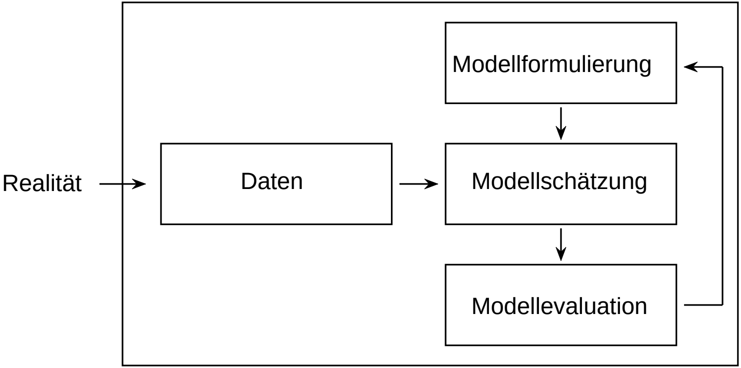
Allgemeines Lineares Modell

BSc Psychologie SoSe 2023

Prof. Dr. Dirk Ostwald

(6) Parameterschätzung

Naturwissenschaft



Modellformulierung

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

Modellschätzung

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

Modellevaluation

$$
T = \frac{c^T\hat{\beta} - c^T\beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}, F = \frac{(\hat{\varepsilon}_0^T\hat{\varepsilon}_0 - \hat{\varepsilon}^T\hat{\varepsilon})/p_1}{\hat{\varepsilon}^T\hat{\varepsilon}/(n-p)} \tag{3}
$$

Standardprobleme Frequentistischer Inferenz

(1) Parameterschätzung

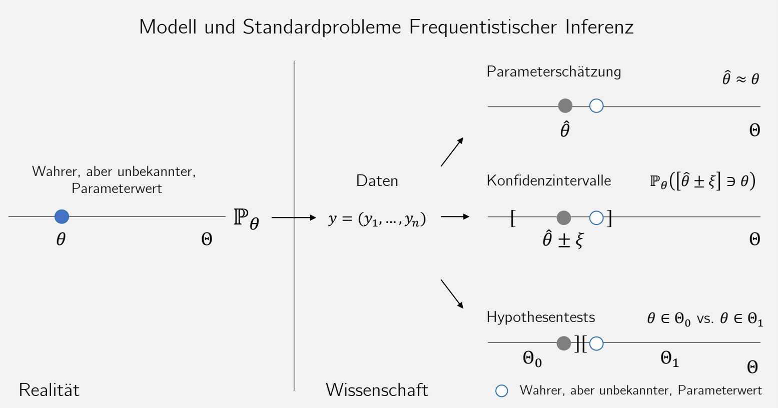
Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für wahre, aber unbekannte, Parameterwerte oder Funktionen dieser abzugeben, typischerweise mithilfe von Daten.

(2) Konfidenzintervalle

Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der angenommenen Verteilung der Daten eine quantitative Aussage über die mit Schätzwerten assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Ziel des Hypothesentestens ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst zuverlässigen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes liegt.



$\theta := (\beta, \sigma^2)$, $\Theta := \mathbb{R}^p \times \mathbb{R}_{>0}$ $\mathbb{P}_\theta(\upsilon) := \mathbb{P}_{\beta,\sigma^2}(\upsilon)$ mit WDF $p_{\beta,\sigma^2}(y) := N(y; X\beta, \sigma^2 I_n)$

Standardannahmen Frequentistischer Inferenz

Gegeben sei das Allgemeine Lineare Modell. 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

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

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

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

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

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

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 $\hat{\beta}^{(1)}$, $\hat{\beta}^{(2)}$, $\hat{\beta}^{(3)}$, $\hat{\beta}^{(4)}$, ... also die Verteilung der Zufallsvariable $\hat{\beta} := (X^TX)^{-1}X^T\upsilon$? 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

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Theorem (Betaparameterschätzer)

Es sei

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

das ALM und es sei

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

der *Betaparameterschätzer*. Dann gilt, dass $\hat{\beta}$ die Summe der Abweichungsquadrate minimiert,

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

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.
- Da $\hat{\beta}$ die Summe der Abweichungsquadrate minimiert, heißt $\hat{\beta}$ auch Kleinste-Quadrate (KQ) Schätzer.
- Die $\tilde{\beta}$ Notation des Maximierungarguments dient lediglich zur Abgrenzung vom w.a.u. β.
- Als ML Schätzer ist $\hat{\beta}$ weiterhin 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

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

minimiert. Dazu halten wir zunächst fest, dass

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

Weiterhin gilt dann auch, dass

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

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 die Summe der Abweichungsquadrate

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

Beweis (fortgeführt)

Es ergibt sich dann

$$
\begin{aligned}
&(\upsilon - X\tilde{\beta})^T(\upsilon - X\tilde{\beta}) \\
&= (\upsilon - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta})^T(\upsilon - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta}) \\
&= ((\upsilon - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta}))^T((\upsilon - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta})) \\
&= (\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta}) + (\upsilon - X\hat{\beta})^T X(\hat{\beta} - \tilde{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T(\upsilon - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \\
&= (\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta}) + 0_p^T(\hat{\beta} - \tilde{\beta}) + (\hat{\beta} - \tilde{\beta})^T 0_p + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \\
&= (\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}).
\end{aligned}
$$

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

$$
(\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \ge 0 \tag{12}
$$

nimmt dieser Term genau dann seinen Minimalwert 0 an, wenn

$$
(\hat{\beta} - \tilde{\beta}) = 0_p \Leftrightarrow \tilde{\beta} = \hat{\beta}. \tag{13}
$$

Also gilt

$$
\hat{\beta} = \arg\min_{\tilde{\beta}} (\upsilon - X\tilde{\beta})^T(\upsilon - X\tilde{\beta}). \tag{14}
$$

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

$$\ell : \mathbb{R}^p \to \mathbb{R}, \tilde{\beta} \mapsto \ln p_{\tilde{\beta}}(y) = \ln N(y; X\tilde{\beta}, \sigma^2 I_n) \tag{15}$$

wobei gilt, dass

$$\begin{aligned}\ln N(y; X\tilde{\beta}, \sigma^2 I_n) &= \ln \left((2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} (\upsilon - X\tilde{\beta})^T (\upsilon - X\tilde{\beta}) \right) \right) \\ &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma^2 I_n| - \frac{1}{2\sigma^2} (\upsilon - X\tilde{\beta})^T (\upsilon - X\tilde{\beta}).\end{aligned} \tag{16}$$

Dabei hängt allein der Term $-\frac{1}{2\sigma^2}(\upsilon - X\tilde{\beta})^T(\upsilon - X\tilde{\beta})$ von $\tilde{\beta}$ ab. Weil aber $(\upsilon - X\tilde{\beta})^T(\upsilon - X\tilde{\beta}) \geq 0$, gilt wird dieser Term aufgrund des negativen Vorzeichen maximal, wenn $(\upsilon - X\tilde{\beta})^T(\upsilon - X\tilde{\beta})$ 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

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}\left((X^T X)^{-1} X^T \upsilon \right) = (X^T X)^{-1} X^T \mathbb{E}(\upsilon) = (X^T X)^{-1} X^T X\beta = \beta. \tag{17}$$

Definition (Erklärte Daten, Residuenvektor, Residuen)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{18}$$

das Allgemeine Lineare Modell und es sei

$$\hat{\beta} := (X^T X)^{-1} X^T \upsilon. \tag{19}$$

der Betaparameterschätzer. Dann heißt der Zufallsvektor

$$\hat{\upsilon} := X\hat{\beta} = X(X^T X)^{-1} X^T \upsilon \tag{20}$$

die *erklärten Daten*, der Zufallsvektor

$$\hat{\varepsilon} := \upsilon - \hat{\upsilon} = \upsilon - X\hat{\beta} \tag{21}$$

heißt *Residuenvektor* und für $i = 1, ..., n$ heißen die Komponenten dieses Zufallsvektors

$$\hat{\varepsilon}_i := \upsilon_i - \hat{\upsilon}_i = \upsilon_i - (X\hat{\beta})_i \tag{22}$$

die *Residuen*.

Bemerkungen

- Die Begriffe sind analog zu den in Einheit (2) Korrelation eingeführten Begriffen.

Theorem (Varianzparameterschätzer)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{23}$$

das ALM in generativer Form. Dann ist

$$\hat{\sigma}^2 := \frac{(\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta})}{n-p} \tag{24}$$

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

Bemerkungen

- Es handelt sich bei $\hat{\sigma}^2$ *nicht* um einen Maximum Likelihood 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 bieten sich für $\hat{\sigma}^2$ auch folgende Schreibweisen an:

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-p} = \frac{1}{n-p}\sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n-p}\sum_{i=1}^n (\upsilon_i - (X\beta)_i)^2 \tag{25}$$

- σ^2 wird also durch die eine skalierte Residualquadratsumme geschätzt.
- Der Maximum Likelihood Schätzer des Varianzparameters ist $\hat{\sigma}^2_{\text{ML}} := \frac{1}{n}\hat{\varepsilon}^T\hat{\varepsilon}$.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Frequentistische Schätzerverteilungen

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,

$$\upsilon_i \sim N(\mu, \sigma^2) \text{ für } i = 1, ..., n. \tag{26}$$

Dann gilt, wie unten gezeigt,

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \upsilon_i =: \bar{\upsilon} \text{ und } \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\upsilon_i - \bar{\upsilon})^2 =: s_\upsilon^2. \tag{27}$$

In diesem Fall ist also der Betaparameterschätzer mit dem Stichprobenmittel $\bar{\upsilon}$ der $\upsilon_1, ..., \upsilon_n$ und der Varianzparameterschätzer mit der Stichprobenvarianz s_υ^2 der $\upsilon_1, ..., \upsilon_n$ identisch.

Für $\hat{\beta}$ ergibt sich

$$
\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T \upsilon \\
&= \left(1_n^T 1_n\right)^{-1} 1_n^T \upsilon \\
&= \left(\begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} \\
&= n^{-1} \sum_{i=1}^{n} \upsilon_i \\
&= \frac{1}{n} \sum_{i=1}^{n} \upsilon_i \\
&=: \bar{\upsilon}.
\end{aligned}
$$

Unabhängige und identisch normalverteilte Zufallsvariablen

Für $\hat{\sigma}^2$ ergibt sich

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-1}\left(\upsilon - X\hat{\beta}\right)^T\left(\upsilon - X\hat{\beta}\right) \\
&= \frac{1}{n-1}\left(\upsilon - 1_n\bar{\upsilon}\right)^T\left(\upsilon - 1_n\bar{\upsilon}\right) \\
&= \frac{1}{n-1}\left(\begin{pmatrix}\upsilon_1\\ \vdots \\ \upsilon_n\end{pmatrix} - \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\bar{\upsilon}\right)^T\left(\begin{pmatrix}\upsilon_1\\ \vdots \\ \upsilon_n\end{pmatrix} - \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\bar{\upsilon}\right) \\
&= \frac{1}{n-1}\begin{pmatrix}\upsilon_1 - \bar{\upsilon} & \cdots & \upsilon_n - \bar{\upsilon}\end{pmatrix}\begin{pmatrix}\upsilon_1 - \bar{\upsilon}\\ \vdots \\ \upsilon_n - \bar{\upsilon}\end{pmatrix} \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(\upsilon_i - \bar{\upsilon}\right)^2 \\
&=: s_\upsilon^2.
\end{aligned}
$$

Unabhängige und identisch normalverteilte Zufallsvariablen

Parameterschätzung

```r
# Modellformulierung
library(MASS)                                          # Multivariate Normalverteilung
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
```

Unabhängige und identisch normalverteilte Zufallsvariablen

Simulation der Schätzerunverzerrtheit

```r
# Modellformulierung
library(MASS)                                             # Multivariate Normalverteilung
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

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Wir betrachten das Modell der einfachen linearen Regression

$$\upsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, ..., n, \tag{28}$$

Dann gilt wie unten gezeigt, dass

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{\upsilon} - \frac{c_{x\upsilon}}{s_x^2}\bar{x} \\ \frac{c_{x\upsilon}}{s_x^2} \end{pmatrix} \text{ und } \hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n (\upsilon_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \tag{29}$$

wobei

- $\bar{x}$ und $\bar{\upsilon}$ die Stichprobenmittel der $x_1, ..., x_n$ und $\upsilon_1, ..., \upsilon_n$, respektive, bezeichnen
- $c_{x\upsilon}$ die Stichprobenkovarianz der $x_1, ..., x_n$ und $\upsilon_1, ..., \upsilon_n$ bezeichnet
- s_x^2 die Stichprobenvarianz der $x_1, ..., x_n$ bezeichnet.

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.

Einfache lineare Regression

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.

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

$$q(\tilde{\beta}) = \sum_{i=1}^{n} (\upsilon_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i))^2 = (\upsilon - X\tilde{\beta})^T (\upsilon - X\tilde{\beta}) \tag{30}$$

hinsichtlich $\tilde{\beta}$ minimieren.

Einfache lineare Regression

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

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})(\upsilon_i - \bar{\upsilon}) &= \sum_{i=1}^{n}(x_i\upsilon_i - x_i\bar{\upsilon} - \bar{x}\upsilon_i + \bar{x}\bar{\upsilon}) \\
&= \sum_{i=1}^{n} x_i\upsilon_i - \sum_{i=1}^{n} x_i\bar{\upsilon} - \sum_{i=1}^{n}\bar{x}\upsilon_i + \sum_{i=1}^{n}\bar{x}\bar{\upsilon} \\
&= \sum_{i=1}^{n} x_i\upsilon_i - \bar{\upsilon}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n}\upsilon_i + n\bar{x}\bar{\upsilon} \\
&= \sum_{i=1}^{n} x_i\upsilon_i - \bar{\upsilon}n\bar{x} - \bar{x}n\bar{\upsilon} + n\bar{x}\bar{\upsilon} \\
&= \sum_{i=1}^{n} x_i\upsilon_i - n\bar{x}\bar{\upsilon} - n\bar{x}\bar{\upsilon} + n\bar{x}\bar{\upsilon} \\
&= \sum_{i=1}^{n} x_i\upsilon_i - n\bar{x}\bar{\upsilon},
\end{aligned}
\tag{31}
$$

Weiterhin halten wir fest, dass

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2x_i\bar{x} + \sum_{i=1}^{n} \bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2.
\end{aligned}
\tag{32}
$$

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

$$
\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T \upsilon \\
&= \left(\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} \\
&= \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n \upsilon_i \\ \sum_{i=1}^n x_i \upsilon_i \end{pmatrix} \\
&= \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} n\bar{\upsilon} \\ \sum_{i=1}^n x_i \upsilon_i \end{pmatrix}.
\end{aligned}
\tag{33}
$$

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

$$
\frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix},
\tag{34}
$$

weil

$$
\begin{aligned}
&\frac{1}{s_x^2}\begin{pmatrix} \frac{s_x^2}{n}+\bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}\begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix} \\
&= \frac{1}{s_x^2}\begin{pmatrix} \frac{ns_x^2}{n} + n\bar{x}^2 - n\bar{x}^2 & \frac{s_x^2 n\bar{x}}{n} + n\bar{x}^2\bar{x} - \bar{x}\sum_{i=1}^n x_i^2 \\ -\bar{x}n + n\bar{x} & -n\bar{x}^2 + \sum_{i=1}^n x_i^2 \end{pmatrix} \\
&= \frac{1}{s_x^2}\begin{pmatrix} s_x^2 & s_x^2\bar{x} - \bar{x}\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right) \\ 0 & \sum_{i=1}^n x_i^2 - n\bar{x}^2 \end{pmatrix} \\
&= \frac{1}{s_x^2}\begin{pmatrix} s_x^2 & s_x^2\bar{x} - \bar{x}s_x^2 \\ 0 & s_x^2 \end{pmatrix} \\
&= \frac{1}{s_x^2}\begin{pmatrix} s_x^2 & 0 \\ 0 & s_x^2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned}
\tag{35}
$$

Es ergibt sich also

$$
\begin{aligned}
\hat{\beta} = \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{s_x^2} & -\frac{\bar{x}}{s_x^2} \\ -\frac{\bar{x}}{s_x^2} & \frac{1}{s_x^2} \end{pmatrix} \begin{pmatrix} n\bar{\upsilon} \\ \sum_{i=1}^n x_i \upsilon_i \end{pmatrix}
&= \begin{pmatrix} \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}\right) n\bar{\upsilon} - \frac{\bar{x}\sum_{i=1}^n x_i \upsilon_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i \upsilon_i}{s_x^2} - \frac{n\bar{x}\bar{\upsilon}}{s_x^2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{n\bar{\upsilon}}{n} + \frac{\bar{x}^2 n\bar{\upsilon}}{s_x^2} - \frac{\bar{x}\sum_{i=1}^n x_i \upsilon_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i \upsilon_i - n\bar{x}\bar{\upsilon}}{s_x^2} \end{pmatrix} \\
&= \begin{pmatrix} \bar{\upsilon} + \frac{\bar{x} n\bar{x}\bar{\upsilon} - \bar{x}\sum_{i=1}^n x_i \upsilon_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i \upsilon_i - n\bar{x}\bar{\upsilon}}{s_x^2} \end{pmatrix} \\
&= \begin{pmatrix} \bar{\upsilon} - \frac{\sum_{i=1}^n x_i \upsilon_i - n\bar{x}\bar{\upsilon}}{s_x^2}\bar{x} \\ \frac{\sum_{i=1}^n x_i \upsilon_i - n\bar{x}\bar{\upsilon}}{s_x^2} \end{pmatrix} \\
&= \begin{pmatrix} \bar{\upsilon} - \frac{c_{x\upsilon}}{s_x^2}\bar{x} \\ \frac{c_{x\upsilon}}{s_x^2} \end{pmatrix}.
\end{aligned}
\tag{36}
$$

Parameterschätzung

```
# Modellformulierung
library(MASS)                                       # Multivariate Normalverteilung
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    = 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       :  0 1
> hat{beta}  :  -1.04 1.07
> sigsqr     :  1
> hat{sigsqr}:  0.337
```

Simulation der Schätzerunverzerrtheit

```r
# Modellformulierung
library(MASS)                                         # Multivariate Normalverteilung
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]   = 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    : ", 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
```

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Theorem (Frequentistische Verteilung des Betaparameterschätzers)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{37}$$

das ALM. Weiterhin sei

$$\hat{\beta} := \left(X^T X\right)^{-1} X^T \upsilon \tag{38}$$

der Betaparameterschätzer. Dann gilt

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X^T X)^{-1}\right). \tag{39}$$

Bemerkungen

- Es gilt also wie bereits gesehen $\mathbb{E}(\hat{\beta}) = \beta$ und außerdem $\mathbb{C}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$.
- Die Varianzen der Komponenten von $\hat{\beta}$ sind die Diagonalelemente von $\mathbb{C}(\hat{\beta})$, also

$$\mathbb{V}(\hat{\beta}_i) = (\sigma^2 (X^T X)^{-1})_{ii} \text{ für } i = 1, ...p. \tag{40}$$

- Die Streuung von $\hat{\beta}$ hängt von σ^2 und der Designmatrix X ab. σ^2 ist ein experimentell nicht zu beinflussender wahrer, aber unbekannter, Parameter X dagegen kann so gewählt werden, um zum Beispiel die Diagonalelemente von $\mathbb{C}(\hat{\beta})$ bei festem σ^2 zu minimieren.

Frequentistische Schätzerverteilungen

<u>Beweis</u>

Das Theorem folgt direkt mit dem Theorem zur linearen Transformation von multivariaten Normalverteilung aus Einheit (4) Normalverteilungen. Speziell gilt hier:

$$\hat{\beta} \sim N\left((X^TX)^{-1}X^TX\beta, (X^TX)^{-1}X^T(\sigma^2 I_n)((X^TX)^{-1}X^T)^T\right). \tag{41}$$

Der Erwartungswertparameter vereinfacht sich dann zu

$$(X^TX)^{-1}X^TX\beta = \beta. \tag{42}$$

Der Kovarianzmatrixparamter vereinfacht sich wie folgt

$$\begin{aligned}(X^TX)^{-1}X^T(\sigma^2 I_n)((X^TX)^{-1}X^T)^T &= (X^TX)^{-1}X^T(\sigma^2 I_n)X(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1}.\end{aligned} \tag{43}$$

Dabei folgt hier die erste Gleichung aus der Tatsache, dass sowohl X^TX als auch ihre Inverse $(X^TX)^{-1}$ symmetrische Matrizen sind.

Damit gilt dann insgesamt aber sofort

$$\hat{\beta} \sim N\left(\beta, \sigma^2(X^TX)^{-1}\right). \tag{44}$$

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \tag{45}$$

das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen. Wir haben bereits gesehen, dass $\hat{\beta} = \bar{\upsilon}$. Das Theorem zur Frequentistischen Verteilung des Betaparameterschätzers impliziert damit

$$\bar{\upsilon} \sim N\left(\mu, \frac{\sigma^2}{n}\right). \tag{46}$$

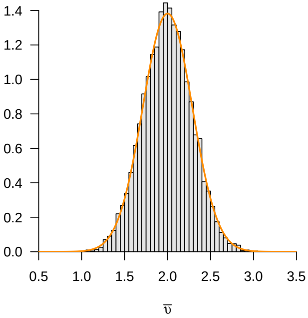
Das Stichprobenmittel von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter μ und Varianzparameter σ^2 ist also normalverteilt mit Erwartungswertparameter μ und Varianzparameter σ^2/n. Wir haben diese Tatsache bereits in Einheit (8) Transformationen der Normalverteilungen in Wahrscheinlichkeitstheorie und Frequentistische Inferenz unter dem Begriff der Mittelwertstransformation kennengelernt.

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

```
# Modellformulierung
library(MASS)                                          # Multivariate Normalverteilung
n        = 12                                          # Anzahl von Datenpunkten
p        = 1                                           # Anzahl von Betparametern
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 Realisierungen des n-dimensionalen ZVs
beta_hat = rep(NaN,nsim)                               # \hat{\beta} Realisierungsarray
for(i in 1:nsim){
  y           = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
  beta_hat[i] = solve(t(X) %*% X) %*% t(X) %*% y   # \hat{\beta} = (X^T)X^{-1}X^T\upsilon
}
```

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

$$\bar{\upsilon} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$



Beispiel (2) Einfache lineare Regression

Es sei

$$
\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta \in \mathbb{R}^2, \sigma^2 > 0. \tag{47}
$$

das ALM Szenario der einfachen linearen Regression. Wir haben bereits gesehen, dass

$$
\sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \text{ mit } s_x^2 := \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{48}
$$

Die Varianz des Offsetparameterschätzers hängt also sowohl von der Summe der quadrierten Differenzen und dem Stichprobenmittel der unabhängigen Variablen $x_1, ..., x_n$ ab, wohingegen die Varianz des Steigungsparameterschätzers nur von der Summe der quadrierten Differenzen der $x_1, ..., x_n$ abhängt. Die Kovarianz von Offset- und Steigungsparameterschätzern hängt vom Mittelwert der $x_1, ..., x_n$ ab.

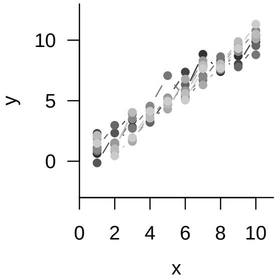
Beispiel (2) Einfache lineare Regression

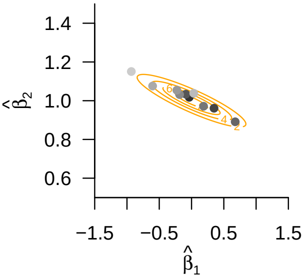
```
# Modellformulierung
library(MASS)                                              # Multivariate Normalverteilung
n        = 10                                              # Anzahl von Datenpunkten
p        = 2                                               # Anzahl von Betparametern
x        = 1:n                                             # 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   = .5                                              # wahrer,aber unbekannter,Varianzparameter

# Frequentistische Simulation
nsim     = 10                                              # Anzahl Realisierungen n-dimensionaler ZV
y        = matrix(rep(NaN,n*nsim), nrow = n)               # y Realisierungsarray
beta_hat = matrix(rep(NaN,p*nsim), nrow = p)               # \hat{\beta} Realisierungsarray
for(i in 1:nsim){
  y[,i]        = mvrnorm(1, X %*% beta, sigsqr*I_n)        # eine Realisierung n-dimensionaler ZV
  beta_hat[,i] = solve(t(X) %*% X) %*% t(X) %*% y[,i] # \hat{\beta} = (X^T)X^{-1}X^T\upsilon
}
```

Beispiel (2) Einfache lineare Regression

$$\upsilon \sim (X\beta, \sigma^2 I_n)$$

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$





Theorem (Frequentistische Verteilung des Varianzparameterschätzers)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{49}$$

das ALM. Weiterhin sei

$$\hat{\sigma}^2 = \frac{(\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta})}{n-p} \tag{50}$$

der Varianzparameterschätzer. Dann gilt

$$\frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2(n-p) \tag{51}$$

Bemerkungen

- Wir verzichten auf einen Beweis. Da es sich bei $(\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta})$ um eine Summe quadrierter normalverteilter Zufallsvariablen handelt, liegt die χ^2-Verteilung im Lichte der χ^2 Transformation aus Einheit (8) Transformationen der Normalverteilung in Wahrscheinlichkeitstheorie und Frequentistische Inferenz zumindest nahe.

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n\times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \tag{52}$$

das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen. Wir haben bereits gesehen, dass in diesem Fall $\hat{\beta}$ mit dem Stichprobenmittel $\bar{\upsilon}$ identisch ist und dass $\hat{\sigma}^2$ mit der Stichprobenvarianz s_υ^2 übereinstimmt.

In Einheit (11) Konfidenzintervalle von Wahrscheinlichkeitstheorie und Frequentistische Inferenz hatten wir für den Fall von n unabhängig und identisch normalverteilten Zufallsvariablen die Statistik

$$U := \frac{n-1}{\sigma^2} S^2 \tag{53}$$

definiert und festgehalten, dass

$$U \sim \chi^2(n-1). \tag{54}$$

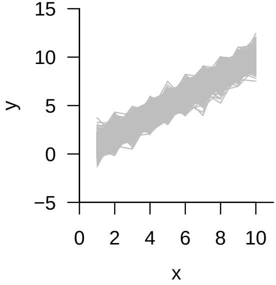
Offenbar ist U für $p = 1$ mit der im obigen Theorem betrachten Zufallsvariable $\frac{n-p}{\sigma^2}\hat{\sigma}^2$ identisch.

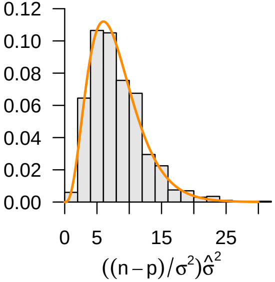
Beispiel (2) Einfache lineare Regression

```
# Modellfumuliertung
library(MASS)                                     # multivariate Normalverteilung
n           = 10                                  # Anzahl von Datenpunkten
p           = 2                                   # Anzahl von Betparametern
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      = .5                                  # wahrer,aber unbekannter,Varianzparameter

# Frequentistische Simulation
nsim        = 1e3                                 # Anzahl Realisierungen n-dimensionaler ZV
y           = matrix(rep(NaN,n*nsim), nrow = n)   # y Realisierungsarray
beta_hat    = matrix(rep(NaN,p*nsim), nrow = p)   # \hat{\beta} Realisierungsarray
sigsqr_hat = rep(NaN, nsim)                       # \hat{\sigma}^2 Realisierungsarray
for(i in 1:nsim){
  y[,i]         = mvrnorm(1, X %*% beta, sigsqr*I_n)   # eine Realisierung n-dimensionaler ZV
  beta_hat[,i]  = solve(t(X) %*% X) %*% t(X) %*% y[,i] # \hat{\beta}     = (X^T)X^{-1}X^T\upsilon
  eps_hat       = y[,i] - X %*% beta_hat[,i]           # \hat{\eps}      = \upsilon-X\hat{\beta}
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat)/(n-p)       # \hat{\sigma}^2 =\hat{\eps}^T\hat{\eps}/(n-p)
}
U = ((n-p)/sigsqr)*sigsqr_hat                     # \chi^2 verteilte Zufallsvariable
```

Beispiel (2) Einfache lineare Regression

$$\upsilon \sim (X\beta, \sigma^2 I_n) \qquad \frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2(n-p)$$





Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie das Theorem zum Betaparameterschätzer wieder.
2. Warum ist der Betaparameterschätzer ein Maximum-Likelihood Schätzer?
3. Geben Sie das Theorem zum Varianzparameterschätzer wieder-
4. Geben Sie die Parameterschätzer bei n unabhängigen und identisch normalverteilten Zufallsvariablen an.
5. Geben Sie die Parameterschätzer bei einfacher linearer Regression an.
6. Geben Sie das Theorem zur Verteilung des Betaparameterschätzers wieder.
7. Geben Sie das Theorem zur Verteilung des Varianzparameterschätzers wieder.

Referenzen

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.