

Prognoseintervalle für y_0 gegeben x_0

- Intervallprognosen für y_0 zur Vertrauenswahrscheinlichkeit $1 - \alpha$ erhält man also analog zu den Intervallprognosen für $E(y_0)$ in der Form

$$\begin{aligned} & [\hat{y}_0 - t_{n-2;1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{e_0}, \hat{y}_0 + t_{n-2;1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{e_0}] \\ & = \left[(\hat{\beta}_1 + \hat{\beta}_2 \cdot x_0) - t_{n-2;1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{e_0}, (\hat{\beta}_1 + \hat{\beta}_2 \cdot x_0) + t_{n-2;1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{e_0} \right]. \end{aligned}$$

- Im Beispiel (Ausgaben in Abhängigkeit vom Einkommen) erhält man zu gegebenem $x_0 = 38$ (in 100 €)

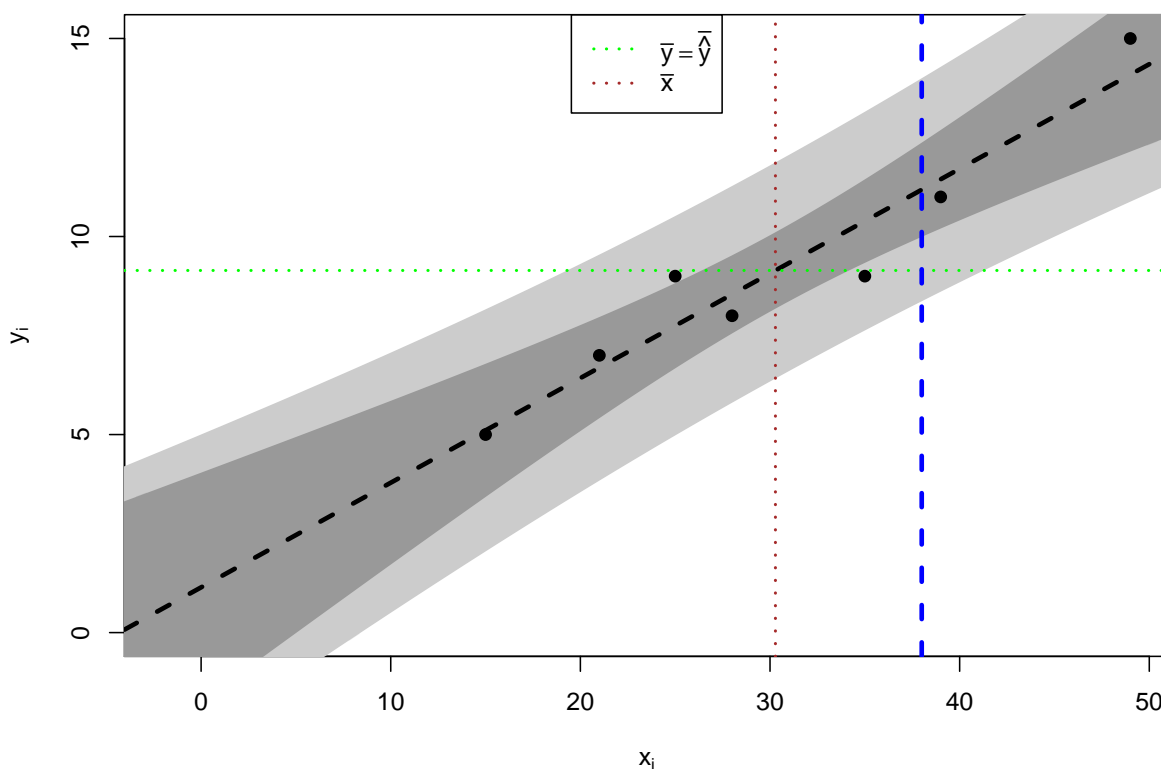
$$\hat{\sigma}_{e_0}^2 = \hat{\sigma}^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n \cdot s_x^2} \right) = 0.9856 \cdot \left(1 + \frac{1}{7} + \frac{(38 - 30.28571)^2}{7 \cdot 114.4901} \right) = 1.1996$$

mit der bereits berechneten Punktprognose $\hat{y}_0 = \widehat{E}(y_0) = 11.1807$ (in 100 €) die zugehörige Intervallprognose für y_0 zur Vertrauenswahrscheinlichkeit 0.95

$$\begin{aligned} & [11.1807 - 2.571 \cdot \sqrt{1.1996}, 11.1807 + 2.571 \cdot \sqrt{1.1996}] \\ & = [8.3648, 13.9966] \text{ (in 100 €)}. \end{aligned}$$

Prognose: Ausgaben in Abhängigkeit vom Einkommen

$\hat{\beta}_1 = 1.14228$, $\hat{\beta}_2 = 0.26417$, $x_0 = 38$, $\hat{y}_0 = 11.1807$, $1 - \alpha = 0.95$



Lineare Modelle mit Statistik-Software R

Beispiel (Ausgaben in Abhängigkeit vom Einkommen)

- Modellschätzung mit aussagekräftiger Zusammenfassung in nur einer Zeile:

```
> summary(lm(y~x))
```

Call:

```
lm(formula = y ~ x)
```

Residuals:

```
      1      2      3      4      5      6      7
-1.3882  0.9134  0.3102 -0.4449 -0.1048 -0.5390  1.2535
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.14225     1.12645   1.014 0.357100
x             0.26417     0.03507   7.533 0.000653 ***
```

Signif. codes:

```
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.9928 on 5 degrees of freedom

Multiple R-squared: 0.919, Adjusted R-squared: 0.9028

F-statistic: 56.74 on 1 and 5 DF, p-value: 0.0006529

Interpretation des Outputs (I)

Residuen, $\hat{\sigma}^2$ und R^2

Residuals:

```
      1      2      3      4      5      6      7
-1.3882  0.9134  0.3102 -0.4449 -0.1048 -0.5390  1.2535
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.14225     1.12645   1.014 0.357100
x             0.26417     0.03507   7.533 0.000653 ***
```

--

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.9928 on 5 degrees of freedom

Multiple R-squared: 0.919, Adjusted R-squared: 0.9028

F-statistic: 56.74 on 1 and 5 DF, p-value: 0.0006529

- Auflistung bzw. Zusammenfassung der Residuen \hat{u}_i**
- Geschätzte Standardabweichung $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$, hier: $\hat{\sigma} = 0.9928 \Rightarrow \hat{\sigma}^2 = 0.9857$**
- Anzahl Freiheitsgrade $n - 2$, hier: $n - 2 = 5 \Rightarrow n = 7$**
- (Multiples) Bestimmtheitsmaß R^2 , hier: $R^2 = 0.919$**

Interpretation des Outputs (II)

Ergebnisse zur Schätzung von β_1 und β_2

Residuals:

	1	2	3	4	5	6	7
Residuals	-1.3882	0.9134	0.3102	-0.4449	-0.1048	-0.5390	1.2535

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.14225	1.12645	1.014	0.357100
x	0.26417	0.03507	7.533	0.000653 ***

--

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9928 on 5 degrees of freedom

Multiple R-squared: 0.919, Adjusted R-squared: 0.9028

F-statistic: 56.74 on 1 and 5 DF, p-value: 0.0006529

- Realisationen von $\hat{\beta}_1, \hat{\beta}_2$, hier: $\hat{\beta}_1 = 1.14225, \hat{\beta}_2 = 0.26417$
- Standardfehler von $\hat{\beta}_1, \hat{\beta}_2$, hier: $\hat{\sigma}_{\hat{\beta}_1} = 1.12645, \hat{\sigma}_{\hat{\beta}_2} = 0.03507$
- t-Statistiken zu Tests auf Signifikanz, hier: zu $\beta_1 : t = 1.014$, zu $\beta_2 : t = 7.533$
- p-Werte zu Tests auf Signifikanz, hier: zu $\beta_1 : p = 0.3571$, zu $\beta_2 : p = 0.000653$

Zusammenhang zwischen p-Werten

zu zweiseitigen und einseitigen Tests bei unter H_0 (um Null) symmetrisch verteilter Teststatistik

- Erinnerung: $t(n)$ - sowie $N(0, 1)$ -Verteilung sind symmetrisch um Null, für die zugehörigen Verteilungsfunktionen F gilt also $F(x) = 1 - F(-x)$ für alle $x \in \mathbb{R}$ und $F(0) = 0.5$, $F(x) < 0.5$ für $x < 0$ sowie $F(x) > 0.5$ für $x > 0$.
- Für die p-Werte p_z der zweiseitigen Tests auf den Mittelwert bei bekannter (Gauß-Test) sowie unbekannter (t -Test) Varianz gilt daher bekanntlich

$$p_z = 2 \cdot \min\{F(x), 1 - F(x)\} = \begin{cases} 2 \cdot F(x) & \text{falls } x < 0 \\ 2 \cdot (1 - F(x)) & \text{falls } x \geq 0 \end{cases},$$

wobei x den realisierten Wert der Teststatistik sowie F die Verteilungsfunktion der Teststatistik unter H_0 bezeichne.

- Für die p-Werte $p_l = F(x)$ zum linksseitigen sowie $p_r = 1 - F(x)$ zum rechtsseitigen Test bei realisierter Teststatistik x gelten demnach die folgenden Zusammenhänge:

$$p_l = \begin{cases} \frac{p_z}{2} & \text{falls } x < 0 \\ 1 - \frac{p_z}{2} & \text{falls } x \geq 0 \end{cases} \quad \text{sowie} \quad p_r = \begin{cases} 1 - \frac{p_z}{2} & \text{falls } x < 0 \\ \frac{p_z}{2} & \text{falls } x \geq 0 \end{cases}.$$

- Somit auch p-Werte zu einseitigen Tests aus R-Output bestimmbar!

Verallgemeinerungen des einfachen linearen Modells

- Zahlreiche Verallgemeinerungen des einfachen linearen Modells möglich.
- Statt einem Regressor mehrere Regressoren \rightsquigarrow multiples Regressionsmodell.
- Statt unabhängiger identisch verteilter Störgrößen (z.B.)
 - ▶ unabhängige Störgrößen mit unterschiedlichen Varianzen,
 - ▶ abhängige (korrelierte) Störgrößen.
- Statt deterministischer Regressoren stochastische Regressoren.
- Statt nur einer Gleichung für einen Regressanden (simultane) Betrachtung mehrerer Regressanden \rightsquigarrow Mehrgleichungsmodelle.
- Über Betrachtung linearer Abhängigkeiten hinaus auch nichtlineare Regressionsmodelle möglich.
- Verallgemeinerungen werden in weiterführenden Vorlesungen diskutiert, insbesondere „Ökonometrie“ (Bachelorstudiengang) und gegebenenfalls „Econometric Methods and Applications“ (Masterstudiengang).