

Lineare Regressionsfunktion

$$M_{Y|X} = \alpha + \beta \cdot X$$

$M_{Y X}$	Variable, die die bedingten Populationsmittelwerte der abhängigen Variablen Y als Funktion von X beschreibt
α	Regressionskonstante α (alpha) in der Population
β	Regressionsgewicht β (beta) in der Population
X	erklärende (bedingende) Variable

$$Y = M_{Y|X} + U = \alpha + \beta \cdot X + U$$

Y	abhängige Variable Y mit den Realisationen y_i
U	Residual variable U mit den Ausprägungen u_i [Populationsresiduen]

Schätzung der Regressionsgeraden mit Stichprobendaten

$$Y = \hat{Y} + E = a + b \cdot X + E$$

\hat{Y} Vorhersagewerte, **Schätzer** der Regressionsgeraden in der Population
a, b **Schätzer** der Regressionskoeffizienten α und β
 E Residualvariable in der Stichprobe

$$y_i = \hat{y}_i + e_i = a + b \cdot x_i + e_i$$

\hat{y}_i Vorhersagewert der i-ten Realisation von Y in der Stichprobe
 e_i i-te Realisation der **Residual**variablen E [Stichprobenresiduen]
 y_i i-te Realisation von Y

Methode der kleinsten Quadrate

OLS-Methode [**O**rdinary **L**east **S**quares]

Minimierungsfunktion

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (a + b \cdot x_i))^2 = \min$$

$$y_i = \hat{y}_i + e_i = a + b \cdot x_i + e_i$$

$$\hat{y}_i = a + b \cdot x_i$$

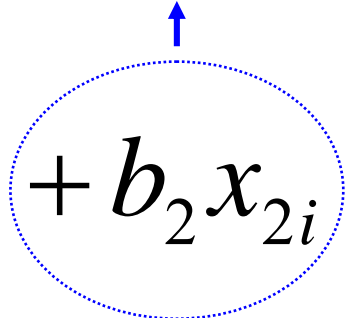
OLS Annahmen

- » Linearitätsannahme
- » Erwartungswert der Residuen gleich Null
- » Residualvariable U ist nicht mit X korreliert
[Werte von X fixiert]
- » Varianzgleichheit (Homoskedastizität) der Residuen
- » Unkorreliertheit der Residuen
- » Residuen sind normalverteilt
- » Gleichung ist identifizierbar
($n > k$, keine linearen Abhängigkeiten)

Spezifikationsfehler

Explizite und implizite Faktoren

$$y_i = a + b_1 x_{1i} + e_i$$

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + e_i$$


Modellzusammenfassung^b

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
1	,661 ^a	,436	,436	1,567	1,594

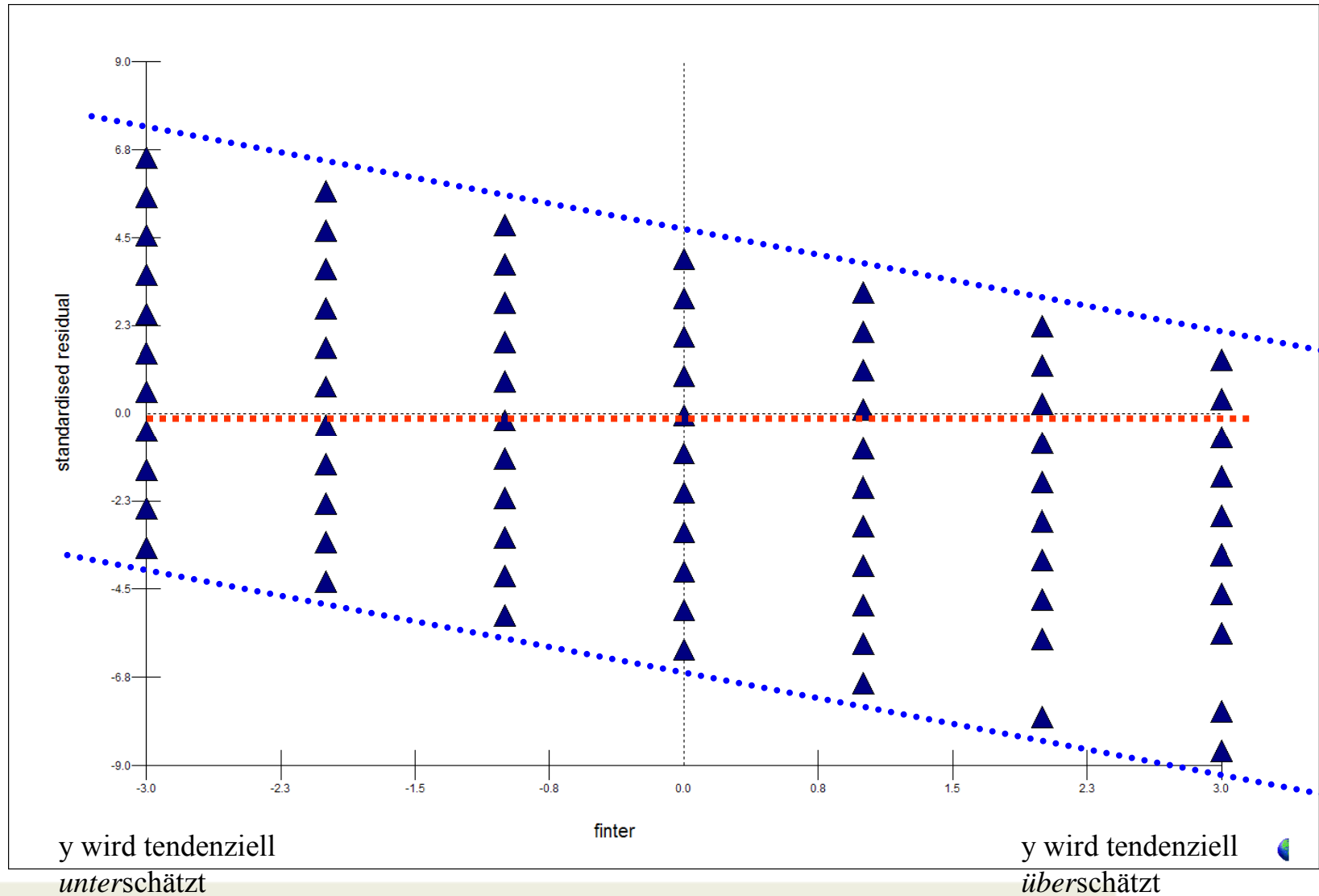
a. Einflußvariablen : (Konstante), FINTER Interesse am Fach

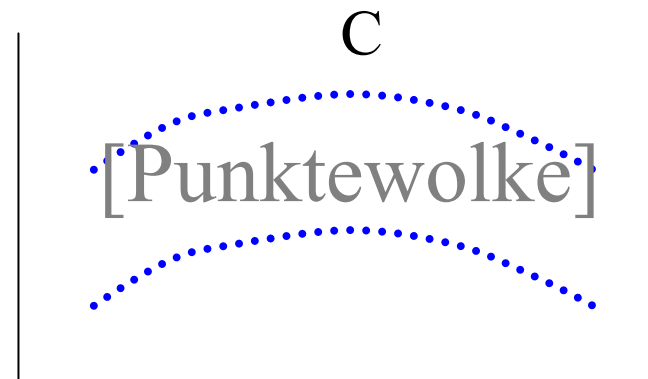
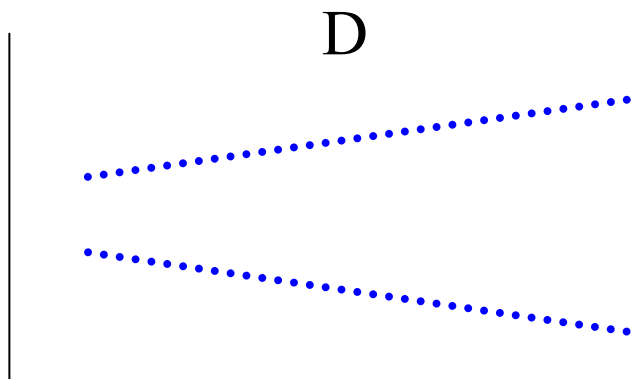
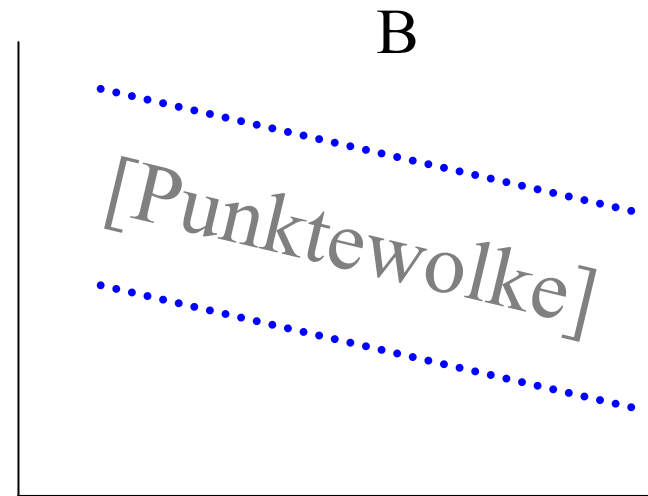
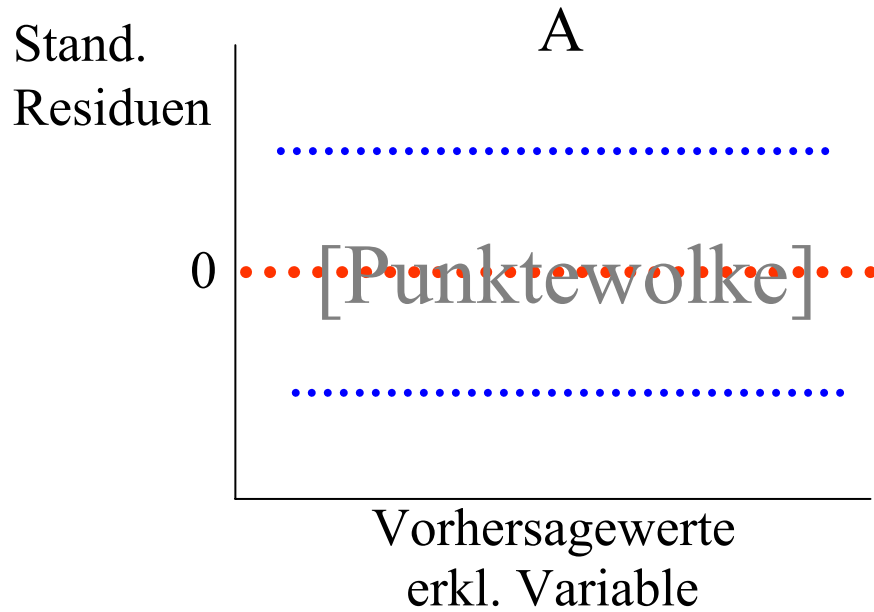
b. Abhängige Variable: GESAMT Gesamturteil

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	7,011	,019		365,791	,000
	FINTER Interesse am Fach	,878	,011	,661	77,555	,000

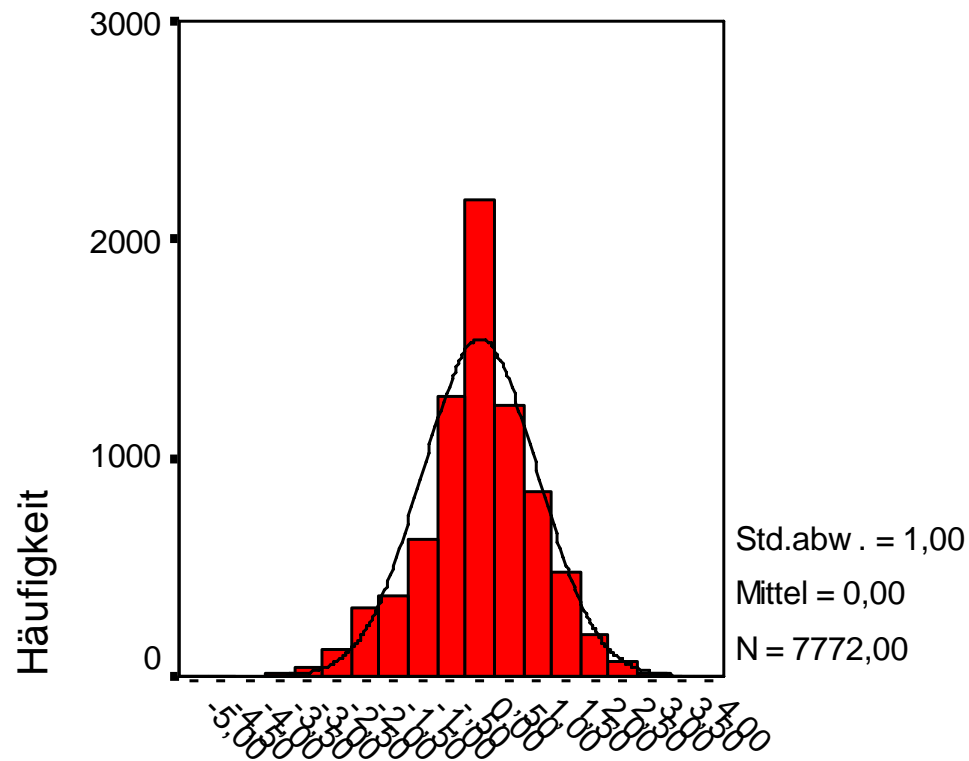
a. Abhängige Variable: GESAMT Gesamturteil





Histogramm

Abhängige Variable: Gesamturteil



Regression Standardisiertes Residuum

Serielle Korrelation / Autokorrelation

Fehler konstituieren autoregressive Serie 1. Ordnung

$$u_t = \rho u_{t-1} + \varepsilon_t \quad |\rho| < 1$$

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=2}^n e_t^2}$$

$$r = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_t^2}$$

ρ geschätzt durch r

$$d = 2(1 - r)$$

Durbin-Watson Statistik

Durbin-Watson Statistik

$$d = 2(1 - r)$$

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

$d < d_L$ weise H_0 zurück

$d > d_U$ weise H_0 nicht zurück

$d_L < d < d_U$ Test ist nicht schlüssig

Kritische Tabellenwerte für d_U und d_L ,
in Abhängigkeit von n und der Zahl k der Prädiktoren

	k = 1		k = 2		...	k = 5	
n	d_L	d_U	d_L	d_U		d_L	d_U
15	1,08	1,36	0,95	1,54	...	0,56	2,21
.
.
30	1,35	1,49	1,28	1,57	...	1,07	1,83
.
100	1,65	1,69	1,63	1,72	...	1,57	1,78