Methode der kleinsten Quadrate

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

Minimierungsfunktion

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

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

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)

Modellzusammenfassung[b]

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

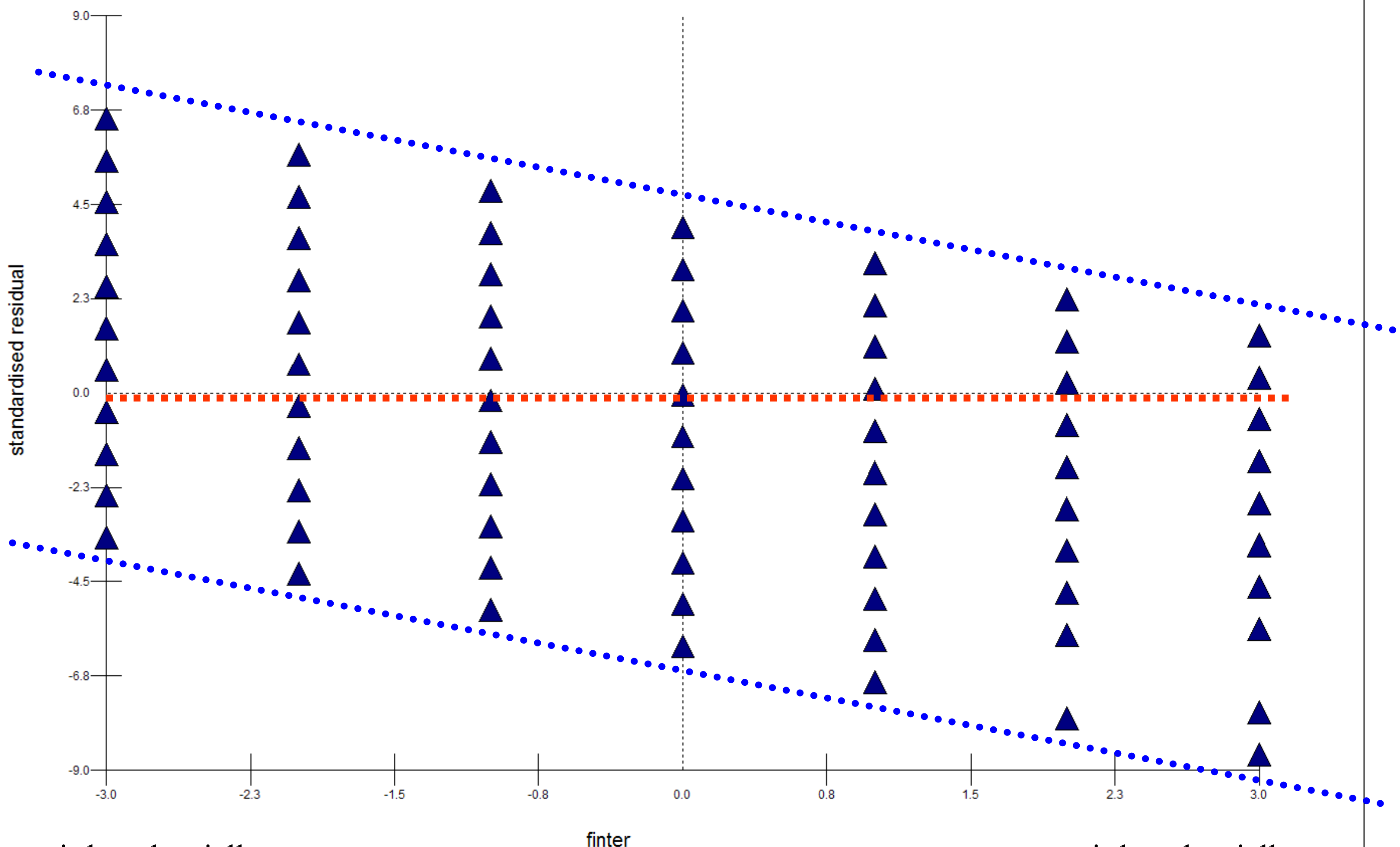
a. Einflußvariablen : (Konstante), FINTER Interesse am Fach

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

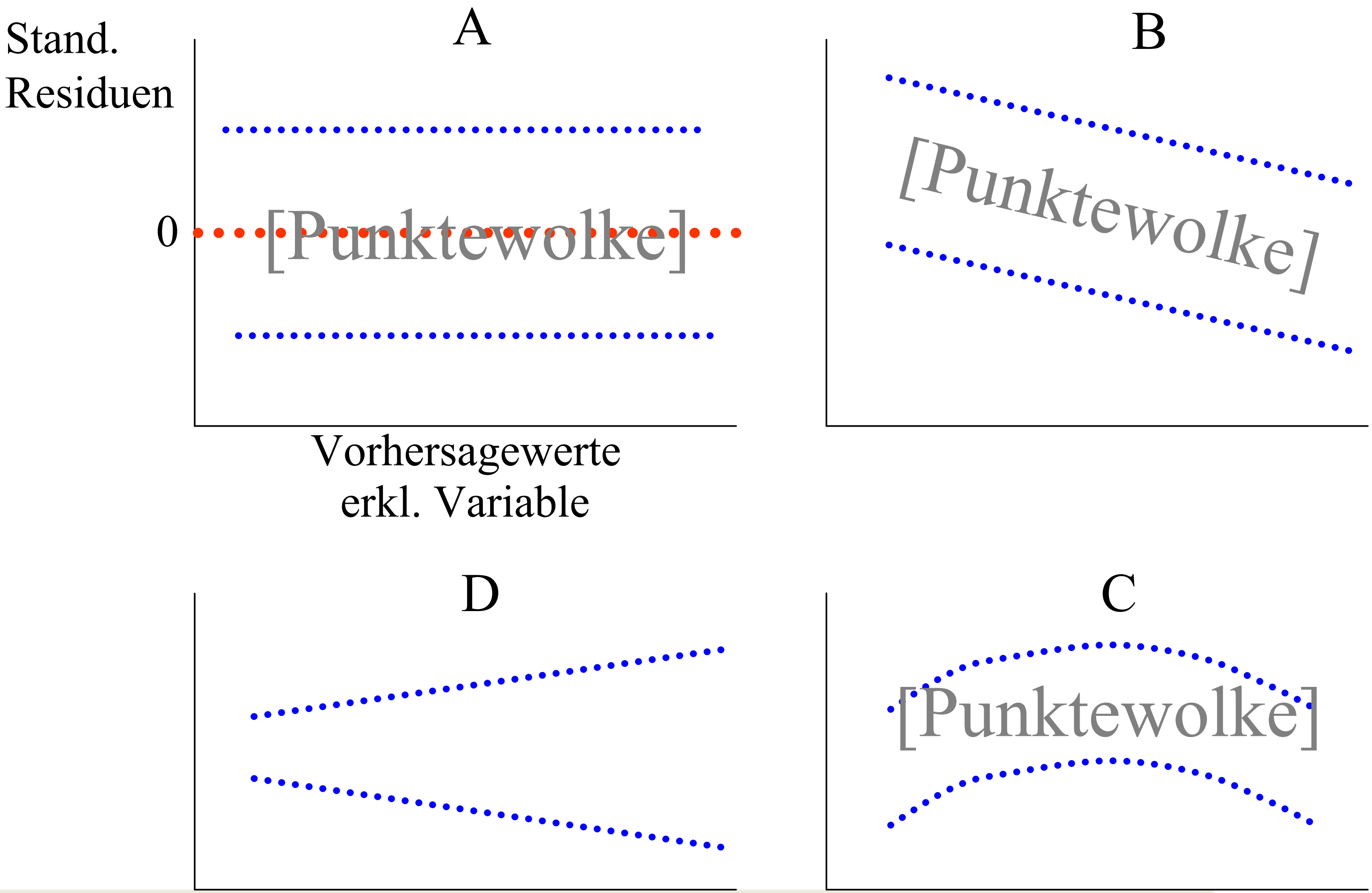
a. Abhängige Variable: GESAMT Gesamturteil



y wird tendenziell *unter*schätzt

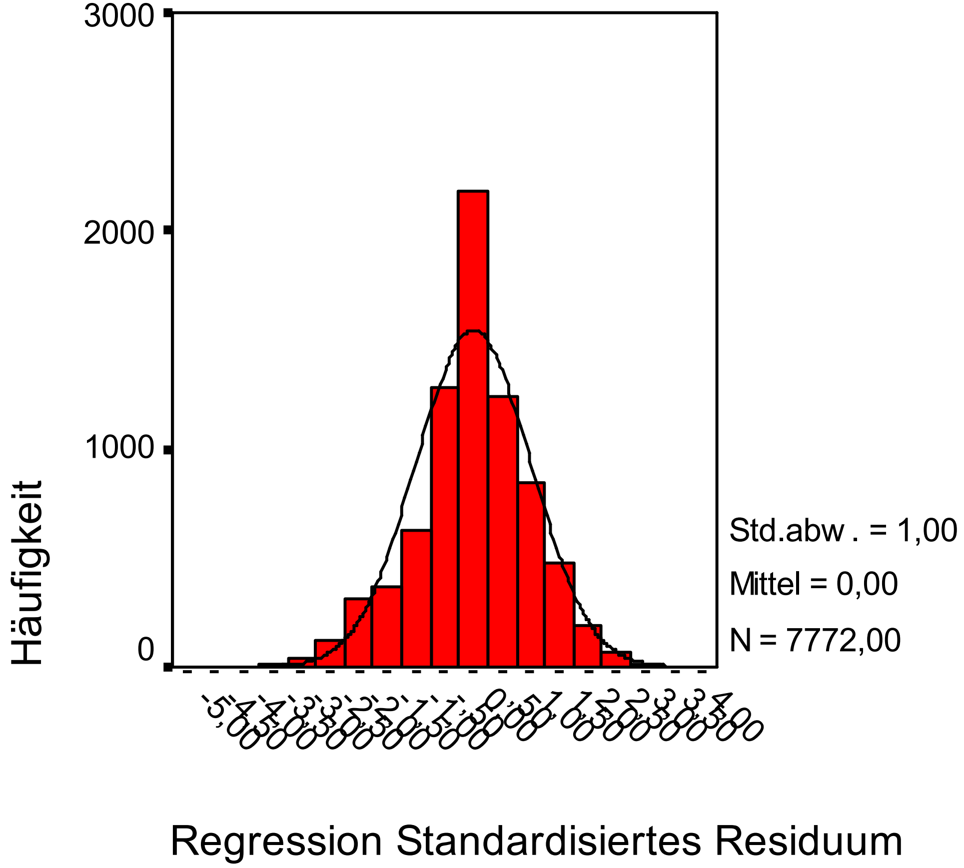
y wird tendenziell *über*schätzt

$e_i = y_i$ (beobachtet) – y_i (geschätzt)


A
B
Stand.
Residuen
0
[Punktewolke]
[Punktewolke]
Vorhersagewerte
erkl. Variable
D
C
[Punktewolke]

Histogramm

Abhängige Variable: Gesamturteil



Serielle Korrelation / Autokorrelation

Fehler konstituieren autoregressive Serie 1. Ordnung

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

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \qquad r = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \qquad r = \frac{\sum_{t=2}^{n} (e_t - 0)(e_{t-1} - 0)}{\sum_{t=1}^{n} (e_t - 0)^2}$$

ρ geschätzt durch r

$$\boxed{d = 2\,(1 - r)}$$

Durbin-Watson Statistik

H_0: $\rho = 0$ H_1: $\rho > 0$

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

1		$1^2=$ 1	1		$1^2=$ 1
1	1×1 = 1	$1^2=$ 1	3	3×1= 3	$3^2=$ 9
2	2×1 = 2	$2^2=$ 4	2	2×3= 6	$2^2=$ 4
2	2×2 = 4	$2^2=$ 4	5	5×2=10	$5^2=$25
3	3×2 = 6	$3^2=$ 9	3	3×5=15	$3^2=$ 9
3	3×3 = 9	$3^2=$ 9	4	4×3=12	$4^2=$16
4	4×3 = 12	$4^2=$ 16	2	2×4= 8	$2^2=$ 4
3	3×4 = 12	$3^2=$ 9	3	3×2= 6	$3^2=$ 9
2	2×3 = 6	$2^2=$ 4	1	1×3= 3	$1^2=$ 1
2	2×2 = 4	$2^2=$ 4	4	4×1= 4	$4^2=$16
1	1×2 = 2	$1^2=$ 1	2	2×4= 8	$2^2=$ 4
Σ	58	62	Σ	75	98

58/62=0.94

75/98= 0.77

Durbin-Watson Statistik

$$\boxed{d = 2\,(1 - r)}$$

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

Durbin-Watson d

	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

Aus Chatterjee, S./B. Price (1977) Regression Analysis by Example. New York: Wiley, p. 223; Kritische Tabellenwerte für Signifikanzniveau von 5%.

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

Durbin-Watson Statistik

$$\boxed{d = 2\,(1 - r)}$$

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

Modellzusammenfassung[b]

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

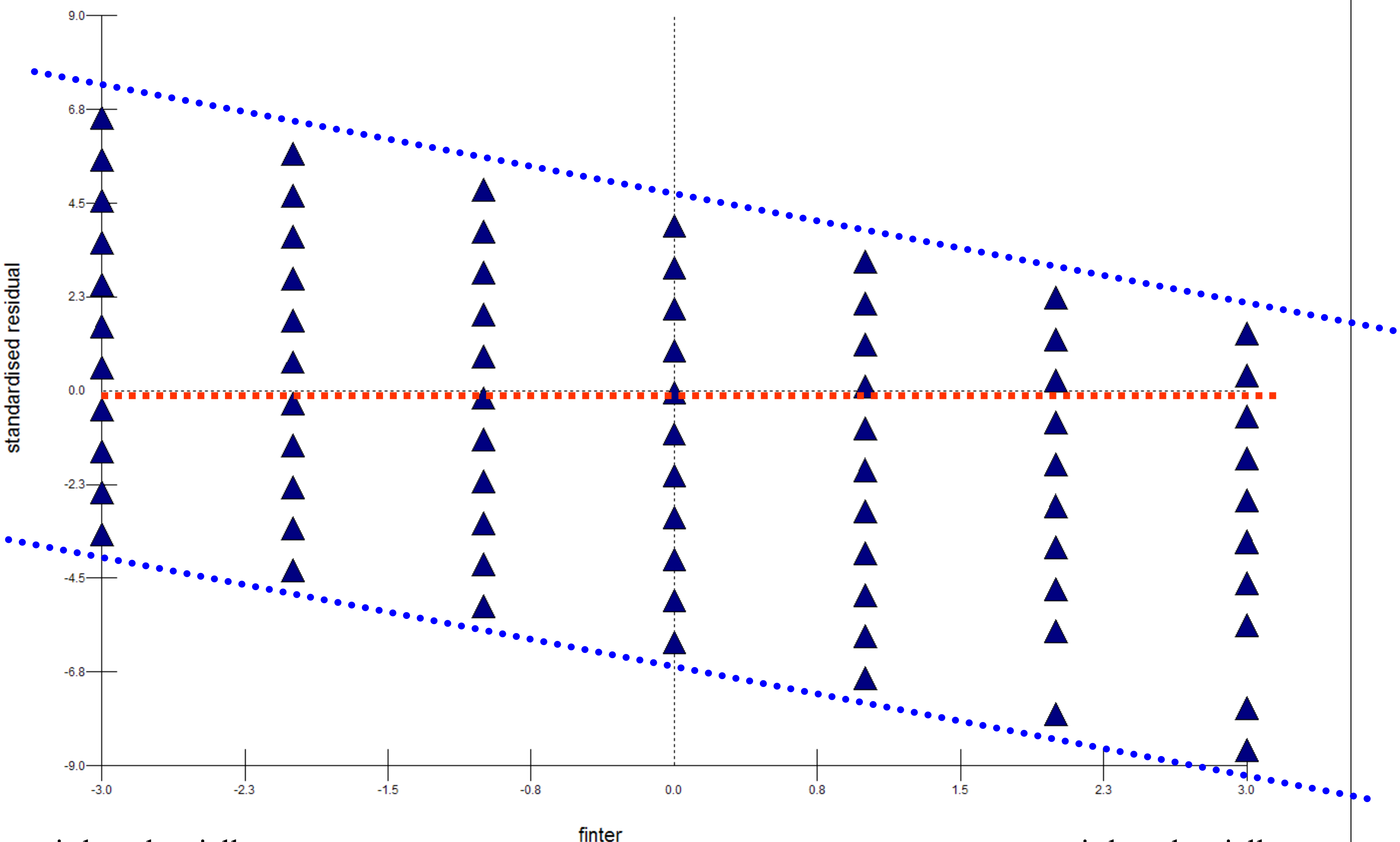
a. Einflußvariablen : (Konstante), FINTER Interesse am Fach

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



y wird tendenziell *unter*schätzt

y wird tendenziell *über*schätzt

$e_i = y_i$ (beobachtet) – y_i (geschätzt)

Durbin-Watson Statistik

$$\boxed{d = 2\,(1 - r)}$$

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

T-Test für den Mittelwertvergleich zwischen **zwei** Gruppen

SPSS:

» Analysieren
- » Mittelwerte vergleichen
  - » T-Test bei unabhängigen Stichproben
    - » Testvariable(n) und Gruppenvariable [eintragen]

Gruppenstatistiken					
	V2	N	Mittel-wert	Standard-abweichung	Standardfehler des Mittelwertes
V1	0	150	5,980	2,3473	,1917
	1	528	7,263	2,3093	,1005

Organisationsqualität und Verfügbarkeit von Informationsmaterialien (N=678)

Test bei unabhängigen Stichproben							
	T-Test für die Mittelwertgleichheit*						
	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
						Untere	Obere
	-5,984	676	,000	-1,283	,2144	-1,7043	-,8622

*unter der Annahme gleicher Varianzen

SPSS:

» Analysieren
  » Mittelwerte vergleichen
    » Einfaktorielle ANOVA
      » Abhängige Variable(n) und Faktor [eintragen]

ONEWAY ANOVA

STUDORGA

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	192,364	1	192,364	35,810	,000
Innerhalb der Gruppen	3631,347	676	5,372		
Gesamt	3823,711	677			

F-Test für den Mittelwertvergleich zwischen **mehr als zwei** Gruppen

ONEWAY ANOVA

STUDORGA

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	492,291	4	123,073	24,863	,000
Innerhalb der Gruppen	3331,420	673	4,950		
Gesamt	3823,711	677			

Organisationsqualität (11stufige Skala von 1 bis 11) und Fakultätszugehörigkeit (5 Fakultäten; N=678)

ANOVA – Varianzanalysetabelle

Streuungs-ursache	Streu-ung	Freiheits-grade	Mittlerer quadratischer Fehler	Prüfgröße
Gruppen (Variabilität *zwischen* den Gruppen)	SQE	$I - 1$	$\frac{SQE}{(I-1)} = MQE$	$F = \frac{MQE}{MQR}$
Residuen Variabilität *innerhalb* der Gruppen	SQR	$n - I$	$\frac{SQR}{(n-I)} = MQR$	

I Anzahl der Gruppen, *n* Zahl der Fälle

F – Test

$$F = \frac{SQE/(I-1)}{SQR/(n-I)} = \frac{\sum_{i=1}^{I} n_i \left(\bar{y}_{i.} - \bar{y}_{..}\right)^2 \Big/ (I-1)}{\sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_{i.}\right)^2 \Big/ (n-I)}$$

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$ gegen $H_1 : \mu_i \neq \mu_{i'}$ für mindestens ein Paar (i, i')

H_0 ist zu verwerfen ist, falls $F > F_{1-\alpha}(I-1, n-I)$.

I Anzahl der Gruppen, n_i Zahl der Fälle in Gruppe *i*

F-Test für den Mittelwertvergleich zwischen **mehr als zwei** Gruppen

ONEWAY ANOVA

STUDORGA

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	492,291	4	123,073	24,863	,000
Innerhalb der Gruppen	3331,420	673	4,950		
Gesamt	3823,711	677			

Organisationsqualität (11stufige Skala von 1 bis 11) und Fakultätszugehörigkeit (5 Fakultäten; N=678)

F – Test

$$F = \frac{SQE/(I-1)}{SQR/(n-I)} = \frac{\sum_{i=1}^{I} n_i\left(\bar{y}_{i.} - \bar{y}_{..}\right)^2 \Big/ (I-1)}{\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{i.}\right)^2 \Big/ (n-I)}$$

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$ gegen $H_1 : \mu_i \neq \mu_{i'}$ für mindestens ein Paar (i, i')

H_0 ist zu verwerfen ist, falls $F > F_{1-\alpha}(I-1, n-I)$.

I Anzahl der Gruppen, n_i Zahl der Fälle in Gruppe *i*

Streuungszerlegung in der Regressionsanalyse

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Gesamtstreuung SQT
= Reststreuung SQR + erklärte Streuung SQE

Gesamtstreuung SS_y
= Reststreuung SS_E + erklärte Streuung $SS_{\hat{y}}$

Chi²-verteilte Zufallsvariable

resultiert, wenn eine standardnormalverteilte Zufallsvariable quadriert wird oder wenn mehrere, voneinander unabhängige standardnormalverteilte Zufallsvariablen quadriert und aufsummiert werden

$$\chi^2_{df=\nu} = \chi^2_{\nu} = Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_{\nu}^2$$

Chi²-Verteilung mit *v* Freiheitsgraden

df = Anzahl Freiheitsgrade (Anzahl der Summanden)

Chi²-Verteilung

Kennwerteverteilung von Variationen und Varianzen (df=n-1)

Siehe Basislektüre (Kühnel/Krebs), Kap. 7.4

$$T = \frac{Z}{\sqrt{\frac{\chi^2}{df}}}$$

T = t-verteilte Zufallsvariable T mit df Freiheitsgraden
Z = standardnormalverteilte Zufallsvariable
χ^2= von Z unabhängige, mit df Freiheitsgraden
chi^2-verteilte Zufallsvariable
df = Zahl der Freiheitsgrade

F-Verteilung

Kennwerteverteilung von Varianzverhältnissen

$$F = \frac{\dfrac{\chi_1^2}{df_1}}{\dfrac{\chi_2^2}{df_2}}$$