungswahrscheinlichkeiten  $P(B_1|A) = 0.9$  für Personen der ersten Klasse und  $P(B_1|\bar{A}) = 0.3$  für Personen der zweiten Klasse. Wir können nun die (unbedingte) Wahrscheinlichkeit, dass eine zufällig aus der Population ausgewählte Person die Aufgabe löst, nach dem Satz von der totalen Wahrscheinlichkeit ausrechnen. Wenden wir ihn auf das vorliegende Beispiel an, gilt:  $P(B_1) = P(B_1|A) \cdot P(A) + P(B_1|\bar{A}) \cdot P(\bar{A}) = 0.9 \cdot 0.8 + 0.3 \cdot 0.2 = 0.72 + 0.06 = 0.78$ .

Für eine zweite Aufgabe gelten die Lösungswahrscheinlichkeiten  $P(B_2|A) = 0.95$  für Personen der ersten Klasse und  $P(B_2|\bar{A}) = 0.1$  für Personen der zweiten Klasse. Die Lösungswahrscheinlichkeiten mögen aber ausschließlich von der Zugehörigkeit zu einer Klasse abhängen. Gegeben eine Person gelte also Unabhängigkeit der Ereignisse „Aufgabe 1 wird gelöst“ und „Aufgabe 2 wird gelöst“. Mit dieser Information können wir nun z. B. die Wahrscheinlichkeit  $P(B_1 \cap B_2|A)$  ausrechnen, dass eine Person in der ersten Klasse beide Aufgaben löst:  $P(B_1|A) \cdot P(B_2|A) = 0.9 \cdot 0.95 = 0.855$ . Die Wahrscheinlichkeit, dass eine Person in der zweiten Klasse beide Aufgaben löst, beträgt  $P(B_1|\bar{A}) \cdot P(B_2|\bar{A}) = 0.3 \cdot 0.1 = 0.03$ .

Wie groß ist nun die (unbedingte) Wahrscheinlichkeit, dass eine zufällig aus der Population ausgewählte Person beide Aufgaben löst? Hier können wir wiederum den Satz von der totalen Wahrscheinlichkeit anwenden:  $P(B_1 \cap B_2) = P(B_1 \cap B_2|A) \cdot P(A) + P(B_1 \cap B_2|\bar{A}) \cdot P(\bar{A}) = 0.855 \cdot 0.8 + 0.03 \cdot 0.2 = 0.684 + 0.006 = 0.69$ .

Regelbox 1. Die wichtigsten Sätze zur bedingten Wahrscheinlichkeit

$$P(A \cap B) = P(A) \cdot P(B | A), \quad \text{falls } P(A) > 0$$

Einfache Produktregel

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdot \dots \cdot P(A_n | A_1 \cap \dots \cap A_{n-1}),$$
  
falls  $P(A_n | A_1 \cap \dots \cap A_{n-1}) > 0$ .

Allgemeine Produktregel  
(Allgemeiner Faktorisierungssatz)

Sind  $A_1, \dots, A_i, \dots, A_n$  paarweise disjunkt,  $B \subset A_1 \cup \dots \cup A_i \cup \dots \cup A_n$ , und  $P(A_i) > 0$ , dann gelten die folgenden beiden Sätze:

$$P(B) = P(B \cap A_1) + \dots + P(B \cap A_n)$$
  
$$= P(B | A_1) \cdot P(A_1) + \dots + P(B | A_n) \cdot P(A_n)$$

Satz  
der totalen Wahrscheinlichkeit

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{P(B | A_1) \cdot P(A_1) + \dots + P(B | A_n) \cdot P(A_n)}$$

Bayes-Theorem

3.6 Zusammenfassende Bemerkungen

In diesem Kapitel wurden die Begriffe „bedingte Wahrscheinlichkeit“ und „Unabhängigkeit“ von Ereignissen eingeführt. Betrachtet man mehr als zwei Ereignisse, so können diese nicht nur paarweise, sondern auch tripelweise etc. unabhängig oder eben auch paarweise, tripelweise, etc. abhängig sein. Zur genaueren Beschreibung der Art und Stärke der stochastischen Abhängigkeit zwischen Ereignissen ist der Begriff der *bedingten Wahrscheinlichkeit* von grundlegender Bedeutung. Die wichtigsten Theoreme zur bedingten Wahrscheinlichkeit sind der Faktorisierungssatz, der Satz der totalen Wahrscheinlichkeit und das Bayes-Theorem. Diese sind nicht nur grundlegend für die Wahrscheinlichkeitstheorie, sondern auch für alle empirischen Wissenschaften, in denen nach Abhängigkeiten zwischen Ereignissen—und darauf aufbauend—zwischen Zufallsvariablen gefragt wird (s. dazu die nächsten beiden Kapitel).

Fragen

- F1. Worin besteht der Unterschied zwischen „Unabhängigkeit“ und „Disjunktheit“ zweier Ereignisse  $A$  und  $B$ ?

leicht
- F2. Wie ist der Ausdruck  $P(A | B)$  zu lesen?

leicht
- F3. Welchem Flächenanteil im Venn-Diagramm der Abbildung 1 entspricht die bedingte Wahrscheinlichkeit  $P(A | B)$ ?

leicht
- F4. Welche Eigenschaften teilt der Begriff der bedingten Wahrscheinlichkeit mit dem der (unbedingten) Wahrscheinlichkeit?

mittel
- F5. Warum wird die Unabhängigkeit zweier Ereignisse über die Gleichung (3.4) und nicht über die Gleichungen (3.3) definiert?

leicht
- F6. Warum genügt es nicht, für die Definition der Unabhängigkeit von drei Ereignissen  $A$ ,  $B$ , und  $C$  ihre paarweise Unabhängigkeit zu postulieren?

mittel
- F7. Warum wird im Faktorisierungssatz  $P(A_1 \cap \dots \cap A_{n-1}) > 0$  vorausgesetzt?

mittel
- F8. Wozu dient die Befragungstechnik der „randomisierten Antwort“?

mittel

als reellwertige Zufallsvariable einführen, die die Lösungszeit für das vorgelegte Rätsel repräsentiert. Diese könnte dann eine wichtige Rolle spielen, wenn man die interindividuellen Unterschiede zwischen den drei Personen betrachten will.

### 4.3 Zufallsvariable

Die Begriffe *Zufallsvariable* und *Verteilung* wurden oben nur auf informelle Weise eingeführt. Daher sollen nun die formalen und allgemeinen Definitionen nachgeholt werden.

Beim Beispiel 1 in Abschnitt 4.2 (Würfelwurf) wurde als Zweck der Einführung von Zufallsvariablen die damit verbundene Informationsreduktion oder Vereinfachung genannt. In formaler Hinsicht zeigt sich diese Vereinfachung wie folgt: Anstatt des relativ komplexen Wahrscheinlichkeitsraums  $\langle \Omega, \mathfrak{P}(\Omega), P \rangle$ , bei dem die Menge  $\Omega$  der möglichen Ergebnisse sechs Elemente und die Menge  $\mathfrak{A}$  der möglichen Ereignisse, die Potenzmenge  $\mathfrak{P}(\Omega)$  bereits  $2^6 = 64$  Elemente hat, betrachtet man nun einen neuen Wahrscheinlichkeitsraum mit der Menge  $\Omega' = \{0, 1\}$  der möglichen Ergebnisse und der Menge  $\mathfrak{A}' = \mathfrak{P}(\Omega') = \{\{0\}, \{1\}, \Omega', \emptyset\}$  der möglichen Ereignisse. Damit wird also die Komplexität erheblich reduziert. Wie bereits betont, haben Zufallsvariablen, deren Werte Zahlen sind, in vielen Fällen darüber hinaus noch den Vorteil, dass man mit ihrer Hilfe relativ einfach Gesetzmäßigkeiten beschreiben kann, die das Zufallsexperiment charakterisieren.

Zur allgemeinen Definition einer Zufallsvariablen  $X: \Omega \rightarrow \Omega'$  benötigen wir neben dem W-Raum  $\langle \Omega, \mathfrak{A}, P \rangle$  und einem beliebigen Wertebereich  $\Omega'$ , in dem  $X$  ihre Werte annimmt, auch eine  $\sigma$ -Algebra  $\mathfrak{A}'$  auf  $\Omega'$ . Bei dem oben aufgeführten Beispiel war  $\mathfrak{A}'$  die Potenzmenge von  $\Omega' = \{0, 1\}$ . Außerdem greifen wir auf den Begriff eines Urbilds zurück. Zur Erinnerung: Das *Urbild*  $X^{-1}(A')$  von  $A'$  unter  $X$  ist das Ereignis  $\{\omega \in \Omega: X(\omega) \in A'\}$ , dass  $X$  einen Wert in der Menge  $A'$  annimmt:

$$X^{-1}(A') := \{\omega \in \Omega: X(\omega) \in A'\}. \quad (4.6)$$

**Definition 1.** Seien  $\langle \Omega, \mathfrak{A}, P \rangle$  ein W-Raum,  $\Omega'$  eine Menge und  $\mathfrak{A}'$  eine  $\sigma$ -Algebra auf  $\Omega'$ . Eine Abbildung  $X: \Omega \rightarrow \Omega'$  heißt *Zufallsvariable*, wenn für das Urbild  $X^{-1}(A')$  von jedem  $A' \in \mathfrak{A}'$  gilt:  $X^{-1}(A') \in \mathfrak{A}$ . Die Menge aller Urbilder  $X^{-1}(A')$  heißt die von  $X$  erzeugte  $\sigma$ -Algebra.

Die Bedingung, dass die Urbilder  $X^{-1}(A')$  Elemente der zugrunde liegenden  $\sigma$ -Algebra  $\mathfrak{A}$  sind, stellt sicher, dass die mit der Zufallsvariablen  $X$  darstellbaren Ereignisse  $X^{-1}(A')$  eine Wahrscheinlichkeit haben, nämlich  $P[X^{-1}(A')]$ , denn das W-Maß  $P$  weist definitionsgemäß allen Elementen aus  $\mathfrak{A}$  ihre Wahrscheinlichkeit zu. Die von  $X$  erzeugte  $\sigma$ -Algebra kann man auch als die durch  $X$  darstellbaren Ereignisse ansehen. In der folgenden Definition werden verschiedene Arten von Zufallsvariablen unterschieden.

Neue Ergebnismenge  $\Omega'$

Neue Ereignismenge  $\mathfrak{A}'$

Urbild  $X^{-1}(A')$

Zufallsvariable

Von  $X$  erzeugte  $\sigma$ -Algebra



*Gewichtung bei zusammengesetzten  
Zufallsvariablen (composite measures)*

### Anwendungsbox 3

Manchmal definiert man neue Zufallsvariablen als einfache oder auch gewichtete Summen bekannter Zufallsvariablen. Ein Beispiel ist der *sozio-ökonomische Status*, in dessen Definition sowohl Bildung als auch Einkommen eingehen. Bei dieser Art der Definition neuer Variablen muss man bei der Gewichtung bedenken, dass nicht nur die explizit verwendeten Gewichte eine Rolle spielen, sondern auch die Varianzen und Kovarianzen der beteiligten Ausgangsvariablen. In Übung 9 findet sich ein Beispiel, in dem zwar gleiche Gewichte bei der Summenbildung verwendet werden, in dem aber trotzdem der erste der beiden Summanden zu .9950 mit der Summe korreliert, der zweite der beiden Summanden aber nur zu .0995, und das trotz gleicher expliziter Gewichtung. Dies liegt daran, dass der erste Summand eine weitaus größere Varianz hat als der zweite.

rierten Gewichten multiplizierten Varianzen plus der mit dem zweifachen Produkt der Gewichte multiplizierten „Kovarianz“ der beiden Variablen.<sup>1</sup> Aus dieser Regel folgt, dass die Varianz einer Summenvariablen nur dann gleich der Summe der Varianzen ist, wenn die beteiligten Variablen *unkorreliert* sind, wenn also  $Cov(X_1, X_2) = 0$  gilt. Für die Differenz  $X_1 - X_2$  *unkorrelierter* numerische Zufallsvariablen folgt aus Regel (v) (mit  $\alpha_1 = 1$  und  $\alpha_2 = -1$ ):

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2), \quad \text{falls } Cov(X_1, X_2) = 0. \quad (5.5)$$

Die Varianz einer Differenzvariablen ist also gleich der *Summe* der Varianzen, *falls die beiden Variablen unkorreliert* sind. Andernfalls und allgemein gilt:

*Varianz einer Differenzvariablen*

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2) - 2Cov(X_1, X_2). \quad (5.6)$$

## 5.3 Kovarianz und Korrelation

Zunächst werden wir die Begriffe „Kovarianz“ und „Korrelation“ einführen und ein kleines Beispiel betrachten. Beide Begriffe, Kovarianz und Korrelation, sind Kennwerte, die das Ausmaß oder die Stärke des *Kovariierens*, also des „*Zusammen-Variierens*“ zweier Zufallsvariablen charakterisieren.

### 5.3.1 Kovarianz

*Grundidee*

Zwei Zufallsvariablen  $X$  und  $Y$  *variieren positiv zusammen* oder *kovariieren positiv*, wenn bei einer positiven Abweichung der Variablen  $X$  von ihrem Erwartungswert  $E(X)$  auch mit einer positiven Abweichung der Variablen  $Y$  von ihrem Erwartungswert  $E(Y)$  zu rechnen ist, und umge-

<sup>1</sup> Dieser Begriff der Kovarianz wird zwar erst im nächsten Abschnitt eingeführt. Der Vollständigkeit halber sei die Regel jedoch schon hier aufgeführt.

Regressor  $X: \Omega \rightarrow \Omega'$  darstellen:  $E(Y|X) = f(X)$ . Dabei weist  $f$  jedem Wert  $x_i$  von  $X$  den Wert  $E(Y|X=x_i)$  zu (s. Abb. 2).

### 6.3.2 Allgemeine Definition der Regression

Aus didaktischen Gründen wurde die obige Definition einer Regression zunächst auf den Fall beschränkt, in dem  $X$  eine Zufallsvariable auf  $\langle \Omega, \mathfrak{A}, P \rangle$  ist, mit endlich vielen Werten  $x_1, \dots, x_n$ , die jeweils eine Wahrscheinlichkeit  $P(X = x_i) > 0$  haben. Die allgemeine Definition gilt nun auch für kontinuierliche Regressanden und Regressoren. Insbesondere schließt sie auch den Fall  $P(X = x_i) = 0$  ein. Man beachte, dass die in den Regelboxen 2 und 3 aufgeführten Rechenregeln für Regressionen und ihr Residuum auch für den hier definierten allgemeinen Fall gelten.

**Definition 5.** Seien  $\langle \Omega, \mathfrak{A}, P \rangle$  ein Wahrscheinlichkeitsraum,  $Y$  eine numerische Zufallsvariable auf  $\langle \Omega, \mathfrak{A}, P \rangle$  mit endlichem Erwartungswert,  $X: \Omega \rightarrow \Omega'$  eine Zufallsvariable auf  $\langle \Omega, \mathfrak{A}, P \rangle$ ,  $\mathfrak{A}'$  eine  $\sigma$ -Algebra auf  $\Omega'$ ,  $\overline{\mathfrak{B}}$  die Borelsche  $\sigma$ -Algebra auf  $\overline{\mathbb{R}}$  und  $X^{-1}(\mathfrak{A}') := \{X^{-1}(A') : A' \in \mathfrak{A}'\}$  die Urbild- $\sigma$ -Algebra von  $X$ . Dann heißt jede numerische Zufallsvariable  $Z: \Omega \rightarrow \overline{\mathbb{R}}$  auf  $\langle \Omega, \mathfrak{A}, P \rangle$  *bedingte Erwartung* oder *Regression* von  $Y$  auf  $X$ , falls sie die folgenden Bedingungen erfüllt:

- (a)  $Z^{-1}(B) \in X^{-1}(\mathfrak{A}')$ , für alle  $B \in \overline{\mathfrak{B}}$ ;
- (b)  $E(I_C \cdot Z) = E(I_C \cdot Y)$ , für alle  $C \in X^{-1}(\mathfrak{A}')$ .

Anstelle von  $Z$  schreibt man für eine Regression von  $Y$  auf  $X$  meist  $E(Y|X)$ .

Wenn wir im Folgenden also von einer Regression  $E(Y|X)$  sprechen, dann wird vorausgesetzt, dass alle oben aufgeführten Bedingungen erfüllt sind. Gemäß Bedingung (a) müssen die Urbilder  $Z^{-1}(B)$  aller Elemente  $B$  der erweiterten Borelschen  $\sigma$ -Algebra  $\overline{\mathfrak{B}}$  Elemente der Urbild- $\sigma$ -Algebra von  $X$  sein. Dies stellt sicher, dass eine Regression  $Z$  von  $Y$  auf  $X$  eine Funktion von  $X$  ist, d. h. dass es eine Funktion  $f: \Omega' \rightarrow \overline{\mathbb{R}}$  gibt derart, dass  $Z$  als Komposition  $f(X)$  von  $f$  mit  $X$  dargestellt werden kann. Die Bedingung (b) dagegen garantiert, dass die bedingte Erwartung  $Z$  den Wert  $Z(\omega) = E(Y|X=x)$  annimmt, falls  $X$  den Wert  $X(\omega) = x$  annimmt und  $x$  eine Wahrscheinlichkeit  $P(X = x) > 0$  hat.

Durch die obige Definition wird die bedingte Erwartung oder Regression nicht völlig eindeutig definiert, insbesondere dann nicht, wenn  $X$  ein kontinuierlicher Regressor ist, dessen Werte ja die Wahrscheinlichkeit  $P(X = x) = 0$  haben. Dies hat zur Folge, dass es dann unterschiedliche „Versionen“ der Regression gibt, die jedoch „fast sicher“, d. h. mit Wahrscheinlichkeit 1, identisch sind. Weiter sei angemerkt, dass der bedingte Erwartungswert  $E(Y|X=x)$  im allgemeinen Fall als Wert der Faktorisierung  $f$  der bedingten Erwartung  $E(Y|X) = f(X)$  definiert ist. Aussagen über die so definierten bedingten Erwartungswerte gelten dann im allgemeinen nur für  $P^X$ -fast alle  $x$ . Zur weiteren Vertiefung sei der mathematisch interessierte Leser z. B. auf Bauer (2002) verwiesen.

**Anwendungsbox 1.**

Über die Internet-Adresse <http://www.wahrscheinlichkeit-und-regression.de> finden Sie ein Programm, mit dem Sie das in 8.1 erwähnte Experiment selbst durchführen können. Dort ist auch beschrieben, wie man das Programm ausführt und welche Voraussetzungen auf Ihrem PC vorhanden sein müssen, und wie man diese herstellen kann, um das Experiment mit diesem Programm durchzuführen. Sie können das Programm direkt aus dem Internet starten oder es lokal speichern.

*Experiment:  
Baldwin-Figuren*

das Potenzgesetz auch auf Reize mit Kontexten verallgemeinerbar ist, wie dies z. B. Bredenkamp (1982, 1984a, 1984b) postuliert hat.

Um eine konkrete Anschauung dieser Fragestellung zu erlangen, können Sie wieder selbst ein kleines Experiment durchführen (s. A-Box 1). Dazu machen Sie einen ähnlichen Versuch wie im letzten Kapitel, allerdings mit dem Unterschied, dass die Reizlinie in einen bestimmten Kontext eingebettet ist, nämlich in zwei Quadrate. Die resultierenden Objekte nennt man Baldwin-Figuren. Bei diesen Figuren treten systematische Täuschungen, d. h. Abweichungen der Länge der Urteilslinien von der Länge der Reizlinien auf (vgl. hierzu Abb. 1).

Bevor wir diese Anwendung weiter verfolgen, werden in den nächsten Abschnitten weitere regressionstheoretische Konzepte eingeführt. Dazu gehören der Begriff der linearen Quasi-Regression und verschiedene Parametrisierungen nichtlinearer Regressionen. Dabei werden die Begriffe zur Verfügung gestellt, die uns dann erlauben werden, die inhaltlichen Fragen zur Baldwin-Täuschung zu beantworten.

## 8.2 Lineare Quasi-Regression

*Lineare Quasi-Regression  
oder „lineare Regression 2. Art“  
oder „lineare Kleinst-Quadrat-  
Regression“*

Ein Begriff, der leicht mit dem Begriff der linearen Regression verwechselt werden kann, ist der der *linearen Quasi-Regression* [oder „lineare Regression 2. Art“ (Müller, 1975) oder „lineare Kleinst-Quadrat-Regression“]. Dieser Begriff wurde in Kapitel 5 schon eingeführt, um die Art der Abhängigkeit anzugeben, deren Stärke durch eine Korrelation beschrieben wird. Dieser Begriff ist aber auch zur Formulierung der Hypothese hilfreich, dass eine Regression  $E(Y | X)$  linear in  $X$  ist.

**Definition 1.** Seien  $X$  und  $Y$  numerische Zufallsvariablen mit endlichen Erwartungswerten und Varianzen auf dem gleichen Wahrscheinlichkeitsraum. Als *lineare Quasi-Regression* wird diejenige lineare Funktion  $Q(Y | X) := \alpha_0 + \alpha_1 \cdot X$  von  $X$  bezeichnet, für deren Residuum

$$v := Y - (\alpha_0 + \alpha_1 \cdot X), \quad \alpha_0, \alpha_1 \in \mathbb{R}, \quad (8.1)$$

die folgenden beiden Gleichungen gelten:

$$E(v) = 0, \quad (8.2)$$

$$\text{Cov}(v, X) = 0. \quad (8.3)$$



Prüfung der Linearitätshypothese

Kenngröße für das Ausmaß der Nichtlinearität:  $R^2_{Y|X} - Q^2_{Y|X}$

Nullhypothese

Signifikanztest

Voraussetzungen

Anwendungsbox 3. Prüfung der Linearität einer Regression II

Sie haben begleitend zum letzten Kapitel ein kleines Experiment durchgeführt, in dem Linienlänge *ohne* Kontexte zu beurteilen waren. In diesem Kapitel wurden Sie dazu angeleitet, ein zweites Experiment durchzuführen, in dem Linienlängen zu beurteilen waren, die *mit* einem Kontext dargeboten wurden. Für die von Ihnen erzeugten Daten aus diesen beiden Experimenten liegen jeweils vier verschiedene Serienreizlinien und damit Werte des Regressors vor und für jede Reizlinie haben Sie in jedem der beiden Experimente einige Urteilslinien erzeugt. Sie können nun für jeden der beiden Datensätze nach Gleichung (8.16) die vier Indikatorvariablen für die Werte des Regressors berechnen und mit einem Programm zur multiplen linearen Regression die Regression der logarithmierten Urteilslinie auf diese vier Indikatorvariablen berechnen. Dabei ist zu beachten, dass Sie eine multiple lineare Regression *ohne die allgemeine Konstante* berechnen lassen, die ja in Gleichung (8.16) nicht vorkommt. Auf diese Weise erhalten Sie Schätzungen für die Parameter der Gleichung (8.16) und eine Schätzung des Determinationskoeffizienten. Mit dem gleichen Programm können Sie die lineare Quasi-Regression analysieren und erhalten dabei Schätzungen für die Koeffizienten der Gleichung (8.6) und des Quasi-Determinationskoeffizienten. Nun können Sie einen Signifikanztest für die Nullhypothese

$$H_0: R^2_{Y|X} - Q^2_{Y|X} = 0$$

durchführen, dass die Regression linear ist. Diese geschieht über die Teststatistik:

$$F = \frac{(\hat{R}^2_{Y|X} - \hat{Q}^2_{Y|X})/(n - 2)}{(1 - \hat{R}^2_{Y|X})/(N - n)}.$$

Dabei sind  $n$  die Anzahl der Parameter in der saturierten Parametrisierung (hier: die 4 verschiedenen Reizlinien), 2 die Anzahl der Parameter in der eingeschränkten, linearen Quasi-Regression und  $N$  die Stichprobengröße (hier: die Gesamtzahl der Urteilslinien). Diese Teststatistik ist mit  $n - 2$  Zähler- und  $N - n$  Nennerfreiheitsgraden  $F$ -verteilt. Dabei setzen wir allerdings folgendes voraus: (a) die Unabhängigkeit der Fehlervariablen untereinander, d. h. der logarithmierten Abweichungen der Urteilslinien von ihrem bedingten Erwartungswert gegeben die jeweilige Länge der Serienreizlinie, (b) die Gleichheit der Fehlervarianzen zwischen den vier Bedingungen (d. h. Längen der Reizlinien) und (c) die Normalverteilung der logarithmierten Urteilslinien innerhalb jeder der vier Serienreize.

Den Determinationskoeffizienten für das saturierte Modell kann man alternativ auch über eine einfaktorielle Varianzanalyse mit vier Gruppen berechnen. Dabei erspart man sich die Bildung der Indikatorvariablen.

Wenn das Stevenssche Potenzgesetz gilt, dürfte bei den Daten des 1. Experiments der Test nicht signifikant werden. Auch beim zweiten Experiment sollte dies so sein, wenn Bredenkamp (1982, 1984a, 1984b) recht hätte. Ich denke allerdings, dass die Nullhypothese hier nicht zutrifft (s. Erdfelder & Steyer, 1984). Ob Sie allerdings ebenfalls zu diesem Schluss kommen, hängt nicht zuletzt auch von der Stichprobengröße ab, d. h. davon, wie viele Urteilslinien Sie insgesamt erzeugt haben, da mit der Zahl der Urteilslinien (der Stichprobengröße) auch die Teststärke steigt.

Wenn die Hypothese der Linearität der Regression für das 2. Experiment verworfen wird, stellt sich natürlich die Frage, von welchem Typ diese Regression dann ist. Sie können diese Frage nach dem gleichen Prinzip wie oben untersuchen. An die Stelle der linearen Quasi-Regression tritt dann eben eine quadratische, kubische Quasi-Regression oder auch eine ganz andere Funktion.

von  $X$ . Auf diese Weise entstehen Zellen, innerhalb derer man die verschiedenen Beobachtungen des Regressanden  $Y$  anordnen kann.

Diese Voraussetzung ist keineswegs immer erfüllt. So ist z. B. auch die einfache quadratische Regression  $E(Y | X)$  linear in  $(X, X^2)$ , wenn gilt:

$$E(Y | X) = \beta_0 + \beta_1 X + \beta_2 X^2. \quad (9.4)$$

*Einfache  
quadratische Regression*

Bei der Betrachtung der bedingten Regressionen setzen wir also immer voraus, dass es keine Funktion  $f(X)$  gibt mit  $Z = f(X)$  und auch keine Funktion  $f(Z)$  mit  $X = f(Z)$ . Nur dann kann bei Konstanthaltung des einen Regressors der andere variieren.

Die wichtigste Konsequenz der Gleichung (9.3) ist, dass jede bedingte Regression  $E_{Z=z}(Y | X)$  von  $Y$  auf  $X$  für einen beliebigen festen Wert  $z$  von  $Z$  eine *lineare Funktion* von  $X$  ist und dass die Steigungskoeffizienten für alle Werte  $z$  von  $Z$  gleich sind, und umgekehrt, dass die bedingte Regression  $E_{X=x}(Y | Z)$  von  $Y$  auf  $Z$  für einen beliebigen festen Wert  $x$  von  $X$  eine *lineare Funktion* von  $Z$  ist.<sup>2</sup> Dies ist der Schlüssel zum Verständnis der zweifachen linearen Regression.

*$E_{Z=z}(Y | X)$  als lineare  
Funktion von  $X$  bei beliebigem  
festen Wert von  $Z$*

*Steigungskoeffizienten für alle  
Werte  $z$  von  $Z$  sind gleich*

Diese bedingten Regressionen sind ganz normale Regressionen, nur dass sie sich auf das bedingte Wahrscheinlichkeitsmaß  $P_{Z=z}$  beziehen. Hat jedes Ereignis  $\{Z = z\} := \{\omega \in \Omega : Z(\omega) = z\}$  eine Wahrscheinlichkeit  $P(\{Z = z\}) > 0$ , so ist das bedingte Wahrscheinlichkeitsmaß  $P_{Z=z}$  auf der zugrunde gelegten  $\sigma$ -Algebra  $\mathfrak{A}$  definiert durch:

$$P_{Z=z}(A) := P(A \cap \{Z = z\}) / P(\{Z = z\}), \quad \text{für alle } A \in \mathfrak{A}.^3 \quad (9.5)$$

*Bedingtes  
Wahrscheinlichkeitsmaß*

Repräsentiert  $Z$  beispielsweise das Geschlecht mit  $z_1 = \text{männlich}$  und  $z_2 = \text{weiblich}$ , dann ist  $E_{Z=z_1}(Y | X)$  eine ganz normale einfache Regression innerhalb der Gruppe der männlichen und  $E_{Z=z_2}(Y | X)$  eine ganz normale einfache Regression innerhalb der Gruppe der weiblichen Personen.

Demnach folgt für einen beliebigen festen Wert  $z$  von  $Z$  aus Gleichung (9.3):

$$E_{Z=z}(Y | X) = (\beta_0 + \beta_2 z) + \beta_1 X, \quad (9.6)$$

*Bedingte lineare Regression*

also eine *bedingte lineare* Regression von  $Y$  auf  $X$  gegeben  $Z = z$ , und zwar mit einem Steigungskoeffizienten  $\beta_1$ , der für alle Werte  $z$  von  $Z$  gleich ist, und dem Ordinatenabschnitt  $\beta_0 + \beta_2 z$ , der für verschiedene Werte  $z$  von  $Z$  verschieden sein kann. Repräsentiert  $Z$  beispielsweise das Geschlecht mit  $Z = 0$  (für männlich) und  $Z = 1$  (für weiblich), dann wird mit Gleichung (9.6) formuliert, dass innerhalb einer Geschlechtsgruppe eine *lineare* Regression vorliegt und dass die Regressionen in beiden Geschlechtsgruppen den gleichen Steigungskoeffizienten  $\beta_1$  haben. Die Ordinatenabschnitte sind dann  $\beta_0$  bei den Männern und  $\beta_0 + \beta_2$  bei den Frauen.

<sup>2</sup> Genau genommen, für fast alle Werte  $z$  von  $Z$  bzw. fast alle Werte  $x$  von  $X$ .

<sup>3</sup> Zu den allgemeineren Bedingungen, unter denen  $P_{Z=z}$  definiert ist, siehe Gänssler und Stute (1977, S. 193 ff).

Kann man zwar die Gültigkeit der Gleichung (9.3) voraussetzen, nicht jedoch Gleichung (9.26), so gilt Gleichung (9.9) für den Determinationskoeffizienten  $R_{Y|X,Z}^2$  (s. Übung 7). Demnach muss also im Fall korrelierter Regressoren auch deren Kovarianz bei der Berechnung der Varianz von  $E(Y|X, Z)$  berücksichtigt werden, jedenfalls dann, wenn die beiden partiellen Regressionskoeffizienten ungleich null sind. Ist mindestens einer von beiden gleich null, so vereinfacht sich Gleichung (9.9) entsprechend.

### 9.3.2 Bedingung für die Linearität der Regression

Kann man nicht voraussetzen, dass  $Z$  von  $X$  regressiv unabhängig ist [s. Gl. (9.26)], und gilt stattdessen

$$E(Z|X) = \gamma_0 + \gamma_1 X, \quad (9.36)$$

so folgt dennoch Gleichung (9.27) und der Koeffizient  $\alpha_1$  der Gleichung (9.27) lässt sich aus den Koeffizienten der Gleichungen (9.3) und (9.36) berechnen:

$$E(Y|X) = (\beta_0 + \beta_2 \gamma_0) + (\beta_1 + \beta_2 \gamma_1) X. \quad (9.37)$$

Demnach ist, wenn man die Gleichungen (9.3) und (9.36) voraussetzt,  $E(Y|X)$  eine *lineare* Regression mit dem Ordinatenabschnitt  $\alpha_0 := \beta_0 + \beta_2 \gamma_0$  und dem Steigungskoeffizienten  $\alpha_1 := \beta_1 + \beta_2 \gamma_1$ . Gleichung (9.37) zeigt, dass  $\alpha_1$  auch dann von Null verschieden sein kann, wenn  $\beta_1$  gleich null ist und umgekehrt. Auch umgekehrte Vorzeichen von  $\alpha_1$  und  $\beta_1$  sind durchaus möglich. Bei der Ableitung der Gleichung (9.37) (s. Übung 9) sieht man, dass  $E(Y|X)$  durchaus keine lineare Funktion von  $X$  sein muss, selbst dann nicht, wenn Gleichung (9.3) gilt. Der Fall, dass trotz Gültigkeit der Gleichung (9.3) die Regression  $E(Y|X)$  *keine* lineare Funktion von  $X$  ist, tritt z. B. dann ein, wenn die Regression  $E(Z|X)$  eine nichtlineare Funktion von  $X$  ist.

## 9.4 Lineare Quasi-Regression

*Echte lineare Regression vs. lineare Quasi-Regression*

*Optimale Linearkombination von  $X$  und  $Z$*

Zum Abschluss wollen wir noch den Begriff der zweifachen *linearen Quasi-Regression* betrachten, der leicht mit dem der „echten“ zweifachen linearen Regression verwechselt werden kann. Der Unterschied zwischen den beiden Begriffen liegt darin, dass die lineare Quasi-Regression eine in einem bestimmten Sinn optimale Linearkombination von  $X$  und  $Z$  ist, gleichgültig, ob die echte Regression eine Linearkombination von  $X$  und  $Z$  ist oder aber eine andere Funktion von  $X$  und  $Z$ .

**Definition 2.** Unter den gleichen Voraussetzungen wie in Definition 1 handelt es sich bei der *zweifachen linearen Quasi-Regression*, die wir

### Antworten

- A1. Die wichtigste Eigenschaft der bezüglich  $Z$  partiellen linearen regressiven Abhängigkeit eines Regressanden  $Y$  von einem Regressor  $X$  ist die Gleichheit der Regressionskoeffizienten  $\beta_1$  der bedingten linearen Regression für alle Werte  $z$  von  $Z$ .
- A2. Falls Gleichung (9.3) vorausgesetzt wird und die Regression  $E(Z|X)$  linear ist, ist auch die einfache Regression  $E(Y|X)$  eine lineare Funktion von  $X$ .
- A3. Falls Gleichung (9.3) vorausgesetzt wird und  $Z$  von  $X$  regressiv unabhängig ist oder der Koeffizient  $\beta_2$  aus Gleichung (9.3) gleich null ist, sind der einfache und der partielle Regressionskoeffizient von  $X$  identisch.
- A4. Wird Gleichung (9.3) vorausgesetzt und sind  $X$  und  $Z$  voneinander regressiv unabhängig [s. Gln. (9.26) und (9.32)], ist der Determinationskoeffizient  $R_{Y|X,Z}^2$  die Summe der Determinationskoeffizienten  $R_{Y|X}^2$  und  $R_{Y|Z}^2$ .
- A5. Im allgemeinen Fall, wenn man nur die Gleichung (9.3), aber keine weiteren Bedingungen voraussetzen kann, kann man den Zähler des Determinationskoeffizienten  $R_{Y|X,Z}^2$  unter Verwendung der Gleichung (9.9) ausrechnen.
- A6. Der wesentliche Unterschied zwischen der Regression  $E(Y|X, Z)$  und der zweifachen linearen Quasi-Regression  $Q(Y|X, Z)$  ist, dass  $Q(Y|X, Z)$  zwar ebenso wie  $E(Y|X, Z)$  das Kleinst-Quadrat-Kriterium minimiert, aber nur bei der Regression  $E(Y|X, Z)$  ist garantiert, dass ihre Werte auch die bedingten Erwartungswerte  $E(Y|X = x, Z = z)$  sind.

### Übungen

leicht  
leicht  
leicht  
schwer

mittel

schwer  
mittel

mittel

mittel  
mittel

- Ü1. Zeigen Sie, dass die Gleichungen (9.13) aus Gleichung (9.11) folgen!
- Ü2. Zeigen Sie, dass die Gleichungen (9.14) aus den Gleichungen (9.13) folgen!
- Ü3. Zeigen Sie, dass die Gleichung (9.15) aus der Gleichung (9.3) folgt!
- Ü4. Zeigen Sie, dass die Gleichungen (9.16) und (9.18) aus der Gleichung (9.3) folgen!
- Ü5. Zeigen Sie, dass die Gleichungen (9.27) bis (9.29) aus den Gleichungen (9.3) und (9.26) folgen!
- Ü6. Zeigen Sie, dass die Gleichung (9.30) aus Gleichung (9.26) folgt!
- Ü7. Zeigen Sie, dass die Gleichung (9.9) aus Gleichung (9.3) folgt und dass auch Gleichung (9.33) folgt, wenn wir auch Gleichung (9.26) und die daraus folgende Gleichung (9.30) voraussetzen können!
- Ü8. Zeigen Sie, dass die Gleichung (9.34) aus den Gleichungen (9.3), (9.26) und (9.32) folgt!
- Ü9. Zeigen Sie, dass die Gleichung (9.37) aus den Gleichungen (9.3) und (9.36) folgt!
- Ü10. Geben Sie ein Zahlenbeispiel, in dem zwar der partielle Regressionskoeffizient  $\beta_1$  aus Gleichung (9.3) positiv ist, aber der einfache Regressionskoeffizient  $\alpha_1$  aus Gleichung (9.27) gleich null ist! [Hinweis: Nutzen Sie Gleichung (9.37)!]

### Lösungen

R-Box 6.2 (vi)  
Gl. (9.11)  
R-Box 6.2 (i)

- L1. Die Gleichungen (9.13) folgen aus Gleichung (9.11), denn
 
$$\begin{aligned} E(\varepsilon|X) &= E[E(\varepsilon|X, Z)|X] \\ &= E(0|X) \\ &= 0. \end{aligned}$$
 Das Entsprechende gilt für  $E(\varepsilon|Z)$ .
- L2. Die Gleichungen (9.14) folgen aus (9.13), da  $E(\varepsilon|X) = 0 + 0 \cdot X$  eine lineare Regression mit dem Steigungskoeffizienten  $\alpha_1 = 0$  ist. Für diesen gilt aber immer  $\alpha_1 = \text{Cov}(X, \varepsilon) / \text{Var}(X)$ . Daraus folgt:  $\text{Cov}(X, \varepsilon) = 0$ . Das entsprechende Argument gilt für  $\text{Cov}(Z, \varepsilon)$ .
- L3. Die Gleichung (9.15) folgt aus Gleichung (9.3), denn
 
$$\begin{aligned} E(Y) &= E[E(Y|X, Z)] \\ &= E(\beta_0 + \beta_1 X + \beta_2 Z) \end{aligned}$$

R-Box 6.2 (iv)  
Gl. (9.3)

(s. R-Box 5.3) und entsprechende Umformungen. Die Umformung der ersten dieser beiden Gleichungen ergibt:

$$\beta_1 = \frac{\text{Cov}(X, Y) - \beta_2 \text{Cov}(X, Z)}{\text{Var}(X)}.$$

Setzen wir dies in die zweite Gleichung ein, erhalten wir

$$\text{Cov}(Z, Y) = \left[ \frac{\text{Cov}(X, Y) - \beta_2 \text{Cov}(X, Z)}{\text{Var}(X)} \right] \cdot \text{Cov}(Z, X) + \beta_2 \text{Var}(Z).$$

Multiplizieren beider Seiten mit  $\text{Var}(X)$  ergibt:

$$\begin{aligned} & \text{Cov}(Z, Y) \cdot \text{Var}(X) \\ &= [\text{Cov}(X, Y) - \beta_2 \text{Cov}(Z, X)] \text{Cov}(Z, X) + \beta_2 \text{Var}(Z) \cdot \text{Var}(X) \\ &= \text{Cov}(X, Y) \cdot \text{Cov}(Z, X) - \beta_2 \text{Cov}(Z, X)^2 + \beta_2 \text{Var}(Z) \cdot \text{Var}(X). \end{aligned}$$

Diese Gleichung lässt sich nun weiter umformen zu:

$$\begin{aligned} & \text{Cov}(Z, Y) \cdot \text{Var}(X) - \text{Cov}(X, Y) \cdot \text{Cov}(Z, X) \\ &= \beta_2 [-\text{Cov}(Z, X)^2 + \text{Var}(Z) \cdot \text{Var}(X)]. \end{aligned}$$

Daraus folgt nun:

$$\beta_2 = \frac{\text{Cov}(Z, Y) \text{Var}(X) - \text{Cov}(X, Y) \text{Cov}(Z, X)}{\text{Var}(X) \text{Var}(Z) - \text{Cov}(Z, X)^2}.$$

Die entsprechende Gleichung für  $\beta_1$  erhält man durch Vertauschung von  $X$  und  $Z$ .

- L5. Die Gleichungen (9.27) bis (9.29) folgen aus den Gleichungen (9.3) und (9.26), denn:

$$\begin{aligned} E(Y|X) &= E[E(Y|X, Z) | X] \\ &= E(\beta_0 + \beta_1 X + \beta_2 Z | X) \\ &= \beta_0 + \beta_1 E(X|X) + \beta_2 E(Z|X) \\ &= \beta_0 + \beta_1 X + \beta_2 E(Z|X) \\ &= [\beta_0 + \beta_2 E(Z)] + \beta_1 X. \end{aligned}$$

- L6. Die Gleichung (9.30) folgt aus der regressiven Unabhängigkeit der Variablen  $Z$  von  $X$  [Gl. (9.26)], denn:

$$\begin{aligned} \text{Cov}(X, Z) &= E(XZ) - E(X)E(Z) \\ &= E[E(XZ|X)] - E(X)E(Z) \\ &= E[XE(Z|X)] - E(X)E(Z) \\ &= E[XE(Z)] - E(X)E(Z) \\ &= E(X)E(Z) - E(X)E(Z) = 0. \end{aligned}$$

- L7. Die Gleichung (9.34) folgt aus Gleichung (9.3), denn:

$$\begin{aligned} \text{Var}[E(Y|X, Z)] &= \text{Var}(\beta_0 + \beta_1 X + \beta_2 Z) \\ &= \beta_1^2 \text{Var}(X) + \beta_2^2 \text{Var}(Z) + 2\beta_1 \beta_2 \text{Cov}(Y, X). \end{aligned}$$

Können wir auch Gleichung (9.26) und die daraus folgende Gleichung (9.30) verwenden, dann folgt:

$$\text{Var}[E(Y|X, Z)] = \beta_1^2 \text{Var}(X) + \beta_2^2 \text{Var}(Z).$$

- L8. Die Gleichung (9.34) folgt aus den Gleichungen (9.3), (9.26) und (9.32), denn aus der Gleichung (9.31) erhält man durch Einsetzen der beiden Gleichungen  $\beta_1 = \text{Cov}(X, Y)/\text{Var}(X)$  und  $\beta_2 = \text{Cov}(Z, Y)/\text{Var}(Z)$ , die bei Gültigkeit der Gleichung (9.26) und (9.32) für die partiellen Regressionskoeffizienten gelten:

$$\text{Var}[E(Y|X, Z)] = \text{Cov}(X, Y)^2 / \text{Var}(X) + \text{Cov}(Y, Z)^2 / \text{Var}(Z)$$

Dividieren durch die Varianz von  $Y$  ergibt:

$$\frac{\text{Var}[E(Y|X, Z)]}{\text{Var}(Y)} = \frac{\text{Cov}(X, Y)^2}{\text{Var}(X) \cdot \text{Var}(Y)} + \frac{\text{Cov}(Y, Z)^2}{\text{Var}(Z) \cdot \text{Var}(Y)}$$

R-Box 6.2 (vi)  
Gl. (9.3)  
R-Box 6.2 (i) bis (iv)  
R-Box 6.2 (v)  
Gl. (9.26)

R-Box 5.3 (i)  
R-Box 6.2 (iv)  
R-Box 6.2 (vii)  
Gl. (9.26)  
R-Box 5.1 (ii)

Gl. (9.3)  
R-Box 5.2 (v)

Gl. (9.30)



Durch Einsetzen der Definitionen der drei Determinationskoeffizienten  $R_{Y|X,Z}^2$ ,  $R_{Y|X}^2$  und  $R_{Y|Z}^2$  (s. R-Box 6.3) erhält man dann die Gleichung (9.34).

L9. Die Gleichung (9.37) folgt aus den Gleichungen (9.3) und (9.36), denn

$$\begin{aligned}
 E(Y|X) &= E[E(Y|X,Z)|X] && \text{R-Box 6.2 (vi)} \\
 &= E[\beta_0 + \beta_1 X + \beta_2 Z|X] && \text{Gl. (9.3)} \\
 &= E(\beta_0|X) + E(\beta_1 X|X) + E(\beta_2 Z|X) && \text{R-Box 6.2 (iii)} \\
 &= \beta_0 + \beta_1 X + \beta_2 E(Z|X) && \text{R-Box 6.2 (i), (v), (ii)} \\
 &= \beta_0 + \beta_1 X + \beta_2 (\gamma_0 + \gamma_1 X) && \text{Gl. (9.36)} \\
 &= (\beta_0 + \beta_2 \gamma_0) + (\beta_1 + \beta_2 \gamma_1) X.
 \end{aligned}$$

L10. Unter Verwendung der Gleichung (9.37) lassen sich solche Beispiele leicht angeben. Sind beispielsweise  $\beta_0 = 100$ ,  $\beta_1 = 10$  und  $\beta_2 = 5$  die Koeffizienten der Gleichung (9.3), so führen die Koeffizienten  $\gamma_0 = 1$  und  $\gamma_1 = -2$  aus Gleichung (9.36) zu einer linearen Regression  $E(Y|X)$  mit dem Steigungskoeffizienten  $\alpha_1 = 0$ .

liche Erwartungswerte und Varianzen haben. (Andernfalls wären diese Kovarianzen gar nicht definiert.)

Die in den Gleichungen (10.6) und (10.7) beschriebenen Eigenschaften des Residuums eignen sich am besten zur empirischen Überprüfung der Frage, ob in einer konkreten Anwendung der betrachtete Regressand  $Y$  von  $X$  tatsächlich  $Z$ -bedingt linear regressiv abhängig ist. Ist für eine spezifizierte Funktion von  $X$  und  $Z$  der Form  $g_0(Z) + g_1(Z) \cdot X$  und die Variable  $\varepsilon := Y - [g_0(Z) + g_1(Z) \cdot X]$  eine der drei Gleichungen  $E(\varepsilon | X, Z) = E(\varepsilon | X) = E(\varepsilon | Z)$  [vgl. Gleichung (10.6) und (10.7)] *nicht* erfüllt, so kann in einem solchen Fall diese Funktion  $g_0(Z) + g_1(Z) \cdot X$  nicht die Regression  $E(Y | X, Z)$  sein.

*Anwendung bei der empirischen  
Überprüfung bedingter  
linearer regressiver  
Abhängigkeit*

### 10.2.3 Einfache Spezialfälle der bedingten linearen regressiven Abhängigkeit

Ist die Regression  $E(Y | X, Z)$  von  $Y$  auf  $X$  und  $Z$  mit der Regression  $E(Y | X)$  von  $Y$  auf  $X$  identisch, d. h. gilt  $E(Y | X, Z) = E(Y | X)$ , so sprechen wir von *bedingter regressiver Unabhängigkeit* des Regressanden  $Y$  von  $Z$ , gegeben  $X$ .

*Bedingte regressive  
Unabhängigkeit*

Die obige Definition der bedingten linearen regressiven Abhängigkeit schließt auch denjenigen Spezialfall ein, in dem  $g_0(Z)$  und  $g_1(Z)$  konstante Funktionen von  $Z$  sind, also für alle Werte  $z$  von  $Z$  den gleichen Wert  $g_0(Z) = \beta_0$  bzw.  $g_1(Z) = \gamma_0$  annehmen. In diesem Fall vereinfacht sich Gleichung (10.3) zu

$$E(Y | X, Z) = \beta_0 + \gamma_0 X = E(Y | X). \quad (10.10)$$

Dies ist also ein Spezialfall der bedingten regressiven Unabhängigkeit des Regressanden  $Y$  von  $Z$ , gegeben  $X$ . Warum wir hier die Koeffizienten dieser einfachen linearen Regression mit  $\beta_0$  bzw.  $\gamma_0$  bezeichnen, wird in den folgenden Abschnitten verständlich [s. insbesondere die Gln. (10.11) und (10.13)].

Sind dagegen

$$g_0(Z) = \beta_0 + \beta_1 Z, \quad \beta_0, \beta_1 \in \mathbb{R}, \quad (10.11)$$

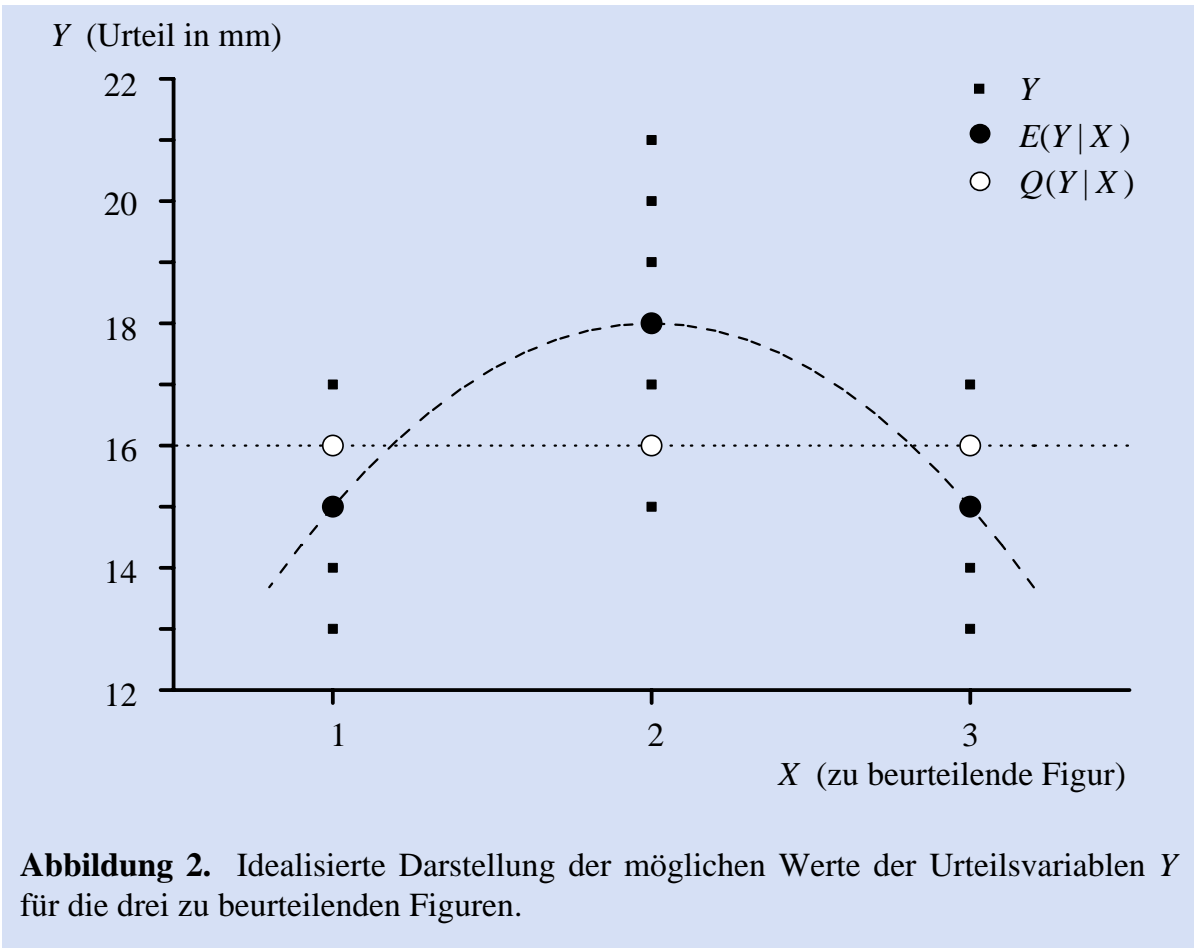
eine lineare Funktion von  $Z$  und  $g_1(Z) = \gamma_0$  eine konstante reelle Funktion, so vereinfacht sich Gleichung (10.3) zu

$$E(Y | X, Z) = \beta_0 + \beta_1 Z + \gamma_0 X, \quad (10.12) \quad \text{Zweifache lineare Regression}$$

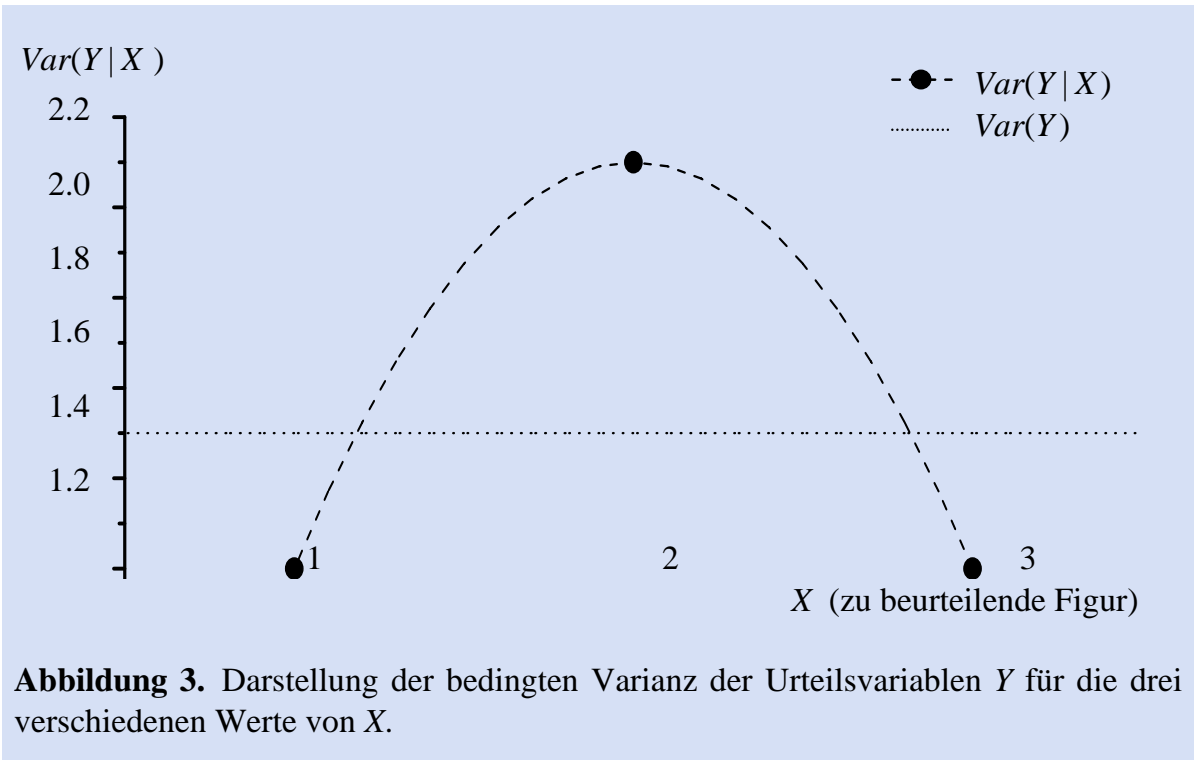
dem im letzten Kapitel behandelten Spezialfall der zweifachen linearen Regression. In den beiden bisher genannten Spezialfällen ist also die Modifikatorfunktion  $g_1(Z)$  eine Konstante.

Gilt sowohl Gleichung (10.11) und ist

$$g_1(Z) = \gamma_0 + \gamma_1 Z, \quad \gamma_0, \gamma_1 \in \mathbb{R}, \quad (10.13)$$



Werte dargestellt. In Abbildung 3 ist nun direkt die *bedingte Varianz* von  $Y$  für die drei verschiedenen Werte von  $X$  aufgetragen. Dabei mache man sich klar, dass die Kurve hier zwar einen ähnlichen Verlauf hat, wie die Darstellung der *bedingten Erwartungswerte* in Abbildung 2, dass in den beiden Abbildungen aber dennoch zwei verschiedene Dinge dargestellt sind. In Abbildung 2 sind die unterschiedlichen bedingten Varianzen an der Streubreite der Werte der Urteilsvariablen zu erkennen, in Abbildung 3 dagegen sind unterschiedlichen bedingten Varianzen direkt als Funktion von  $X$  eingetragen



- leicht

F5. Was weiß man über die Beziehung zwischen dem Erwartungswert der bedingten Kovarianz zweier numerischer Zufallsvariablen  $Y_1$  und  $Y_2$  gegeben  $X$  und der Kovarianz der beiden Residuen  $\varepsilon_1 = Y_1 - E(Y_1 | X)$  und  $\varepsilon_2 = Y_2 - E(Y_2 | X)$ ?
- mittel

F6. Warum korrelieren im Fall linearer Regressionen  $E(Y_i | X) = \alpha_{i0} + \alpha_{i1} X$ ,  $i = 1, 2$ , diese beiden Regressionen miteinander zu 1?
- mittel

F7. Welche Rechenregeln für (unbedingte) Varianzen und Kovarianzen sind die Analoga zu (v), (vi), (xi) und (xii) der Regelbox 1?

Antworten

- A1. Die bedingte Kovarianz zweier numerischer Zufallsvariablen  $Y_1$  und  $Y_2$  gegeben  $X$  ist definiert als die bedingte Erwartung  $E(\varepsilon_1 \cdot \varepsilon_2 | X)$  des Produkts der beiden Residuen  $\varepsilon_1 = Y_1 - E(Y_1 | X)$  und  $\varepsilon_2 = Y_2 - E(Y_2 | X)$  gegeben  $X$ .
- A2. Ein Wert der bedingten Kovarianz zweier numerischer Zufallsvariablen  $Y_1$  und  $Y_2$  gegeben  $X$  gibt an, wie stark die durch eine lineare Funktion beschreibbare Abhängigkeit zwischen  $Y_1$  und  $Y_2$  bei gegebenem Wert  $x$  des Regressors  $X$  ist.
- A3. Die bedingte Kovarianz  $Cov(Y_1, Y_2 | X = x)$  ist eine Zahl, wohingegen die bedingte Kovarianz  $Cov(Y_1, Y_2 | X)$  eine Zufallsvariable ist, deren Werte die bedingten Kovarianzen  $Cov(Y_1, Y_2 | X = x)$  sind.
- A4. Ein Wert  $Var(Y | X = x)$  der bedingten Varianz  $Var(Y | X)$  gibt an, wie stark die Werte des Regressanden  $Y$  um den bedingten Erwartungswert  $E(Y | X = x)$  herum variieren, d. h.  $Var(Y | X = x)$  ist ein Kennwert für die Dispersion der Verteilung des Regressanden  $Y$  an der Stelle  $x$  von  $X$ .
- A5. Der Erwartungswert der bedingten Kovarianz zweier numerischer Zufallsvariablen  $Y_1$  und  $Y_2$  gegeben  $X$  ist gleich der Kovarianz der beiden Residuen  $\varepsilon_1 = Y_1 - E(Y_1 | X)$  und  $\varepsilon_2 = Y_2 - E(Y_2 | X)$ .
- A6. Die linearen Regressionen  $E(Y_i | X) = \alpha_{i0} + \alpha_{i1} X$  korrelieren zu 1, weil es sich jeweils um *lineare Funktionen* einer numerischen Zufallsvariablen  $X$  handelt. Eine numerische Zufallsvariable korreliert mit sich selbst zu 1 und die Korrelation ist invariant unter linearen Transformationen.
- A7. Der Regel (v) entspricht die Regel (v) der Regelbox 5.2 und der Regel (vi) entspricht die Regel (iii) der Regelbox 5.2. Bei gegebenem Wert  $x$  ist nämlich auch  $f(x)$  eine Konstante. Weiter entspricht die Regel (xi) der Regel (iii) der Regelbox 5.3 und der Regel (xii) entspricht die Regel (v) der Regelbox 5.3.

Übungen

- mittel

Ü1. Berechnen Sie die Varianz des Residuums für das in den Abbildungen 1 bis 3 dargestellte Beispiel unter Verwendung der Rechenregel aus Regelbox 1!
- mittel

Ü2. Leiten Sie die Gleichung  $Cov(Y_1, Y_2 | X) = E(Y_1 \cdot Y_2 | X) - E(Y_1 | X) \cdot E(Y_2 | X)$  aus der Definition der bedingten Kovarianz ab.
- mittel

Ü3. Leiten Sie die Gleichung  $E[Var(Y | X)] = Var(\varepsilon) = E(\varepsilon^2)$  unter Verwendung der Rechenregeln für Erwartungswerte und Eigenschaften des Residuums  $\varepsilon$  her!
- mittel

Ü4. Das Modell kongenerischer Variablen (s. z. B. Steyer & Eid, 2001, Kap. 13 – 15) kann man durch die beiden Gleichungen  $E(Y_i | \eta) = \lambda_{i0} + \lambda_{i1} \eta$ ,  $\lambda_{i0}, \lambda_{i1} \in \mathbb{R}$ , und  $Cov(\varepsilon_i, \varepsilon_j) = 0$ ,  $i \neq j$ , definieren, wobei  $\eta$  eine latente (Zufalls-)Variable ist und  $\varepsilon_i := Y_i - E(Y_i | \eta)$ . Zeigen Sie unter Verwendung der in den Kapiteln 5 und 6 angegebenen Rechenregeln, dass aus  $Cov(Y_i, Y_j | \eta) = 0$ ,  $i \neq j$ , die Gleichung  $Cov(\varepsilon_i, \varepsilon_j) = 0$ ,  $i \neq j$ , folgt.
- mittel

Ü5. Zeigen Sie, dass
$$Std(\varepsilon) = Std(Y) \cdot \sqrt{1 - R_{Y|X}^2}$$
gilt, wobei  $\varepsilon := Y - E(Y | X)$ .
- schwer

Ü6. Leiten Sie die Gleichung (12.9) aus der Gleichung (12.8) ab.  
Hinweis: Zeigen Sie zunächst, dass  $Cov[Y_1, E(Y_2 | X)] = Cov[E(Y_1 | X), E(Y_2 | X)]$  gilt. Unter Verwendung dieser Beziehung, des Ergebnisses aus Übung 5 sowie der Definition der Korrelation und des Determinationskoeffizienten  $R_{Y|X}^2$  können Sie dann ausgehend von Gleichung (12.8) die Gleichung (12.9) herleiten.

$$\mathbf{A}\mathbf{B} = \begin{pmatrix} 1 & 3 & 2 \\ 0 & -8 & 1 \\ 1 & 3 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -8 \end{pmatrix} = \begin{pmatrix} 1 & 6 & -16 \\ 0 & -16 & -8 \\ 1 & 6 & -40 \end{pmatrix}.$$

(d)  $\mathbf{B}$  hat drei Spalten und  $\mathbf{A}$  drei Zeilen. Also ist auch das Produkt  $\mathbf{B}\mathbf{A}$  definiert:

$$\mathbf{B}\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -8 \end{pmatrix} \begin{pmatrix} 1 & 3 & 2 \\ 0 & -8 & 1 \\ 1 & 3 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 \\ 0 & -16 & 2 \\ -8 & -24 & -40 \end{pmatrix}.$$

(e)  $\mathbf{A}$  hat drei Spalten und  $\mathbf{C}$  drei Zeilen. Das Produkt  $\mathbf{A}\mathbf{C}$  ist also definiert:

$$\mathbf{A}\mathbf{C} = \begin{pmatrix} 1 & 3 & 2 \\ 0 & -8 & 1 \\ 1 & 3 & 5 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ -2 & 3 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -2 & 7 \\ 16 & -25 \\ -2 & 4 \end{pmatrix}.$$

(f)  $\mathbf{C}$  hat zwei Spalten, aber  $\mathbf{A}$  hat drei Zeilen. Das Produkt  $\mathbf{C}\mathbf{A}$  ist folglich nicht definiert.

(g)  $\mathbf{C}$  ist keine quadratische Matrix, kann also nicht invertiert werden.

(h)  $\mathbf{B}$  ist eine Diagonalmatrix, auf deren Hauptdiagonalen sämtliche Komponenten ungleich null sind. Ihre Inverse existiert und berechnet sich wie folgt:

$$\mathbf{B}^{-1} = \begin{pmatrix} \frac{1}{1} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{8} \end{pmatrix}.$$

L2. (a) Die vom Modell implizierte Struktur der einzelnen Kovarianzen  $\sigma_{Y_i Y_j} = \text{Cov}(Y_i, Y_j)$  berechnet sich wie folgt:

$$\begin{aligned} \sigma_{Y_i Y_j} &= \text{Cov}(Y_i, Y_j) \\ &= \text{Cov}(\lambda_i + \eta + \varepsilon_i, \lambda_j + \eta + \varepsilon_j) \\ &= \text{Cov}(\eta + \varepsilon_i, \eta + \varepsilon_j) \\ &= \text{Cov}(\eta, \eta) + \text{Cov}(\eta, \varepsilon_j) + \text{Cov}(\varepsilon_i, \eta) + \text{Cov}(\varepsilon_i, \varepsilon_j) \end{aligned}$$

Da laut Aufgabenstellung  $\eta$  und die einzelnen Fehlervariablen sowie jeweils zwei voneinander verschiedene Fehlervariablen unkorreliert sind, gelten:

$$\begin{aligned} \text{für } i=j: \quad \sigma_{Y_i Y_i} &= \text{Cov}(\eta, \eta) + \text{Cov}(\varepsilon_i, \varepsilon_i) = \sigma_\eta^2 + \sigma_{\varepsilon_i}^2 \\ \text{für } i \neq j: \quad \sigma_{Y_i Y_j} &= \text{Cov}(\eta, \eta) = \sigma_\eta^2. \end{aligned}$$

Folglich hat die Varianz-Kovarianzmatrix die Struktur:

$$\Sigma_{yy} = \begin{pmatrix} \sigma_\eta^2 + \sigma_{\varepsilon_1}^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{\varepsilon_2}^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{\varepsilon_3}^2 \end{pmatrix}.$$

(b) Betrachtet man  $Y_i$  als die  $i$ -te Zeile des Vektors  $\mathbf{y}$ , so erhält man den gesamten Vektor  $\mathbf{y}$  als:

$$\mathbf{y} = \lambda_0 + \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (\boldsymbol{\eta}) + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}.$$

Aus Annahme (2) erhält man die Kovarianzmatrix  $\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\varepsilon}) = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$ , und aus Annahme (3) die Varianz-Kovarianzmatrix  $\text{Var}(\boldsymbol{\varepsilon})$ :

Annahme (1)  
R-Box 5.3 (iv)  
R-Box 5.3 (v)



$$E(\mathbf{y} | \mathbf{x}) = E(\mathbf{y} | \mathbf{z}) = \mathbf{z}' \boldsymbol{\gamma} = (1 \ X_1 \ \dots \ X_m) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}. \quad (14.11)$$

Auf diese noch einfachere Darstellung der multiplen linearen Regression werden bei der Einführung des Allgemeinen Linearen Modells in diesem Kapitel zurückgreifen.

#### 14.1.4 Identifikation der Regressionskoeffizienten

Zur Bestimmung von  $\beta_0$  und der Komponenten von  $\boldsymbol{\beta} = (\beta_1 \dots \beta_m)'$  greift man (wie bei der Betrachtung von nur zwei Regressoren  $X_1$  und  $X_2$ ) auf die Erwartungswerte des Regressanden und der Regressoren sowie die Kovarianzmatrizen  $\Sigma_{xx}$  und  $\Sigma_{xy}$  zurück. Für die Konstante  $\beta_0$  ergibt sich (s. Übung 1)

$$\beta_0 = E(\mathbf{y}) - E(\mathbf{x})' \boldsymbol{\beta} = E(Y) - [E(X_1) \ \dots \ E(X_m)] \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}. \quad (14.12)$$

*Identifikation  
der Regressionskonstanten*

Für die Bestimmung von  $\beta_0$  benötigt man also wiederum, neben den Erwartungswerten der beteiligten Variablen, auch die Regressionskoeffizienten  $\beta_1, \dots, \beta_m$ .

Die Regressionskoeffizienten  $\beta_1, \dots, \beta_m$  lassen sich ausgehend von der Kovarianzmatrix  $\Sigma_{xy}$  bestimmen. Mit  $\boldsymbol{\varepsilon} := \mathbf{y} - E(\mathbf{y} | \mathbf{x}) = \mathbf{y} - E(\beta_0 + \mathbf{x}' \boldsymbol{\beta})$  ergibt sich nach den Regeln (v) bis (vii) der Regelbox 13.3

$$\begin{aligned} \Sigma_{xy} &= \text{Cov}(\mathbf{x}, \mathbf{y}) = \text{Cov}(\mathbf{x}, \beta_0 + \mathbf{x}' \boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \text{Cov}(\mathbf{x}, \beta_0 + \boldsymbol{\beta}' \mathbf{x} + \boldsymbol{\varepsilon}) \\ &= \text{Cov}(\mathbf{x}, \mathbf{x}) \boldsymbol{\beta} = \Sigma_{xx} \boldsymbol{\beta}. \end{aligned}$$

Der Vektor  $\boldsymbol{\beta}$  der Regressionsgewichte lässt sich bestimmen, indem man diese Gleichung nach  $\boldsymbol{\beta}$  auflöst. Dies geschieht durch die Multiplikation beider Seiten mit der Inversen  $\Sigma_{xx}^{-1}$  der Kovarianzmatrix der Regressoren. Zur Erinnerung: Diese Inverse existiert, wenn keine der Variablen  $X_1, \dots, X_m$  eine Linearkombination der übrigen ist (s. z. B. Graybill, 1983; Schmidt & Trenkler, 1998; Searle, 1982; Searle & Willet, 2001; Zurmühl & Falk, 1992; s. dazu auch Abschnitt 13.3).

Nach Multiplikation beider Seiten der obigen Gleichung mit der Inversen  $\Sigma_{xx}^{-1}$  erhält man zunächst

$$\Sigma_{xx}^{-1} \Sigma_{xx} \boldsymbol{\beta} = \Sigma_{xx}^{-1} \Sigma_{xy}.$$

Da  $\Sigma_{xx}^{-1} \Sigma_{xx} = \mathbf{I}$  die Einheitsmatrix ist, folgt daraus

$$\boldsymbol{\beta} = \Sigma_{xx}^{-1} \Sigma_{xy}. \quad (14.13)$$

*Allgemeine Formel  
für die Berechnung  
der Regressionskoeffizienten*

$$\text{Var}[E(\mathbf{y}|\mathbf{x})] = \boldsymbol{\beta}' \text{Var}(\mathbf{x}) \boldsymbol{\beta} = \boldsymbol{\beta}' \boldsymbol{\Sigma}_{\mathbf{xx}} \boldsymbol{\beta}. \quad (14.14) \quad \text{Varianz der Regression}$$

Dabei sind  $\text{Var}(\mathbf{x}) = \boldsymbol{\Sigma}_{\mathbf{xx}}$  die  $(m \times m)$  Varianz-Kovarianzmatrix der Regressoren  $X_1, \dots, X_m$  und  $\boldsymbol{\beta}$  der Spaltenvektor der Regressionskoeffizienten  $\beta_1, \dots, \beta_m$ . Der Determinationskoeffizient ergibt sich dann wie folgt:

$$R_{Y|X_1, \dots, X_m}^2 = [\boldsymbol{\beta}' \boldsymbol{\Sigma}_{\mathbf{xx}} \boldsymbol{\beta}] / \text{Var}(Y). \quad (14.15) \quad \text{Determinationskoeffizient}$$

Im Fall mit zwei Regressoren  $X_1$  und  $X_2$  erhält man die bereits aus Kapitel 9 bekannte Gleichung

$$R_{Y|X_1, X_2}^2 = [\beta_1^2 \text{Var}(X_1) + \beta_2^2 \text{Var}(X_2) + 2\beta_1\beta_2 \text{Cov}(X_1, X_2)] / \text{Var}(Y).$$

Für den Spezialfall, dass alle Regressoren *paarweise unkorreliert* sind, dass also für alle Kovarianzen zweier Regressoren  $X_i$  und  $X_j$  mit  $i \neq j$  gilt:  $\text{Cov}(X_i, X_j) = 0$ , vereinfacht sich der Ausdruck für den multiplen Determinationskoeffizienten. Die Varianz-Kovarianz-Matrix  $\text{Var}(\mathbf{x}) = \boldsymbol{\Sigma}_{\mathbf{xx}}$  der Regressoren ist dann eine  $(m \times m)$ -Diagonalmatrix. Für den Determinationskoeffizienten  $R_{Y|X_1, \dots, X_m}^2$  erhält man dann den Ausdruck

$$R_{Y|X_1, \dots, X_m}^2 = \left( \sum_{i=1}^m \beta_i^2 \text{Var}(X_i) \right) \frac{1}{\text{Var}(Y)}.$$

Die Gleichung

$$R_{Y|X_1, \dots, X_m}^2 = \sum_{i=1}^m R_{Y|X_i}^2$$

*Spezialfall  
paarweiser Unkorreliertheit  
der Regressoren*

jedoch gälte erst dann, wenn auch  $E(X_i|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m) = E(X_i)$  für alle Regressoren  $X_i$  gälte.

## 14.2 Multiple lineare Quasi-Regression

Nicht jede Linearkombination  $\beta_0 + \beta_1 X_1 + \dots + \beta_m X_m$  von Regressoren ist tatsächlich eine Regression, auch dann nicht, wenn die Koeffizienten  $\beta_0, \beta_1, \dots, \beta_m$  im Sinne des Kleinst-Quadrat-Prinzips optimal sind. Oft hat man eine Hypothese, dass eine solche optimale Linearkombination zugleich auch die Regression  $E(Y|X_1, \dots, X_m)$  ist. Um eine solche Hypothese überhaupt formulieren zu können, müssen wir also auch hier zwischen der multiplen *linearen Quasi-Regression* und der (echten) multiplen linearen Regression unterscheiden. Der Unterschied zwischen den beiden Begriffen liegt darin, dass die multiple lineare Quasi-Regression eine in einem bestimmten Sinn optimale Linearkombination der Regressoren  $X_1, \dots, X_m$  ist, gleichgültig, ob die echte Regression eine Linearkombination der  $X_1, \dots, X_m$  ist oder eine andere Funktion dieser Regressoren.

Regressoren diejenigen selektiert werden sollen, die einen möglichst hohen Varianzanteil des Regressanden  $Y$  mit einer möglichst kleinen Zahl von Regressoren erklärt werden soll.

### Fragen

- |        |   |
|--------|---|
| leicht | F1. Warum ist $\mathbf{x}' \boldsymbol{\beta} = \boldsymbol{\beta}' \mathbf{x}$ ?   |
| leicht | F2. Wie viele Parameter hat eine saturierte Parametrisierung der dreifachen Regression $E(Y   X_1, X_2, X_3)$ , wenn jeder der drei Regressoren jeweils nur zwei verschiedenen Werte hat?   |
| mittel | F3. Geben Sie ein einfaches Beispiel an, in dem die wahre regressive Abhängigkeit des Regressanden $Y$ von den Regressoren in $\mathbf{x}$ komplizierter als linear ist. Woran könnte man sehen, dass für dieses Beispiel die multiple lineare Quasi-Regression nicht gleich der echten Regression ist? |
| leicht | F4. Wie kann man die Differenz $R_{Y X_1, \dots, X_m}^2 - Q_{Y X_1, \dots, X_{m-p}}^2$ interpretieren?  |
| leicht | F5. Was bedeutet die Annahme der Homoskedastizität im ALM?  |
| leicht | F6. Welche Verteilungsannahme wird im ALM gemacht?  |
| mittel | F7. Was versteht man unter einer saturierten Parametrisierung einer Regression?   |
| mittel | F8. In welchem Kapitel wurde schon mal eine nichtlineare regressive Abhängigkeit durch eine multiple lineare Regression dargestellt und um welche Art der Abhängigkeit ging es dabei?   |
| schwer | F9. Unter welcher Bedingung sind die Erwartungswerte und die Varianzen der Beobachtungen $Y_i$ , die einen identischen Zeilenvektor in der Designmatrix $\mathbf{X}$ aufweisen, gleich und welche Besonderheit gilt bei zufälliger Ziehung innerhalb der Wertekombinationen der Regressoren?            |
| leicht | F10. Wie geht man beim schrittweisen Verfahren der Modellsuche vor?   |

### Antworten

- A1. Hier wird ein Zeilenvektor mit einem Spaltenvektor multipliziert. Dabei sind Produkte von jeweils zwei Zahlen aufzuaddieren. Bei einem Produkt zweier Zahlen spielt aber die Reihenfolge keine Rolle, d. h.  $a b = b a$ . Eine andere Begründung ist, dass es sich bei  $\mathbf{x}' \boldsymbol{\beta}$  um eine Zahl handelt, die mit ihrer Transponierten identisch ist. Daher gilt:  $\mathbf{x}' \boldsymbol{\beta} = (\mathbf{x}' \boldsymbol{\beta})' = \boldsymbol{\beta}' \mathbf{x}$  (s. R-Box 13.1).
- A2. Die saturierte Parametrisierung hat dann acht Parameter.
- A3. Ein einfaches Beispiel dafür ist  $E(Y | X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$  mit  $\beta_3 \neq 0$ . Die zweifache lineare Quasi-Regression hätte die Form  $Q(Y | X_1, X_2) = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2$ . Für die Wertekombinationen von  $X_1$  und  $X_2$  würde dann nicht mehr gelten:  $E(v | X_1 = x_{1i}, X_2 = x_{2j}) = 0$ . Sind die beiden Regressoren dichotom, dann hätte man vier solcher Wertekombinationen, innerhalb derer man jeweils den Erwartungswert des Residuums  $v$  überprüfen könnte.
- A4. Die Differenz  $R_{Y|X_1, \dots, X_m}^2 - Q_{Y|X_1, \dots, X_{m-p}}^2$  kann man als den durch die  $p$  zusätzlichen Regressoren zusätzlich erklärten Varianzanteil interpretieren.
- A5. Die Annahme der Homoskedastizität im ALM bedeutet, dass alle Beobachtungen  $Y_i$  die gleiche Varianz haben. Innerhalb einer Wertekombination der Regressoren ist diese Annahme immer erfüllt. Zwischen verschiedenen Wertekombination der Regressoren kann diese Annahme jedoch durchaus falsch sein.
- A6. Die Verteilungsannahme im ALM ist, dass die Beobachtungen  $Y_i$  unabhängig und multivariat normalverteilt sind.
- A7. Eine saturierte Parametrisierung einer Regression ist eine Funktion des Regressors, in der genauso viele Parameter vorkommen, wie der Regressor Werte hat. Bei einer Regression mit mehreren Regressoren ist es eine Funktion der Regressoren, in der genauso viele Parameter vorkommen, wie die Regressoren Wertekombinationen haben.

**Tabelle 5.** Beispiel für eine homogene Population

Person	$\mu_{1,u}$	$\mu_{2,u}$	$\mu_{1,u} - \mu_{2,u}$
$u_1$	110	100	10
$u_2$	110	100	10
$u_3$	110	100	10
$u_4$	110	100	10
$u_5$	110	100	10
$u_6$	110	100	10
$u_7$	110	100	10
$u_8$	110	100	10
Mittel	110	100	10

gar bei jeder einzelnen Person einer Population ein *positiver* Effekt einer Behandlung vor, aber dennoch resultiert eine *negative* Mittelwertedifferenz, wenn man diese Personen in zwei Gruppen, die der Behandelten und die der Nichtbehandelten, aufteilt und dabei nicht sorgfältig auf die Vergleichbarkeit der beiden Gruppen achtet. Iseler (1996) hat ein analoges Paradoxon für Mediane vorgestellt und gezeigt, dass Randomisierung derartige Paradoxa nicht verhindert, wenn man die kausalen Effekte mit Medianen anstatt mit Erwartungswerten definieren würde (s. auch Iseler, 1997).

Die hier dargestellten Beispiele haben deutlich gemacht, dass wir nicht *per se* an Mittelwerten oder Mittelwerteunterschieden interessiert sind. Worüber wir eigentlich etwas erfahren *wollen*, sind die *individuellen kausalen Effekte*, d. h. die Differenzen zwischen Erwartungswerten einer Person bzgl. eines Regressanden  $Y$  in den verschiedenen Ausprägungen des betrachteten Regressors  $X$ . Worüber wir unter günstigen Umständen etwas erfahren *können*, ist der *Durchschnitt* dieser individuellen kausalen Effekte in einer Gesamtpopulation oder in verschiedenen Subpopulationen. Diese günstigen Umstände kann man in einem *randomisierten Experiment* herstellen, in dem man die Personen den experimentellen Bedingungen nach dem Zufallsprinzip zuordnet. Damit wird die oben erwähnte Vergleichbarkeit der experimentellen Gruppen garantiert, wobei wir natürlich voraussetzen, dass die Randomisierung nicht durch systematischen Ausfall der Personen konterkariert wird (vgl. hierzu Cook & Campbell, 1979).

Dabei ist auch zu bedenken, dass wir hier nicht von der Stichprobe reden, sondern von den Verhältnissen in der Population. In einer Stichprobe können natürlich auch bei randomisierter Zuordnung der Personen auf die experimentellen Bedingungen zufällige Mittelwerteunterschiede vorkommen, die in fünf von hundert Fällen auch auf dem 5%-Niveau statistisch signifikant sind. Die dadurch auftretenden Fehler sind jedoch unvermeidbar, solange wir Entscheidungen unter Unsicherheit treffen müssen. Vermeidbar sind jedoch die systematischen Fehler, die wir begehen, wenn wir Mittelwerte nichtvergleichbarer Gruppen vergleichen.

Was wir erfahren wollen und was wir erfahren können

Randomisiertes Experiment ermöglicht die Ermittlung durchschnittlicher kausaler Effekte

*schlecht* ( $Z_3$ ). Alle vor der Behandlung erhebbarer Variablen, die die *Krankheitskosten in den fünf Jahren nach der Behandlung* ( $Y$ ) prädictieren könnten, sind hier zu bedenken. Bei dieser Strategie wäre also ein wichtiges Ziel der empirischen Untersuchungen eine möglichst einfache Variable  $Z$  zu finden, für deren Werte  $z$  jeweils die Gleichung

$$E_{Z=z}(Y | X, U) = E_{Z=z}(Y | X) \quad (17.14)$$

gilt.

## 17.4 Berechnung des durchschnittlichen kausalen Effekts in der Gesamtpopulation

Das folgende Theorem ist entscheidend für die Berechnung durchschnittlicher kausaler Effekte  $ACE(1, 2)$  in der Gesamtpopulation, wenn man davon ausgehen kann, dass die bedingten Regressionen  $E_{Z=z}(Y | X)$  kausal unverfälscht sind. Selbst wenn die Regression  $E(Y | X)$  und ihre Werte verfälscht sind, kann man unter bestimmten Voraussetzungen die kausal unverfälschten Erwartungswerte  $CUE(Y | X = x)$  und daraus die (unbedingten) durchschnittlichen kausalen Effekte  $ACE(1, 2)$  berechnen. Die theoretische Grundlage dazu liefert das folgende Theorem.

**Theorem 4.** Es mögen die in Definition 1 genannten Voraussetzungen gelten. Wenn gelten:

(a) für jeden Wert  $z$  einer Variablen  $Z$  ist die bedingte Regression  $E_{Z=z}(Y | X)$  kausal unverfälscht,

und

(b)  $E(Y | X, U, Z) = E(Y | X, U)$ ,

dann folgt für jeden Wert  $x$  von  $X$ :

$$CUE(Y | X = x) = \sum_z E_{Z=z}(Y | X = x) P(Z = z). \quad (17.15)$$

*Berechnung  
der kausal unverfälschten  
bedingten Erwartungswerte*

**Bemerkungen.** (i) Das Theorem gibt zunächst nur an, wie man unter den Voraussetzungen (a) und (b) die unverfälschten Erwartungswerte  $CUE(Y | X = x)$  aus den bedingten Erwartungswerten  $E_{Z=z}(Y | X = x)$  und den Wahrscheinlichkeiten  $P(Z = z)$  berechnen kann. Die Differenz  $CUE(Y | X = x_1) - CUE(Y | X = x_2)$  zweier unverfälschter Erwartungswerte ist dann jedoch gleich dem durchschnittlichen kausalen Effekt  $ACE(1, 2)$  in der Gesamtpopulation, d. h.:

$$ACE(1, 2) = CUE(Y | X = x_1) - CUE(Y | X = x_2)$$

(ii) Die Voraussetzung (a) ist z. B. unter den in den Theoremen 1 bis 3 genannten Bedingungen gegeben.



*Rasch-Homogenität*

$$P(Y_i = 1/U) = \frac{\exp(\xi - \kappa_i)}{1 + \exp(\xi - \kappa_i)} . \quad (18.11)$$

Diese erste Annahme heißt *Rasch-Homogenität*, weil damit angenommen wird, dass die Lösung aller betrachteten Items von einer einzigen unidimensionalen latenten (Fähigkeits-)Variablen, nämlich  $\xi$ , abhängt.

Eine zweite Annahme ist die *bedingte (oder lokale) stochastische Unabhängigkeit*

*Lokale**stochastische Unabhängigkeit*

$$P(Y_i = 1 \mid U, \mathbf{y}_{-i}) = P(Y_i = 1 \mid U) \quad (18.12)$$

*Fähigkeit der Person**und Schwierigkeit des Items*

jedes Items  $Y_i$  von den *anderen* Items *gegeben die Personvariable*  $U$ . Dabei bezeichnet  $\mathbf{y}_{-i} := (Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_m)$  den Vektor der anderen Items. Ziele der Anwendung eines solchen Modells sind zum einen die Schätzung der *Fähigkeit* der betrachteten Person und der damit verbundenen Unsicherheit, d. h. des damit verbundenen Standardschätzfehlers; zum anderen aber auch die Schätzung der *Schwierigkeit*  $\kappa_i$  des jeweiligen Items  $i$  und des damit verbundenen Standardschätzfehlers.

In der KTT hat dieses Modell seine Entsprechung im Modell essentiell  $\tau$ -äquivalenter Variablen. Dort wird allerdings die True-score-Variable  $\tau_i$ , nicht ihr Logit  $\ln [\tau_i / (1 - \tau_i)] = \ln (P(Y_i = 1 \mid U) / [1 - P(Y_i = 1 \mid U)])$  [so heißt das Argument der Exponentialfunktion in Gleichung (18.11)], in eine Personvariable  $\eta$  und einen Itemparameter  $\lambda_i$  zerlegt (s. Steyer & Eid, 2001).

Ein etwas weniger restriktives Modell ist das *Birnbaum-Modell*

*Birnbaum-Modell*

$$P(Y_i = 1/U) = \frac{\exp(\beta_i(\xi - \kappa_i))}{1 + \exp(\beta_i(\xi - \kappa_i))} , \quad (18.13)$$

das neben der Schwierigkeit  $\kappa_i$  auch die *Diskrimination*  $\beta_i$  als zweiten Itemparameter beinhaltet. Das Birnbaum-Modell hat in der KTT seine Entsprechung im Modell  $\tau$ -kongenerischer Variablen.

Will man auch Items betrachten, deren Lösung durch Raten möglich ist, so kommt das logistische *Drei-Parameter-Modell mit Rateparameter*

*Logistisches**Drei-Parameter-Modell  
mit Rateparameter*

$$P(Y_i = 1/U) = \gamma_i + (1 - \gamma_i) \cdot \frac{\exp(\beta_i(\xi - \kappa_i))}{1 + \exp(\beta_i(\xi - \kappa_i))} , \quad (18.14)$$

in Frage. Zu beiden letztgenannten Modellen gehört natürlich auch die Annahme der lokalen stochastischen Unabhängigkeit.

**18.2.2 Weiterführende Literatur**

Das Rasch-Modell wurde von Georg Rasch (1960) entwickelt und von Gerhard Fischer (1974) im deutschsprachigen Raum eingeführt. Neuere Lehrbücher dazu sind Amelang und Zielinski (1997), Rost (1996), Kumbinger (1987) sowie Steyer und Eid (2001). Eine klassische Einführung in eindimensionale Modelle der Item-response-Theorie bieten Hambleton

anz  $\sigma_{\xi}^2$  der Latent-trait-Variablen gleich den vier Kovarianzen der Observablen  $Y_{i1}$  und  $Y_{j2}$ .

### 18.3.3 Weiterführende Literatur

Als Einführungsartikel empfiehlt sich Steyer et al. (1992). Steyer et al. (1999) geben einen Überblick über verschiedene Forschungsfragen, die mit LST-Modellen untersucht werden können und über vorliegende Anwendungen in verschiedenen Gebieten der Psychologie. Die Beziehung zu Wachstumskurvenmodellen wird von Tisak und Tisak (2000) herausgearbeitet. Eid (1995) stellt eine Erweiterung der LST-Theorie für Probit-Modelle dar. Wichtig sind in diesem Kontext auch die von Eid (2000) entwickelten Modelle zur Behandlung von Methodenfaktoren, die dann nützlich werden, wenn nicht perfekt parallele Testformen innerhalb jeder Messgelegenheit vorgelegt werden oder wenn grundsätzlich verschiedene Erhebungsmethoden, wie z. B. Selbst- und Fremdrating, innerhalb jeder Messgelegenheit die gleiche Latent-state-Variable erfassen sollen.

## 18.4 Logistische Latent-state-trait-Modelle

Sowohl die Modelle der KTT als auch die der LST-Theorie setzen bereits vorliegende numerische Testwertvariablen voraus. In der Item-response-Theorie dagegen setzen die Modelle auf der Itemebene an und spezifizieren die Antwort- bzw. Lösungswahrscheinlichkeit in Abhängigkeit von der oder den als relevant erachteten latenten Variablen, wie z. B. die *Fähigkeit*, und von Parametern, welche die Items charakterisieren, wie z. B. die *Itemschwierigkeit*.

*Ansatz auf Item-Ebene*

Das bereits in Abschnitt 18.2.1 dargestellte Rasch-Modell lässt sich zu einem logistischen LST-Modells verallgemeinern, in dem man auch dem Problem situativer Effekte und der Interaktion zwischen Personen und Situationen Rechnung trägt. Dabei geht allerdings die Eindimensionalität der latenten Personvariablen verloren, wenn man mindestens zwei Messgelegenheiten betrachten will. Das in Zusammenfassungsbox 2 beschriebene Zufallsexperiment wird auch hier vorausgesetzt und es wird auch die dort eingeführte Notation verwendet.

*Ziel:  
zusätzliche Berücksichtigung  
situativer und interaktiver Effekte*

Das einfachste Modell lässt sich durch drei Annahmen definieren. Die erste Annahme,

$$P(Y_{it} = 1/U, S_t) = \frac{\exp(\xi + \zeta_t - \kappa_{it})}{1 + \exp(\xi + \zeta_t - \kappa_{it})}, \quad (18.26)$$

*Verallgemeinerte  
Rasch-Homogenität*

nennen wir die *Verallgemeinerte Rasch-Homogenität*. Dieser Annahme zufolge wird die Lösungswahrscheinlichkeit durch den Wert der Traitvariablen  $\xi$  und den situativen und/interaktiven Effekt, den Wert des Latent-state-Residuum  $\zeta_t$  sowie durch die Itemschwierigkeit  $\kappa_{it}$  determiniert. (In speziellen Fällen kann man deren Invarianz über die Zeit postulieren:  $\kappa_{it} = \kappa_i, t = 1, \dots, n$ .)



<http://www.springer.com/978-3-540-43911-0>

Wahrscheinlichkeit und Regression

Steyer, R.

2003, XIV, 329 S., Hardcover

ISBN: 978-3-540-43911-0