12. Kleinste-Quadrate-Schätzer und Konfidenzintervalle

12.1 Kleinste-Quadrate (KQ)- Schätzung

12.2 Erwartungswert

12.3 Varianz

12.4 Konfidenzintervalle

12.1 Kleinste-Quadrate (KQ)- Schätzung

- Gegeben: lineares Modell der Form

$$y(\boldsymbol{x}) = \boldsymbol{\beta}'\boldsymbol{x} + \epsilon = \hat{y}(\boldsymbol{x}) + \epsilon, \quad \epsilon \sim i.i.d.$$

  welches *alle Gauß-Markow-Annahmen mit Ausnahme der gaußverteilten Residualterme* erfüllt

- Gegeben sind ebenfalls n vollständige Datensätze aller exogenen Faktoren und der endogenen Variablen, welche die Datenspezifikation erfüllen:

$$\{\boldsymbol{p}_i = (x_{i0}, ..., x_{iJ}, y_i)', \quad i = 1, ..., n\}$$

- Gesucht ist ein Parameterschätzer $\hat{\boldsymbol{\beta}}$, welcher die Fehlerquadratsumme zwischen den deterministischen Modellvoraussagen $\boldsymbol{x}_i'\boldsymbol{\beta}$ und der beobachteten endogenen Variable y_i bezüglich der Parameter minimiert:

$$\boxed{\hat{\boldsymbol{\beta}} = \text{arg min}_{\boldsymbol{\beta}} S(\boldsymbol{\beta})}$$

  wobei

$$S(\boldsymbol{\beta}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}).$$

Bestimmung des KQ-Schätzers

$$\begin{aligned} S &= (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}) \\ \text{[Distributivität} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\,\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + (\mathbf{X}\,\boldsymbol{\beta})'\mathbf{X}\,\boldsymbol{\beta} \\ \text{[Transpositionsregeln} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \\ \text{[Transpositionsregeln} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{y}) + \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta} \end{aligned}$$

Ableiten nach $\boldsymbol{\beta}$ und Nullsetzen mit den Regeln $\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{a}) = \boldsymbol{a}$ und $\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{A}\,\boldsymbol{\beta}) = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}$, wobei $\mathbf{A} = \mathbf{X}'\mathbf{X}$:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = 0 - 2\mathbf{X}'\boldsymbol{y} + 2\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \overset{!}{=} 0$$

$$\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} = \mathbf{X}'\boldsymbol{y} \quad |\cdot(\mathbf{X}'\mathbf{X})^{-1}$$

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{y}}$$

12.2 Erwartungswert des KQ-Schätzers

- Einzige Zufallsquelle: Residualterm $\boldsymbol{\epsilon}$ gemäß $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- Berücksichtige, dass der KQ-Schätzer linear in $\boldsymbol{y}$ ist:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\ &= \underline{\underline{\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}}}\end{aligned}$$

Erwartungswert

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\epsilon}) = \boldsymbol{\beta}$$

Der KQ-Schätzer ist **unverzerrt** unter der milden Bedingung $E(\boldsymbol{\epsilon}) = \mathbf{0}$

12.3 Varianz des KQ-Schätzers

- Gauß-Markow Bedingungen $\rightarrow \epsilon \sim$ i.i.d$N(0, \sigma^2) \rightarrow \hat{\boldsymbol{\beta}}$ ist normalverteilt (**Warum?**)
- $\Rightarrow$ komplette Verteilung ist vollständigg definiert durch die **Varianz-Kovarianz-Matrix** $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$

$$
\begin{array}{rcl}
\mathbf{V}_{\hat{\boldsymbol{\beta}}} & \overset{\text{def}}{=} & E\left((\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\right) \\
\text{[setze } \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon} \rightarrow] & = & E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}\right)'\right) \\
\text{[Matrixregeln} \rightarrow] & = & E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
[E(.) \text{ wirkt nur auf } \boldsymbol{\epsilon} \rightarrow] & = & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
\text{[Gauß-Markow} \rightarrow] & = & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_\epsilon^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
\text{[Def Matrixinverse} \rightarrow] & = & \sigma_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1}
\end{array}
$$

Die Varianz-Kovarianz-Matrix hängt nur von den *exogenen* Daten ab!

Ergebnisse

- KQ-Schätzer:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$$

- Falls $\epsilon \sim$ i.i.d. $\Rightarrow$ **Varianz-Kovarianz-Matrix** der Schätzfehler kann als **Hesse-Matrix H** der Fehlerquadratsumme $S(\boldsymbol{\beta})$ geschrieben werden:

$$\begin{aligned} \mathbf{V}_{\hat{\boldsymbol{\beta}}} &= E\left((\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = 2\sigma^2\mathbf{H}^{-1}, \\ H_{jk} &= \left.\frac{\partial^2 S}{\partial\beta_j\,\partial\beta_k}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 2(\mathbf{X}'\mathbf{X})_{jk} \end{aligned}$$

- Varianz der Schätzfehler: $V(\hat{\beta}_j) = V_{jj}$
- Korrelation der Schätzfehler: $\text{Corr}(\hat{\beta}_j, \hat{\beta}_k) = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}}$
- Die normalisierten Schätzfehler sind standardnormalverteilt: $\frac{\hat{\beta}_j-\beta_j}{\sqrt{V_{jj}}} \sim N(0,1)$

Schätzung der Residualvarianz

Residualvarianz σ^2 ist nicht bekannt $\Rightarrow$ muss ebenfalls geschätzt werden!

$$\hat{\sigma}^2 = \frac{1}{n-J-1} \sum_i \left(y_i - \hat{y}(\boldsymbol{x}_i)\right)^2 = \frac{S(\hat{\beta})}{n-J-1}$$

bzw $\quad \hat{\sigma}^2 = \dfrac{\boldsymbol{y}' \left(\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \boldsymbol{y}}{n-J-1} \Rightarrow$ Summe quadrierter Normalverteilungen

$\Rightarrow$ *geschätzte* Fehlerstatistik (*"nicht einmal der Fehler ist genau bekannt"*):

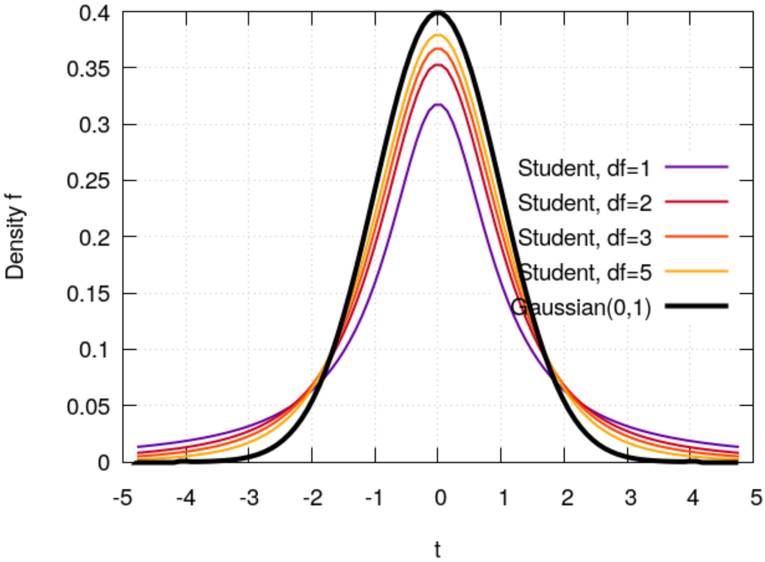
- Schätzung der Varianze-Kovarianz-Matrix:

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}} = 2\hat{\sigma}^2 \mathbf{H}^{-1} = \hat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

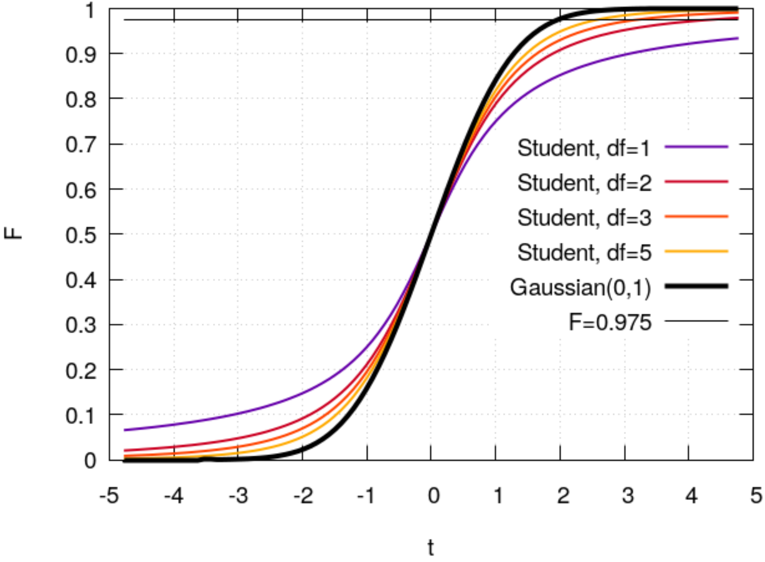
- Die normalisierten geschätzten Fehler sind **student-t** verteilt:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}_{jj}}} \sim T(n-1-J)$$

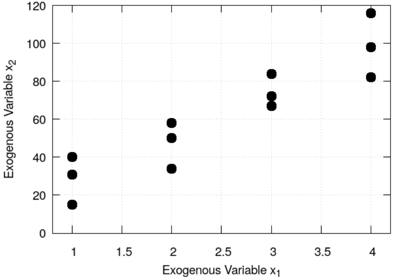
Dichtefunktionen der Student-t Verteilungen vs. Standardnormalverteilung



Verteilungsfunktionen der Student-t Verteilungen vs. Standardnormalverteilung



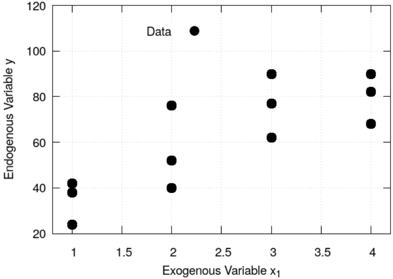
Beispiel korrelierter Schätzfehler: die bekannte Hotelbuchungsaufgabe

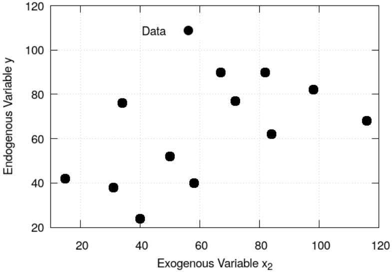


- Hotelbuchungen aus VL 11:

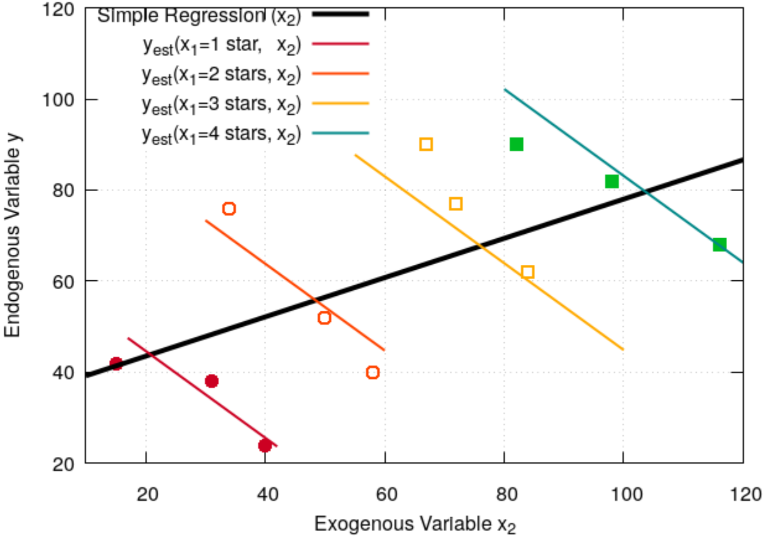
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- Exogene Faktoren: $x_0 = 1$, x_1: Qualitäts-Proxy [# stars]; x_2: Preis [€/Nacht].
- Endogene Variable: Buchungsrate [%]
- Nachfrage positiv mit Qualitätsproxy und Preis korreliert

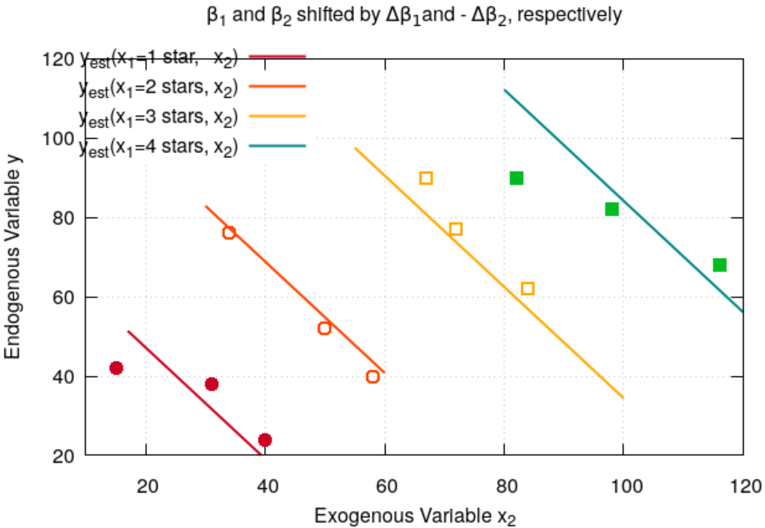




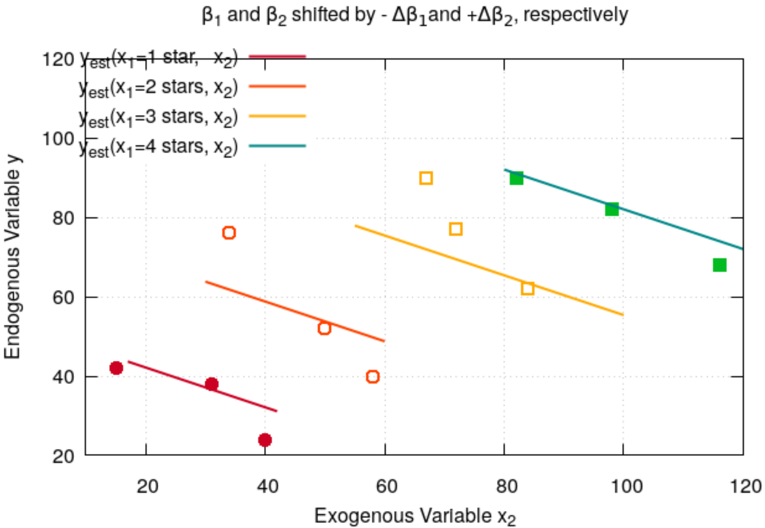
Residualfehler bei KQ-Fit der Parameter



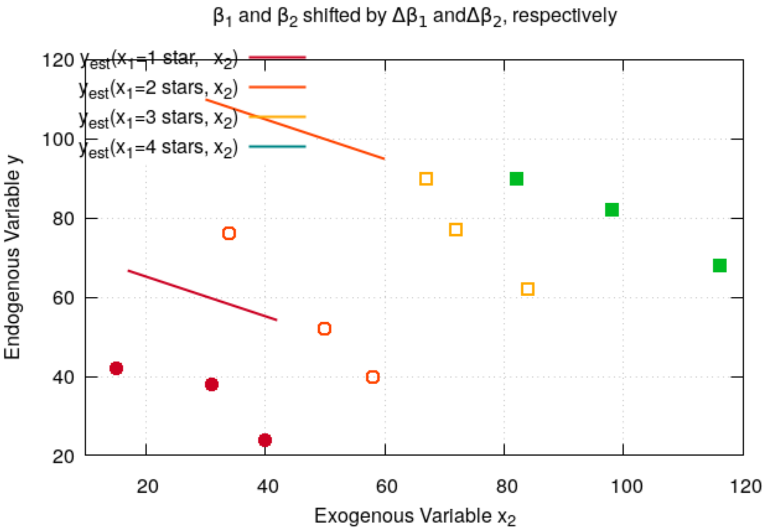
Abweichung vom Fit I: kleiner Effekt, falls β_1 und β_2 gegensätzlich fehlgefittet sind



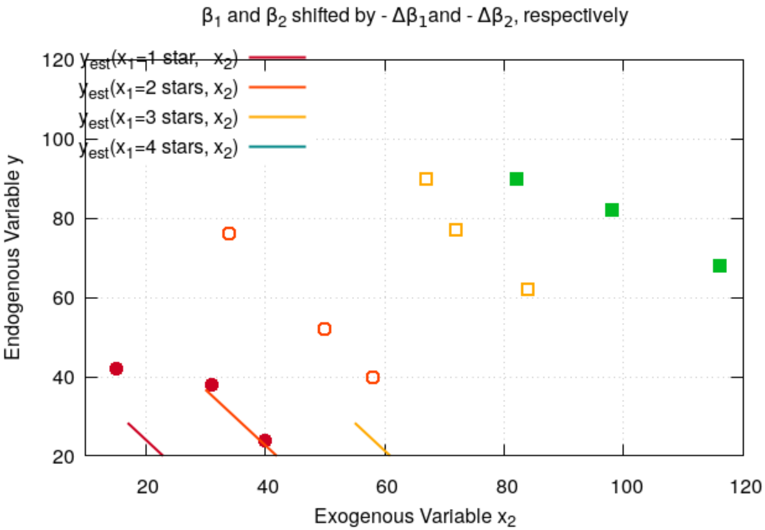
Abweichung vom Fit II: kleiner Effekt, falls β_1 und β_2 gegensätzlich fehlgefittet sind



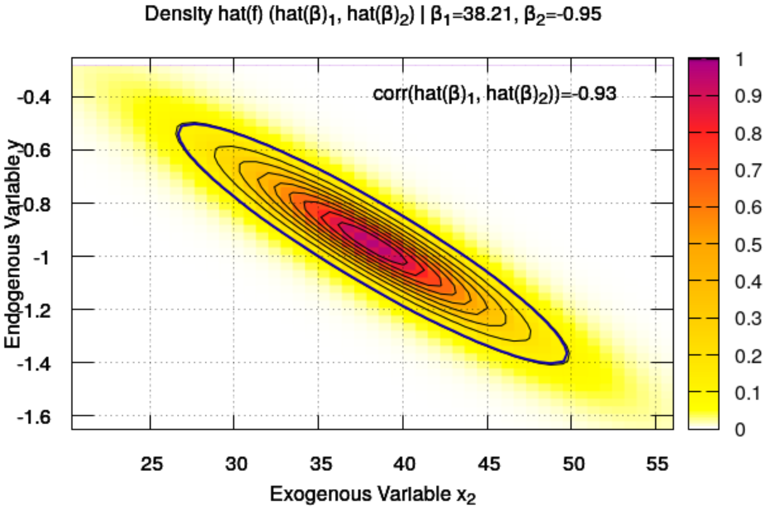
Abweichung vom Fit III: Großer Effekt bei positiven Abweichungen von β_1 und β_2



Abweichung vom Fit IV: Großer Effekt bei negativen Abweichungen von β_1 und β_2



Ergebnis: negative Korrelation zwischen den Schätzfehlern von β_1 und β_2



Spezialfall 1: Mittelwertschätzung (keine exogenen Faktoren)

- Model: $y = \beta_0 + \epsilon := \mu + \epsilon$
- System matrix: $\mathbf{X} = (1, 1, ..., 1)'$
- KQ estimator:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n}, \quad \mathbf{X}'\boldsymbol{y} = \sum_i y_i = n\bar{y},$$
$$\hat{\beta}_0 = \hat{\mu} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \bar{y}$$

- Variance: $V_{00} = V(\hat{\mu}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{n}, \quad \hat{V}_{00} = \frac{\hat{\sigma}^2}{n}$
- Distribution of the estimator (if $\epsilon \sim i.i.d N(\mu, \sigma^2)$)

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{V_{00}}} = \frac{\bar{y} - \mu}{\sigma} \sqrt{n} \sim N(0, 1),$$
$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{V}_{00}}} = \frac{\bar{y} - \mu}{\hat{\sigma}} \sqrt{n} \sim T(n-1)$$

Spezialfall 2: Lineare Einfachregression

- Model (with $x_1 = x$): $y = \beta_0 + \beta_1 x + \epsilon$
- System matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{pmatrix}$$

- KQ estimator (with $s_x^2 = 1/n(\sum x_i^2 - n\bar{x}))$:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{ns_x^2}\begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \quad \mathbf{X}'\boldsymbol{y} = \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix}$$

$$\begin{aligned} \hat{\beta}_1 &= \left(-\frac{\bar{x}}{ns_x^2}, \frac{1}{ns_x^2}\right)\begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}} = \frac{s_{xy}}{s_x^2}, \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned}$$

Simple linear regression (ctnd)

- Variance-covariance matrix (assuming w/o loss of generality $\bar{x} = 0$):

$$\mathbf{V}\,(\hat{\boldsymbol{\beta}}) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{ns_x^2} \end{pmatrix}$$

- Variance of the estimator $\hat{y}(x)$ (x is deterministic):

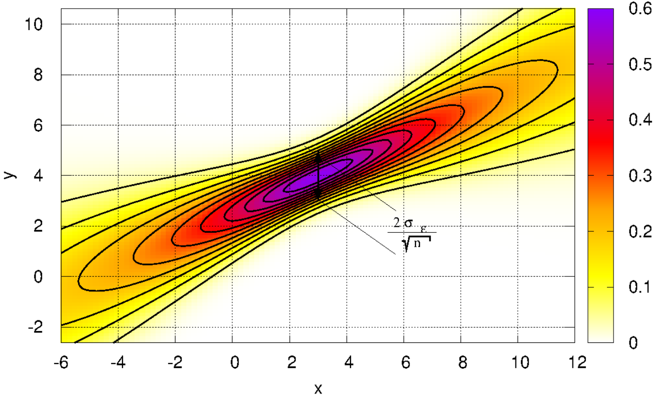
$$V\,(\hat{y}(x)) = V\,(\hat{\beta}_0 + \hat{\beta}_1 x) = V_{00} + x^2 V_{11} + 2x V_{01} = \frac{\sigma^2}{n}\left(1 + \frac{x^2}{s_x^2}\right)$$

- Distribution of the estimator for $y(x)$:

$$\hat{y}(x) \sim N\big(y(x), V\,(\hat{y}(x))\big)$$

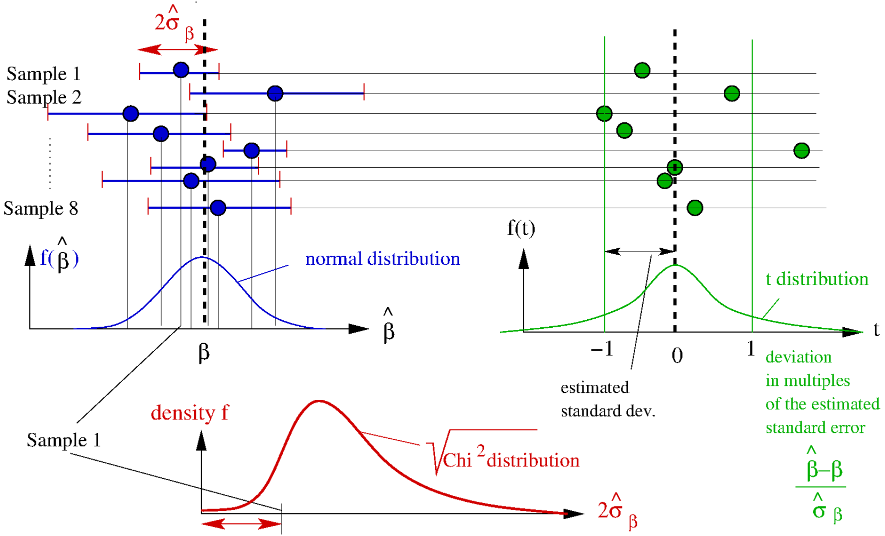
If σ^2 has to be estimated by $\hat{\sigma}^2$, the normalized estimators for β_0, β_1 and $y(x)$ are $\sim T(n-2)$.

Probability density for $\hat{y}(x)$ for simple linear regression

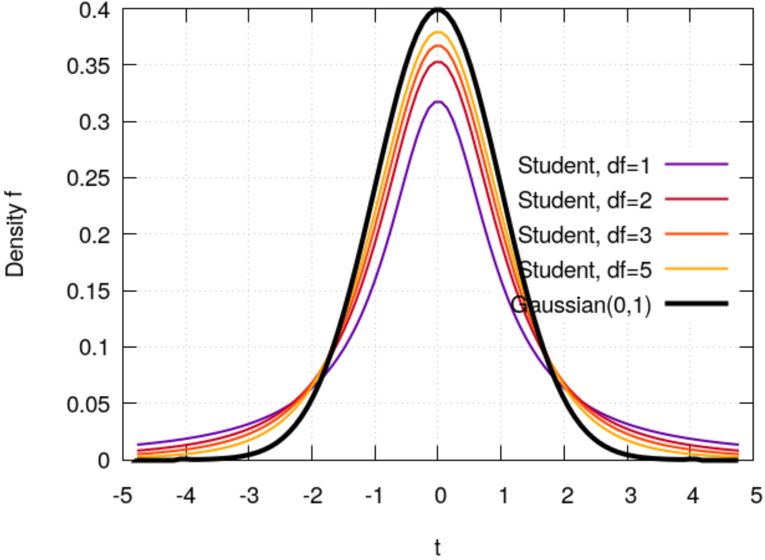


- If the Gauß-Markov assumptions apply, the model estimation errors $\hat{y}(x) - y(x)$ are Gaussian distributed
- The expectation and variance depends on x; the standard error is hyperbola-shaped.

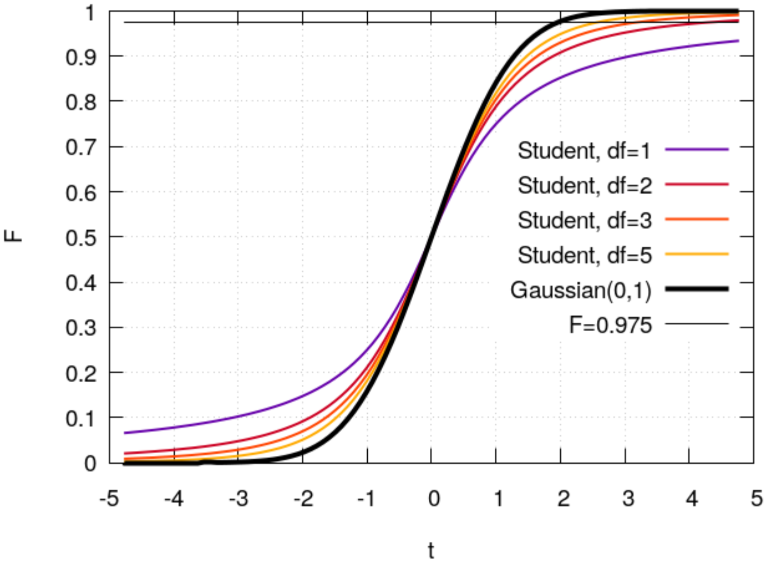
12.4 Konfidenzintervalle der Parameterwerte: Herkunft der 1 Student-t Verteilung



Dichtefunktionen der Student-t Verteilungen vs. Standardnormalverteilung



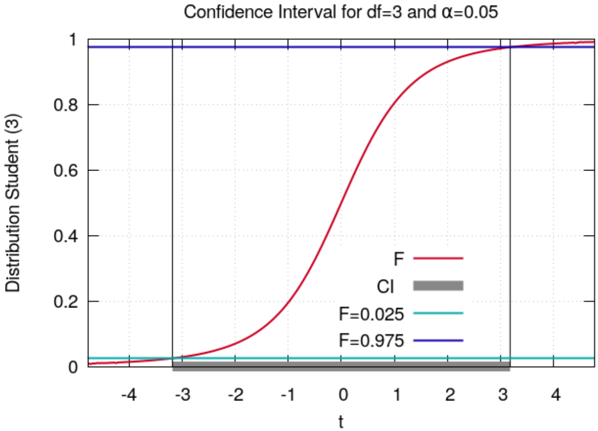
Verteilungsfunktionen der Student-t Verteilungen vs. Standardnormalverteilung



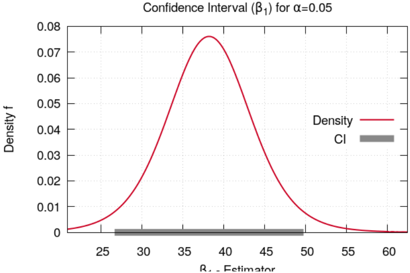
Konfidenzintervalle (KI)

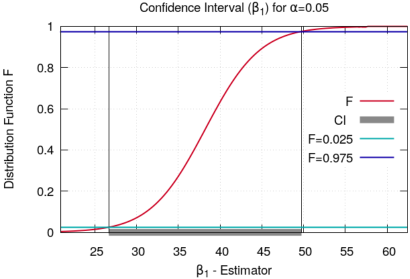
$$\mathrm{CI}_{\beta_j}^{(\alpha)} : \beta_j \in \left[\hat{\beta}_j - \Delta\hat{\beta}_j, \hat{\beta}_j + \Delta\hat{\beta}_j \right], \quad \Delta\hat{\beta}_j = t_{1-\alpha/2}^{(n-J-1)} \hat{\sigma}_{\hat{\beta}_j}.$$

- $t_{1-\alpha/2}$: Quantil (Inverse der Verteilungsfunktion)
- "Unschärferelation der KI: Eine schärfere Aussage (ein kleineres KI) impliziert größere (α-) Fehler und umgekehrt"



Hotelbeispiel: KI für die Wertschätzung der Qualität bzw Sternezahl β_1





Modell: $y(\boldsymbol{x}) = \sum_j \beta_j x_j + \epsilon$

Faktoren:
$x_0 = 1$, x_1: #stars, x_2: price

Konfidenzintervall CI:

$$\beta_1 \in \left[\hat{\beta}_1 - \Delta\hat{\beta}_1^{(\alpha)}, \hat{\beta}_1 + \Delta\hat{\beta}_1^{(\alpha)}\right]$$

$$\Delta\hat{\beta}_1^{(\alpha)} = t_{1-\alpha/2}^{(n-3)} \sqrt{\hat{V}(\hat{\beta}_1)}$$

$$\hat{V}(\hat{\beta}_1) = \hat{\sigma}_\epsilon^2 \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]_{11}$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-3} \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2$$