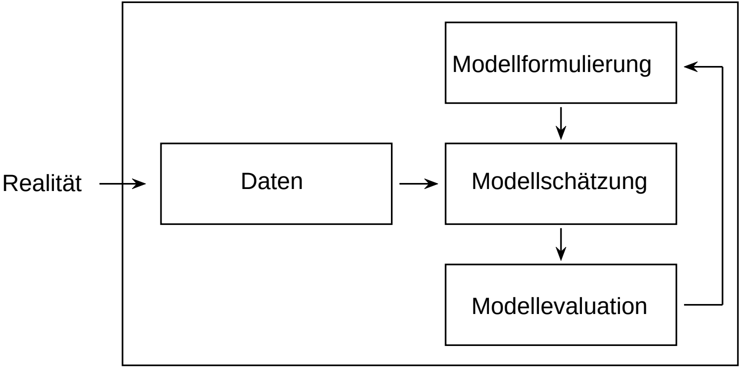
Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(6) Modellschätzung

Naturwissenschaft



Überblick

Modellformulierung

$$y = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{1}$$

Modellschätzung

$$\hat{\beta} = (X^T X)^{-1} X^T y, \hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \tag{2}$$

Modellevaluation

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}, F = \frac{(\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon})/p_2}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)} \tag{3}$$

Überblick

Standardprobleme Frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für die wahren, aber unbekannten, Parameterwerte (oder eine Funktion derer) abzugeben, typischerweise basierend auf der Beobachtung einer Datenrealisierung.

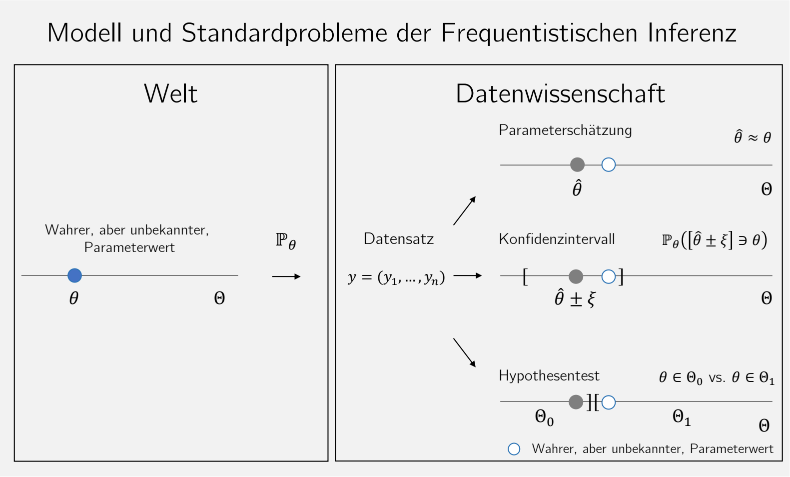
(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst sinnvollen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert, sich in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Überblick



Modell und Standardprobleme der Frequentistischen Inferenz
Welt
Wahrer, aber unbekannter, Parameterwert
θ
Θ
$\mathbb{P}_\theta$
Datenwissenschaft
Datensatz
$y = (y_1, \dots, y_n)$
Parameterschätzung
$\hat{\theta} \approx \theta$
$\hat{\theta}$
Θ
Konfidenzintervall
$\mathbb{P}_\theta([\hat{\theta} \pm \xi] \ni \theta)$
$\hat{\theta} \pm \xi$
Θ
Hypothesentest
$\theta \in \Theta_0$ vs. $\theta \in \Theta_1$
Θ_0
Θ_1
Θ
Wahrer, aber unbekannter, Parameterwert

Überblick

Standardannahmen Frequentistischer Inferenz

Gegeben sei ein statistisches Modell mit. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen der Daten des Modells ist. Aus Frequentistischer Sicht kann man unendlich oft Datensätze basierend auf einem Modell generieren und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. den Betaparameterschätzer

Datensatz (1) : $y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right)^T$ mit $\hat{\beta}^{(1)} = (X^TX)^{-1}X^Ty^{(1)}$

Datensatz (2) : $y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right)^T$ mit $\hat{\beta}^{(2)} = (X^TX)^{-1}X^Ty^{(2)}$

Datensatz (3) : $y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right)^T$ mit $\hat{\beta}^{(3)} = (X^TX)^{-1}X^Ty^{(3)}$

Datensatz (4) : $y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right)^T$ mit $\hat{\beta}^{(4)} = (X^TX)^{-1}X^Ty^{(4)}$

Datensatz (5) : $y^{(5)} = ...$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme der Datenverteilung. Was zum Beispiel ist die Verteilung von $\hat{\beta}^{(1)}$, $\hat{\beta}^{(2)}$, $\hat{\beta}^{(3)}$, $\hat{\beta}^{(4)}$, $\ldots$ also die Verteilung der Zufallsvariable $\hat{\beta} := (X^TX)^{-1}X^Ty$? Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Theorem (Betaparameterschätzer)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{4}$$

das ALM in generativer Form und es sei

$$\hat{\beta} := (X^T X)^{-1} X^T y. \tag{5}$$

der *Betaparameterschätzer*. Dann gilt, dass

$$\hat{\beta} = \arg\min_{\tilde{\beta}} (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \tag{6}$$

und dass $\hat{\beta}$ ein unverzerrter Maximum Likelihood Schätzer von $\beta \in \mathbb{R}^p$ ist.

Bemerkungen

- Das Theorem gibt ein Formel an, um β anhand von Designmatrix und Daten zu schätzen.
- Als ML Schätzer ist $\hat{\beta}$ konsistent, asymptotisch normalverteilt und asymptotisch effizient.
- Wir sehen später, dass $\hat{\beta}$ sogar normalverteilt ist.
- Außerdem hat $\hat{\beta}$ die "kleinste Varianz" in der Klasse der linearen unverzerrten Schätzer von β.
- Letztere Eigenschaft ist Kernaussage des *Gauss-Markov Theorems*, auf das wir hier nicht näher eingehen wollen.
- Für eine Diskussion und einen Beweis des Gauss-Markov Theorems siehe z.B. Searle (1971), Kapitel 3.

Allgemeine Theorie

Beweis

(1) Wir zeigen in einem ersten Schritt, dass $\hat{\beta}$ die Summe der Abweichungsquadrate

$$(y - X\tilde{\beta})^T(y - X\tilde{\beta}) \tag{7}$$

minimiert. Dazu halten wir zunächst fest, dass

$$\hat{\beta} = (X^TX)^{-1}X^Ty \Leftrightarrow X^TX\hat{\beta} = X^Ty \Leftrightarrow X^Ty - X^TX\hat{\beta} = 0_p \Leftrightarrow X^T(y - X\hat{\beta}) = 0_p. \tag{8}$$

Weiterhin gilt dann auch, dass

$$X^T(y - X\hat{\beta}) = 0_p \Leftrightarrow \left(X^T(y - X\hat{\beta})\right)^T = 0_p^T \Leftrightarrow (y - X\hat{\beta})^TX = 0_p^T \tag{9}$$

Weiterhin halten wir ohne Beweis fest, dass für jede Matrix $X \in \mathbb{R}^{n\times p}$ gilt, dass

$$z^TX^TXz \geq 0 \text{ für alle } z \in \mathbb{R}^p. \tag{10}$$

Wir betrachten nun für festes y und ein beliebiges $\tilde{\beta}$ die Summe der Abweichungsquadrate

$$(y - X\tilde{\beta})^T(y - X\tilde{\beta}). \tag{11}$$

Allgemeine Theorie

Beweis (fortgeführt)

Es ergibt sich

$$
\begin{aligned}
(y - X\tilde{\beta})^T(y - X\tilde{\beta}) &= (y - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta})^T(y - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta}) \\
&= ((y - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta}))^T((y - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta})) \\
&= \quad (y - X\hat{\beta})^T(y - X\hat{\beta}) + (y - X\hat{\beta})^T X(\hat{\beta} - \tilde{\beta}) \\
&\quad + (\hat{\beta} - \tilde{\beta})^T X^T(y - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \\
&= \quad (y - X\hat{\beta})^T(y - X\hat{\beta}) + 0_p^T(\hat{\beta} - \tilde{\beta}) \\
&\quad + (\hat{\beta} - \tilde{\beta})^T 0_p + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \\
&= (y - X\hat{\beta})^T(y - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta})
\end{aligned}
\tag{12}
$$

Auf der rechten Seite obiger Gleichung ist nur der zweite Term von $\tilde{\beta}$ abhängig. Da für diesen Term gilt, dass

$$
(\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \geq 0 \tag{13}
$$

nimmt dieser Term genau dann seinen Minimalwert 0 an, wenn

$$
(\hat{\beta} - \tilde{\beta}) = 0_p \Leftrightarrow \tilde{\beta} = \hat{\beta}. \tag{14}
$$

Also gilt

$$
\hat{\beta} = \arg\min_{\tilde{\beta}} (y - X\tilde{\beta})^T(y - X\tilde{\beta}). \tag{15}
$$

Beweis (fortgeführt)

(2) Um zu zeigen, dass $\hat{\beta}$ ein Maximum Likelihood Schätzer ist, betrachten wir für festes $y \in \mathbb{R}^n$ und festes $\sigma^2 > 0$ die Log-Likelihood Funktion

$$\ell : \mathbb{R}^p \to \mathbb{R}, \tilde{\beta} \mapsto \ln p_{\tilde{\beta}}(y) = \ln N(y; X\tilde{\beta}, \sigma^2 I_n) \tag{16}$$

wobei gilt, dass

$$\begin{aligned} \ln N(y; X\tilde{\beta}, \sigma^2 I_n) &= \ln \left((2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\tilde{\beta})^T (y - X\tilde{\beta}) \right) \right) \\ &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma^2 I_n| - \frac{1}{2\sigma^2}(y - X\tilde{\beta})^T (y - X\tilde{\beta}). \end{aligned} \tag{17}$$

Dabei hängt allein der Term $-\frac{1}{2\sigma^2}(y - X\tilde{\beta})^T (y - X\tilde{\beta})$ von $\tilde{\beta}$ ab. Weil aber $(y - X\tilde{\beta})^T (y - X\tilde{\beta}) \geq 0$, gilt wird dieser Term aufgrund des negativen Vorzeichen maximal, wenn $(y - X\tilde{\beta})^T (y - X\tilde{\beta})$ minimal wird. Dies ist aber wie oben gezeigt genau für $\tilde{\beta} = \hat{\beta}$ der Fall.

(3) Die Unverzerrtheit von $\hat{\beta}$ schließlich ergibt sich aus

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}\left((X^T X)^{-1} X^T y \right) = (X^T X)^{-1} X^T \mathbb{E}(y) = (X^T X)^{-1} X^T X \beta = \beta. \tag{18}$$

Theorem (Varianzparameterschätzer)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{19}$$

das ALM in generativer Form. Dann ist

$$\hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \tag{20}$$

ein unverzerrter Schätzer von $\sigma^2 > 0$.

Bemerkungen

- Es handelt sich bei $\hat{\sigma}^2$ *nicht* um einen ML Schätzer von σ^2.
- Für einen Beweis siehe z.B. Searle (1971), Kapitel 3 oder Rencher and Schaalje (2008), Kapitel 7.
- Mit Definition des *Residuenvektors* und der *Residuen*

$$\hat{\varepsilon} := y - X\hat{\beta} \text{ und } \hat{\varepsilon}_i := y_i - (X\hat{\beta})_i, i = 1, ..., n, \tag{21}$$

  respektive, bieten sich für $\hat{\sigma}^2$ auch folgende Schreibweisen an:

$$\hat{\sigma}^2 = \frac{1}{n-p}\hat{\varepsilon}^T\hat{\varepsilon} = \frac{1}{n-p}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - (X\beta)_i)^2 \tag{22}$$

- σ^2 wird also durch die summierten quadrierten Residuen geschätzt.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Unabhängige und identisch normalverteilte Zufallsvariablen

Wir betrachten das Szenario von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter $\mu \in \mathbb{R}$ und Varianzparameter σ^2,

$$y_i \sim N(\mu, \sigma^2) \text{ für } i = 1, ..., n. \tag{23}$$

Dann gilt, wie unten gezeigt,

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} y_i =: \bar{y} \text{ und } \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 =: s_y^2. \tag{24}$$

In diesem Fall ist also der Betaparameterschätzer mit dem Stichprobenmittel $\bar{y}$ der $y_1, ..., y_n$ und der Varianzparameterschätzer mit der Stichprobenvarianz s_y^2 der $y_1, ..., y_n$ identisch.

Unabhängige und identisch normalverteilte Zufallsvariablen

Für $\hat{\beta}$ ergibt sich

$$
\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T y \\
&= \left(1_n^T 1_n\right)^{-1} 1_n^T y \\
&= \left(\begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
&= n^{-1} \sum_{i=1}^{n} y_i \\
&= \frac{1}{n} \sum_{i=1}^{n} y_i \\
&=: \bar{y}.
\end{aligned}
$$

Unabhängige und identisch normalverteilte Zufallsvariablen

Für $\hat{\sigma}^2$ ergibt sich

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-1}\left(y - X\hat{\beta}\right)^T\left(y - X\hat{\beta}\right) \\
&= \frac{1}{n-1}\left(y - 1_n\bar{y}\right)^T\left(y - 1_n\bar{y}\right) \\
&= \frac{1}{n-1}\left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\bar{y}\right)^T\left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\bar{y}\right) \\
&= \frac{1}{n-1}\begin{pmatrix} y_1 - \bar{y} & \cdots & y_n - \bar{y} \end{pmatrix}\begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \\
&=: s_y^2.
\end{aligned}
$$

Parameterschätzung

```
# Libraries
library(MASS)                                   # Multivariate Normalverteilung

# Modellformulierung
n          = 12                                 # Anzahl Datenpunkte
p          = 1                                  # Anzahl Betaparameter
X          = matrix(rep(1,n), nrow = n)         # Designmatrix
I_n        = diag(n)                            # n x n Einheitsmatrix
beta       = 2                                  # wahrer, aber unbekannter, Betaparameter
sigsqr     = 1                                  # wahrer, aber unbekannter, Varianzparameter

# Datenrealisierung
y          =  mvrnorm(1, X %*% beta, sigsqr*I_n)  # eine Realisierung eines n-dimensionalen ZVs

# Parameterschätzung
beta_hat   = solve(t(X) %*% X) %*% t(X) %*% y    # Betaparameterschätzer
eps_hat    = y - X %*% beta_hat                 # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) /(n-p)    # Varianzparameterschätzer

# Ausgabe
cat("beta       : ", beta,
    "\nhat{beta}  : ", beta_hat,
    "\nsigsqr     : ", sigsqr,
    "\nhat{sigsqr}: ", sigsqr_hat)
```

```
> beta       :  2
> hat{beta}  :  1.64
> sigsqr     :  1
> hat{sigsqr}:  1.12
```

Simulation der Schätzerunverzerrtheit

```
# Libraries
library(MASS)                                               # Multivariate Normalverteilung

# Modellformulierung
n         = 12                                              # Anzahl Datenpunkte
p          = 1                                              # Anzahl Betaparameter
X          = matrix(rep(1,n), nrow = n)                     # Designmatrix
I_n        = diag(n)                                        # n x n Einheitsmatrix
beta       = 2                                              # wahrer, aber unbekannter, Betaparameter
sigsqr     = 1                                              # wahrer, aber unbekannter, Varianzparameter

# Frequentistische Simulation
nsim       = 1e4                                            # Anzahl Datenrealisierungen
beta_hat   = rep(NaN,nsim)                                  # \hat{\beta} Realisierungsarray
sigsqr_hat = rep(NaN,nsim)                                  # \hat{sigsqr} Realisierungsarray
for(i in 1:nsim){                                           # Simulationsiterationen
  y             = mvrnorm(1, X %*% beta, sigsqr*I_n)        # Datenrealisierung
  beta_hat[i]   = solve(t(X) %*% X) %*% t(X) %*% y          # Betaparameterschätzer
  eps_hat       = y - X %*% beta_hat[i]                     # Residuenvektor
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat) /(n-p)           # Varianzparameterschätzer
}

# Ausgabe
cat("Wahrer, aber unbekannter, Betaparameter                  : ", beta,
    "\nGeschätzter Erwartungswert des Betaparameterschätzers    : ", mean(beta_hat),
    "\nWahrer, aber unbekannter, Varianzparameter               : ", sigsqr,
    "\nGeschätzter Erwartungswert des Varianzparameterschätzers : ", mean(sigsqr_hat))
```

```
> Wahrer, aber unbekannter, Betaparameter                  :  2
> Geschätzter Erwartungswert des Betaparameterschätzers    :  2
> Wahrer, aber unbekannter, Varianzparameter               :  1
> Geschätzter Erwartungswert des Varianzparameterschätzers :  0.998
```

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Einfache lineare Regression

Wir betrachten das generative Modell der einfachen linearen Regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, ..., n, \tag{25}$$

Dann gilt

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{y} - \frac{c_{xy}}{s_x^2}\bar{x} \\ \frac{c_{xy}}{s_x^2} \end{pmatrix} \text{ and } \hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \tag{26}$$

wobei $\bar{x}$ und $\bar{y}$ die Stichprobenmittel der $x_1, ..., x_n$ und $y_1, ..., y_n$, respektive, bezeichnen, c_{xy} die Stichprobenkovarianz der $x_1, ..., x_n$ und s_x^2 die Stichprobenvarianz der $x_1, ..., x_n$ bezeichnen, wie unten gezeigt.

Wir halten fest, dass für eine parametrische Designmatrixspalte sich der entsprechende Betaparameterschätzer aus der Stichprobenkovarianz der respektiven Spalte mit den Daten geteilt durch die Stichprobenvarianz der entsprechenden Spalte ergibt und somit einer "standardisierten" Stichprobenkovarianz entspricht. Wie in (1) Regression sind die Bezeichnungen "Stichproben"kovarianz und "Stichproben"varianz bezüglich der $x_1, ..., x_n$ hier lediglich formal gemeint, da keine Annahme zugrundeliegt, dass die $x_1, ..., x_n$ Realisierungen von Zufallsvariablen sind. Die $x_1, ..., x_n$ sind vorgegebene Werte. Ein Vergleich mit den Parametern der Ausgleichsgerade in (1) Regression zeigt weiterhin die Identität der Betaparameterschätzerkomponenten $\hat{\beta}_0$ und $\hat{\beta}_1$ mit den dort unter dem Kriterium der Minimierung der quadrierten vertikalen Abweichungen hergeleiteten Parametern. Dies überrascht nicht, da sowohl $\hat{\beta}$ als auch die Parameter der Ausgleichsgerade den Wert

$$q(\tilde{\beta}) = \sum_{i=1}^{n}(y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i))^2 = (y - X\tilde{\beta})^T(y - X\tilde{\beta}) \tag{27}$$

hinsichtlich $\tilde{\beta}$ minimieren.

Einfache lineare Regression

Um die Form des Betaparameterschätzers herzuleiten, halten wir zunächst fest, dass

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n}(x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y}) \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \bar{y} - \sum_{i=1}^{n} \bar{x} y_i + \sum_{i=1}^{n} \bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i - \bar{x} \sum_{i=1}^{n} y_i + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y} n \bar{x} - \bar{x} n \bar{y} + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y},
\end{aligned}
\tag{28}
$$

Weiterhin halten wir fest, dass

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2x_i\bar{x} + \sum_{i=1}^{n} \bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2.
\end{aligned}
\tag{29}
$$

Einfache lineare Regression

Aus der Definition von $\hat{\beta}$ ergibt sich

$$
\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T y \\
&= \left(\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
&= \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix} \\
&= \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.
\end{aligned}
\tag{30}
$$

Die Inverse von $X^T X$ ist gegeben durch

$$
\frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix},
\tag{31}
$$

Einfache lineare Regression

weil

$$
\begin{aligned}
&\frac{1}{s_x^2}\begin{pmatrix}\frac{s_x^2}{n}+\bar{x}^2 & -\bar{x}\\ -\bar{x} & 1\end{pmatrix}\begin{pmatrix} n & n\bar{x}\\ n\bar{x} & \sum_{i=1}^n x_i^2\end{pmatrix}\\
&=\frac{1}{s_x^2}\begin{pmatrix}\frac{ns_x^2}{n}+n\bar{x}^2-n\bar{x}^2 & \frac{s_x^2 n\bar{x}}{n}+n\bar{x}^2\bar{x}-\bar{x}\sum_{i=1}^n x_i^2\\ -\bar{x}n+n\bar{x} & -n\bar{x}^2+\sum_{i=1}^n x_i^2\end{pmatrix}\\
&=\frac{1}{s_x^2}\begin{pmatrix} s_x^2 & s_x^2\bar{x}-\bar{x}\left(\sum_{i=1}^n x_i^2-n\bar{x}^2\right)\\ 0 & \sum_{i=1}^n x_i^2-n\bar{x}^2\end{pmatrix}\\
&=\frac{1}{s_x^2}\begin{pmatrix} s_x^2 & s_x^2\bar{x}-\bar{x}s_x^2\\ 0 & s_x^2\end{pmatrix}\\
&=\frac{1}{s_x^2}\begin{pmatrix} s_x^2 & 0\\ 0 & s_x^2\end{pmatrix}\\
&=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.
\end{aligned}
\tag{32}
$$

Einfache lineare Regression

Es ergibt sich also

$$
\begin{aligned}
\hat{\beta} = \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{s_x^2} & -\frac{\bar{x}}{s_x^2} \\ -\frac{\bar{x}}{s_x^2} & \frac{1}{s_x^2} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix} & = \begin{pmatrix} \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}\right) n\bar{y} - \frac{\bar{x}\sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i}{s_x^2} - \frac{n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\
& = \begin{pmatrix} \frac{n\bar{y}}{n} + \frac{\bar{x}^2 n\bar{y}}{s_x^2} - \frac{\bar{x}\sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\
& = \begin{pmatrix} \bar{y} + \frac{\bar{x}n\bar{x}\bar{y} - \bar{x}\sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\
& = \begin{pmatrix} \bar{y} - \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2}\bar{x} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\
& = \begin{pmatrix} \bar{y} - \frac{c_{xy}}{s_x^2}\bar{x} \\ \frac{c_{xy}}{s_x^2} \end{pmatrix}.
\end{aligned}
\tag{33}
$$

Parameterschätzung

```r
# Libraries
library(MASS)                                           # Multivariate Normalverteilung
library(matlib)                                         # Matrizenrechnung

# Modellformulierung
n          = 10                                         # Anzahl Datenpunkte
p          = 2                                          # Anzahl Betaparameter
x          = 1:n                                        # Prädiktorwerte
X          = matrix(c(rep(1,n),x), nrow = n)            # Designmatrix
I_n        = diag(n)                                    # n x n Einheitsmatrix
beta       = matrix(c(0,1), nrow = p)                   # wahrer, aber unbekannter, Betaparameter
sigsqr     = 1                                          # wahrer, aber unbekannter, Varianzparameter

# Datenrealisierung
y          = mvrnorm(1, X %*% beta, sigsqr*I_n)         # eine Realisierung eines n-dimensionalen ZVs

# Parameterschätzung
beta_hat   = inv(t(X) %*% X) %*% t(X) %*% y             # Betaparameterschätzer
eps_hat    = y - X %*% beta_hat                         # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) /(n-p)            # Varianzparameterschätzer

# Ausgabe
cat("beta       : ", beta,
    "\nhat{beta}  : ", beta_hat,
    "\nsigsqr     : ", sigsqr,
    "\nhat{sigsqr}: ", sigsqr_hat)
```

```
> beta       :  0 1
> hat{beta}  :  -1.04 1.07
> sigsqr     :  1
> hat{sigsqr}:  0.337
```

Simulation der Schätzerunverzerrtheit

```r
# Libraries
library(MASS)                                         # Multivariate Normalverteilung
library(matlib)                                       # Matrizenrechnung

# Modellformulierung
n          = 10                                       # Anzahl Datenpunkte
p          = 2                                        # Anzahl Betaparameter
x          = 1:n                                      # Prädiktorwerte
X          = matrix(c(rep(1,n),x), nrow = n)          # Designmatrix
I_n        = diag(n)                                  # n x n Einheitsmatrix
beta       = matrix(c(0,1), nrow = p)                 # wahrer, aber unbekannter, Betaparameter
sigsqr     = 1                                        # wahrer, aber unbekannter, Varianzparameter

# Frequentistische Simulation
nsim       = 1e4                                      # Anzahl Realisierungen des n-dimensionalen ZVs
beta_hat   = matrix(rep(NaN,p*nsim), nrow = p)        # \hat{\beta} Realisierungsarray
sigsqr_hat = rep(NaN,nsim)                            # \hat{sigsqr} Realisierungsarray
for(i in 1:nsim){                                     # Simulationsiterationen
  y             = mvrnorm(1, X %*% beta, sigsqr*I_n) # Datenrealisierung
  beta_hat[,i]  = inv(t(X) %*% X) %*% t(X) %*% y      # Betaparameterschätzer
  eps_hat       = y - X %*% beta_hat[,i]              # Residuenvektor
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat) /(n-p)     # Varianzparameterschätzer
}

# Ausgabe
cat("Wahrer, aber unbekannter, Betaparameter                  : ", beta,
    "\nGeschätzter Erwartungswert des Betaparameterschätzers    : ", rowMeans(beta_hat),
    "\nWahrer, aber unbekannter, Varianzparameter               : ", sigsqr,
    "\nGeschätzter Erwartungswert des Varianzparameterschätzers : ", mean(sigsqr_hat))
```

```
> Wahrer, aber unbekannter, Betaparameter                  :  0 1
> Geschätzter Erwartungswert des Betaparameterschätzers    :  -0.00191 1
> Wahrer, aber unbekannter, Varianzparameter               :  1
> Geschätzter Erwartungswert des Varianzparameterschätzers :  1.01
```

Selbstkontrollfragen

1. Geben Sie das Betaparameterschätzer Theorem wieder.
2. Geben Sie das Varianzparameterschätzer Theorem wieder.
3. Geben Sie die Beta- und Varianzparameterschätzer des ALM Szenarios von n unabhängigen und identisch normalverteilten Zufallsvariablen wieder.
4. Geben Sie die Beta- und Varianzparameterschätzer des ALM Szenarios der einfachen linearen Regression wieder.
5. Wie unterscheiden sich die Betaparameterschätzer des ALM Szenarios der einfachen linearen Regression und die Parameter der Ausgleichsgerade aus Einheit (1) Regression?
6. Simulieren Sie die Unverzerrtheit des Varianzparameterschätzers im ALM Szenario von n unabhängigen und identisch normalverteilten Zufallsvariablen in einem R Skript.
7. Simulieren Sie die Unverzerrtheit des Varianzparameterschätzers im ALM Szenario der einfachen linearen Regression in einem R Skript.

References

Rencher, Alvin C., and G. Bruce Schaalje. 2008. *Linear Models in Statistics*. 2nd ed. Hoboken, N.J: Wiley-Interscience.

Searle, S. R. 1971. *Linear Models*. A Wiley Publication in Mathematical Statistics. New York: Wiley.