

In Kap. 9 und 10 haben wir statistische Modelle betrachtet, bei denen die Beobachtungen als Realisierungen unabhängiger und normalverteilter Zufallsvariablen betrachtet wurden. Einen allgemeinen Rahmen für Modelle mit Normalverteilungsannahmen bietet das *normale lineare Modell*. In Abschn. 11.1 führen wir das normale lineare Modell ein und diskutieren die ANOVA als Spezialfall. In Abschn. 11.2 betrachten wir einen weiteren prominenten Spezialfall, das einfache lineare Regressionsmodell.

11.1 Das normale lineare Modell

Die einfaktorielle ANOVA, obwohl allgemeiner als der t -Test, ist ebenfalls nur ein Spezialfall des wesentlich größeren normalen linearen Modells, in dessen allgemeinem Rahmen eine ganze Reihe von Fragestellungen mit den gleichen statistischen Werkzeugen behandelt werden können. Wir formulieren daher allgemein folgende Definition:

Definition 11.1 (Normales lineares Modell)

Es sei \mathcal{M} ein k -dimensionaler Untervektorraum des \mathbb{R}^n mit $1 < k < n$. Das normale lineare Modell ist gegeben durch einen Zufallsvektor \mathfrak{X} , mit

$$\mathfrak{X} = \mu + \sigma \mathfrak{Z},$$

wobei $(\mu, \sigma^2) \in \Theta := \mathcal{M} \times \mathbb{R}^+$ und $\mathfrak{Z} \sim N_n(0, E_n)$.

Im allgemeinen Fall des normalen linearen Modells machen wir zunächst keine expliziten Annahmen an den Modellraum \mathcal{M} und den Nullhypothesenraum. Der Modellraum \mathcal{M} ist dann problemabhängig zu formulieren, beispielsweise im Rahmen der ANOVA durch (10.1).

Parameterschätzer Wiederum ist ein Schätzer für μ gegeben durch die orthogonale Projektion von \mathbf{x} auf \mathcal{M} ,

$$\hat{\mu}(\mathbf{x}) := \arg \min_{\mu \in \mathcal{M}} \|\mathbf{x} - \mu\|^2 = \mathcal{P}_{\mathcal{M}} \mathbf{x}, \quad (11.1)$$

den *Kleinste-Quadrate-Schätzer* (engl. *least squares estimator*, LSE), für μ . Der LSE ist zum einen erwartungstreu für μ . Denn wegen $\hat{\mu}(\mathfrak{X}) = \mu + \sigma \mathcal{P}_{\mathcal{M}} \mathfrak{Z}$ gilt $\mathbb{E}_{(\mu, \sigma^2)}[\hat{\mu}(\mathfrak{X})] = \mu$. Weiterhin ist $\hat{\mu}$ sogar MLE für μ . Denn für $\mu = (\mu_1, \dots, \mu_n)^t \in \mathcal{M}$ gilt für die Likelihoodfunktion

$$L(\mathbf{x}, (\mu, \sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_i)^2\right),$$

d. h., für jedes mögliche σ^2 wird $L(\mathbf{x}, (\mu, \sigma^2))$ genau dann maximal, wenn $\|\mathbf{x} - \mu\|^2$ minimal wird.

Als Schätzer für σ^2 betrachten wir wieder

$$\hat{\sigma}^2(\mathbf{x}) := \frac{\|R(\mathbf{x})\|^2}{n - k}, \quad (11.2)$$

mit Residuum $R(\mathbf{x}) := \mathbf{x} - \mathcal{P}_{\mathcal{M}} \mathbf{x} = \mathcal{P}_{\mathcal{M}^\perp} \mathbf{x}$. Dann gilt im Rahmen des Modells wegen $\mu \in \mathcal{M}$ die Beziehung $R(\mathfrak{X}) = \sigma \mathcal{P}_{\mathcal{M}^\perp} \mathfrak{Z}$. Wegen $\|\mathcal{P}_{\mathcal{M}^\perp} \mathfrak{Z}\|^2 \sim \chi^2(n - k)$ ist damit $\mathbb{E}_{(\mu, \sigma^2)}[\hat{\sigma}^2(\mathfrak{X})] = \sigma^2$, d. h., $\hat{\sigma}^2$ ist erwartungstreu für σ^2 .

F-Test im normalen linearen Modell Im normalen linearen Modell wird analog zur ANOVA ein F -Test der Nullhypothese konstruiert, dass μ in einem Untervektorraum \mathcal{H} von \mathcal{M} liegt, mit $\dim(\mathcal{H}) < \dim(\mathcal{M})$ ($\mathcal{H} \leftrightarrow$ „Nullhypothese“). Wir zerlegen dazu wieder

$$\mathbb{R}^n = \underbrace{\mathcal{H} \oplus \mathcal{E}}_{\mathcal{M}} \oplus \mathcal{M}^\perp \quad \text{und} \quad \mathbf{x} = \mathcal{P}_{\mathcal{H}} \mathbf{x} + \mathcal{P}_{\mathcal{E}} \mathbf{x} + \mathcal{P}_{\mathcal{M}^\perp} \mathbf{x}.$$

Wir erinnern hier auch wieder an die Abb. 10.2, in welcher die Diagonale \mathcal{D} nun im allgemeinen Fall des linearen Modells durch den Hypothesenraum \mathcal{H} ersetzt werden muss. Darüber hinaus bleiben das Bild und die Denkweise die gleiche. Folglich finden wir mit den gleichen Argumenten wie bei der ANOVA, dass für die F -Statistik unter der Nullhypothese gilt

$$F(\mathfrak{X}) = \frac{\|\mathcal{P}_{\mathcal{E}}\mathfrak{X}\|^2/\dim(\mathcal{E})}{\|\mathcal{P}_{\mathcal{M}^\perp}\mathfrak{X}\|^2/\dim(\mathcal{M}^\perp)} = \frac{\sigma^2\|\mathcal{P}_{\mathcal{E}}\mathfrak{Z}\|^2/\dim(\mathcal{E})}{\sigma^2\|\mathcal{P}_{\mathcal{M}^\perp}\mathfrak{Z}\|^2/\dim(\mathcal{M}^\perp)} \sim \mathcal{F}(\dim(\mathcal{E}), \dim(\mathcal{M}^\perp)),$$

vgl. Herleitung von (10.2).

Interpretation: Die Abweichung von der Nullhypothese \mathcal{H} (Zähler) ist dann groß, wenn die Variabilität der Beobachtungen, die nicht durch das Modell erklärt wird (Nenner), klein ist. Formal formulieren wir folgendes Lemma:

Lemma 11.2 (F -Test im normalen linearen Modell)

Es sei \mathcal{M} ein Untervektorraum des \mathbb{R}^n ($\dim(\mathcal{M}) < n$) und ein normales lineares Modell gegeben durch

$$\mathfrak{X} = \mu + \sigma \mathfrak{Z},$$

mit $(\mu, \sigma^2) \in \Theta := \mathcal{M} \times \mathbb{R}^+$ und $\mathfrak{Z} \sim N_n(0, E_n)$. Weiter sei \mathcal{H} ein Untervektorraum von \mathcal{M} ($1 < \dim(\mathcal{H}) < \dim(\mathcal{M})$), sodass $\mathbb{R}^n = \mathcal{H} \oplus \mathcal{E} \oplus \mathcal{M}^\perp$. Es sei eine Nullhypothese gegeben durch

$$H_0 : (\mu, \sigma^2) \in \mathcal{H} \times \mathbb{R}^+.$$

Zudem sei $\alpha \in (0, 1)$, sowie q_α das α -Quantil der $\mathcal{F}(\dim(\mathcal{E}), \dim(\mathcal{M}^\perp))$ -Verteilung. Dann ist die F -Statistik

$$F(\mathbf{x}) := \frac{\|\mathcal{P}_{\mathcal{E}}\mathbf{x}\|^2/\dim(\mathcal{E})}{\|\mathcal{P}_{\mathcal{M}^\perp}\mathbf{x}\|^2/\dim(\mathcal{M}^\perp)}$$

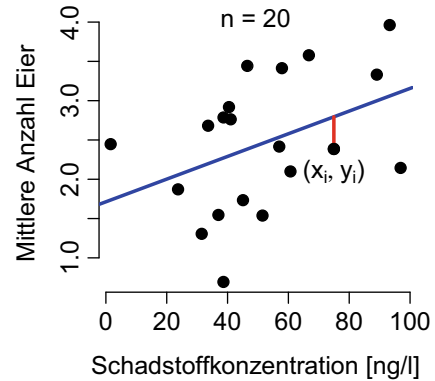
eine Teststatistik für einen exakten Test der Nullhypothese H_0 zum Niveau α mit Ablehnungsbereich $\mathcal{R}(\alpha) := [q_{1-\alpha}, \infty)$.

11.2 Einfache lineare Regression

Ein weiteres Beispiel eines normalen linearen Modells ist die einfache lineare Regression. Dabei wird ein linearer Zusammenhang zwischen Beobachtungen $\mathbf{y} = (y_1, \dots, y_n)^t$ und einer erklärenden Variable $\mathbf{x} = (x_1, \dots, x_n)^t$ untersucht. Das Wort *einfach* bedeutet hier, dass nur eine erklärende Variable \mathbf{x} involviert ist.

Motivation und Beispiel Wir betrachten folgendes Beispiel: Gewisse Stoffe können die reproduktive Aktivität von Organismen beeinflussen. Um dies an einem neuen Schadstoff zu untersuchen, entnimmt jemand an $n = 20$ Stellen eines Fließgewässers Wasserproben und misst dort die Konzentration des Schadstoffs sowie die mittlere Anzahl Eier, die Schnecken einer bestimmten Art in diesem Wasser pro Zentimeter Körperlänge in einem gewissen

Abb. 11.1 Situation der linearen Regression, mit Regressionsgerade (blau) und einem Residuum (rot)



Zeitraum produzieren. Für Messstelle i sei die Konzentration mit x_i und die mittlere Anzahl Eier pro Länge mit y_i bezeichnet. Das liefert uns *Datenpaare* $(x_i, y_i)_{i=1, \dots, n}$. In Abb. 11.1 ist für $n = 20$ ein Beispiel solcher y -Werte gegen die x -Werte aufgetragen.

Wir beobachten einen *positiven* Zusammenhang: Je höher die Konzentration, desto höher die mittlere Anzahl Eier. Dies suggeriert, dass eine erhöhte Konzentration dieses Stoffs i. Allg. mit einer erhöhten reproduktiven Aktivität dieser Schneckenart einhergeht. Aber die Variabilität ist groß, und viele y -Werte liegen auch unter dem Referenzwert bei Schadstoffkonzentration 0. Daher möchten wir die Frage beantworten: Kann – bzw. mit welcher Wahrscheinlichkeit kann – ein mindestens so deutlicher Zusammenhang durch Zufall beobachtet werden, obwohl es in der zugehörigen Population eigentlich gar keinen Zusammenhang zwischen den beiden Variablen gibt?

Die Grundidee der linearen Regression ist es, die y_i durch eine geeignete affin-lineare Funktion der x_i möglichst gut zu beschreiben (blaue *Regressionsgerade*). Lägen alle Punkte auf einer Geraden, so ließe sich die mittlere Anzahl Eier perfekt durch die Konzentration beschreiben. Die beobachteten Abweichungen der y_i von der Regressionsgeraden modellieren wir stochastisch.

Modell der einfachen linearen Regression Es sei $\mathbf{z} := (Z_1, \dots, Z_n)^t \sim N_n(0, E_n)$ und $(\beta_0, \beta_1, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$, sowie $(x_1, \dots, x_n)^t \in \mathbb{R}^n$. Das Modell der einfachen linearen Regression hat die Form

$$Y_i = \beta_0 + \beta_1 x_i + \sigma Z_i, \quad (11.3)$$

für $i = 1, \dots, n$. Zu beachten ist, dass die x_i hier im Unterschied zu den y_i nicht als zufällig modelliert, sondern als gegeben angenommen werden. Die Y_1, \dots, Y_n sind also unabhängige Zufallsvariablen mit $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$. Es sei $\mathbf{y} = (y_1, \dots, y_n)^t \in \mathbb{R}^n$. Werden die Parameter β_0 und β_1 aus \mathbf{y} durch die Methode der kleinsten Quadrate geschätzt und die Schätzungen mit $\hat{\beta}_0(\mathbf{y})$ und $\hat{\beta}_1(\mathbf{y})$ bezeichnet, so nennen wir die Gerade

$$\{(x, y) \in \mathbb{R}^2 \mid y = \hat{\beta}_0(\mathbf{y}) + \hat{\beta}_1(\mathbf{y}) \cdot x\} \quad (11.4)$$

die *Regressionsgerade* des Modells. Eine typische statistische Frage ist hier zum Beispiel: Ist in Wahrheit die Steigung $\beta_1 = 0$? In diesem Falle wäre der Erwartungswert aller Y_i konstant β_0 und hinge damit nicht vom Wert von x_i ab.

Um zu verstehen, dass wir uns eigentlich im bekannten Szenario des linearen Modells befinden, stellen wir (11.3) in Vektorschreibweise dar. Mit $\beta := (\beta_0, \beta_1)^t$ ist

$$\mathfrak{Y} = \beta_0 \mathbb{1} + \beta_1 x + \sigma \mathfrak{Z} = \overbrace{\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}^{=:C} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \sigma \mathfrak{Z} = C\beta + \sigma \mathfrak{Z}. \quad (11.5)$$

Die Matrix C nennen wir die *Systemmatrix* und die Komponenten von β die *Regressionskoeffizienten* des Modells.

Wir erkennen nun, dass durch (11.5) ein lineares Modell gemäß Definition 11.1 beschrieben wird, mit $\mu := C\beta$. Im Sinne der Definition haben wir noch den Modellraum $\mathcal{M} \subset \mathbb{R}^n$ zu spezifizieren, also sämtliche Werte, die μ annehmen kann. Wir finden

$$\mathcal{M} = \{C\beta \mid \beta \in \mathbb{R}^2\}. \quad (11.6)$$

Um technische Schwierigkeiten zu vermeiden, nehmen wir an, dass C vollen Rang hat, d. h., $n \geq 2$, und die Komponenten von \mathbf{x} sind nicht konstant, sodass die Spalten von C linear unabhängig sind. Dies impliziert die Isomorphie von \mathbb{R}^2 und \mathcal{M} , und man spricht dann auch von einem *identifizierbaren* statistischen Modell, d. h., es gibt keine zwei verschiedenen Parameterkombinationen, die das gleiche Modell beschreiben. Insbesondere finden wir $\dim(\mathcal{M}) = 2$. Wegen $\mu = C\beta$ wird der LSE für β durch den LSE für μ definiert. Letzterer war laut (11.1) gegeben als $\hat{\mu}(\mathbf{y}) = \mathcal{P}_{\mathcal{M}}\mathbf{y}$.

Definition 11.3 (Einfaches Regressionsmodell und LSE der Regressionskoeffizienten)

Es sei $n > 2$, und C sei die Systemmatrix aus (11.5), wobei diese vollen $\text{Rang}(C) = 2$ habe. Weiter sei $\mathfrak{Z} \sim N_n(0, E_n)$. Das einfache lineare Regressionsmodell ist gegeben durch einen Zufallsvektor \mathfrak{Y} mit

$$(i) \quad \mathfrak{Y} = C\beta + \sigma \mathfrak{Z}, \quad \text{mit} \quad (\beta, \sigma^2) \in \Theta := \mathbb{R}^2 \times \mathbb{R}^+.$$

Mit Wahl von $\mu := C\beta$ und $\mathcal{M} := \{C\beta \mid \beta \in \mathbb{R}^2\}$ ist dies äquivalent zu

$$(ii) \quad \mathfrak{Y} = \mu + \sigma \mathfrak{Z}, \quad \text{mit} \quad (\mu, \sigma^2) \in \mathcal{M} \times \mathbb{R}^+.$$

In Darstellung (ii) sei $\hat{\mu}$ der LSE für μ . In (i) heißt dann $\hat{\beta}$ der LSE für β , wenn gilt

$$C\hat{\beta} = \hat{\mu}.$$

Interpretation des LSE und der Regressionsgeraden Wir diskutieren die Bedeutung des LSE und der Regressionsgeraden. In Abb. 11.2 ist die Darstellung der Beobachtungen \mathbf{y} als Punktwolke (a) der Darstellung als Vektor (b) gegenübergestellt.

Schritt 1: Wir denken zunächst vektorwertig. Der Schätzer $\hat{\beta}$ ist für $\mathbf{y} \in \mathbb{R}^n$ unter allen $\beta \in \mathbb{R}^k$ als der Minimierer von $\|\mathbf{y} - C\beta\|^2$ definiert. Das kleinste Abstandsquadrat ist das Quadrat des Residuums $R(\mathbf{y}) = \mathbf{y} - \mathcal{P}_{\mathcal{M}}(\mathbf{y}) = \mathbf{y} - C\hat{\beta}(\mathbf{y})$ (rot).

Schritt 2: Wie ist dieses Residuum nun im Kontext der Punktwolke zu verstehen? Dafür schreiben wir das Residuenquadrat als $\|R(\mathbf{y})\|^2 = \sum_{i=1}^n (y_i - (\hat{\beta}_0(\mathbf{y}) + \hat{\beta}_1(\mathbf{y}) \cdot x_i))^2$, denn der i -te Summand ist gerade das Quadrat der i -ten Komponente von $\mathbf{y} - C\hat{\beta}(\mathbf{y})$. Das bedeutet, dass die Regressionsgerade (blau) durch diejenige affin-lineare Funktion beschrieben wird, durch die die Summe sämtlicher Abstandsquadrate von Beobachtung y_i (schwarze Punkte) und deren Vorhersage via Regressionsgerade, d. h., $\hat{\beta}_0(\mathbf{y}) + \hat{\beta}_1(\mathbf{y}) \cdot x_i$ (blaue Punkte), minimiert wird. Das Residuum ist hier also anhand der roten Strecken zu verstehen.

Schritt 3: Zurück zur vektorwertigen Denkweise. Fassen wir jene Vorhersagen wiederum als einen Vektor zusammen, genauer durch den Vektor, dessen i -te Komponente durch $\hat{\beta}_0(\mathbf{y}) + \hat{\beta}_1(\mathbf{y}) \cdot x_i$ gegeben ist, dann entspricht dieser gerade der Projektion $\mathcal{P}_{\mathcal{M}}(\mathbf{y}) = C\hat{\beta}(\mathbf{y})$ (blau).

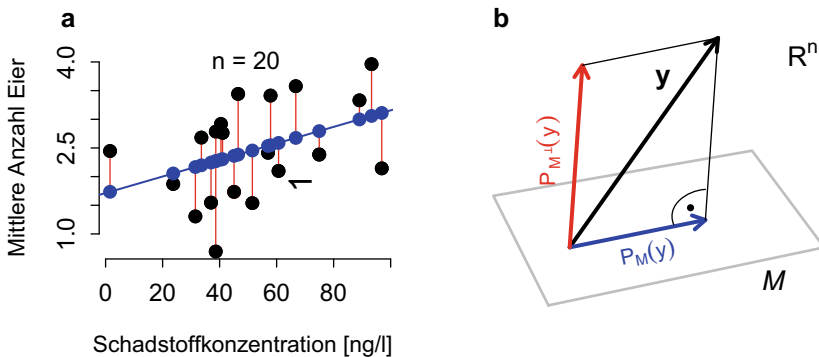
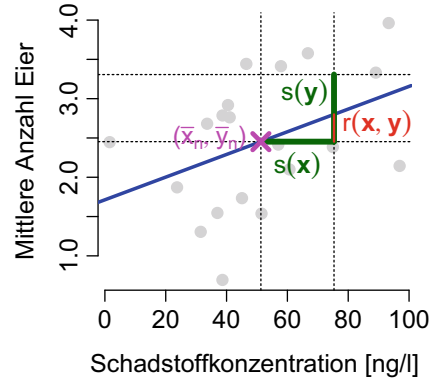


Abb. 11.2 Zwei geometrische Darstellungen der Regression, **a** n Punkte (x_i, y_i) im \mathbb{R}^2 , **b** ein Vektor \mathbf{y} im \mathbb{R}^n

Abb. 11.3 Interpretation der Regressionsgeraden



Wir diskutieren noch zwei schöne Eigenschaften der Regressionsgeraden, visualisiert in Abb. 11.3.

1. Der sogenannte Schwerpunkt (\bar{x}_n, \bar{y}_n) (magentafarben) der Vektoren \mathbf{x} und \mathbf{y} liegt auf der Regressionsgeraden.

Das erkennen wir durch eine geometrische Überlegung, und zwar durch Projektion von \mathbf{y} auf die Diagonale \mathcal{D} . Denn da $\mathbb{1} \in \mathcal{M}$, gilt $(\mathbf{y} - C\hat{\beta}(\mathbf{y})) \perp \mathbb{1}$, sodass

$$0 \stackrel{(*)}{=} \sum_{i=1}^n (y_i - \hat{\beta}_0(\mathbf{y}) - \hat{\beta}_1(\mathbf{y}) \cdot x_i) = n\bar{y}_n - n\hat{\beta}_0(\mathbf{y}) - n\hat{\beta}_1(\mathbf{y}) \cdot \bar{x}_n,$$

und folglich $\bar{y}_n = \hat{\beta}_0(\mathbf{y}) + \hat{\beta}_1(\mathbf{y}) \cdot \bar{x}_n$. Der Schwerpunkt erfüllt also immer die Geradengleichung.

2. Die Steigung $\hat{\beta}_1$ der Regressionsgeraden schreibt sich auch als

$$\hat{\beta}_1(\mathbf{y}) = r(\mathbf{x}, \mathbf{y}) \frac{s(\mathbf{y})}{s(\mathbf{x})}. \quad (11.7)$$

Das bedeutet, dass mit einer Erhöhung der Schadstoffkonzentration um eine Standardabweichung $s(\mathbf{x})$ im Mittel eine Veränderung der mittleren Anzahl Eier um $r(\mathbf{x}, \mathbf{y})s(\mathbf{y})$ einhergeht. Den Faktor r bezeichnet man als die empirische Korrelation von \mathbf{x} und \mathbf{y} , gegeben durch

$$r(\mathbf{x}, \mathbf{y}) := \frac{(1/(n-1)) \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{s(\mathbf{x})s(\mathbf{y})},$$

wobei der Zähler als empirische Kovarianz von \mathbf{x} und \mathbf{y} bezeichnet wird. Man beachte die Analogie zu den theoretischen Größen in (2.10) und (2.11). Es gilt $|r(\mathbf{x}, \mathbf{y})| \leq 1$, denn laut der Cauchy-Schwarz-Ungleichung ist $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$.

Die Darstellung der Steigung in (11.7) lässt sich ebenfalls durch die geometrischen Überlegungen herleiten. Da $\mathbf{x} \in \mathcal{M}$, gilt $(\mathbf{y} - C\hat{\beta}(\mathbf{y})) \perp \mathbf{x}$, sodass

$$\begin{aligned} 0 &= \sum_{i=1}^n \left(y_i - (\hat{\beta}_0(\mathbf{y}) + \hat{\beta}_1(\mathbf{y}) \cdot x_i) \right) x_i = \sum_{i=1}^n \left(y_i - \hat{\beta}_0(\mathbf{y}) - \hat{\beta}_1(\mathbf{y}) \cdot x_i \right) (x_i - \bar{x}_n) \\ &= \sum_{i=1}^n \left(y_i - \hat{\beta}_1(\mathbf{y}) \cdot x_i \right) (x_i - \bar{x}_n) \\ &= \sum_{i=1}^n \left((y_i - \bar{y}_n) - \hat{\beta}_1(\mathbf{y})(x_i - \bar{x}_n) \right) (x_i - \bar{x}_n). \end{aligned}$$

Dabei haben wir in der zweiten Gleichung (*) aus Teil 1 genutzt (\bar{x}_n ist konstant), und in der dritten und vierten Gleichung lediglich Konstanten eingeführt und dabei die Null addiert, denn $\sum (x_i - \bar{x}_n) = 0$. Auflösen nach $\hat{\beta}_1(\mathbf{y})$ ergibt

$$\hat{\beta}_1(\mathbf{y}) = \frac{1/(n-1)}{1/(n-1)} \cdot \frac{\sum_{i=1}^n (y_i - \bar{y}_n)(x_i - \bar{x}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} = r(\mathbf{x}, \mathbf{y}) \frac{s(\mathbf{y})}{s(\mathbf{x})}.$$

Geschlossene Form der Parameterschätzer In Definition 11.3 haben wir im Kontext des Regressionsmodells den LSE $\hat{\beta}$ für den Vektor der Regressionskoeffizienten β kennengelernt. Er ist der Minimierer des Längenquadrats $\|\mathbf{y} - C\beta\|^2$, wobei über sämtliche $\beta \in \mathbb{R}^2$ minimiert wird. Das folgende Lemma liefert uns sogar eine geschlossene Form.

Lemma 11.4 (LSE im Regressionsmodell)

Im Regressionsmodell (Def. 11.3) ist der LSE $\hat{\beta}$ von β gegeben durch

$$\hat{\beta}(\mathbf{y}) = (C^T C)^{-1} C^T \mathbf{y}.$$

Beweis Es sei $\mathbf{y} \in \mathbb{R}^n$. Aufgrund der Minimierungseigenschaft des LSE steht das Residuum $\mathbf{y} - C\hat{\beta}(\mathbf{y})$ senkrecht auf \mathcal{M} . Da für alle $\beta \in \mathbb{R}^k$ gilt, dass $C\beta \in \mathcal{M}$, folgt:

$$\begin{aligned} \mathbf{y} - C\hat{\beta}(\mathbf{y}) &\perp C\beta \Leftrightarrow (\mathbf{y} - C\hat{\beta}(\mathbf{y}))^T C\beta = 0 \\ &\Leftrightarrow \mathbf{y}^T C = \hat{\beta}(\mathbf{y})^T C^T C \quad | \text{transponieren} \\ &\Leftrightarrow C^T \mathbf{y} = C^T C \hat{\beta}(\mathbf{y}) \\ &\Leftrightarrow \hat{\beta}(\mathbf{y}) = (C^T C)^{-1} C^T \mathbf{y} \end{aligned}$$

Insbesondere ist der Schätzer wohldefiniert, denn $C^T C$ ist genau dann invertierbar, wenn C maximalen Rang hat.

Der aus dem allgemeinen linearen Modell bekannte erwartungstreue Schätzer $\hat{\sigma}^2$ für σ^2 , vgl. (11.2), lässt sich dann durch $\hat{\beta}$ ausdrücken:

$$\hat{\sigma}^2(\mathbf{y}) = \frac{\|R(\mathbf{y})\|^2}{n-2} = \frac{\sum_{i=1}^n (y_i - (\hat{\beta}_0(\mathbf{y}) + \hat{\beta}_1(\mathbf{y}) \cdot x_i))^2}{n-2}. \quad (11.8)$$

F-Test im Regressionsmodell Wir wollen nun zur Ausgangsfrage zurückkommen. Ist die Tatsache, dass die beobachtete Steigung positiv ist, leicht durch Zufall zu erklären, wenn sie eigentlich null ist, d. h., wenn es eigentlich gar keinen Zusammenhang gibt?

Wir testen dazu die Nullhypothese, dass die Steigung β_1 verschwindet, also dass der Erwartungswert aller Y_i konstant ist. In diesem Fall ist $\mu = C\beta = \beta_0 \mathbb{1}$ und daher findet sich der mit der Nullhypothese assoziierte Teilraum von \mathcal{M} wieder als die Diagonale \mathcal{D} ,

$$H_0 : \mu \in \mathcal{D},$$

oder äquivalent $H_0 : (\beta, \sigma^2) \in (\mathbb{R} \times \{0\}) \times \mathbb{R}^+$. Für die F -Statistik zerlegen wir dann mal wieder $\mathcal{M} = \mathcal{D} \oplus \mathcal{E}$ und finden unter der Nullhypothese, dass

$$F(\mathfrak{Y}) = \frac{\|\mathcal{P}_{\mathcal{E}}\mathfrak{Y}\|^2 / \dim(\mathcal{E})}{\|\mathcal{P}_{\mathcal{M}^\perp}\mathfrak{Y}\|^2 / \dim(\mathcal{M}^\perp)} \stackrel{H_0}{\sim} \mathcal{F}(\dim(\mathcal{E}), \dim(\mathcal{M}^\perp)),$$

mit $\dim(\mathcal{M}) = 2$ und daher $\dim(\mathcal{E}) = 1$ und $\dim(\mathcal{M}^\perp) = n - 2$.

Lemma 11.5 (Einfache lineare Regression)

Es sei das einfache lineare Regressionsmodell gemäß Definition 11.3 gegeben. Mit Zerlegung $\mathcal{M} = \mathcal{D} \oplus \mathcal{E}$ sei eine Nullhypothese formuliert durch

$$H_0 : \mu \in \mathcal{D},$$

wobei \mathcal{D} die Diagonale bezeichne. Zudem sei $\alpha \in (0, 1)$, sowie q_α das α -Quantil der $\mathcal{F}(1, n-2)$ -Verteilung. Dann ist die F -Statistik

$$F(\mathbf{y}) := \frac{\|\mathcal{P}_{\mathcal{E}}\mathbf{y}\|^2}{\|\mathcal{P}_{\mathcal{M}^\perp}\mathbf{y}\|^2 / (n-2)}$$

eine Teststatistik für einen Test der Nullhypothese H_0 zum Niveau α mit Ablehnungsbereich $\mathcal{R}(\alpha) = [q_{1-\alpha}, \infty)$.

Da der Nenner der F -Statistik dem Schätzer $\hat{\sigma}^2(\mathbf{y})$ aus (11.8) gleicht und wir den Zähler aufgrund der Orthogonalität der involvierten Teilräume darstellen können als

$\|\mathcal{P}_{\mathcal{E}}\mathbf{y}\|^2 = \|\mathcal{P}_{\mathcal{M}}\mathbf{y}\|^2 - \|\mathcal{P}_{\mathcal{D}}\mathbf{y}\|^2 = \|C\hat{\beta}(\mathbf{y})\|^2 - \|\bar{y}_n\mathbf{1}\|^2$, finden wir mit Lemma 11.4 die geschlossene Form

$$F(\mathbf{y}) = \frac{\|C(C^t C)^{-1}C^t \mathbf{y}\|^2 - n\bar{y}_n^2}{\left[\sum_{i=1}^n (y_i - (\hat{\beta}_0(\mathbf{y}) + \hat{\beta}_1(\mathbf{y}) \cdot x_i))^2\right] / (n-2)}. \quad (11.9)$$

Die wesentliche Botschaft dieses Abschnittes ist, dass wir uns im Kontext des normalen linearen Modells bewegen und wir deswegen wieder die Verteilung der F -Statistik unter der Nullhypothese anhand der Projektionen herleiten können.

Wir betonen, dass, obwohl die ANOVA aus Lemma 10.1 und die einfache lineare Regression aus Lemma 11.5 zwei völlig verschiedenen Fragestellungen nachgehen, beide Modelle als Spezialfälle des normalen linearen Modells, Definition 11.1 und Lemma 11.2, betrachtet werden können. Technisch passiert immer das Gleiche, nur angewandt auf verschiedene Modellräume \mathcal{M} , die problemabhängig gewählt werden.

Anwendung Für die Daten $\mathbf{y} = (y_1, \dots, y_n)^t$ und $\mathbf{x} = (x_1, \dots, x_n)^t$ mit $n = 20$ aus Abb. 11.1 testen wir nun die Nullhypothese, dass die wahre Steigung β_1 null ist, d. h., dass die mittlere Reproduktion der Schneckenart für alle beobachteten Konzentrationen des Schadstoffs gleich ist. Unter der Nullhypothese liegt also der Erwartungswert von \mathfrak{Y} in \mathcal{D} , d. h., der Erwartungswert aller Komponenten Y_i nimmt den gleichen Wert an – unabhängig vom Wert des zugehörigen x_i . Zum Test dieser Nullhypothese berechnen wir $F(\mathbf{y}) \approx 3.56$ durch Formel (11.9), wobei wir den Schätzer $\hat{\beta}$ durch Lemma 11.4 erhalten. Nach Lemma 11.5 ist nun $F(\mathbf{y}) \approx 3.56$ in Bezug zur $\mathcal{F}(1, 18)$ -Verteilung zu setzen. Deren 0.95-Quantil beträgt etwa 4.41, sodass die Nullhypothese auf dem 5 %-Niveau nicht abgelehnt werden kann. Mit $P(\mathbf{y}) = \mathbb{P}_{H_0}(F(\mathfrak{Y}) \geq 3.56) \approx 0.075$ ist die Verträglichkeit der Daten mit Nullhypothese dennoch relativ gering. Wiederum sei die Bedeutung der korrekten Interpretation betont: Die Nullhypothese hier nicht abzulehnen, bedeutet nicht automatisch, dass die Nullhypothese stimmt! Es kann also gefährlich sein, so zu handeln, als träfe die Nullhypothese zu.

Verallgemeinerung des Regressionsmodells Wir haben das einfache lineare Regressionsmodell $\mathfrak{Y} = C\beta + \sigma\mathfrak{Z}$ mit $n \times 2$ -Systemmatrix C kennengelernt, und dieses auch im Kontext des normalen linearen Modells $\mathfrak{Y} = \mu + \sigma\mathfrak{Z}$ mit Modellraum $\mathcal{M} = \{C\beta \mid \beta \in \mathbb{R}^2\}$ verstanden.

Andersherum können wir ausgehend von einem beliebigen linearen Modell $\mathfrak{Y} = \mu + \sigma\mathfrak{Z}$, mit k -dimensionalem Modellraum \mathcal{M} , siehe Definition 11.1, auch immer eine Darstellung in Regressionsmodellschreibweise $\mathfrak{Y} = C\beta + \sigma\mathfrak{Z}$, mit Modellraum $\mathcal{M} = \{C\beta \mid \beta \in \mathbb{R}^k\}$, finden. Dafür haben wir lediglich die $n \times k$ Systemmatrix C zu spezifizieren. Abhängig von C ist dann der Vektor $\beta \in \mathbb{R}^k$ der Regressionskoeffizienten zu interpretieren.

Beispielsweise haben wir für die ANOVA bei k Gruppen den Modellraum als

$$\mathcal{M} := \{(\underbrace{\mu_1, \dots, \mu_1}_{n_1 \text{ mal}}, \underbrace{\mu_2, \dots, \mu_2}_{n_2 \text{ mal}}, \dots, \underbrace{\mu_k, \dots, \mu_k}_{n_k \text{ mal}})^t \mid (\mu_1, \dots, \mu_k)^t \in \mathbb{R}^k\}$$

formuliert, siehe (10.1). Wir suchen dann eine $n \times k$ -Matrix C , sodass gilt $\mathcal{M} = \{C\beta \mid \beta \in \mathbb{R}^k\}$. Wir können dazu beispielsweise jeder Gruppe eine Spalte von C zuordnen. In der ersten Spalte stehen dann in den ersten n_1 Zeilen Einsen und sonst nur Nullen. In der zweiten Spalte stehen zunächst n_1 Nullen, gefolgt von n_2 Einsen und $(n - n_1 - n_2)$ Nullen, usw., und in der letzten Spalte kommen zunächst $(n - n_k)$ Nullen, und lediglich die letzten n_k Einträge sind Einsen. Wir finden dann für die i -te Komponente β_i des Vektors β der Regressionskoeffizienten, dass $\beta_i = \mu_i$ für alle $i = 1, \dots, k$, sodass die i -te Komponente des LSE $\hat{\beta}$ als der Mittelwert der i -ten Gruppe zu verstehen ist. Die Komponenten von β sind also immer problemabhängig zu interpretieren. Im Kontext des einfachen linearen Regressionsmodells waren sie ja als Abszissenabschnitt und Steigung verstanden.

Diese Darstellung als Regressionsmodell erweist sich als sehr nützlich für vielfältige erweiterte Anwendungen des normalen linearen Modells. So könnte man etwa beim Beispiel der Schnecken zusätzlich die Wassertemperatur als eine zweite erklärende Variable berücksichtigen, denn eine erhöhte Temperatur könnte ja ebenfalls einen Effekt auf die mittlere Anzahl der Eier vermuten lassen. Die Systemmatrix würde dann um eine zusätzliche dritte Spalte der Temperaturwerte erweitert werden. Technisch läuft dann vieles analog. Der interessierte Leser sei zur weiterführenden Lektüre etwa an Pruscha (2000) verwiesen.