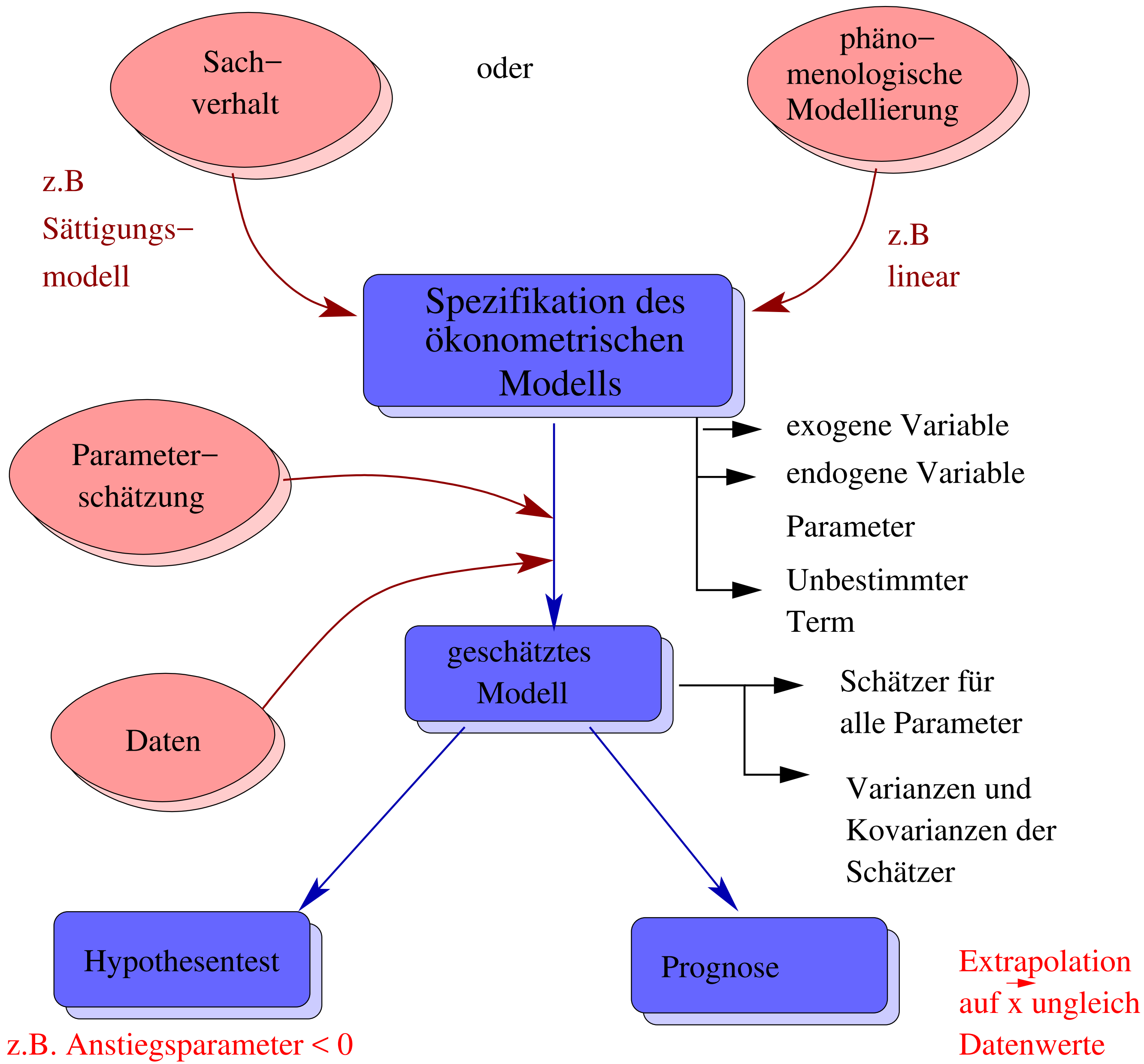
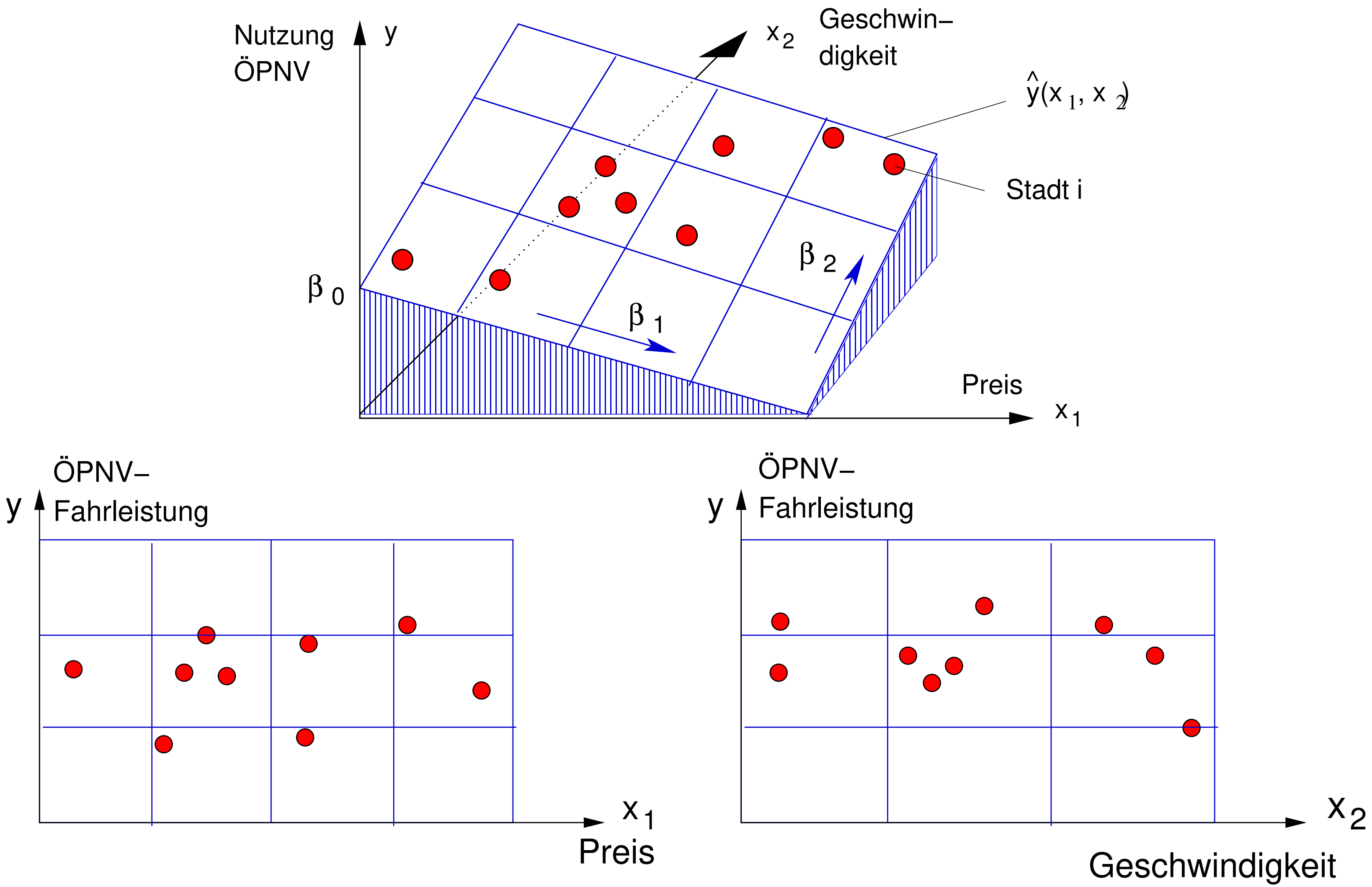
Flussdiagramm der ökonometrischen Methode

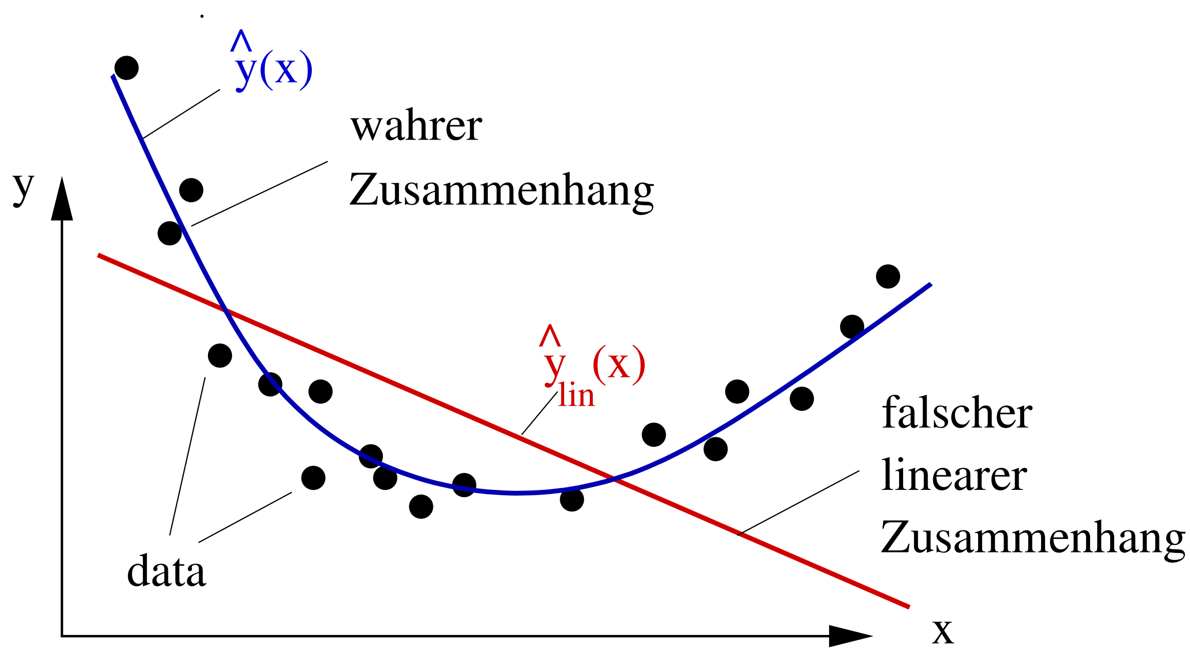


Modellspezifikation I: Funktionale Spezifikation 1



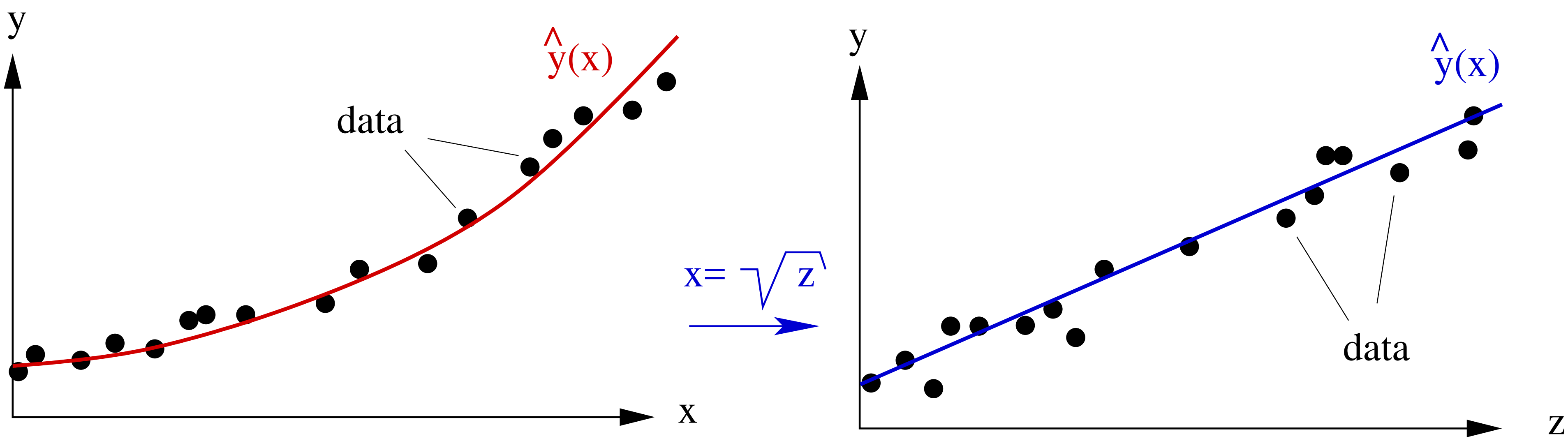
1. Alle relevanten Einflussfaktoren sind berücksichtigt (oben, nicht aber unten)

Modellspezifikation I: Funktionale Spezifikation 2



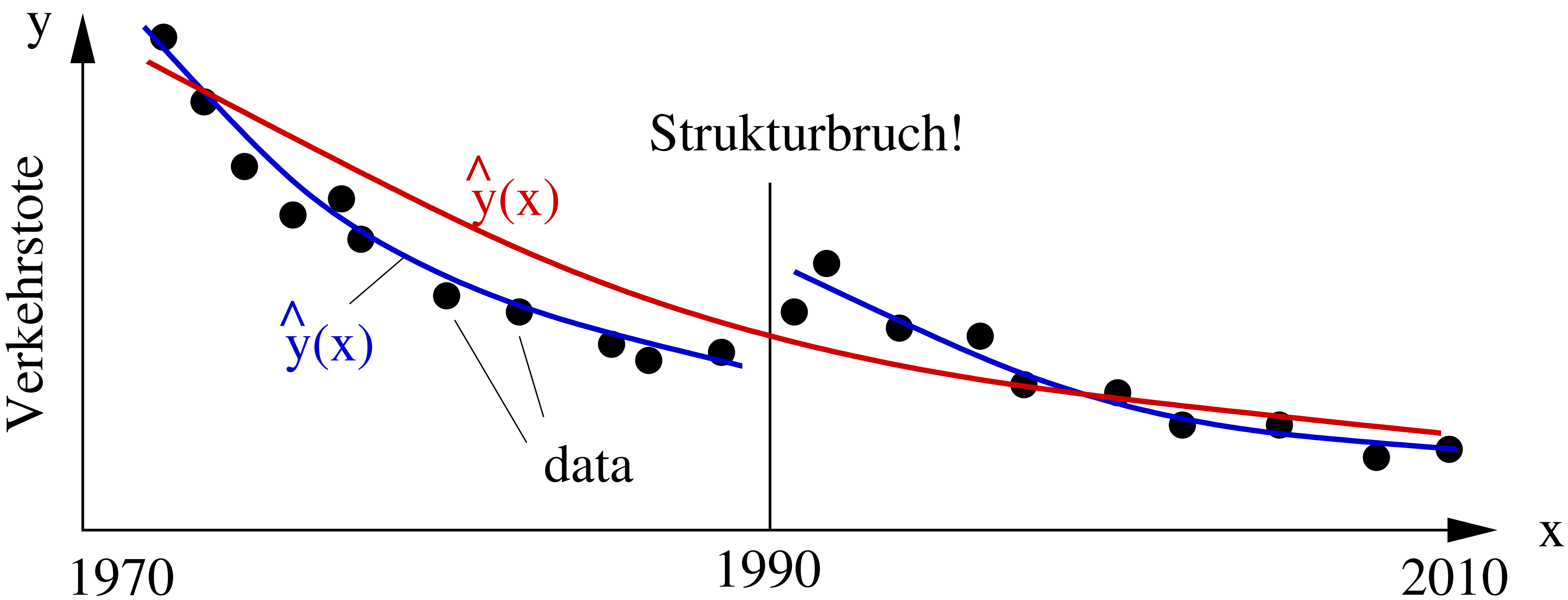
2. Das Modell ist linear, was hier nicht erfüllt ist

Modellspezifikation I: Funktionale Spezifikation 2



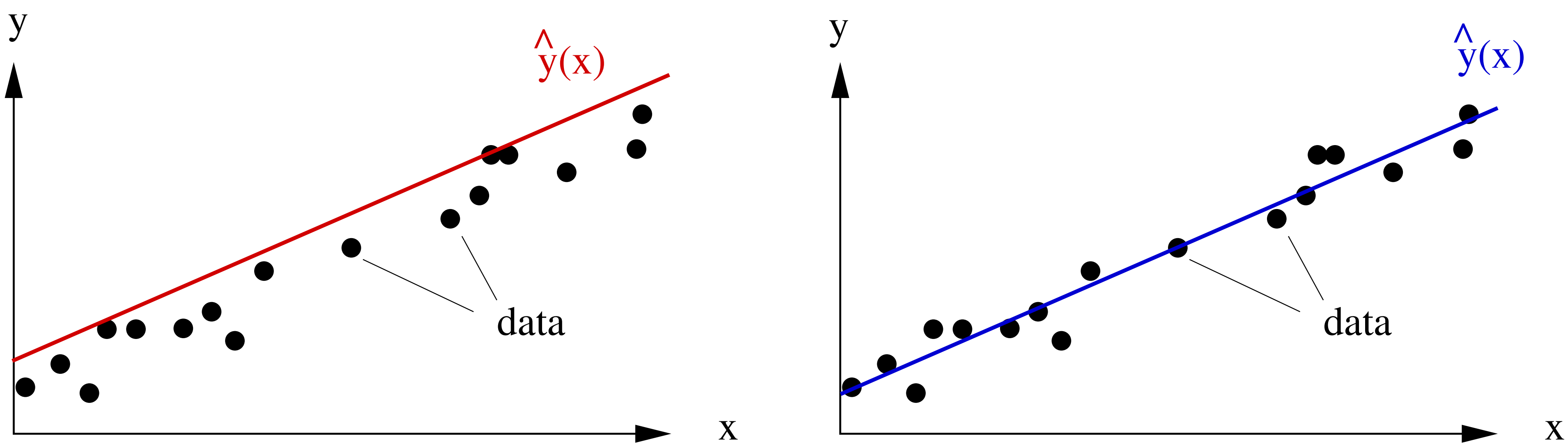
Manchmal kann das Modell durch Transformationen der exogenen und/oder endogenen Variablen linearisiert werden

Modellspezifikation I: Funktionale Spezifikation 3



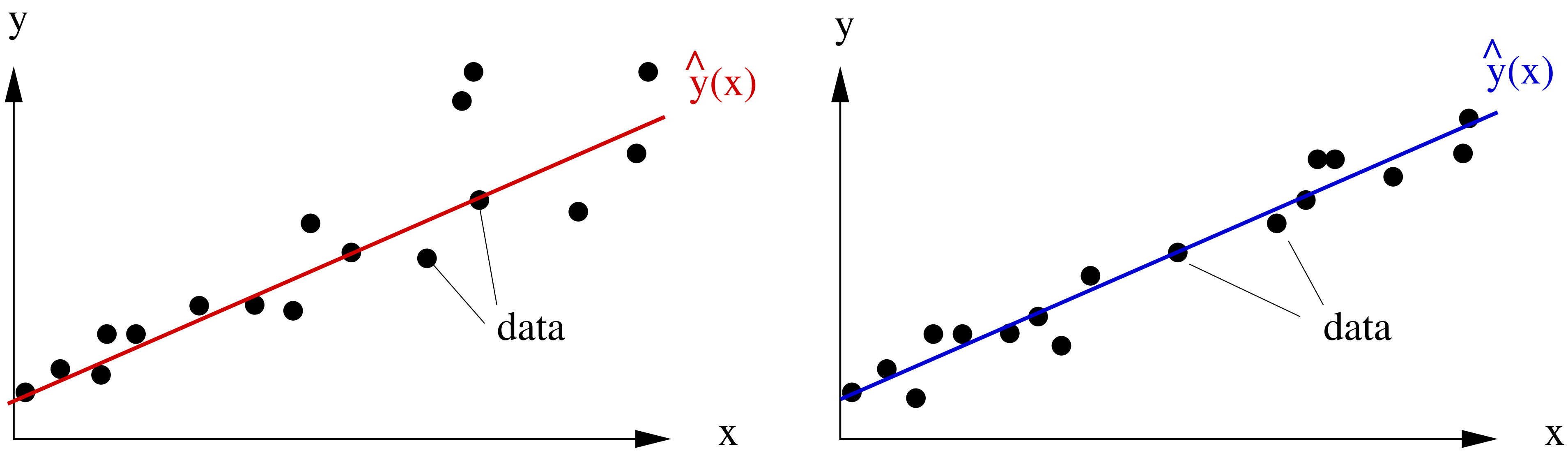
3. Homogenitätskriterium (z.B. kein Strukturbruch im Raum der exogenen Variablen, wie hier gezeigt)

Modellspezifikation II: Statistische Spezifikation 1



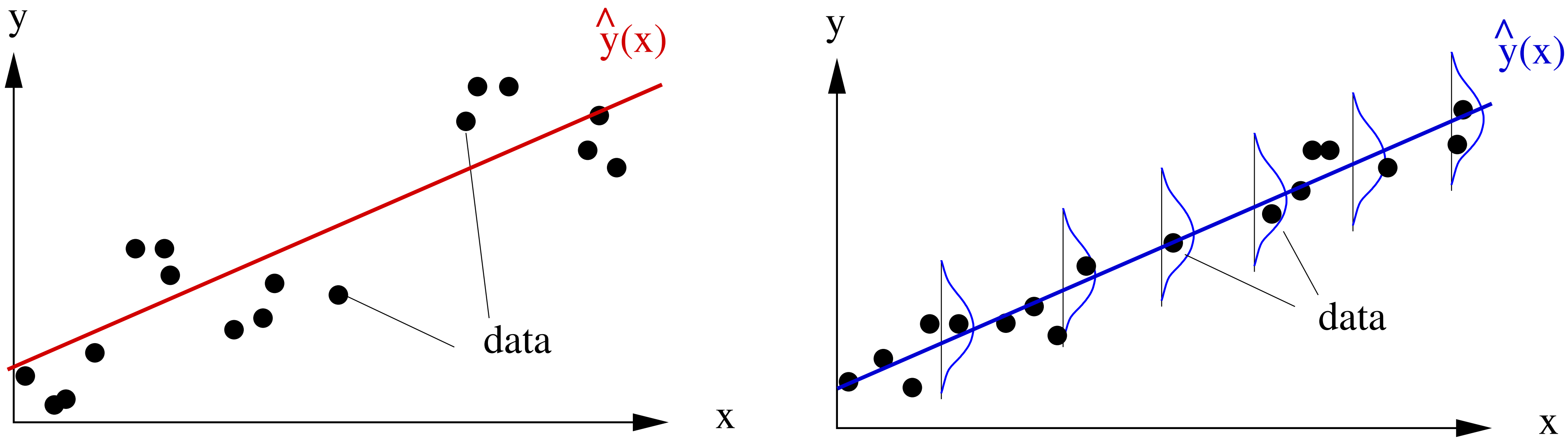
1. Der Erwartungswert der Störgröße muss verschwinden.

Modellspezifikation II: Statistische Spezifikation 2



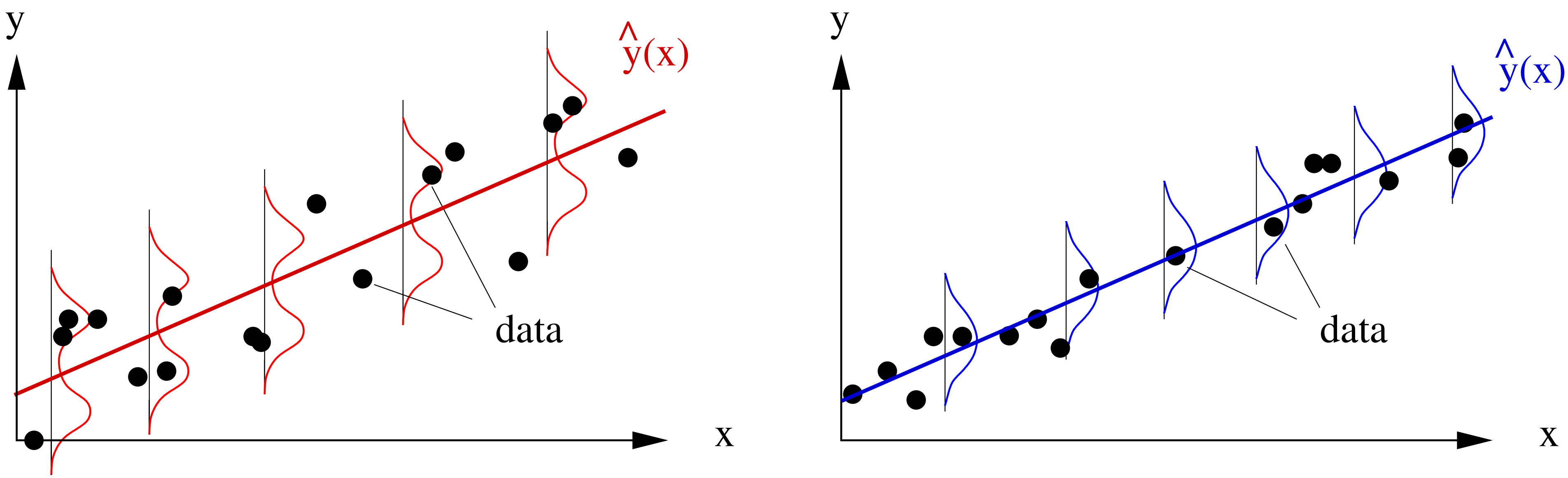
2. Der Residualterm ϵ ist homoskedastisch (rechts), nicht etwa heteroskedastisch (links)

Modellspezifikation II: Statistische Spezifikation 3



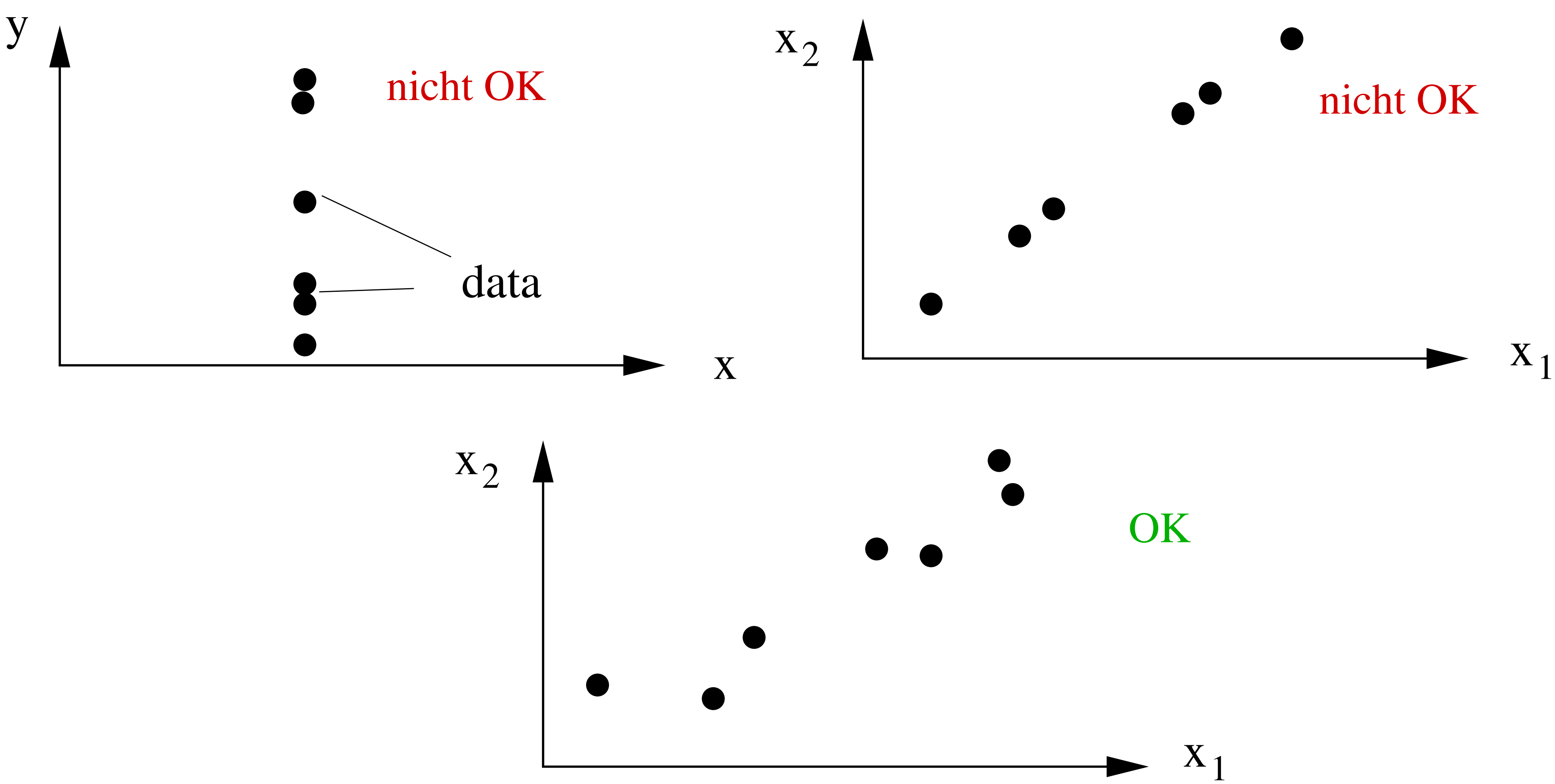
Keine Korrelationen von ϵ bezüglich x_i oder y (rechts),
während das Modell links fehlspezifiziert ist

Modellspezifikation II: Statistische Spezifikation 4



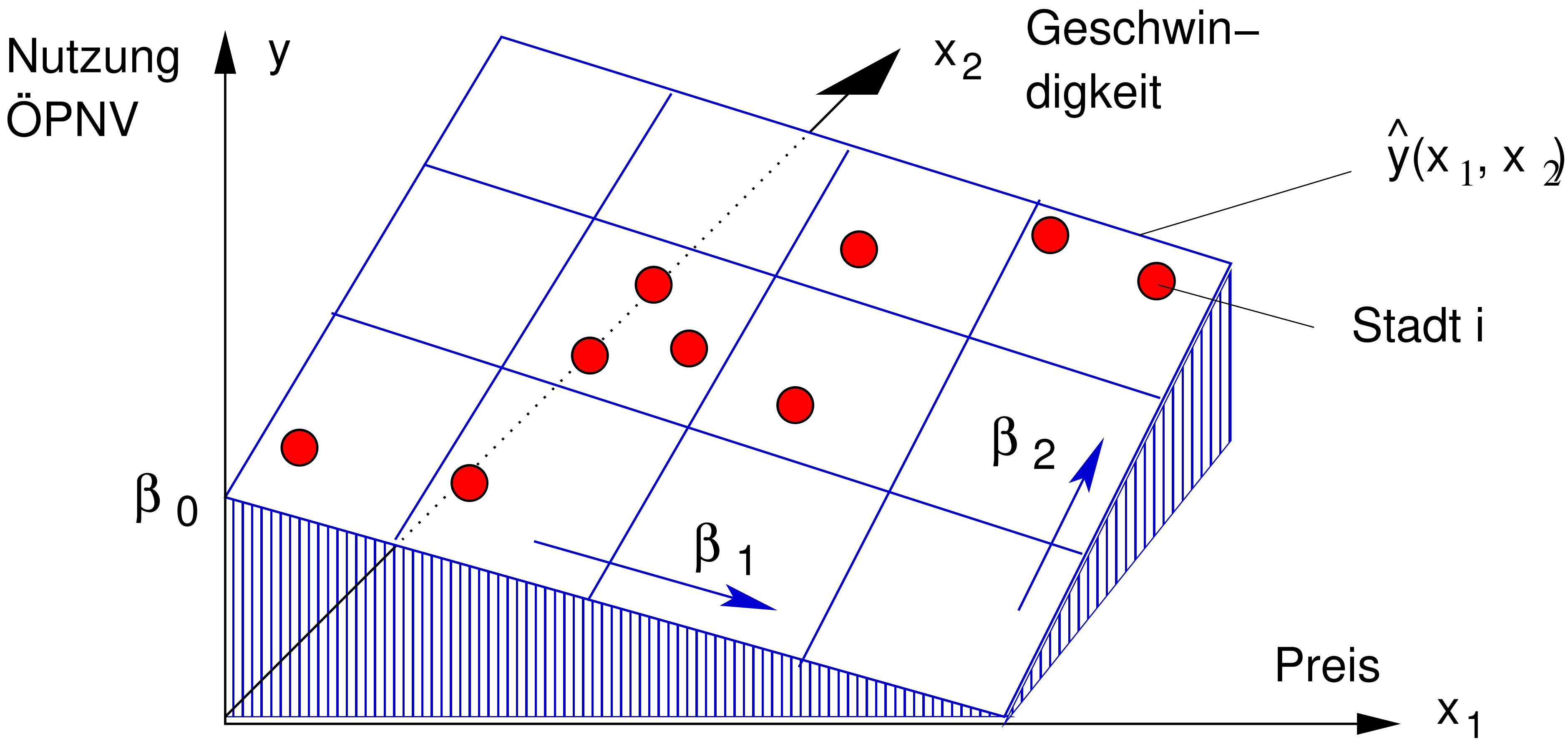
Der Residualterm ϵ ist gaußverteilt (rechts),
nicht etwa bimodal verteilt (links)

Modellspezifikation III: Datenspezifikation



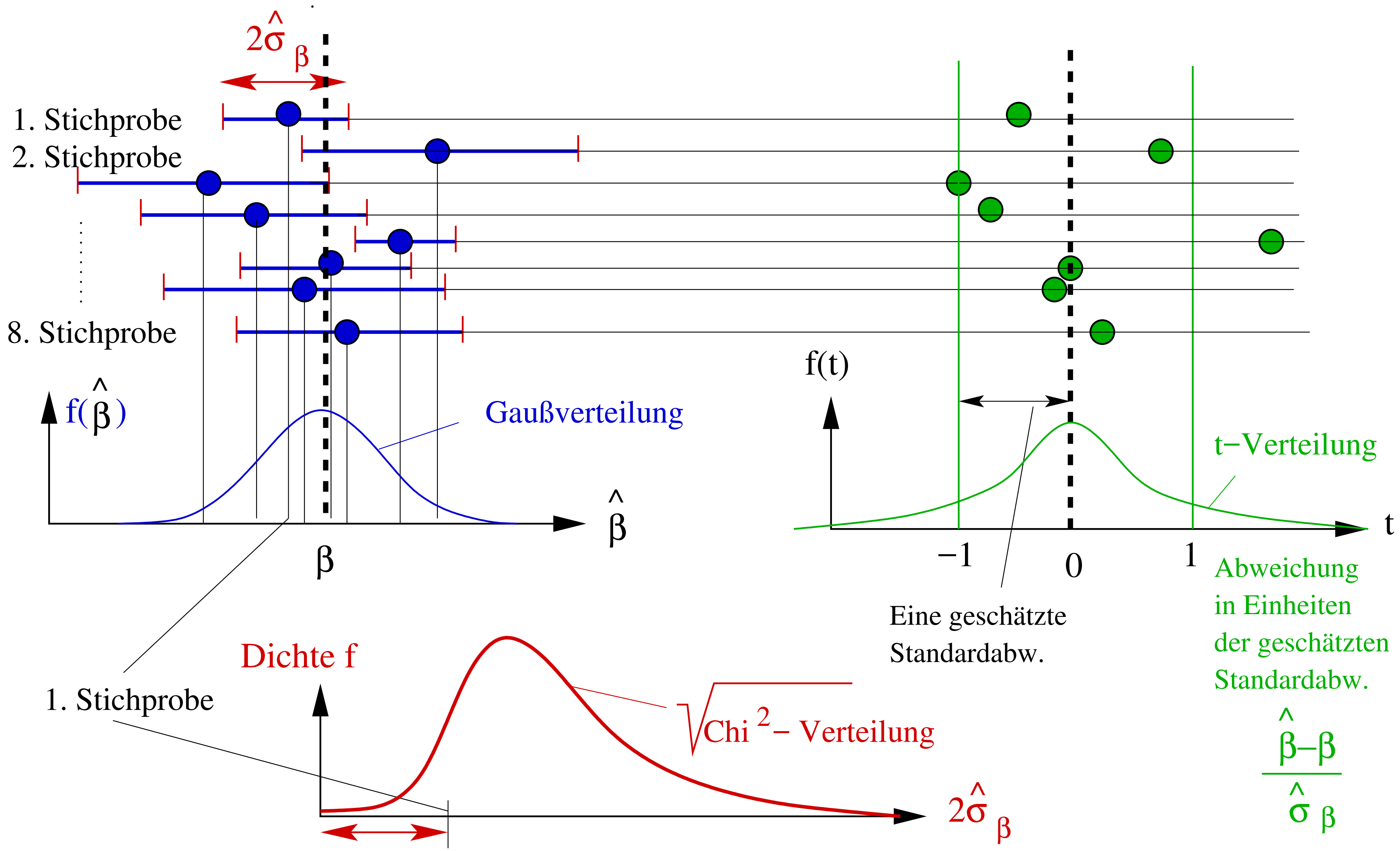
Keine der exogenen Variablen darf sich als Linearkombination aus Konstanten und anderen exogenen Variablen darstellen lassen (oben); nichtperfekte Korrelationen sind aber erlaubt (unten)

Lineares Modell mit zwei exogenen Variablen (schematisch)

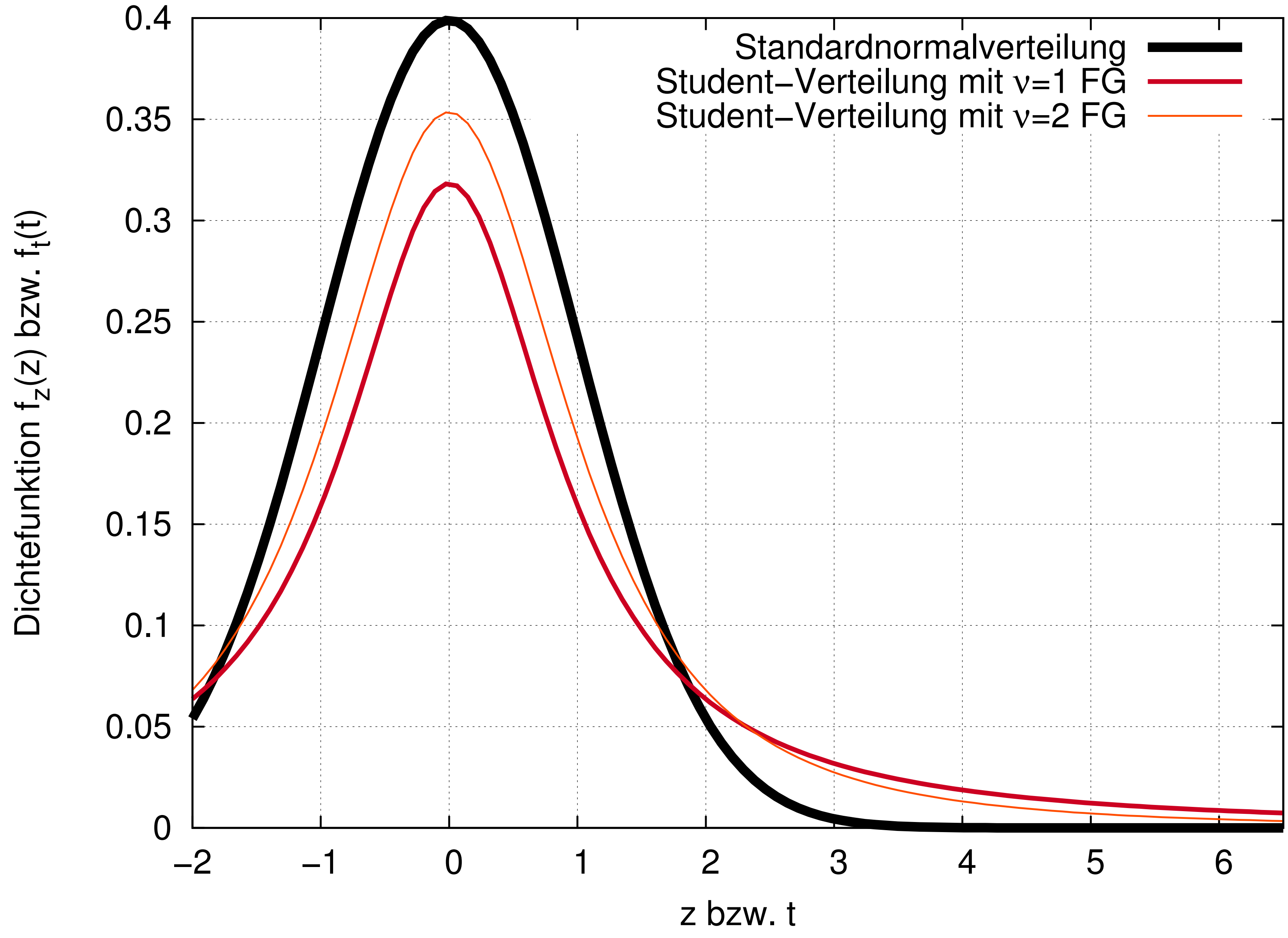


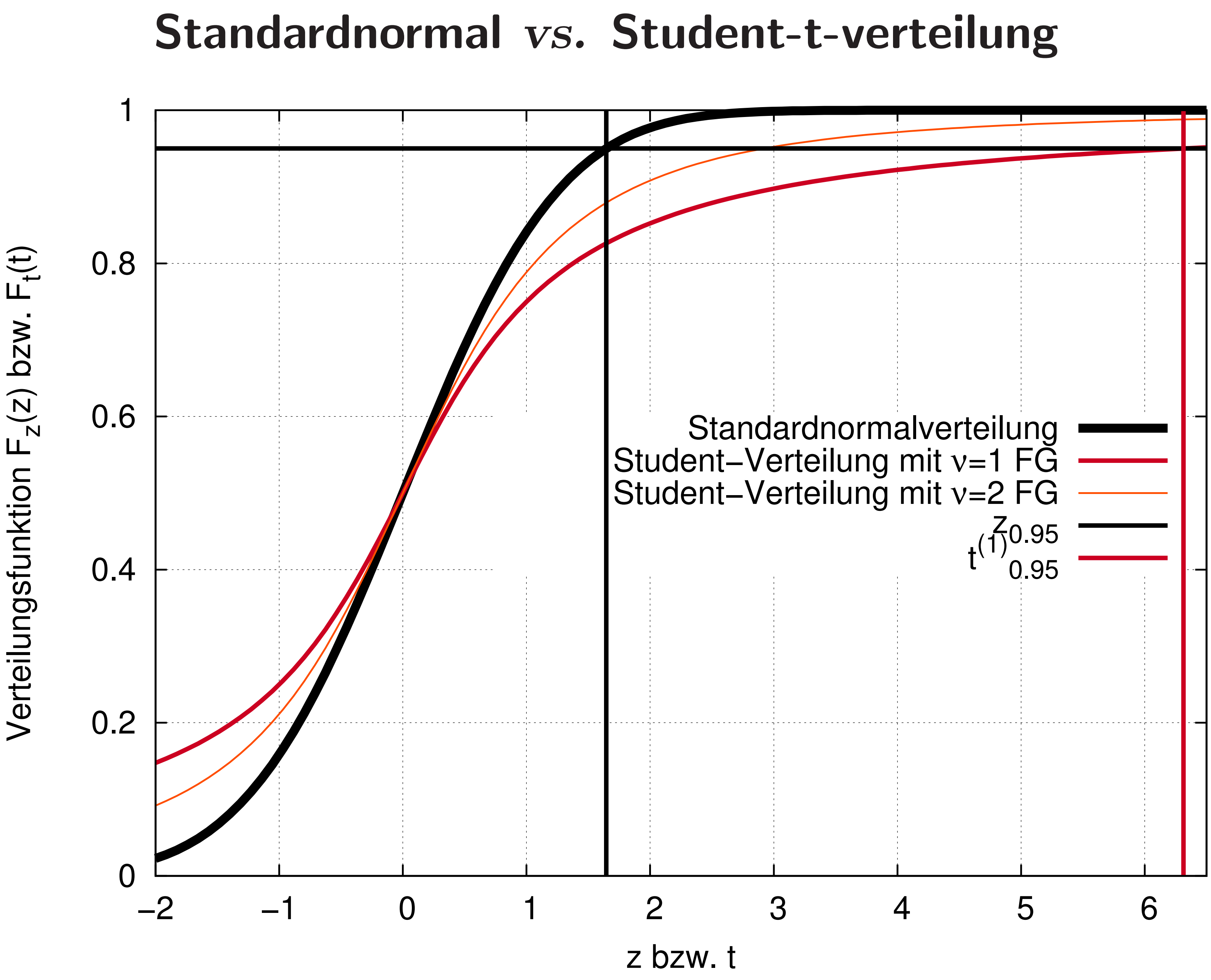
Die Daten gehorchen hier dem linearen Modell exakt!

Konfidenzintervalle und die Entstehung der Student-Verteilung



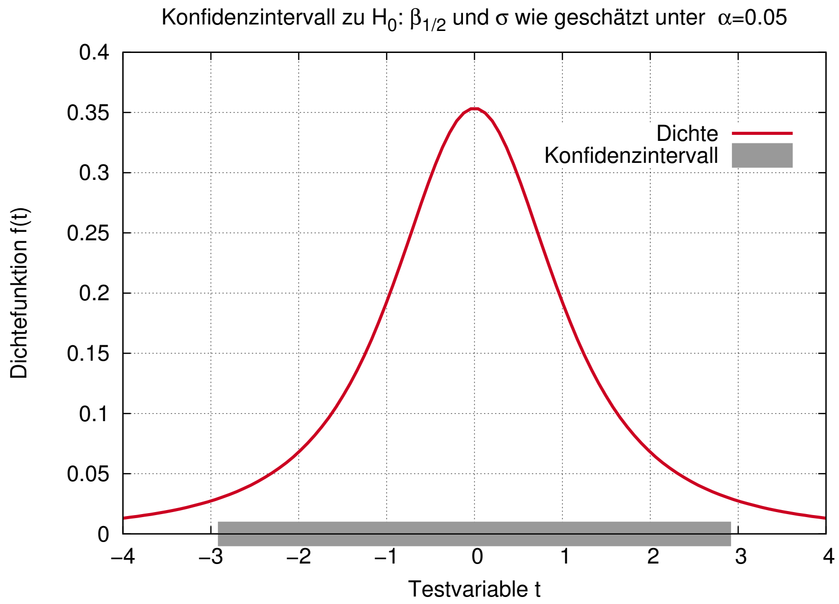
Dichten der Standardnormal *vs.* Student-t-verteilung

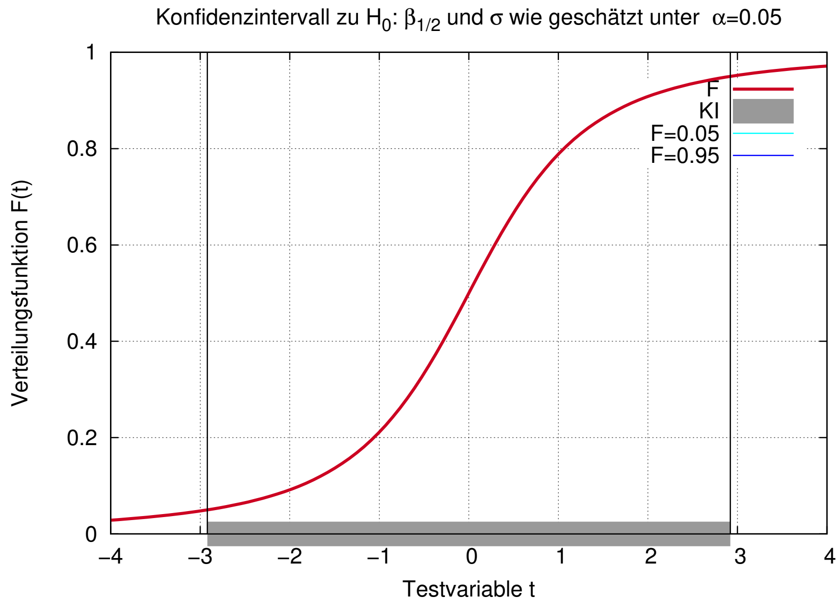




Standardnormal vs. Student-t-verteilung
Verteilungsfunktion $F_z(z)$ bzw. $F_t(t)$
z bzw. t
Standardnormalverteilung
Student−Verteilung mit ν=1 FG
Student−Verteilung mit ν=2 FG
$z_{0.95}$
$t^{(1)}_{0.95}$
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
2
3
4
5
6

Konfidenzintervalle





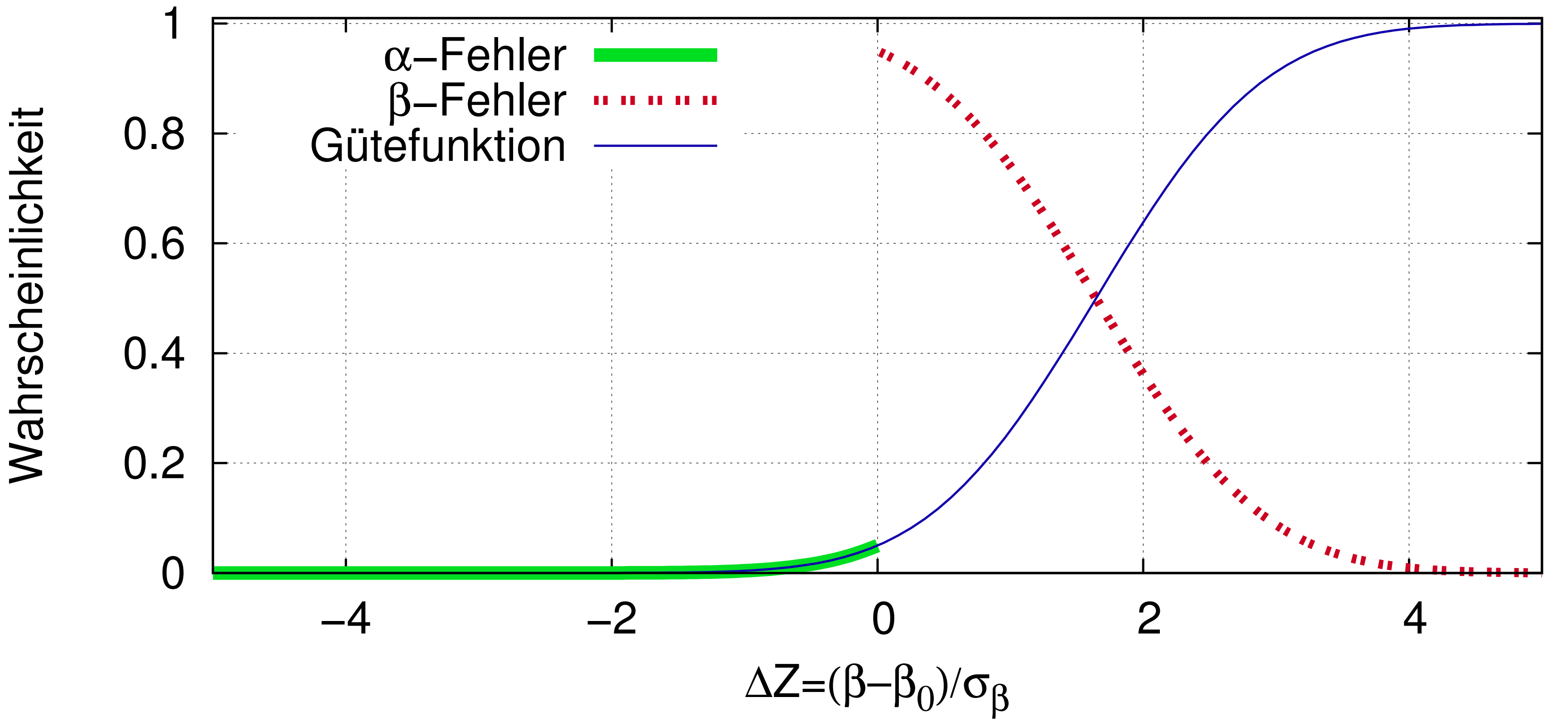
KI zu einer Fehlerwahrscheinlichkeit $\alpha = 5\,\%$ für $n - J - 1 = 2$ Freiheitsgrade.

Fehler erster und 2. Art allgemein

	H_0 nicht abgelehnt	H_0 abgelehnt
H_0 trifft zu	✓	Fehler erster Art
H_0 trifft nicht zu	Fehler zweiter Art	✓

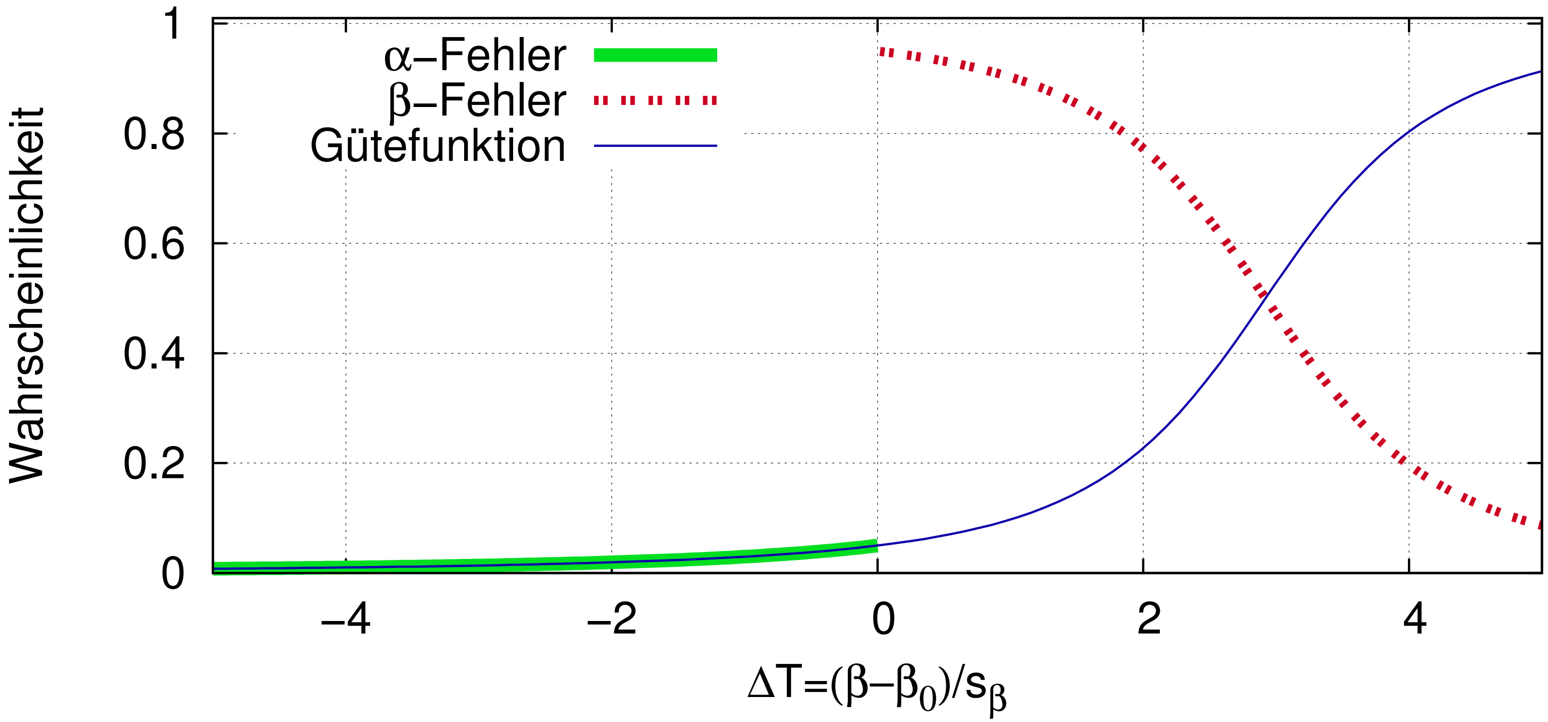
Definition der Fehler erster und zweiter Art bei Signifikanztests

Fehler erster und 2. Art bei H_0: $\beta \leq \beta_0$



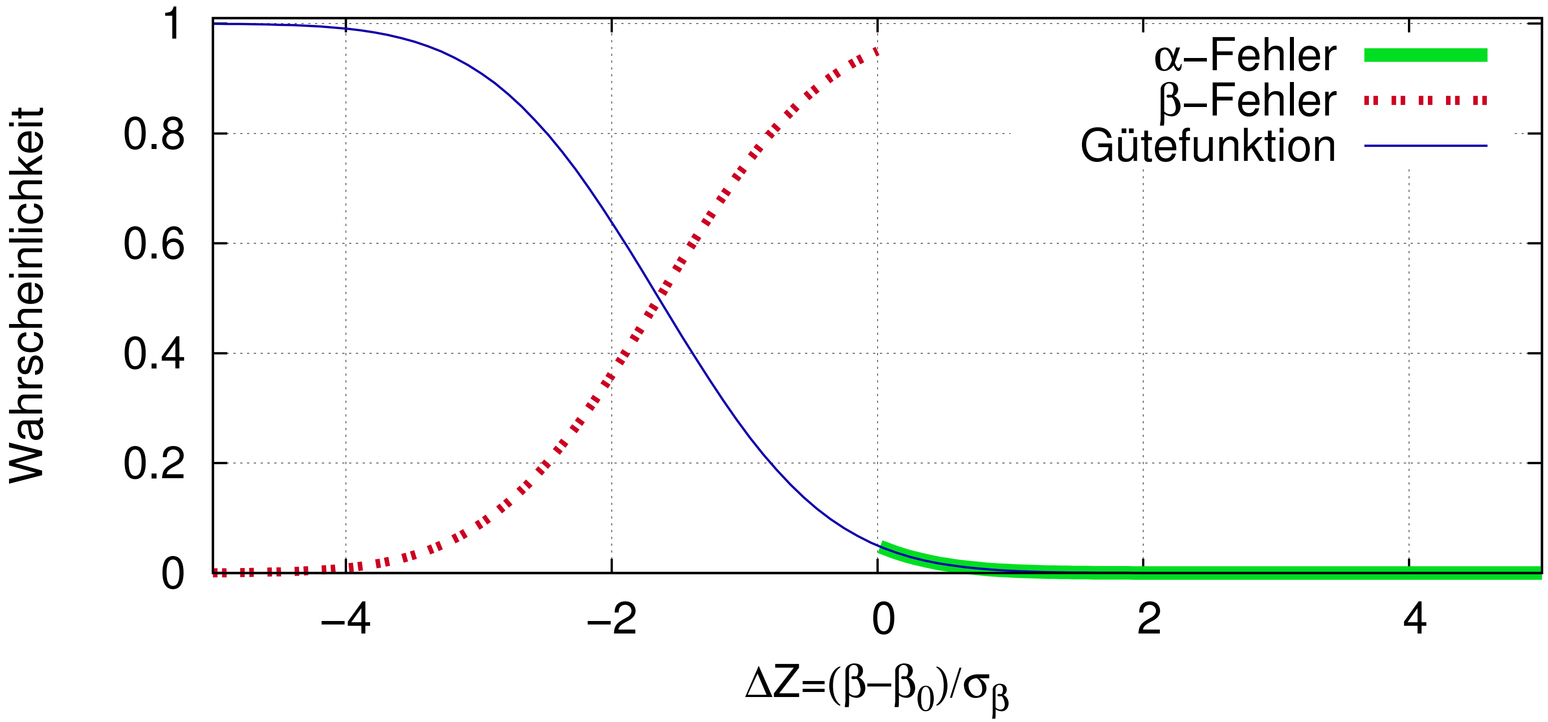
Einseitiger Test auf $<$, $\leq$ in Abhängigkeit des skalierten Abstandes $\Delta z = (\beta_j - \beta_{0j})/\sigma_{\hat{\beta}_j}$ des wahren Parameterwertes vom Grenzwert der Nullhypothese (bekannte Varianz des Schätzers, $\alpha = 0.1$)

Fehler erster und 2. Art bei H_0: $\beta \leq \beta_0$



Das Gleiche bei unbekannte Varianz und $n - J - 1 = 2$ Freiheitsgraden. Der skalierte Abstand ist nun $\Delta t = (\beta_j - \beta_{0j})/\hat{\sigma}_{\hat{\beta}_j}$.

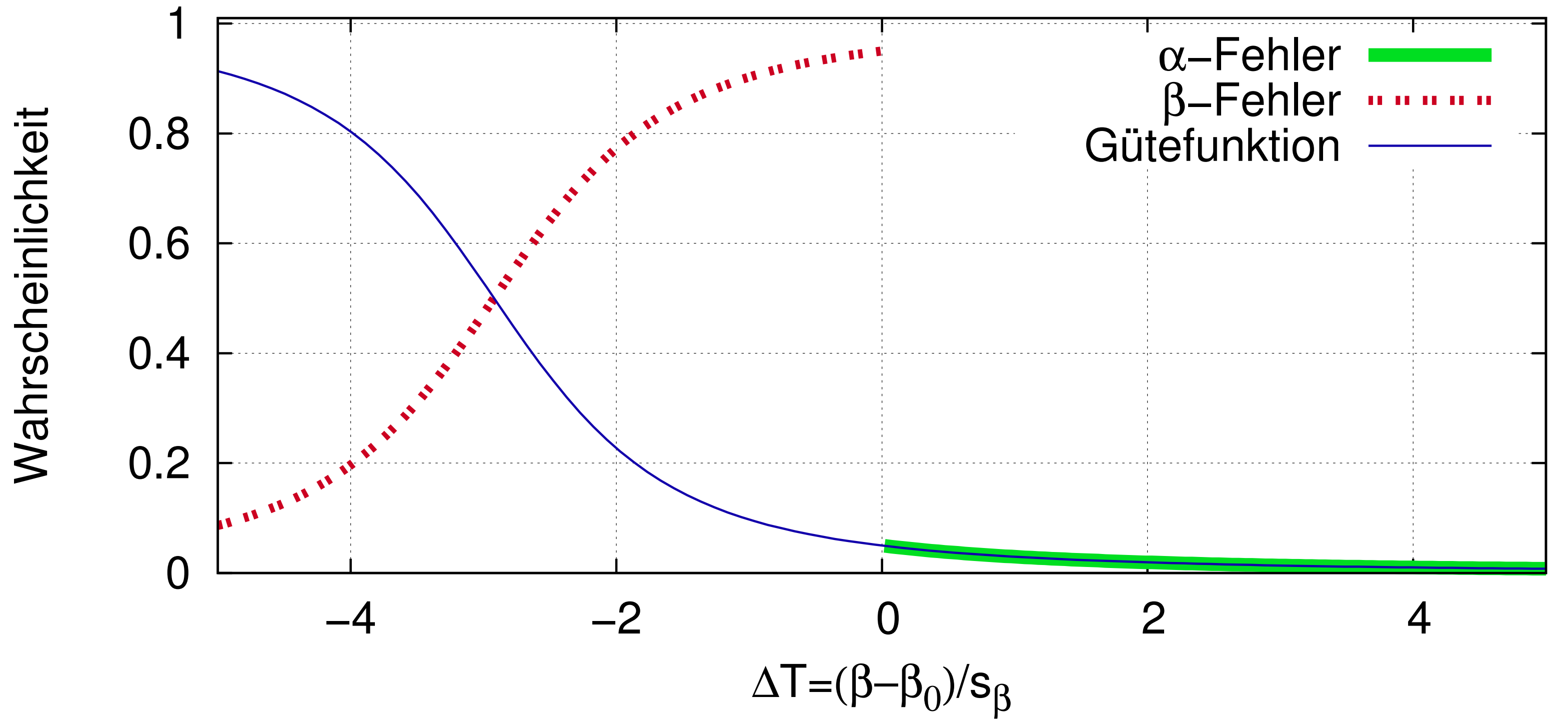
Fehler erster und 2. Art bei H_0: $\beta \geq \beta_0$



Einseitiger Test auf $>$, $\geq$

(bekannte Varianz, $\alpha = 0.1$)

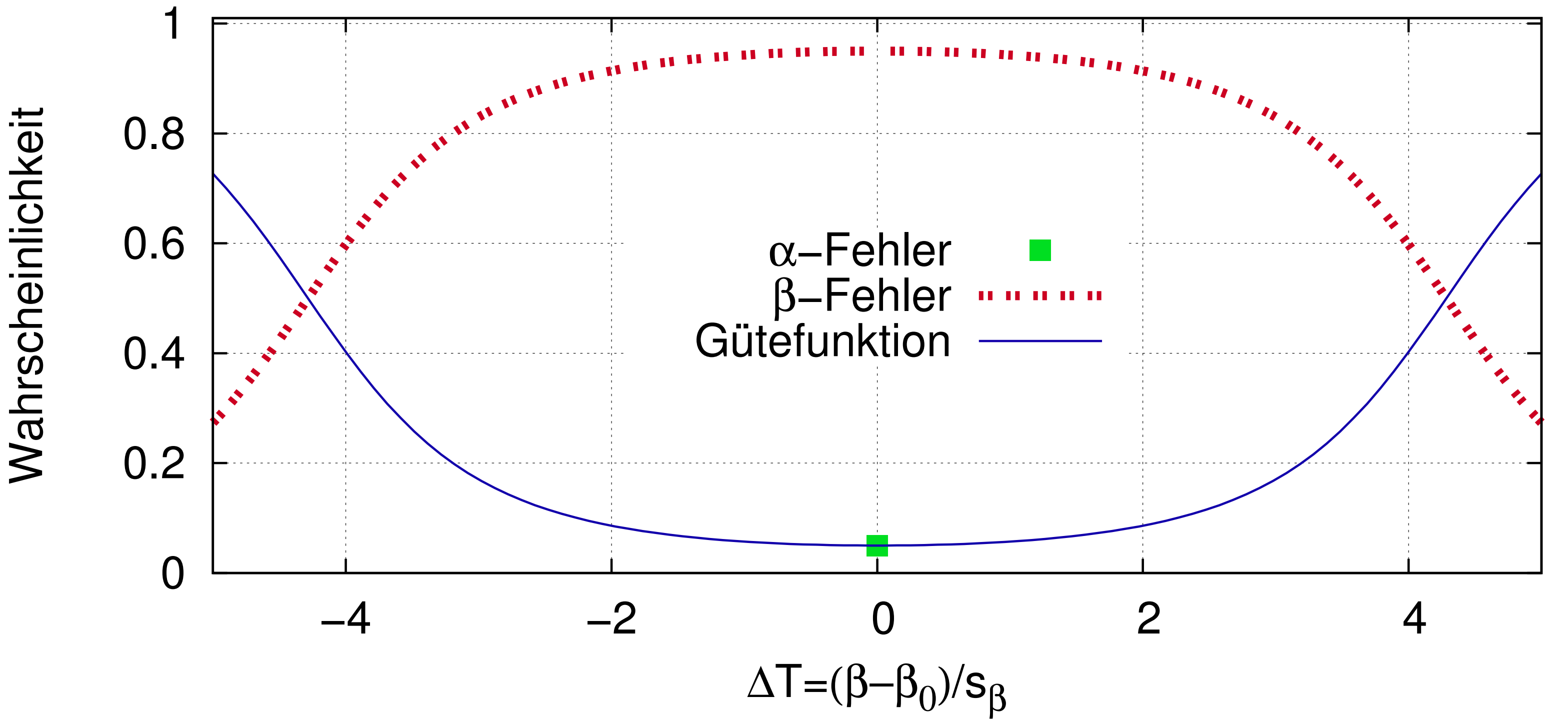
Fehler erster und 2. Art bei H_0: $\beta \geq \beta_0$



Einseitiger Test auf $>$, $\geq$

(unbekannte Varianz, $n - J - 1 = 2$ Freiheitsgrade, $\alpha = 0.1$)

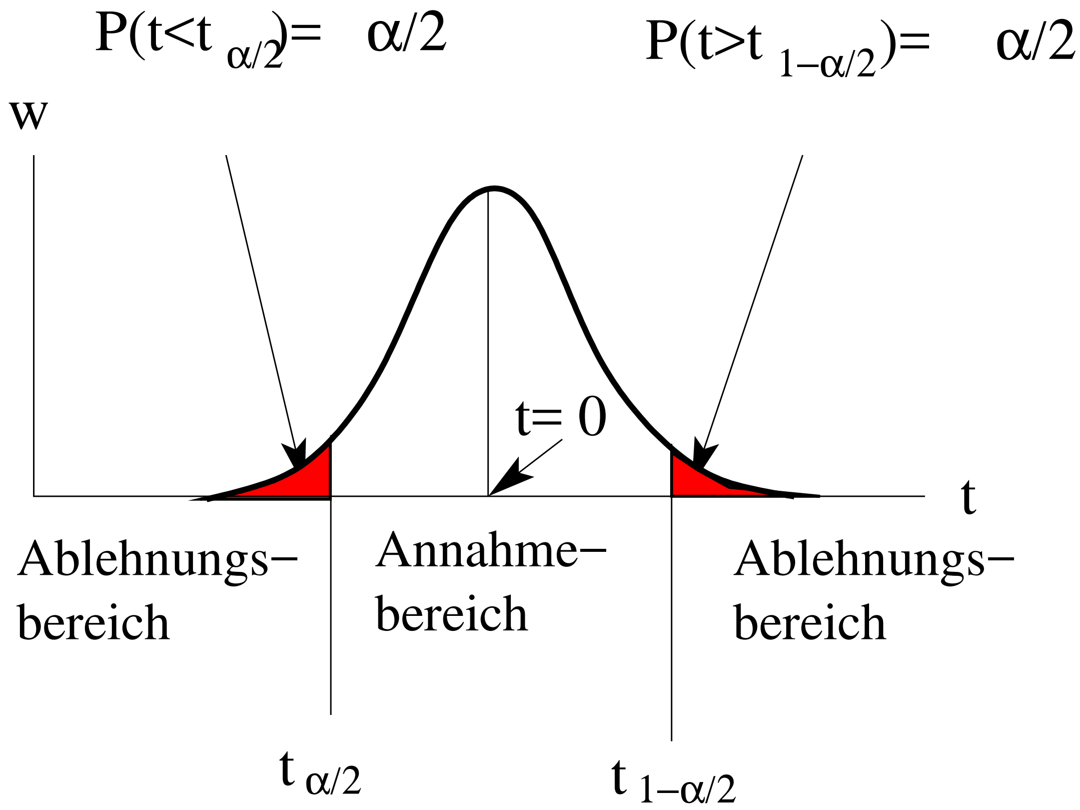
Fehler erster und 2. Art bei H_0: $\beta = \beta_0$



Zweiseitiger Test auf Gleichheit

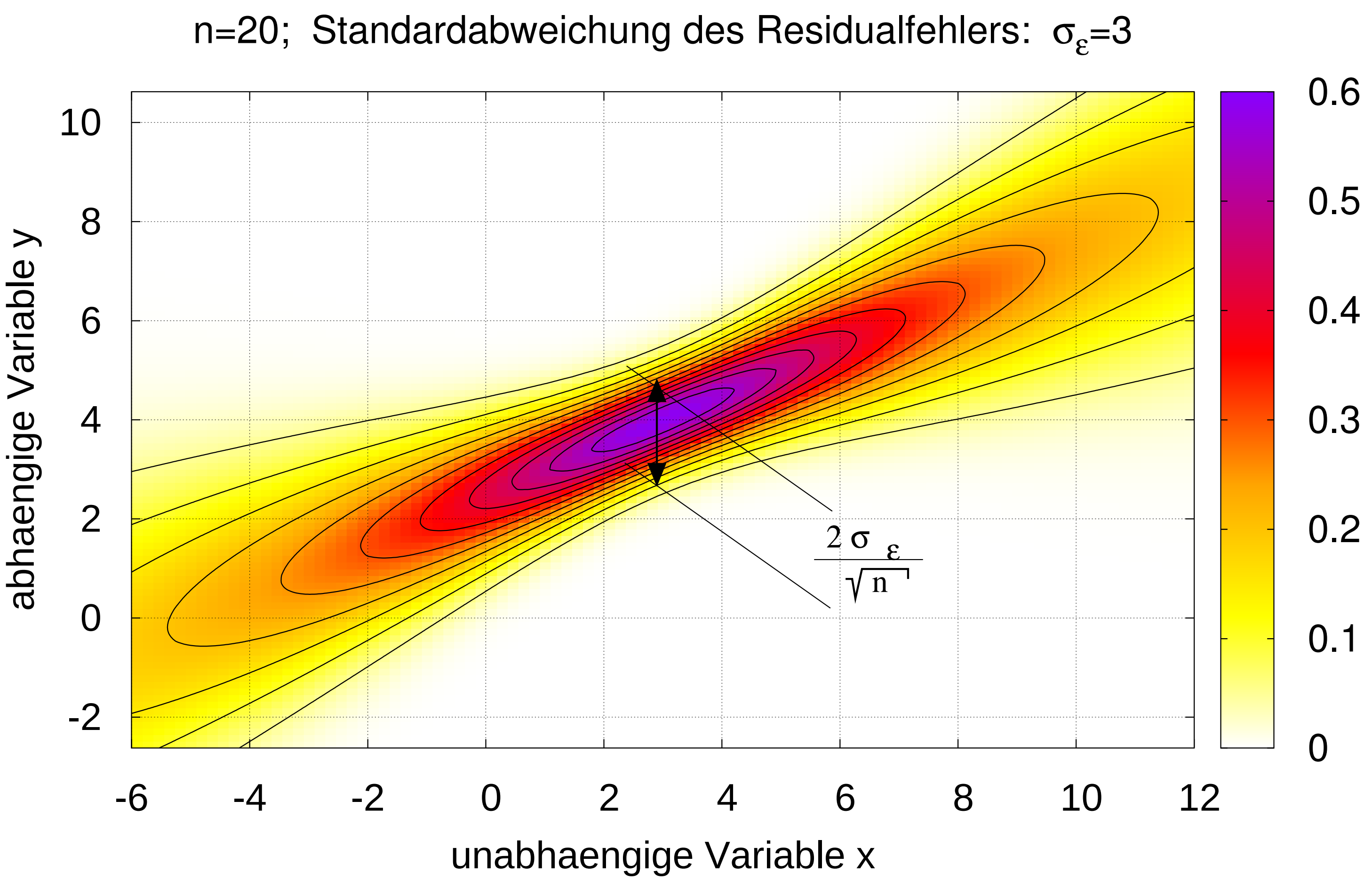
(unbekannte Varianz, $n - J - 1 = 2$ Freiheitsgrade, $\alpha = 0.1$)

Fehler erster und 2. Art allgemein



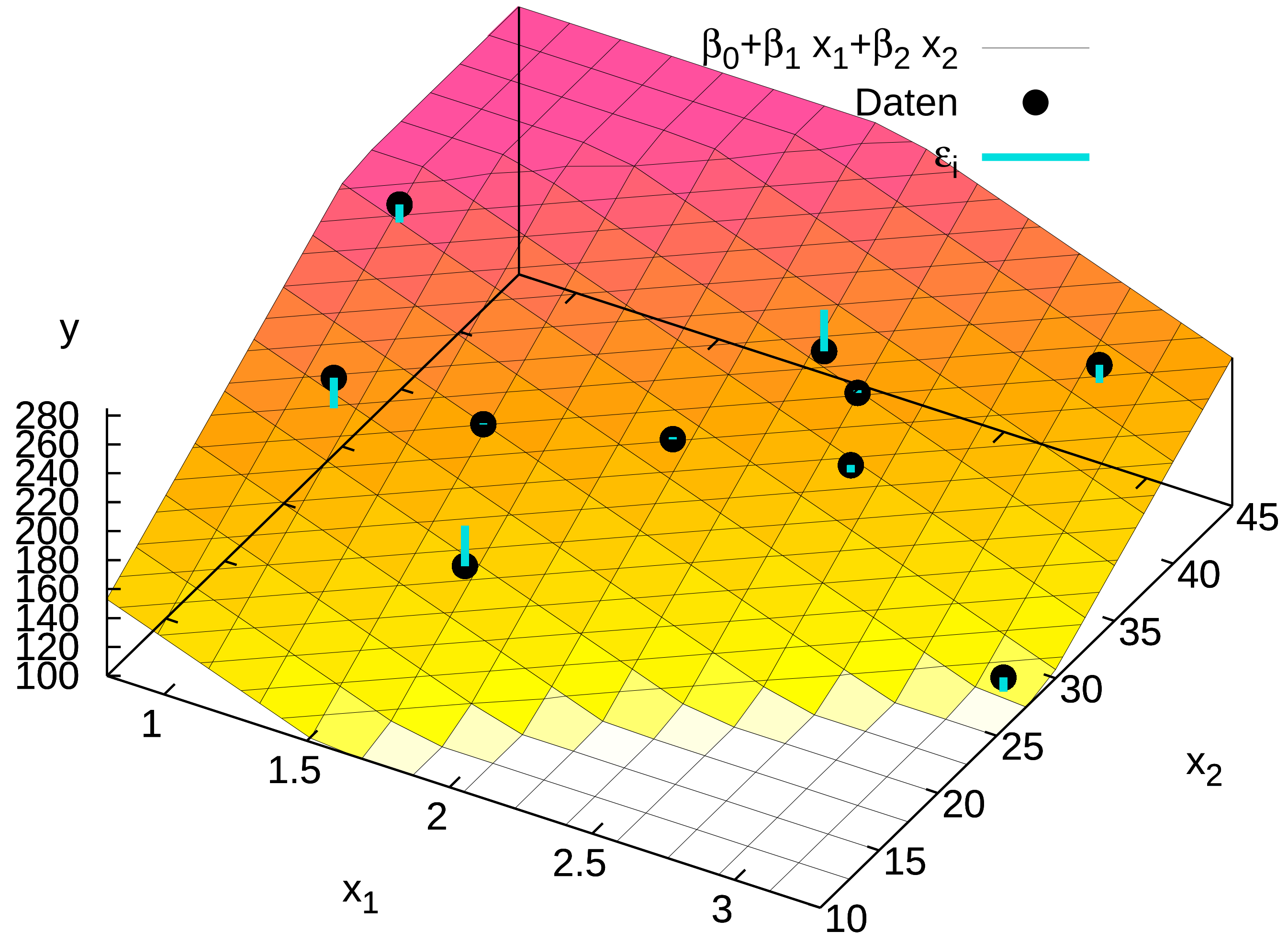
Annahme- und Ablehnungsbereiche bei zweiseitigen Tests (Tests einer Punkt-Hypothese). Die Verteilungsfunktion ist nur bei Zutreffen von H_0 gültig!

Parameter-Schätzfehler (bedingte W-Dichte) bei linearer Einfachregression



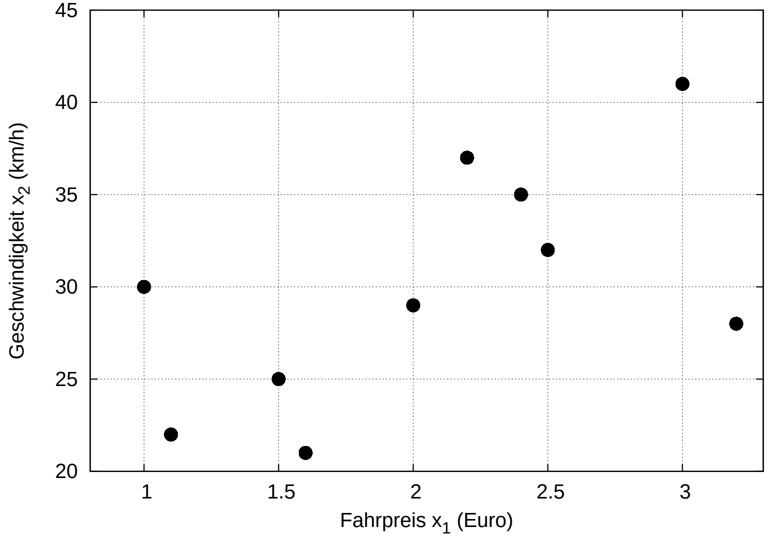
Die Residualfehler sind i.i.d. verteilt

Konkretes Beispiel: ÖPNV-Nutzung bei 10 Städten

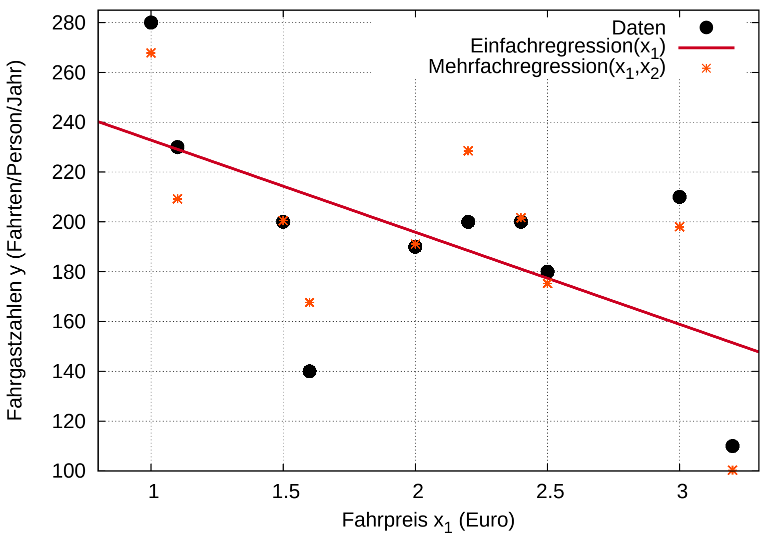


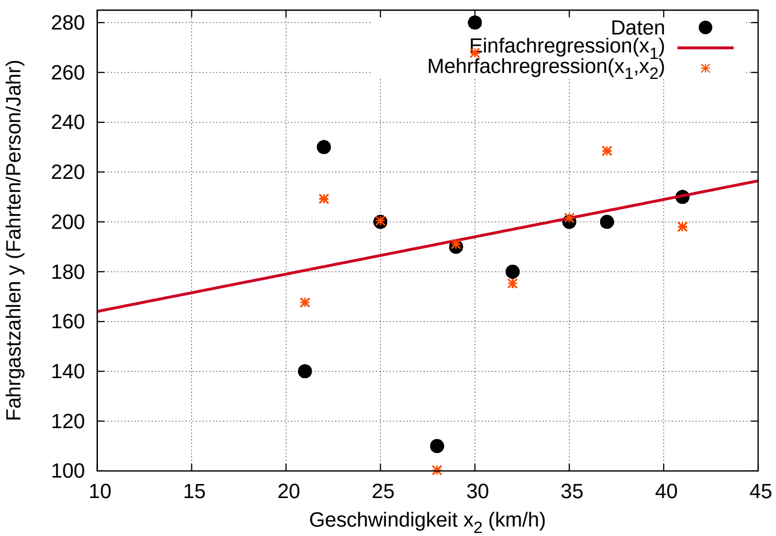
Türkisgrünen Striche: unbestimmten Anteile ϵ_i
(positiv, wenn Über dem Datenpunkt)

Projizierte Streudiagramme



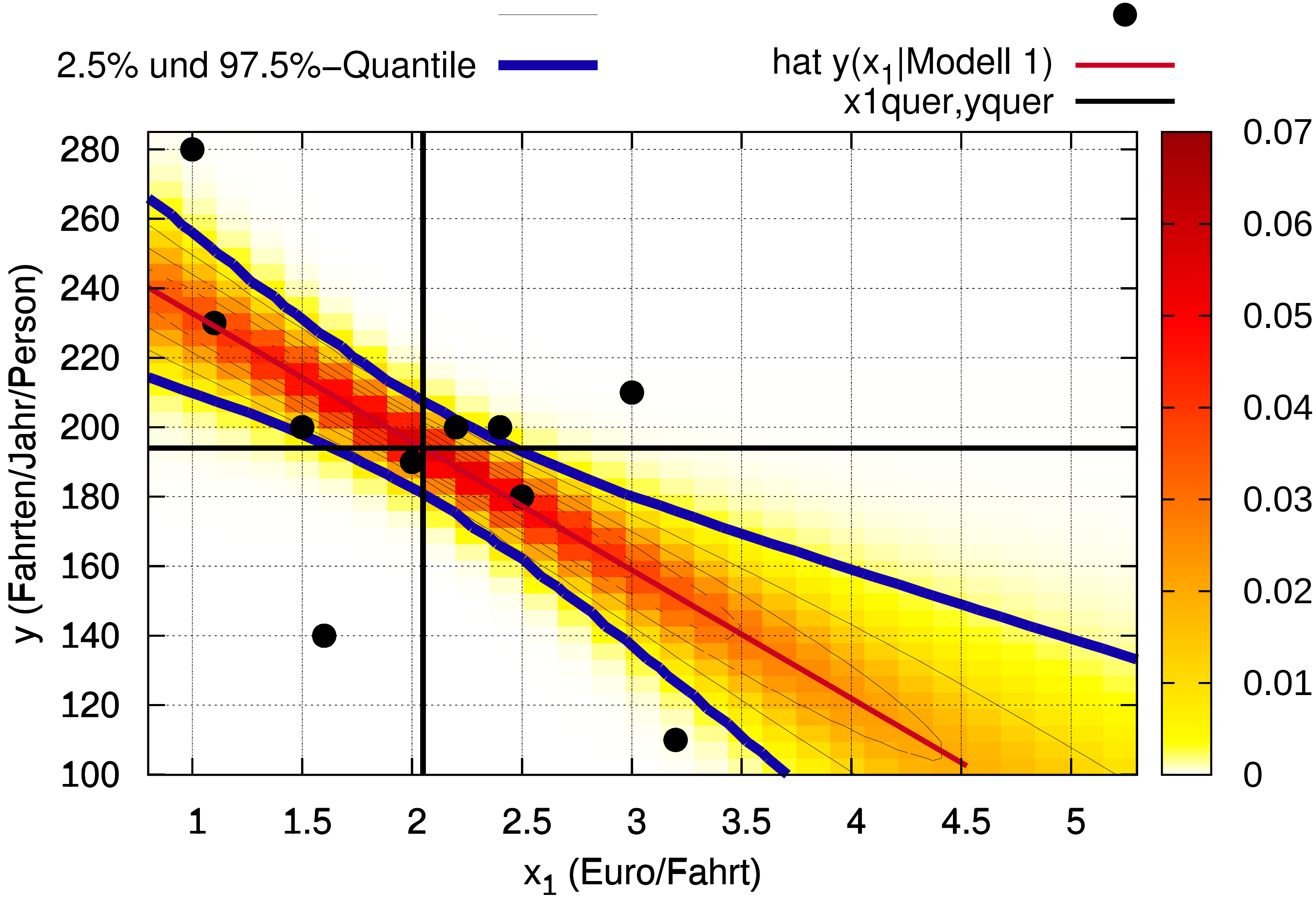
Die exogenen Variablen sind korreliert, aber nicht perfekt kolinear



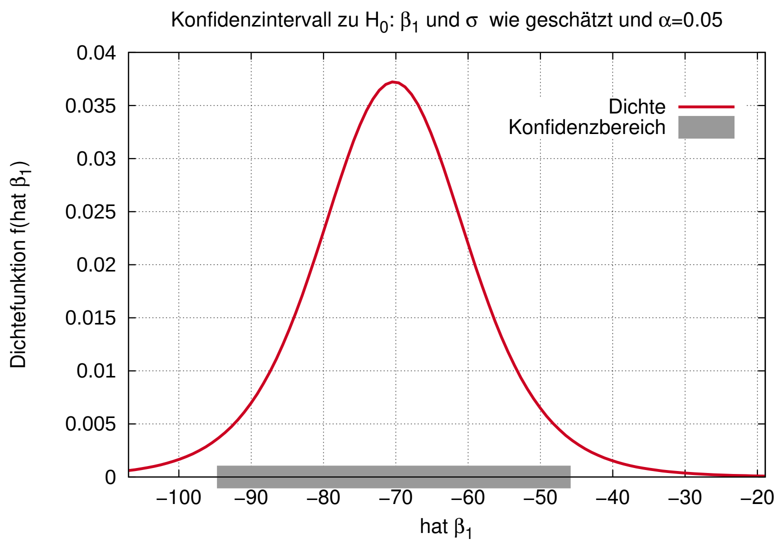


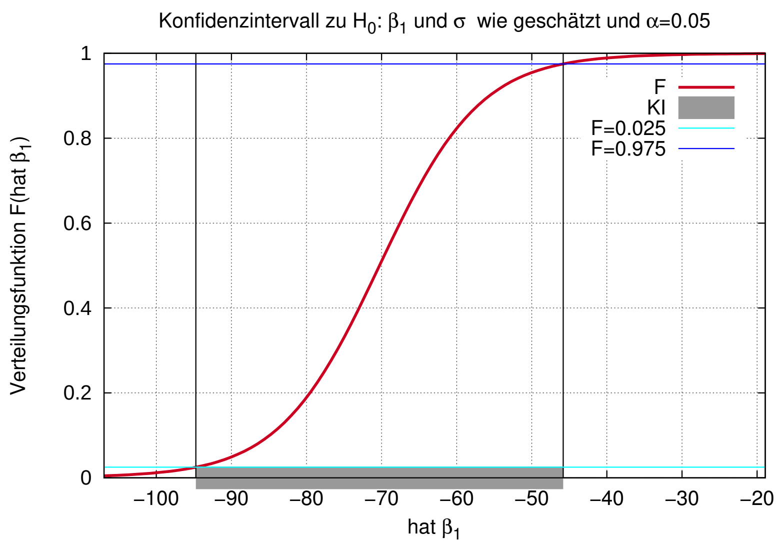
x_1=Preis (Euro), x_2=Geschwindigkeit (km/h), y = Nutzungszahl

Akzeptanzintervalle des Teilmodells M_1 (nur x_1)

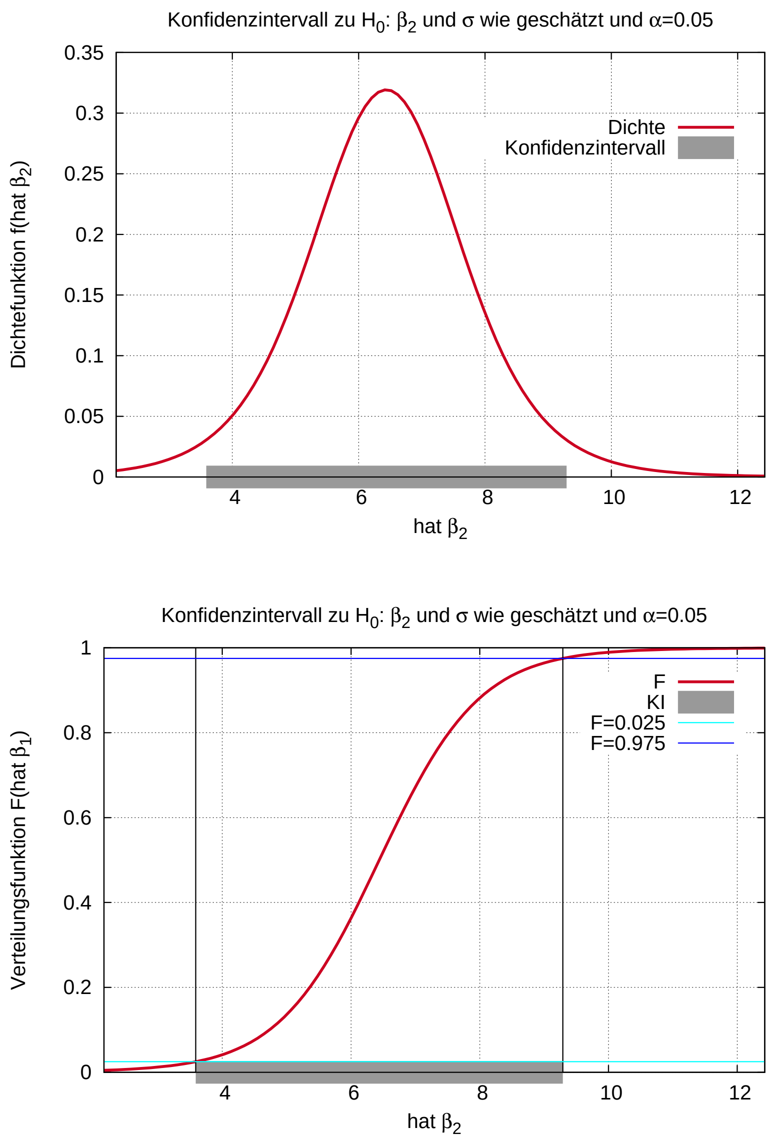


Konfidenzintervall zu H_{01}: $\beta_1 = \hat{\beta}_1 = 70$ (volles Modell)

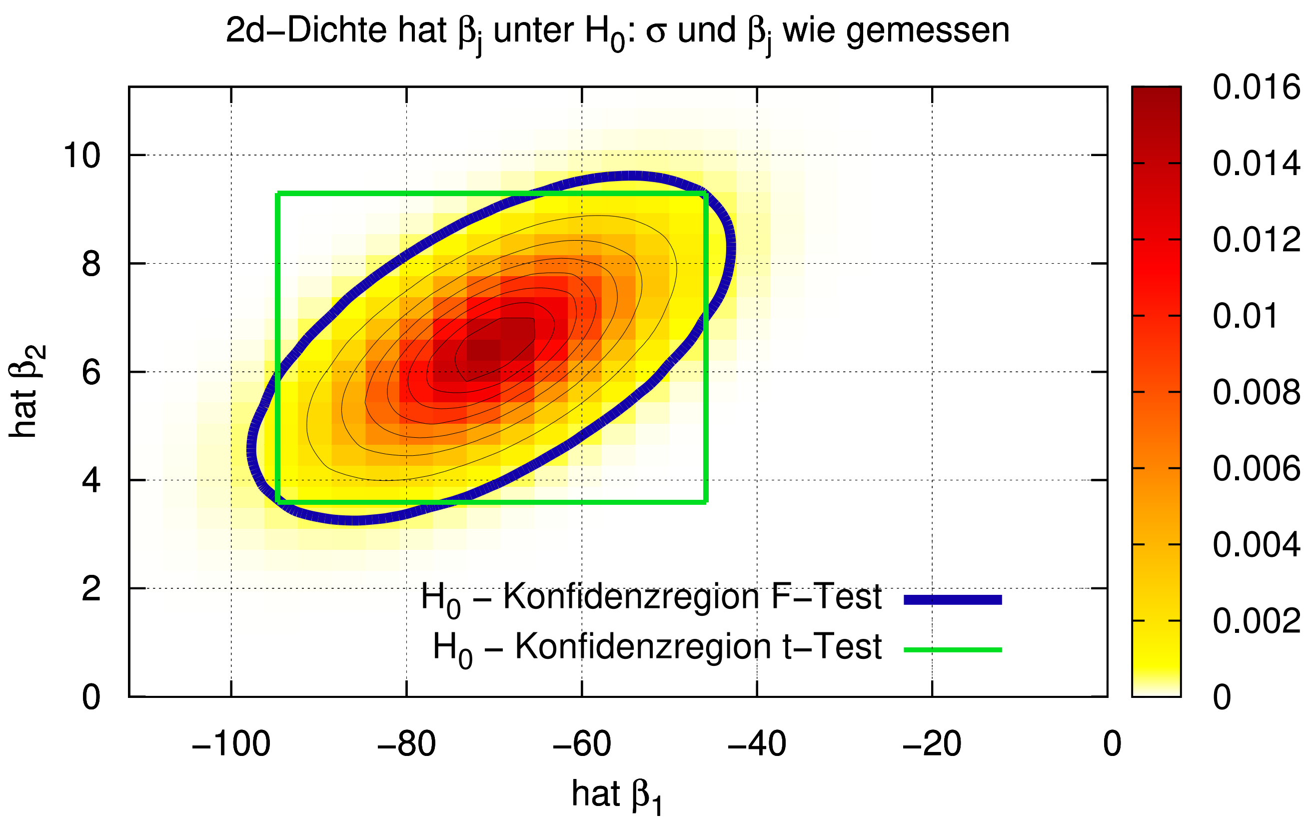




Konfidenzintervall zu H_{02}: $\beta_2 = \hat{\beta}_2 = 6.5$ (volles Modell)



Likelihoodfunktion der Anstiegsparameter

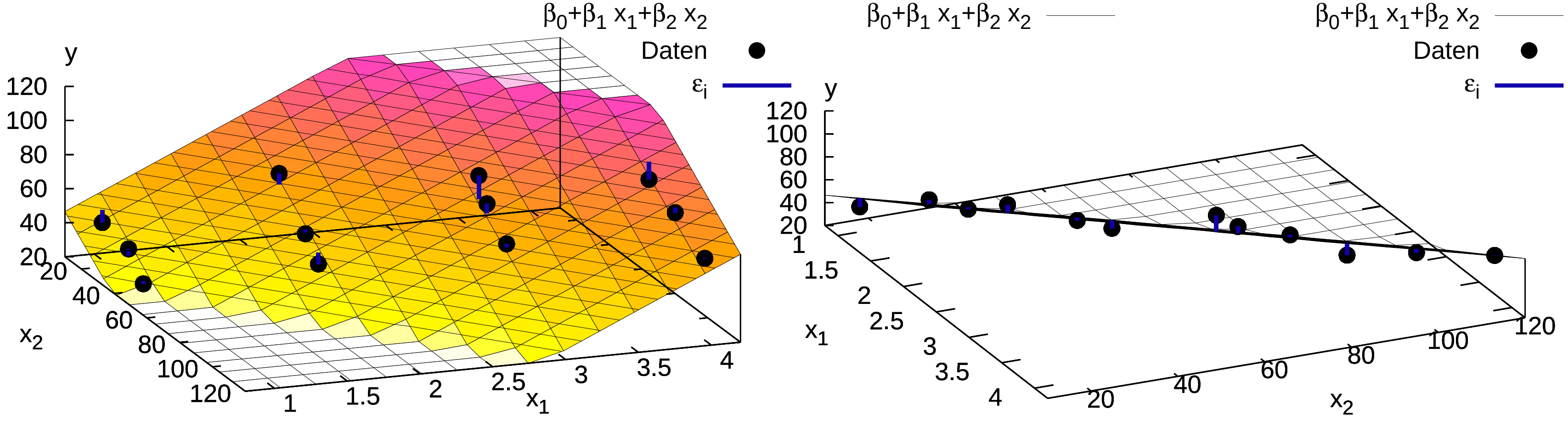


- t-Test: Die zwei separaten Nullhypothesen $H_{01} : \beta_1 = \beta_{10}$ und $H_{02} : \beta_2 = \beta_{20}$ sind beide erfüllt

- F-Test für die verbundene Nullhypothese $H_0^* : \beta_1 = \beta_{10}$, $\beta_2 = \beta_{20}$

Korrelation der Schwankungsbreiten von $\hat{\beta}_1$ und $\hat{\beta}_2$:
$r_{\hat{\beta}_1,\hat{\beta}_2} = 0.60.$

Hotelbeispiel I: Geschätztes Modell und Residualfehler

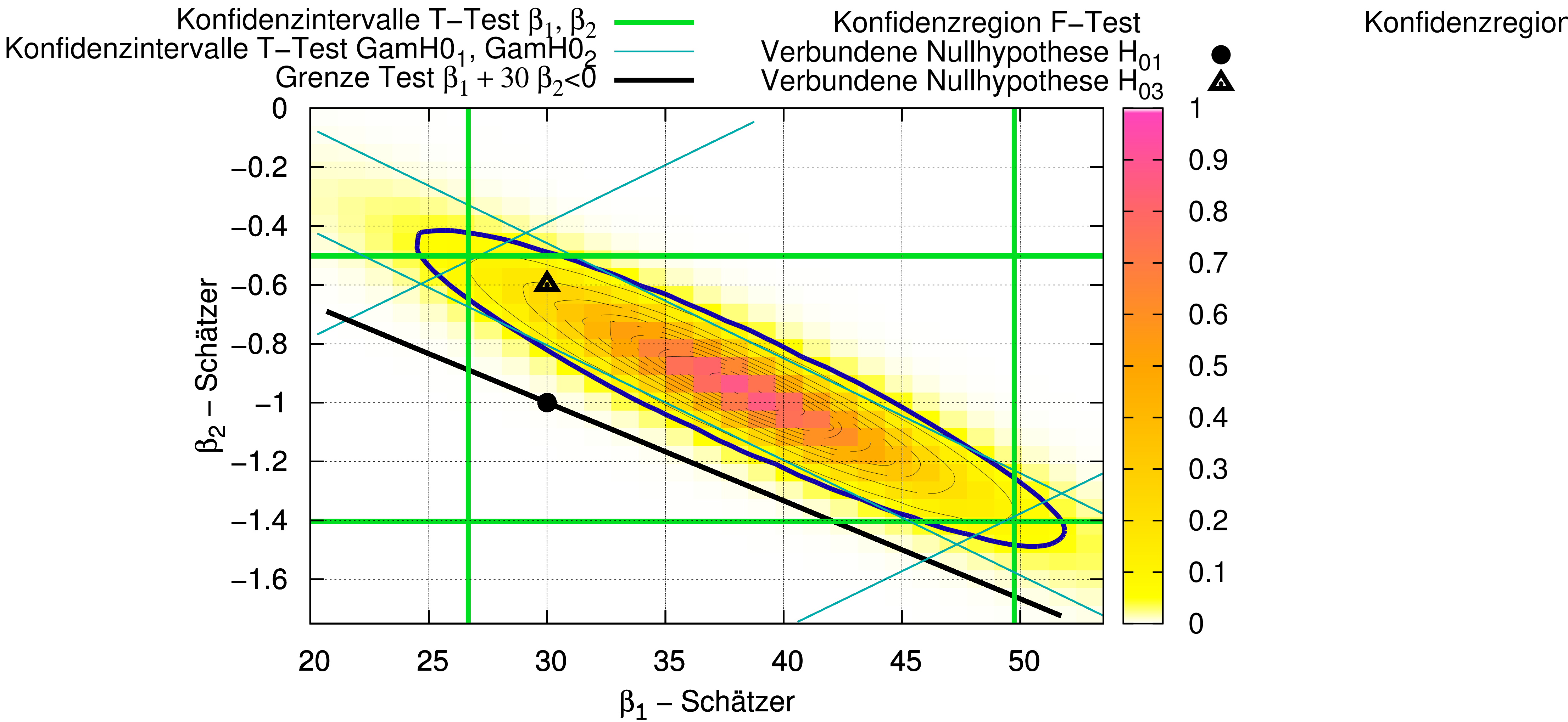


Zwei Perspektiven der Ebene des deterministischen Teils des geschätzten Modells

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2,$$

mit $\hat{\beta}_0 = 25.5$, $\hat{\beta}_1 = 38.2$ und $\hat{\beta}_2 = -0.953$ sowie die Abweichung $\epsilon_i = y_i - \hat{y}_i$ der Datenpunkte vom Modell

Hotelbeispiel II: Zweidimensionale Konfidenzregionen

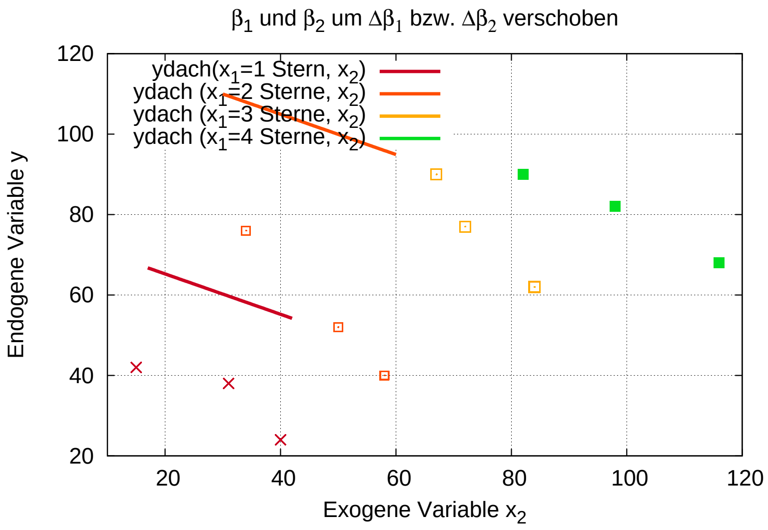


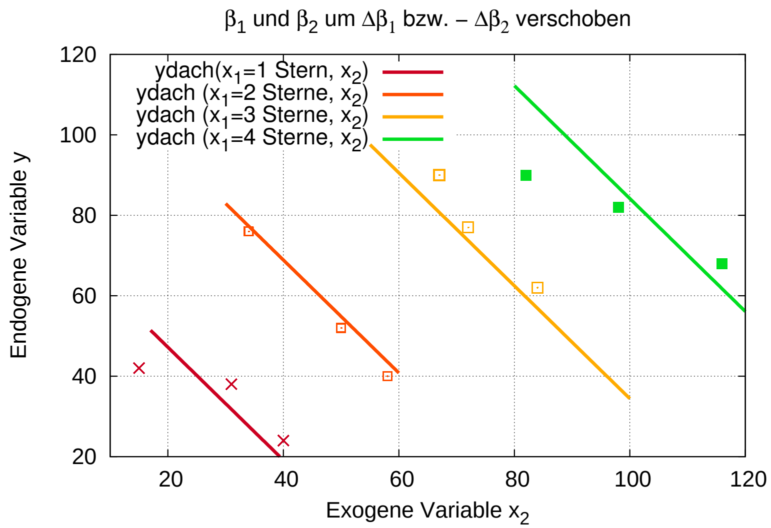
Verbundene Nullhypothesen:

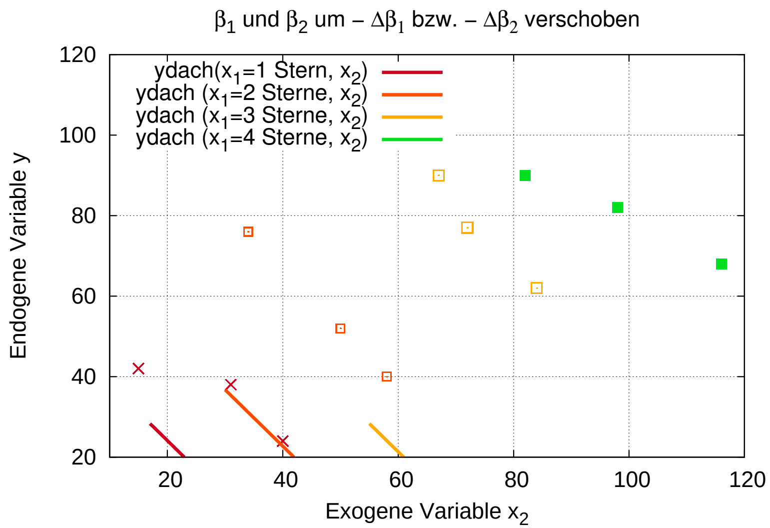
$\bullet = H_{01} : \beta_{10} = 30$ und $\beta_{20} = -1$

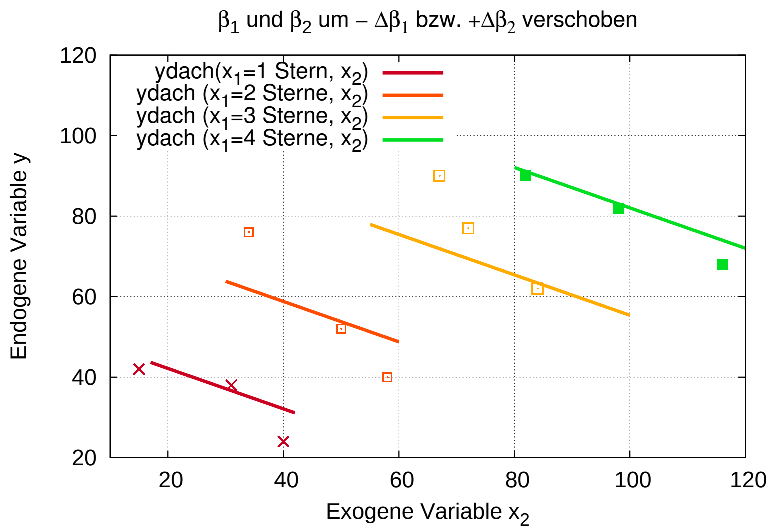
$\triangle = H_{02} : \beta_{10} = 34$ und $\beta_{20} = -1$

Hotelbeispiel III: Falsch geschätzt!



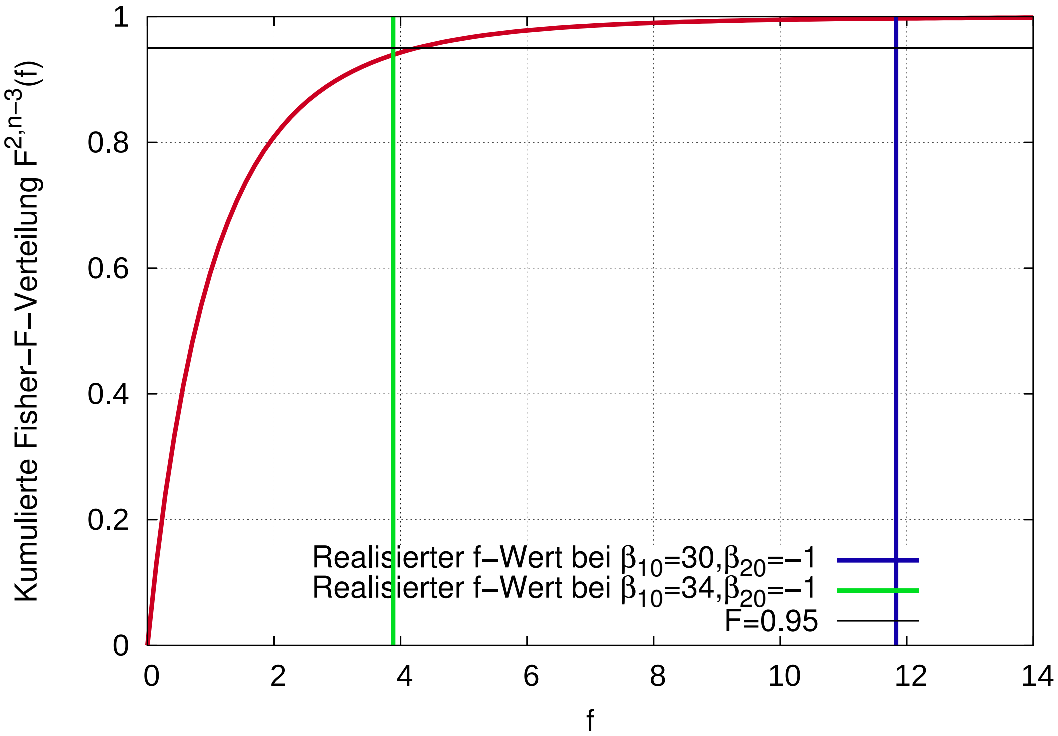






Übereinstimmung zwischen Modell und Daten für vier verschiedene Parametrisierungen

Hotelbeispiel IV: F-Test zweier verbundenen Nullhypothesen



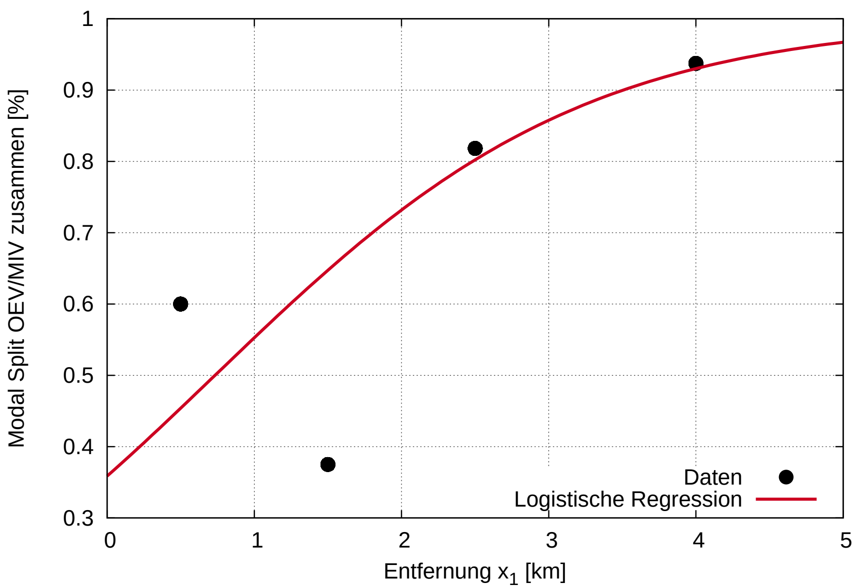
Verbundene Nullhypothesen:

$\bullet = H_{01} : \beta_{10} = 30$ und $\beta_{20} = -1$

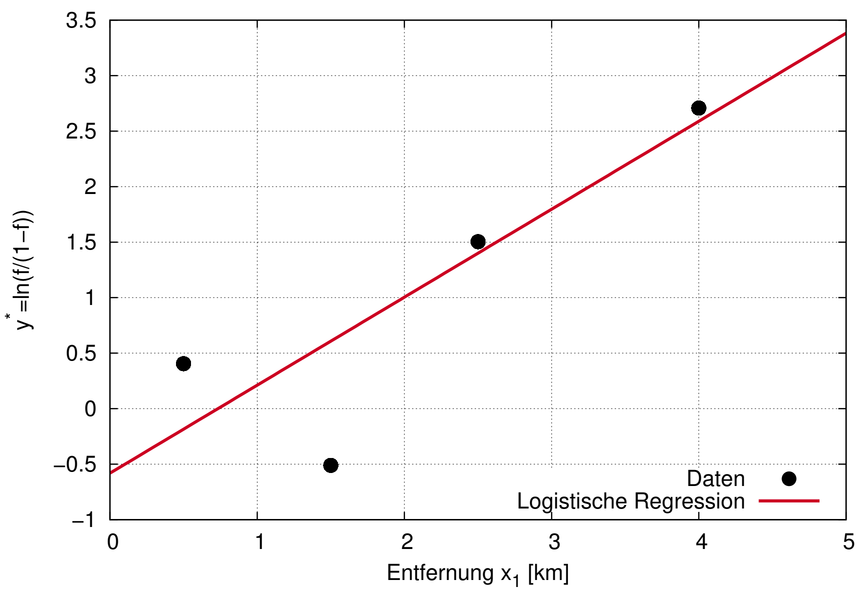
$\triangle = H_{02} : \beta_{10} = 34$ und $\beta_{20} = -1$

Logistische Regression mit naiver LSE-Schätzung der log-Odd-Ratios: RC-Umfrage WS14/15 und WS15/16 kumuliert

Daten und Ergebnis
mit 4 Entfernungsklassen



Unbeobachtete Variable
$y^* = \ln(f_1/(1-f_1))$

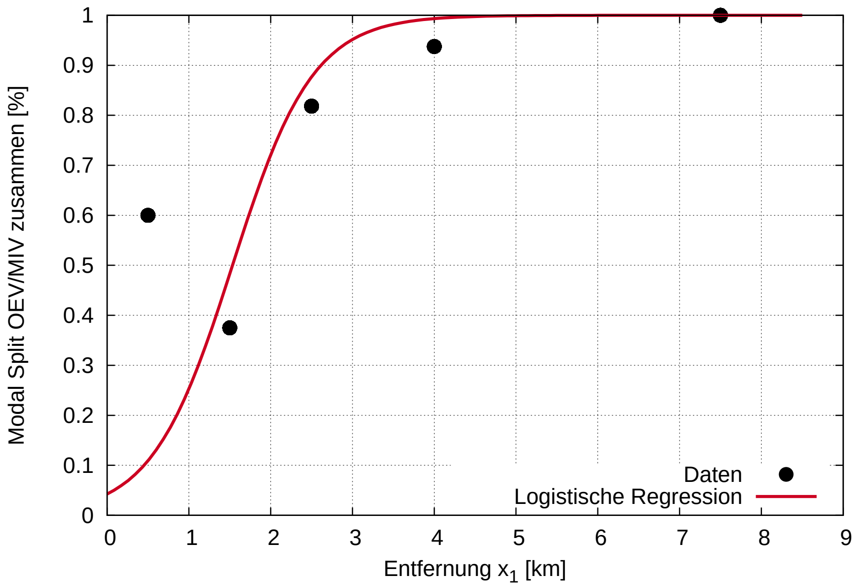


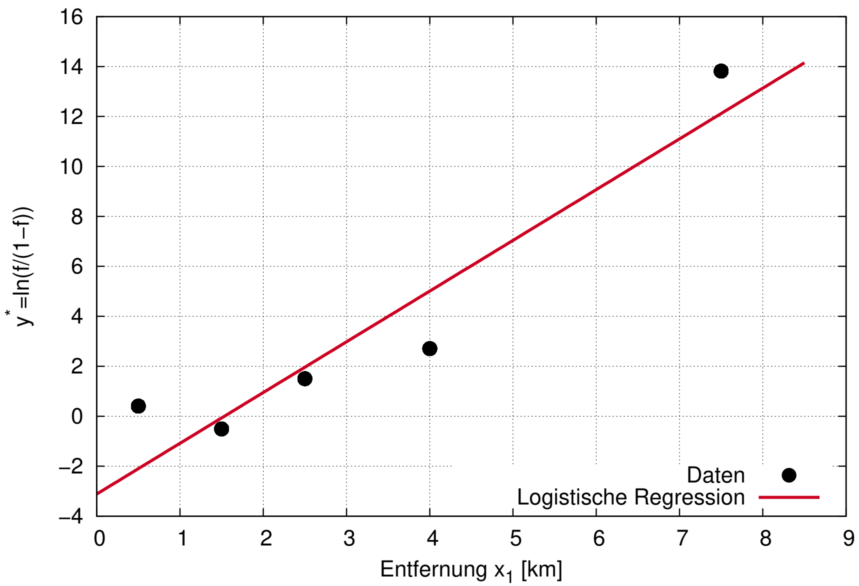
$$\beta_0 = -0.58, \quad \beta_1 = 0.79$$

Logistische Regression
mit naiver LSE-Schätzung der log-Odd-Ratios:
5. Datenpunkt addiert mit f=0.9999

Daten und Ergebnis
mit 4 Entfernungsklassen

Unbeobachtete Variable
$y^* = \ln(f_1/(1-f_1))$

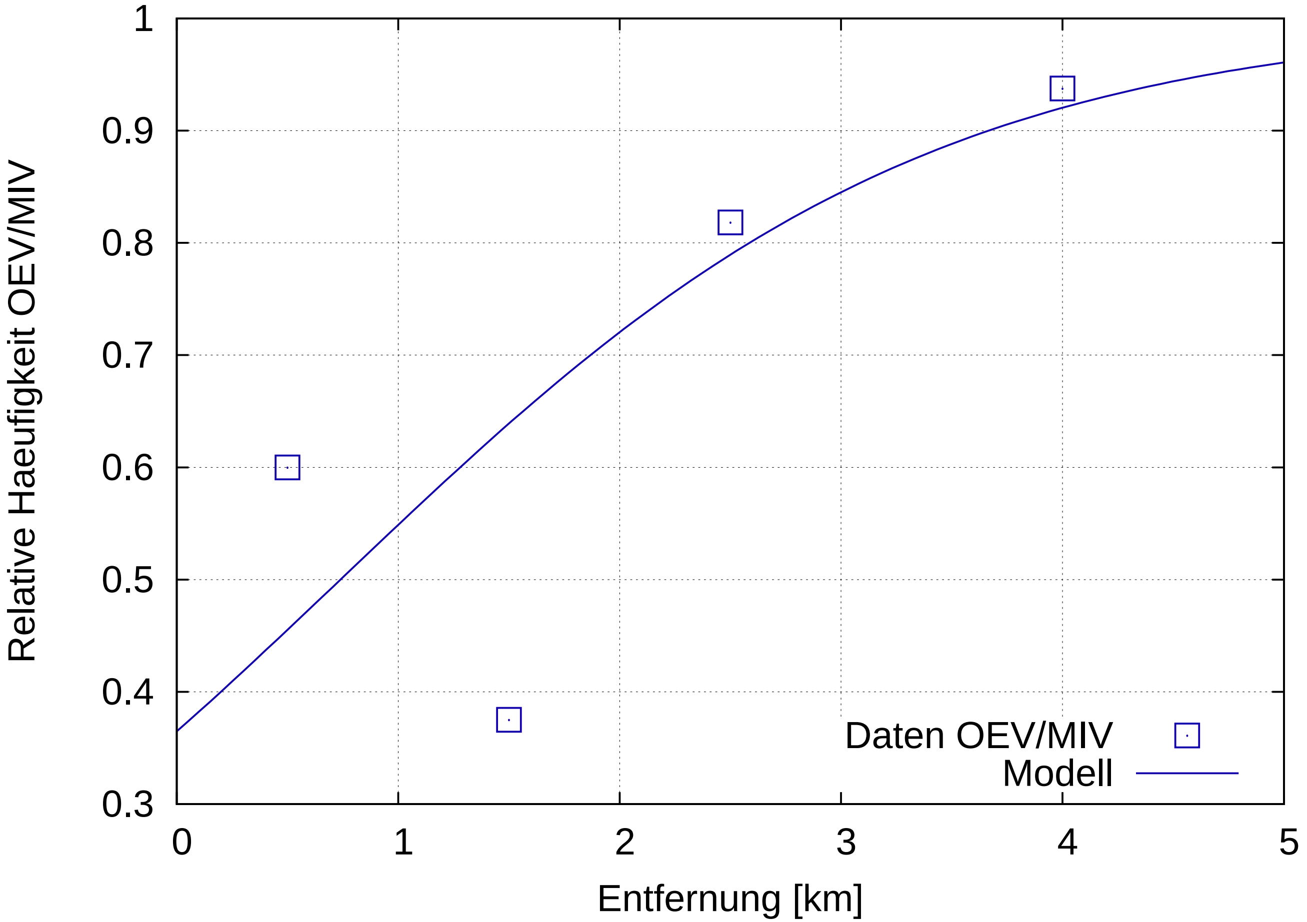




$$\beta_0 = -3.12, \quad \beta_1 = 2.03$$

Vergleich: “echte” Maximum-Likelihood-Schätzung

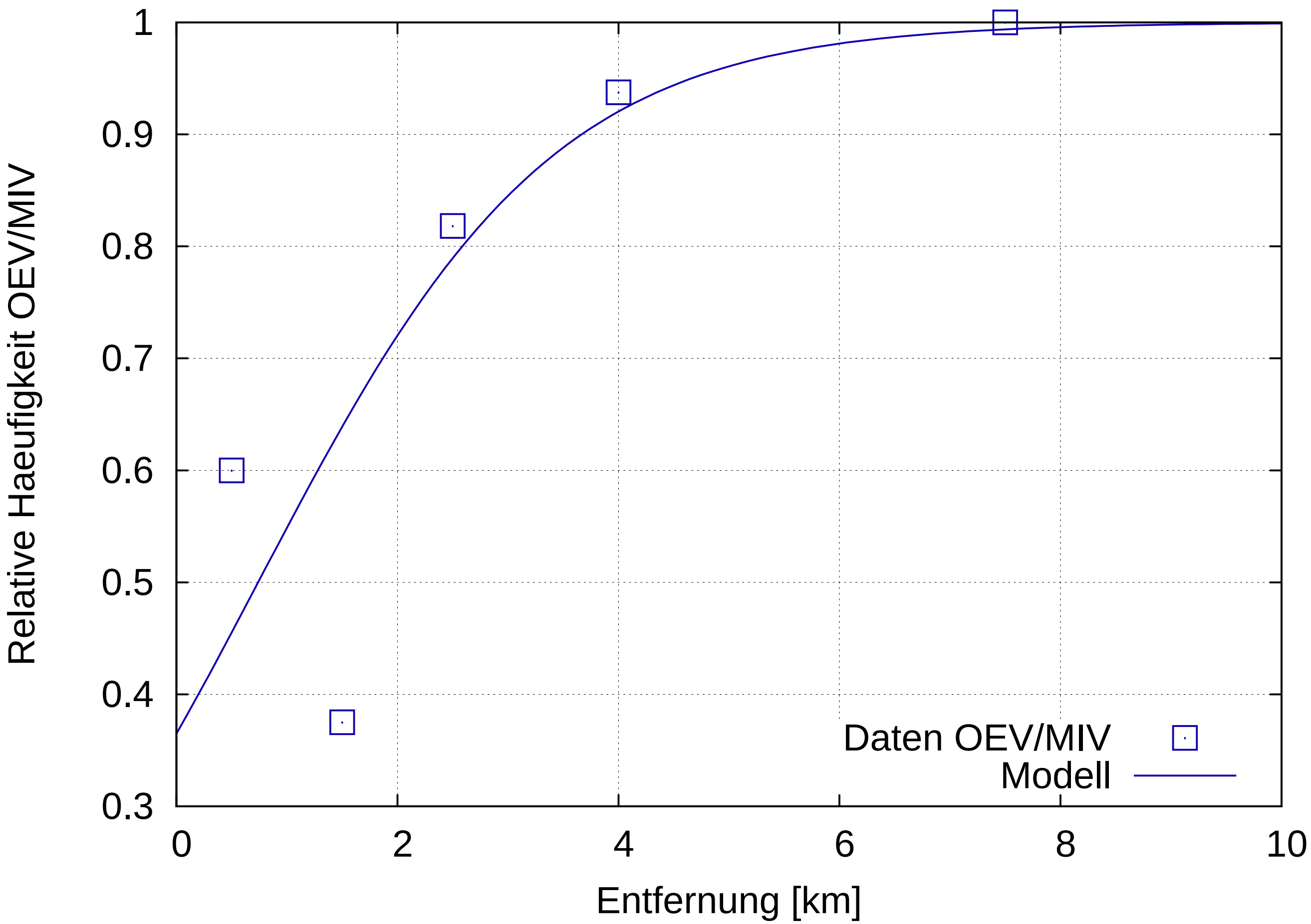
Alternativen 0 (kein ÖV) und 1 (ÖV) $V_i(r) = \beta_0 \delta_{i1} + \beta_1 r \delta_{i1}$



$$\beta_0 = -0.50 \pm 0.65,$$
$$\beta_1 = +0.71 \pm 0.30$$

Vergleich: “echtes” Maximum-Likelihood-Schätzung mit 5. Datenpunkt

Alternativen 0 (kein ÖV) und 1 (ÖV) $V_i(r) = \beta_0 \delta_{i1} + \beta_1 r \delta_{i1}$



$$\beta_0 = -0.55 \pm 0.63,$$
$$\beta_1 = +0.75 \pm 0.27$$