3 Multiple lineare Regression

3.1 Modell und Statistik

a Die Abhängigkeit einer Zielgrösse von einer Ausgangsgrösse kann in einem einfachen Streudiagramm dargestellt werden. Oft wird dadurch das Wesentliche des Zusammenhangs sofort sichtbar. Die ganze Methodik der einfachen Regression wird dann nur noch zur Erfassung der Genauigkeit von Schätzungen und Vorhersagen gebraucht – in Grenzfällen auch zur Beurteilung, ob der Einfluss von X auf Y „signifikant“ sei.

Wenn der Zusammenhang zwischen einer Zielgrösse und **mehreren Ausgangsgrössen** $X^{(1)}, X^{(2)}, \ldots, X^{(m)}$ erfasst werden soll, reichen grafische Mittel nicht mehr aus. Das Modell der Regression lässt sich aber ohne Weiteres verallgemeinern zu

$$Y_i = h\left\langle x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(m)} \right\rangle + E_i \ .$$

Über die zufälligen Fehler E_i macht man die gleichen Annahmen wie früher. Für h ist die einfachste Form wieder die lineare,

$$h\left\langle x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(m)} \right\rangle = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \ldots + \beta_m x_i^{(m)} \ .$$

Sie führt zum Modell der **multiplen linearen Regression**. Die Parameter sind die so genannten **Koeffizienten** β_0, $\beta_1, \ldots, \beta_m$ der Ausgangs-Variablen und die Varianz σ^2 der zufälligen Abweichungen E_i. Die Koeffizienten β_1, $\beta_2, \ldots, \beta_m$ sind die „Steigungen in Richtung der x-Achsen“. Den „Achsenabschnitt“ (für die Y-Achse) bezeichnen wir mit β_0 statt mit α wie in der einfachen Regression; das wird später die Notation vereinfachen.

b ▷ *Im **Beispiel der Sprengungen** wurde nicht nur in unterschiedlicher Distanz vom Messort gesprengt, sondern es wurden auch verschiedene Ladungen verwendet (siehe Abbildung 1.1.b). Das multiple lineare Regressionsmodell mit $m = 2$ Ausgangs-Variablen lautet*

$$Y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + E_i \ .$$

Wieder ist eine lineare Beziehung nicht für die ursprünglichen Variablen, sondern – wenn schon – für die logarithmierten Werte plausibel. Wir verwenden also $Y = \log_{10}\langle \mathit{Erschütterung} \rangle$, $X^{(1)} = \log_{10}\langle \mathit{Distanz} \rangle$ und $X^{(2)} = \log_{10}\langle \mathit{Ladung} \rangle$. Eine Formulierung des Modells, die der Programmeingabe näher steht, lautet

$$\texttt{log10(ersch)}_i = \beta_0 + \beta_1 \texttt{log10(dist)}_i + \beta_2 \texttt{log10(ladung)}_i + E_i \ .$$

c Die übliche **Schätzung** der Koeffizienten β_j erfolgt wie in der einfachen Regression über die **Methode der Kleinsten Quadrate**. Ihre Verteilung ist mit Hilfe von Linearer Algebra nicht schwierig zu bestimmen(Anhänge 3.4 und 3.5), und darauf werden wieder Tests und Vertrauensintervalle aufgebaut. Auch die Streuung σ^2 wird auf die gleiche Weise wie vorher behandelt (siehe 2.2.n). Hier wollen wir sofort die Interpretation der Ergebnisse diskutieren.

d ▷ *Eine* ***Computer-Ausgabe*** *für das* ***Beispiel der Sprengungen*** *zeigt Tabelle 3.1.d. (Es wurden zunächst von den sechs Messorten nur die ersten vier berücksichtigt, die gut zueinander passen.) Die Tabelle enthält die Schätzungen der Koeffizienten in der Kolonne „Value", die geschätzte Standardabweichung des Fehlers und die nötigen Angaben für Tests, auf die wir gleich zurückkommen.*

Coefficients:	Value	Std. Error	t value	Pr(> |t|)	
(Intercept)	2.8323	0.2229	12.71	0.000	***
log10(dist)	-1.5107	0.1111	-13.59	0.000	***
log10(ladung)	0.8083	0.3042	2.66	0.011	*

St.dev. of Error = 0.1529 on 45 degrees of freedom
Multiple R-Squared: 0.8048
F-statistic: 92.79 on 2 and 45 degrees of freedom
p-value 1.11e-16

Tabelle 3.1.d: Computer-Output für das Beispiel der Sprengungen

e Bevor wir P-Werte interpretieren können, sollten wir überlegen, **welche Fragen** zu stellen sind. In den Beispielen könnten wir fragen (wenn es nicht so eindeutig wäre), ob die Distanz und die Ladung die Erschütterung, respektive die Basizität das Wachstum, überhaupt beeinflussen. Allgemeiner: Beeinflusst die **Gesamtheit der Ausgangsgrössen** die Zielgrösse? Die Nullhypothese lautet: „Alle β_j sind $= 0$." Den entsprechenden Test findet man in den beiden letzten Zeilen der Tabelle 3.1.d. Es wird eine Testgrösse gebildet, die eine F-Verteilung hat; man spricht vom F-Test.

Bei einer einzigen Ausgangsgrösse ist die Frage, ob sie einen Einfluss auf die Zielgrösse hat, mit dem Test der Nullhypothese $\beta = 0$ zu prüfen. Der „F-Test", der in Tabelle 2.3.e auch aufgeführt wird, gibt in diesem Fall immer die gleiche Antwort – ist äquivalent – zum t-Test, der dort besprochen wurde.

f* Die Testgrösse ist $T = \left(\mathrm{SSQ}^{(R)}/m\right)/\left(\mathrm{SSQ}^{(E)}/(n-p)\right)$. Dabei ist die „Quadratsumme der Regression" $\mathrm{SSQ}^{(R)} = \mathrm{SSQ}^{(Y)} - \mathrm{SSQ}^{(E)}$ die Differenz zwischen der „Quadratsumme der Zielgrösse" oder „totalen Quadratsumme"$\mathrm{SSQ}^{(Y)} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$ und der „Quadratsumme der Fehler" $\mathrm{SSQ}^{(E)} = \sum_{i=1}^{n} R_i^2$. Ferner ist $p = m+1$ die Zahl der Koeffizienten. Falls kein Achsenabschnitt β_0 im Modell erscheint, ist $p = m$ und $\mathrm{SSQ}^{(Y)} = \sum_{i=1}^{n} Y_i^2$. Die Freiheitsgrade der F-Verteilung sind m und $n-p$.

g ▷ *Etliche Programme liefern auch eine so genannte Varianzanalyse-Tabelle. Tabelle 3.1.g zeigt entsprechend ausführlichere Angaben für das* ***Beispiel der basischen Böden*** *(1.1.g). In dieser Tabelle wird der genannte F-Test in der Zeile „Regression" ausgewiesen; der P-Wert in dieser Zeile gibt Auskunft über die Signifikanz.*

Coefficients:

	Value	Std. Error	t value	Pr(> |t|)
(Intercept)	19.7645	2.6339	7.5039	0.0000
pH	-1.7530	0.3484	-5.0309	0.0000
lSAR	-1.2905	0.2429	-5.3128	0.0000

Residual standard error: $\widehat{\sigma} = 0.9108$ on $n - p = 120$ degrees of freedom
Multiple R-Squared: $R^2 = 0.5787$

Analysis of variance

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Regression	$m =$ 2	$SSQ^{(R)} = 136.772$	68.386	$T = 82.43$	0.0000
Residuals	$n - p = 120$	$SSQ^{(E)} =$ 99.554	$\widehat{\sigma}^2 = 0.830$		*P-Wert*
Total	122	$SSQ^{(Y)} = 236.326$			

Tabelle 3.1.g: Computer-Output für das Beispiel der basischen Böden mit Varianzanalyse-Tabelle und der im folgenden verwendeten Notation

h Die Grösse „Multiple R-Squared“ ist das Quadrat der so genannten **multiplen Korrelation**, der Korrelation zwischen den Beobachtungen Y_i und den **angepassten Werten** *(fitted values)*

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i^{(1)} + \widehat{\beta}_2 x_i^{(2)} + \ldots + \widehat{\beta}_m x_i^{(m)} .$$

Man kann zeigen, dass die nach Kleinsten Quadraten geschätzten Koeffizienten nicht nur die Quadratsumme der Residuen minimieren, sondern auch die Korrelation zwischen den angepassten Werten und den Beobachtungen der Zielgrösse maximieren; der maximale Wert ist die multiple Korrelation. Das Streudiagramm in Abbildung 3.1.h soll diese Korrelation veranschaulichen.

Die quadrierte multiple Korrelation wird auch **Bestimmtheitsmass** genannt, da sie den „durch die Regression bestimmten“ Anteil der Streuung der Y-Werte misst,

$$R^2 = SSQ^{(R)}/SSQ^{(Y)} = 1 - SSQ^{(E)}/SSQ^{(Y)} .$$

i Die Frage nach dem **Einfluss der einzelnen Variablen** $X^{(j)}$ muss man genau stellen. Der t-Wert und der P-Wert in derjenigen Zeile der Tabelle 3.1.d (oder des ersten Teils von 3.1.g), die $X^{(j)}$ entspricht, prüft, ob diese Variable aus dem Modell weggelassen werden kann, also ob die Nullhypothese $\beta_j = 0$ mit den Daten verträglich ist.

Die letzte Spalte der Tabelle enthält die übliche symbolische Darstellung der **Signifikanz**: Drei Sternchen *** für hoch signifikante Testergebnisse (P-Wert unter 0.1%), zwei Sternchen für P-Werte zwischen 0.1% und 1%, ein Sternchen für gerade noch signifikante Ergebnisse (1% bis 5 %), einen Punkt für nicht ganz signifikante Fälle (P-Wert unter 10%) und gar nichts für Zeilen mit P-Wert über 10%. Das erleichtert in grossen Tabellen das Auffinden von signifikanten Resultaten.

Im **Beispiel der basischen Böden** zeigt sich unter anderem, dass die zweite Art der

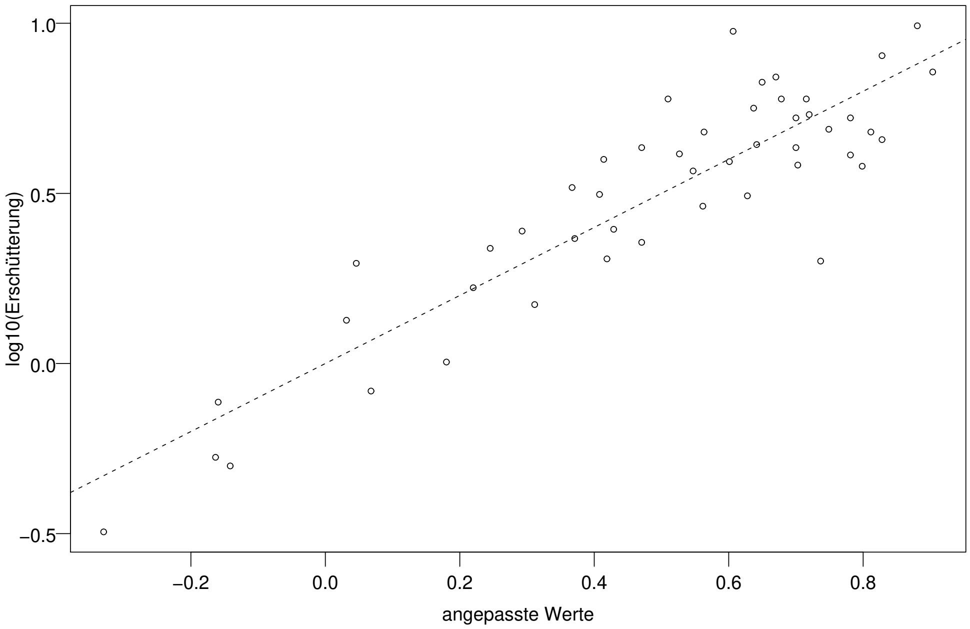

Abbildung 3.1.h: Streudiagramm der beobachteten und der angepassten Werte im Beispiel der Sprengungen

Erfassung der Basizität, also $X^{(2)}$, einen Teil der Variabilität von Y erfasst, der durch den pH-Wert $X^{(1)}$ nicht „erklärt" wird.

Die Frage, wie stark $X^{(2)}$ für sich allein, ohne Konkurrenz von $X^{(1)}$, mit Y zusammenhängt, lässt sich mit einer einfachen Regression beantworten und wird im Computer-Output der multiplen Regressionsrechnung nicht geprüft.

j Mit den Angaben der Tabelle lässt sich auch ein **Vertrauensintervall** für einen Koeffizienten β_j angeben. Es hat wie üblich die Form $\widehat{\beta}_j \pm q \;\mathrm{se}^{(\beta_j)}$, wobei $\widehat{\beta}_j$ und $\mathrm{se}^{(\beta_j)}$ in Tabelle 3.1.d unter „Value" und „Std. Error" zu finden sind, während der kritische Wert $q = q_{0.975}^{t_{n-2}}$ in einer Tabelle der t-Verteilung zu finden ist.

Einige Programme geben die Vertrauensintervalle direkt an.

k ▷ *Im **Beispiel der Sprengungen** erhält man für den Koeffizienten von* `log10(dist)` *das Vertrauensintervall* $-1.5107 \pm 2.014 \cdot 0.1111 = -1.5107 \pm 0.2237 = [1.2869, 1.7345]$. *Nun ist der Wert -2, den wir bisher als von der Theorie vorgegeben dargestellt haben, nicht mehr im Vertrauensintervall enthalten. Der Wert -2 entspricht der ungehinderten Ausbreitung der Energie in drei Dimensionen – die Energie ist dann umgekehrt proportional zur Kugeloberfläche und damit zum quadriereten Radius. Wenn die Energie an gewissen Schichten reflektiert wird, dann ist eine weniger starke Abnahme mit der Distanz plausibel.*

l In diesem Skript wird eine neue Grösse eingeführt, die einerseits die Spalte „t value" ersetzt und andererseits die Berechnung der Vertrauensintervalle erleichtert. Die t-Werte werden eigentlich nicht mehr gebraucht, um den Test auf $\beta_j = 0$ durchzuführen, da ja die p-Werte angegeben werden. Immerhin geben sie eine andere Art der „Stärke der Signifikanz“ an: Wenn sie wesentlich grösser als etwa 2 sind, dann ist der Effekt entsprechend stark gesichert, denn das 95 %-Quantil einer t-Verteilung mit nicht allzu wenigen Freiheitsgraden ist ungefähr 2. Vor allem für klar signifikante Effekte kann das eine quantitative Beurteilung erleichtern, da der p-Wert dann einfach „sehr klein“ wird.

Machen wir das exakt und führen den „**t-Quotienten**“ *(t-ratio)* ein,

$$\widetilde{T}_j = \frac{\widehat{\beta}_j}{\mathrm{se}^{(\beta_j)} \cdot q_{0.975}^{(t_k)}} = T / q_{0.975}^{(t_k)} \; .$$

Die Stärke der Signifikanz wird jetzt nicht mehr durch Vergleich mit „ungefähr 2“, sondern mit exakt 1 beurteilt; wenn $\widetilde{T}_j$ betragsmässig grösser als 1 ist, ist der Koeffizient signifikant. $\widetilde{T}_j$ sagt direkt, wie weit innerhalb oder ausserhalb des Vertrauensintervalls der Wert 0 liegt – im Verhältnis zur halben Länge des Intervalls. Ist der Wert 0.8, so liegt 0 innerhalb des Vertrauensintervalls, und zwar um 20% seiner halben Länge. Ist $\widetilde{T}_j = 1.2$, so liegt 0 um gleich viel ausserhalb des Intervalls. Anders ausgedrückt, ermöglicht $\widetilde{T}_j$, das Vertrauensintervall zu berechnen: Die halbe Breite des Intervalls ist $\widehat{\beta}_j / \widetilde{T}_j$ und deshalb das Vertrauensintervall selbst

$$\widehat{\beta}_j \cdot (1 \pm 1/\widetilde{T}_j) \; .$$

Tabelle 3.1.l zeigt eine Tabelle mit dieser Grösse, bezeichnet als „signif“ und wir erhalten das Vertrauensintervall für den Koeffizienten von `log10(dist)` aus $-1.511(1 \pm 1/6.75) = -1.511 \pm 0.224$, ohne das Quantil der t-Verteilung nachsehen oder abrufen zu müssen. Die Tabelle enthält ausserdem eine Spalt mit den „Freiheitsgraden“ (df), die im gegenwärtigen Zusammenhang immer gleich 1 sind, und zwei weiteren Grössen, die gleich noch erklärt werden.

Coefficients:

	coef	stcoef	signif	R2.x	df	p.value
(Intercept)	2.832	0.000	6.31	NA	1	0.000
log10(dist)	-1.511	-0.903	-6.75	0.01659	1	0.000
log10(ladung)	0.808	0.176	1.32	0.01659	1	0.011

St.dev. of Error = 0.1529 on 45 degrees of freedom
Multiple R-Squared: 0.8048
F-statistic: 92.79 on 2 and 45 degrees of freedom
p-value 1.11e-16

Tabelle 3.1.l: Resultat der S-Funktion `regr` für das Beispiel der Sprengungen

* Man könnte auch $1/\widetilde{T}_j$ als neue Grösse einführen und würde damit die Bildung des Kehrwertes bei der Berechnung des Vertrauensintervalls vermeiden. Das wäre aber als Mass für die Signifikanz ungeeignet, da ein schwacher Effekt zu einer unbegrenzten Zahl führen würde, während ein sehr stark gesicherter Effekt zu einer sehr kleinen Zahl führt.

m Eine weitere nützliche Grösse für jede X-Variable, die von einigen Programmen angegeben wird, ist der **standardisierte Regressions-Koeffizient** („stcoef" in der Tabelle)

$$\widehat{\beta}_j^* = \widehat{\beta}_j \cdot \mathrm{sd}\left\langle X^{(j)} \right\rangle / \mathrm{sd}\langle Y \rangle \ .$$

(sd steht für die Standardabweichung.) Es ist der Koeffizient, den man erhält, wenn man alle X-Variablen und die Zielgrösse auf Mittelwert 0 und Varianz 1 standardisiert und das Modell mit den neuen Grössen anpasst. In einer einfachen Regression ist die so standardisierte Steigung gleich der Korrelation. In der multiplen Regression messen die standardisierten Koeffizienten ebenfalls die Stärke des Einflusses der einzelnen Ausgangs-Variablen auf die Zielgrösse, unabhängig von den Masseinheiten oder Streuungen der Variablen. Ändert man $X^{(j)}$ um eine Standardabweichung $\mathrm{sd}\langle X^{(j)} \rangle$, dann ändert sich der geschätzte Wert der Zielgrösse um $\widehat{\beta}_j^*$ Standardabweichungen $\mathrm{sd}\langle Y \rangle$.

n* Schliesslich erscheint in der Tabelle unter der Spalte „R2.x" ein Mass für die so genannte Kollinearität zwischen den X-Variablen. Wenn eine X-Variable stark mit den anderen zusammenhängt, führt das zu Schwierigkeiten bei der Interpretation und zu grossen Ungenauigkeiten bei der Schätzung der betroffenen Koeffizienten. Genaueres folgt in 5.3.l und 5.4.

Das hier verwendete Mass für diese Schwierigkeit wird bestimmt, indem man die Regression jeder X-Variablen $X^{(j)}$ gegen alle anderen X-Variablen durchführt und das entsprechende Bestimmtheitsmass R_j^2 notiert. Auch wenn eine X-Variable, als Zielgrösse verwendet, allen Annahmen des entsprechenden Regressionsmodells widersprechen sollte, gibt das Bestimmtheitsmass einen brauchbaren Hinweis auf das Problem der Kollinearität. Der Minimalwert 0 sagt, dass $X^{(j)}$ mit den anderen Ausgangsgrössen nicht (linear) zusammenhängt. Das Maximum 1 tritt auf, wenn $X^{(j)}$ von den anderen X-Variablen vollständig linear abhängt. In diesem Fall tritt sogar ein numerisches Problem auf, da die Koeffizienten nicht mehr eindeutig schätzbar sind (wie in 3.2.f).

Ein häufig verwendetes Mass für die Kollinearität ist der „Variance Inflation Factor"(VIF), der gleich $1/(1 - R_j^2)$ ist. Sein Minimum ist 1; er kann beliebig gross werden.

3.2 Vielfalt der Fragestellungen

a Die Ausgangs-Variablen $X^{(1)}$ und $X^{(2)}$ sind in den Beispielen kontinuierliche Messgrössen wie die Zielvariable. Das braucht allgemein nicht so zu sein.

Im Modell der multiplen Regression werden keine einschränkenden Annahmen über die X-Variablen getroffen. Sie müssen von keinem bestimmten Datentyp sein und schon gar nicht einer bestimmten Verteilung folgen. Sie sind ja nicht einmal als Zufallsvariable eingesetzt.

b* Im Beispiel der basischen Böden sind die Bodenwerte wohl ebenso zufällig wie die Baumhöhen. Für die Analyse können wir trotzdem so tun, als ob die Basizität vorgegeben wäre. Eine formale Begründung besteht darin, dass die Verteilungen gemäss Modell als bedingte Verteilungen, gegeben die $x_i^{(j)}$-Werte, aufgefasst werden.

c Eine Ausgangs-Variable kann beispielsweise **binär**, also auf die Werte 0 und 1 beschränkt sein. Ist sie die einzige X-Variable, dann wird das Modell zu $Y_i = \beta_0 + E_i$ für $x_i = 0$ und $Y_i = \beta_0 + \beta_1 + E_i$ für $x_i = 1$. Das Regressionsmodell ist dann äquivalent zum Modell von zwei unabhängigen Stichproben, von denen ein allfälliger Unterschied der Lage interessiert – eine sehr übliche, einfache Fragestellung in der Statistik.

Das sieht man folgendermassen: Oft werden bei zwei Stichproben die Beobachtungen mit zwei Indices versehen: Y_{ki} ist die $i$te Beobachtung der $k$ten Gruppe ($k = 1$ oder 2) und $Y_{ki} \sim \mathcal{N}\langle \mu_k, \sigma^2 \rangle$. Es sei nun $x_{ki} = 0$, falls $k = 1$ ist, und $x_{ki} = 1$ für $k = 2$. Dann ist $Y_{ki} \sim \mathcal{N}\langle \beta_0 + \beta_1 x_{ki}, \sigma^2 \rangle$, mit $\beta_0 = \mu_1$ und $\beta_1 = \mu_2 - \mu_1$. Wenn man die Beobachtungen wieder mit einem einzigen Index durchnummeriert, ergibt sich das Regressionsmodell mit der binären x-Variablen.

d ▷ *Im **Beispiel der Sprengungen** wurde die Messstelle je nach Arbeitsfortschritt verändert. Es ist plausibel, dass die örtlichen Gegebenheiten bei den Messstellen einen Einfluss auf die Erschütterung haben.*

Betrachten wir zunächst den Fall von nur zwei Messstellen! Ein einfaches Modell lautet wie in 3.1.b

$$Y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + E_i \ ,$$

*wobei $X^{(1)}$ die logarithmierte Distanz sei und $X^{(2)}$ die binäre Variable, die die Messstelle bezeichnet, beispielsweise durch die Werte 0 für die erste und 1 für die zweite Messstelle. Das Modell beschreibt zwei Geraden $y = \beta_0 + \beta_1 x^{(1)}$ für die erste und $y = (\beta_0 + \beta_2) + \beta_1 x^{(1)}$ für die zweite Messstelle. Für beide Messstellen ist die gleiche Steigung β_1 wirksam; deshalb sind die beiden Geraden **parallel**. Dass die Geraden parallel sein sollen, ist eine Annahme, die in unserem Beispiel recht plausibel erscheint. Auf den allgemeineren Fall kommen wir zurück (3.2.u).*

e ▷ *Nun waren es aber vier Stellen, die wie üblich in einer willkürlichen Reihenfolge durchnummeriert wurden. Es ist sinnlos, die so entstehende Variable „Stellennummer" als Ausgangs-Variable $X^{(j)}$ ins Modell aufzunehmen, da eine* lineare *Abhängigkeit der Erschütterung von der Stellen-Nummer kaum plausibel ist.*

Eine solche Ausgangs-Variable mit **nominalem** oder **kategoriellem Wertebereich** wird auch **Faktor** genannt. Um sie in ein Regressionsmodell einzubeziehen, führt man für jeden möglichen Wert (jede Stelle) eine **„Indikatorvariable"** ein,

$$x_i^{(j)} = \begin{cases} 1 & \text{falls } i \text{ te Beobachtung aus der } j \text{ ten Gruppe,} \\ 0 & \text{sonst.} \end{cases} \quad .$$

Ein Modell für mehrere Gruppen j von Beobachtungen mit verschiedenen Erwartungswerten μ_j (aber sonst gleicher Verteilung) kann man schreiben als

$$Y_i = \mu_1 x_i^{(1)} + \mu_2 x_i^{(2)} + \ldots + E_i$$

mit unabhängigen, gleich verteilten E_i. Setzt man $\mu_j = \beta_j$, so steht das multiple Regressionsmodell da, allerdings ohne Achsenabschnitt β_0.

Eine binäre Variable, die eine Gruppenzugehörigkeit ausdrückt, wird als **dummy variable** bezeichnet. Eine nominale Ausgangs-Variable führt so zu einem **„Block" von dummy Variablen**.

f ▷ *Im Beispiel kommt dieser Block zu den beiden andern Ausgangs-Variablen hinzu (und die Nummerierung j der $X^{(j)}$ mag sich dadurch verändern). Das Modell kann man so schreiben:*

$$\begin{aligned}\texttt{log10(ersch)}_i &= \beta_0 + \beta_1 \texttt{log10(dist)}_i + \beta_2 \texttt{log10(ladung)}_i \\ &\quad + \gamma_1 \texttt{St1}_i + \gamma_2 \texttt{St2}_i + \gamma_3 \texttt{St3}_i + \gamma_4 \texttt{St4}_i + E_i\end{aligned}$$

g Ein technischer Punkt: In diesem Modell lassen sich die Koeffizienten prinzipiell **nicht eindeutig** bestimmen (vergleiche 3.4.h). Es verändern sich nämlich die „Modellwerte" $h\left\langle x_i^{(1)}, ...x_i^{(m)}\right\rangle$ nicht, wenn man zu allen γ_k eine Konstante dazuzählt und sie von β_0 abzählt. Eine so gebildete Kombination von Koeffizienten passt also sicher genau gleich gut zu den Beobachtungen. Man sagt deshalb, die Parameter seien **nicht identifizierbar**.

Um die Sache eindeutig zu machen, braucht man entweder **Nebenbedingungen** oder man lässt eine dummy Variable weg. Eine einfache Lösung besteht darin, $\gamma_1 = 0$ zu setzen oder, anders gesagt, die Variable `St1` nicht ins Modell aufzunehmen. (In der Varianzanalyse werden wir auf das Problem zurückkommen und auch andere Abhilfen diskutieren.)

h ▷ *Die numerischen Ergebnisse zeigt Tabelle 3.2.h. Die t- und P-Werte, die zu den „dummy" Variablen* `St2` *bis* `St4` *angegeben werden, haben wenig Bedeutung. Bei unserer Wahl von $\gamma_1 = 0$ zeigen sie, ob der Unterschied zwischen der entsprechenden Stelle und Stelle 1 signifikant sei.*

Coefficients:

	Value	Std. Error	t value	Pr(> |t|)	Signif
(Intercept)	2.51044	0.28215	8.90	0.000	***
log10(dist)	-1.33779	0.14073	-9.51	0.000	***
log10(ladung)	0.69179	0.29666	2.33	0.025	*
St2	0.16430	0.07494	2.19	0.034	*
St3	0.02170	0.06366	0.34	0.735	
St4	0.11080	0.07477	1.48	0.146	

Residual standard error: 0.1468 on 42 degrees of freedom
Multiple R-Squared: 0.8322
F-statistic: 41.66 on 5 and 42 degrees of freedom
the p-value is 3.22e-15

Tabelle 3.2.h: Computer-Ausgabe im Beispiel Sprengungen mit 3 Ausgangs-Variablen i ▷ *Um die Idee grafisch veranschaulichen zu können, unterdrücken wir die Variable* `ladung`*, indem wir nur Beobachtungen mit* `ladung`*=2.6 berücksichtigen. Abbildung 3.2.i zeigt die Beobachtungen und das angepasste Modell:* ***Für jede Stelle*** *ergibt sich* ***eine Gerade****, und da für die verschiedenen Stellen im Modell die gleiche Steigung bezüglich der Variablen* `log(dist)` *vorausgesetzt wurde, sind die angepassten Geraden* ***parallel****.*

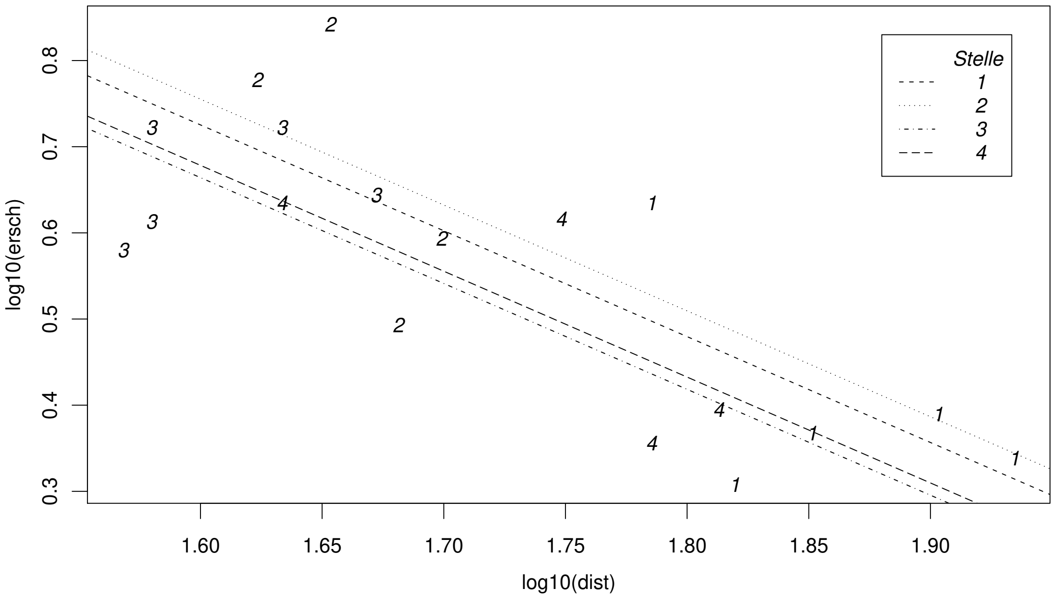

Abbildung 3.2.i: Beobachtungen und geschätzte Geraden im Beispiel der Sprengungen

j Es gibt eine sehr nützliche vereinfachte **Notation**, in der solche Modelle aufgeschrieben werden, die „**Modell-Formeln**". Das Modell im Beispiel wird geschrieben als

$$\texttt{log10(ersch)} \sim \texttt{log10(dist)} + \texttt{log10(ladung)} + \texttt{St}\ .$$

Die Indices, die Koeffizienten und der Fehlerterm werden weggelassen. Das Plus-Zeichen hat jetzt natürlich eine andere Bedeutung als üblich; es verbindet nicht mehr Zahlen, sondern Ausgangs-Variable – in ursprünglicher oder transformierter Form.

Die Sprache der Modell-Formeln eignet sich zur Eingabe in Programm-Pakete. Für die Variable `St` muss dem Programm bekannt sein, dass es sich um eine nominale Variable oder einen so genannten Faktor (siehe Varianzanalyse) handelt. Es konstruiert sich dann die entsprechenden dummy Variablen selber. `St` ist also ein **Term** in der Modell-Formel, der eine ganze Gruppe von X-Variablen umfasst, die in ihrer Bedeutung zusammengehören.

In einigen Programmen können in der Modellangabe keine Transformationen festgelegt werden. Man muss dann zuerst transformierte Variable `lersch`=`log10(ersch)` und analog `ldist` und `lladung` erzeugen. Das Modell lautet dann `lersch` $\sim$ `ldist` + `lladung` + `St`.

k Die Ausgangsgrössen erscheinen nun in verschiedenen Formen, die wir mit verschiedenen Ausdrücken bezeichnen wollen: Eine **Ausgangsgrösse** oder **Ausgangs-Variable** ist eine Grösse, von der angenommen wird, dass sie mit der Zielgrösse zusammenhängt, und für die deshalb eine geeignete Form gesucht wird, in der sie in das lineare Regressionsmodell einbezogen werden soll. Das kann in transformierter Form geschehen oder, wenn es eine nominale Variable ist, in Form mehrerer dummy-Variablen. Die X-Variablen, wie sie im linearen Modell erscheinen, nennt man auch **Regressoren**. Ein **Term** in der Modell-Formel kann ein einzelner Regressor sein oder eine Gruppe von zusammengehörigen Regressoren, die als Einheit betrachtet werden. Neben den Faktoren werden solche Gruppen vor allem Wechselwirkungen mit Faktoren sein, die bald eingeführt werden (3.2.t).

l Man wird die Frage stellen, ob die Messstelle (`St`) überhaupt einen Einfluss auf die Erschütterung habe. „Kein Einfluss“ bedeutet, dass die Koeffizienten aller entsprechenden Indikator-Variablen null sind, $\gamma_1 = 0$, $\gamma_2 = 0$, $\gamma_3 = 0$, $\gamma_4 = 0$. Den üblichen Test für diese Hypothese wollen wir allgemeiner aufschreiben.

m **F-Test zum Vergleich von Modellen**. Die Frage sei, ob die q Koeffzienten β_{j_1}, β_{j_2}, ..., β_{j_q} in einem linearen Regressionsmodell gleich null sein könnten.

- Nullhypothese: $\beta_{j_1} = 0$ und $\beta_{j_2} = 0$ und ... und $\beta_{j_q} = 0$
- Teststatistik:
$$T = \frac{(\mathrm{SSQ}^{(E)*} - \mathrm{SSQ}^{(E)})/q}{\mathrm{SSQ}^{(E)}/(n-p)} ;$$
$\mathrm{SSQ}^{(E)*}$ ist die Quadratsumme des Fehlers im „kleinen“ Modell, die man aus einer Regression mit den verbleibenden $m-q$ X-Variablen erhält, und p die Anzahl Koeffizienten im „grossen“ Modell ($= m+1$, falls das Modell einen Achsenabschnitt enthält, $= m$ sonst).
- Verteilung von T unter der Nullhypolthese: $T \sim \mathcal{F}_{q,n-p}$, F-Verteilung mit q und $n-p$ Freiheitsgraden.

Der Test heisst F-Test zum Vergleich von Modellen. Allerdings kann nur ein kleineres Modell mit einem grösseren verglichen werden, in dem alle X-Variablen des kleinen wieder vorkommen, also mit einem „umfassenderen“ Modell. Der früher besprochene F-Test für das gesamte Modell (3.1.e) ist ein Spezialfall: das „kleine“ Modell besteht dort nur aus dem Achsenabschnitt β_0.

n Zurück zur Prüfung des Einflusses einer nominalen erklärenden Variablen: Die besseren Programme liefern den entsprechenden Test gleich mit, indem sie in einer Tabelle den F-Test für die einzelnen Terme in der Modellformel zusammenstellen (Tabelle 3.2.n).

	Df	Sum of Sq	RSS	F Value	Pr(F)
log10(dist)	1	1.947	2.851	90.4	4.9e-12
log10(ladung)	1	0.117	1.022	5.44	0.025
Stelle	3	0.148	1.052	2.283	0.093

Tabelle 3.2.n: Tests für die Effekte der einzelnen Terme im Beispiel der Sprengungen

Für die ersten beiden erklärenden Variablen gibt diese Tabelle die gleiche Auskunft wie die vorhergehende (3.2.h). Der „F Value" ist gleich dem quadrierten „t value" von damals, und die entsprechenden Tests sind äquivalent. Die dritte Zeile vergleicht das umfassende Modell mit dem Modell ohne `St` als erklärende Variable. Sie zeigt, dass der Einfluss der Stelle nicht signifikant ist.

o* Achtung! Oft wird in einer genau gleich aussehenden Tabelle ein anderer Test durchgeführt, der im Allgemeinen wenig Bedeutung hat. Es wird nämlich in der eingegebenen Reihenfolge der Terme im Regressionsmodell schrittweise geprüft, ob der betreffende Term eine Verbesserung gegenüber dem vorhergehenden Modell, ohne diesen Term, bringt. Nur für den letzten Term in der Tabelle erhält man also den gewünschten Test.

p ▷ *Wenn kontinuierliche Variable und Faktoren als Ausgangsgrössen im Modell stehen, muss man üblicherweise die nützliche Information aus zwei verschiedenen Tabellen zusammensuchen: Aus Tabelle 3.1.d, liest man die Koeffizienten der kontinuierlichen Variablen ab und schaut sich auch ihren P-Wert für den Test gegen $\beta_j = 0$ an, und in der vorhergehenden Tabelle (3.2.n), die man extra verlangen muss, sucht man den P-Wert für die Faktoren. Das Resultat der Funktion* `regr` *zeigt beides in einer Tabelle (Tabelle 3.2.p). Die geschätzten Koeffizienten des Faktors erscheinen unterhalb der Haupttabelle.*

```
Call:
regr(formula = log10(ersch) ~ log10(dist) + log10(ladung) + Stelle,
     data = d.spreng14)
Terms:
                  coef  stcoef  signif    R2.x df p.value
(Intercept)    2.5104  0.0000  4.4090      NA  1   0.000
log10(dist)   -1.3378 -0.7993 -4.7106 0.24825  1   0.000
log10(ladung)  0.6918  0.1510  1.1555 0.02409  1   0.025
Stelle             NA      NA  0.8986 0.08884  3   0.093

Coefficients for factors:
$Stelle
     1      2      3      4
0.0000 0.1643 0.0217 0.1108

St.dev.error:  0.147   on 42 degrees of freedom
Multiple R^2:  0.832  Adjusted R-squared:   NA
F-statistic:   41.7   on 5 and 42 d.f.,  p.value: 3.22e-15
```

Tabelle 3.2.p: Ergebnisse der Funktion `regr` für das Beispiel der Sprengungen

q In den üblichen Darstellungen der Resultate (3.2.h) werden Koeffizienten für Faktoren in der gleichen Tabelle wie für kontinuierliche Variable gezeigt. Je nach „Codierung" sind diese aber nicht die Effekte γ_k der einzelnen Werte des Faktors (3.2.g), sondern kaum interpretierbare Grössen, die als Koeffizienten von erzeugten Variablen auftreten. Für die Koeffizienten werden dann, wie für die kontinuierlichen Variablen, t- und P-Werte angegeben, die nur bei geeigneter Codierung („treatment" oder „sum" in S) mit der entsprechenden Vorsicht sinnvoll zu interpretieren sind.

r* Die Spalte „signif" liefert für eine kontinuierliche Variable, wie beschrieben (3.1.l), das Verhältnis $\widetilde{T}_j$ zwischen dem geschätzten Koeffizienten und seiner Signifikanzgrenze. Die Grösse soll für Faktoren so definiert sein, dass sie eine ähnliche anschauliche Bedeutung erhält. Es sei (für irgendeinen Test) die „**z-ratio**" das Quantil der Standard-Normalverteilung, das dem P-Wert entspricht, dividiert durch den entsprechenden kritischen Wert $q^{(\mathcal{N})}\langle 0.95\rangle = 1.96$,

$$\widetilde{T} = q^{(\mathcal{N})}\langle 1-p\rangle \Big/ q^{(\mathcal{N})}\langle 0.95\rangle \ .$$

(Die t-ratio für kontinuierliche Variable ist zwar nicht genau gleich diesem Wert, aber für nicht allzu kleine Anzahlen von Freiheitsgraden sehr ähnlich.)

Fox and Monette (1992) verallgemeinern den Variance Inflation Factor für Faktoren. Hier wird dieser verallgemeinerte VIF verwendet und „in die R^2-Skala umgerechnet nach der Formel $R^2 = 1 - 1/\mathrm{VIF}$.

s* Allgemeinere Vergleiche von Modellen können nicht automatisch erfolgen, da es zu viele Möglichkeiten gibt und das Programm die interessanten kaum erraten kann. In umfassenden Programmen kann man die interessierenden Vergleiche angeben und erhält dann die gewünschten Testergebnisse. Sonst muss man sich die nötigen Quadratsummen aus zwei Computer-Ausgaben heraussuchen und mit der obenstehenden Formel den Wert der Testgrösse und den P-Wert bestimmen.

t Im Modell 3.2.f zeigt sich der Einfluss der Stelle nur durch eine additive Konstante. Der Wechsel von einer Messstelle zu einer anderen „darf" also nur zur Folge haben, dass sich die logarithmierten Erschütterungen um eine Konstante vergrössern oder verkleinern; die Geraden in 3.2.d müssen **parallel** sein. Es ist natürlich denkbar, dass der Zusammenhang zwischen Erschütterung einerseits und Distanz und Ladung andererseits sich zwischen den Stellen auf kompliziertere Art unterscheidet.

Eine nahe liegende Variante wäre, dass sich die Steigungskoeffizienten β_1 und β_2 für verschiedene Messstellen unterscheiden. Man spricht dann von einer **Wechselwirkung** zwischen Distanz und Stelle oder zwischen Ladung und Stelle. Das ist eine allgemeinere Frage als die folgende einfache, die immer wieder auftaucht.

u **Sind zwei Geraden gleich?** Oder unterscheiden sie sich im Achsenabschnitt, in der Steigung oder in beidem? Um diese Frage zu untersuchen, formulieren wir als Modell

$$Y_i = \alpha + \beta\, x_i + \Delta\alpha\, g_i + \Delta\beta\, x_i g_i + E_i$$

wobei g_i die „Gruppenzugehörigkeit" angibt: $g_i = 0$, falls die Beobachtung i zur einen Geraden, $g_i = 1$, falls sie zur anderen gehört. Für die Gruppe mit $g_i = 0$ entsteht die Gerade $\alpha + \beta x_i$, für $g_i = 1$ kommt $(\alpha + \Delta\alpha) + (\beta + \Delta\beta)x_i$ heraus. Die beiden Geraden stimmen in der Steigung überein, wenn $\Delta\beta = 0$ ist. Sie stimmen gesamthaft überein, wenn $\Delta\beta = 0$ und $\Delta\alpha = 0$ gelten. (Der Fall eines gleichen Achsenabschnitts bei ungleicher Steigung ist selten von Bedeutung.)

Das Modell sieht zunächst anders aus als das Grundmodell der multiplen Regression. Wir brauchen aber nur $x_i^{(1)} = x_i$, $x_i^{(2)} = g_i$ und $x_i^{(3)} = x_i g_i$ zu setzen und die Koeffizienten α, β, $\Delta\alpha$, $\Delta\beta$ als β_0, β_1, β_2, β_3 zu bezeichnen, damit wieder die vertraute Form dasteht.

Die Nullhypothese $\Delta\beta = 0$ lässt sich mit der üblichen Tabelle testen. Der Test für „$\Delta\alpha = 0$ und $\Delta\beta = 0$" ist ein weiterer Fall für den F-Test zum Vergleich von Modellen.

v Das Beispiel zeigt, dass die x-Variablen im Modell in irgendeiner Weise aus ursprünglichen erklärenden Variablen ausgerechnet werden können. So darf beispielsweise auch $X^{(2)} = (X^{(1)})^2$ sein. Das führt zur **quadratischen Regression**,

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + E_i \; .$$

Abbildung 3.2.v zeigt die Anpassung dieses Modells im Beispiel der basischen Böden (Beobachtungen mit pH>8.5 wurden weggelassen).
In gleicher Weise können auch höhere Potenzen eingeführt werden, was zur **polynomialen Regression** führt.

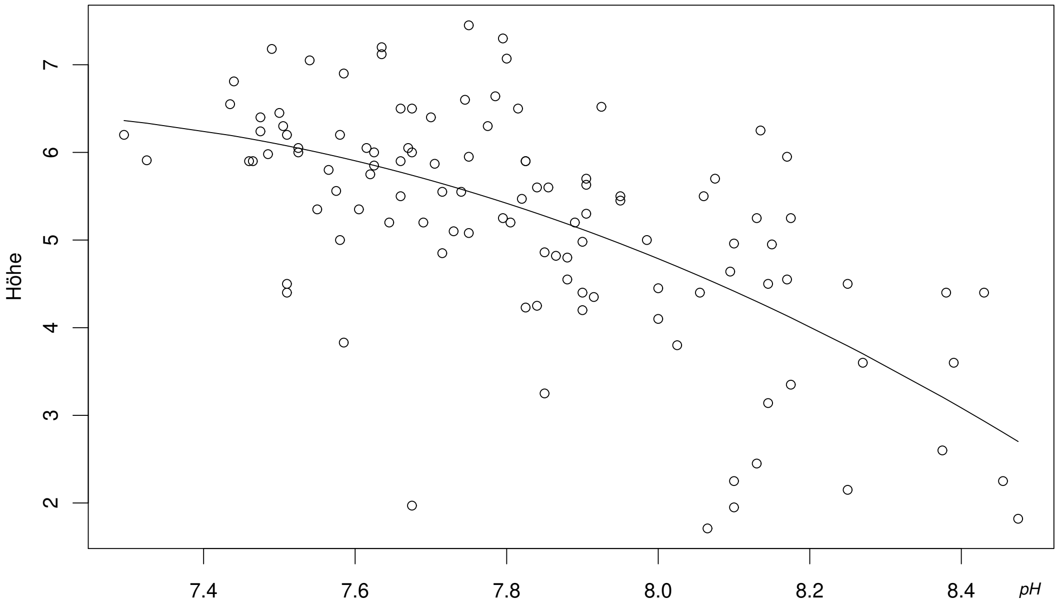

Abbildung 3.2.v: Quadratische Regression im Beispiel der basischen Böden

* Da jede glatte Funktion sich durch eine Polynom-Reihe annähern lässt, wird die polynomiale Regression oft eingesetzt, wenn man über die Art der Abhängigkeit zwischen einer erklärenden Variablen und einer Zielgrösse „keine" Annahmen treffen will. Es gibt dafür aber unter dem Stichwort **Glättung** oder **smoothing** oder **nichtparametrische Regression** geeignetere Methoden.

w Nun geraten die Begriffe durcheinander: Eine quadratische Regression wird als (multiple) lineare Regression bezeichnet! – **Das Wort *linear* im Begriff der multiplen linearen Regression bezieht sich *nicht* auf eine lineare Beziehung zwischen Y und den $X^{(j)}$, sondern darauf, dass die *Koeffizienten* linear in der Formel vorkommen!**

x Dieser Abschnitt hat gezeigt, dass das Modell der multiplen linearen Regression viele Situationen beschreiben kann, wenn man die X-Variablen geeignet wählt:

- Transformationen der X- (und Y-) Variablen können aus ursprünglich nichtlinearen Zusammenhängen lineare machen.
- Ein Vergleich von zwei Gruppen lässt sich mit einer zweiwertigen X-Variablen, von mehreren Gruppen mit einem „Block“ von dummy Variablen als multiple Regression schreiben. Auf diese Art werden nominale erklärende Variable in ein Regressionsmodell aufgenommen.
- Die Vorstellung von zwei verschiedenen Geraden für zwei Gruppen von Daten kann als ein einziges Modell hingeschrieben werden – das gilt auch für mehrere Gruppen. Auf allgemeinere Wechselwirkungen zwischen erklärenden Variablen kommen wir zurück (4.6.g).
- Die polynomiale Regression ist ein Spezialfall der multiplen linearen (!) Regression.

3.3 Multiple Regression ist viel mehr als viele einfache Regressionen

a Die multiple Regression wurde eingeführt, um den Einfluss mehrerer erklärender Grössen auf eine Zielgrösse zu erfassen. Ein verlockender, einfacherer Ansatz zum gleichen Ziel besteht darin, für jede erklärende Variable eine einfache Regression durchzuführen. Man erhält so ebenfalls je einen geschätzten Koeffizienten mit Vertrauensintervall. In der Computer-Ausgabe der multiplen Regression stehen die Koeffizienten in einer einzigen Tabelle. Ist das der wesentliche Vorteil?

Die Überschrift über diesen Abschnitt behauptet, dass der Unterschied der beiden Ansätze – mehrere einfache gegen eine multiple Regressionsanalyse – viel grundlegender ist. Das soll im Folgenden begründet werden.

b ▷ ***Modifiziertes Beispiel der Sprengungen.*** *Um Unterschiede der beiden möglichen Arten der Auswertungen zu demonstrieren, wurde der Datensatz der Sprengungen auf die Stellen 3 und 6 und Distanzen kleiner als 100 m eingeschränkt. Tabelle 3.3.b zeigt die numerischen Resultate der einfachen Regressionen der logarithmierten Erschütterung auf die logarithmierte Distanz und zum Vergleich das Resultat der multiplen Regression mit den erklärenden Variablen log(Distanz), log(Ladung) und Stelle.*

Die einfache Regression liefert einen völlig unplausiblen Wert für den Koeffizienten der logarithmierten Distanz, mit einem Vertrauensintervall von $[-0.1316 \pm 2.037 \cdot 0.3260] = [-0.80, 0.53]$. *Mit dem multiplen Modell ergibt sich für diesen Koeffizienten ein Intervall von* $[-0.72687 \pm 2.042 \cdot 0.35503] = [-1.45, -0.002]$, *das mit den Ergebnissen verträglich ist, die der gesamte Datensatz lieferte (3.2.h).*

```
----------------------------------------------------------------------
(i)
lm(formula = log10(ersch) ~ log10(dist), data = dd)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8976     0.5736   1.565    0.127
log10(dist)  -0.1316     0.3260  -0.404    0.689

Residual standard error: 0.2134 on 32 degrees of freedom
Multiple R-Squared: 0.00507,    Adjusted R-squared: -0.02602
F-statistic: 0.1631 on 1 and 32 degrees of freedom,     p-value: 0.689
----------------------------------------------------------------------
(ii)
lm(formula = log10(ersch) ~ log10(dist) + log10(ladung) + stelle,
   data = dd)
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.19297    0.58161   2.051  0.04908 *
log10(dist)        -0.72687    0.35503  -2.047  0.04947 *
log10(ladung)       1.49261    0.44162   3.380  0.00203 **
stelle6             0.16956    0.08604   1.971  0.05803 .

Residual standard error: 0.1813 on 30 degrees of freedom
Multiple R-Squared: 0.3269,     Adjusted R-squared: 0.2596
F-statistic: 4.856 on 3 and 30 degrees of freedom,  p-value: 0.00717
--------------------------------------------------------------------
```

Tabelle 3.3.b: Ergebnisse für die (i) einfache Regressionen der logarithmierten Erschütterung auf die logarithmierte Distanz und für die (ii) multiple Regression mit Distanz, Ladung und Stelle.

In Abbildung 3.3.b sind geschätzte Steigungen für die einfache Regression eingezeichnet – sowohl für beide Stellen zusammen als auch für die getrennte Auswertung. Die beiden weiteren, parallelen Geraden haben die Steigung, die sich aus der multiplen Regression ergibt, und geben die angepassten Werte für eine mittlere Ladung wieder. (Die Wechselwirkung zwischen log10(Distanz) und der Stelle, die einer unterschiedlichen Steigung der beiden Geraden entspricht, erwies sich als nicht signifikant.)

c ▷ *An künstlichen Beispielen lassen sich solche Effekte noch klarer veranschaulichen. In Abbildung 3.3.c sind für den Fall einer kontinuierlichen erklärenden Variablen $X^{(1)}$ und einer Gruppierungsvariablen $X^{(2)}$ vier mögliche Fälle aufgezeichnet. Die gestrichelten Geraden zeigen das Modell, nach dem die Beobachtungen erzeugt wurden: Zwei parallele Geraden mit Steigung β_1 und einem vertikalen Abstand von β_2. Die Beobachtungen der beiden Gruppen tragen verschiedene Symbole. Die ausgezogene Gerade stellt das Resultat einer einfachen Regression von Y auf $X^{(1)}$ dar; das schmale Rechteck am rechten Rand zeigt den Unterschied zwischen den Gruppenmittelwerten der Zielgrösse, was der einfachen Regression von Y gegen $X^{(2)}$ entspricht. Die Gerade und das Rechteck zeigen also das Resultat, das man erhält, wenn man die beiden Regressoren $X^{(1)}$ und $X^{(2)}$ je mit einfacher Regression „abhandelt“.*

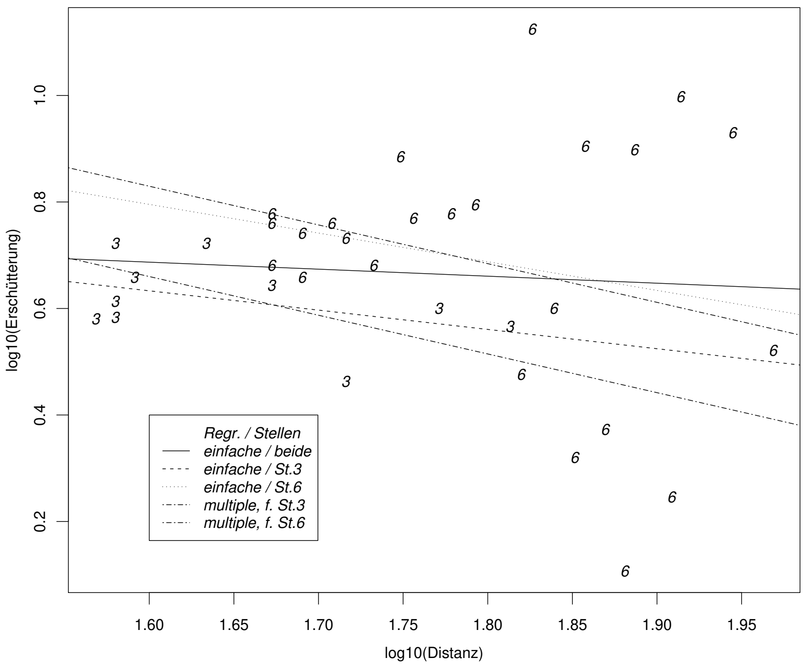

Abbildung 3.3.b: Daten des eingeschränkten Beispiels der Sprengungen (Stellen 3 und 6) mit geschätzten Regressionsgeraden: Die eingezeichneten Geraden stehen einerseits für die einfachen Regressionen, für beide Stellen zusammen wie auch separat gerechnet; andererseits erscheinen zwei parallele Geraden, die die angepassten Werte gemäss multipler Regression für eine mittlere Ladung für die beiden Stellen wiedergeben.

Die Ergebnisse der multiplen Regression sind nicht eingezeichnet; sie widerspiegeln das Modell ziemlich genau. Die vier Fälle zeigen die Schwierigkeiten der Interpretation von einfachen Regressionen drastisch:

(A) *Beide Variablen haben einen positiven Effekt, $\beta_1 > 0$, $\beta_2 > 0$. Die geschätzte Steigung und der Unterschied der Gruppenmittelwerte werden zu gross.*

(B) *Kein Effekt der kontinuierlichen erklärenden Variablen $X^{(1)}$. Die geschätzte Gerade erhält ihre Steigung durch den Unterschied zwischen den Gruppen.*

(C) *Entgegengesetzte Effekte, $\beta_1 < 0$, $\beta_2 > 0$. Die geschätzte Steigung zeigt einen positiven Effekt der kontinuierlichen erklärenden Variablen $X^{(1)}$ auf die Zielgrösse, während er in Wirklichkeit negativ ist!*

(D) *Hier sind die Effekte so eingerichtet, dass sie sich gegenseitig aufheben. Man wird fälschlicherweise schliessen, dass keine der beiden Variablen einen Einfluss auf Y hat.*

d Wenn wir uns das Modell der multiplen Regression vergegenwärtigen, wird klar, wie der Unterschied zu den Ergebnissen der einfachen Regression entsteht: Der Koeffizient β_1 beispielsweise gibt an, um wie viel sich der erwartete Wert der Zielgrösse erhöht, wenn $X^{(1)}$ um 1 erhöht wird – und alle anderen erklärenden Variablen gleich bleiben. Im Beispiel bleibt die Ladung und die Stelle gleich; wir erhalten also die Steigung der Geraden innerhalb der Stelle bei konstanter Ladung – und gehen, wenn die Wechselwirkung im Modell fehlt, davon aus, dass diese für beide Stellen gleich ist.

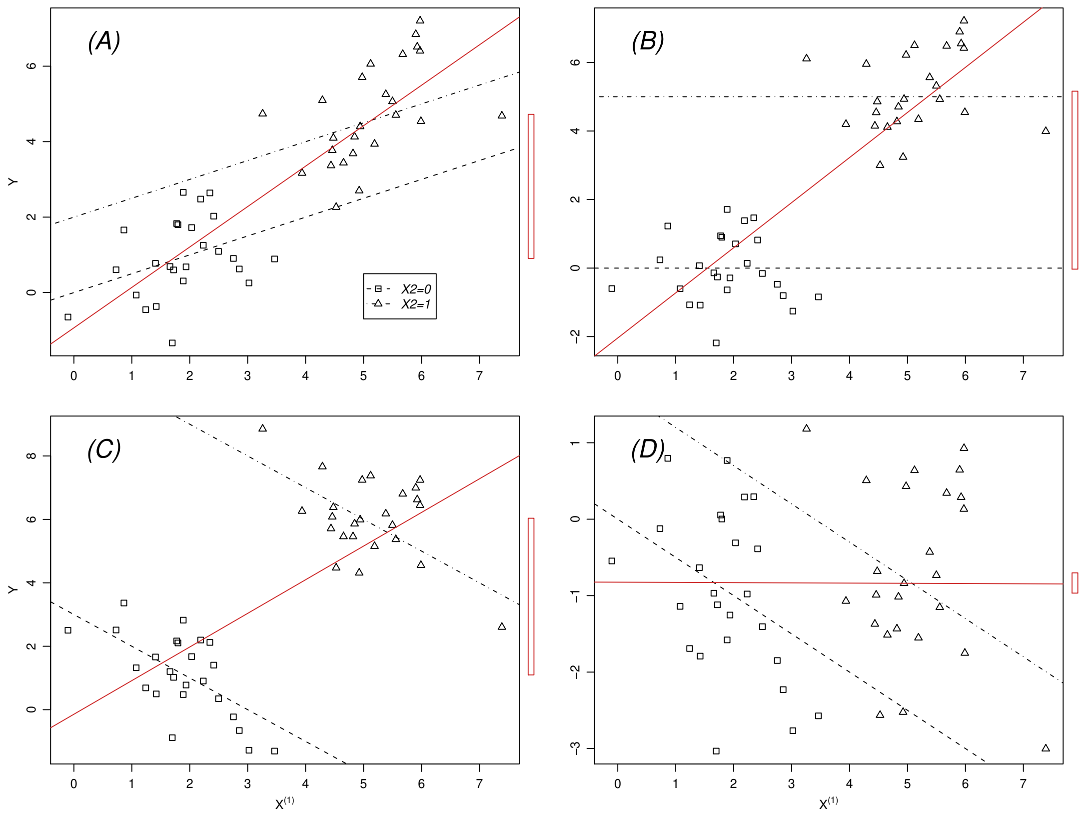

Abbildung 3.3.c: Einfache und multiple Regression für eine Gruppierungsvariable (binäre Variable) und eine kontinuierliche erklärende Variable

Betrachten wir die einfache Regression der Zielgrösse auf $X^{(1)}$, dann wird sich die Bedeutung von β_1 ändern. Die zweite ausgewählte Stelle wurde bei grösseren Distanzen erfasst als die erste und führte trotzdem tendenziell zu gleich hohen Erschütterungen. Teilweise lag das daran, dass auch stärker geladen wurde. Wenn $X^{(1)}$ um 1 erhöht wird, kommen im Datensatz tendenziell Beobachtungen mit höherer Ladung und anderer Stellenzugehörigkeit zum Zuge, und daher sinkt der Erschütterungswert kaum. Die Effekte der erklärenden Variablen werden vermischt.

e Ist eine kontinuierliche erklärende Variable $X^{(2)}$ mit $X^{(1)}$ positiv korreliert, dann wird sich bei einer Erhöhung von $X^{(1)}$ um 1 erwartungsgemäss auch $X^{(2)}$ erhöhen, was einen zusätzlichen Effekt auf die Zielgrösse hat. (* Der Effekt, ausgedrückt durch den Koeffizienten β_2 im multiplen Modell und dem „Regressionskoeffizienten von $X^{(2)}$ auf $X^{(1)}$, $\beta_{21} = \mathrm{cov}\langle X^{(1)}, X^{(2)}\rangle \,/\, \mathrm{var}\langle X^{(1)}\rangle$, beträgt $\beta_2\beta_{21}$.) Analoges gilt, wenn $X^{(1)}$ sich für die verschiedenen Werte einer nominalen erklärenden Grösse $X^{(2)}$ im Mittel wesentlich unterscheidet.

Diese Betrachtung zeigt allgemeiner, dass die **Bedeutung der Regressionskoeffizienten** prinzipiell davon abhängt, welche erklärenden Grössen im Modell auftreten.

Beachten Sie, dass wir vom Modell gesprochen haben, dass also dieses Problem nicht mit der Schätzung zusammenhängt.

f Grundlegend für alle Wissenschaften ist die Suche nach **Ursache-Wirkungs-Beziehungen**. Bekanntlich kann aus statistischen Korrelationen nicht auf solche Beziehungen geschlossen werden. Dennoch besteht eine wichtige Anwendung der Regression darin, Indizien für solche Beziehungen zu sammeln. Zwei Arten von Schlüssen sind üblich:

g Erste Schlussweise: Falls ein Koeffizient in einem Regressionsmodell **signifikant** von Null verschieden ist und eine ursächliche Wirkung der Zielgrösse auf die erklärende Grösse aus prinzipiellen Überlegungen heraus ausgeschlossen werden kann (die Erschütterung kann die Distanz zum Sprengort nicht beeinflussen!), dann wird dies als **Nachweis für eine vermutete ursächliche Wirkung** der erklärenden Grösse auf die Zielgrösse interpretiert.

h Oft kommt aber eine Korrelation zwischen einer erklärenden Variablen und der Zielgrösse dadurch zustande, dass **beide von einer dritten** Grösse Z verursacht werden.

Dies ist besonders häufig, wenn die Daten als **Zeitreihe** entstehen. Die Zahl der Neugeborenen hat im 20. Jahrhundert in den hochentwickelten Ländern abgenommen. Das lässt sich gut mit der Abnahme der Störche erklären... Die Zeit ist hier nicht die eigentliche Ursache der beiden Phänomene, sondern die Ursachen für den Niedergang der Anzahl Störche und der Anzahl Babies haben sich mit der Zeit ebenfalls verändert. Die Zeit kann dann die Ursachen in dieser Betrachtung (teilweise) vertreten.

Solche Situationen werden auch als **indirekte Zusammenhänge**, indirekte Korrelationen oder **Schein-Korrelationen** bezeichnet.

i Wenn die Grösse Z im Modell als erklärende Variable auftaucht, dann verfälschen die durch sie erfassten indirekten Wirkungen die Koeffizienten der anderen erklärenden Variablen nicht. Im Idealfall wird man also **alle denkbaren ursächlichen Variablen** für die betrachtete Zielgrösse als erklärende Variable **ins Modell aufnehmen**; dann stellt ein signifikanter Koeffizient von $X^{(1)}$ ein starkes Indiz für eine Ursache-Wirkungsbeziehung dar.

j Eine noch bessere Basis für eine solche Interpretation bilden, wenn sie möglich sind, **geplante Versuche**, in denen unter sonst gleichen Bedingungen nur die fragliche Variable $X^{(1)}$ variiert wird. Dann kann man die Wirkung direkt messen. Am überzeugendsten ist aber natürlich immer noch der konkrete **Nachweis eines Wirkungs-Mechanismus**.

k Zweite Schlussweise: Wenn ein Koeffizient **nicht signifikant** ist, wird dies oft als Nachweis betrachtet, dass die entsprechende erklärende Grösse **keinen Einfluss** auf die Zielgrösse habe. Dies ist in mehrfacher Hinsicht ein Fehlschluss:

- Wie bei allen statistischen Tests ist die Beibehaltung der Nullhypothese kein Beweis, dass sie gilt.
- Die vorher erwähnten Effekte von nicht ins Modell einbezogenen Einflussgrössen können auch dazu führen, dass eine ursächliche Wirkung durch indirekte Zusammenhänge gerade **kompensiert** wird (vergleiche das Beispiel!).
- Der Einfluss einer erklärenden Grösse kann nicht-linear sein. Dann kann man mit einer geeigneten Transformation (4.4, 4.6.c) oder mit Zusatztermen (4.6.d) zu einem genaueren Modell kommen.

l Die **am klarsten interpretierbare Antwort** auf die Frage nach einer Wirkung einer erklärenden Variablen auf die Zielgrösse erreicht man also, wenn man

- in einem geeignet **geplanten Versuch** die Variable gezielt verändert.

... oder, falls das nicht geht,

- möglichst alle denkbaren ursächlichen Grössen ins Modell aufnimmt,
- die Linearität der Zusammenhänge überprüft (siehe 4.4, 4.2.h),
- ein *Vertrauensintervall* für den Koeffizienten liefert – statt eines P-Wertes. Dieses gibt bei fehlender Signifikanz an, wie gross der Effekt dennoch sein könnte.

m Indirekte Effekte, wie sie hier als Gründe für falsche Interpretationen angeführt wurden, können nicht vorkommen, wenn die **erklärenden Grössen** selbst **nicht zusammenhängen** – wenigstens nicht linear – genauer: wenn sie „orthogonal" sind. Wir könnten von *unkorreliert* reden, wenn die erklärenden Grössen Zufallsvariable wären. „Orthogonal" heisst also: wenn wir trotz allem die empirische Korrelation zwischen den Variablen ausrechnen, so erhalten wir null. Wir kommen auf die Schwierigkeiten von „korrelierten" erklärenden Variablen in 5.4 zurück.

Wenn das möglich ist – namentlich bei geplanten Versuchen – ist deshalb sehr zu empfehlen, die $x_i^{(j)}$-Werte so zu wählen, dass die Orthogonalität erfüllt wird. Näheres wird in der Versuchsplanung besprochen.

n Wenn alle erklärenden Variablen in diesem Sinne orthogonal zueinander sind, dann kann man zeigen, dass die *Schätzungen* der Koeffizienten der einfachen Regressionen genau die geschätzten Werte des multiplen Modells geben müssen. Trotzdem lohnt sich das multiple Modell, da die geschätzte Standardabweichung der Fehler kleiner wird und daduch die **Vertrauensintervalle kürzer** und die **Tests eher signifikant** werden.

o Zusammenfassend: Ein multiples Regressionsmodell sagt mehr aus als viele einfache Regressionen – im Falle von korrelierten erklärenden Variablen sogar **viel mehr**.

3.4 Modell und Schätzungen in Matrix-Schreibweise

a Es ist Zeit, wieder etwas Theorie zu behandeln. Es wird sich lohnen, auch für praktisch orientierte Leute. Sie wollen ja nicht nur Rezepte auswendig lernen. Für Rezepte gibt es Bücher. Theorie stellt Zusammenhänge her. Etliche Probleme, die in der praktischen Anwendung der Regression auftreten können, lassen sich mit Hilfe der Theorie besser verstehen.

Die Theorie, die hier folgt, zeigt die Nützlichkeit von Linearer Algebra, von Matrizen und Vektoren. Sie werden die hier eingeführten Begriffe und Methoden in der multivariaten Statistik und bei den Zeitreihen wieder antreffen.

Bevor wir zufällige Vektoren und Matrizen betrachten, empfiehlt es sich, die gewöhnliche Vektor- und Matrixalgebra in Erinnerung zu rufen. Was für die folgenden Abschnitte wichtig ist, fasst Anhang 3.A zusammen.

b Das Modell der multiplen Regression, $Y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \ldots + \beta_m x_i^{(m)} + E_i$, wollen wir mit Hilfe von Vektoren und Matrizen formulieren.

Dazu müssen wir zuerst den Begriff des „Vektors von Zufallsvariablen“ oder der „vektoriellen Zufallsvariablen“ oder des „**Zufallsvektors**“ einführen: Es handelt sich einfach um eine Zusammenfassung von mehreren Zufallsvariablen,

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \text{und} \qquad \underline{E} = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix} .$$

Man verwendet also Spaltenvektoren. (Drucktechnisch platzsparender wären Zeilenvektoren, und deshalb schreibt man oft den transponierten Vektor hin, $\underline{Y} = [Y_1, \ldots, Y_n]^T$; ^{T} steht für transponiert.)

c Die Koeffizienten β_j können wir auch als Vektor schreiben, und die erklärenden Variablen $x_i^{(j)}$ zu einer Matrix zusammenfassen:

$$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} \qquad \text{und} \qquad \boldsymbol{X} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \ldots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & \ldots & x_2^{(m)} \\ \vdots & & & \vdots \\ x_n^{(1)} & x_n^{(2)} & \ldots & x_n^{(m)} \end{bmatrix} .$$

Schliesslich brauchen wir noch den Vektor, der aus lauter Einsen besteht, $\underline{1} = [1, 1, \ldots, 1]^T$. Jetzt wird das Regressionsmodell einfach zu

$$\underline{Y} = \beta_0 \underline{1} + \boldsymbol{X} \underline{\beta} + \underline{E} .$$

Was heisst das? Auf beiden Seiten des Gleichheitszeichens stehen Vektoren. Das i-te Element des Vektors rechts ist $\beta_0 \cdot 1 + \sum_j \beta_j x_i^{(j)} + E_i$, und das ist laut Modell gleich dem i-ten Element von $\underline{Y}$.

d Die Vektor-Gleichung ist noch nicht ganz einfach genug! Damit β_0 noch verschwindet, erweitern wir $\boldsymbol{X}$ um eine Kolonne von Einsen und $\underline{\beta}$ um das Element β_0:

$$\widetilde{\boldsymbol{X}} = [\underline{1}\ \boldsymbol{X}\,] = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \ldots & x_1^{(m)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \ldots & x_2^{(m)} \\ & \vdots & & & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & \ldots & x_n^{(m)} \end{bmatrix} \qquad \underline{\widetilde{\beta}} = \begin{bmatrix} \beta_0 \\ \underline{\beta} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}$$

Jetzt gilt

$$\underline{Y} = \widetilde{\boldsymbol{X}} \underline{\widetilde{\beta}} + \underline{E} .$$

Wenn das Modell keinen Achsenabschnitt enthält, setzen wir $\widetilde{\boldsymbol{X}} = \boldsymbol{X}$ und $\underline{\widetilde{\beta}} = \underline{\beta}$.

e Auf das Modell folgt die **Schätzung**. In der einfachen Regression haben wir das Prinzip der Kleinsten Quadrate angewandt. Die **Residuen**, die zu einem Parameter-Vektor $\underline{\widetilde{\beta}}^*$ gehören, sind

$$R_i = Y_i - (\beta_0^* + \sum\nolimits_j \beta_j^* x_i^{(j)}) \ .$$

Wir können auch sie zu einem Vektor zusammenfassen und erhalten

$$\underline{R} = \underline{Y} - \widetilde{\boldsymbol{X}} \underline{\widetilde{\beta}}^* \ .$$

(Wenn $\underline{\widetilde{\beta}}^* = \underline{\widetilde{\beta}}$ ist, sind die R_i gerade die Zufalls-Fehler E_i.)
Die Summe der Quadrate $\sum_i R_i^2$ kann man schreiben als

$$Q\langle\underline{\widetilde{\beta}}^*\rangle = \sum\nolimits_i R_i^2 = \underline{R}^T \underline{R}$$

(und das ist auch die quadrierte Norm des Vektors $\underline{R}$). Diesen Ausdruck wollen wir also minimieren. Dass dies aus dem Prinzip der Maximalen Likelihood folgt, wurde in 2.A.0.a gezeigt.

f Wir wollen dasjenige $\underline{\widetilde{\beta}}^*$ finden, für das $Q\langle\underline{\widetilde{\beta}}^*\rangle$ minimal wird, und es als Schätzung von $\underline{\widetilde{\beta}}$ verwenden. Eine klare Schreibweise für diese Aufgabe, die man vermehrt verwenden sollte, ist

$$\underline{\widehat{\beta}} = \arg\min_{\underline{\widetilde{\beta}}} \left\langle Q\langle\underline{\widetilde{\beta}}\rangle \right\rangle \ .$$

Minimieren läuft oft über Ableiten und null Setzen. Man kann Regeln für Ableitungen von und nach Vektoren herleiten und einsetzen. Wir kommen aber auch mit gewöhnlichen Ableitungen durch, wenn es auch etwas mühsam wird. Es ist

$$\partial Q\langle\underline{\widetilde{\beta}}\rangle / \partial\beta_j = \sum\nolimits_i \partial R_i^2 / \partial\beta_j = 2 \sum\nolimits_i R_i \partial R_i / \partial\beta_j$$

und

$$\partial R_i / \partial\beta_j = \partial \left(Y_i - (\beta_0 + \sum\nolimits_j \beta_j x_i^{(j)}) \right) \Big/ \partial\beta_j = -x_i^{(j)}$$

(wenn man $x_i^{(0)} = 1$ setzt, gilt dies auch für $j = 0$), also

$$\partial Q\langle\underline{\widetilde{\beta}}\rangle / \partial\beta_j = -2 \sum\nolimits_i R_i x_i^{(j)} = -2\, (\widetilde{\boldsymbol{X}}^T \underline{R})_j \ .$$

Die Ableitungen (für $j = 0, 1, \ldots, m$) sollen gleich 0 sein.

g Das können wir gleich als Vektor hinschreiben, $\widetilde{\boldsymbol{X}}^T \underline{R} = \underline{0}$. Einsetzen führt zu

$$\widetilde{\boldsymbol{X}}^T (\underline{Y} - \widetilde{\boldsymbol{X}}\underline{\widehat{\beta}}) = \underline{0} \quad \Rightarrow \quad \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}} \underline{\widehat{\beta}} = \widetilde{\boldsymbol{X}}^T \underline{Y} \;.$$

Die letzte Gleichung hat einen Namen: Sie heisst „die **Normal-Gleichung*en***" – es sind ja p Gleichungen, in eine Vektoren-Gleichung verpackt.

Links steht eine quadratische, symmetrische Matrix,

$$\boldsymbol{C} = \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}} \;,$$

multipliziert mit dem gesuchten Vektor $\underline{\widehat{\beta}}$, rechts ein Vektor, $\widetilde{\boldsymbol{X}}^T \underline{Y}$.

Bei der Auflösung dieser Gleichung macht sich die lineare Algebra erstmals richtig bezahlt: Wir multiplizieren die Gleichung von links mit der Inversen von $\boldsymbol{C}$, $\boldsymbol{C}^{-1}$, und erhalten

$$\underline{\widehat{\beta}} = \boldsymbol{C}^{-1} \widetilde{\boldsymbol{X}}^T \underline{Y} \;.$$

h Dazu müssen wir voraussetzen, dass $\boldsymbol{C}$ invertierbar oder nicht-singulär (oder regulär oder von vollem Rang) ist. Sonst? Sonst ist die Lösung des Problems der Kleinsten Quadrate nicht eindeutig, und man muss mit komplizierteren Methoden dahintergehen (mit verallgemeinerten Inversen).

Das Prinzip der Kleinsten Quadrate führt also nicht immer zu einer eindeutigen Lösung.

Das ist nicht nur ein theoretisches Problem! Wenn $\boldsymbol{C}$ nicht invertierbar ist, heisst das, dass das Regressions-Modell selbst schlecht formuliert ist, dass nämlich die Parameter nicht eindeutig sind, also verschiedene Parameter-Kombinationen genau das gleiche Modell festlegen. Man spricht von **nicht identifizierbaren Parametern**. Das Modell wird dann besser so geändert, dass man wieder eindeutig weiss, was ein Parameter bedeuten soll. (Einen solchen Fall haben wir in 3.2.g angetroffen.)

Das Problem kann auch „fast" auftreten. Wir kommen darauf unter dem Stichwort „Kollinearität" zurück (5.3.l).

i Schreiben Sie die letzte Formel für die einfache lineare Regression (2.2.c) auf und zeigen Sie, dass sie mit 2.2.c übereinstimmt! Das ist nützlich, um die allgemeinere Formel besser zu verstehen und um etwas lineare Algebra zu üben.

3.5 Verteilung der geschätzten Regressionskoeffizienten

a Die geschätzten Regressionskoeffizienten lassen sich also in Matrixform sehr kurz schreiben,

$$\underline{\widehat{\beta}} = \widetilde{\boldsymbol{C}} \underline{Y} \;, \qquad \widetilde{\boldsymbol{C}} = \boldsymbol{C}^{-1} \widetilde{\boldsymbol{X}}^T \;.$$

Wenn wir jetzt ein Element $\widehat{\beta}_j$ des Vektors $\underline{\widehat{\beta}}$ herausgreifen, so lässt sich dieses also auch als Summe ausdrücken,

$$\widehat{\beta}_j = \sum\nolimits_{i=1}^{n} \widetilde{\boldsymbol{C}}_{ji} Y_i \;.$$

Die $\widetilde{\boldsymbol{C}}_{ji}$ sind feste Zahlen, die Y_i Zufallsvariable. Wie in der Einführung über Wahrscheinlichkeitsrechnung gezeigt wird, ist eine solche „Linearkombination" von normalverteilten Zufallsvariable wieder normalverteilt, und es bleibt noch, den Erwartungswert und die Varianz zu bestimmen.

b Der Erwartungswert ist gemäss der allgemeinen Formel $\mathcal{E}\langle\sum_i a_i Y_i\rangle = \sum_i a_i\,\mathcal{E}\langle Y_i\rangle$ gleich

$$\mathcal{E}\langle\widehat{\beta}_j\rangle = \sum\nolimits_{i=1}^{n} \widetilde{\boldsymbol{C}}_{ji}\,\mathcal{E}\langle Y_i\rangle = \sum\nolimits_{i=1}^{n} \widetilde{\boldsymbol{C}}_{ji} \sum\nolimits_k X_i^{(k)}\beta_k \; .$$

Das sieht sehr kompliziert aus. Wir nehmen wieder die Matrixrechnung zu Hilfe. Die Doppelsumme ist gleich dem $j$ten Element von

$$\widetilde{\boldsymbol{C}}\,\boldsymbol{X}\underline{\beta} = \boldsymbol{C}^{-1}\widetilde{\boldsymbol{X}}^T\boldsymbol{X}\underline{\beta} = \boldsymbol{C}^{-1}\boldsymbol{C}\underline{\beta} = \underline{\beta} \; ,$$

also gleich β_j.

c Für die Varianz einer Summe von unabhängigen Zufallsvariablen lautet die allgemeine Formel $\mathrm{var}\langle\sum_i a_i Y_i\rangle = \sum_i a_i^2\,\mathrm{var}\langle Y_i\rangle$. Einsetzen ergibt

$$\mathrm{var}\langle\widehat{\beta}_j\rangle = \sum\nolimits_{k=1}^{n}\left(\widetilde{\boldsymbol{C}}_{jk}\right)^2 \mathrm{var}\langle Y_k\rangle = \sigma^2\sum\nolimits_{k=1}^{n}\left(\widetilde{\boldsymbol{C}}_{jk}\right)^2 \; .$$

Die Summe der Quadrate ist gleich dem $j$ten Diagonalelement von

$$\begin{aligned}\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{C}}^T &= \boldsymbol{C}^{-1}\widetilde{\boldsymbol{X}}^T(\boldsymbol{C}^{-1}\widetilde{\boldsymbol{X}}^T)^T = \boldsymbol{C}^{-1}\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{X}}(\boldsymbol{C}^{-1})^T \\ &= \boldsymbol{C}^{-1}\boldsymbol{C}(\boldsymbol{C}^{-1})^T = (\boldsymbol{C}^{-1})^T \; .\end{aligned}$$

Da $\boldsymbol{C}$ symmetrisch ist (und wir sowieso nur die Diagonalelemente betrachten), kann man das Transponieren weglassen. Also ist

$$\mathrm{var}\langle\widehat{\beta}_j\rangle = \sigma^2\left(\boldsymbol{C}^{-1}\right)_{jj} \; .$$

d Mit etwas mehr Theorie kann man auch Kovarianzen zwischen den geschätzten Koeffizienten $\widehat{\beta}_j$ erhalten. Diese Überlegungen gehören zum Thema der Multivariaten Statistik und werden im entsprechenden Block behandelt.

3.A Anhang: Grundbegriffe der Linearen Algebra

a **Matrizen.** Matrix, genauer $n \times m$-Matrix:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

Zeilen $i = 1, ..., n$, Spalten $j = 1, ..., m$. Elemente a_{ij}.

Quadratische Matrix: Gleiche Anzahl Zeilen und Spalten, $n = m$.

Symmetrische Matrix: Es gilt $a_{ij} = a_{ji}$.

Diagonale einer quadratischen Matrix: Die Elemente $[a_{11}, a_{22}, ..., a_{nn}]$.

Diagonalmatrix: Eine, die „nur aus der Diagonalen besteht", $d_{ij} = 0$ für $i \neq j$.

$$\boldsymbol{D} = \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_{nn} \end{bmatrix}$$

b **Transponierte Matrix**: Wenn man Zeilen und Spalten einer Matrix $\boldsymbol{A}$ vertauscht, erhält man die transponierte Matrix $\boldsymbol{A}^T$:

$$\boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$$

Bemerkungen:

1. Es gilt offensichtlich $(\boldsymbol{A}^T)^T = \boldsymbol{A}$ (vgl. die zweimal gewendete Matratze).
2. Für symmetrische Matrizen gilt $\boldsymbol{A}^T = \boldsymbol{A}$.

c **Vektoren.** Vektor, genauer Spaltenvektor: n Zahlen, unter einander geschrieben.

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Elemente b_i.

d **Transponierte Vektoren**: Spaltenvektoren werden zu Zeilenvektoren, wenn man sie transponiert:

$$\underline{b}^T = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}^T = [b_1, b_2, ..., b_n] \; .$$

Drucktechnisch platzsparender als Spaltenvektoren sind Zeilenvektoren, und deshalb schreibt man Spaltenvektoren oft als transponierte Zeilenvektoren hin: $\underline{b} = [b_1, b_2, ..., b_n]^T$.

e **Einfache Rechenoperationen.** Addition und Subtraktion: Geht nur bei gleichen Dimensionen. Man addiert oder subtrahiert die einander entsprechenden Elemente.

Multiplikation mit einer Zahl (einem „Skalar"): Jedes Element wird multipliziert. Division durch eine Zahl ebenso.

Recht oft trifft man in der Statistik und anderswo auf so genannte **Linearkombinationen** von Vektoren. Das ist ein schöner Name für Ausdrücke der Form

$$\lambda_1 \underline{b}_1 + \lambda_2 \underline{b}_2$$

+ eventuell weitere solche Terme – man addiert Vielfache der beteiligten Vektoren.

f **Matrix-Multiplikation.** Matrizen können nur multipliziert werden, wenn die Dimensionen passen: $\boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{B}$ ist definiert, wenn die Anzahl Spalten von $\boldsymbol{A}$ gleich der Anzahl Zeilen von $\boldsymbol{B}$ ist. Dann ist

$$c_{ik} = \sum\nolimits_{j=1}^{m} a_{ij} b_{jk}$$

Beispiel:

$$\begin{bmatrix} 2 & 1 \\ -1 & 0 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 1 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 3 + 1 \cdot 4 & 2 \cdot 1 + 1 \cdot (-2) \\ (-1) \cdot 3 + 0 \cdot 4 & (-1) \cdot 1 + 0 \cdot (-2) \\ 3 \cdot 3 + 1 \cdot 4 & 3 \cdot 1 + 1 \cdot (-2) \end{bmatrix} = \begin{bmatrix} 10 & 0 \\ -3 & -1 \\ 13 & 1 \end{bmatrix}$$

Bemerkungen:

1. Im Beispiel ist $\boldsymbol{B} \cdot \boldsymbol{A}$ nicht definiert, da $\boldsymbol{B}$ 2 Spalten, $\boldsymbol{A}$ aber 3 Zeilen hat.
2. Wenn $\boldsymbol{A} \cdot \boldsymbol{B}$ und $\boldsymbol{B} \cdot \boldsymbol{A}$ beide definiert sind, sind die beiden im allgemeinen verschieden, $\boldsymbol{A} \cdot \boldsymbol{B} \neq \boldsymbol{B} \cdot \boldsymbol{A}$! Matrizen dürfen nicht vertauscht werden.
3. Es kann $\boldsymbol{A} \cdot \boldsymbol{B} = \boldsymbol{0}$ sein, obwohl weder $\boldsymbol{A} = \boldsymbol{0}$ noch $\boldsymbol{B} = \boldsymbol{0}$ ist.
4. Es gilt das Assoziativgesetz: $(\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{C})$
5. Es gilt das Distributivgesetz: $\boldsymbol{A} \cdot (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \cdot \boldsymbol{B} + \boldsymbol{A} \cdot \boldsymbol{C}$ und ebenso $(\boldsymbol{A} + \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{C} + \boldsymbol{B} \cdot \boldsymbol{C}$.
6. Transponieren eines Produktes: Es ist

   $$(\boldsymbol{A} \cdot \boldsymbol{B})^T = \boldsymbol{B}^T \cdot \boldsymbol{A}^T$$

   Man muss also beim Transponieren die Reihenfolge vertauschen!
7. Das Produkt $\boldsymbol{A} \cdot \boldsymbol{A}^T$ ist immer symmetrisch.

g All das gilt auch für Vektoren: Wenn $\underline{a}$ und $\underline{b}$ Spaltenvektoren sind, ist

$$\underline{a} \cdot \underline{b}^T = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \dots & a_1 b_m \\ a_2 b_1 & a_2 b_2 & \dots & a_2 b_m \\ \vdots & & & \vdots \\ a_n b_1 & a_n b_2 & \dots & a_n b_m \end{bmatrix} .$$

Wenn sie gleiche Länge haben, ist

$$\underline{a}^T \cdot \underline{b} = \sum_i a_i \cdot b_i \ .$$

„Matrix mal Spaltenvektor“ ergibt (falls definiert) einen Spaltenvektor: $\boldsymbol{A} \cdot \underline{b} = \underline{c}$.

h Die **Länge eines Vektors** ist die Wurzel aus $\sum_i a_i^2$. Man bezeichnet sie oft mit $\|\underline{a}\|$. Man kann schreiben

$$\|\underline{a}\|^2 = \underline{a}^T \cdot \underline{a} \ .$$

i Die **Einheitsmatrix** (der Dimension m) ist definiert als Diagonalmatrix mit lauter Einsen:

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Sie lässt bei Multiplikation Matrizen unverändert: $\boldsymbol{I} \cdot \boldsymbol{A} = \boldsymbol{A}$, $\boldsymbol{A} \cdot \boldsymbol{I} = \boldsymbol{A}$.

j **Inverse Matrix.** Wenn $\boldsymbol{A}$ quadratisch ist und $\boldsymbol{B} \cdot \boldsymbol{A} = \boldsymbol{I}$ gilt, heisst $\boldsymbol{B}$ die zu $\boldsymbol{A}$ inverse Matrix; man schreibt $\boldsymbol{B} = \boldsymbol{A}^{-1}$.

Bemerkungen:

1. Es gilt dann auch $\boldsymbol{A} \cdot \boldsymbol{B} = \boldsymbol{I}$. Wenn also $\boldsymbol{B} = \boldsymbol{A}^{-1}$ ist, ist auch $\boldsymbol{A} = \boldsymbol{B}^{-1}$.
2. Es gibt nicht zu jeder quadratischen Matrix $\boldsymbol{A}$ eine Inverse. Wenn es eine gibt, heisst $\boldsymbol{A}$ **regulär**, und es gibt nur eine Inverse. Wenn es keine Inverse gibt, heisst $\boldsymbol{A}$ **singulär**.
3. Es ist $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$.
4. Inverses eines Matrix-Produkts: Wenn $\boldsymbol{A}$ und $\boldsymbol{B}$ quadratisch sind, ist

   $$(\boldsymbol{A} \cdot \boldsymbol{B})^{-1} = \boldsymbol{B}^{-1} \cdot \boldsymbol{A}^{-1}$$

   Die Reihenfolge muss also vertauscht werden, wie beim Transponieren!
5. Es ist $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$. Man schreibt oft kurz $\boldsymbol{A}^{-T}$.

k **Lineares Gleichungssystem.** Kurz zusammengefasst: Das Gleichungssystem

$$
\begin{aligned}
a_{11}\beta_1 + a_{12}\beta_2 + ... + a_{1m}\beta_m &= y_1 \\
a_{21}\beta_1 + a_{22}\beta_2 + ... + a_{2m}\beta_m &= y_2 \\
\ldots \quad & \quad .. \\
a_{m1}\beta_1 + a_{m2}\beta_2 + ... + a_{mm}\beta_m &= y_m
\end{aligned}
$$

(für die β_j) lässt sich schreiben als

$$\boldsymbol{A}\underline{\beta} = \underline{y}$$

(für $\underline{\beta}$). Es hat genau eine Lösung, wenn $\boldsymbol{A}$ regulär ist, also wenn die Inverse $\boldsymbol{A}^{-1}$ existiert. Dann ist

$$\underline{\beta} = \boldsymbol{A}^{-1}\underline{y}$$

diese Lösung.

l Wenn die **Matrix** $\boldsymbol{A}$ **singulär** ist, dann gibt es eine Zeile $[a_{i1}, a_{i2}, ..., a_{im}]$, die sich als Linearkombination der andern schreiben lässt. Die entsprechende Gleichung führt entweder zu einem Widerspruch (keine Lösung) oder ist überflüssig (unendlich viele Lösungen). Man spricht von **linearer Abhängigkeit** der Zeilen der Matrix oder der Gleichungen.

(Wenn die Matrix singulär ist, gibt es auch eine Spalte, die sich als Linearkombination der andern schreiben lässt. Es sind also auch die Spaltenvektoren linear abhängig.)

3.S S-Funktionen

a **Modell-Formeln** dienen dazu, Modelle von Regressionen und Varianzanalysen aller Art und auch Modelle der multivariaten Statistik festzulegen. Sie sind dadurch gekenntzeichnet, dass sie das Zeichen $\sim$ enthalten. Solche Ausdrücke bilden eine spezielle Klasse von S-Objekten, genannt `formula`-Objekte. Regressions- und Varianzanalyse-Funktionen verlangen jeweils als erstes Argument eine solche `formula`.

Bei Regressions- und Varianzanalyse-Modellen steht links von diesem Zeichen die Zielgrösse und rechts die Ausgangsgrössen. In der einfachsten Form lautet ein multiples Regressionsmodell

```
y ~ x1 + x2
```

Das Zeichen + erhält hier eine neue Bedeutung. Es werden nicht `x1` und `x2` zusammengezählt, sondern die beiden Variablen werden als Ausgangsvariable im Modell erkannt. In mathematischer Schreibweise entsteht also der Ausdruck β_1`x1` + β_2`x2`. Automatisch wird ein Fehlerterm $+E$ hinzugefügt. Ebenso ein **Achsenabschnitt** β_0, wenn man ihn nicht ausdrücklich unterdrückt, indem man -1 einfügt, also beispielsweise `y ~ -1 + x1 + x2` schreibt. So entspricht also der Ausdruck `y ~ x1 + x2` dem Regressionsmodell

$$\mathrm{y}_i = \beta_1 \mathrm{x1}_i + \beta_2 \mathrm{x2}_i + E_i \ .$$

Wie schon in 2.S.0.c erwähnt, können **Transformationen** direkt in die Formel geschrieben werden,

```
log10(ersch) ~ log10(dist) + log10(ladung)
```

b **Faktoren** oder nominale Ausgangsgrössen können (wie in 3.2.j erwähnt) ebenfalls direkt in die S-Formel geschrieben werden. Die Regressionsfunktion verwandelt solche Variable zuerst in die entsprechende Anzahl von Dummy-Variablen (3.2.h). Normalerweise sind solche Variable im `data.frame` als `factor` gekenntzeichnet und werden deshalb automatisch richtig behandelt. Wenn eine numerische Variable, beispielsweise mit den Werten 1, 2, 3, 4, als Faktor interpretiert werden soll, braucht man die Funktion `factor`. Wäre die Stelle im Beispiel in `d.spreng` nicht als Faktor gespeichert, so könnte man durch

```
log10(ersch) ~ log10(dist) + log10(ladung) + factor(St)
```

das richtige Modell dennoch erhalten.

In 3.2.g von **Nebenbedingungen** gesprochen, die nötig sind, um bei Faktoren zu einem eindeutigen Modell zu kommen. Diese können verschieden gewählt werden. Um sicher zu gehen, dass die dort erwähnte Lösung gewählt wird, dass also einfach die erste Dummy-Variable weggelassen wird, muss man in der Regressionsfunktion das Argument `contrasts="treatment"` setzen. In R ist diese Wahl der Weglasswert (default), in S-Plus wird eine numerisch bessere aber für die Interpretation unbrauchbare Nebenbedingung gewählt. Genaueres folgt in der Varianzanalyse.

c **Wechselwirkungen** zwischen Variablen (3.2.t) können in der `formula` ebenfalls einfach angegeben werden, und zwar mit einem Ausdruck der Form `x1:x2`,

`log10(ersch) ~ log10(dist) + St + log10(dist):St`

Da in den Modellen Wechselwirkungen immer nur zwischen Variablen einbezogen werden sollen, die auch als Einzelterme („Haupteffekte" im Gegensatz zu Wechselwirkungen) auftreten, gibt es eine Kurzschreibweise. `x1*x2` bedeutet das Gleiche wie `x1+x2+x1:x2`. Das vorhergende Modell kann deshalb kurz als

`log10(ersch) ~ log10(dist) * St`

angegeben werden.

d Wie man sieht, erhält nicht nur das Zeichen + eine neue Bedeutung, wenn es in einer `formula` erscheint, sondern auch * und : ; sie bezeichnen Wechselwirkungen. (In der Varianzanalyse werden auch `^` und `/` für Abkürzungen üblicher Modellstrukturen benützt werden.) Manchmal möchte man aber * auch als Multiplikationszeichen verstanden wissen. Wenn man beispielsweise eine in cm gemessene Variable in inches ausdrücken will, braucht man 2.51*`x` als Ausgangsgrösse. Man kann diese einfache Transformation mit Hilfe der Funktion `I()` angeben durch `y~I(2.51*x)`.

e **Funktion `lm, summary`.** Die Funktionen `lm` und `summary` produzieren die gleichen Resultate wie in der einfachen Regression (2.S.0.g), mit zusätzlichen Zeilen in der Koeffizienten-Tabelle, die dem erweiterten Modell entsprechen.

f **Funktion `drop1`.** Wenn eine Ausgangsgrösse und damit ein Term in der Modell-Formel einen Faktor beinhaltet, sind die Tests für die einzelnen Koeffizienten nicht sinnvoll. Ihre Bedeutung hängt ja von den Nebenbedingungen, also von den `contrasts` ab. Der sinnvolle Test, der prüft, ob der ganze Term nötig sei (3.2.m), wird von der Funktion `drop1` durchgeführt.

```
> drop1(r.lm, test="F")
```

Die Funktion berechnet primär ein Kriterium mit Namen AIC, das wir später für die Modellwahl brauchen werden (5.2.e). Wenn das Argument `test` nicht angegeben wird, wird kein Test durchgeführt.

g Einige Eigenheiten dieser Methodik erscheinen dem Autor dieser Beschreibung wenig benutzerfreundlich. Beispielsweise ist nicht einzusehen, weshalb das Objekt, das `lm` produziert, wenig Nützliches zeigt, wenn man es direkt ausgibt, sondern zuerst die generische Funktion `summary` darauf angewendet werden muss. Will man die Resultate weiter verwenden, so sind einige interessante Ergebnisse, wie die geschätzte Standardabweichung $\widehat{\sigma}$ der Fehler, nicht im Ergebnis von `lm` enthalten, sondern erst im Ergebnis von `summary(r.lm)`. Leider enthält auch das `summary` nicht das, was für die Interpretation gebraucht wird. Vertrauensintervalle, standardisierte Koeffizienten und die R_j^2-Werte müssen mit zusätzlichen Funktionen ermittelt werden. Für nominale Ausgangsgrössen muss, wie erwähnt, `drop1` aufgerufen werden.

Ich habe daher eine neue grundlegende Funktion geschrieben, die Klasse von Objekten erzeugt, welche wiederum durch verbesserte Methoden der generischen Funktionen `print` und `plot` dargestellt werden. Die neuen Funktionen beruhen selbstverständlich auf den grundlegenden Funktionen von S. Die neue Klasse „erbt" auch die Methoden von `lm`, soweit keine speziellen Methoden zu generischen Funktionen nötig wurden.

h **Funktion `regr`.** Die Funktion `regr` hat die gleichen Argumente wie `lm`. Sie erzeugt ein Objekt der Klasse `regr`, das alle interessanten Resultate der Anpassung enthält.

```
> r.regr <- regr(log10(ersch)~log10(dist)+log10(ladung)+stelle,
  data=d.spreng)
```

Die wichtigsten Resultate sieht man durch Eintippen von

```
> r.regr
```

Das Hauptresultat ist eine Tabelle, die für alle erklärenden Variablen den Test für die Nullhypothese "kein Einfluss" prüft. Für Variable mit einem Freiheitsgrad wird neben dem geschätzten Koeffizienten die standardisierte Version angegeben. Statt dem Standarfehler wird eine nützliche Grösse angegeben, mit der das Vertrauensintervall einfach berechnet werden kann (3.1.l).

Für Terme mit mehreren Freiheitsgraden wird in der Haupttabelle nur der F-Test angegeben. Die geschätzten Koeffizienten folgen anschliessend an die Tabelle. Sie sind direkt interpretierbar, ohne dass bekannt sein muss, mit welchen Kontrasten Faktoren codiert werden.

Weitere Vorteile der Funktion `regr` werden sich bei der Residuen-Analyse und bei den Methoden für andere Regressionsmodelle zeigen.

i **Resultate von `regr`**

- Aufruf, mit dem das Objekt erzeugt wurde;
- „Haupttabelle“ mit den Spalten
  - `coef`: die geschätzten Koeffizienten $\widehat{\beta}_j$ für Variable mit einem einzigen Freiheitsgrad,
  - `stcoef`: die standardisierten Koeffizienten $\widehat{\beta}_j^* = \widehat{\beta}_j \cdot \mathrm{sd}\langle X^{(j)}\rangle / \mathrm{sd}\langle Y\rangle$,
  - `Rx2`: Das Mass R_j^2 für Kollinearität,
  - `df`: Anzahl Freiheitsgrade,
  - `signif`: Für Variable mit einem einzigen Freiheitsgrad wird hier die t-ratio $= T/q_{0.975}^{(t_k)}$, der Quotient aus der klassischen t-Test-Statistik und ihrer Signifikanzgrenze, angegeben. Die Nullhypothese $\beta_j = 0$ wird abgelehnt, wenn die t-ratio betragsmässig grösser als 1 ist.

    Für Faktoren und andere Terme mit mehr als einem Freiheitsgrad liefert die Spalte eine monotone Transformation der Teststatistik des F-Tests, deren Wert ebenfalls mit 1 verglichen werden kann, siehe 3.2.r.
  - `p value`: Der P-Wert für den durchgeführten Test.
- Falls Faktoren oder andere Terme mit mehr als einem Freiheitsgrad vorkommen, folgen die geschätzten Koeffizienten.
- Es folgen die Angaben über die geschätzte Standardabweichung des Zufallsterms (mit einer sinnvollen Bezeichnung!), das Bestimmtheitsmass und der Gesamt-Test.
- Falls das Argument `correlation=TRUE` gesetzt wird, folgt die Korrelationsmatrix der geschätzten Koeffizienten (siehe `summary.lm`)

j **Funktionen `residuals, fitted`.** Die Residuen und die angepassten Werte sind als Komponenten in der Resultat-Liste von `lm` oder `regr` enthalten. Man kann sie also als `t.r$residuals` resp. `t.r$fitted.values` ansprechen. Eleganter, weil auch in anderen Modellen anwendbar, ist die Anwendung der Funktionen („Extraktor-Funktionen") `residuals` und `fitted` (oder synonym `resid`, `fitted.values`. Man schreibt also beispielsweise `residuals(t.r)`, um die Residuen zu erhalten. Achtung: Bei `lm` ist, wenn die Daten fehlende Werte (NA) enthalten, der Residuen-Vektor kürzer als die Daten, ausser wenn `na.action=na.replace` gesetzt wurde. Dann enthält der Residuenvektor selbst NAs für jene Beobachtungen, die für die Regressionsrechnung nicht verwendet wurden.

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.