9.2 Multiple Lineare Regression

↪ Mehrere Regressoren : $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i$

↪ Zusammenfassung in Matrizenschreibweise:

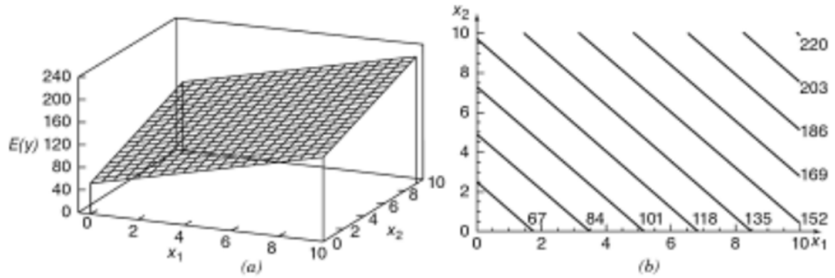
$$
\begin{aligned}
\mathbf{Y} &= (Y_1, \ldots, Y_n)^T \\
&= \beta_0 \mathbf{X}_0 + \beta_1 \mathbf{X}_1 + \ldots + \beta_k \mathbf{X}_k + \varepsilon \\
&= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\
\boldsymbol{\beta} &= (\beta_0, \beta_1, \ldots, \beta_k)^T \\
\boldsymbol{\varepsilon} &= (\varepsilon_1, \ldots, \varepsilon_n)^T \\
\mathbf{X} &= \begin{pmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{pmatrix}, \quad \mathbf{X}_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{X}_j = \begin{pmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{nj} \end{pmatrix}, 1 \leq j \leq k
\end{aligned}
$$

	A	B	C	D	E	F
1	beta	2,000	1,000	0,100		
2	delivery	const	n.prod	distance	delTime	hat_deltime
3	1	1	7	560	16,68	65
4	2	1	3	220	11,5	27
5	3	1	3	340	12,03	39
6	4	1	4	80	14,88	14
7	5	1	6	150	13,75	23
23	21	1	10	140	17,9	26
24	22	1	26	810	52,32	109
25	23	1	9	450	18,75	56
26	24	1	8	635	19,83	73,5
27	25	1	4	150	10,75	21

↪ Umfasst auch den Spezialfall der einfachen linearen Regression.

↪ Modell beschreibt Hyperebene im Raum der erklärenden Variablen.

$$E(Y|X_1, X_2) = 50 + 20X_1 + 7X_2$$

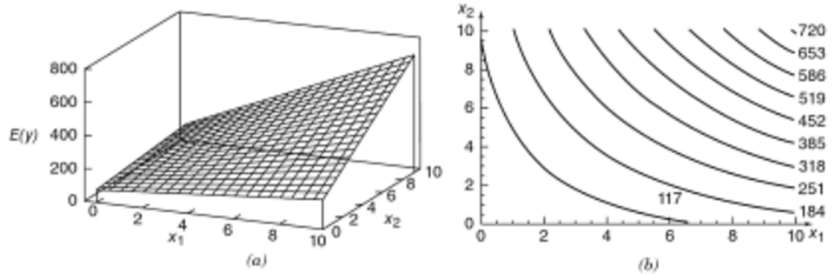


↪ β_i: Änderung in $E(Y)$ bei Erhöhung von X_i um 1 Einheit (ceteris paribus)

↪ Auch polynomielle Zusammenhänge und Interaktionen zwischen Variablen lassen sich abbilden, solange die Funktion linear in den Parametern ist, z.B.

1) $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \varepsilon$
2) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \ldots + \varepsilon$

$$E(Y|X_1, X_2) = 50 + 20 \cdot X_1 + 7 \cdot X_2 + 5 \cdot X_1 X_2$$



Parameterschätzung wieder mit KQ-Methode

Mit $\epsilon_i = Y_i - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})$, d.h. $\boldsymbol{\varepsilon} = \mathrm{Y} - \boldsymbol{X}\boldsymbol{\beta}$

Minimiere in $\beta = (\beta_0, \ldots, \beta_k)$: $K(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$

$$
\begin{aligned}
&= (\mathrm{Y} - \boldsymbol{X}\boldsymbol{\beta})^T (\mathrm{Y} - \mathrm{X}\boldsymbol{\beta}) \\
&= \mathrm{Y}^T\mathrm{Y} - \boldsymbol{\beta}^T\mathrm{X}^T\mathrm{Y} - \mathrm{Y}^T\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathrm{X}^T\mathrm{X}\boldsymbol{\beta} \\
&= \mathrm{Y}^T\mathrm{Y} - \boldsymbol{\beta}^T\mathrm{X}^T\mathrm{Y} - \boldsymbol{\beta}^T\mathrm{X}^T\mathrm{Y} + \boldsymbol{\beta}^T\mathrm{X}^T\mathrm{X}\boldsymbol{\beta} \\
&= \mathrm{Y}^T\mathrm{Y} - 2\boldsymbol{\beta}^T\mathrm{X}^T\mathrm{Y} + \boldsymbol{\beta}^T\mathrm{X}^T\mathrm{X}\boldsymbol{\beta}
\end{aligned}
$$

Das führt auf die sog. Normalgleichungen:

$$\frac{\partial K}{\partial \boldsymbol{\beta}} = -2\mathrm{X}^T\mathrm{Y} + 2\mathrm{X}^T\mathrm{X}\boldsymbol{\beta} \stackrel{!}{=} \bar{0} \quad \Leftrightarrow \quad \mathrm{X}^T\mathrm{X}\boldsymbol{\beta} = \mathrm{X}^T\mathrm{Y}$$

$\Rightarrow$ KQ-Schätzer ist $\hat{\boldsymbol{\beta}} = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathrm{Y}$, falls $(\mathrm{X}^T\mathrm{X})^{-1}$ existiert.

Optimierung mit Excel (Matrixfunktionen)

	A	B	C	D	E	F	G	H	I	J
1	beta	2,341	1,616	0,014			R^2	0,9596		
2	delivery	const	n.prod	distance	delTime	hat_deltime	residuals	SS_Res	SS_T	SS_R
3	1	1	7	560	16,68	21,708	-5,0281	25,282	32,536	0,457
4	2	1	3	220	11,5	10,354	1,1464	1,314	118,461	144,730
5	3	1	3	340	12,03	12,080	-0,0498	0,002	107,205	106,177
6	4	1	4	80	14,88	9,956	4,9244	24,249	56,310	154,464
7	5	1	6	150	13,75	14,194	-0,4444	0,197	74,546	67,070
23	21	1	10	140	17,9	20,514	-2,6142	6,834	20,106	3,496
24	22	1	26	810	52,32	56,007	-3,6865	13,590	896,164	1130,474
25	23	1	9	450	18,75	23,358	-4,6076	21,230	13,206	0,948
26	24	1	8	635	19,83	24,403	-4,5729	20,911	6,523	4,076
27	25	1	4	150	10,75	10,963	-0,2126	0,045	135,350	130,449
28	sum	25	219	10232	559,6	559,6	0,0000	233,732	5784,543	5550,811
29		X^T*X			X^T*Y					
30	normal	25	219	10232	559,6					
31	equations	219	3055	133899	7375,44					
32		10232	133899	6725688	337071,69					

B30:E32 =MMULT(MTRANS(B3:D27);B3:E27)

B1:D1 =MTRANS(MMULT(MINV(B30:D32);E30:E32))

F3:F27 =MMULT(B3:D27;MTRANS(B1:D1))

Lieferzeitenanalyse für Getränke-/Süßigkeitenautomaten

Y: Lieferzeit (Minuten), `delTime`

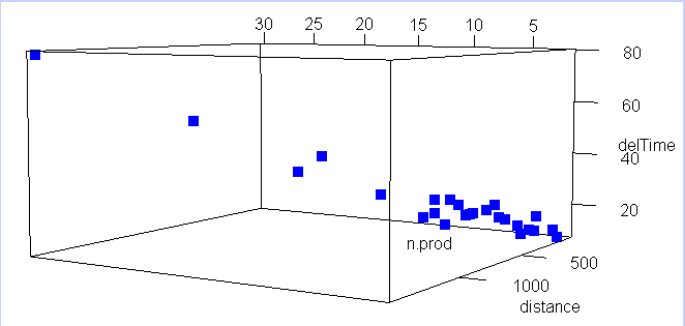
X_1: Anzahl nachzufüllender Produkte, `n.prod`

X_2: Fußweg (ft), `distance`

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

	n.prod	distance	delTime		n.prod	distance	delTime
1	7	560	16.68	14	6	462	19.75
2	3	220	11.50	15	9	448	24.00
3	3	340	12.03	16	10	776	29.00
4	4	80	14.88	17	6	200	15.35
5	6	150	13.75	18	7	132	19.00
6	7	330	18.11	19	3	36	9.50
7	2	110	8.00	20	17	770	35.10
8	7	210	17.83	21	10	140	17.90
9	30	1460	79.24	22	26	810	52.32
10	5	605	21.50	23	9	450	18.75
11	16	688	40.33	24	8	635	19.83
12	10	215	21.00	25	4	150	10.75
13	4	255	13.50				

$$\mathrm{X} = \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ 1 & 3 & 340 \\ \vdots & \vdots & \vdots \\ 1 & 8 & 635 \\ 1 & 4 & 150 \end{bmatrix}, \quad \mathrm{Y} = \begin{bmatrix} 16.68 \\ 11.50 \\ 12.03 \\ \vdots \\ 19.83 \\ 10.75 \end{bmatrix}$$



$$
\begin{aligned}
X^T X &= \begin{bmatrix} 1 & 1 & \dots & 1 \\ 7 & 3 & \dots & 4 \\ 560 & 220 & \dots & 150 \end{bmatrix} \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 150 \end{bmatrix} \\
&= \begin{bmatrix} 25 & 219 & 10.232 \\ 219 & 3.055 & 133.899 \\ 10.232 & 133.899 & 6725.688 \end{bmatrix}
\end{aligned}
$$

$$
X^T Y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 7 & 3 & \dots & 4 \\ 560 & 220 & \dots & 150 \end{bmatrix} \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 10.75 \end{bmatrix} = \begin{bmatrix} 559.60 \\ 7375.44 \\ 337072 \end{bmatrix}
$$

$$X^T X = \begin{bmatrix} 25 & 219 & 10.232 \\ 219 & 3.055 & 133.899 \\ 10.232 & 133.899 & 6725.688 \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} 559.60 \\ 7375.44 \\ 337072 \end{bmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 2.341 \\ 1.616 \\ 0.014 \end{bmatrix}$$

$$\hat{y} = 2.341 + 1.616 \cdot x_1 + 0.014 \cdot x_2$$

Somit beträgt die durchschnittliche Lieferzeit 2.341 Minuten und verlängert sich im Schnitt mit jedem nachzufüllendem Produkt um 1.616 Minuten und mit jedem zusätzlichen Fuß (ft) Wegstrecke um 0.014 Minuten.

R-Code und zugehöriger Output

```
45
46  require(robustbase)
47
48  data(delivery)
49
50  summary(lm.deli <- lm(delTime ~ n.prod+distance, data = delivery))
51  summary(lm.deli <- lm(delTime ~ ., data = delivery))
```

48:15 (Untitled)

Console ~/

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.341231   1.096730   2.135 0.044170 *
n.prod      1.615907   0.170735   9.464 3.25e-09 ***
distance    0.014385   0.003613   3.981 0.000631 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared:  0.9596, Adjusted R-squared:  0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

Wie ist der Output zu interpretieren? (siehe Hypothesentests)

Umsatz an einer Tankstelle

Der Inhaber einer Kette von 5 freien Tankstellen möchte wissen, ob und wie sich der Tagesgewinn aus seinen Umsätzen erklären lässt. Hierbei unterteilt er diese in Kraftstoff (K) und in Sonstige (S) Umsätze. Ergebnisse eines Tages: (in €):

Umsatz K x_1	6 000	2 500	8 500	6 500	9 500
Umsatz S x_2	7 000	6 500	3 000	7 000	7 500
Gewinn y	3 000	4 000	2 000	3 000	3 500

Erklärn

~~Modellieren~~ Sie den Gewinn des Tankstellenbetreibers unter gemeinsamer Berücksichtigung beider Umsatzkategorien!

Erklärungsmodell: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U, \quad \|Y - X\beta\|^2 \stackrel{!}{=} \min_\beta$

$$X = \begin{pmatrix} 1 & 6\,000 & 7\,000 \\ 1 & 2\,500 & 6\,500 \\ 1 & 8\,500 & 3\,000 \\ 1 & 6\,500 & 7\,000 \\ 1 & 9\,500 & 7\,500 \end{pmatrix}, \qquad Y = \begin{pmatrix} 3\,000 \\ 4\,000 \\ 2\,000 \\ 3\,000 \\ 3\,500 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 5 & 33\,000 & 31\,000 \\ 33\,000 & 247\,000\,000 & 200\,500\,000 \\ 31\,000 & 200\,500\,000 & 205\,500\,000 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 5{,}61 & 3{,}05 \cdot 10^{-4} & 5{,}6 \cdot 10^{-4} \\ & 3{,}58 \cdot 10^{-8} & 1{,}1 \cdot 10^{-8} \\ & & 7{,}86 \cdot 10^{-6} \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 15\,500 \\ 97\,750\,000 \\ 100\,250\,000 \end{pmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} \cdot X^T Y = \begin{pmatrix} 2\,161.889 \\ -0.117 \\ 0.276 \end{pmatrix}$$

$$\Rightarrow \quad \hat{y}_i = 2\,161.889 - 0.117 \cdot x_{i,1} + 0.276 \cdot x_{i,2}$$

Eigenschaften der KQ-Schätzer

Es gilt:

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}\right] \\
&= E\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})\right] \\
&= E\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}+(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\right] \\
&= E\left[\mathbf{I}\boldsymbol{\beta}\right]+(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[\boldsymbol{\varepsilon}] \\
&= \boldsymbol{\beta}+0=\boldsymbol{\beta}
\end{aligned}
$$

$\Rightarrow \hat{\boldsymbol{\beta}}$ ist ein erwartungstreuer Schätzer für $\boldsymbol{\beta}$

$\hat{\boldsymbol{\beta}}$ ist bester linearer erwartungstreuer Schätzer, d.h. $\hat{\boldsymbol{\beta}}$ hat die kleinste Varianz in der Klasse aller erwartungstreuen Schätzer, die Linearkombinationen der erklärenden Variablen darstellen (ohne Beweis, [4]).

Eigenschaften der KQ-Schätzer (Forts.)

Weiterhin gilt für die Kovarianzmatrix:

$$
\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= var\left[(\mathsf{X}^T\mathsf{X})^{-1}\mathsf{X}^T\mathsf{Y}\right] \\
&= (\mathsf{X}^T\mathsf{X})^{-1}\mathsf{X}^T var(\mathsf{Y})\left[(\mathsf{X}^T\mathsf{X})^{-1}\mathsf{X}^T\right]^T \\
&= (\mathsf{X}^T\mathsf{X})^{-1}\mathsf{X}^T \cdot (\sigma^2\mathsf{I}) \cdot \left[(\mathsf{X}^T\mathsf{X})^{-1}\mathsf{X}^T\right]^T \\
&= \sigma^2(\mathsf{X}^T\mathsf{X})^{-1}\mathsf{X}^T\mathsf{X}(\mathsf{X}^T\mathsf{X})^{-1} \\
&= \sigma^2(\mathsf{X}^T\mathsf{X})^{-1} = \sigma^2 C, \qquad \text{mit } C = (\mathsf{X}^T\mathsf{X})^{-1}, \text{ d.h.}
\end{aligned}
$$

$$
\text{d.h. } var(\hat{\beta}_j) = \sigma^2 C_{jj}
$$

$$
cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 C_{ij}
$$

$\rightarrow$ Schätzung von σ?

Hat-Matrix $\mathrm{H} = \mathrm{X}(\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T \in \mathbb{R}^{n \times n}$

H projiziert Y in vorhergesagte Werte: $\hat{\mathrm{Y}} = \mathrm{X}\hat{\boldsymbol{\beta}} = \mathrm{X}(\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathrm{Y} = \mathrm{HY}$,

(Vulgo: „Die Hat-Matrix setzt (der Zielgröße) y den Hut auf.“)

Eigenschaften der Hat-Matrix

- $\mathrm{H}^2 = \mathrm{H}$ und $(\mathrm{I} - \mathrm{H})^2 = \mathrm{I} - \mathrm{H}$ (Übung)
- $tr(\mathrm{H}) = k + 1$ (die Spur $tr(H)$ ist die Summe der Diagonalelemente von H)
- $\varepsilon = \mathrm{Y} - \hat{\boldsymbol{Y}} = \mathrm{Y} - \mathrm{X}\hat{\boldsymbol{\beta}} = \mathrm{Y} - \mathrm{HY} = (\mathrm{I} - \mathrm{H})\mathrm{Y}$

Dazu Eigenschaften der Spur:

- $\hookrightarrow$ $tr(A \pm B) = tr(A) \pm tr(B)$ falls A, B quadratisch sind.
- $\hookrightarrow$ $tr(AB) = tr(BA)$, falls beide Matrixprodukte gebildet werden können und AB, BA quadratisch sind.
- $\hookrightarrow$ $tr(\mathrm{H}) = tr(\mathrm{X}(\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T) = tr(\mathrm{X}^T\mathrm{X}(\mathrm{X}^T\mathrm{X})^{-1}) = tr(I_{k+1}) = k + 1$

Übung: Zeigen Sie:

$$H^2 = H \text{ sowie } (I - H)^2 = I - H.$$

Es gilt:

$$\begin{aligned} H^2 &= X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T \\ &= X\underbrace{(X^TX)^{-1}(X^TX)}_{=I}(X^TX)^{-1}X^T \\ &= X(X^TX)^{-1}X^T = H \end{aligned}$$

Analog gilt:

$$\begin{aligned} (I - H)^2 &= \underbrace{I^2}_{=I} - 2\underbrace{IH}_{=H} + \underbrace{H^2}_{=H} \\ &= I - 2H + H = I - H. \end{aligned}$$

Schätzung von σ^2

$\hat{\sigma}^2 = \frac{SS_{res}}{n-p}$ (mit $p = k+1$) ist ein erwartungstreuer Schätzer für σ^2.

$$
\begin{aligned}
SS_{res} &= \varepsilon^T \varepsilon = ((\mathrm{I}-\mathrm{H})\mathrm{Y})^T(\mathrm{I}-\mathrm{H})\mathrm{Y} = \mathrm{Y}^T(\mathrm{I}-\mathrm{H})\mathrm{Y} = \mathrm{Y}^T\mathrm{Y} - \hat{\beta}^T\mathrm{X}^T\mathrm{Y} \\
E(SS_{res}) &= E(\mathrm{Y}^T(\mathrm{I}-\mathrm{H})\mathrm{Y}) \\
&= E((\mathrm{Y}-E(\mathrm{Y}))^T(\mathrm{I}-\mathrm{H})(\mathrm{Y}-E(\mathrm{Y})) \\
&= E(tr((\mathrm{Y}-E(\mathrm{Y}))^T(\mathrm{I}-\mathrm{H})(\mathrm{Y}-E(\mathrm{Y})))) \\
&= E(tr((\mathrm{I}-\mathrm{H})(\mathrm{Y}-E(\mathrm{Y}))(\mathrm{Y}-E(\mathrm{Y}))^T) \\
&= tr((\mathrm{I}-\mathrm{H}) \cdot E((\mathrm{Y}-E(\mathrm{Y}))(\mathrm{Y}-E(\mathrm{Y}))^T)) \\
&= tr((\mathrm{I}-\mathrm{H}) \cdot Cov(Y)) \\
&= tr((\mathrm{I}-\mathrm{H}) \cdot \sigma^2 I_n) \\
&= (n-k-1)\sigma^2
\end{aligned}
$$

Lieferzeitenanalyse

Schätzung der Varianz σ^2 der Residuen.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.341231   1.096730   2.135 0.044170 *
n.prod      1.615907   0.170735   9.464 3.25e-09 ***
distance    0.014385   0.003613   3.981 0.000631 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared:  0.9596, Adjusted R-squared:  0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

$$SS_{res} = \mathbf{Y}^T\mathbf{Y} - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{Y}$$

$$\mathbf{Y}^T\mathbf{Y} = \sum_{i=1}^{25} y_i^2 = 18\,310.63$$

$$\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{Y} = [2.341 \quad 1.616 \quad 0.014] \begin{bmatrix} 559.60 \\ 7\,375.44 \\ 337\,072 \end{bmatrix} = 18\,076.9$$

$$SS_{res} = 18\,310.63 - 18\,076.9 = 233.73$$

$$\hat{\sigma}^2 = \frac{SS_{res}}{n-p} = \frac{233.73}{25-3} = 10.62, \quad \hat{\sigma} = 3.259$$

Umsatz an einer Tankstelle

Schätzen Sie die Varianz σ^2 der Residuen im Tankstellen-Beispiel.

$$
\begin{aligned}
\mathbf{Y}^T\mathbf{Y} &= \sum_{i=1}^{5} y_i^2 = 50\,250\,000 \\
\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{Y} &= [2\,161.889 \quad -0.117 \quad 0.276] \cdot \begin{bmatrix} 15\,500 \\ 97\,750\,000 \\ 100\,250\,000 \end{bmatrix} \\
&= 49\,727\,846 \\
SS_{res} &= 50\,250\,000 - 49\,727\,846 = 522\,153.8 \\
\hat{\sigma}^2 &= \frac{SS_{res}}{n-p} = \frac{522153.8}{5-3} = 261\,076.9, \quad \hat{\sigma} = 510.96
\end{aligned}
$$

Maximum-Likelihood-Schätzer

- die KQ-Schätzer im multiplen linearen Regressionsmodell entsprechen den Maximum-Likelihood-Schätzern bei Gültigkeit der Modellannahmen

$$
\begin{aligned}
\mathsf{Y} &= \mathsf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\
\boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \sigma^2 \mathsf{I}), \quad \text{d.h.} \\
f(\varepsilon_i) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right)
\end{aligned}
$$

$$
\Rightarrow L(\boldsymbol{\varepsilon}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} f(\varepsilon_i) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right)
$$

$$
L(\mathsf{Y}, \mathsf{X}, \boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}(\mathsf{Y} - \mathsf{X}\boldsymbol{\beta})^T(\mathsf{Y} - \mathsf{X}\boldsymbol{\beta})\right)
$$

Log-Likelihood:

$$\begin{aligned}
\ln L(\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) &= \ln\left(\frac{1}{(2\pi)^{n/2}\sigma^n}\right) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= 0 - \left(\frac{n}{2}\ln(2\pi) + n\ln(\sigma)\right) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})
\end{aligned}$$

$\Rightarrow$ Für fixes σ wird $\ln L(\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2)$ maximal für minimales $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$

$\Rightarrow$ ML-Schätzer ist $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ (s.o.)

ML-Schätzer für σ^2:

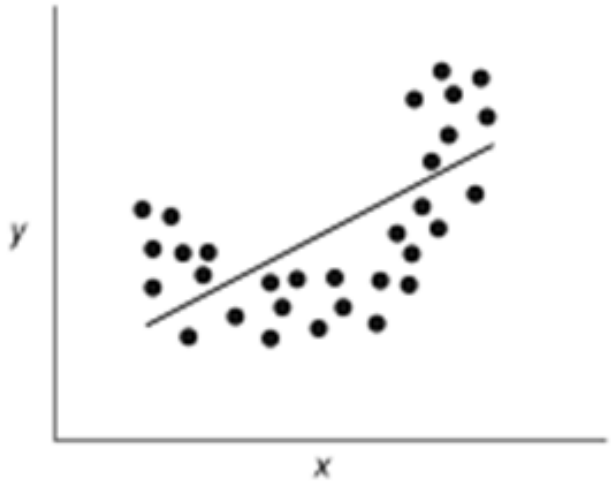
$$\begin{aligned}
\frac{\partial \ln(Y, X, \hat{\beta}, \sigma^2)}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}(Y - X\hat{\beta})^T(Y - X\hat{\beta}) \stackrel{!}{=} 0 \\
\Leftrightarrow \quad \frac{n}{\sigma} &= \frac{1}{\sigma^3}(Y - X\hat{\beta})^T(Y - X\hat{\beta}) \\
\Leftrightarrow \quad n\sigma^2 &= (Y - X\hat{\beta})^T(Y - X\hat{\beta})
\end{aligned}$$

$$\Rightarrow \quad \hat{\sigma}^2 = \frac{(Y - X\hat{\beta})^T(Y - X\hat{\beta})}{n} = \frac{SS_{res}}{n}$$

- Der ML-Schätzer für σ^2 ist nicht erwartungstreu.

9.3 Hypothesentests in der linearen Regression

- Wie ist die Modellgüte? Ist das Modell geeignet, um die Zusammenhänge zu beschreiben?
- Von welcher Bedeutung sind die erklärenden Variablen?
- Überprüfung anhand statistischer Hypothesentests
- Annahmen: $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$



Varianzzerlegung

$$SS_T = SS_R + SS_{res} \quad \text{d.h.}$$

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

$$\mathrm{Y}^T\mathrm{Y} - n\bar{\mathrm{Y}}^2 = \hat{\beta}^T\mathrm{X}^T\mathrm{Y} - n\bar{\mathrm{Y}}^2 + \mathrm{Y}^T\mathrm{Y} - \hat{\beta}^T\mathrm{X}^T\mathrm{Y}$$

SS_T : Gesamtvariabilität (Total sum of squares)

SS_R : Variabilität der Regression (sum of sq. due to regression)

SS_{res} : Variabilität der Residuen (residual sum of squares)

Die Varianzzerlegung ist (nur) korrekt, wenn der konstante Effekt β_0 im Modell ist (Eins-Spalte in X)

Beispiel mit Excel (fortges.)

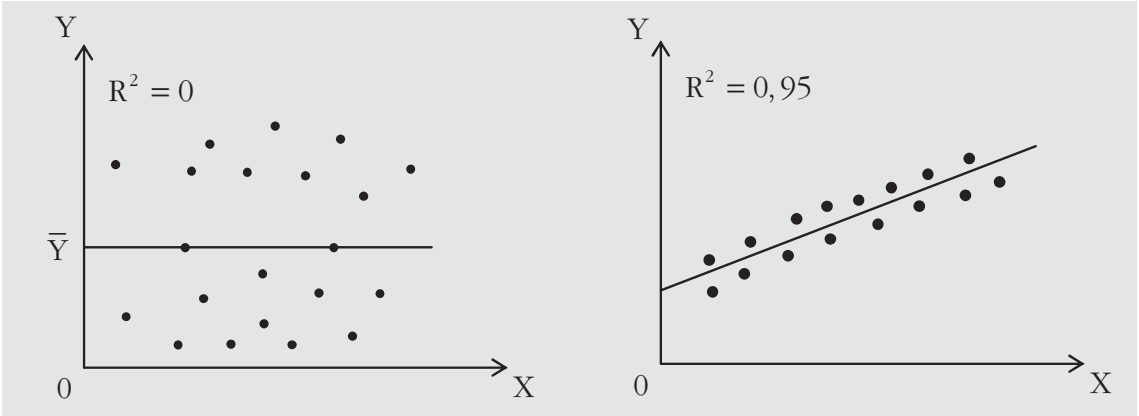
	A	B	C	D	E	F	G	H	I	J
1	beta	2,341	1,616	0,014			R^2	0,9596		
2	delivery	const	n.prod	distance	delTime	hat_deltime	residuals	SS_Res	SS_T	SS_R
3	1	1	7	560	16,68	21,7081	-5,02805188	25,281	32,536	0,457
4	2	1	3	220	11,5	10,3536	1,14636478	1,314	118,461	144,730
5	3	1	3	340	12,03	12,0798	-0,04978497	0,002	107,205	106,177
6	4	1	4	80	14,88	9,95571	4,9242914	24,249	56,310	154,462
7	5	1	6	150	13,75	14,1945	-0,44445881	0,198	74,546	67,069
23	21	1	10	140	17,9	20,5143	-2,61427203	6,834	20,106	3,496
24	22	1	26	810	52,32	56,0066	-3,68657762	13,591	896,164	1130,478
25	23	1	9	450	18,75	23,3576	-4,60757746	21,230	13,206	0,948
26	24	1	8	635	19,83	24,4028	-4,57281023	20,911	6,523	4,076
27	25	1	4	150	10,75	10,9626	-0,21262929	0,045	135,350	130,448
28	sum	25	219	10232	559,6	559,6	-0,00044424	233,732	5784,543	5550,795

g

Bestimmtheitsmaß R^2

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{res}}{SS_T}, 0 \leq R^2 \leq 1$$

- die Fähigkeit der Regressionsgeraden, die Variabilität in Y zu erklären, ist umso besser, je kleiner die Residualvariabilität im Vergleich zur Gesamtvariabilität ist.



- $R^2 = 0$: Regressionsgerade erklärt nicht besser als der Mittelwert.
- Je größer R^2, desto besser ist der "Fit" der Regressionsgeraden.

Lieferzeitenanalyse

```
58  summary(lm.deli <- lm(delTime ~ n.prod+distance, data = delivery))
```

58:67 (Untitled)

Console C:/DatenM/Repositories/Datenanalyse/R/

```
Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared:  0.9596, Adjusted R-squared:  0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

```
53  summary(lm.deli <- lm(delTime ~ n.prod, data = delivery))
```

53:58 (Untitled)

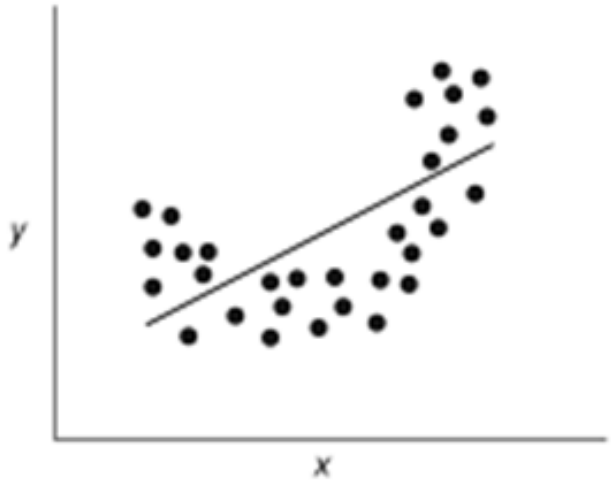
Console C:/DatenM/Repositories/Datenanalyse/R/

```
Residual standard error: 4.181 on 23 degrees of freedom
Multiple R-squared:  0.9305, Adjusted R-squared:  0.9275
F-statistic: 307.8 on 1 and 23 DF,  p-value: 8.22e-15
```

- Vorsicht bei der Interpretation von R^2
- R^2 kann sich nicht verschlechtern bei Hinzunahme weiterer erklärender Variablen

Varianzzerlegung und Bestimmtheitsmaß

- Mit weiteren erklärenden Variablen lässt sich R^2 künstlich erhöhen
- Eine Erhöhung von k führt zu einer Verringerung der Freiheitsgrade, d.h. zu ungenaueren Koeffizientenschätzungen
- R^2 kann nicht die Angemessenheit des linearen Modelles beurteilen, d.h. es können nichtlineare Zusammenhänge vorliegen



Adjustiertes Bestimmtheitsmaß $\tilde{R}^2$

- Verwende adjustiertes Bestimmtheitsmaß $\tilde{R}^2$

$$R^2 = 1 - \frac{SS_{res}}{SS_T} \qquad \rightsquigarrow \qquad \tilde{R}^2 = 1 - \frac{SS_{res}/(n-k-1)}{SS_T/(n-1)}$$

- „Bestrafe“ die Hinzunahme weiterer erklärender Variablen.
- Verwendung von erwartungstreuen Schätzern für die wahren Variabilitätswerte.

Lieferzeitenanalyse

```
52
53 summary(lm.deli <- lm(delTime ~ n.prod, data = delivery))
```

53:58 (Untitled)

Console C:/DatenM/Repositories/Datenanalyse/R/

```
Residual standard error: 4.181 on 23 degrees of freedom
Multiple R-squared:  0.9305, Adjusted R-squared:  0.9275
F-statistic: 307.8 on 1 and 23 DF,  p-value: 8.22e-15
```

```
58 summary(lm.deli <- lm(delTime ~ n.prod+distance, data = delivery))
```

58:67 (Untitled)

Console C:/DatenM/Repositories/Datenanalyse/R/

```
Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared:  0.9596, Adjusted R-squared:  0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

- Hier verbessert das Hinzufügen der zweiten Variablen die Erklärung der Gesamtvariabilität tatsächlich (allerdings nur geringfügig).

Umsatz an einer Tankstelle

Prüfen Sie die Modellgüte im Tankstellen-Beispiel ($\hat{\beta} = (2\,161.9, -0.117, 0.276)^T$)

Umsatz K x_1	6 000	2 500	8 500	6 500	9 500
Umsatz S x_2	7 000	6 500	3 000	7 000	7 500
Gewinn y	3 000	4 000	2 000	3 000	3 500

$$
\begin{aligned}
SS_T &= \mathrm{Y}^T\mathrm{Y} - \frac{1}{n}\left(\sum_{i=1}^{5} y_i\right)^2 = 2\,200\,000 \\
SS_{res} &= \mathrm{Y}^T\mathrm{Y} - \hat{\beta}^T\mathrm{X}^T\mathrm{Y} = 522\,153.8 \\
R^2 &= 1 - \frac{SS_{res}}{SS_T} = 1 - \frac{522\,153.8}{2\,200\,000} = 0.763 \\
\tilde{R}^2 &= 1 - \frac{SS_{res}/(5-2-1)}{SS_T/(5-1)} = 1 - \frac{522\,153.8/2}{2\,200\,000/4} = 0.525
\end{aligned}
$$

F-Test auf Modellgüte

Besteht ein linearer Zusammenhang zwischen der abhängigen Variablen und mindestens einer der erklärenden Variablen?

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0 \quad \text{für mind. ein } j$$

Teststatistik/-entscheidung: Es gilt unter H_0

$$\begin{aligned} SS_T &= SS_R + SS_{res} \quad \text{und } SS_R, SS_{res} \text{ sind st.u.} \\ SS_R/\sigma^2 &\sim \chi^2_k \\ SS_{res}/\sigma^2 &\sim \chi^2_{n-k-1} \\ \Rightarrow F_0 &= \frac{SS_R/k}{SS_{res}/(n-k-1)} = \frac{n-k-1}{k} \cdot \frac{R^2}{1-R^2} \sim \mathcal{F}_{k,n-k-1} \end{aligned}$$

H_0 wird abgelehnt, falls $F_0 > \mathcal{F}_{1-\alpha,k,n-k-1}$

	A	B	C	D	E	F	G	H	I	J
1	beta	2,341	1,616	0,014			R^2	0,9596		
2	delivery	const	n.prod	distance	delTime	hat_deltime	residuals	SS_Res	SS_T	SS_R
3	1	1	7	560	16,68	21,7081	-5,02805188	25,281	32,536	0,457
4	2	1	3	220	11,5	10,3536	1,14636478	1,314	118,461	144,730
5	3	1	3	340	12,03	12,0798	-0,04978497	0,002	107,205	106,177
6	4	1	4	80	14,88	9,95571	4,9242914	24,249	56,310	154,462
7	5	1	6	150	13,75	14,1945	-0,44445881	0,198	74,546	67,069
23	21	1	10	140	17,9	20,5143	-2,61427203	6,834	20,106	3,496
24	22	1	26	810	52,32	56,0066	-3,68657762	13,591	896,164	1130,478
25	23	1	9	450	18,75	23,3576	-4,60757746	21,230	13,206	0,948
26	24	1	8	635	19,83	24,4028	-4,57281023	20,911	6,523	4,076
27	25	1	4	150	10,75	10,9626	-0,21262929	0,045	135,350	130,448
28	sum	25	219	10232	559,6	559,6	-0,00044424	233,732	5784,543	5550,795

$$F_0 = \frac{SS_R/k}{SS_{res}(n-k-1)} = \frac{5550.8/2}{233.732/(24-2-1)} \approx 261.2 \approx \frac{22}{2}\frac{0.9596}{1-0.9596} = \frac{n-k-1}{k}\frac{R^2}{1-R^2}$$

```
58  summary(lm.deli <- lm(delTime ~ n.prod+distance, data = delivery))
```

58:67 (Untitled)

Console C:/DatenM/Repositories/Datenanalyse/R/

```
Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared:  0.9596, Adjusted R-squared:  0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

Umsatz an einer Tankstelle

Bestimmen Sie die F-Statistik, sowie den zugehörigen p-Wert zum Tankstellen-Beispiel. Entscheiden Sie anhand Ihrer Ergebnisse, ob ein (linearer) Zusammenhang zwischen (mind.) einem der Umsätze und dem Gewinn nachweisbar ist.

$$SS_R = SS_T - SS_{res} = 2\,200\,000 - 522\,154 = 1\,677\,846$$

$$\Rightarrow F_0 = \frac{SS_R/k}{SS_{res}/(n-k-1)} = \frac{1\,677\,846/2}{522\,154/(5-2-1)} = 3.213$$

Alternative Berechnung von F_0:

$$F_0 = \frac{n-k-1}{k} \cdot \frac{R^2}{1-R^2} = \frac{5-2-1}{2} \cdot \frac{0.7626574}{1-0.7626574} = 3.21$$

Zugehöriger p-Wert: `1 - pf(3.213, 2, 5 - 2 - 1)= 0.237`
Es konnte somit nicht nachgewiesen werden, dass mindestens eine der beiden Umsatz-Merkmale einen (linearen) Einfluss auf den Gewinn hat.

F-Test auf Einfluss von Parametergruppen

Hat ein - festgelegtes - Set von Variablen einen Einfluss auf die abhängige Variable?
Konkret mit $1 \leq i_1 < \cdots < i_m \leq k$:

$$H_0 : \beta_{i_1} = \beta_{i_2} = \cdots = \beta_{i_m} = 0$$

Teststatistik / Testentscheidung:

- Unter H_0 entfallen m Parameter im Modell, d.h. $\mathrm{Y} = \mathrm{X}_H \beta_H + \varepsilon$, dabei entstehen β_H bzw. X_H durch Streichung der Einträge bzw. Spalten $i_1 + 1, \ldots, i_m + 1$ aus β bzw. X.
  Zugehörige KQ-Schätzung

$$\hat{\beta}_H = (\mathrm{X}_H^T \mathrm{X}_H)^{-1} \mathrm{X}_H^T \mathrm{Y}, \qquad \hat{\mathrm{Y}}_H = \mathrm{X}_H \hat{\beta}_H$$

- Teststatistik $V = \dfrac{\|\hat{\mathrm{Y}} - \hat{\mathrm{Y}}_H\|^2 / m}{SS_{res}/(n-k-1)} \sim F = \mathcal{F}_{m,n-k-1}$
- H_0 wird abgelehnt, falls $V > \mathcal{F}_{1-\alpha,m,n-k-1}$ (p-value $1 - F(v)$)

Anwendung: Tests für kategorielle Einflussfaktoren (s.u.)

Tests für die Regressionskoeffizienten

- *Welche* Regressionskoeffizienten sind wichtig?

$$H_0 : \beta_j = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0$$

Prinzipiell ein Spezialfall des Parametergruppentests (F-Test), aber gleichwertig als t-Test durchführbar.

Teststatistik / Testentscheidung:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}_{jj}}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

- H_0 wird abgelehnt, falls $|t_0| > t_{1-\alpha/2, n-k-1}$
- Test beurteilt den Beitrag von β_j bedingt auf das Vorliegen der anderen Koeffizienten

```
45
46  require(robustbase)
47
48  data(delivery)
49
50  summary(lm.deli <- lm(delTime ~ n.prod+distance, data = delivery))
51  summary(lm.deli <- lm(delTime ~ ., data = delivery))
```

48:15 (Untitled)

Console ~/

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.341231   1.096730   2.135 0.044170 *
n.prod      1.615907   0.170735   9.464 3.25e-09 ***
distance    0.014385   0.003613   3.981 0.000631 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared:  0.9596, Adjusted R-squared:  0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

Umsatz an einer Tankstelle

Welche der Koeffizienten $\hat{\boldsymbol{\beta}} = \begin{pmatrix} 2\,161.889 & -0.117 & 0.276 \end{pmatrix}^T$ haben einen signifikanten Einfluss ($\alpha = 5\%$) auf das Modell?

Kritischer Wert: $t_{1-\alpha/2,n-k-1} = 4.303$ Teststatistiken:

$$t_{\beta_0} = \frac{\hat{\beta}_0}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^T\mathbf{X})^{-1}_{11}}} = \frac{2\,161.889}{\sqrt{261\,076.9 \cdot 5.68}} = 1.775$$

$$t_{\beta_1} = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^T\mathbf{X})^{-1}_{22}}} = \frac{-0.117}{\sqrt{261\,076.9 \cdot 3.58 \cdot 10^{-8}}} = -1.211$$

$$t_{\beta_2} = \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^T\mathbf{X})^{-1}_{33}}} = \frac{0.276}{\sqrt{261\,076.9 \cdot 7.86 \cdot 10^{-8}}} = 1.926$$

Da jeweils $|t_{\beta_j}| \leq t_{1-\alpha/2,n-k-1}$ gilt, wird die Nullhypothese $\beta_j = 0$ für jeden der drei Koeffizienten beibehalten.