5 Modellwahl

5.1 Problemstellung

a Von der wissenschaftlichen **Fragestellung** und vom Vorwissen her gibt es verschiedene Arten, die Regressions-Analyse einzusetzen:

1. Im „Idealfall“ ist bereits klar, dass die Zielgrösse Y von den gegebenen Regressoren $X^{(1)}, \ldots, X^{(m)}$ linear abhängt. Man interessiert sich für eine klassische Fragestellung über die Koeffizienten der Regressoren, also für einen Test einer Nullhypothese (z. B. $\beta_j = 0$), eine Punkt- oder Intervallschätzung für einen oder mehrere Koeffizienten oder allenfalls für Vorhersage-Intervalle. Die entsprechenden Methoden haben wir behandelt.
2. Im anderen Extremfall dient die Studie dazu, Zusammenhänge zwischen der Zielgrösse Y und den Ausgangs-Variablen überhaupt erst zu erforschen. Man weiss nicht, ob und in welcher Form die Ausgangs-Variablen die Zielgrössen beeinflussen. Oft hat man dann für eine recht grosse Zahl potentieller Einflussgrössen „vorsorglich“ Daten erhoben.
3. Manchmal liegt die Fragestellung dazwischen:
   - Man ist eigentlich nur am Einfluss eines einzigen Regressors interessiert, aber unter Berücksichtigung der Effekte von anderen Ausgangs-Variablen (um indirekte Einflüsse zu vermeiden). Beispiel: Wirkung eines Medikamentes.
   - Man weiss einiges aus früheren Studien und aus theoretischen Überlegungen und will zusätzliche Erkenntnisse gewinnen.

In 2. und 3. stellt sich – in unterschiedlichem Ausmass – die Frage der **Modellwahl**: Welche Ausgangs-Variablen sollen in welcher Form in der Modell-Gleichung der linearen Regression erscheinen?

b ▷ ***Beispiel Baukosten von Atomkraftwerken**. Die Baukosten von 32 Kernkraftwerken, die in den Jahren 1967-71 in den USA entstanden, wurden untersucht (Quelle: Cox and Snell (1981)). Eine Fragestellung war, ob eine partielle Kostengarantie des Generalunternehmers zu Einsparungen führe. Als weitere erklärende Angaben für die Baukosten wurden die in Tabelle 5.1.b aufgeführten Variablen notiert. – Das Beispiel ist zwar schon in die Jahre gekommen, und die Anzahl Beobachtungen ist prekär klein. Es zeigt aber die Chancen und Schwierigkeiten der Modellwahl recht schön.*

c Erinnern Sie sich, dass die $x^{(j)}$ in der Modellgleichung $Y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \ldots + \beta_m x_i^{(m)} + E_i$ nicht unbedingt die **ursprünglich** beobachteten oder gemessenen Grössen, die wir zur Unterscheidung mit $u^{(k)}$ bezeichnen wollen, sein müssen; es können **transformierte** Grössen (z. B. $x^{(j)} = \log_{10}\langle u^{(j)} \rangle$) sein oder Funktionen von mehreren ursprünglichen Grössen (z. B. $x^{(j)} = u^{(k)} \cdot u^{(\ell)}$). Auch die Zielgrösse Y kann durch

Bez.	Bedeutung	Typ	Transf.
`K`	Baukosten	Betrag	log
`G`	Grösse	Betrag	log
`D`	Datum der Baubewilligung	kontin.	–
`WZ`	Wartezeit zwischen Antrag und Baubewilligung	Betrag	–
`BZ`	Bauzeit: Zeit bis Inbetriebnahme	Betrag	–
`Z`	Zweitwerk: früheres Werk auf gleichem Gelände	binär	–
`NE`	Werk steht im Nordosten der USA	binär	–
`KT`	Werk arbeitet mit Kühlturm	binär	–
`BW`	Reaktor hergestellt durch Babcock-Wilcox	binär	–
`N`	Anzahl Werke, die das gleiche Ingenieur-Team bereits erbaut hat, +1	Anzahl	Wurzel
`KG`	Partielle Kostengarantie des Generalunternehmers	binär	–

Tabelle 5.1.b: Die Variablen des Beispiels Baukosten

geeignete Transformation oder Standardisierung aus einer oder mehreren ursprünglich gemessenen Variablen gewonnen werden.

d ▷ *Im Beispiel führen allgemeine Überlegungen (siehe 4.4.g) zu den in Tabelle 5.1.b aufgeführten Transformationen als Ausgangsgrössen. Die Wartezeit und die Bauzeit wurden, obwohl es sich um Beträge (positive Zahlen) handelt, nicht logarithmiert, da es gemäss Zinseszins-Rechnung sinnvoll ist, einen linearen Einfluss dieser Zeiten auf die logarithmierten Kosten anzunehmen. Es sind auch andere Transformationen denkbar, und solche sollen ja auf Grund der Residuenanalyse immer wieder in Betracht gezogen werden.*

Das lineare Regressionsmodell mit allen transformierten Variablen, das „volle Modell", lautet im Beispiel also in Modellschreibweise

$$\texttt{log10(K)} \sim \texttt{log10(G)} + \texttt{D} + \texttt{WZ} + \texttt{BZ} + \texttt{Z} + \texttt{NE} + \texttt{KT} + \texttt{BW} + \texttt{sqrt(N)} + \texttt{KG}$$

oder ausführlich

$$\begin{aligned} \log_{10}\langle \texttt{K}_i \rangle &= \beta_0 + \beta_1 \log_{10}\langle \texttt{G}_i \rangle + \beta_2 \texttt{D}_i + \beta_3 \texttt{WZ}_i + \beta_4 \texttt{BZ}_i \\ &\quad + \beta_5 \texttt{Z}_i + \beta_6 \texttt{NE}_i + \beta_7 \texttt{KT}_i + \beta_8 \texttt{BW}_i + \beta_9 \sqrt{\texttt{N}_i} + \beta_{10} \texttt{KG}_i + E_i \;. \end{aligned}$$

e ▷ *Tabelle 5.1.e zeigt die Computer-Ausgabe für das Beispiel. Es können mindestens 5 Variable als überflüssig angesehen werden. Auch die Kostengarantie ist „schwach nicht-signifikant". Ist die Frage damit schon beantwortet? Wir werden das Beispiel noch weiter verfolgen. Schliesslich kann es um viel Geld gehen.*

Coefficients:

	Value	Std. Error	t value	Pr(> $\|t\|$)	Signif
(Intercept)	-6.02586	2.34729	-2.57	0.018	*
log10(G)	0.69254	0.13713	5.05	0.000	***
D	0.09525	0.03580	2.66	0.015	*
WZ	0.00263	0.00955	0.28	0.785	
BZ	0.00229	0.00198	1.16	0.261	
Z	-0.04573	0.03561	-1.28	0.213	
NE	0.11045	0.03391	3.26	0.004	**
KT	0.05340	0.02970	1.80	0.087	.
BW	0.01278	0.04537	0.28	0.781	
sqrt(N)	-0.02997	0.01780	-1.68	0.107	
KG	-0.09951	0.05562	-1.79	0.088	.

Tabelle 5.1.e: Computer-Ausgabe für das volle Modell im Beispiel Baukosten

5.2 Wichtigkeit eines einzelnen Terms

a Ist ein bestimmter Term $\beta_j x^{(j)}$ im Modell nötig? nützlich? überflüssig? – Die Beantwortung dieser Frage bildet einen **Grundbaustein für die Modellwahl**.

Als Hypothesen-Prüfung haben wir diese Frage schon gelöst: Wir wissen, wie man die Nullhypothese $\beta_j = 0$ prüft (mit dem t-Test). Diese Antwort tönt aber besser, als sie ist, denn es ergibt sich das Problem des **multiplen Testens**.

b Bei der Suche nach einem geeigneten Modell werden meistens einige bis viele Entscheidungen der erwähnten Art getroffen. Extremfall: Man habe 20 Regressoren („X-Variable"), und ein einziger Koeffizient sei „signifikant" (auf dem 5%-Niveau) von 0 verschieden. Dann entspricht das auf Grund der Wahrscheinlichkeit eines „Fehlers erster Art" der Erwartung für den Fall, dass überhaupt kein Regressor einen Einfluss auf Y hat!

c Dazu kommt ein weiteres, kleineres Problem: Man müsste die Voraussetzungen der Normalverteilung und der Unabhängigkeit der Fehler prüfen, wenn man die P-Werte der t-Tests zum Nennwert nehmen wollte.

d Man kann also nicht behaupten, dass ein Term mit signifikantem Test-Wert einen „statistisch gesicherten" Einfluss auf die Zielgrösse habe.

Statt die Tests für strikte statistische Schlüsse zu verwenden, begnügen wir uns damit, die P-Werte der t-Tests für die Koeffizienten (oder direkt die t-Werte) zu benützen, um die *relative* Wichtigkeit der entsprechenden Regressoren anzugeben, insbesondere um die „wichtigste" oder die „unwichtigste" zu ermitteln.

e Eine nominale Variable (ein „Faktor", also eine Variable mit mehreren möglichen Werten, die keine natürliche Ordnung zeigen) kann, wie in 3.2.e erklärt, in mehrere Indikator-Variable oder *dummy variables* verwandelt werden; wir reden von einem **Block von Indikator-Variablen**.

▷ *(Das Beispiel enthält (leider) keine nominale Variable. Die fünf binären Variablen sind zwar Indikator-Variable, aber nicht im Sinne der „dummy variables“ eines Faktors verknüpft.)*

Wenn gefragt wird, ob man eine nominale Ausgangs-Variable ins Modell einbeziehen soll oder nicht, muss man für den ganzen Block der entsprechenden Indikator-Variablen prüfen, ob alle weggelassen werden können. Das geschieht mit dem F-Test zum Vergleich von Modellen (3.2.m). Sein P-Wert kann mit den P-Werten der anderen Variablen „notfalls“ verglichen werden. (Besser eignet sich ein Vergleich mit den so genannten C_p-Werten, die in 5.3.g eingeführt werden.)

5.3 Automatisierte Verfahren zur Modellwahl

a Mit Hilfe eines Masses für die relative Nützlichkeit eines einzelnen Terms in der Regressionsgleichung können Strategien der Modellwahl formuliert werden:

- **Schrittweise rückwärts**. Man geht vom Modell aus, in dem alle in Frage kommenden Regressoren enthalten sind. (Das ist nur möglich, wenn die Zahl dieser Variablen kleiner ist als die Zahl der Beobachtungen – sie sollte bedeutend kleiner sein, sagen wir mindestens fünfmal kleiner.) Nun kann man schrittweise den „unwichtigsten“ wegnehmen, solange er unwichtig genug erscheint. Wo die entsprechende Grenze der „Wichtigkeit“, also des P-Wertes, liegen soll, ist kaum generell festzulegen. Die Schranke 0.05 für den P-Wert ist wegen des Problems des multiplen Testens nicht sinnvoller als andere (niedrigere) Werte.

b ▷ *Im **Beispiel der Baukosten** ist gemäss Tabelle 5.1.e die Variable* `WZ` *die unwichtigste. Wenn sie weggelassen wird, ergeben sich neue t- und P-Werte und damit eine neue Reihenfolge. Die P-Werte sind jetzt*

log10(G)	*0.000*	*Z*	*0.213*	*BW*	*0.852*
D	*0.000*	*NE*	*0.003*	*sqrt(N)*	*0.092*
BZ	*0.262*	*KT*	*0.082*	*KG*	*0.084*

Das Maximum zeigt die Variable `BW`*, die also als nächste zu eliminieren ist. So werden der Reihe nach zunächst die Variablen* `BW`, `BZ`, `Z`, $\sqrt{\texttt{N}}$ *und* `KT` *weggelassen. Nun ist, wie Tabelle 5.3.b zeigt, der Einfluss der Kostengarantie hochsignifikant. Also doch!*

Coefficients:

	Value	Std. Error	t value	$\Pr(>\lvert t\rvert)$	Signif
(Intercept)	-3.4612	1.1458	-3.02	0.005	**
log10(G)	0.6629	0.1295	5.12	0.000	***
D	0.0610	0.0160	3.82	0.001	***
NE	0.0831	0.0330	2.52	0.018	*
KG	-0.1844	0.0424	-4.35	0.000	***

Tabelle 5.3.b: Computer-Ausgabe für das durch schrittweise Elimination reduzierte Modell im Beispiel Baukosten

c • **Schrittweise vorwärts**. Analog zum schrittweisen Rückwärts-Verfahren kann man vom „leeren“ Modell (kein Regressor) zu immer grösseren kommen, indem man schrittweise einen zusätzlichen Term (einen Regressor oder einen Faktor in Form des entsprechenden Blockes von dummy Variablen) hinzunimmt, und zwar in jedem Schritt denjenigen, der (von den verbleibenden) am „wichtigsten“ ist. Dieses Verfahren hatte in den Anfangszeiten der multiplen Regression eine grundlegende Bedeutung, da es einen minimalen Rechenaufwand erfordert.

d ▷ *Im **Beispiel** zeigt die Kostengarantie* `KG` *die grösste einfache Korrelation mit den logarithmierten Baukosten und wird deshalb als erste Variable ins Modell aufgenommen! Es folgen* $\log_{10}\langle \texttt{G} \rangle$, `D`, `NE` *und* `KT`. *Der letzte Schritt führt zu einem formal nicht-signifikanten Koeffizienten. Wir lassen also* `KT` *wieder weg und haben das gleiche Modell wie vorher erreicht.*

Nun sind wir von der Bedeutsamkeit der Kostengarantie langsam überzeugt, nicht wahr?

e • **„Alle Gleichungen“ *(all subsets)***. Gehen wir wie beim Rückwärts-Verfahren von einem festen Satz von m möglichen Regressoren aus. Mit diesen Variablen lassen sich prinzipiell 2^m mögliche lineare Modell-Gleichungen bilden; man kann für jede Variable wählen, ob sie in der Gleichung erscheinen soll oder nicht. Der Computer kann alle möglichen Gleichungen an die Daten anpassen und nach einem geeigneten Kriterium die beste oder die paar besten suchen. (Intelligente Algorithmen vermeiden es, alle Gleichungen durchzurechnen.)

Im Folgenden bezeichnen wir die Anzahl Regressoren in einem in Frage stehenden Modell mit m'. Analog zu früher sei $p' = m'+1$, falls das Modell einen Achsenabschnitt β_0 enthält und $= m'$ im gegenteiligen Fall.

f Als **Kriterien** können die folgenden Grössen verwendet werden:

1. „Bestimmtheitsmass“ R^2 oder multiple Korrelation R,
2. Wert der Teststatistik für das gesamte Modell (F-Test),
3. zur F-Teststatistik gehöriger P-Wert,
4. geschätzte Varianz $\widehat{\sigma}^2$ der Fehler (oder Standardabweichung $\widehat{\sigma}$).

Für eine feste Anzahl m' von Regressoren führen alle diese (und auch die unten aufgeführten) Kriterien zur gleichen Ordnung unter den $\binom{m}{m'}$ möglichen Modellen (da jedes sich aus jedem andern – für festes m' – über eine monotone Funktion ausrechnen lässt); es werden also von allen die gleichen Modelle als die besten ausgewählt.

g Beim Vergleich zwischen **Modellen mit verschieden vielen Koeffizienten** gibt es Unterschiede:

Das Bestimmtheitsmass R^2 kann nicht abnehmen, wenn ein Term zur Modellgleichung hinzugefügt wird.

* Es misst ja im grösseren Modell das Quadrat der maximalen Korrelation zwischen Y und einer geschätzten Regressions-Funktion $\beta_0+\beta_{j_1}x^{(j_1)}+\ldots+\beta_{j_{m'+1}}x^{(j_{m'+1})}$. Die Variable $x^{(j_{m'+1})}$ weglassen heisst $\beta_{j_{m'+1}} = 0$ setzen. Das Maximum unter dieser Nebenbedingung kann nicht grösser sein als ohne Bedingung.

Trotzdem ist ein grösseres Modell ja nicht unbedingt besser als ein kleineres. Sonst wäre ja das vollständige Modell immer das beste. Es sind deshalb Kriterien vorgeschlagen worden, die automatisch auch unter Gleichungen mit verschieden vielen Termen eine sinnvolle Wahl der besten vornehmen:

5. **Korrigiertes Bestimmtheitsmass** R^2 (adjusted R^2): $R^2_{\text{adj}} = 1 - \frac{n-1}{n-p'}(1-R^2)$
6. C_p **von Mallows**. Dieses verbreitete Kriterium minimiert in gewisser Weise einen mittleren Vorhersagefehler. Es ist definiert als
$$C_{p'} := \text{SSQ}^{(E)}/\widehat{\sigma}_m^2 + 2p' - n = (n-p')(\text{MSQ}^{(E)}/\widehat{\sigma}_m^2 - 1) + p' \ ,$$
wobei $\text{MSQ}^{(E)} = \text{SSQ}^{(E)}/(n-p')$ das „mittlere Quadrat des Fehlers“ ist und $\widehat{\sigma}_m$ die Schätzung von σ im grössten Modell.
7. Das Informations-Kriterium AIC von Akaike (und Varianten davon). Es ist AIC $\approx$ C_p plus eine Konstante.

Diese Kriterien zeichnen jeweils ein Modell als das beste aus. Oft sind sie sich nicht einig in bezug auf die Anzahl Terme. Innerhalb der Gleichungen mit gleicher Anzahl Terme führen sie, wie erwähnt, zur gleichen Ordnung wie die erste Liste, sind sich also auch untereinander einig.

Häufig, aber nicht immer, ist jedes dieser „besten“ auch unter den Modellen zu finden, die die schrittweisen Verfahren liefern.

h Einen grafischen Überblick über die Modelle und die zugehörigen Kriterienwerte vermittelt ein Streudiagramm der Kriterienwerte gegen die Anzahl Koeffizienten p' im Modell (Abbildung 5.3.h). Da dies für das Kriterium C_p eingeführt wurde (Daniel and Wood, 1980) wird die Grafik „C_p-Plot“ genannt.

i ⊳ *Im Beispiel würden laut dem C_p-Kriterium zusätzlich zu den in Tabelle 5.3.b erwähnten Variablen noch* `KT` *und* $\sqrt{\texttt{N}}$ *ins Modell einbezogen. In diesem Modell beträgt der P-Wert für die Kostengarantie 0.049 – ein nur noch ganz knapp signifikantes Resultat also! Die Frage, ob die Kostengarantie zu Einsparungen führe, wird also verschieden beantwortet, je nach den zusätzlichen erklärenden Variablen im Modell. Wir kommen auf diesen Punkt zurück (5.5.g).*

j Das „beste“ Modell ist aber noch lange nicht das „richtige“ oder das „wahre“ Modell! Wenn man Daten auf Grund eines bestimmten Modells simuliert, werden (je nach Streuung der Fehler, Anzahl Beobachtungen, Grösse der Modell-Koeffizienten und „Verteilung“ der Regressoren, genannt „design“) mehr oder weniger oft andere Modelle als „beste“ ausgelesen. **Das „beste Modell“ wird also vom Zufall mitbestimmt!**
Deshalb soll man immer **mehrere Modelle in Betracht ziehen**, die von den Kriterien als „gut“ – nicht viel schlechter als das „beste“ – bewertet werden.

Wie viel schlechter? Leider gibt die Statistik darauf keine Antwort. (Eine kleine Hilfe ist der Test für einzelne Koeffizienten, siehe oben.)

k* Eher peinlich berührt es, zu erwähnen, dass die meisten Programme zur Modellwahl mit den in 5.2.e erwähnten **Blöcken von Indikator- oder dummy-Variablen** (und anderen Variablen-Blöcken) nicht richtig umgehen. Es werden die einzelnen Indikator-Variablen als völlig unzusammenhängend behandelt. Die „beste“ Gleichung enthält daher oft eine oder einige, aber nicht alle Indikator-Variablen eines Blocks – ein unsinniges Ergebnis.

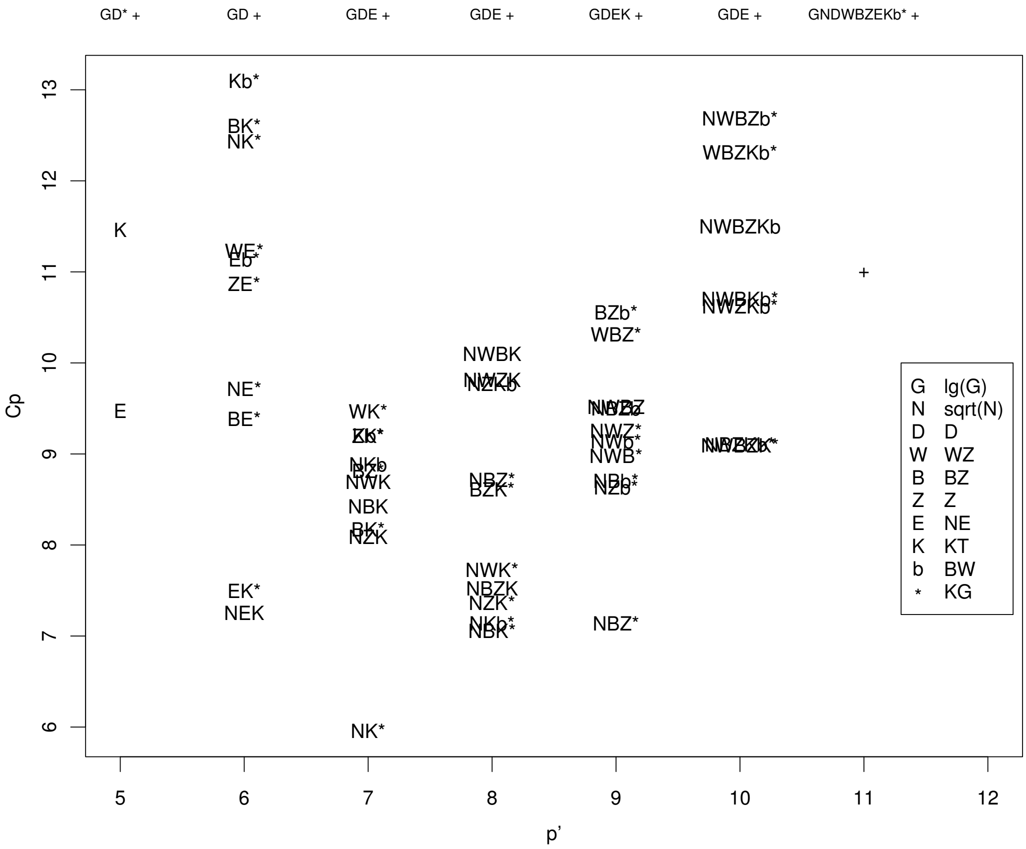

Abbildung 5.3.h: C_p-Plot für das Beispiel der Baukosten

l **Hohe Korrelationen zwischen Regressoren** oder allgemeinere Formen von **Kollinearität** führen zwar zu Problemen mit der Interpretation, sind aber von der Theorie her zugelassen. Im Vorwärts- und Rückwärts-Verfahren ist es in solchen Fällen häufig vom Zufall abhängig, welche der beteiligten Variablen als erste weggelassen respektive aufgenommen wird. Wenn alle Gleichungen untersucht werden, gibt es in diesem Fall jeweils Gruppen von ähnlich geeigneten.

Wir untersuchen diese Erscheinung im nächsten Abschnitt noch genauer.

m Als Ergebnis der Modellwahl kann man die Teilmenge der ausgewählten Terme aus allen Termen des vollständigen Modells ansprechen – eine zufällige Menge also. Wenn man die Daten leicht verändert, wird diese Teilmenge in gewissen Fällen sprunghaft ändern, indem beispielsweise ein Regressor $X^{(j)}$ wegfällt. Man kann auch sagen, der entsprechende Koeffizient β_j springe auf 0. Das ist keine wünschenswerte Eigenschaft. Es gibt deshalb Verfahren, für die die Koeffizienten stetig von den Daten abhängen.

n Die Idee des Verfahrens namens **Lasso** (siehe Hastie, Tibshirani and Friedman, 2001) besteht darin, das Kriterium „Kleinste Quadrate“, das ja bei der Bestimmung der Koeffizienten minimiert wird, durch einen „**Bestrafungsterm**“ für die Grösse der Koeffizienten zu versehen. Man spricht im Englischen von „penalized regression“. Damit die Grössen der Koeffizienten vergleichbar sind, benützt man standardisierte Koeffizienten β_j^* (siehe 3.1.m). Hier wird ausnahmsweise keine Quadratsumme als Mass der Grösse benützt, sondern die Summe der Absolutbeträge. Man minimiert also

$$Q\langle\underline{\beta};\lambda\rangle = \sum\nolimits_i R_i^2 + \lambda \sum\nolimits_j |\beta_j^*| \,.$$

Die Grösse λ steuert, wie stark die Grösse der Koeffizienten gegenüber der Residuen-Quadratsumme ins Gewicht fallen soll.

Man kann das Problem der Minimierung von Q auch formulieren als Minimierung der Quadratsumme der Residuen unter Einhaltung einer Schranke für die Grösse der Koeffzienten. Man minimiert also $\sum_i R_i^2$ unter einer Nebenbedingung der Form $\sum_j |\beta_j^*| < c$. Jeder Lösung dieses zweiten Problems, mit bestimmtem c, entspricht eine Lösung des ersten Problems mit einem gewissen λ, das von c abhängt. Die Gesamtheit der Lösungen für alle verschiedenen c im zweiten Fall ist also gleich der Gesamtheit der Lösungen für alle verschiedenen λ im ersten Fall.

Wenn c so gross ist, dass die Kleinste-Quadrate-Schätzungwerte $\widehat{\beta}_j$ die Nebenbedingung erfüllen, also $\sum_j |\widehat{\beta}_j| \le c$, dann ergibt sich keine Änderung. Wird c kleiner gewählt, dann werden die Koeffizienten demgegenüber verkleinert oder „gegen 0 geschrumpft“. Um c in einem sinnvollen Bereich zu wählen, setzt man deshalb besser $b = c\big/\sum_j |\widehat{\beta}_j|$ fest auf einen Wert zwischen 0 und 1.

Die Art der Nebenbedingung führt dazu, dass bald der erste Koeffizient exakt gleich 0 wird und mit kleineren c-Werten immer mehr Koeffizienten verschwinden. Dadurch entsteht eine Modellselektions-Reihe wie in einem schrittweisen Rückwärts-Verfahren.

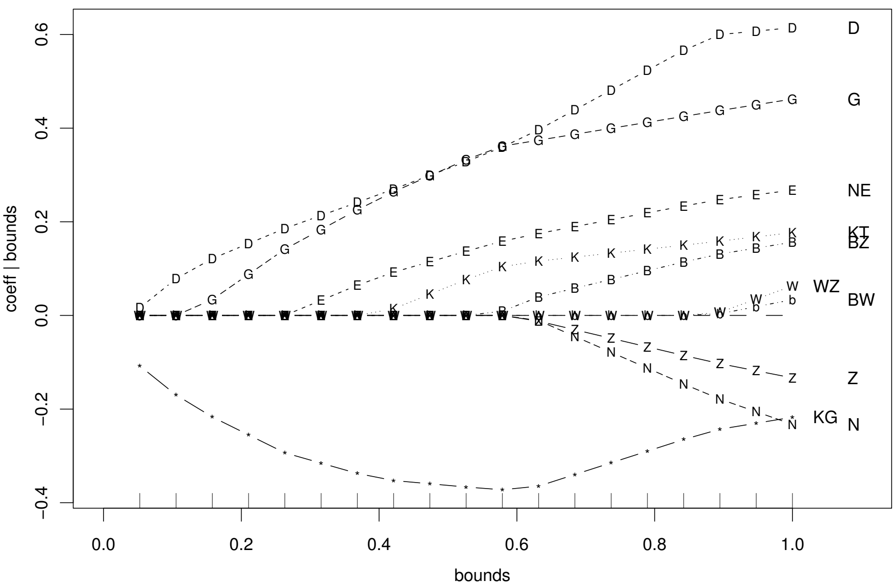

Abbildung 5.3.o: Lasso im Beispiel der Kernkraftwerke: standardisierte Koeffizienten in Abhängigkeit der relativen Schranke b

o ▷ *Abbildung 5.3.o zeigt, wie die geschätzten standardisierten Koeffizienten von der relativen Schranke b abhängen. Wenn man von $b = 1$ rückwärts geht, wird zunächst der Koeffizient von* `BW` *gleich 0, dann derjenige von* `WZ`*, dann* `Z`*, $\sqrt{\texttt{N}}$ und* `BZ`*. Ein merkwürdiges Verhalten zeigt ausgerechnet der Koeffizient der Kostengarantie* `KG`*: Er ist im Bereich von mittleren Schranken am bedeutendsten.*

5.4 Kollinearität

a Der Begriff der Kollinearität stammt aus der linearen Algebra. Das Modell lautete in Matrix-Schreibweise $\underline{Y} = \widetilde{\boldsymbol{X}}\underline{\widetilde{\beta}} + \underline{E}$ (3.4.d), und die Schätzung war $\underline{\widehat{\beta}} = \boldsymbol{C}^{-1}\widetilde{\boldsymbol{X}}^T\underline{Y}$ (3.4.g). Man braucht also die Inverse der Matrix $\boldsymbol{C} = \widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{X}}$.
Die Matrix $\boldsymbol{C}$ ist **singulär**, wenn die Spalten der **Design-Matrix** $\widetilde{\boldsymbol{X}}$ **kollinear** sind,

$$\begin{aligned}\boldsymbol{C} \text{ singulär} \quad &\Longleftrightarrow \quad \text{es gibt Zahlen } \underline{\widetilde{c}} = [c_0,\, c_1, \ldots, c_p] \text{ mit } \widetilde{\boldsymbol{X}}\underline{c} = \underline{0} \quad (\underline{c} \neq \underline{0}) \\ &\Longleftrightarrow \quad \text{es gibt ein } j \text{ und Zahlen } [c_0,\, c_1, \ldots, c_p] \text{ mit } \quad \widetilde{x}_i^{(j)} = \sum_{k\neq j} \widetilde{c}_k \widetilde{x}_i^{(k)} \;.\end{aligned}$$

In diesem Fall sind die Parameter im Modell nicht eindeutig zu bestimmen. Wegen

$$\widetilde{\boldsymbol{X}}\underline{\widetilde{\beta}} = \widetilde{\boldsymbol{X}}\,(\underline{\widetilde{\beta}} + \gamma\underline{c}) \quad \text{mit beliebigem } \gamma$$

gilt: Wenn $\underline{\widehat{\beta}}$ ein Schätzwert von $\underline{\widetilde{\beta}}$ ist, dann führt $\underline{\widehat{\beta}}+\gamma\underline{c}$ zu den gleichen Abweichungen $\underline{R}$ und ist deshalb ein gleich guter Schätzwert. Die Kleinste-Quadrate-Schätzung ist also nicht eindeutig, und etliche Programme steigen aus.

b Das Problem kann gelöst werden, indem man eine x-Variable, $x^{(j)}$, also eine Spalte in der Design-Matrix, streicht – falls die verbleibende Matrix immer noch singulär ist, streicht man eine weitere, usw. (Man muss jeweils eine Spalte $\widetilde{x}^{(j)}$ wählen, für die die erwähnte Gleichung $\widetilde{x}_i^{(j)} = \sum_{k\neq j} \widetilde{c}_k \widetilde{x}_i^{(k)}$ erfüllt ist.) Die Verteilungen, die das Modell beschreibt, bleiben damit eigentlich die gleichen, nur die Parametrisierung ändert, und damit die Interpretation der Parameter.

c Wenn solche lineare Beziehungen zwischen den x-Variablen nicht exakt, aber näherungsweise gelten, sind die Parameter zwar formell identifizierbar, aber „schlecht bestimmt“. Man spricht dann in der Statistik immer noch von **Kollinearität**.

Ein anschauliches einfaches Beispiel bilden zwei stark korrelierte x-Variable, z. B. $x^{(1)}$ und $x^{(2)}$. Abbildung 5.4.c zeigt einen solchen Datensatz.

d **Welches sind die Auswirkungen von Kollinearität?**
Im dargestellten Beispiel ist die Ebene, die dem linearen Regressionsmodell entspricht, in der einen Richtung, „entlang des Zauns“ gut, in der anderen (quer zum „Zaun“) schlecht bestimmt. Die Koeffizienten von $x^{(1)}$ und $x^{(2)}$, die Steigungen der Schnittgeraden der Ebene mit der „Aufriss-“ und „Seitenriss-Ebene“ ($x^{(1)}$-Y- und $x^{(2)}$-Y-Ebene), sind dann ebenfalls mit grosser Unsicherheit behaftet. Das führt zu grossen Standardfehlern für die geschätzten Koeffizienten. Deshalb kann man auf Grund des t-Tests (siehe 3.1.i) meistens die eine oder die andere Variable aus dem Modell streichen – aber oft nicht beide gleichzeitig!

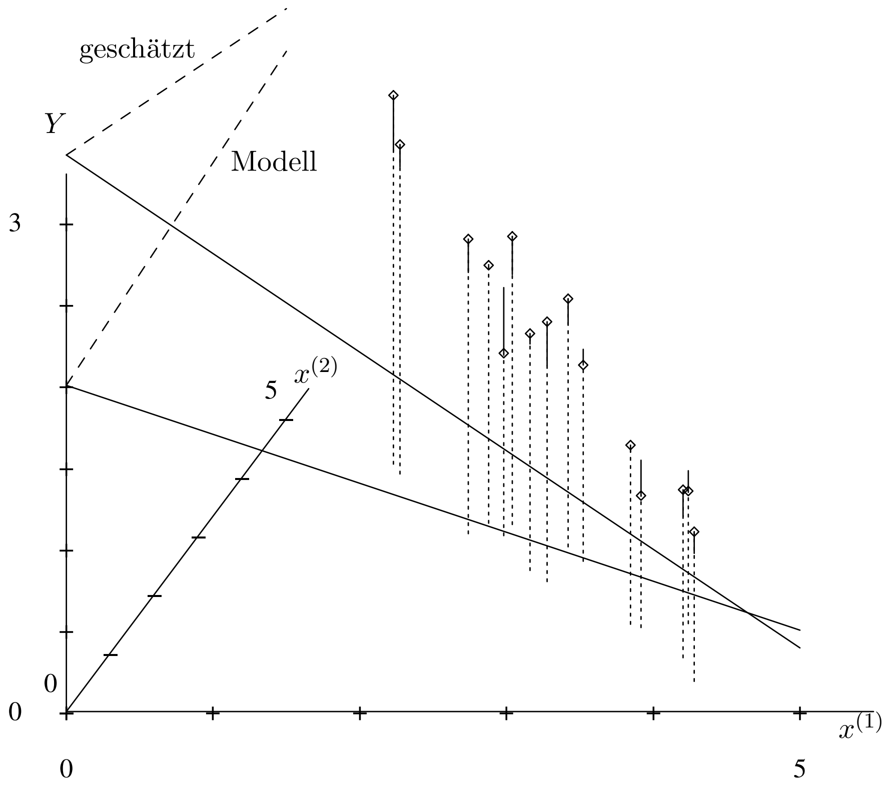

Abbildung 5.4.c: Kollinearität durch zwei stark korrelierte x-Variable. Die Y-Werte sind entsprechend dem „Modell“ simuliert. Eingezeichnet ist auch die „geschätzte“ Ebene.

e Die „Höhe“ der Ebene ist im Bereich der Daten mit der üblichen Genauigkeit durch diese bestimmbar, und in der Verlängerung des „Zauns“ recht gut extrapolierbar. An diesen Orten sind also Vorhersagen mit vernünftiger Genauigkeit anzugeben. Auf beiden Seiten des „Zauns“ nimmt aber die Genauigkeit rapide ab!

f **Wie entdeckt man Kollinearität?**
Die Probleme zeigen sich in den Standardfehlern, also auch in der Länge von Vertrauens-Intervallen und Prognose-Intervallen deutlich – sofern man darauf achtet!

Wir können aber auch direkter feststellen, ob eine Beziehung $\widetilde{x}_i^{(j)} \approx \widetilde{\sum}_{k \neq j} \widetilde{c}_k \widetilde{x}_i^{(k)}$ (annähernd) erfüllt ist. Das ist ein Regressionsproblem. Das Bestimmtheitsmass R_j^2 der Regression von $x^{(j)}$ auf alle übrigen erklärenden Variablen zeigt, wie stark eine solche Beziehung ist und ist also ein sinnvolles **Mass für Kollinearität**, das erst noch angibt, welche Variable „das Problem verursacht“.

Ein Mass, das man in Programmen findet, ist der so genannte **variance inflation factor** $\mathrm{VIF}_j = 1/(1 - R_j^2)$.

g **Was tun gegen Kollinearität?**
Wenn immer möglich, soll man die Beobachtungen so durchführen, dass das Problem vermieden wird. Bei Experimenten geben die x-Variablen die Versuchsbedingungen an. Kollinearität lässt sich durch geeignete Wahl der Versuchsbedingungen vermeiden.

h Können die Versuchsbedingungen nicht gewählt werden, dann kann man zu anderen X-Variablen übergehen, die besser bestimmte Koeffizienten ergeben.

Im Beispiel der beiden stark korrelierten Variablen ersetzt man diese durch ihre Summe und Differenz oder durch andere einfache Linearkombinationen, die nicht-kollineare neue Variable liefern.

Es gibt immer viele Möglichkeiten von linearen Transformationen, die zu „unkorrelierten“ x-Variablen führen. Für die Anwendung ist wesentlich, dass die neuen x-Variablen und damit ihre Koeffizienten leicht **interpretierbar** bleiben.

i Immer hilft das folgende Rezept:

- Die wichtigste Variable, sagen wir $x^{(1)}$, wird beibehalten;
- $x^{(2)}$ wird durch die Residuen einer Regression von $x^{(2)}$ auf $x^{(1)}$ ersetzt, also durch „den Teil von $x^{(2)}$, der von $x^{(1)}$ nicht erklärt wird“;
- Wenn die Kollinearität nicht von einem Paar von stark korrelierten Variablen stammt, sondern drei oder mehr Variable beteiligt sind, kann man allgemein die x-Variable mit dem höchsten R_j^2 wählen und durch Residuen bezüglich der Regression auf die anderen erklärenden Variablen ersetzen – und auch hier Modellwahl anwenden.

j Eine einfachere Lösung besteht darin, dass man die Variable mit dem höchsten R_j^2 aus dem Modell entfernt. (Das wird man oft auf Grund des t-Tests sowieso tun, siehe 5.3.l.)

k* In der Literatur wird auch ein Verfahren unter dem Namen „**ridge regression**“ vorgeschlagen. Ich finde es wenig hilfreich; die Ergebnisse sind schlecht interpretierbar.

5.5 Strategien der Modellwahl

a Die automatisierten Verfahren zur Modellwahl genügen für eine befriedigende explorative Analyse aus verschiedenen Gründen nicht:

- Wie erwähnt (5.3.j), ist die Auswahl der Variablen in der besten Gleichung entsprechend jedem Kriterium selbst **vom Zufall abhängig**, und man muss zumindest neben diesem „besten“ Modell die „**fast gleich guten**“ in Betracht ziehen. Um diese zu finden, ist die „all subsets“-Rechnung immerhin sehr hilfreich.
- Wir sind von einem festen Satz von Regressoren ausgegangen. Im Kapitel Residuen-Analyse haben wir gesehen, dass oft **Variable transformiert oder quadratische oder Wechselwirkungsterme eingeführt** werden sollten. Wollte man alle diese Möglichkeiten von Anfang an zum „festen Satz von Regressoren“ hinzufügen, dann würde dies schon bei wenigen ursprünglichen Ausgangs-Variablen zu einer übergrossen Zahl von Regressoren führen. Solche Zusatzterme müssen daher mit anderen Mitteln auf ihre Eignung geprüft werden.
- Manchmal liefern die Verfahren Modelle, die mit dem gesicherten **Fachwissen** nicht übereinstimmen. Beispielsweise kann der geschätzte Koeffizient eines Regressors ein Vorzeichen haben, das „nicht stimmen kann“. Bevor man eine ganz neue Theorie entwickelt, wird man weitere Modelle prüfen wollen.

b Zur Modellwahl braucht es also eine **„Strategie“**, die allerdings noch selten formuliert und diskutiert wird. Sie wird eher als Kunst angesehen, die allenfalls durch Beispiele zu vermitteln sei.

c Die Modellwahl findet innerhalb eines gesamten Ablaufs der Datenanalyse statt, deren „nullter“ Schritt immer lautet:

0. Daten kennenlernen und bereinigen. Man macht sich mit der genauen **Bedeutung aller Variablen** bekannt und legt kurze, informative **Variablennamen** fest, die alle Beteiligten gut interpretieren können. Dann überprüft man **unmögliche oder unplausible Werte** und Ausreisser für alle Variablen im Datensatz, korrigiert wenn nötig und setzt verbleibende unmögliche Werte auf „fehlend“. In diesem Zusammenhang bewährt es sich (wenn die Zahl der Variablen nicht allzu gross ist), die **Streudiagramm-Matrix** aller Variablen (mindestens der Variablen mit stetigem oder geordnetem Wertebereich) zu studieren.

Schliesslich untersucht man die Häufigkeiten und Auffälligkeiten des Auftretens von **fehlenden Werten.** Wenn sie mit spürbarer Häufigkeit auftreten, muss eine eigene Strategie zu ihrer Behandlung festgelegt werden, die wir hier nicht besprechen wollen.

Wer hier zu wenig investiert, büsst später!

d Wir werden sehen, dass die geeignete Strategie vom Zweck der Studie abhängt (vergleiche 5.1.a). Gehen wir zunächst davon aus, dass es der Zweck der Studie sei, **die erklärenden Variablen zu identifizieren**, die die Zielgrösse beeinflussen.

Dieses Ziel ist nicht so klar, wie es zunächst tönt. Am befriedigsten wäre es, die **Ursachen** für die Werte der Zielvariablen zu finden. Das ist aber mit einer explorativen Analyse von Daten nicht zu erreichen, sondern nur mit geplanten Versuchen, soweit solche möglich sind (siehe Versuchsplanung).

Es geht also darum, ein Modell zu finden, das die vorliegenden Daten gut beschreibt und möglichst keine systematischen Abweichungen übriglässt – die zufälligen sind nicht zu vermeiden.

e **Eine Strategie** zur Analyse solcher Daten kann etwa so aussehen:

1. “First aid” Transformationen. Allgemeine statistische Gesichtspunkte (4.4.g) und spezifisches Fachwissen führen für jede Variable zu einer plausiblen „Skala“ – oft einer transformierten ursprünglichen Grösse (englisches Stichwort *re-expression*).

2. Ein grosses Modell. Man passt eine Gleichung an, die vermutlich zu viele erklärende Variable enthält, nämlich

- alle Variablen, falls deren Anzahl höchstens einen Fünftel der Anzahl Beobachtungen ausmacht (* allenfalls setzt man gar ein „general additive model“ an),
- alle Variablen, die entsprechend Plausibilitäts-Überlegungen und Fachwissen einen Einfluss auf die Zielgrösse haben könnten,
- die Variablen, die mit einem „Schrittweise-Vorwärts-Verfahren“ mit grosszügigem Abbruchkriterium (hohem P-Wert) ausgewählt werden.

Falls gemäss Fachwissen Wechselwirkungen zwischen erklärenden Variablen erwartet werden, sollen diese ebenfalls einbezogen werden.

Wenn möglich sollten robuste Schätzmethoden verwendet werden.

3. Überprüfung des zufälligen Teils:

- Ausreisser in den Residuen,
- Verteilung der Residuen,
- Gleichheit der Varianzen,
- Unabhängigkeit der Fehler.

Es kann auf Grund der Ergebnisse angezeigt sein,

- die Zielgrösse zu transformieren,
- Gewichte einzuführen,
- robuste(re) Methoden zu verwenden, soweit dies nicht schon sowieso geschieht,
- Blöcke in der zeitlichen Abfolge (oder geographischen Anordnung) zu bilden und eine entsprechende nominale erklärende Variable einzuführen, um serielle Korrelationen mit dem funktionalen Teil statt mit korrelierten Fehlern E_i zu beschreiben,
- Schätzmethoden zu verwenden, die den Korrelationen Rechnung tragen.

Allerdings müssen die Modell-Voraussetzungen für das angegebene Analyse-Ziel nur grob erfüllt sein.

4. Nicht-Linearitäten. Streudiagramme der Residuen gegen die erklärenden Variablen können zu Transformationen der erklärenden Variablen oder zu quadratischen Termen führen.

5. Automatisierte Variablen-Wahl mit „all subsets“, notfalls mit schrittweisem Rückwärts-Verfahren.

6. Variable hinzufügen. Streudiagramme der Residuen gegen die erklärenden Variablen, die nicht im Modell sind – auch gegen jene, die gerade eliminiert wurden – und wie in Schritt 4 verfahren.

7. Einflussreiche Beobachtungen. Man sucht multivariate Ausreisser im Raum der x-Variablen, also hohe Leverage-Werte H_{ii}, und überprüft allgemein einflussreiche Beobachtungen (* mit robusten Methoden).

8. Kritik mit Fachwissen. Wenn das Modell Terme enthält, die unplausibel sind oder deren geschätzter Koeffizient das „falsche“ Vorzeichen hat, lässt man sie weg, sofern sich dadurch die Anpassung nicht allzu stark verschlechtert.

9. Anpassung prüfen. Man vergleicht die geschätzte Varianz der Fehler im Modell mit einer anderen Schätzung, beispielsweise einer minimalen, sicher vorhandenen Streuung (Messgenauigkeit) oder einer Schätzung aus wiederholten oder „benachbarten“ Messungen (4.8.a). Falls dieser Vergleich befriedigend ausfällt, kann man zu Schritt 12 gehen.

10. Wechselwirkungen. Man prüft, ob Wechselwirkungsterme zwischen den Variablen, die bereits im Modell sind, zur Verbesserung der Anpassung führen. Wechselwirkungen mit Variablen, die mangels Einfluss auf die Zielgrösse nicht ins Modell aufgenommen werden, sind unerwünscht und selten nützlich (siehe Cox and Snell, 1981, S. 126). Wenn solche ins Modell aufgenommen werden, nimmt man auch die beteiligten (nicht-signifikanten) erklärenden Variablen wieder ins Modell auf

11. Revision. Falls sich das Modell seit Schritt 4 merklich verändert hat, geht man dorthin oder gar zu Schritt 3 zurück.

12. Entfernte Terme überprüfen. Wenn in Schritt 8 Terme unterdrückt wurden, muss man nochmals überprüfen, wie wichtig sie jetzt erscheinen.

f Die Strategie soll sich nach dem Zweck der Studie richten. Die Absicht sei nun, **eine Hypothese zu überprüfen**, genauer wollen wir beispielsweise überprüfen, ob der Koeffizient von $x^{(1)}$ null sein kann.

Dann wird man die Strategie anpassen:

1. Daten-Transformation (soweit von der Fragestellung her zugelassen), wie oben.

2-7. In gewissen Fällen ist auch hier eine Modellwahl möglich oder nötig. Man folgt dann den Schritten 2-7 der vorhergehenden Strategie, aber mit „Nebenbedingungen“:

- $X^{(1)}$ bleibt immer im Modell,
- man kümmert sich nur um Variable, die eine merkliche Vergrösserung von R^2 bewirken oder die mit $X^{(1)}$ korreliert sind,
- eventuell ist die Transformation der Zielgrösse und von $X^{(1)}$ von der Fragestellung her nicht erlaubt.

8. **Kollinearitäten.** Genaue Überprüfung der X-Variablen im Modell, die mit $X^{(1)}$ korreliert sind („kritische“ X-Variable). **Aufgepasst:** Die Fragestellung selbst ändert sich, wenn man Variable ins Modell einbezieht, die mit der zu testenden Variablen korreliert sind. Die Beurteilung des Modells vom Fachwissen her ist daher hier unumgänglich.

9. **Annahmen über die Zufalls-Fehler** überprüfen. Gegebenenfalls muss man die Testmethode anpassen (generalized least squares, robuster Test, ...). Die Einhaltung der Voraussetzungen ist hier wichtig.

10. **Test-Resultate.** Man berechnet die P-Werte für die Modelle mit und ohne kritische Variable.

g ▷ *Im **Beispiel der Atomkraftwerke** liegt eine solche Fragestellung vor. Es soll ja herausgefunden werden, ob die Kostengarantie einen (vermindernden) Einfluss auf die Zielgrösse Kosten hat. Verschiedene Modelle haben zwiespältige Antworten geliefert. Die Variable `N`, die zählt, wie viele Werke das gleiche Ingenieur-Team bereits erbaut hat, ist eine „kritische“ Variable. Mit fachlicher Beurteilung kommt man zu einem überraschend klaren Ergebnis, das wir aber hier nicht ausführen wollen.*

h Ein dritter Zweck: **Vorhersage**. Hier ist noch keine Strategie formuliert. Es kommt bei dieser Fragestellung nur darauf an, gute angepasste Werte zu erhalten. Kollinearitäten sind unwichtig. Für Prognose-Intervalle ist die Form der Verteilung der Fehler wesentlich.

5.S S-Funktionen

a Die Wichtigkeit eines Terms in der Modellgleichung wird von `drop1` geprüft, siehe 3.S.0.f. Diese Funktion liefert nicht nur Test-Resultate (wenn man `test="F"` setzt), sondern (vor allem) einen AIC-Wert (5.3.g), der den Vergleich zwischen Modellen mit verschiedenen Anzahlen von Regressoren ermöglicht.

Analog zu `drop1` gibt es eine Funktion `add1`, die prüft, ob Terme zum bestehenden Modell hinzugefügt werden sollen.

b **Funktion `step`.** Die schrittweisen Verfahren sind in der Funktion `step` implementiert. Als erstes Argument verlangt `step` ein `lm`- (oder `regr`-) Resultat. Wenn nichts weiteres gegeben wird, dann läuft die Modellwahl schrittweise rückwärts. Man kann aber als Argument `scope=~.+X5+X6` zusätzliche Terme (`X5` und `X6`)angeben und auch festlegen, dass gewisse Terme in allen Modellen vorkommen müssen (`scope=list( lower=~X1, upper=~.+X5+X6)`). Will man ein Vorwärts-Verfahren vom „leeren Modell“ an durchführen, dann muss man zunächst „das leere Modell anpassen“, also `t.r <- lm(Y~1, data=...)` eingeben. Beispiel:

```
> t.r <- lm(K 1,data=d.nuk)
```

```
> t.rs <- step(t.r,
  scope=paste(" ",paste(names(d.nuk)[-1],collapse="+")))
```

Das schrittweise Verfahren stoppt, wenn die Grösse `AIC` nicht mehr abnimmt. Oft will man sehen, welche Variablen in weiteren Schritten eliminiert würden. Dazu kann man das Argument `k=100` benützen. Dann ist zwar `AIC` nicht mehr, was es sein soll, aber das Rückwärts-Verfahren läuft weiter, meistens bis zum leeren Modell.

c **Funktion `regsubsets`.** `library(leaps)` Ermöglicht die Prüfung aller Gleichungen (all subsets).

```
> t.ras <- regsubsets(K .,data=d.nuk,nbest=3)
```

```
> summary(t.ras)
```

Mit `nvmax=` maximale Anzahl Regressoren und mit `force.in=` kann man den Aufwand reduzieren und deshalb (noch) grössere Modelle verarbeiten.

```
> t.ras <- regsubsets(x=d.nuk[,-1], y=d.nuk[,"K"],
  force.in=c("G","D"), nvmax=8,nbest=3)
```

d **Funktion `update`.** Die Idee der Funktion `update` ist es, einzelne Modell-Spezifikationen ändern zu können und auf einfache Art eine neue Modell-Anpassung zu erwirken. Beispielweise führt

```
> update(t.r,formula= .-BW)
```

zu einem Modell, das sich von dem in `t.r` abgespeicherten Modell-Ergebnis nur dadurch unterscheiden, dass der Term `BW` im Modell weggelassen wird. – Allerdings kann es gerade so effizient und transparent sein, mit „copy-paste“ den vorhergehenden Aufruf von `lm` zu duplizieren und abzuändern.

e Die **Lasso-Methode** ist im package `library(lasso2)` unter dem Namen `l1ce` implementiert. Die Standardisierung der Variablen muss man selber organisieren. Das Argument `bound` legt die relative Schranke b fest (ausser man setzt `absolute.t=TRUE`). Man kann diesem Argument mehrere Werte geben (einen Vektor), beispielsweise `bound=seq(0.05,1,0.05)` und erhält dann eine ganze Liste von Regressionsresultaten. Mit `plot(...)` erhält man eine Darstellung der erhaltenen Koeffizienten in Abhängigkeit von der Schranke.

```
> t.r  <- l1ce(K .,data=t.d, bound=seq(0.05,1,0.05))
> plot(t.r)
> summary(t.r[[5]])
```

Literaturverzeichnis

Agresti, A. (1990). *Categorical Data Analysis*, Wiley, N.Y.

Agresti, A. (1996). *Introduction to categorical data analysis*, Wiley Series in Probability & Math. Statistics, Wiley, New York.

Christensen, R. (1990). *Log-linear models*, Springer, N.Y.

Cleveland, W. S. (1994). *The Elements of Graphing Data*, 2nd edn, Hobart Press, Summit, New Jersey.

Clogg, C. C. and Shihadeh, E. S. (1994). *Statistical models for ordinal variables*, Sage, Thousand Oaks, CA.

Cohen, A. (1980). On the graphical display of the significant components in a two-way contingency table, *Communications in Statistics – Theory and Methods* **A9**: 1025–1041.

Collet, D. (1991, 1999). *Modelling binary data*, Chapman & Hall/CRC Press LLC, Boca Raton, Florida.

Cook, R. D. and Weisberg, S. (1999). *Applied regression including computing and graphics*, Wiley, N.Y.

Cox, D. R. (1989). *Analysis of Binary Data*, 2nd edn, Chapman and Hall, London.

Cox, D. R. and Snell, E. J. (1981). *Applied Statistics*, Chapman and Hall, London.

Daniel, C. and Wood, F. S. (1980). *Fitting Equations to Data*, 2nd edn, Wiley, N.Y.

Davies, P. (1995). Data features, *Statistica Neerlandica* **49**: 185–245.

Devore, J. L. (1991). *Probability and Statistics for Engineering and the Sciences*, 3rd edn, Duxbury Press, Belmont, California.

Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*, 2nd edn, Chapman and Hall, London.

Draper, N. and Smith, H. (1998). *Applied Regression Analysis*, 3rd edn, Wiley, N.Y.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer-Verlag, New York.

Fox, J. and Monette, G. (1992). Generalized collinearity diagnostics, *Journal of the American Statistical Association* **87**: 178–183.

Fuller, W. A. (1987). *Measurement Error Models*, Wiley, N.Y.

Haaland, P. D. (1989). *Experimental Design in Biotechnology*, Marcel Dekker, N.Y.

Hartung, J., Elpelt, B. und Klösener, K. (1998). *Statistik. Lehr- und Handbuch der angewandten Statistik*, 11. Aufl., Oldenbourg, München.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer-Verlag, New York.

Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*, Wiley, N.Y.

Linder, A. und Berchtold, W. (1982). *Statistische Methoden II: Varianzanalyse und Regressionsrechnung*, Birkhäuser, Basel.

Lindsey, J. K. (1995). *Modelling Frequency and Count Data*, number 15 in *Oxford Statistical Science Series*, Clarendon Press, Oxford.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn, Chapman and Hall, London.

Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics*, Addison-Wesley, Reading, Massachusetts.

Myers, R. H., Montgomery, D. C. and Vining, G. G. (2001). *Generalized Linear Models. With Applications in Engineering and the Sciences*, Wiley Series in Probability and Statistics, Wiley, NY.

Ryan, T. P. (1997). *Modern Regression Methods*, Series in Probability and Statistics, Wiley, N.Y. *includes disk*

Sachs, L. (1997). *Angewandte Statistik*, 8. Aufl., Springer, Berlin.

Sen, A. and Srivastava, M. (1990). *Regression Analysis; Theory, Methods, and Applications*, Springer-Verlag, N.Y.

Stahel, W. A. (2000). *Statistische Datenanalyse: Eine Einführung für Naturwissenschaftler*, 3. Aufl., Vieweg, Wiesbaden.

Stahel, W. A. (2002). *Statistische Datenanalyse: Eine Einführung für Naturwissenschaftler*, 4. Aufl., Vieweg, Wiesbaden.

van der Waerden, B. L. (1971). *Mathematische Statistik*, 3. Aufl., Springer, Berlin.

Vincze, I. (1984). *Mathematische Statistik mit industriellen Anwendungen*, Band1, 2, 2. Aufl., Bibliograhisches Institut, Mannheim.

Weisberg, S. (1990). *Applied Linear Regression*, 2nd edn, Wiley, N.Y.

Wetherill, G. (1986). *Regression Analysis with Applications*, number 27 in *Monographs on Statistics and Applied Probability*, Chapmann and Hall, London.